



HAL
open science

Éléments de Neurogéométrie, II

Jean Petitot

► **To cite this version:**

| Jean Petitot. Éléments de Neurogéométrie, II. A paraître. hal-04722501

HAL Id: hal-04722501

<https://hal.science/hal-04722501v1>

Submitted on 5 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

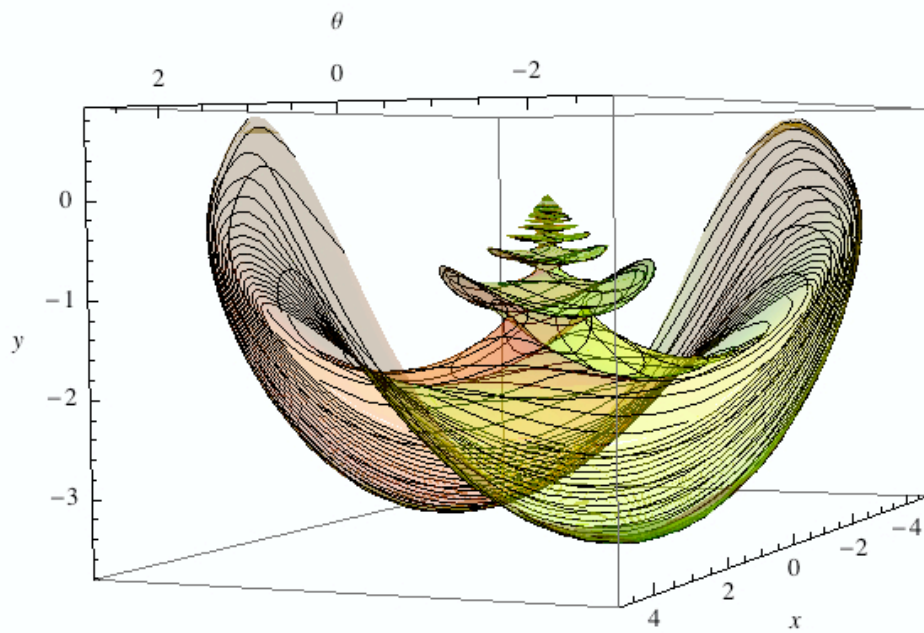
Copyright

Jean Petitot



ÉLÉMENTS DE NEUROGÉOMÉTRIE

Volume 2



2024

Image de couverture :

Une partie du front d'onde sous-riemannien de rayon $3\pi/2$ du groupe de Lie $SE(2)$ des déplacements du plan.

Aux étoiles qui ont guidé cette
circumnavigation :

René Thom, David Mumford,
Daniel Bennequin pour la géométrie
des formes,

William Hoffman, Jan Koenderink
pour la géométrie de la perception,

Andrei Agrachev, Richard
Montgomery pour la géométrie
sous-riemannienne,

Michel Imbert, Alain Berthoz,
Yves Frégnac pour les neurosciences.

Je remercie Giovanna Citti et Alessandro Sarti qui ont
considérablement enrichi les modèles initiaux de la neurogéométrie
ainsi qu'Ugo Boscain, Jean-Paul Gauthier, Dario Prandi et Yuri
Sachkov qui en ont beaucoup approfondi certains aspects théoriques.

Je remercie les personnalités, collègues et amis qui, à des titres divers,
ont été des compagnons de route et/ou des inspirateurs de certains
épisodes de ce périple : Dmitri Alekseevsky, Frédéric Barbaresco,
Davide Barbieri, Davide Barilari, André Bellaïche, Henri Berestycki,
Michael Berry, Michel Bitbol, Paul Bourguin, Paul Bressloff, Marius
Buliga, Marc Chaperon, Alain Chenciner, Jack Cowan, Remco Duits,
Jean-Pierre Dupuy, Olivier Faugeras, Sara Franceschelli, Michael
Friedman, Vasiliki Liantou, Giuseppe Longo, Jean Lorenceau, Stéphane
Mallat, Matilde Marcolli, Joël Merker, Jean-Michel Morel, Jacques
Ninio, Pawel Nurowski, Marco Panza, Jean-Michel Roy, Thomas
Ryckman, Jean-Michel Salanskis, Jean-Jacques Szczeciniarz, Bernard
Teissier, Emmanuel Trélat, Steven Zucker, Hervé Zwirn.

Je remercie du fond du cœur mon épouse Anne qui m'a accompagné et
soutenu pendant toutes ces années.

Paris, le 8 avril 2024

© Jean Petitot, 2024

Cet ouvrage a été (imparfaitement) saisi en L^AT_EX par l'auteur au format *amsbook* avec le logiciel *Scientific Word* de MacKichan Software.

La traduction anglaise doit paraître chez Springer dans la série *Lectures Notes in Morphogenesis* dirigée par Alessandro Sarti.

Quelques études du même auteur pertinentes pour la Neuro-géométrie

- o *Les Catastrophes de la Parole. De Roman Jakobson à René Thom*, Maloine, Paris, 1985.
- o *Physique du Sens*, Éditions du CNRS, Paris, 1992.
- o Vers une Neuro-géométrie. Fibrations corticales, structures de contact et contours subjectifs modaux, *Numéro spécial de Mathématiques, Informatique et Sciences Humaines*, 145, 1999, 5-101, École des Hautes Études en Sciences Sociales, Paris.
- o *Morphogenesis of Meaning*, (trans. F. Manjali), Peter Lang, Bern, 2003.
- o Neurogéométrie des architectures fonctionnelles de la vision, *Mathématiques et Vision, Journée annuelle de la Société Mathématique de France*, 24 juin 2006, 69-128.
- o *Neurogéométrie de la vision. Modèles mathématiques et physiques des architectures fonctionnelles*, Les Éditions de l'École Polytechnique, Distribution Ellipses, Paris, 2008.
- o *Cognitive Morphodynamics. Dynamical Morphological Models of Constituency in Perception and Syntax*, (with R. Doursat), Peter Lang, Bern, 2011.
- o Landmarks for neurogeometry, *Neuromathematics of Vision*, (G. Citti, A. Sarti eds), Springer, Berlin, Heidelberg, 1-85, 2014.
- o *Elements of Neurogeometry* (Volume 1), *Functional Architectures of Vision*, Lecture Notes in Morphogenesis, Springer, 2017.

- o (dir.) *Logos et Théorie des Catastrophes*, Colloque de Cerisy à partir de l'œuvre de René Thom, Éditions Patino, Genève, 1988.
- o (ed., with F. Varela, J.-M. Roy and B. Pachoud) *Naturalizing Phenomenology. Issues in Contemporary Phenomenology and Cognitive Science*, Stanford University Press, 1999.
- o (ed. with J. Lorenceau), "Neurogeometry and Visual Perception", Special issue of the *Journal of Physiology-Paris*, 97, 2003.
- o (ed., with M. Bitbol and P. Kerszberg) *Constituting Objectivity. Transcendental Perspectives on Modern Physics*, The Western Ontario Series in Philosophy of Science, vol. 74, Springer, Berlin, New York, 2009.
- o (ed. with A. Sarti and G. Citti), "Neuromathematics of vision", Special Issue of the *Journal of Physiology-Paris*, 103, 1-2, 2009.

Présentation

Ce second volume des *Éléments de Neurogéométrie* est consacré à la modélisation mathématique des nombreuses données expérimentales sur le cortex visuel primaire présentées dans le premier volume. Il enrichit notablement les modèles d'architecture fonctionnelle corticale déjà introduits dans le volume de 2008 *Neurogéométrie de la Vision*.

Le premier pilier en est la théorie à la Cartan des formes différentielles, des connexions et des “espaces de jets”. Les structures dites de contact y jouent un rôle prépondérant. D'abord celle de l'espace des 1-jets des courbes planes, structure invariante sous l'action du groupe de Heisenberg polarisé, ensuite celle du groupe de Lie $SE(2)$ des déplacements du plan. La théorie des groupes de Lie constitue le second pilier de l'ouvrage.

Ces structures de contact sont munies d'une géométrie sous-riemannienne naturelle dont les géodésiques sont proposées comme modèles des contours illusoires modaux qui constituent l'une des énigmes de la perception visuelle. Avec les méthodes variationnelles de la théorie du contrôle optimal qui permettent de les étudier, ces géodésiques forment le troisième pilier du volume.

Enfin, le quatrième pilier en est l'analyse harmonique non commutative reposant sur la théorie des représentations des groupes d'invariance. Associée à la théorie des “états cohérents” et de la diffusion sous-riemannienne, elle permet de comprendre comment la géométrie du connectome pilote l'analyse du signal optique par les profils récepteurs des neurones visuels.

Une particularité de l'ouvrage est d'approfondir ces structures modélisantes, d'une part en amont, en présentant leur histoire captivante et d'autre part, en aval, en compilant certains de leurs développements techniques dépassant leur implication dans les modèles neurogéométriques.

S'adressant à des biomathématiciens possédant déjà une double compétence minimale en neurosciences et en mathématiques et intéressés aux formalismes géométriques de l'analyse différentielle, cette longue étude voudrait promouvoir en neurosciences des concepts mathématiques permettant de géométriser le connectome visuel primaire et, en sens inverse, de promouvoir une interprétation neuronale des fondements de la géométrie.

Éléments de Neurogéométrie

Volume 2

Géométrie sous-riemannienne et analyse harmonique non commutative du cortex visuel primaire

Jean Petitot

Centre d'Analyse et de Mathématique Sociales
École des Hautes Études en Sciences Sociales

Paris, 2024

Table des matières

Chapitre 1. Introduction	1
1. Remarques liminaires	1
2. Modèles théoriques VS IA computationnelle	2
3. Géométrie du connectome	4
4. La difficulté théorique principale	5
5. La Neurogéométrie comme géométrie “from within”	7
5.1. Idéalité transcendantale et matérialisme neuronal	7
5.2. La “twofold way” de la géométrie	7
5.3. Les structures neurogéométriques de base	8
5.4. Données manifestes “overt” et activités neuronales “covert”	9
5.5. Un phénomène noyau	11
6. Quelques repères historiques du parcours	11
6.1. Les différentes étapes	11
6.2. Géométrie et Cognition	13
7. Résumé du Volume I	14
7.1. Chapitre 1 du Vol I	14
7.2. Chapitre 2 du Vol I	14
7.3. Chapitre 3 du Vol I	14
7.4. Chapitre 4 du Vol I	15
7.5. Chapitre 5 du Vol I	16
7.6. Chapitre 6 du Vol I	18
8. Nos choix méthodologiques et pédagogiques	19
8.1. Un marteau-pilon pour casser une noisette?	19
8.2. Modèles particuliers et structures générales	20
8.3. Une perspective propédeutique	21
8.4. Ouvrages de référence et/ou traité didactique	21
8.5. L’encadrement des modèles en amont et en aval	22
8.6. Un guide cartographique pour une circumnavigation	23
8.7. L’usage d’une conception “internaliste” des mathématiques	23
8.8. Une pédagogie mathématique inhabituelle et risquée	24
8.9. Réquisits	25

8.10.	Beaucoup de calculs pour rien?	25
8.11.	Interdisciplinarité et polyscientificité intrinsèques	27
8.12.	Neurogéométrie et formalismes physiques	27
8.13.	La vertu pédagogique d'un peu de redondance	28
8.14.	Les notations	28
8.15.	Bibliographie	29
8.16.	Un traité quadriparti	30
9.	Résumé du volume	30
9.1.	Le fil directeur	30
9.2.	Chapitre 2	32
9.3.	Chapitre 3	34
9.4.	Chapitre 4	37
9.5.	Chapitre 5	40
9.6.	Chapitre 6	41
9.7.	Chapitre 7	44
9.8.	Chapitre 8	46
9.9.	Chapitre 9	48
9.10.	Chapitre 10	53
9.11.	Chapitre 11	54
9.12.	Chapitre 12	55
9.13.	Chapitre 13	57
9.14.	Chapitre 14	58
9.15.	Chapitre 15	61
9.16.	Chapitre 16	65
9.17.	Chapitre 17	77
9.18.	Chapitre 18	89
9.19.	Chapitre 19 : Conclusion	90
Partie I : Géométrie à la Lie-Cartan		93
Chapitre 2. La géométrie différentielle de $V1$ comme fibration		95
1.	Brefs rappels expérimentaux (Vol I)	95
1.1.	Champs récepteurs et analyse en ondelettes	95
1.2.	La structure hypercolumnaire de l'aire $V1$	97
1.3.	L'organisation en pinwheels de $V1$	98
1.4.	Connexions horizontales et structure de contact	102
1.5.	Intégration des contours et champ d'association	104
1.6.	Un double vocabulaire	106
2.	$V1$ comme fibration	106
3.	Méthodologie des modèles géométriques	107
4.	Neurotopologie	108

4.1.	Le problème d'un espace immanent global	109
4.2.	Un minimum de topologie	110
4.3.	Topologie et Gestalttheorie : ouvert/fermé et fond/forme	113
4.4.	L'implémentation neuronale de la topologie et son immanence	116
4.5.	Topologie et homotopie	125
5.	Interlude philosophique : vers un transcendantalisme neuronal	127
5.1.	Problèmes inverses et "révolution copernicienne"	127
5.2.	Intuition, schème, concept	128
5.3.	Kant \implies Helmholtz \implies aujourd'hui	129
6.	De la topologie à la géométrie différentielle	130
6.1.	Variétés différentiables et fibrés tangents	130
6.1.1.	Variétés et cartes locales	130
6.1.2.	Vecteurs tangents	131
6.1.3.	Ponctuel, infinitésimal, local, global	132
6.2.	Le calcul différentiel classique	133
6.2.1.	Les infinitésimales leibniziennes et leur paradoxe	133
6.2.2.	Quelques réponses au paradoxe	134
6.3.	Opérateurs de dérivation et formes différentielles	139
6.3.1.	Dualité "spectrale" entre espace et fonctions	139
6.3.2.	Les vecteurs tangents comme opérateurs de dérivation	140
6.3.3.	L'algèbre de Lie des champs de vecteurs	141
6.3.4.	Les 1-formes différentielles	143
6.4.	L'implémentation neuronale de la structure différentiable	144
7.	Élie Cartan et le calcul des formes différentielles	145
7.1.	Une vision exceptionnelle	145
7.2.	La notion de forme différentielle	146
7.2.1.	Formes multilinéaires antisymétriques	146
7.2.2.	Intégration des 1-formes	147
7.2.3.	Intégration des k -formes et déterminants	149
7.2.4.	Le produit extérieur	150
7.2.5.	Le produit intérieur	152
7.2.6.	Intégration invariante et signification géométrique intrinsèque	153
7.2.7.	La dérivation extérieure	154
7.2.8.	La nilpotence $d^2 = 0$	159
7.2.9.	Formes et cohomologie	160
7.2.10.	Chaînes, cycles, bords et homologie	161
7.2.11.	Le théorème de Stokes : dualité homologie/cohomologie	164
7.2.12.	Dérivées de Lie	166
7.3.	Le texte fondamental d'Élie Cartan de 1899	168
7.3.1.	Un retournement structural du calcul intégral	168

7.3.2.	Un calcul symbolique	170
7.3.3.	Le lien avec les intégrales multidimensionnelles	171
7.3.4.	La dérivée extérieure des formes de Pfaff	175
7.3.5.	Les “dérivées” d’ordre supérieur	178
8.	Formes différentielles et physique classique	179
8.1.	L’analyse vectorielle classique	179
8.1.1.	Gradients, rotationnels, divergences	179
8.1.2.	Applications classiques du théorème de Stokes	182
8.1.3.	L’émergence du non commutatif	183
8.2.	Digression : l’exemple des équations de Maxwell	185
8.2.1.	L’unification des lois : Gauss, Thomson, Faraday, Ampère	185
8.2.2.	Le type des entités “champs”	189
8.2.3.	La synthèse des formules (1) et (3) et l’espace-temps	191
8.2.4.	La synthèse des formules (2) et (4) et la théorie de Hodge	193
8.2.5.	Calculs symboliques et objectivité physique	195
9.	Faisceaux et cohomologie	196
9.1.	Le concept de faisceau	196
9.2.	Phénoménologie du remplissage et du recollement	197
9.3.	Formalisation du remplissage et du recollement	201
9.4.	Cohomologies de Čech et de de Rham	204
Chapitre 3.	$V1$ comme structure de contact	209
1.	Structures de contact et vision	209
2.	Relevées legendriennes, forme de contact et espace des 1-jets	210
2.1.	L’espace des 1-jets	210
2.2.	Relevées legendriennes et condition d’intégrabilité	213
2.3.	La forme ω et l’équation $\omega = 0$	216
2.4.	$SE(2)$ -invariance	219
2.5.	Structure de contact et courbure	221
2.6.	La non-intégrabilité de la structure de contact	221
2.7.	La “simplexité” de la notion de jet	224
2.8.	De Frobenius à Hörmander	226
2.9.	Champs et points caractéristiques des surfaces de V	227
2.9.1.	Définition et exemple	227
2.9.2.	Le champ Δ sur C	228
2.9.3.	Le champ D sur S	232
2.9.4.	Les cusps le long du contour apparent	234
2.10.	Structure de contact et EDOs	240
3.	Champs de Reeb, structures symplectiques, métriques et complexes	241
3.1.	Structures symplectiques	241
3.2.	Symplectisation de la structure de contact	242

3.3.	Formes $f\omega$, champs de Reeb et métriques	243
3.4.	Structures (presque) complexes	244
3.4.1.	Conditions générales	244
3.4.2.	L'exemple de \mathbb{V}_J	247
3.4.3.	Modification de l'exemple	249
3.4.4.	Le cas de la métrique euclidienne	254
4.	Courbure, 2-jets et structure d'Engel	256
4.1.	Structure d'Engel dans le cas $\{x, y, p\}$	257
4.2.	Structure d'Engel dans le cas $\{x, y, \theta\}$	257
5.	La notion cruciale d'élément de contact	258
5.1.	Les "écailles de poisson" sont de bonnes géomètres (Klein)	258
5.2.	Des "point processors" peuvent être de bons géomètres	262
Chapitre 4. Géométriser l'intégrabilité : une "révolution copernicienne"		265
1.	Que signifie une "révolution copernicienne"	265
1.1.	De Copernic à Galois	265
1.2.	La similitude des exemples	267
1.3.	L'exemple initial (Copernic, Kepler, Newton)	267
1.4.	L'exemple de Kant et de la relativité galiléenne	268
1.5.	L'exemple de Galois	272
1.6.	Notre exemple de neurotopologie	282
1.7.	Notre exemple de l'intégrabilité	283
2.	Le problème de Pfaff de Monge à Cartan : la période héroïque	283
2.1.	Le "moment fécond" de la géométrie de l'intégrabilité	283
2.2.	Quelques repères biographiques	284
2.3.	EDO et EDP	292
2.3.1.	EDO	292
2.3.2.	EDP	294
2.4.	Les EDT : problème direct et problème inverse	295
2.5.	La "véritable" intégrabilité chez Monge	301
2.6.	Autour du Mémoire de Pfaff (les années 1810-1820)	304
2.6.1.	L'idée générale de l'intégration d'une équation de Pfaff	304
2.6.2.	Le résultat principal	305
2.6.3.	Difféomorphismes et élimination d'éléments différentiels	306
2.6.4.	La méthode de Pfaff	307
2.6.5.	La méthode du <i>Methodus</i>	312
2.6.6.	L'apport de Jacobi	315
2.7.	Les années 1860-1880	316
2.7.1.	Grassmann	316
2.7.2.	Natani	318
2.7.3.	Clebsch	318

2.7.4.	Frobenius	318
2.7.5.	Lie	321
2.7.6.	Darboux	322
3.	La systématisation de Cartan et Goursat	327
3.1.	Classe et réduction	327
3.2.	Un exemple	330
3.3.	L'intégration des structures de contact	336
3.3.1.	Solutions génériques	336
3.3.2.	Solutions singulières	338
3.3.3.	EDP du premier ordre avec contraintes	339
3.3.4.	Transformations de contact	339
3.4.	Les invariants intégraux de Poincaré-Cartan	341
3.5.	La pleine géométrisation de l'intégrabilité	343
3.5.1.	Éléments de contact intégraux et associés	343
3.5.2.	Cas d'une seule équation de Pfaff ($s = 1$)	348
3.5.3.	Prolongements de variétés intégrales	349
3.5.4.	Inégalités dimensionnelles	349
3.5.5.	La solution générale dans le cas $n = 5$	352
3.5.6.	Éléments caractéristiques	355
3.5.7.	Éléments singuliers	355
Chapitre 5.	V_1 comme groupe de Lie	359
1.	Rendre \mathcal{K} invariante	359
2.	Groupes et algèbres de Lie (prélude)	360
2.1.	Définition	360
2.2.	L'algèbre de Lie sur l'espace tangent $\mathcal{G} = T_e G$	362
2.3.	La parallélisabilité	363
3.	\mathbb{V}_J comme groupe de Heisenberg polarisé \mathbb{H}_{pol}	364
3.1.	La loi de groupe	364
3.2.	\mathbb{V}_J comme produit semi-direct $\mathbb{R}^2 \rtimes \mathbb{R}$	366
3.3.	La condition d'Hörmander	369
3.4.	Représentations matricielles	370
4.	Algèbres de Lie nilpotentes VS semi-simples (introduction)	372
5.	Le groupe de Heisenberg non polarisé \mathbb{H}	374
5.1.	L'exponentielle du groupe	374
5.2.	\mathbb{H} et le problème isopérimétrique	376
5.3.	$\mathbb{H}_{pol} \simeq \mathbb{V}_J$ et \mathbb{H} : les coordonnées canoniques	377
5.4.	Le groupe de Heisenberg \mathbb{H} et la mécanique quantique	378
6.	Représentations adjointes et co-adjointes	380
6.1.	Invariance par translations à gauche et non-holonomie	380
6.2.	Les représentations Ad, ad, Ad^*, ad^* et leurs orbites	382

6.2.1.	Représentation adjointe	382
6.2.2.	Représentation co-adjointe	383
6.3.	La forme symplectique de Kirillov	385
7.	\mathbb{V}_S comme groupe des déplacements du plan	387
8.	Extension en dimension 3	391
9.	Exponentielle générale	393
10.	La forme de Maurer-Cartan	396
Chapitre 6. Géométrie des connexions		399
1.	Au “cœur” de l’espace : géométrie différentielle et groupes de Lie	399
2.	Les structures de contact comme connexions	402
3.	Connexions et dérivation covariante : mini <i>vademecum</i>	403
3.1.	Cadre général et variantes	403
3.1.1.	Un calcul différentiel fibré	403
3.1.2.	Horizontal/vertical, externe/interne	404
3.1.3.	La base M et ses fibrés de repères	404
3.1.4.	Fibrés principaux	407
3.1.5.	Fibrés vectoriels	409
3.1.6.	Symétries internes	410
3.1.7.	Fibrés quelconques	410
3.2.	Transport parallèle infinitésimal	410
3.2.1.	L’intuition géométrique initiale	411
3.2.2.	Das Raumproblem	411
3.3.	Connexions générales et espaces horizontaux	414
3.4.	Dérivation covariante (prélude)	416
3.5.	Connexions principales	417
3.6.	Les potentiels vecteurs et le formalisme physique	418
3.7.	Connexions linéaires ou affines (avec “déluge d’indices”)	419
4.	Théories de jauge	424
4.1.	Théories de jauge abéliennes (équations de Maxwell)	424
4.2.	Théories de jauge non abéliennes	426
5.	La dialectique “interne/externe”	428
5.1.	Connexions sur le fibré des repères	428
5.2.	Le problème de la torsion	431
6.	Géométrie riemannienne	432
6.1.	Métriques et courbures	432
6.2.	Géodésiques riemanniennes	435
6.3.	L’exemple de la sphère \mathbb{S}^2	437
6.4.	Note sur la relativité générale	438
6.5.	Note sur parallélisme “absolu” (Cartan-Einstein)	439
7.	Géométrie sous-riemannienne (prélude)	443

8. Groupes de Carnot et variétés de Heisenberg	445
8.1. Les groupes de Carnot et leur propriétés d'homogénéité	446
8.2. Groupes de Carnot et champ d'association	450
8.3. Variétés de Heisenberg, nilpotentisation et cônes tangents	450
Chapitre 7. Géométrie des groupes et algèbres de Lie	455
1. Géométrie de la nilpotence	455
2. Algèbres résolubles	457
3. Structure des algèbres de Lie générales	459
3.1. Idéaux et forme de Killing	459
3.2. Mesure de Haar et unimodularité	463
3.3. La classification des algèbres de Lie réelles de dimension ≤ 3	464
3.3.1. $\dim \mathcal{G}^{(1)} = 0$	465
3.3.2. $\dim \mathcal{G}^{(1)} = 1$	465
3.3.3. $\dim \mathcal{G}^{(1)} = 2$	467
3.3.4. $\dim \mathcal{G}^{(1)} = 3$	470
3.4. Algèbre enveloppante et algèbre symétrique	479
3.4.1. Aspect algébrique	480
3.4.2. Aspect fonctionnel	481
3.5. Retour sur la formule de Baker-Campbell-Hausdorff	485
4. Métriques invariantes sur les groupes de Lie (prélude)	488
4.1. Métriques riemanniennes L -invariantes en D3 (Milnor)	488
4.2. Géodésiques homogènes	502
4.3. Métriques sous-riemanniennes L -invariantes en D3 (Agrachev)	504
4.3.1. Les invariants métriques δ et κ	504
4.3.2. Cas $\delta \neq 0$	509
4.3.3. Cas $\delta = 0$	514
Partie II : Géométrie sous-riemannienne	517
Chapitre 8. Contours illusoires modaux et modèles variationnels	519
1. Complétion modale VS complétion amodale	519
2. Les contours illusoires modaux	520
2.1. Exemples de contours subjectifs modaux	521
2.1.1. Contours illusoires du premier ordre	521
2.1.2. Les deux types d'intégration	526
2.1.3. Contours illusoires du second ordre	530
2.1.4. Certains corrélats neuronaux des CIs	531
2.1.5. Présence des CIs chez de nombreuses espèces	535
2.2. Quelques modèles classiques de CIs	537
2.3. Méthodes de mesure des CIs	540

2.4.	Une expérience sur les K -contours courbes (avec J. Ninio)	543
2.5.	Psychophysique et traitement d'image	546
2.6.	Neurogéométrie de la contribution de $V1$	547
3.	Les premiers modèles variationnels (Ullman, Horn, Mumford)	547
3.1.	“L’intelligence des lignes courbes”	547
3.2.	Les classes de modèles	551
3.3.	Les modèles de Ullman et Horn	552
3.4.	Le modèle des elasticæ de Mumford	552
4.	Des elasticæ dans \mathbb{R}^2 aux géodésiques dans \mathbb{V}	555
4.1.	Relèvements dans les espaces de jets	555
4.2.	Contours illusoires et surfaces minimales	556
4.3.	Lagrangiens et métriques	557
4.4.	“Géodésiques legendriennes” dans le fibré de contact \mathbb{V}_J	557
Chapitre 9. L’univers des modèles variationnels classiques		559
1.	Quelques repères biographiques	560
2.	Quadratures et rectifications	565
2.1.	Quadratures	565
2.2.	Rectifications	566
3.	La rectification de l’ellipse	567
4.	Développements en séries	569
5.	Fonctions circulaires et trigonométrie	569
5.1.	Des intégrales circulaires aux fonctions circulaires	570
5.2.	Les formules d’addition	572
5.3.	Extension à \mathbb{R}	574
5.4.	Extension à \mathbb{C}	575
5.5.	Fonctions analytiques et conditions de Cauchy-Riemann	576
6.	La lemniscate	578
6.1.	La géométrie de la lemniscate	578
6.2.	Le Comte de Fagnano et Euler	581
6.3.	Une trigonométrie “lemniscatique” (Gauss, Abel)	585
7.	Le pendule et les fonctions elliptiques	587
7.1.	Les équations du mouvement	587
7.2.	Les fonctions elliptiques $\operatorname{sn}(u)$, $\operatorname{cn}(u)$ et $\operatorname{dn}(u)$	589
7.3.	La limite $k \rightarrow 1$	594
7.4.	Translations, dérivations et formules d’addition	597
7.5.	Extension à \mathbb{R} et à \mathbb{C}	599
8.	Calcul des variations : le <i>Methodus</i> d’Euler (1744)	600
9.	Mécanique et Lagrangiens	615
9.1.	D’Euler à Lagrange	615
9.2.	Équations d’E-L et relevées legendriennes	616

9.3.	L'exemple newtonien	618
10.	Les elasticæ : de Bernouilli et Euler à Kirchoff	620
10.1.	Le problème initial	621
10.2.	L'explication variationnelle	621
10.3.	L'explication mécanique causale	624
10.4.	La classification (<i>enumeratio</i>) des solutions	625
10.5.	Géométrie des elasticæ et mécanique du pendule	633
11.	Intégrales, fonctions et courbes elliptiques	634
11.1.	La classification de Legendre	634
11.2.	Vers les courbes elliptiques	637
Chapitre 10. Contours illusoires modaux et géodésiques "legendriennes"		641
1.	Les différents modèles	641
2.	Les équations d'Euler-Lagrange (I)	642
2.1.	L'essai de la métrique euclidienne	642
2.2.	Le lagrangien non contraint	644
2.3.	Le multiplicateur de Lagrange	646
2.3.1.	La méthode	646
2.3.2.	L'équation différentielle pour la tangente $g = f'$	646
2.3.3.	Les classes de solutions	650
3.	Intégration numérique dans le cas symétrique $A = 0, B \neq 0$	650
4.	Remarques sur le cas $A \neq 0, B = 0$	653
5.	Équation différentielle sur la courbure	656
6.	Les équations d'Euler-Lagrange (II)	660
7.	Le modèle du fibré en cercles	662
8.	Les courbes de moindre énergie	665
9.	Comparaison avec les elasticæ	666
Chapitre 11. Passage à $SE(2)$ et au fibré principal		669
1.	Justification du passage	669
2.	Les repères mobiles et le fibré principal $SE(2) = \mathbb{R}^2 \rtimes SO(2)$	670
3.	Les représentations adjointe et co-adjointe de $G = SE(2)$	672
3.1.	L'algèbre de Lie de $G = SE(2)$	672
3.2.	Changements de repère et automorphismes intérieurs	674
3.3.	La représentation adjointe	676
3.4.	Les orbites de la représentation adjointe	677
3.5.	La représentation co-adjointe	677
3.6.	Les orbites de la représentation co-adjointe	679
4.	Le formalisme d'Élie Cartan	681
4.1.	La forme de Maurer-Cartan	681
4.2.	Les équations de Maurer-Cartan	683

5.	Les problèmes variationnels d'après Bryant et Griffiths	684
5.1.	Le système de Pfaff des relevées dans le fibré principal G	684
5.2.	La fonctionnelle du problème variationnel	686
5.3.	Les équations d'Euler-Lagrange (III)	687
5.4.	L'interprétation géométrique des équations d'Euler-Lagrange	691
5.5.	Retour sur les modèles du fibré en cercle et des elasticæ	694
Chapitre 12. Lagrangiens, hamiltoniens et contrôle optimal		695
1.	Une double généralisation	695
2.	La formulation hamiltonienne	695
2.1.	La transformation de Legendre	696
2.1.1.	Cas élémentaires	696
2.1.2.	Cas général	701
2.2.	Les conditions de Legendre et de Weierstrass	705
3.	Théorème de Noether et application moment	705
3.1.	Le théorème de Noether	706
3.1.1.	Version lagrangienne	706
3.1.2.	Version hamiltonienne	706
3.2.	L'application moment	707
3.3.	Symétries et mécanique a priori	708
4.	La théorie du contrôle optimal	710
4.1.	Une vaste généralisation	710
4.2.	Conditions et contraintes	712
5.	Le principe du maximum de Pontryagin (PMP)	713
5.1.	Énoncé	713
5.2.	Un exemple élémentaire	717
5.3.	Quelques idées de la preuve (d'après A. Lewis)	719
5.3.1.	Variations et cônes tangents	719
5.3.2.	Cônes tangents et ensembles atteignables	721
5.3.3.	Un premier théorème	723
5.3.4.	Le théorème général	725
6.	Géodésiques, caustiques et solutions lagrangiennes	726
6.1.	Un détour par les caustiques en optique	728
6.1.1.	Un exemple	729
6.1.2.	Géométrie des surfaces	730
6.1.3.	Solutions asymptotiques de l'équation des ondes	732
6.1.4.	Solutions lagrangiennes	734
6.2.	Un détour par les ellipsoïdes	736
6.2.1.	La sphère	736
6.2.2.	Les ellipsoïdes de révolution	741
6.2.3.	Les ellipsoïdes triaxiaux	745

6.2.4. Généralisations	752
Chapitre 13. Retour sur les <i>elasticæ</i> : strates de Maxwell et points conjugués (d'après Y. Sachkov)	755
1. Les équations de Hamilton	755
2. Lien avec le pendule	760
3. Coordonnées de Jacobi	761
4. Stratification des espaces cotangents	765
5. Intégration des équations pour $k \in [0, 1]$ (pendule oscillant)	766
6. Quelques calculs dans le cas $\varphi_0 = 0$ ($k \in [0, 1]$)	768
6.1. Géométrie générale	768
6.2. Catastrophes de conflits	771
6.3. Exemple $k = 0.5$	771
6.4. Exemple $k \sim 0.707107$ (<i>elastica</i> rectangulaire)	772
6.5. Exemple $k = 0.8$	773
6.6. Exemple $k \sim 0.8551$	774
6.7. Exemple $k \sim 0.8858$	774
6.8. Exemple $k = k_0 \sim 0.908909$ ($2E(k) - K(k) = 0$, lemniscate)	776
6.9. Exemple $k \sim 0.9624$	776
6.10. Exemple $k : 0.99999 \rightarrow 1$	777
7. Intégration des équations pour $k > 1$ (pendule tournant)	778
8. Quelques calculs dans le cas $\varphi_0 = 0$ ($k > 1$)	780
9. Résumé	782
10. Symétries et points de Maxwell	785
11. Cut locus et points conjugués	789
Chapitre 14. Contours illusoires modaux et géodésiques sous-riemanniennes	791
1. Passage à la géométrie sous-riemannienne	791
2. Les géodésiques sous-riemanniennes de \mathbb{V}_J	793
2.1. Le groupe de Heisenberg polarisé	793
2.1.1. L'hamiltonien des géodésiques	793
2.1.2. Calcul des géodésiques	796
2.1.3. La multiplicité des géodésiques	800
2.1.4. La sphère et le front d'onde sous-riemanniens	802
2.1.5. Singularités des sphères sous-riemanniennes	804
2.2. Le groupe de Heisenberg non polarisé (Beals, Gaveau, Greiner)	808
2.2.1. Le groupe \mathbb{H}	808
2.2.2. Les géodésiques de \mathbb{H} : hamiltonien et lagrangien	811
2.2.3. Les calculs de Beals, Gaveau et Greiner	814
2.2.4. Cut locus et points conjugués	818
3. Les géodésiques des espaces de jets	820

4.	Les géodésiques sous-riemanniennes de \mathbb{V}_S (Agrachev, Sachkov)	822
4.1.	Le système hamiltonien sur \mathbb{V}_S	822
4.2.	Intégration dans le cas d'une impulsion de module 1	826
4.3.	Du portrait de phase aux géodésiques	829
4.4.	Images de géodésiques	836
4.4.1.	Considérations préliminaires	836
4.4.2.	De C_1 vers C_3	837
4.4.3.	Arrêt sur C_3	841
4.4.4.	De C_3 à C_2	843
4.4.5.	Représentations toriques	846
4.4.6.	Un feuilletage géodésique compliqué	851
4.5.	Exponentielle, cut locus, points conjugués	852
4.6.	Récapitulation	867
4.7.	Les sphères et les fronts d'onde de $\mathbb{V}_S = SE(2)$	872
4.7.1.	Structure générale	872
4.7.2.	Le cas $R = \frac{\pi}{2}$	881
4.7.3.	Le cas $R = \frac{3\pi}{2}$	886
4.7.4.	Le cas $R = \frac{3\pi}{2}$: représentations toriques	892
4.8.	Non commutativité et complexité	904
5.	Cusps des projections des géodésiques	910
5.1.	Le modèle \mathbb{V}_S^\emptyset	910
5.1.1.	La contrainte d'élimination des cusps	910
5.1.2.	L'hamiltonien de \mathbb{V}_S^\emptyset	912
5.1.3.	Les portraits de phases	913
5.1.4.	Les cônes accessibles	916
5.1.5.	Le fibré en cercles	918
5.2.	Le modèle $\mathbb{V}_P = SE(2) / (\sim \text{mod } \pi)$	918
5.3.	Cusps ou points anguleux ?	919
5.4.	La bistabilité des contours illusoirs	920
6.	Quelques autres résultats	922
6.1.	Variations des géodésiques	922
6.2.	Modèles stochastiques	925
6.3.	Autres groupes de Lie	927
6.4.	Caustiques génériques (Agrachev, Gauthier, Zakalyukin)	928
6.5.	Embarquement pour l'univers des jets	930
Chapitre 15. Géométrie des distributions de jets		931
1.	Généralités sur les distributions d'hyperplans tangents	931
1.1.	Distributions générales	931
1.2.	Courbure d'une distribution	932
1.3.	Typologie des distributions	933

1.4.	Le principe de Thom : stratifications et transversalité	937
1.5.	Formes normales	938
1.6.	Classification des germes de 1-formes sur \mathbb{R}^3	940
1.6.1.	Les cas simples	940
1.6.2.	Les cas plus compliqués (Zhitomirskii)	943
2.	Géométrie de contact et géométrie symplectique	946
2.1.	Les variétés de contact et leur symplectisation	946
2.2.	Formes normales de Darboux	949
2.2.1.	Les formules	949
2.2.2.	Structures symplectiques, riemanniennes et complexes	950
2.2.3.	Sous-variétés lagrangiennes	953
2.2.4.	Forme de contact de Darboux	953
2.2.5.	Relevées legendriennes	954
2.2.6.	Stabilité structurelle des structures de contact	954
2.3.	Champs de contact	955
2.3.1.	Champs de contact sur \mathbb{V}_J	955
2.3.2.	La fonction génératrice	959
2.3.3.	Relèvement des champs de l'espace de base	961
2.3.4.	Surfaces convexes	963
3.	Géométrie des espaces de jets et distributions de Goursat	966
3.1.	Jets et géométrie des équations différentielles	966
3.2.	Les dérivées "cachées" de Leibniz à Montgomery	966
3.3.	Distributions et drapeaux de Goursat	967
3.4.	Apparition de singularités (Giaro-Kumpera-Ruiz)	971
3.5.	La récurrence K-R	973
3.6.	Géométrie des singularités (Montgomery-Zhitomirskii)	975
3.7.	Le "Monstre" classifiant	981
3.7.1.	La tour infinie des complétions projectives $\mathbb{P}_{\mathfrak{M}}^s$ des $J^s(\mathbb{R}, \mathbb{R})$	981
3.7.2.	Le relèvement de singularités	985
3.7.3.	Points singuliers verticaux	988
3.7.4.	Courbes et points critiques	992
4.	Structures de contact globales	994
4.1.	Local/global : une dialectique fondamentale	994
4.2.	Contraintes d'existence de structures de contact globales	995
4.3.	Le problème de la classification.	997
4.4.	Trajectoires fermées du champ de Reeb et homologie de contact	997
4.5.	Courbes pseudo-holomorphes de Gromov	1000
4.6.	La dimension 3 et la théorie des nœuds (Bennequin)	1001
4.6.1.	Nœuds et feuilletages caractéristiques	1001
4.6.2.	L'exemple de \mathbb{S}^3 (fibration de Hopf et quaternions)	1001

4.7.	Le théorème de Bennequin	1005
4.8.	La méthode des surfaces convexes (Giroux)	1013
4.9.	Prolongements	1018
Partie III : Analyse harmonique non commutative et diffusion		1019
Chapitre 16. Mini <i>vademecum</i> d'analyse fonctionnelle		1021
1.	Géométrie et analyse fonctionnelle en analyse harmonique	1021
2.	Mesure et intégration	1023
2.1.	Espaces mesurables	1023
2.2.	Intégration	1026
2.3.	Théorème de Radon-Nykodym-Lebesgue	1027
3.	Espaces $L^p(X, \nu)$	1028
4.	Espaces de Hilbert	1036
4.1.	Produit scalaire	1036
4.2.	Projecteurs	1038
4.3.	Théorème de Riesz et auto-dualité	1038
4.4.	Théorème de "séparation" de Hahn-Banach	1039
4.5.	Bases hilbertiennes	1040
4.6.	Non compacité de la boule unité	1041
4.7.	Topologies	1041
5.	Les distributions de Laurent Schwartz	1042
6.	L'exemple des séries de Fourier	1046
7.	Les opérateurs et leur décomposition spectrale	1074
7.1.	Opérateurs continus et bornés	1074
7.2.	Opérateurs non bornés et fermés	1077
7.3.	Des opérateurs symétriques aux opérateurs auto-adjoints	1078
7.3.1.	L'opérateur adjoint	1078
7.3.2.	Opérateurs auto-adjoints en dimension finie	1078
7.3.3.	Opérateurs symétriques	1079
7.3.4.	Opérateurs auto-adjoints	1082
7.3.5.	Opérateurs essentiellement auto-adjoints	1082
7.3.6.	Conditions de symétrie et d'adjonction	1084
7.4.	Décomposition spectrale des opérateurs auto-adjoints	1086
7.5.	Vecteurs propres approximatifs et théorème de Stone	1090
8.	Quantification géométrique	1095
8.1.	L'obstruction initiale	1095
8.1.1.	Les conditions de Dirac	1095
8.1.2.	Cas classique	1097
8.1.3.	Cas quantique	1099
8.1.4.	Comment passer du commutatif au non commutatif?	1100

8.2.	Préquantification	1102
8.3.	Polarisation	1113
8.3.1.	Polarisation verticale de $T^*\mathbb{R}^n$	1113
8.3.2.	Polarisation horizontale de $T^*\mathbb{R}^n$	1115
8.3.3.	Polarisations kähleriennes	1117
8.3.4.	Noyaux BKS	1120
8.3.5.	Stone-von Neumann : de Schrödinger à Fourier et Bargmann	1121
8.4.	Demi-formes et demi-densités	1122
8.5.	Conditions discrètes de quantification	1124
8.6.	Exemples	1126
8.6.1.	Le fibré cotangent du cercle	1126
8.6.2.	Les orbites de la représentation co-adjointe de $SE(2)$	1128
8.6.3.	La fibration lagrangienne du plan épointé	1131
8.6.4.	L'oscillateur harmonique	1132
8.7.	Approximations semi-classiques	1134
8.7.1.	La quantification de Weyl	1134
8.7.2.	Retour sur les caustiques en optique	1135
8.7.3.	De l'optique à la mécanique quantique	1140
9.	Opérateurs pseudo-différentiels (OΨD)	1141
9.1.	Généralités	1141
9.2.	Intégrales oscillantes	1149
9.3.	Opérateurs elliptiques	1150
9.4.	Opérateurs hyperboliques	1151
9.5.	Opérateurs intégraux de Fourier (OIF)	1152
9.6.	Analyse microlocale	1152
9.7.	Opérateurs pseudo-différentiels sur les variétés de Heisenberg	1152
9.8.	Passage à l'analyse harmonique	1156
Chapitre 17. Analyse harmonique non commutative, états cohérents et diffusion		
	sous-riemannienne	1159
1.	Représentations de groupes	1159
1.1.	Généralités	1159
1.1.1.	Un concept fondamental	1159
1.1.2.	Définition basique	1160
1.1.3.	Groupes finis	1161
1.1.4.	Groupes de Lie	1164
1.2.	Représentations dans des espaces de Hilbert	1167
1.2.1.	Représentations générales	1167
1.2.2.	Extension des représentations	1168
1.3.	Représentations unitaires de \mathbb{R}	1169
1.3.1.	Groupes à 1-paramètre et générateurs infinitésimaux	1169

1.3.2.	Représentations de \mathcal{G} et de G	1173
2.	Représentations de carré intégrable et états cohérents	1175
2.1.	Retour sur les exemples du Volume I	1175
2.1.1.	Transformée de Fourier	1175
2.1.2.	Transformée de Gabor	1183
2.1.3.	Transformée en ondelettes	1186
2.1.4.	Synthèses des méthodes	1188
2.1.5.	Profils récepteurs	1188
2.2.	Le problème de la discrétisation	1188
2.2.1.	Les trames de Gabor	1189
2.2.2.	Trames de Gabor dans \mathbb{V}_S (Liontou-Marcolli)	1194
3.	Vers l'analyse harmonique généralisée	1198
3.1.	Groupes de Lie compacts	1198
3.2.	Groupes de Lie abéliens	1200
3.3.	Séries de Fourier	1201
4.	Etats cohérents	1201
4.1.	Idée de la théorie générale	1201
4.2.	Précisions sur quelques exemples	1207
4.3.	Les représentations régulières et quasi-régulières	1210
4.3.1.	Représentations quasi-régulières des produits semi-directs	1210
4.3.2.	L'exemple de $\mathbb{V}_J = \mathbb{H}_{pol}$	1212
4.3.3.	L'exemple de $\mathbb{V}_S = SE(2)$	1213
4.4.	Unirreps de carré intégrable de produits semi-directs	1213
4.5.	Généralisations	1215
5.	Analyse harmonique et groupe de Heisenberg	1215
5.1.	Représentations de Schrödinger (Stone-von Neumann)	1216
5.2.	Forme intégrale de Weyl	1217
5.3.	Versions du théorème de Stone-von Neumann	1218
5.4.	Irréductibilité des représentations de Schrödinger	1220
5.5.	Reconstruction du groupe de Heisenberg	1221
5.6.	Esquisse de preuve du théorème	1223
5.6.1.	Systèmes d'imprimitivité	1224
5.6.2.	La preuve	1226
5.6.3.	Généralisation aux produits semi-directs	1229
5.7.	Lien avec la transformée de Fourier	1230
5.8.	Généralisations	1231
5.9.	Lien avec les travaux de Kirillov	1231
5.10.	Représentations de Segal-Bargmann	1232
5.10.1.	La transformée de Bargmann	1232
5.10.2.	Opérateurs de création et d'annihilation	1233

5.10.3.	Dérivation de la transformée de Bargmann	1236
5.10.4.	Le noyau reproduisant et l'état cohérent	1238
5.10.5.	Structure de l'isométrie	1239
5.10.6.	La représentation de \mathbb{H} dans \mathcal{B}	1241
5.10.7.	Lien avec le noyau de la chaleur	1242
5.10.8.	La généralisation de Brian Hall	1243
5.10.9.	Le cas du cercle	1245
5.10.10.	La transformée de Bargmann de $SE(2)$	1247
6.	Analyse harmonique sur les groupes nilpotents	1247
6.1.	Niveau 2	1248
6.2.	Niveau > 2	1249
6.3.	Représentations induites (précisions)	1251
6.4.	De Mackey à Kirillov (méthode des orbites co-adjointes)	1254
6.5.	Transformée de Fourier non commutative	1257
7.	Analyse harmonique sur les groupes (semi-)simples	1258
8.	Analyse harmonique et groupe euclidien $SE(2)$	1261
8.1.	Unirreps et orbites co-adjointes de $SE(2)$	1261
8.2.	Les représentations de l'algèbre de Lie	1264
8.3.	La transformée de Bargmann de $SE(2)$ (Barbieri-Citti)	1265
9.	Analyse harmonique et diffusion sous-riemannienne	1268
9.1.	La diffusion euclidienne	1268
9.1.1.	Les solutions fondamentales	1268
9.1.2.	Diffusion et géodésiques	1273
9.1.3.	Le noyau de la chaleur	1274
9.1.4.	Formule de la trace	1276
9.1.5.	Diffusion, mouvement brownien et calcul stochastique	1276
9.1.6.	Repères historiques du mouvement brownien	1281
9.1.7.	Passage aux cas riemanniens et sous-riemanniens	1282
9.2.	La diffusion riemannienne	1282
9.3.	Le volume intrinsèque en diffusion sous-riemannienne	1287
10.	Diffusion sous riemannienne dans \mathbb{V}_J et \mathbb{V}_S	1288
10.1.	Le cas Heisenberg	1288
10.1.1.	Noyau de Heisenberg (Gaveau, Beals, Greiner)	1288
10.1.2.	Asymptotique pour $\mathfrak{r} \rightarrow 0^+$	1296
10.1.3.	Généralisations	1299
10.1.4.	Diffusion, points conjugués et cut locus	1299
10.1.5.	Noyau de Heisenberg polarisé (Hulanicki)	1302
10.2.	La transformée de Fourier non commutative pour \mathbb{V}_J	1303
10.3.	La transformée de Fourier non commutative pour $\mathbb{V}_S = SE(2)$	1306
10.4.	Confluence entre \mathbb{V}_J et \mathbb{V}_S	1308

11. Inpainting par diffusion sous-riemannienne : le modèle semi-discret	1310
11.1. Un exemple d'inpainting	1310
11.2. Le modèle théorique de reconstruction	1312
11.3. Produits semi-directs semi-discrets	1313
11.4. Résultats de finitude	1315
Chapitre 18. Sur quelques prolongements de la neurogéométrie	1319
1. Le cut locus des contours $2D$	1319
1.1. La réalité neuronale du cut locus	1319
1.2. Squelettisation des formes	1323
1.2.1. D'Harry Blum à René Thom	1324
1.2.2. Après David Marr	1326
1.2.3. Cut locus et fonction distance	1326
1.2.4. Implémentation neuronale	1327
1.2.5. Squelettisation et constituance méréologique	1327
1.2.6. Cut locus multi-échelle	1330
2. Patterns de vision entoptique	1331
2.1. Les premières données de Heinrich Klüver	1332
2.2. Architecture fonctionnelle et réseaux de neurones	1335
2.3. Analyse spectrale et bifurcations	1340
3. Stéréopsie, kinesthésie, motricité	1344
3.1. La neurogéométrie de la stéréopsie	1344
3.2. Le contrôle kinesthésique du regard	1346
3.3. La neurogéométrie de la motricité	1352
4. Extensions nécessaires des modèles	1353
4.1. Stimuli artificiels-simples VS stimuli naturels-complexes	1353
4.2. Multi-fibrations et interactions entre cartes	1354
4.3. Du niveau mésoscopique au niveau microphysique	1354
4.4. Évo/Dévo	1355
Chapitre 19. Conclusion	1357
1. Conscience visuelle et neurogéométrie	1359
1.1. Réductionnisme VS conscience phénoménale	1359
1.2. La conscience visuelle comme émergence	1362
1.3. Retour sur la vision entoptique	1364
1.3.1. L'accord $2 \leftrightarrow 3$	1365
1.3.2. L'émergence $1 \rightarrow 2$ comme corrélat neuronal	1367
1.4. Émergentisme dynamique	1368
2. Esthétique transcendantale et neurogéométrie	1369
2.1. Les deux eidétiques et leurs fondations	1369
2.2. Synthétique a priori et neurosciences	1370

2.3. Neurogéométrie et exposition métaphysique	1371
2.4. Géométrie et réflexion : “décompiler” le calcul neuronal	1372
2.5. Intuitif VS Conceptuel	1374
2.6. Géométrie et physiologie	1374
Bibliographie	1379

CHAPITRE 1

Introduction

Ce volume est la seconde partie de notre traité consacré aux *Éléments de Neurogéométrie*. Il fait suite au premier volume sur les *Architectures fonctionnelles de la vision* dont la traduction anglaise [442] est parue en 2017 chez Springer. Ces deux volumes approfondissent considérablement notre *Neurogéométrie de la Vision* de 2008 [435].

1. Remarques liminaires

1. Les progrès spectaculaires des techniques d'enregistrement et d'imagerie ont transformé le cerveau de "boîte noire" en boîte un peu plus transparente. Cette révolution des moyens d'observation a conduit à une complète transformation de tout ce qui implique des activités mentales puisque l'on peut désormais se proposer d'en explorer scientifiquement les activités neuronales sous-jacentes et/ou corrélatives. D'où la prolifération naturelle et justifiée du préfixe "neuro" devant tout un ensemble de disciplines classiques, qu'elles soient psychologiques, linguistiques, esthétiques, sociologiques, anthropologiques, économiques, etc.

2. Deux aspects essentiels des réseaux neuronaux constituant les différentes aires cérébrales primaires sont leur modularité (leur composition à partir de modules neuro-anatomiquement définissables) et leur connectivité (axones, arbres dendritiques, synapses). Cette dernière est très spécifique pour chaque réseau et définit ce que l'on appelle son "*architecture fonctionnelle*". Si l'on pense l'activité neuronale sous-jacente et/ou corrélative à une activité mentale comme une implémentation dans un hardware, alors on peut dire que la spécificité des architectures fonctionnelles signifie que les hardware neuronaux de ces aires primaires sont *dédiés*.

3. Sur le plan de la modélisation, l'univers des neuromathématiques s'est évidemment considérablement développé. Le domaine d'investigation est immense :

- (i) modèles de champs récepteurs et de profils récepteurs neuronaux inspirés des analyses en ondelettes en théorie du signal ;
- (ii) modèles d'activités sous-limaires et supra-limaires (trains de potentiels d'action) de neurones individuels en termes d'équations différentielles dérivant de celles de Hodgkin et Huxley ou FitzHugh et Nagumo ;

- (iii) énormes systèmes d'équations différentielles modélisant des réseaux neuronaux complexes dont on étudie les propriétés de synchronisation, les rythmes, etc. ;
- (iv) modèles de dynamiques globales décrivant des interactions entre plusieurs aires.

4. Dans cet univers neuromathématique, ce que nous avons proposé d'appeler *neurogéométrie* concerne un aspect très précis et très limité, mais fondamental et problématique, de la perception visuelle. C'est un fait d'expérience que la perception visuelle, avec son rapport constitutif à la motricité, est fortement structurée géométriquement et que sa phénoménologie et ses intuitions sont à la source de nombre de théories mathématiques : la géométrie métrique euclidienne classique depuis l'antiquité, la géométrie projective depuis la Renaissance, la géométrie différentielle depuis le calcul infinitésimal et intégral au XVII^e siècle (cela remonte en fait à Archimède), la géométrie riemannienne (faut-il rappeler les liens entre Riemann et Herbart, ainsi que la réponse de Helmholtz [249] au texte fondateur de Riemann [464]), ou encore la réinterprétation des différentes géométries en termes de théorie des groupes (faut-il rappeler des textes de Poincaré comme [450]). Il y a une énorme masse de travaux successifs, accumulés génération après génération, sur l'enracinement de différentes théories mathématiques dans la phénoménologie de la perception visuelle.¹

Conformément à ce qui a été dit plus haut, toutes ces traditions scientifiques d'une richesse et d'une diversité incomparables doivent être entièrement repensées en termes de fondements neuronaux.

C'est en particulier le cas pour la géométrie différentielle des courbes (et aussi des surfaces mais pour simplifier nous ne parlerons ici que des courbes). Dans les scènes visuelles nous voyons partout des formes délimitées par des contours qui sont des lignes avec des tangentes, des courbures, des points d'inflexion, des points de rebroussement, des coins, des croisements. Nous savons parfaitement décrire toutes ces données locales en termes de calcul différentiel. Nous avons tous appris à l'école comment se calcule la tangente en un point à une courbe comme limite de sécantes ou la courbure comme inverse du rayon du cercle osculateur.

Mais quelle peut être l'implémentation neuronale de tels algorithmes différentiels ? La neurogéométrie concerne la compréhension *des architectures fonctionnelles dédiées implémentant un calcul différentiel neuronal*.

2. Modèles théoriques VS IA computationnelle

La neurogéométrie se focalise donc sur certains modèles mathématiques *théoriques* de certaines structures neuronales de plus en plus richement documentées sur

1. Cf. l'Introduction de notre ouvrage [442].

le plan expérimental. Mais le statut de tels modèles est quelque peu problématique dans le contexte contemporain des neurosciences cognitives. En effet, parallèlement aux progrès considérables des résultats expérimentaux, deux types de techniques se sont massivement déployées.

1. La puissance computationnelle des calculateurs a permis d'envisager l'intégration de systèmes énormes d'équations différentielles de neurones individuels connectés entre eux et de viser une modélisation très fine sans chercher pour autant à en faire émerger des structures mathématiques *théoriques* fondamentales de niveau méso et/ou macro.

2. L'Intelligence Artificielle fondée sur des réseaux de neurones multicouches énormes capables d'apprentissage profond (deep learning) à partir d'un nombre vertigineux de données (big data) a permis une ingénierie de capacités cognitives remarquables, souvent supérieures à celles des spécialistes humains. Dans ces réseaux de neurones, la connaissance n'est pas explicite, conceptuelle et théorique; elle est encodée dans les poids synaptiques des connexions du réseau et donc implicite, numérique et distribuée. C'est une "intelligence collective" de grandes populations de neurones élémentaires.

Le point (1) a suscité de vifs débats dans la communauté internationale des neurosciences cognitives. Il ne serait pas pertinent ici d'entrer vraiment dans la discussion mais citons néanmoins des notes critiques d'Yves Frégnac ancien directeur de l'Unité de Neurosciences, Information et Complexité (UNIC) du CNRS, "Big data and the industrialization of neuroscience" (dans *Science*, [196]) et "How Blue is the Sky" (dans *eNeuro*, [197]). L'auteur discute les deux grands projets européens le *Blue Brain Project* (BBP, 2005-2013) et le *Human Brain Project* (HBP, 2013-...) fondés tous deux sur l'espoir d'une simulation *in silico*, d'une émulation digitale, du cerveau humain. Il s'agit selon lui d'une industrialisation du datamining et de l'exploration du cerveau en neurosciences reposant sur

"the technological prowess of artificial intelligence algorithms and high-information processing".

D'un coté le défi est de simuler les cerveaux biologiques naturels et, d'un autre côté, il est de construire des cerveaux artificiels et virtuels desquels on pourrait, en faisant varier de nombreux paramètres, extraire une foule de nouvelles données quasi-expérimentales permettant d'observer, mais pas forcément de théoriser, l'émergence de structures de niveau méso et macro. Comme le note Yves Frégnac

"The knowledge quest is no longer to find the mechanisms of mind and emerging consciousness, but rather to build a multiscale description of the physical constitution of the Brain."

Des critiques analogues, également très aiguës, du computationnel, des Big Data et de l'IA se trouvent dans le chapitre III du "Cauchemar de Prométhée" [337] de

Giuseppe Longo. Cet ouvrage contient aussi de nombreuses réflexions remarquables sur des problèmes de philosophie des sciences ayant trait aux différences ontologiques entre physique et biologie. Nous ne saurions trop en conseiller la lecture.

3. Géométrie du connectome

En fait, en utilisant une notion assez récente, bien postérieure à nos travaux, on pourrait dire que ce volume 2 vise à proposer des outils de géométrie différentielle à la Lie-Cartan et de géométrie sous-riemannienne du “connectome”² de l’aire $V1$ considéré au niveau *fonctionnel méso*.

Le connectome des aires cérébrales corticales et sous-corticales est l’ensemble de leurs connexions vu comme un système complexe *sui generis*. Pour le cerveau humain, sa notion a été promue par Olaf Sporns, Giulio Tononi et Rolf Kötter en 2005 dans un article “The Human Connectome : A Structural Description of the Human Brain” [512]. Cette connectivité est d’une complexité incommensurable et possède plusieurs niveaux de structuration, mais il est vital de la comprendre car, ainsi que l’expliquent les auteurs,

“The connectome will significantly increase our understanding of how functional states emerge from their underlying structural substrate (and) how human cognitive function emerges from neuronal structure and dynamics.”

1. Au niveau micro, le connectome est le système hypercomplexe des neurones, des axones, des dendrites, des synapses, avec leur cytoarchitecture, leurs différents composants (gaines de myéline, etc.) et leurs nombreux paramètres physiologiques. Il est extrêmement mal connu pour les organismes supérieurs car d’un niveau beaucoup trop fin. Rappelons que le cerveau humain comprend environ 10^{11} neurones et 10^{15} connexions synaptiques.

2. Au niveau macro, le connectome correspond aux faisceaux et aux voies (pathways) reliant entre elles les différentes aires. Il représente une connectivité globale, structurelle et fonctionnelle dont la figure 1 donne une idée. L’image est obtenue par résonance magnétique de diffusion détectant la diffusion de molécules d’eau le long des axones. Cette méthode ne permet pas de visualiser les connexions en tant que telles mais seulement des trajets de diffusion et des champs locaux d’orientations de fibres nerveuses. Des faisceaux de fibres peuvent néanmoins être en partie approximativement reconstruits par des techniques computationnelles de “tractographie” (cf. l’article [345] de Klaus Maier-Hein).

3. Au niveau méso ce sont les patterns de connexions entre petits modules fonctionnellement spécialisés d’une aire ou d’une sous-aire, et aussi des structures comme des “bandes” (“stripes”) et des “blobs” qui deviennent le principal objet d’étude. Les

2. Le connectome regroupe les connexions, leurs faisceaux et leurs voies comme le génome regroupe les gènes.

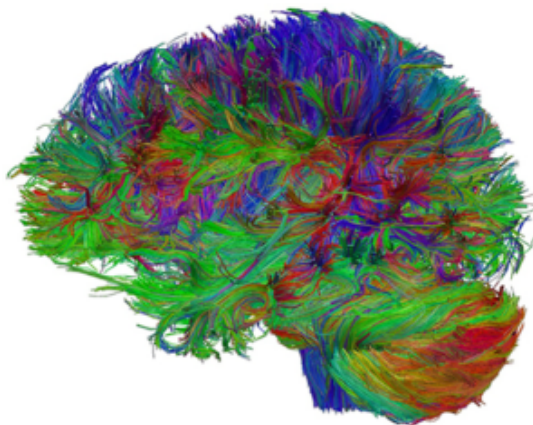


FIGURE 1. Image du connectome humain obtenue par résonance magnétique de diffusion. (GIN-IMN, CEA, CNRS-Université de Bordeaux)

(hyper)colonnes d’orientation de $V1$ traitées dans le Vol I en fournissent un exemple typique. Ainsi que l’expliquent Olaf Sporns et ses coauteurs dans [512]

“Minicolumns may be a sensible choice for neural elements at the mesoscale of the connectome because they may represent basic fonctionnal elements that are crucial for cortical information processing.”

C’est bien ce genre de connectivité fonctionnelle de niveau méso qu’il s’agit de modéliser en neurogéométrie de façon à mieux comprendre la genèse neuronale de la géométrie perceptive. Mais, insistons-y, la perspective est théorique et mathématique plutôt que technique et computationnelle. Non seulement elle conduit à des structures de mathématique pure hautement non triviales mais elle permet de repenser en profondeur le sens de telles structures puisqu’elle permet de commencer à comprendre leur genèse neuronale.

4. La difficulté théorique principale

Le problème clef est de comprendre comment les aires visuelles primaires peuvent implémenter un *calcul différentiel* à partir de données rétiniennes “pixelisées”. Dans la mesure où toute donnée différentielle concerne des petites variations de position, les aires visuelles primaires concernées doivent être rétintopiques et les modèles de leurs architectures fonctionnelles doivent être élaborées sur cette base. Les progrès considérables des méthodes d’enregistrement d’activités neuronales depuis les travaux pionniers de Vernon Mountcastle, David Hubel et Torsten Wiesel et des méthodes de l’imagerie optique *in vivo* (“in vivo optical imaging based on activity-dependent intrinsic signals”) depuis les travaux pionniers d’Amiram Grinvald et Tobias Bonhöffer ont permis d’accéder à des renseignements précis sur ces architectures

fonctionnelles. La question devient donc de comprendre comment celles-ci peuvent se relier au calcul différentiel. Nous allons voir que pour ce faire il faut utiliser, pour des raisons à la fois empiriques et fondationnelles, des formulations non classiques du calcul différentiel, formulations sophistiquées dues à de grands géomètres comme Pfaff, Frobenius, Lie, Darboux, (Élie) Cartan, Weyl ou Goursat.

La difficulté principale vient du fait que les neurones ne détectent pas des données (“features”, “cues”) locales mais plutôt des données *ponctuelles*, ou mieux σ -ponctuelles – “point-like” – à l’échelle σ définie par la taille de leurs champs et profils récepteurs. Beaucoup de profils récepteurs sont des dérivées de Gaussiennes G^3 et l’échelle σ correspond à la largeur de G . Bref, quel que soit leur codage (par taux de décharge de potentiels d’action ou autre) les neurones codent des valeurs numériques σ -ponctuelles. Comme dit Jan Koenderink, ce sont des “point processors”.

Quand on idéalise mathématiquement cette situation on passe à la limite $\sigma = 0$, i.e. à des données véritablement ponctuelles. Or, si l’on s’en tient aux intuitions classiques, il est apparemment *impossible* de construire un calcul différentiel à partir de données ponctuelles finies. Un calcul différentiel neuronal serait donc *antinomique*.

Pour avoir une chance de dépasser cette antinomie, il faut pouvoir résoudre deux problèmes qui plongent jusqu’aux racines du calcul différentiel.

1. Comprendre comment des données différentielles (orientations, directions, tangentes, courbures, etc.), c’est-à-dire locales au sens de multi-ponctuelles peuvent être malgré tout reformulées en termes ponctuels au moyen de *formats* et de structures appropriées. Dans les formulations “sophistiquées” que nous venons d’évoquer, ce problème s’appelle celui du “prolongement”.

2. Comprendre comment ces données différentielles locales prolongeant des données ponctuelles peuvent être intégrées en structures géométriques globales (lignes, bords, contours, formes, etc.). C’est le problème de l’intégration.

Le second problème correspond aux intuitions classiques. Sa difficulté dans notre contexte vient de son lien avec le premier problème qui, lui, est complètement non classique. Il s’agit en effet, répétons-le, de savoir si l’on peut formuler le calcul différentiel dans un format ponctuel.

Jan Koenderink a beaucoup insisté sur ce point crucial :

“Geometrical features [are] multilocal objects, i.e., in order to compute [boundary or curvature] the processor would have to look at different positions simultaneously, whereas in the case of jets⁴ it could establish a format that provides the information by addressing a single location. Routines accessing a single location may aptly be called *point processors*, those accessing multiple locations *array processors*. The difference is crucial in the sense that point

3. Cf. par exemple [431].

4. Voir plus bas la définition du concept crucial de “jet”.

processors need no geometrical expertise at all, whereas array processors do (e.g., they have to know the environment or neighbours of a given location).” ([299], p.374)

C’est bien le point. Les aires visuelles primaires rétinotopiques doivent effectuer leurs tâches géométriques en ne mettant en jeu aucune “expertise” géométrique (aucun “geometrical knowhow”). Mais cela n’est possible que si la géométrie concernée s’y trouve neuralemement câblée de façon à ce que “être activé” devienne équivalent à “faire de la géométrie différentielle”.

Et c’est précisément à cela que sert une architecture fonctionnelle.

5. La Neurogéométrie comme géométrie “from within”

La Neurogéométrie concerne donc la géométrisation *interne* – “immanente” et *ab quo* – des connectivités de la vision de bas niveau, et non pas la géométrie classique “transcendante” et *ad quem*, de l’espace perçu externe. Pour utiliser une belle expression de Mikhaïl Gromov (à propos de la géométrie sous-riemannienne), elle vise à modéliser l’espace perçu *from within*, à partir de l’implémentation neuronale de ses structures constitutives.

5.1. Idéalité transcendantale et matérialisme neuronal

En neurogéométrie, tout ce qui n’est pas neuralemement implémenté n’existe pas et cela implique que tous les concepts mathématiques utilisés opératoirement dans les modèles doivent avoir une contrepartie matérielle. Unifier dans les modèles d’un côté des idéalités géométriques relevant, comme chez Kant, d’une “Esthétique transcendantale” (cadre général des sensations et perceptions), avec, d’un autre côté, des processus neuronaux matériels est un défi scientifique *et* philosophique considérable qui est l’équivalent pour le couple “idéalités géométriques/cerveau” de ce qu’a pu être le défi des ordinateurs pour le couple “idéalités logiques/machine de Turing”.

Qui plus est, sur le plan empirique, les implémentations peuvent varier de façon significative suivant les espèces, et la même structure fonctionnelle abstraite peut être réalisée matériellement de façon différente dans les couches corticales chez le rat, le furet, le tupaya (tree shrew), le douroucoulis (owl monkey), le chat, le macaque, l’homme, etc.

5.2. La “twofold way” de la géométrie

La chaîne d’implications “architectures fonctionnelles \implies propriétés géométriques de l’espace perçu externe” est un processus causal de type “structures neurales \implies genèse d’une spatialité”. Mais si les structures neurales sont elles-mêmes géométrisables au moyen d’autres structures géométriques *sui generis*, alors, une

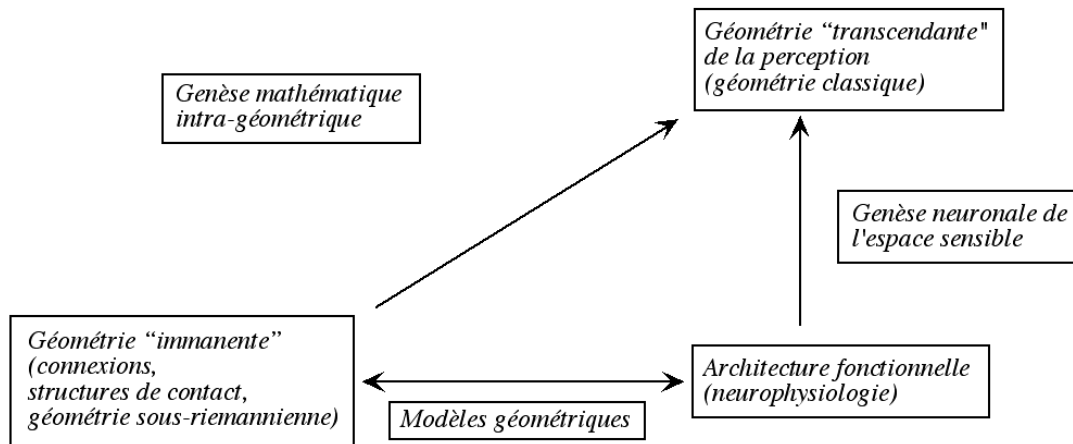


FIGURE 2. Schéma des liens entre géométrie "immanente" et géométrie "transcendante".

fois ainsi modélisée, la genèse neuronale de l'espace externe *s'internalise dans les mathématiques elles-mêmes* pour venir s'identifier à une genèse mathématique d'une géométrie macro, globale et "transcendante" à partir d'une géométrie méso, locale et "immanente" se globalisant par prolongement et intégration. Le schéma de la figure 2 explique cette interaction entre différents niveaux d'opérativité de la géométrie.

Or il se trouve que cette genèse intra-mathématique "from within" de la géométrie différentielle classique a été élaborée à partir de la seconde moitié du XIX^e siècle par certains des géomètres les plus innovateurs de leur temps, ceux que nous avons cités (Pfaff, Frobenius, Lie, Darboux, Cartan, Weyl, Goursat) ou, plus proches de nous, Whitney, Ehresmann, Thom, Mather, Arnold, Gromov, Bennequin, Agrachev, Montgomery, et plusieurs autres⁵.

5.3. Les structures neurogéométriques de base

La neurogéométrie se fonde sur la reconnaissance du fait que tout un ensemble de résultats empiriques très pointus en neurosciences de la vision, en particulier les *cartes d'orientation* du cortex visuel primaire (aires $V1$ et $V2$, appelées aussi aires 17 et 18 chez le chat) sont descriptibles en utilisant des concepts comme ceux (nous allons y revenir très longuement) :

1. de la *fibration* des orientations au-dessus du plan rétinien (structuration des neurones dits "simples" de $V1$ en hypercolonnes d'orientation) ;

5. Pour des précisions sur les géomètres dont la neurogéométrie s'est inspirée, cf [431], [435], [442] et leurs bibliographies.

2. de sa modélisation de base par le fibré \mathbb{V} des 1-jets des courbes régulières par morceaux dans le plan ;
3. de l’action naturelle du groupe $SE(2)$ des déplacements du plan sur \mathbb{V} ;
4. de la *structure de contact* de ce fibré \mathbb{V} des 1-jets ;
5. de la structure du groupe non commutatif \mathbb{H}_{pol} (*groupe de Heisenberg polarisé*) sous laquelle la structure de contact est invariante par translations à gauche ;
6. de la géométrie *sous-riemannienne* naturelle de cette structure de contact \mathbb{H}_{pol} -invariante.

Cela est intéressant au niveau de l’histoire des sciences. Il existe plusieurs exemples bien connus où des résultats empiriques très pointus et des constructions théoriques très spécialisées ont conduit d’abord à reconnaître dans un champ particulier des structures mathématiques profondes déjà élaborées par les mathématiciens purs, puis ensuite à les appliquer dans des modèles. Deux exemples bien connus sont :

- (i) Gravitation et astrophysique \implies idée d’espace courbe chez Einstein \implies reconnaissance de la géométrie riemannienne préexistante et du rôle de la courbure de Ricci \implies modèles d’univers dérivés des équations d’Einstein-Hilbert du champ gravitationnel.
- (ii) Raies spectrales des atomes (formule de Balmer-Rydberg) \implies nouveau calcul introduit par Heisenberg dans son article fondateur de 1925 sur la “Réinterprétation des relations entre Cinématique et Mécanique en Mécanique quantique” [246]⁶ \implies reconnaissance par Max Born et Pascual Jordan qu’il s’agissait d’un calcul non commutatif matriciel \implies modèles des phénomènes quantiques en termes d’opérateurs sur des espaces de Hilbert.

Ici nous introduisons un troisième exemple, évidemment beaucoup plus modeste et limité :

- (iii) Expériences d’enregistrement à partir d’électrodes \implies idée des “engrafted variables” et premières cartes d’orientation et de direction (Hubel et Wiesel, Braitenberg, Swindale, neuro-imagerie) \implies reconnaissance que ces cartes implémentent des fibrés de jets et des structures de contact \implies modèles neurogéométriques.

5.4. Données manifestes “overt” et activités neuronales “covert”

Pour éviter de possibles mésinterprétations de l’ensemble de notre projet, il nous faut insister avec vigueur dans cette Ouverture sur les difficultés majeures, tant scientifiques qu’épistémologiques, que rencontre l’explicitation *générale* des *corrélations* entre, d’un côté, les descriptions de données phénoménales manifestes et, d’un autre

6. “Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen”.

côté, les explications neurocognitives en termes d'activités neuronales "internes" non manifestes, "cachées" dans la boîte cérébrale. On pourrait parler d'une opposition de type overt/covert.

Les descriptions de données phénoménales manifestes peuvent être par exemple des mesures objectives de paramètres cinématiques, dynamiques, etc. des mouvements oculaires, des mouvements de la tête, etc. ou bien des descriptions eidétiques phénoménologiques fondées sur l'analyse réflexive des vécus de l'expérience (celles de Husserl sont admirables)

Les explications neuronales se font quant à elles en termes de propagation de potentiels d'action dans des systèmes de connexions excitatrices ou inhibitrices, systèmes internes à certaines aires ou reliant entre elles des aires différentes et systèmes organisés par des architectures fonctionnelles spécifiques.

Les corrélations sont indéniables entre les descriptions "overt" et les explications "covert" mais leur explicitation soulève des problèmes abyssaux. En effet, elles ne se bornent pas du tout à des correspondances locales du genre "telle petite population de neurones neuromoteurs de tel module mesure par sa 'firing rate' moyenne telle valeur de tel paramètre cinématique ou autre". Les paramètres élémentaires locaux des descriptions cinématiques et autres sont corrélés à des *cascades globales* d'activité de plusieurs aires avec des hiérarchies très compliquées de commandes et de contrôle.

Dans notre ouvrage de 2008 [435], chapitre 13, section 4 nous avons illustré ce fait fondamental en nous appuyant sur l'ouvrage d'Alain Berthoz [53] *Le sens du mouvement*. Nous y reviendrons dans notre partie conclusive à la section 3.2. Le lecteur curieux peut s'y référer dès à présent pour prendre la mesure de la complexité neuronale globale mise en jeu par un simple mouvement local du globe oculaire. Le contrôle kinesthésique des saccades oculaires déclenchées des neurones excitateurs phasiques (*NEP*) est particulièrement fascinant. Il met en jeu le système vestibulaire et sa "centrale gravito-inertielle", les aires médio-temporale et médio-temporale supérieure, le champ oculomoteur frontal, les noyaux dorsolatéral et dorsomédian, les noyaux du tractus optique, le noyau terminal médian, le noyau terminal dorsal. Des neurones inhibiteurs "pauseurs" (*P*) inhibent les *NEP* et pour avoir lieu la saccade doit les inhiber (inhibition d'inhibition). Cela met en jeu le colliculus supérieur (*CS*) et le champ oculomoteur frontal. Mais le *CS* est lui-même contrôlé de façon inhibitrice par les ganglions de la base et l'ensemble de cette structure de contrôle possède une architecture fonctionnelle précise "fibrée" au-dessus de l'espace des directions de l'œil. Il y a en plus un troisième niveau inhibiteur du contrôle et un autre mécanisme inhibiteur régulateur des saccades dépendant lui du cervelet. Et enfin il existe le contrôle cortical des saccades.

Ce réseau global d'interactions suffit à montrer la complexité des corrélations entre les descriptions de phénomènes manifestes et leurs explications en termes d'activité neuronales internes.

5.5. Un phénomène noyau

Notre ouvrage n’abordera pas de telles complexités. Il ne le pourrait d’ailleurs pas car elles sont précisément trop complexes et qu’il faut commencer par un commencement plus simple. C’est pourquoi nous nous sommes focalisés sur le “phénomène noyau” d’une simplicité apparente absolue : qu’est-ce qu’une ligne régulière par morceaux au niveau du cortex visuel primaire ? Quelle est la corrélation géo-neuro entre, d’un côté, la description géométrique classique en termes de tangentes, de courbures, de point anguleux, etc., et, d’un autre côté, les activités neuronales des aires primaires dont les architectures fonctionnelles *sui generis* ont été abondamment analysées dans le Vol I ?

6. Quelques repères historiques du parcours

6.1. Les différentes étapes

Le lecteur intéressé trouvera dans la section 1.3. de la Préface du premier volume et dans nos travaux cités en bibliographie, en particulier la longue étude [440], des détails sur l’histoire de la neurogéométrie et ses différentes étapes. Résumons-les brièvement.

1. Pour nous, le contexte précurseur a été, dans les années 1970, celui des progrès remarquables effectués en théorie des singularités et en dynamique qualitative : espaces de jets (Thom, Mather, Arnold⁷), propriétés de stabilité VS instabilité structurelles, déploiements universels de singularités, structure des systèmes dynamiques compliqués (Thom, Smale, Ruelle), bifurcations et ruptures de symétries (Golubitsky), etc. Les applications en physique et en chimie ont été nombreuses et fondamentales : caustiques en optique (Berry, Bennequin), défauts des milieux ordonnés, phénomènes critiques, transitions de phases (Cf. le colloque de Cerisy de 1982 [339] *Logos et théorie des catastrophes*).⁸ Cela nous a conduit à nos premières applications à la géométrisation de la *phénoménologie de la perception visuelle*.⁹
2. En 1990, un “tournant” vers les neurosciences a été accompli avec notre participation à la création du DEA de Sciences cognitives (EHESS, Paris VI, École Polytechnique) par Michel Imbert. Elle nous a ouvert de riches collaborations avec Yves Frégnac et Jean Lorenceau (UNIC), Alain Berthoz, Jacques Droulez, Chantal Milleret (LPPA), et nous a inspiré de nouveaux

7. Voir l’article principes de Thom [527], l’article [529] et aussi notre compilation [412] et sa bibliographie.

8. Nous avons eu le privilège de suivre en détail ces progrès en suivant les séminaires hebdomadaires de René Thom à l’IHÉS.

9. Une synthèse en anglais de ces premiers travaux se trouve dans [429].

cours et séminaires à l'EHESS puis à Polytechnique. Grâce à Bernard Teissier (spécialiste de théorie des singularités dont nous étions proche depuis le début des années 1970), l'émergence de ces recherches proprement neurogéométriques modélisant l'implémentation neuronale de la phénoménologie de la perception a pu bénéficier de colloques interdisciplinaires à la Fondation des Treilles (1993–94–98). Des liens ont été établis avec le développement des algorithmes d'analyse des signaux en ondelettes (Stéphane Mallat, pour ce thème voir [348]). D'autres liens ont également été tissés avec les travaux de David Mumford (voir [381]), Jean-Michel Morel, Olivier Faugeras, Giuseppe Longo. Notre première synthèse de la neurogéométrie unifiant nos travaux de la décade 1990 a été publiée en 1999 [428].

3. En 2001 une collaboration a été nouée avec Alessandro Sarti et Giovanna Citti de l'Université de Bologne intéressés par le développement de notre modèle géodésique pour les *contours illusives courbes* à la Kanizsa. C'est de cette période que date l'introduction des modèles de géométrie sous-riemannienne ([431], [121]). Une deuxième synthèse de la neurogéométrie a été publiée en 2003 dans le numéro spécial (97, 2 – 3) *Neurogeometry and Visual Perception* du *Journal of Physiology-Paris* (J. Lorenceau, J.P., Y. Frégnac, eds). De nombreuses conférences ont été organisées par Giovanna Citti et Alessandro Sarti (Bologne, Pise) et de nouvelles relations établies avec (entre autres) Jack Cowan et Paul Bressloff, Steve Zucker, Fred Wolf, Remco Duits (pour tous ces thèmes, voir par exemple [79], [45], [159]). Notre collègue Hélène Frankowska nous mis en contact avec Andrei Agrachev et son groupe de théorie du contrôle et de géométrie sous-riemannienne (Jean-Paul Gauthier, Ugo Boscain, Yuri Sachkov, Igor Moiseev, Davide Barilari, Emmanuel Trélat, pour ce thème, voir [10], [69]) et il s'ensuivit de nombreuses nouvelles collaborations avec plusieurs spécialistes de ces disciplines. Comme nous le verrons, elles se trouvent au cœur de ce volume.
4. C'est après des années de cours et de séminaires, après de très nombreux colloques et conférences et après la publication de *Neurogéométrie de la Vision* [435], que nous avons commencé à penser à un traité d'introduction à la Neurogéométrie qui serait beaucoup plus étoffé en amont sur le plan de l'histoire des structures mathématiques et en aval sur celui de développements plus techniques que ceux utilisés dans nos modèles. Il s'agissait d'offrir aux neuromathématiciens une circumnavigation neurobiologiquement inspirée dans ces merveilleux univers géométriques. C'était aussi une façon pour nous de partager avec la communauté des sciences cognitives une partie de l'enchantement que nous ont procuré depuis toujours les œuvres des mathématiciens créateurs.

6.2. Géométrie et Cognition

Le développement de la neurogéométrie dans les années 1990 fait partie d'un intérêt de plus en plus accentué pour les relations entre géométrie et cognition. Le développement des sciences cognitives a conduit à revenir sur tout un ensemble de réflexions d'immenses géomètres comme Bernhard Riemann, Henri Poincaré, Federico Enriques ou Hermann Weyl (et aussi de divers psychologues et physiologistes tout aussi éminents comme Hermann von Helmholtz, les phénoménologues et les gestaltistes) sur les divers modes d'*intuitions* spatiales. Ces réflexions fondationnelles "intuitives" étaient restées assez marginales car l'approche axiomatique des théories exerçait une véritable hégémonie sur la question des fondements des mathématiques. On ne voyait guère quels sens rigoureux donner aux intuitions puisque, même si l'on faisait l'hypothèse (scientifiquement tout à fait acceptable) que leur origine était neuronale on se heurtait à l'obstacle insurmontable du cerveau comme "boîte noire". Il aura donc fallu la révolution des méthodes d'imagerie pour complètement inverser la situation et aborder scientifiquement les fondements intuitifs de la géométrie en tant que manifestation d'architectures fonctionnelles neuronales expérimentalement accessibles.

En 1999 nous avons fondé avec Giuseppe Longo et Bernard Teissier à l'École Normale Supérieure un atelier *Géométrie et Cognition* qui a approfondi ces idées. Son "manifesto" [338] expose l'aller-retour Géométrie \rightarrow Cognition (modèles mathématiques de la cognition) VS Cognition \rightarrow Géométrie (fondements cognitifs des mathématiques). Un Colloque de Cerisy de 2006 *Logique et interaction ; vers une géométrie de la cognition* aborda ces problèmes et Bernard Teissier y proposa une réflexion "Géométrie et Cognition : l'exemple du continu" [525] sur l'"isomorphisme de Poincaré-Berthoz" qui est emblématique de la révolution neurocognitive. L'idée est que la notion géométrique classique de droite s'enracine dans un proto-objet géométrique intuitif au sens Poincaré dont l'implémentation neuronale est fort complexe. Ce proto-objet primitif fusionne la "droite vestibulaire" et la "droite visuelle".¹⁰ La première est définie par les mouvements sans activité vestibulaire, c'est-à-dire les mouvements rectilignes uniformes puisque le système vestibulaire détecte les accélérations. La seconde est définie par le transport parallèle d'une direction. La "fusion" de ces deux "droites perceptives" s'effectue au moyen d'algorithmes neuronaux sophistiqués étudiés par des spécialistes comme Alain Berthoz. Bernard Teissier insiste sur le fait que cette "racine naturelle" et "primale" de la notion mathématique de droite lui donne son *sens*.

De façon générale, il existe plusieurs géométries implémentées dans les modules neuronaux sensori-moteurs et vestibulaires et leurs interactions soulèvent des problèmes neuromathématiques fascinants.

10. Qui est l'objet de base de cet ouvrage.

La référence à Alain Berthoz est scientifique mais aussi institutionnelle. En effet son laboratoire du Collège de France, le LPPA (Laboratoire de la Perception et de l'Action), a été l'un des lieux privilégiés de l'accueil des géomètres (par exemple Daniel Bennequin) et des phénoménologues (par exemple Jean-Luc Petit ¹¹) intéressés par la façon dont les neurosciences pouvaient révolutionner la compréhension des racines intuitives de la géométrie.

7. Résumé du Volume I

Le volume I sera référé comme Vol I. Pour que le lecteur puisse bien coordonner les deux volumes, nous en résumons ici brièvement les chapitres. Nous le faisons aussi comme des prolégomènes à quelques commentaires sur nos choix méthodologiques et pédagogiques.

7.1. Chapitre 1 du Vol I

Le chapitre 1 était une introduction générale portant sur les sciences cognitives et sur la façon dont la neurogéométrie de la vision s'y insère.

7.2. Chapitre 2 du Vol I

Le chapitre 2 explicitait le type très particulier de géométrie que représente la neurogéométrie et la situait dans le contexte des sciences cognitives de la perception, dans celui des systèmes complexes et dans celui de l'histoire de la géométrie. À partir de là, le Vol I a été consacré à des données expérimentales de neurophysiologie ainsi qu'à un premier niveau de modélisation. Ces données expérimentales nombreuses et détaillées sont évidemment *indispensables* à la compréhension des modèles neurogéométriques.

7.3. Chapitre 3 du Vol I

Dans le chapitre 3, nous avons traité des champs et des profils récepteurs des neurones visuels, depuis les photorécepteurs et les cellules ganglionnaires de la rétine jusqu'aux neurones de V1 en passant par ceux du corps genouillé latéral. Nous avons expliqué comment ils agissent sur le signal optique comme des filtres et nous avons abordé le problème de leur linéarité ou non-linéarité. Le traitement qu'ils effectuent du signal dans le cas de linéarité relève en grande partie de ce que l'on appelle en théorie du signal une *analyse en ondelettes*. Nous avons donné quelques précisions sur ce concept fondamental, puis nous avons évoqué la façon dont les profils récepteurs peuvent être interprétés dans le cadre de la théorie de l'information comme un moyen d'optimiser le traitement des images *naturelles*, images dont la statistique est très

11. Cf. leur ouvrage commun [55] "Physiologie de l'Action et Phénoménologie".

particulière. Dans une certaine mesure, le formatage géométrique du signal au moyen d'ondelettes d'un certain type (essentiellement des dérivées de gaussiennes) est une façon d'optimiser la *compression* des images naturelles.

7.4. Chapitre 4 du Vol I

Dans le chapitre 4, nous avons présenté un premier ensemble de données expérimentales sur *l'architecture fonctionnelle* de l'aire $V1$ et, en particulier, sur ce que l'on appelle sa structure en "*pinwheels*" (roues d'orientation). Si l'on simplifie leur activité très compliquée, on peut dire que la majorité des neurones dits "simples" de $V1$ détectent des positions et des orientations¹² dans le champ visuel, ceux détectant les diverses orientations pour une position donnée se regroupant en micromodules fonctionnels anatomiquement définissables appelés des *hypercolonnes* d'orientation organisées en "*pinwheels*". En ce sens, $V1$ implémente une approximation discrète de la *fibration* $\pi : \mathbb{V} = R \times \mathbb{P}^1 \rightarrow R$ ayant pour base le plan rétinien R (que l'on peut prendre isomorphe à \mathbb{R}^2) et pour fibre la droite projective \mathbb{P}^1 des orientations du plan. Cette fibration \mathbb{V} est de dimension abstraite 3 (deux degrés de liberté pour la position et un degré de liberté pour l'orientation).

La notion géométrique de fibration est une première notion géométrique clé qui permet de modéliser l'architecture fonctionnelle "verticale" rétino-géniculo-corticale de $V1$. Nous avons expliqué que la fibration π a pour modèle celle, bien connue des géomètres, de ce que l'on appelle les *1-jets* des courbes du plan $\mathbb{V}_J = \mathbb{R}^2 \times \mathbb{P}^1 \rightarrow \mathbb{R}^2$. L'application rétinotopique projetant les courbes du champ visuel sur des chaînes d'activation de neurones de $V1$ s'identifie ainsi géométriquement à un processus dit de "*relèvement legendrien*" des courbes du plan dans le fibré des 1-jets \mathbb{V}_J .

Une partie importante de ce chapitre 4 a été consacrée au fait que l'implémentation du fibré des 1-jets \mathbb{V}_J dans l'organisation en *pinwheels* de $V1$ repose sur une *réduction dimensionnelle*, réduction nécessaire à la résolution d'un problème de "*collapse*" dimensionnel : les couches bidimensionnelles de $V1$ implémentent en fait une fibration de dimension abstraite (au moins) 3. Une fois réduit, $V1$ apparaît alors comme un champ d'orientations en dimension 2 (appelé "carte d'orientation" par les neurophysiologistes), champ dont les *singularités* sont les centres des *pinwheels*. Nous avons étudié ces singularités, donné leurs formes normales et précisé les distorsions et les défauts de leurs réseaux. Nous avons présenté de nombreuses données expérimentales fines à leur sujet ainsi que quelques hypothèses sur leurs origines évolutionnistes comme par exemple la minimisation des câblages.

Une façon de modéliser ces cartes d'orientation est de les traiter comme des *champs de phases* analogues à ceux que l'on rencontre en optique et dont les singularités (appelées dans ce cas "*dislocations*") ont été bien analysées par des spécialistes

12. Les neurophysiologistes utilisent le terme "orientation" pour désigner les droites du plan avec un angle défini modulo π .

comme Michael Berry. Ces champs sont des superpositions de solutions de l'équation de Helmholtz pour un nombre d'onde relié de façon précise à la maille du réseau de pinwheels. Ils permettent de construire des modèles fort intéressants comme ceux proposés par Fred Wolf et Théo Geisel.

Mais un problème vient du fait que, dans ces modèles de champs de phases, la sélectivité à l'orientation doit être nulle aux singularités. Or de nombreux résultats expérimentaux montrent que ce n'est pas le cas. Pour pallier cette difficulté, nous avons présenté alors un autre modèle reposant sur le concept géométrique d'*éclatement*. Ce modèle alternatif converge, lorsque la maille du réseau de pinwheels tend vers 0, vers le fibré des 1-jets.

Nous avons également exposé la façon dont la fibration modélisant la variable d'orientation interfère avec d'autres fibrations (d'autres "cartes") modélisant d'autres variables comme la direction, la dominance oculaire, la phase, la fréquence spatiale ou la couleur. Pour la fréquence, nous avons présenté en particulier le modèle du dipôle introduit par Daniel Bennequin et ses collaborateurs.

Ce sont donc des fibrations d'assez grande dimension $2 + m$, $m \geq 3$, que $V1$ implémente dans des couches bidimensionnelles. D'où le problème fort intéressant de savoir comment peut bien s'exprimer *l'indépendance* de ces différentes variables. L'une des hypothèses les plus plausibles repose sur un principe de *transversalité* : la transversalité des lignes de niveau de deux cartes indépendantes est maximale (en fait quasiment orthogonale) dans les zones du plan de base où les deux gradients sont *conjointement* forts.

Nous avons conclu ce chapitre par quelques données

- (i) sur le recollement des hémisphères cérébraux à travers les connexions calleuses et
- (ii) sur le traitement primaire de la couleur dans les "blobs" de $V1$.

7.5. Chapitre 5 du Vol I

Dans le chapitre 5, nous avons enrichi et complété les données rassemblées au chapitre 4 en étudiant la *seconde* composante de l'architecture fonctionnelle de $V1$ et des autres aires visuelles, à savoir ce que l'on appelle les connexions "horizontales" *cortico-corticales*. Pour les hypercolonnes d'orientation, ces connexions relient des neurones détectant des orientations en des positions différentes qui sont *parallèles* et même non seulement parallèles mais *approximativement alignées*. Ce "transport parallèle" renforcé par une certaine coaxialité est confirmé par de nombreuses expériences psychophysiques sur ce que David Field, Anthony Hayes et Robert Hess ont appelé le *champ d'association*. Nous en exposons les principes et les résultats.

Nous montrons alors que, en première approximation, le transport parallèle peut être considéré comme une implémentation neurophysiologique de ce que l'on appelle

en géométrie différentielle la *structure de contact* du fibré \mathbb{V}_J des 1-jets de courbes dans le plan \mathbb{R}^2 . \mathbb{V}_J est isomorphe à un espace \mathbb{R}^3 de coordonnées (x, y, p) , (x, y) étant les coordonnées spatiales et p la coordonnée angulaire $\tan(\theta)$. Sa structure de contact se trouve au cœur de l'approche neurogéométrique. Elle est invariante sous l'action du groupe $SE(2)$ des déplacements du plan euclidien. En effet, elle est constituée par le champ de plans \mathcal{K} engendré par l'action de $SE(2)$ sur le plan K_0 des (x, p) passant par l'origine (équation $y = 0$). Comme K_0 est le plan $y = 0$ noyau de la 1-forme différentielle dy , \mathcal{K} est donc le noyau de la 1-forme différentielle $\omega = dy - p dx$ obtenue en translatant dy par les éléments de $SE(2)$.

Qui plus est, la structure de contact est associée à une structure de *groupe non commutatif* sur \mathbb{V}_J , l'action de $SE(2)$ s'identifiant alors aux translations à gauche de \mathbb{V}_J sur lui-même. Ce groupe est isomorphe au *groupe de Heisenberg "polarisé"*, groupe bien connu des physiciens. Muni de son champ de plans de contact, c'est un groupe nilpotent appartenant à la classe de ce que l'on appelle les groupes de Carnot. Nous avons utilisé ce fait pour transférer à la neurogéométrie un ensemble important de beaux résultats de physique mathématique.

La définition d'une métrique $SE(2)$ -invariante (c'est-à-dire invariante par translations à gauche) sur les plans de contact définit alors ce que l'on appelle une géométrie *sous-riemannienne*. Nous avons ainsi montré que, dans cette première approche neurogéométrique, le cadre mathématique naturel de la modélisation de l'architecture fonctionnelle de $V1$ est fourni par la géométrie de contact de \mathbb{V}_J et la géométrie sous-riemannienne qui lui est étroitement associée. Ce type de géométrie sous-riemannienne, aussi dite de Carnot-Carathéodory, a été profondément développée par de nombreux géomètres. Pour notre part, Misha Gromov et d'autres spécialistes comme Andrei Agrachev, Richard Beals, André Bellaïche, Ugo Boscain, Giovanna Citti, Jean-Paul Gauthier, Bernard Gaveau, Peter Greiner, Richard Montgomery, Pierre Pansu, Yuri Sachkov, Alessandro Sarti, Robert Strichartz, nous ont beaucoup inspiré.

La géométrie sous-riemannienne de la structure de contact est au cœur de ce second volume. Elle permet d'expliquer de façon naturelle certains phénomènes énigmatiques comme par exemple celui de la construction par le système visuel de *contours illusoires* à très longue portée. Ces phénomènes de complétion de données sensorielles lacunaires par les mécanismes intégratifs de la perception sont connus depuis le début de la théorie de la Gestalt et ont donné lieu, en particulier après les célèbres travaux de Gaetano Kanizsa, à d'innombrables recherches et modèles. En ce qui concerne les modèles, on est naturellement conduit à des modèles *variationnels*. Nous proposons quant à nous un modèle dans le fibré de contact \mathbb{V}_J qui repose sur l'hypothèse que les contours illusoires sont des *géodésiques* sous-riemanniennes.

Après cette première analyse de la géométrie de contact de $V1$ et de la géométrie sous-riemannienne qui lui est associée dans le modèle \mathbb{V}_J , nous avons dit un mot

du “*prolongement*” de cette structure que l’on appelle la structure d’Engel et qui correspond aux *2-jets* des courbes dans le plan \mathbb{R}^2 . Ce prolongement doit être pris en compte si l’on veut pouvoir rendre compte des données expérimentales montrant qu’il existe dans le système visuel primaire non seulement des détecteurs d’orientation (tangentes et 1-jets) mais également des détecteurs de *courbure* (cercles osculateurs et 2-jets).

Pour conclure ce chapitre 5, nous avons indiqué quelques propriétés de l’architecture fonctionnelle des aires $V2$, $V4$ (pour la couleur), $V5$ ou MT (pour le mouvement). Nous discutons également le modèle de Swindale pour les *directions*.

Enfin nous avons donné quelques informations sur le contrôle génétique de la *morphogenèse neuronale* des architectures fonctionnelles dont nous avons analysé la géométrie. La neurogenèse des voies visuelles et le guidage axonal sont des merveilles d’auto-organisation qui permettent d’entrevoir les racines de l’origine des représentations spatiales.

7.6. Chapitre 6 du Vol I

Pour conclure ce premier volume nous avons effectué dans un sixième et dernier chapitre une transition vers les thèmes mathématiques qui sont développés dans ce second volume.

Nous y esquissons d’abord le calcul explicite des géodésiques du modèle \mathbb{V}_J en utilisant les outils de la théorie du contrôle. Nous adaptons pour cela au cas du groupe de Heisenberg *polarisé* les résultats effectués pour le groupe de Heisenberg \mathbb{H} par une pléiade de spécialistes comme ceux déjà cités et d’autres comme Jean-Michel Bismut, Luca Capogna, Vladimir Gershkovich, John Mitchell, Jean-Jacques Risler ou Anatoly Vershik à la suite des travaux pionniers de maîtres comme Lars Hörmander, Elias Stein ou Gerald Folland. On peut ainsi calculer assez simplement tous les éléments de la géométrie sous-riemannienne de \mathbb{V}_J : géodésiques, sphère unité, front d’onde, caustique, cut locus, points conjugués, etc.

Toutefois, le modèle le plus naturel n’est pas celui, \mathbb{V}_J , construit sur l’espace des 1-jets avec l’action de $SE(2)$, mais celui, \mathbb{V}_S , construit sur $SE(2)$ lui-même. C’est le passage, standard en géométrie, d’un fibré sur lequel opère un groupe au fibré *principal* associé. Avec Alessandro Sarti et Giovanna Citti, nous avons insisté sur l’importance de ce second modèle. Or, le groupe $SE(2)$ des déplacements du plan, dit aussi groupe des roto-translations ou “shift-twist action”, n’est pas nilpotent. Sa “nilpotentisation” qui définit son “cône tangent” à l’origine est bien isomorphe au groupe de Heisenberg polarisé \mathbb{H}_{pol} , mais, globalement, sa géométrie sous-riemannienne est très différente. Nous avons constaté avec étonnement qu’elle n’était pas connue jusqu’ici bien que $SE(2)$ soit pourtant un groupe tout à fait classique et élémentaire. À la suite de discussions, Andrei Agrachev et ses collaborateurs

(en particulier Yuri Sachkov, Jean-Paul Gauthier, Ugo Boscain et Igor Moiseev) ont explicité cette géométrie. Nous résumons leurs calculs de géodésiques.

Nous nous arrêtons également sur le fait que les géométries de contact sous-riemanniennes modélisent ici une architecture fonctionnelle de connexions entre des neurones qui agissent comme des filtres. Cela signifie que le cadre mathématique naturel de la perception visuelle de bas niveau est celui où *l'analyse harmonique non commutative* sur le groupe $SE(2)$ se relie à sa géométrie sous-riemannienne. Comme l'analyse harmonique s'effectue à travers une ondelette "mère" translatée par $SE(2)$, c'est-à-dire par ce que l'on appelle un *état cohérent*, c'est donc le lien entre états cohérents et géométrie sous-riemannienne qui s'impose comme cadre théorique naturel de la vision de bas niveau.

Puis nous indiquons comment le modèle variationnel des *elasticæ* et son interprétation *stochastique* proposés par David Mumford peuvent être reformulés dans ce nouveau contexte. Comme l'ont montré Giovanna Citti, Alessandro Sarti, Remco Duits et Markus van Almsick, cela conduit à des algorithmes d'advection-diffusion décrits par une équation de Fokker-Planck que l'on peut calculer explicitement pour le modèle \mathbb{V}_J (le calcul dans \mathbb{V}_S est nettement plus compliqué).

Ces techniques relèvent de la théorie générale du *noyau de la chaleur* (heat kernel) pour les laplaciens *hypoelliptiques* des variétés sous-riemanniennes. La littérature à leur sujet est énorme depuis les travaux pionniers d'Hörmander. Nous indiquons en particulier les résultats de Giovanna Citti et du groupe d'Andrei Agrachev sur les laplaciens hypoelliptiques et les noyaux de la chaleur des modèles \mathbb{V}_J et \mathbb{V}_S . Il faut pour cela expliciter les représentations unitaires irréductibles de ces groupes (*i.e.* leur espace *dual* muni de la mesure de Plancherel) et les transformées de Fourier des laplaciens. Cela fait intervenir les *fonctions d'Hermite* dans le cas de \mathbb{V}_J et les *fonctions de Mathieu* dans le cas de \mathbb{V}_S .

Nous indiquons enfin comment l'on peut interpoler entre \mathbb{V}_J et \mathbb{V}_S au moyen d'une famille *continue* de modèles sous-riemanniens.

8. Nos choix méthodologiques et pédagogiques

Disons maintenant un mot sur la façon assez particulière dont est structuré ce volume.

8.1. Un marteau-pilon pour casser une noisette ?

La première chose qui interrogera sans doute le lecteur peu familier de la modélisation dans les sciences très formalisées sera la mobilisation de structures mathématiques d'assez haut niveau pour modéliser des phénomènes visuels somme toute assez simples comme la détection de contours, même si lorsqu'il s'agit de contours virtuels la "simplicité" est moins évidente. Mais ce serait oublier que cette faculté est en

fait le résultat d'une évolution extraordinairement longue et complexe depuis l'apparition au précambrien des premières cellules photosensibles grâce aux molécules d'opsines. Considérer la phylogenèse du système visuel suffit pour se convaincre que la mathématisation des algorithmes visuels apparemment les plus simples n'a aucune raison d'être elle-même simple. Comprendre comment des organismes multicellulaires génétiquement codés sont devenus à même, par une suite immense de mutations génétiques et d'adaptations, de percevoir les lignes est un défi abyssal.

Il n'y a donc pas lieu de s'étonner d'une certaine complexité.

8.2. Modèles particuliers et structures générales

La clé de notre méthodologie sera le dialogue entre modèles particuliers et structures générales. Nos modèles neurogéométriques dérivés des données expérimentales sont des cas particuliers et spécifiques de concepts et de structures géométriques généraux beaucoup plus riches, profonds et techniques. En mathématiques, les spécialistes cherchent souvent la généralité et donc la complexité abstraite devient vite très grande. Ici nous traiterons assez en détail les cas particuliers dont nous avons besoin pour la modélisation en effectuant explicitement certains calculs techniques conceptuellement conformes à ces structures générales. Nous espérons qu'ils permettront à des lecteurs motivés surtout par la modélisation en mathématiques appliquées de s'initier à ces arcanes de mathématiques pures. Nous choisissons donc une pédagogie par l'exemple qui permet de surmonter bien des écueils de compréhension, tant il est vrai qu'un bon exemple remplace de longs discours.

Mais nous insisterons également sur les aspects plus conceptuels et structurels. Le rapport aux mathématiques se doit d'être à la fois calculatoire et structural, la dialectique entre structures et calculs étant essentielle. On affirme souvent que le structuralisme a été introduit par Galois qui parlait de

“sauter à pied joint sur les calculs, grouper les opérations, les classer suivant leur difficulté et non suivant leur forme”,

puis développé par l'Ecole allemande (Dedekind, Hilbert, Noether, Artin, etc.) jusqu'à l'école française de Bourbaki.¹³ Ce n'est qu'en partie vrai. Car quand on veut expliciter un cas particulier, il faut en revenir de façon constructive à des calculs.

13. Sur le structuralisme en mathématiques, le lecteur pourra consulter notre étude [414] sur la philosophie mathématique d'Albert Lautman. et notre texte [245] avec Gerhard Heinzmann sur le rôle fonctionnel des structures chez Bourbaki. Pour deux exemples approfondis, il pourra consulter notre réflexion [414] sur l'étonnante démonstration de la conjecture de Taniyama-Shimura-Weil par Andrew Wiles en 1993-1994 ainsi que notre résumé [442] des profondes propositions d'Alain Connes pour la démonstration de l'hypothèse de Riemann.

8.3. Une perspective propédeutique

Ceci dit, tant les modèles particuliers que les structures générales feront intervenir des outils géométriques assez peu connus des biologistes, même de ceux qui ont une formation mathématique. Et si l'on veut s'y initier à partir de traités de géométrie différentielle, de groupes de Lie, de géométrie sous-riemannienne ou d'analyse harmonique, même introductifs, destinés à des mathématiciens, on se heurte à un mur infranchissable de technicité.

La finalité de cet ouvrage est donc en partie *propédeutique*. Il voudrait essayer de rendre accessible à des neurobiologistes (et éventuellement à d'autres spécialistes de sciences naturelles) ayant un bagage préliminaire suffisant ce genre de mathématiques déjà assez "avancées". Pour cela il a besoin d'effectuer un certain nombre de *compilations* de résultats existants mais des compilations orientées par leurs *nouvelles* applications neurophysiologiques, applications qui exigent des *transferts* de formalismes.

Les ouvrages propédeutiques motivés par des transferts de formalismes déjà existants vers de nouvelles applications dans de nouveaux domaines empiriques sont assez fréquents. Citons quelques exemples.

(i) La relativité générale a conduit à des ouvrages d'introduction à des aspects sophistiqués de la géométrie riemannienne (connexions de Levi-Civita, courbure de Ricci, identité de Bianchi, etc.), déjà bien connus des géomètres depuis longtemps, mais visant des applications cosmologiques originales.

(ii) La mécanique quantique a conduit à des ouvrages d'introduction à des aspects sophistiqués de l'algèbre linéaire et des opérateurs dans les espaces de Hilbert (analyse spectrale, représentations de groupes, etc.), déjà connus des algébristes et des analystes depuis longtemps, mais visant des applications microphysiques originales (raies spectrales des atomes, etc.).

(iii) Les réseaux de neurones ont conduit à des ouvrages d'introduction à des aspects sophistiqués de la physique statistique (verres de spins, etc.), déjà bien connus des physiciens depuis longtemps, mais visant des applications neurocognitives originales (apprentissage, catégorisations, classifications, etc.).

On pourrait dire que ce volume est un ouvrage d'introduction à des aspects sophistiqués de la géométrie des formes différentielles de Cartan et des groupes de Lie, déjà bien connus des géomètres depuis longtemps, mais visant des applications neuroperceptives originales.

8.4. Ouvrages de référence et/ou traité didactique

De par la nature même des applications empiriques que nous attendons ici des modèles mathématiques, il n'est pas possible de présenter ces modèles sous la forme

d'un traité didactique de mathématique scolairement organisé. La logique structurant ce volume est celle des données expérimentales et ne peut donc pas être déductive. À chaque étape, il faut introduire de nouveaux outils formels (formes différentielles et structures de contact, groupes de Lie et leurs représentations, modèles variationnels et géométries sous-riemanniennes, analyse harmonique non commutative, états cohérents et diffusion, etc.) et si tous ces thèmes devaient être présentés dans un traité de mathématique ils ne le seraient certainement pas de cette façon et dans cet ordre. Nous nous en excusons auprès des mathématiciens mais donner la priorité à la nature de l'objet empirique nous a semblé légitime.

Pour compenser ce défaut de systématisme déductif dans la présentation des structures formelles mobilisées, nous insistons sur leurs inventeurs, leur histoire et leurs sources en donnant une place importante à de nombreux ouvrages de référence, même anciens (et encore en latin) comme ceux d'Euler ou de Pfaff.

Nous donnons également une place importante à des synthèses historiques et conceptuelles effectuées par des mathématiciens passionnés de pédagogie et qui nous ont semblé particulièrement remarquables. Certaines sont des "traités" et des "leçons", voire de véritables "bibles", qui sont plus que des cours universitaires.

8.5. L'encadrement des modèles en amont et en aval

En fait, nous encadrons les modèles de deux façons. D'abord par tout un ensemble de références historiques. Il ne s'agit pas de s'engager dans la très riche historiographie érudite des structures mathématiques utilisées mais simplement de donner une idée de leur "généalogie". Cela nous paraît être essentiel pour comprendre la profondeur de ces innovations mathématiques depuis l'invention du calcul intégral-différentiel par Newton et Leibniz.

Cette volonté de "retour amont" nous conduira à revenir sur certains textes fondateurs assez anciens, textes bien connus dans leur fond mais sans doute peu connus du lecteur dans leur lettre. Nous pensons que cela peut être très utile pour deux sortes de raisons.

(i) Des raisons pédagogiques tant il est vrai que la genèse d'une idée fait intimement partie de la compréhension de son contenu.

(ii) Des raisons de fond. L'applicabilité de structures géométriques à la production *neuronal* interne de la géométrie externe ne va vraiment pas de soi et soulève les innombrables questions tant scientifiques qu'épistémologiques qui se tressent dans cet ouvrage. En retournant à la genèse de ces structures nous constaterons souvent que leurs inventeurs avaient la conscience aiguë de revenir aux sources les plus profondes, les plus cachées et jusque-là les plus incomprises de la géométrie.¹⁴ Cela nous permettra de mieux comprendre les raisons de l'applicabilité.

14. "Enfouies dans les profondeurs de l'âme" comme disait Kant. Nous reviendrons plusieurs fois sur cette expression.

Mais d'un autre côté nous souhaitons aussi permettre au lecteur d'avoir une idée des théories générales, en général très difficiles, qui prolongent nos modèles et constituent une part importante de la géométrie moderne. Nous espérons ainsi que, doublement éclairé par un amont historique et un aval spécialisé, le lecteur pourra mieux évaluer la portée des modèles

Ces références prendront souvent, répétons-le, la forme d'une compilation. Les informations qu'elles condensent sont assez facilement accessibles mais complètement dispersées dans des articles et des ouvrages spécialisés en général inaccessibles aux non spécialistes.

8.6. Un guide cartographique pour une circumnavigation

On peut donc considérer ces "Éléments" comme une sorte de guide cartographique pour une longue circumnavigation qui pourrait permettre à certains spécialistes de neurosciences de s'aventurer sur l'océan des géométries pour rejoindre quelques îles enchantées sans se perdre, faute de cartes, dans des immensités sans repères.

8.7. L'usage d'une conception "internaliste" des mathématiques

Notre choix de suivre un fil historique menant d'un amont (Leibniz-Euler) vers l'aval (travaux contemporains) fait usage d'une conception plutôt "internaliste" des mathématiques puisqu'elle considère qu'il existe des problématiques difficiles et hérissées de difficultés (par exemple l'intégration des équations différentielles) dont la compréhension et la résolution progressent parfois sur une longue durée. Cela est particulièrement flagrant en mathématiques où beaucoup de problèmes fondamentaux sont séculaires, voire millénaires. On peut suivre le fil des découvertes et des inventions et voir évoluer les hypothèses, les concepts, les théories en fonction de la situation interne des problématiques.

Cela ne veut évidemment pas dire que les nombreux facteurs externes (institutionnels, politiques, culturels, etc.) dans le contexte desquels se développent les problématiques et leur logique interne ne sont pas importants. Mais ils sont peu pertinents pour notre propos dans la mesure où il s'agit pour nous de comprendre de quelles structures géométriques nous avons besoin pour géométriser les architectures fonctionnelles en neurosciences.

Mais nous encourageons le lecteur, à partir des quelques indications que nous donnerons, par exemple dans la section 2.2 du chapitre 4 à propos du "Problème de Pfaff", à se plonger dans la biographie des auteurs que nous mobiliserons. Elle est souvent particulièrement intéressante et traversée de façon poignante par la "grande histoire".

Par exemple, on voit bien dans la vie de Sophus Lie s'entremêler les conflits des diverses Écoles géométriques tant en France qu'en Allemagne, conflits de nature aussi bien théorique ("géométrisation" des équations différentielles VS "arithmétisation" de la géométrie, paradigmes "analytiques" VS paradigmes "synthétiques") qu'institutionnels (l'"école" de Paris, l'"école" de Berlin, l'"école" de Göttingen, etc.), tout cela mêlé à la guerre franco-allemande de 1870. Mais ce qui nous importe est le fil "interne" conduisant de Monge à Cartan à travers une suite d'innovations prodigieuses même si celles-ci étaient immergées dans des contextes "externes" institutionnels, politiques, culturels extrêmement divers et souvent conflictuels.

8.8. Une pédagogie mathématique inhabituelle et risquée

Cette stratégie commandera nos choix pédagogiques. Nous traiterons un peu en détail les cas particuliers et nous indiquerons quelques éléments des démonstrations dans les cas les plus simples. Nous indiquerons les cas où les modèles concrets sont facilement généralisables et aussi les cas où l'investissement théorique nécessaire à la généralisation est beaucoup plus difficile. Dans ces derniers cas, nous ne pourrions évidemment pas donner de démonstration des grands théorèmes structuraux, ce serait trop spécialisé. Mais nous essayerons néanmoins d'en donner quelques idées.

Évidemment, ce choix pédagogique par l'exemple faisant que certaines notions générales de base ne seront introduites que *bien après* certains exemples heurtera sans doute les vrais mathématiciens qui préfèrent un exposé déductif. Mais il est fonctionnel pour notre propos et nous a été imposé par notre expérience. Évidemment, une présentation non déductive prenant pour guide ce dont on a besoin pour les modèles concrets, ainsi que de nombreux retours amonts historiques et de nombreux excursions vers des problèmes contemporains plus difficiles ont l'*inconvenient* d'obliger à quelques rétroactions dans le décours du texte. Mais ils ont aussi le grand *avantage* de permettre "d'immerger" *progressivement* la neurogéométrie dans son contexte naturel de géométrie différentielle. Nous voudrions que des spécialistes de neuromathématiques puissent ainsi s'initier à des concepts profonds et des techniques difficiles qui ont montré leur extraordinaire puissance en physique.

Nous avons conscience de prendre le risque de mécontenter tout le monde. Les mathématiciens trouveront le propos mal structuré au niveau de la présentation et trop trivial sur le plan du contenu alors que les neurobiologistes trouveront le propos mathématiquement trop technique et expérimentalement trop réduit et élémentaire.¹⁵ Mais la "dérive des continents" entre les disciplines est telle qu'il faut bien prendre des risques si l'on veut jeter un pont entre les rives opposées.

15. Nous parlons en effet surtout de la rétine et de *V1* à partir de données déjà jugées trop simples voire simplistes et nous ne disons presque rien sur les innombrables données concernant les autres aires.

8.9. Réquisits

Les réquisits mathématiques du volume sont standard, de niveau licence¹⁶.

1. théorie des ensembles : applications, injections, surjections, bijections ;
2. espaces topologiques : ouverts/fermés, continuité, recouvrements ouverts, homéomorphismes, complexes simpliciaux ;
3. variétés différentiables : dérivées, dérivées partielles, applications différentiables, applications linéaires tangentes ;
4. espaces métriques, isométries ;
5. groupes, sous-groupes, groupes quotient, anneaux, corps, idéaux ;
6. algèbre linéaire : espaces vectoriels, matrices, produits tensoriels, formes linéaires.

Des notions plus avancées comme celles de fibration et de section ou d'éclatement, ainsi que des structures essentielles comme les espaces de jets ou le groupe $SE(2)$ des déplacements du plan produit semi-direct $\mathbb{R}^2 \rtimes SO(2)$ sont définies dans le Vol I et redéfinies ici.

8.10. Beaucoup de calculs pour rien ?

Il y aura dans ce volume nombre de calculs explicites. Ils restent souvent implicites dans les articles et ouvrages de mathématiques lorsqu'ils sont jugés être trop élémentaires. Ils sont cependant en général très instructifs. En les refaisant intégralement, nous avons vérifié qu'ils sont tous effectuels avec un bagage mathématique de premier cycle.

Cette attention portée aux calculs fait partie de nos choix pédagogiques concernant la dialectique entre modèles particuliers et structures générales. En dehors de quelques grands théorèmes pour nous cruciaux nous développons peu les démonstrations de résultats structurels généraux. En revanche nous développons le détail calculatoire de certains modèles particuliers.

Le lecteur neurobiologiste pourra se demander quelle peut bien être l'utilité de tels calculs puisque les données expérimentales accessibles ne sont pas suffisamment précises pour bien les valider numériquement. On peut apporter à cette réserve plusieurs réponses.

1. C'est de façon tout à fait naturelle et neurophysiologiquement justifiée que l'on en arrive à certaines équations différentielles précises et il est alors tout à fait normal de chercher à les résoudre, même si cette résolution est incroyablement compliquée. C'est une banalité dans les sciences naturelles mathématisées : les données

16. Cf. par exemple l'indétrônable cours de licence de Jean-Pierre Ramis et André Warusfel [459].

expérimentales et les réflexions théoriques conduisent à des équations qu'il est souvent difficile, voire impossible, d'intégrer. Mais personne n'expliquera à Laplace ou à Poincaré qu'il est sans intérêt de comprendre quelque chose au problème gravitationnel à trois corps du Soleil, de la Terre et de la Lune.

2. D'ailleurs, même au niveau géométrique et arithmétique le plus élémentaire et le plus intuitif, la complexité est immédiatement présente. Qu'il y a-t-il apparemment de plus simple que le groupe des rotations dans \mathbb{R}^3 ? Et pourtant ce groupe compact non commutatif est déjà très subtil comme tout le monde peut l'expérimenter en jouant avec le Rubik cube. Et que dire de la distribution des nombres premiers qui hantent les mathématiques depuis les grecs!

3. Les équations différentielles que nous rencontrerons sont assez analogues à celles qu'ont rencontrées les premiers grands spécialistes du calcul différentiel et intégral lorsqu'ils ont formulé dans ce nouveau cadre formel des problèmes géométriques classiques (par exemple "rectifier" un arc d'ellipse) ou des problèmes mécaniques (par exemple le mouvement d'un pendule). Ces problèmes étaient apparemment simples mais leur explication s'est révélée être d'une complexité insoupçonnée. Il en ira de même pour nous ici.

4. Affronter la complexité computationnelle de modèles de phénomènes empiriques avec une précision supérieure à ce que permet de tester l'observation expérimentale peut être pragmatiquement utile lorsque l'on passe du naturel à l'artificiel et au technologique.¹⁷ Certains cas seront ici assez spectaculaires. Pour ne prendre qu'un exemple, nos modèles sont en grande partie des modèles de la façon dont, à cause de leurs architectures fonctionnelles spécifiques, les aires visuelles primaires *complètent* les inputs sensoriels en percepts. Lorsque l'on passe à la vision artificielle bio-inspirée et au traitement d'image, la complétion devient le problème de l'*inpainting* et il est utile de disposer d'algorithmes computationnellement puissants. Nous verrons que nos modèles permettent des complétions apparemment impossibles comme celle de la figure 1 du chapitre 17. De telles performances montrent que les calculs ne se font pas pour rien.

5. Et puis il y a une autre raison, toute différente et plus subtile, à l'intérêt d'explicitier les calculs. La forme précise des calculs est un guide pour comprendre comment ils peuvent être effectués en étant implémentés dans un hardware *dédié*, ici une architecture fonctionnelle neurale spécifique.

17. Il n'est peut-être pas très utile pour la vie de tous les jours de pouvoir calculer avec une extrême précision les trajectoires de notre système planétaire, mais cela est certainement nécessaire si l'on veut envoyer la sonde Rosetta (mission de l'ESA l'Agence Spatiale Européenne, démarrée en 2004) déposer Philae sur sa comète cible 67P/Churyumov-Gerasimenkola en effectuant pendant une dizaine d'années quatre manœuvres de "fronde gravitationnelle" ("slingshots", trois avec la Terre et un avec Mars), pour arriver enfin sur l'orbite de Jupiter à environ cinq cents millions de kilomètres du Soleil, avant d'orbiter autour de la comète.

6. Ceci dit, la technicité un peu rébarbative de certains calculs sera tempérée par quelques figures permettant de visualiser, lorsque c'est possible, les structures géométriques en jeu.

8.11. Interdisciplinarité et polyscientificité intrinsèques

Comme nous le verrons de façon récurrente, les formalismes que nous mobiliserons seront des formalismes qui, souvent, ont été mis au point dans le passé pour expliquer des phénomènes physiques. Nous pourrions laisser cette interdisciplinarité des modèles en arrière-plan, mais nous avons au contraire choisi de la mettre en lumière, et cela pour des raisons précises.

De façon générale, les applications des mathématiques sont interdisciplinaires et des modèles semblables apparaissent dans des domaines empiriques extrêmement différents. Cela ne signifie évidemment pas qu'il existe une parenté ontologique entre ces domaines, mais cela témoigne d'une communauté de structures formelles. Selon nous, la compréhension de l'interdisciplinarité et de la polyscientificité intrinsèques de ces dernières est épistémologiquement essentielle. L'hétérogénéité empirique va de pair avec une unité formelle et n'est pas un éclectisme.

Le lecteur ne s'étonnera donc pas de rencontrer des sections sur les équations de Maxwell de l'électromagnétisme, ou sur la relativité générale et le problème de l'espace à la Einstein-Cartan-Weyl, ou sur la mécanique quantique, son groupe de Heisenberg et ses opérateurs auto-adjoints dans des espaces de Hilbert, ou encore sur les équations de diffusion et les noyaux de la chaleur. Nous expliquerons pourquoi ces formalismes apparaissent de façon naturelle dans la géométrisation des connectivités neuronales, la construction de contours illusoires ou l'analyse du signal par des profils récepteurs. Notre espoir est de convaincre que les neurosciences en fournissent un immense domaine d'application.

8.12. Neurogéométrie et formalismes physiques

Précisons un peu ce point. Étant donnée la nouveauté des données expérimentales qui ne sont accessibles que depuis la révolution de l'imagerie, la neurogéométrie en est à un stade initial analogue à celui de la mécanique dans la première moitié du XVIII^e siècle, celle des Bernouilli et des Euler. Nous le répéterons plusieurs fois comme un leitmotiv. À l'époque il s'agissait d'intégrer les nouvelles équations différentielles qui permettaient de calculer des formes naturelles mécaniques aussi "simples" que la forme d'une "chaînette", le mouvement d'un pendule ou la déformation élastique d'une barre métallique. C'était un énorme défi. Ces phénomènes empiriques "simples" ne pouvaient être formalisés que par des structures compliquées (comme les fonctions elliptiques) que les scientifiques de l'époque ont dû découvrir.

La simplicité empirique n’a souvent rien à voir avec la simplicité formelle. Nous allons retrouver cette grande vérité de la mécanique classique dans la neurogéométrie d’aujourd’hui. La découverte des architectures fonctionnelles visuelles fournit un nouveau cadre théorique, conceptuel et mathématique, dans lequel modéliser des phénomènes simples comme les extractions de bords délimitant des objets dans des images ou des contours illusoire courbes. Les travaux de David Mumford ou de Yuri Sachkov sur les *elasticæ* en vision répondent, deux siècles et demi après, à ceux d’Euler sur les *elasticæ* en mécanique.

8.13. La vertu pédagogique d’un peu de redondance

Étant donnés nos choix pédagogiques de présentation de nos modèles et de leurs structures mathématiques, avec la priorité donnée au fil directeur de la neurophysiologie, aux rétrospections vers l’amont et aux prospections vers l’aval, nous aurons à revenir de façon récurrente sur certains thèmes au fur et à mesure qu’ils seront repris pour être approfondis tant dans leur contenu que dans leur contexte en suivant le cours naturel de la neurophysiologie. Cela impliquera naturellement une certaine *redondance*, et parfois même de courtes répétitions.

Mais notre expérience d’enseignement dans le DEA de Sciences cognitives et à l’École Polytechnique nous a montré qu’une certaine redondance possède une vertu pédagogique. Même si elle n’est pas habituelle dans les traités de mathématiques didactiques purement déductifs, nous n’hésiterons donc pas à y avoir recours quand cela semblera pertinent.

8.14. Les notations

En ce qui concerne les notations il faut gérer l’équilibre entre l’explicitation complète et la lisibilité. Très souvent les notations complètes seraient un peu lourdes mais heureusement le contexte permet de ne pas toujours tout expliciter. Si par exemple une entité E dépend de plusieurs variables x, y, z et de plusieurs paramètres a, b, c , il faudrait partout écrire $E(x, y, z; a, b, c)$. Mais lorsque le contexte est clair, on écrira E ou $E(x)$ si seule la variable x est pertinente dans le contexte, ou $E(b)$ si seul le paramètre b est pertinent, etc.

Par ailleurs nous essayons de garder un équilibre entre d’un côté des notations propres à des modèles particuliers et d’un autre côté des notations conventionnelles de structures générales. Par exemple nous notons V une variété différentiable modèle de l’aire visuelle primaire $V1$. Mais une notation conventionnelle pour les variétés différentiables est plutôt M (“Mannigfaltigkeit” en allemand, “manifold” en anglais). De même les distributions de contact dans les fibrés tangents sont notés \mathcal{K} ou \mathcal{C} . Mais des distributions plus générales sont plutôt notées \mathfrak{D} .

Nous n'avons pas pu éviter certains problèmes de redondance dans les notations. Ils viennent de la dialectique entre modèles particuliers et structures générales. En effet nous traitons souvent plusieurs modèles qui sont différents cas spécifiques d'une même structure générale. D'où un dilemme : soit garder autant que faire se peut les mêmes notations générales et sacrifier la spécificité notationnelle, ce qui conduit à affaiblir les particularités des différents modèles, soit changer parfois de notations pour souligner les spécificités et éviter les confusions. Après réflexion, nous avons opté pour la seconde solution.

À cela s'ajoute que nos modèles appartiennent à des domaines mathématiques très balisés historiquement où il existe de nombreuses notations conventionnelles (comme on apprend à l'école à utiliser le symbole x de variable dans les équations, ou le symbole G pour un groupe). Par exemple en mécanique hamiltonienne, il est traditionnel d'utiliser des variables q_i pour dénoter les positions et des variables p_i pour dénoter les moments conjugués. En théorie du contrôle il est courant d'utiliser des λ_i plutôt que des p_i . Nous avons essayé de réaliser un compromis équilibré entre les alternatives en compétition pour que le lecteur soucieux d'approfondissements n'ait pas à effectuer une trop fatigante gymnastique notationnelle.

Qui plus est, nous proposons parfois au lecteur de revenir à la lettre de textes fondateurs particulièrement importants et cela nous conduit naturellement à préserver certaines de leurs spécificités notationnelles.

Et puis enfin les lettres de l'alphabet sont peu nombreuses et, même en utilisant des styles (calligraphique, gothique, grec, etc.), restent très largement insuffisantes pour assurer l'univocité des notations dans un ouvrage aussi long.

Bref, un des buts de cet ouvrage propédeutique est de permettre aux lecteurs biomathématiciens non spécialistes de géométrie différentielle d'accéder à des textes plus spécialisés dont ils sont séparés par une falaise de technicité. Or ces textes ne sont pas du tout homogènes et présentent de nombreuses variantes notationnelles. Les homogénéiser complètement rendrait leur accès difficile car le lecteur devrait effectuer de fastidieux recodages. Nous avons donc préféré garder quelques hétérogénéités justifiées par les différentes traditions savantes, mais en les minimisant.

8.15. Bibliographie

Comme il est normal, la bibliographie est essentiellement constituée d'un côté d'articles spécialisés concernant les sujets techniques abordés et d'un autre côté de traités plus génériques présentant plus globalement des problématiques entières. Mais c'est de façon délibérée que nous avons introduit de façon systématique d'autres types d'items :

- (i) des références à des textes originaux anciens mais fondateurs (par ex. Euler, Pfaff, Cartan, etc.) ; ce retour amont aux sources est essentiel pour comprendre la généalogie des modèles ;
- (ii) des références à des séminaires, comme les séminaires Bourbaki, présentant, de façon certes spécialisée mais adressée à un public professionnel assez large de mathématiciens, de nouveaux concepts et de nouveaux résultats particulièrement importants.

8.16. Un traité quadriparti

En définitive ce traité tresse quatre fils.

1. Un fil modélisateur proposant de formaliser les architectures fonctionnelles et les activités synchronisées du cortex visuel primaire en termes de structures géométriques et fonctionnelles (au sens mathématique) assez sophistiquées. C'est le fil directeur du traité.
2. Un fil d'histoire des mathématiques revenant de façon assez approfondie sur la généalogie de ces structures.
3. Un fil philosophique fondationnel sur l'origine neuronale de la géométrie perceptive.
4. Un fil épistémologique sur les conditions de possibilité (transcendantes) de l'application, mieux, de l'implication, d'idéalités mathématiques abstraites dans une réalité empirique matérielle (ici la réalité neuronale).

9. Résumé du volume

Résumons maintenant de façon un peu précise le contenu de ce volume.

9.1. Le fil directeur

Le cortex visuel de bas niveau géométrise globalement les données sensorielles locales venant de la rétine. Il le fait à travers

1. des neurones visuels qui agissent localement sur le signal comme des filtres à travers leurs champs et profils récepteurs,
2. des connexions qui permettent aux activations neuronales de se propager (ou de s'inhiber) afin de produire une dynamique d'activité globale du module neuronal concerné.

Il s'agit de comprendre mathématiquement cette dialectique entre une géométrie implémentée dans une connectivité et une analyse fonctionnelle implémentée dans des profils récepteurs.

En ce qui concerne la partie analyse fonctionnelle, on sait depuis longtemps (le début des années 1980), à partir de très nombreuses données expérimentales

qui se sont accumulées et qui sont explicitées en détail dans le Vol I, qu'il s'agit (du moins au niveau d'une approximation linéaire) d'une analyse en ondelettes du signal optique.

En ce qui concerne la connectivité, on sait également depuis longtemps (la fin des années 1970 avec un approfondissement des expériences de Hubel et Wiesel de 1959 par les Braitenberg, puis les années 1990 avec la révolution de l'imagerie *in vivo*) qu'il existe une architecture fonctionnelle très précise de V1 en hypercolonnes d'orientation. Les cartes d'orientation avec leur fascinante organisation en "pinwheels" ainsi que bien d'autres cartes ont été exposées en détail dans le Vol I. Mais les structures géométriques qui s'y trouvent implémentées n'ont pas été à l'époque thématiques comme telles, sans doute par ce qu'elles relèvent de théories géométriques de haut niveau, certes bien connues des géomètres purs, mais inconnues de la plupart des mathématiciens appliqués travaillant en neurosciences (du moins à l'exception de quelques spécialistes de la vision comme William Hoffman et Jan Koenderink).

La Neurogéométrie commence, nous l'avons expliqué plus haut, par la reconnaissance que ces cartes d'orientation implémentent, dans des couches neuronales de dimension 2, la discrétisation d'une structure abstraite de dimension 3, à savoir la structure de contact des courbes planes dans le champ visuel. À partir de là, la "dialectique" empirique devient une dialectique mathématique entre une géométrie de contact et une analyse en ondelettes. Sa clarification et son exploration exigent un fort investissement mathématique et ce volume essaye de fournir une introduction aux outils théoriques nécessaires.

Nous suivons les niveaux de structure intervenant successivement.

(i) La structure topologique qui constitue le niveau le plus primaire de cohésion spatiale.

(ii) La structure différentiable qui constitue le niveau géométrique où opère le calcul des formes différentielles d'Élie Cartan, l'outil de base de tous nos modèles de connectivité.

(iii) La structure de contact des espaces de jets qui permet de comprendre comment des données locales (infinitésimales) peuvent être intégrées par des processeurs ponctuels à condition de rajouter des dimensions supplémentaires qui sont en quelque sorte des "dérivés cachés".

(iv) La structure de groupe de Lie qui permet de penser les structures de contact comme invariantes par translation (à gauche) par rapport à une structure de groupe.

(v) Les méthodes variationnelles et la géométrie sous-riemannienne permettant de construire les modèles géodésiques des contours illusoires. Il s'agit d'une part très importante du volume. Plusieurs chapitres y sont consacrés.

(vi) L'analyse harmonique non commutative faisant le lien entre ces infrastructures géométriques et l'analyse du signal au moyen d'états cohérents (les profils récepteurs des neurones).

Ces thèmes sont tressés avec des chapitres d'introduction aux théories mathématiques fondamentales dont procèdent les modèles avec leurs différents niveaux de structure.

(a) La géométrisation de l'intégrabilité de données différentielles de Pfaff à Cartan, puis les espaces de jets d'Ehresmann à Thom. Seule cette "révolution copernicienne" de la géométrisation du calcul intégral-différentiel permet selon nous d'être en accord avec le constat particulièrement énigmatique que l'évolution biologique a pu développer des systèmes neuronaux pouvant implémenter des éléments de calcul différentiel.

(b) La théorie des connexions¹⁸ et des dérivations covariantes qui est la base depuis Cartan et Weyl du passage du local au global en géométrie. C'est là que se jouent les fondements de la géométrie c'est-à-dire le passage des niveaux simplement différentiables aux niveaux proprement géométriques. Cela approfondit considérablement la révolution de la géométrie riemannienne.

(c) La structure et la géométrie des groupes et algèbres de Lie.

(d) Les modèles variationnels, leurs lagrangiens et leurs hamiltoniens, et la théorie du contrôle optimal. Elle est à la base des modèles fondés sur un principe d'optimisation et de notre interprétation géodésique des contours illusoires.

(e) Les espaces de jets généralisés et les distributions de Goursat, les groupes de Carnot et les variétés de Heisenberg qui généralisent la structure de contact des espaces de 1-jets et de 2-jets.

(f) Un mini *vademecum* d'analyse fonctionnelle concernant les opérateurs dans les espaces de Hilbert qui permettent de comprendre la façon dont une géométrie opère sur des fonctions et comment on passe de structures géométriques différentielles à des algèbres d'opérateurs. C'est un grand principe physique et philosophique.

(g) Les représentations de groupes, les états cohérents et l'analyse harmonique non commutative qui approfondissent encore cette dualité entre géométrie sous-riemannienne et analyse fonctionnelle.

9.2. Chapitre 2

Le chapitre 2 est consacré au premier niveau de structure de $V1$ comme fibration, la structure de fibration présentée section 2 étant le point de départ de la neurogéométrie. Répétons encore une fois que l'architecture fonctionnelle de $V1$ découverte par Hubel et Wiesel implémente dans des couches neuronales de dimension concrète 2 une structure abstraite de dimensions 3, la dimension supplémentaire par rapport à la dimension rétinienne étant la variable d'orientation opérant fonctionnellement comme "dérivée cachée", celle donnant les directions des tangentes des courbes planes.

18. Au sens mathématique.

Après une courte section de méthodologie (3) le chapitre commence par le niveau de structure topologique (4) qui a son importance car beaucoup de phénomènes perceptifs (bien étudiés depuis la Gestalttheorie et la phénoménologie), comme la dualité fond/forme, sont de nature topologique et reposent sur l'opposition ouvert/fermé. Nous expliquons la façon dont la topologie de l'espace visuel peut s'obtenir par recollement de domaines locaux. Un tel recollement (plus précisément un tel recouvrement d'un espace topologique par des ouverts) est codé par une structure combinatoire appelée son "nerf". Si l'on connaît le recouvrement, on connaît son nerf. C'est le problème direct. Si réciproquement on connaît un code susceptible d'être un nerf peut-on reconstruire un recouvrement dont il est effectivement le nerf et, surtout, peut-on retrouver la dimension de l'espace topologique sous-jacent ? Cela est essentiel parce que le cortex visuel primaire "ne sait pas" que la rétine est bi-dimensionnelle, la dimension 2 étant encodée dans la connectivité.

Nous faisons une incise, section 5, sur le fait que la neurogéométrie est philosophiquement un transcendantalisme neuronal réactivant dans toute sa complexité la question kantienne si débattue des formes de l'intuition sensible. Nous l'avons déjà noté dès le début du Vol I en nous référant à des neurophysiologistes et neuropsychologues comme John O'Keefe et Stanislas Dehaene qui se réclament explicitement de Kant.

Nous abordons ensuite à la section 6 le niveau de structure différentiable qui est le niveau de base où se situe la neurogéométrie. Cette section est purement pédagogique et rappelle brièvement les notions de variété différentiable et de fibré tangent, la façon dont les champs de vecteurs tangents sont interprétables comme des opérateurs de dérivation, leur algèbre de Lie et leur dualité avec les formes différentielles.

En tant qu'implémentation neuronale d'éléments de géométrie différentielle, la neurogéométrie conduit évidemment à remonter aux racines mêmes du calcul différentiel dans une nouvelle perspective. C'est pourquoi nous remontons effectivement, en y accordant une importance particulière, à la notion d'infinitésimale leibnizienne où s'introduit déjà l'idée que les infinitésimales sont des symboles régis par des règles de calcul *sui generis*.

La section 7 consacrée au calcul des formes différentielles d'Élie Cartan est particulièrement importante. Ce calcul est un calcul symbolique général et presque toutes les théories mathématiques qu'utilise la neurogéométrie reposent sur son extraordinaire réinterprétation du calcul différentiel. Comme nous l'avons déjà dit, Lie, Cartan, Weyl sont les initiateurs d'un tournant structural du problème de l'intégration des équations différentielles (problème de Pfaff), très analogue à celui effectué par Galois pour la résolution par radicaux des équations algébriques. On passe de la quête directe des solutions d'une équation différentielle à l'explicitation des conditions de possibilité de sa résolubilité.

Nous présentons l'intégration des formes différentielles, leurs produits extérieur et intérieur, leur dérivation extérieure, leur dérivée de Lie, et leur nilpotence qui est à la base de leur cohomologie et enfin le théorème de Stokes reliant homologie et cohomologie. Tous ces éléments sont devenus scolaires et se trouvent dans tous les manuels.

Les formes différentielles de Cartan ayant de nombreux précurseurs dans la physique classique (gradients, rotationnels, divergences), nous consacrons la section 8 à quelques rappels à ce propos et insistons sur l'importance du théorème de Stokes. Nous rappelons en particulier, à titre d'exemple pédagogique, l'évolution de la formulation des équations de Maxwell.

Nous concluons le chapitre par un aperçu sur la cohomologie des formes différentielles et sur la notion de faisceau.

9.3. Chapitre 3

Le chapitre 3 est consacré au fait que la structure de fibration de $V1$ contient une structure de contact, nommément la structure de contact des 1-jets des courbes du plan (le plan rétinien). Le point de départ, section 1, est que

- (i) les neurones simples de $V1$ sont des détecteurs d'éléments de contact (a, p) composés d'une position rétinienne a et d'une orientation p et
- (ii) que l'architecture fonctionnelle de $V1$ matérialise, par sa connectivité et l'organisation en pinwheels de ses cartes d'orientation, la structure de contact de dimension 3 qui connecte entre eux ces éléments de contact.

Ce fait fondamental pressenti par William Hoffman et Jan Koenderink dans les années 1980, avant la découverte de la structure en pinwheels dans les années 1990, est l'exemple princeps du fait qu'une architecture fonctionnelle neuronale peut implémenter une structure géométrique d'un type plus raffiné que les intuitions géométriques standard.

La section 2 explique la structure de contact du fibré \mathbb{V}_J des 1-jets des courbes du plan de base ainsi que la façon dont ces courbes planes se relèvent dans le fibré en des courbes gauches d'éléments de contact en "bonne continuation" les uns par rapport aux autres, ce que l'on appelle des "relevées legendriennes".

Si l'on connaît une courbe plane γ d'équation $y = f(x)$ on peut la relever en une courbe $\Gamma = (x, y = f(x), p = f'(x))$ dans le fibré des 1-jets de coordonnées (x, y, p) . C'est le problème "direct". Le problème "inverse" consiste, étant donnée une courbe $\Gamma = (x, y = f(x), p(x))$ dans le fibré \mathbb{V}_J , de savoir si elle est la relevée legendrienne d'une courbe plane γ . La condition est que $p(x) = f'(x)$, autrement dit que la 1-forme $\omega_J = dy - p dx$ soit égale à zéro. ω_J est la 1-forme définissant la structure de contact et l'équation $\omega_J = 0$ exprime l'intégrabilité des éléments de contact $(x, y = f(x), p(x))$. Cette structure est invariante sous l'action du groupe

$SE(2)$ des isométries du plan, produit semi-direct $\mathbb{R}^2 \rtimes SO(2)$ des translations et des rotations.

Le point essentiel est le suivant. Quand on pense à une courbe plane $y = f(x)$ dans un plan R en termes d'analyse différentielle, on pense la courbe comme l'enveloppe de ses tangentes de pente $\frac{dy}{dx} = f'(x)$. Mais une tangente est une donnée locale (infinitésimale) et non pas ponctuelle et ne peut donc pas être détectée par des processeurs ponctuels calculant des valeurs en un point a du plan. Si l'on est restreint à des processeurs ponctuels, on peut néanmoins surmonter cette obstruction

- (i) en ajoutant une variable indépendante p supplémentaire (et donc en ajoutant une variété supplémentaire P de dimension 1),
- (ii) en considérant des courbes dans $\mathbb{V} = R \times P$,
- (iii) en imposant une contrainte d'intégrabilité forçant ces courbes à être des relevées legendriennes.

On en arrive ainsi à une équivalence entre d'un côté une analyse par des processeurs locaux non ponctuels dans le plan et d'un autre côté l'analyse par des processeurs ponctuels dans le fibré des 1-jets muni de sa structure de contact. Les deux analyses sont équivalentes mais leurs implémentations matérielles sont très différentes. C'est la seconde implémentation qui convient à la neurogéométrie parce que, répétons-le, les neurones sont (à l'échelle de résolution qui leur est propre) des processeurs ponctuels. L'évolution biologique a progressivement doté certains organismes de structures sophistiquées et complexes composées d'une entrée sensorielle (la surface rétinienne) et de modules post-sensoriels (corticaux) supplémentaires munis d'une architecture fonctionnelle spécifique.

Le noyau de $\omega_J = dy - p dx$ est une distribution \mathcal{K} de plans tangents K_x , dits plans de contact, dans la variété \mathbb{V} de dimensions 3 et cette distribution est très spéciale. Les plans K_x "tournent" très vite, trop vite pour admettre des surfaces intégrales. La distribution \mathcal{K} est dite "complètement non intégrable" ou "non holonome" et ne satisfait pas à une condition d'intégrabilité, dite de Frobenius, qui lui permettrait d'avoir des surfaces intégrales. Cette "non holonomie" assez contre-intuitive peut être une surprise pour les non-mathématiciens.

Les intégrales de \mathcal{K} sont donc de dimension 1 (les relevées legendriennes) et non pas 2 bien que \mathcal{K} soit un champ de plans. Cela est dû techniquement au fait que le commutateur $[X, Y]$ de champs de vecteurs différents dans \mathcal{K} n'est pas dans \mathcal{K} mais transverse à \mathcal{K} . Autrement dit, l'algèbre de Lie engendrée par les plans de contact est celle du fibré tangent $T\mathbb{V}$ tout entier. Cette condition opposée à la condition d'intégrabilité de Frobenius s'appelle la condition d'Hörmander. Elle ouvre tout un univers de géométries plus sophistiquées que les géométries intuitives dont nous avons l'habitude.

Comme il n'existe pas de surface intégrale de \mathcal{K} , si S est une surface dans \mathbb{V} et si $x \in S$ les intersections entre les plans tangents $T_x S$ de S et les plans de contact K_x

de \mathcal{K} sont en général réduites à des droites et définissent un feuilletage de dimension 1 de S dont les singularités sont les points x où $T_x S = K_x$. Nous explicitons en détail l'exemple instructif de la sphère unité \mathbb{S}^2 .

La section suivante (3) approfondit la notion de structure de contact. Géométriquement parlant, la structure de contact est la distribution de plans \mathcal{K} . Mais quand on la considère comme le noyau d'une 1-forme différentielle ω_J on ajoute un niveau de structure supplémentaire puisque, si f est une fonction sur \mathbb{V} partout non nulle, les 1-formes ω_J et $f\omega_J$ ont le même noyau et définissent donc la même distribution \mathcal{K} . Ce supplément de structure est représentable par le champ de vecteurs χ sur \mathbb{V} , dit champ de Reeb, défini par les deux conditions

- (i) $\omega_J(\chi) = 1$, et
- (ii) $d\omega_J(\chi, X) = 0$ pour tout champ X dans \mathcal{K} .

Cela revient à ajouter un nouveau degré de liberté qui est un paramètre local d'échelle correspondant au groupe multiplicatif des nombres réels > 0 , ce que l'on appelle un champ de jauge. La structure de contact $(\mathbb{V}, \mathcal{K}, \omega_J, \chi)$ se trouve par conséquent plongée (avec l'échelle 1) dans une variété de dimension 4 sur laquelle ω_J et les changements d'échelle définissent une structure symplectique dite symplectisée de la structure de contact.

On entre ainsi dans l'immense univers de la géométrie symplectique qui est celui de toute la mécanique classique depuis Hamilton. Elle est bien connue des géomètres et des physiciens mais en général peu connue dans les autres domaines d'application des mathématiques. La neurogéométrie plonge certains éléments de neuromathématiques dans cet univers captivant.

Il existe des liens très étroits entre structures symplectiques, structures métriques et structures complexes (au sens de structures définies sur le corps \mathbb{C} des nombres complexes et pas seulement sur le corps \mathbb{R} des réels). Si M est une variété symplectique de dimension $2n$ munie de sa 2-forme symplectique Ω , la formule de base est $\Omega(X, Y) = g(X, JY)$, où X et Y sont deux champs de vecteurs tangents à M , g le tenseur métrique et J un opérateur "multiplication par i " (rotation directe de $\frac{\pi}{2}$) faisant des plans tangents $T_x M$ de dimension $2n$ sur \mathbb{R} (donc isomorphes à \mathbb{R}^{2n}) des espaces complexes de dimension n sur \mathbb{C} (donc isomorphes à \mathbb{C}^n). Nous étudions avec un certain détail les effets des modifications de ω_J qui laissent \mathcal{K} invariant mais modifient le champ de Reeb, ainsi que la métrique et la structure complexe du symplectisé.

Après ces approfondissements de la géométrie de contact des 1-jets, nous disons quelques mots dans la section 4 sur les 2-jets qui détectent non seulement des éléments de contact (a, p) interprétables comme des directions tangentes à des courbes dans le plan de base mais des éléments de contact du "deuxième ordre" (a, p, k) interprétables comme des courbures $k(x)$ de ces courbes. Il semble en effet qu'il existe des données expérimentales attestant de l'existence de neurones

spécialisés dans la détection de courbures. Cela conduit dans le cadre de la géométrie des espaces de jets à une structure bien connue qui s'appelle la structure d'Engel (Engel était un important disciple de Lie).

La dernière section (5) de ce chapitre est plus historique. Elle résume l'avènement et le développement de la notion cruciale d'élément de contact au cours du XIX^e siècle. Elle revient sur l'histoire de la découverte qu'un calcul intégral-différentiel peut être équivalent à une géométrie de connectivité entre éléments de contact. Cette idée profonde a été introduite par Sophus Lie avec un sens aigu de sa nouveauté et du retournement que cela représentait. Il s'agissait en effet de passer de l'analyse à la géométrie en géométrisant le calcul différentiel.

9.4. Chapitre 4

Le chapitre 4 est un chapitre historique entièrement consacré à la “révolution copernicienne” de la géométrisation des conditions d'intégrabilité des équations différentielles au XIX^e siècle. Il est justifié par le fait que les modèles de neurogéométrie conduisent à repenser ces progrès comme un véritable “tournant” fondationnel révélant “l'origine” de la géométrie.

Depuis Copernic, la notion de révolution copernicienne (section 1) est utilisée pour exprimer un changement de point de vue drastique sur une certaine classe de phénomènes. Pour la cosmologie c'est le passage du géocentrisme à l'héliocentrisme dans l'étude des trajectoires planétaires “errantes” (en amont de Copernic, l'héliocentrisme remonte en fait à Aristarque). En philosophie, Kant s'en réclame explicitement pour expliquer le statut transcendantal de la notion d'objectivité et son passage aux conditions de possibilité de l'expérience. En mathématiques l'exemple princeps est celui de Galois qui a explicité les conditions de possibilité de résolubilité par radicaux des équations algébriques.

Nous essayons de dégager des traits communs à ces différentes révolutions coperniciennes, en particulier la découverte d'une “structure d'arrière-plan” immanente à la résolution d'un certain problème. Cette structure “synthétique a priori” opérant comme cadre pour la diversité des données empiriques concernées est en général formulée conceptuellement. Mais quand elle devient mathématisable (comme c'est le cas avec le groupe de la relativité galiléenne en mécanique classique ou chez Galois avec les groupes de permutation des racines d'une équation algébrique), son contenu mathématique formalise les conditions de résolubilité du problème posé et devient donc déterminant pour ses solutions. Parfois même tellement déterminant qu'il engendre des solutions. Nous en explicitons les exemples en nous attardant un peu sur celui de Galois.

Mais l'exemple qui nous intéresse pour la neurogéométrie est celui de la géométrisation des conditions de possibilité de l'intégrabilité des équations différentielles. Il est appelé communément le problème de Pfaff car il a été bien formulé par Johann

Pfaff. Dans la section 2 nous en retraçons l’histoire héroïque. Les grands innovateurs et créateurs sont Monge, Pfaff, Plücker, Grassmann, Natani, Clebsch, Darboux, Lie, Goursat, Engel, Cartan. Le problème est présenté comme celui de l’intégration d’une équation aux différentielles totales (EDT) à n variables de la forme $\omega = \sum_{i=1}^{i=n} A_i(x_1, \dots, x_n) dx_i = 0$, ce qui deviendra plus tard une équation $\omega = 0$ où ω est une 1-forme sur une variété M de dimension n . Le problème traité précédemment des relevées legendriennes des courbes planes dans l’espace des 1-jets en est un exemple simple et typique.

Quand $\omega = dF$ est la différentielle totale d’une fonction $F(x_1, \dots, x_n)$, le problème est trivial. L’EDT est intégrable et les solutions sont les hypersurfaces d’équation $F = \text{constante}$. Lorsque F est connue on sait calculer dF . C’est le problème “direct”, celui de la différentiation. Le problème de Pfaff est le problème “inverse”, celui de l’intégration. On se donne une EDT à n variables et on se demande si on peut l’intégrer. Pour que ω soit exacte, i.e. de la forme dF , il faut que $\frac{\partial A_i}{\partial x_j} = \frac{\partial A_j}{\partial x_i}$ car pour les fonctions $\frac{\partial^2 F}{\partial x_i \partial x_j} = \frac{\partial^2 F}{\partial x_j \partial x_i}$. Si ω n’est pas exacte on se demande alors s’il existe un facteur intégrant, c’est-à-dire une fonction ρ partout non nulle telle que $\rho\omega$ soit exacte. En effet $\rho\omega = 0$ est alors équivalent à $\omega = 0$. Mais l’existence d’une telle fonction ρ impose également de fortes conditions aux coefficients A_i . Si celles-ci sont satisfaites, la recherche d’un facteur intégrant est déjà en soi un problème difficile qui a constitué un important thème de recherche au XVIII^e siècle par exemple chez Euler.

Mais les vraies difficultés commencent quand il n’existe pas de facteur intégrant. On disait alors au XVIII^e siècle que l’EDT n’était pas intégrable. Mais Gaspard Monge fut le premier à remarquer que rechercher un facteur intégrant $\rho\omega = dF$ c’était chercher des solutions $F = \text{constante}$ et donc des hypersurfaces de dimension $n - 1$. Il eut alors l’idée géniale que, dans le cas non intégrable, l’EDT $\omega = 0$ pouvait quand même avoir des solutions mais de dimension *inférieure* à $n - 1$ et que, dans ce sens, les EDT “sont toutes susceptibles d’une véritable intégration”. Et de façon véritablement prophétique Monge expliqua que cette généralisation du problème des “conditions d’intégralité” aux cas considérés jusque-là comme non intégrables ouvrait “une chaîne immense” et “un nouveau champ à l’analyse et à la géométrie”. C’est à partir de là que Pfaff comprit qu’il s’agissait de déterminer, à partir de certaines propriétés (encore à découvrir) sur les coefficients A_i et sur leurs dérivées, quelle était la dimension maximale des solutions possibles. Un immense domaine mathématique en sortira.

C’est avec le *Methodus* de Pfaff, mémoire de 1815, que commence véritablement la révolution copernicienne. Pfaff démontre un théorème de réduction disant qu’il y a toujours des solutions au moins de dimension m si $n = 2m$ ou $n = 2m + 1$. Sa méthode repose sur des changements de variables permettant d’éliminer le plus

possible de différentielles dans l'EDT $\omega = 0$ et de construire ce que l'on appellera plus tard une forme normale.

Fidèle à notre pédagogie historique nous entrons un peu en détail dans le *Methodus* de Pfaff (rédigé en latin comme cela se faisait encore à l'époque). Il s'agit d'un texte exceptionnel où Pfaff explique que, pour résoudre le problème de l'intégrabilité des EDT, il a dû faire un retour amont jusqu'aux sources du calcul différentiel et prendre un "nouveau chemin". Le tournant s'effectue en considérant qu'une EDT $dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy = p dx + q dy$ (où z est une fonction $z(x, y)$) doit être interprétée dans un premier temps non pas comme une équation à 3 variables (x, y, z) mais bien comme une équation à 5 variables (x, y, z, p, q) et donc comme une équation à 5 différentielles dont les coefficients de dp et dq sont nuls. Pfaff s'inscrit ici dans la filiation d'Euler qui considérait déjà les dérivées partielles des fonctions comme de nouvelles variables indépendantes ne devant être interprétées comme de "vraies" dérivées partielles que dans un second temps. Traiter les dérivées partielles comme des variables indépendantes, augmenter la dimension du système et ajouter a posteriori des contraintes faisant de ces variables supplémentaires des "dérivées cachées" comme a pu le dire aujourd'hui le grand spécialiste de géométrie sous-riemannienne Richard Montgomery, est une refondation géométrique du calcul différentiel, un "nouveau départ" et un "nouveau chemin".

Notre structure de contact $\omega_J = 0$ des 1-jets des courbes planes est un cas particulier de système de Pfaff, d'où la légitimité de cette enquête historique.

Une génération après Pfaff, en 1833, Hamilton réinterprétera la mécanique de Lagrange avec la même méthode en considérant les vitesses (les dérivées des variables de configuration par rapport au temps) comme des variables indépendantes (les "moments"). Il passera, comme on dit maintenant, de "l'espace de configuration" M à "l'espace des phases" qu'est le fibré cotangent T^*M de dimension double et il introduira des contraintes exprimées par la structure symplectique canonique de T^*M .

Pfaff développe en détail plusieurs exemples de formes normales d'EDT obtenues par des changements de variables appropriés. Elles révèlent la structure abstraite sous-jacente à l'EDT et expriment les conditions de possibilité de son intégrabilité.

Après Pfaff, on développera ses idées jusqu'à l'émergence d'une seconde révolution dans les années 1870, période où s'effectuera le véritable tournant de la géométrisation "intrinsèque" (c'est-à-dire invariante par changements de variables) de l'intégrabilité avec l'introduction des "éléments de contact". Tout cela aboutira à la reformulation complète de la théorie dans le langage des formes différentielles. La mise au point de l'algèbre linéaire (en particulier antisymétrique) par Hamilton, Cayley Grassmann et tant d'autres permit de formuler tout un ensemble de conditions en termes de formes différentielles. En particulier Frobenius introduisit en 1877 son fameux "covariant bilinéaire" interprété plus tard comme la dérivée

extérieure $d\omega$ et découvrit la condition $\omega \wedge d\omega = 0$ qui, si elle n'est pas satisfaite, rend impossible l'existence de solutions de dimension $n - 1$ de l'EDT $\omega = 0$. Cela résout le problème de Monge. Lie et Darboux parachèveront la géométrisation de l'intégrabilité avec l'introduction des éléments de contact et des transformations de contact. Nous commentons le mémoire de Darboux de 1882 "Sur le problème de Pfaff" où sont expliquées clairement les structures d'algèbre de Lie, de crochet de Lie et de crochet de Poisson.

Nous concluons ce chapitre historique par la section 3 présentant la systématization du problème de Pfaff par Cartan et Goursat. Elle traite des théorèmes sur la réduction et la classification des systèmes de Pfaff, le dégagement de leurs structures de base (symplectiques en dimension paire et de contact en dimension impaire). À titre d'illustration, nous reprenons un exemple non trivial de Cartan à cinq variables, puis nous entrons assez longuement dans les arcanes techniques de la résolution des systèmes de Pfaff. On y prend conscience de l'extrême richesse de la théorie.

9.5. Chapitre 5

Le chapitre 5 est consacré à l'étude du fait que la structure de contact \mathbb{V}_J est invariante par les translations à gauche d'une structure de groupe de Lie non commutatif sur \mathbb{V}_J (section 1). Ce fait justifie une introduction pédagogique à la théorie des groupes de Lie G et de leurs algèbres de Lie \mathcal{G} . L'algèbre \mathcal{G} exprime sur le plan tangent à l'origine $T_e G$ de G la version infinitésimale de la loi de groupe au moyen des commutateurs que sont les crochets de Lie $[X, Y]$ entre deux vecteurs tangents $X, Y \in T_e G = \mathcal{G}$. Cette théorie est devenue omniprésente en géométrie. La section 2 en est un prélude.

La structure de contact \mathbb{V}_J des 1-jets des courbes planes correspond au groupe de Heisenberg polarisé qui est l'un des exemples les plus connus de groupe de Lie. C'est un produit semi-direct des translations du plan (x, y) par des p opérant comme $p(x, y) = (x, y + px)$. Dans la section 3, nous explicitons assez longuement ses propriétés, ses représentations matricielles, sa propriété de nilpotence, sa variante non polarisée, ses représentations adjointe et co-adjointe ainsi que leurs orbites (section 6). Si \mathcal{G} est une algèbre de Lie, sa représentation adjointe $ad_X(Y) = [X, Y]$ est définie par ses commutateurs.

À partir de cet exemple particulièrement connu nous introduisons la théorie de Kirillov montrant que les orbites de la représentation co-adjointe sont naturellement munies d'une structure symplectique.

Le groupe de Heisenberg étant essentiel en mécanique quantique, nous en disons un mot. Cet excursus physique nous permet d'insister sur la profondeur de ces formalismes.

Puis le chapitre passe, section 7, du modèle nilpotent \mathbb{V}_J au modèle $\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1 = SE(2)$ qui correspond au fibré principal associé à \mathbb{V}_J . Sa 1-forme de contact

est $\omega_S = -\sin(\theta) dx + \cos(\theta) dy$ et s'obtient en multipliant $\omega_J = dy - p dx$ par $\cos(\theta)$ puisque $p = \frac{\sin(\theta)}{\cos(\theta)}$. Cela définit la même distribution \mathcal{K} de plans de contact mais pas du tout le même champ de Reeb. \mathbb{V}_S a l'intérêt de ne plus privilégier des axes dans le plan de base. Le groupe \mathbb{V}_S n'est plus nilpotent, \mathbb{V}_J en est la structure tangente à l'origine et nous en faisons l'analyse algébrique et géométrique.

Nous indiquons en plus l'extension en dimension 3 de ces modèles (section 8) et la structure du fibré de dimension 5 qui décrit le trièdre mobile de Frenet (tangente, normale, binormale) le long des courbes gauches de \mathbb{R}^3 . Les relevées legendriennes sont dans ce cas les courbes intégrales d'une distribution de sous-espaces tangents de dimension 3 définie par deux 1-formes.

Nous revenons aussi section 9 sur la notion générale d'exponentielle d'une algèbre de Lie \mathcal{G} dans son groupe de Lie G et nous évoquons le théorème de Cartan-von Neumann.

Nous concluons enfin ce chapitre par une section 10 consacrée à la forme de Maurer-Cartan Λ qui est associée à tout groupe de Lie et satisfait la formule universelle $d\Lambda + [\Lambda, \Lambda] = 0$.

9.6. Chapitre 6

Le chapitre 6 est consacré à un nouveau niveau de structure de nos modèles. Après la structure différentiable, après la structure de contact et après la structure de groupe et d'algèbre de Lie, nous introduisons un niveau "métrique". Mais ce niveau est plus subtil que celui de la géométrie euclidienne et riemannienne car la métrique n'y est pas définie sur tout l'espace mais seulement sur la distribution des plans de contact \mathcal{K} qui, rappelons-le, est complètement non intégrable. Cette métrique partielle restreinte à des éléments de contact de dimensions 2 n'est donc même pas définie sur des surfaces à cause précisément de la non intégrabilité. Il s'agit d'une situation qui sera sans doute non intuitive pour nombre de lecteurs.

De telles métriques sur une distribution non intégrable d'hyperplans tangents sont dites "sous-riemanniennes" et pour y introduire il faut utiliser la théorie des connexions de Cartan-Weyl. Cartan a distingué trois grandes conceptions de la géométrie :

- (i) celle de Klein et de son programme d'Erlangen reposant sur la notion de groupe d'isométries ;
- (ii) celle de Riemann localisant la géométrie euclidienne ;
- (iii) et enfin celle du "transport parallèle infinitésimal" fondée sur la possibilité de transporter les vecteurs tangents d'une variété différentiable d'un point à un autre, procédé que Cartan a appelé une "connexion".

La section 1 explique que c'est la troisième conception qui est adaptée à la neurogéométrie puisque, ainsi que nous l'avons montré en détail dans le Vol I, les

connexions (au sens neurophysiologique) cortico-corticales à longue portée entre les neurones simples de $V1$ implémentent précisément un transport parallèle d'une hypercolonne d'orientation à une autre.

En fait (section 2) le modèle $(\mathbb{V}_J, \mathcal{K})$ est bien une connexion au sens de Cartan lorsque l'on considère la distribution \mathcal{K} de plans de contact non pas au-dessus du plan de base (x, y) mais, par dualité projective (comme nous l'avons expliqué dans le Vol I) au-dessus du plan (x, p) muni de sa structure symplectique faisant de p le moment conjugué de x .

Cela justifie, en accord avec nos choix pédagogiques, de présenter un petit *va-demecum* de la théorie des connexions et des dérivations covariantes comme calcul différentiel "fibré". Tel est le rôle de la section 3. Les fibres "verticales" sont considérées comme des variables "internes" et les points x de l'espace de base M comme des variables externes. Une connexion consiste alors, étant donné un fibré $\pi : \mathfrak{F} \rightarrow M$, à définir en chaque point $\mathfrak{p} \in \mathfrak{F}$ un relèvement $C_{\mathfrak{p}}$ dans $T_{\mathfrak{p}}\mathfrak{F}$ de l'espace tangent à la base $T_{\pi(\mathfrak{p})}M$ en $\pi(\mathfrak{p})$ qui soit transverse aux fibres. $C_{\mathfrak{p}}$ s'appelle le "plan horizontal" en ce point \mathfrak{p} . Nous explicitons les cas du fibré des repères de M (ce qui redonne la formule de Maurer-Cartan), celui des fibrés principaux associés à un groupe G (ils sont essentiels en physique dans les théories de jauge) et celui des fibrés vectoriels.

La notion de connexion constitue selon nous le véritable fondement de toute géométrie globale cohérente où l'on peut comparer les géométries infinitésimales des différents points. Elle s'impose de façon évidente en neurogéométrie pour des raisons expérimentales. Elle est fondationnelle, encore plus que la géométrie riemannienne. Nous saisissons cette occasion pour revenir sur le problème de l'espace physique, le fameux *Raumproblem*, qui à partir de Riemann et Helmholtz a abouti aux conceptions de Cartan, Weyl et Hilbert dans leur dialogue avec Einstein.

Une fois bien établie la notion de connexion, le calcul différentiel se trouve complètement repensé en termes de dérivation covariante ∇_X . Nous explicitons à ce propos les formules de base des connexions affines (avec leurs biens connus "déluges d'indices"), leurs 2-formes de courbure $\mathfrak{R}(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X, Y]}$ et leur torsion

Comme la notion de connexion est devenue omniprésente en théorie des champs avec les théories de jauge, qu'elles soient abéliennes (équations de Maxwell) ou non abéliennes (théorie quantique des champs, lagrangiens de Yang-Mills), nous donnons un aperçu sur ces dernières dans la section 4. Il nous semble en effet pertinent de faire de temps à autre un petit excursus dans les applications à la physique fondamentale des formalismes que nous utilisons pour la neurogéométrie.

Dans le cas des connexions sur le fibré tangent TM ou sur le fibré des repères affines de l'espace de base M la situation est un peu spéciale parce que les éléments des fibres sont des déplacements infinitésimaux dans M . Comme nous l'expliquons dans la section 5, "l'interne" des fibres "verticales" est en fait une transformation

infinitésimale de “l’externe horizontal”. Cela permet de mixer ces deux composantes. Une telle possibilité est fondamentale en relativité générale et se trouve à l’origine des symboles de Christoffel. La connexion ∇_X agit maintenant sur les vecteurs tangents $Y \in TM$ et l’on peut définir le transport parallèle des vecteurs tangents le long des courbes de M . La courbure $\mathfrak{R}(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X, Y]}$ satisfait l’identité de Bianchi $\nabla \mathfrak{R} \equiv 0$. Les connexions sur TM ont une torsion qui vient de la possibilité de mixer l’interne et externe. Sa formule est $T(X, Y) = \nabla_X(Y) - \nabla_Y(X) - [X, Y]$.

Dans ce contexte, on retrouve facilement la géométrie riemannienne et la connexion de Levi-Civita qui est de torsion nulle et pour laquelle la métrique g est parallèle au sens où $\nabla g \equiv 0$. Nous explicitons donc dans la section 6 les bases de la géométrie riemannienne : identité de Bianchi, courbure de Ricci, géodésiques $\gamma(s)$. Ces dernières sont les courbes qui transportent parallèlement à eux-mêmes leurs vecteurs tangents unitaires : $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$. Nous donnons section 6.3 l’exemple de la sphère \mathbb{S}^2 et retrouvons à partir des formules générales le fait que les géodésiques sont les grands cercles.

À propos de la géométrie riemannienne nous revenons un peu plus en détail sur son utilisation en relativité générale et sur l’origine de l’équation d’Einstein reliant, côté physique, la conservation du tenseur T d’impulsion-énergie $\nabla T = 0$ (∇ connexion de Levi-Civita) et, côté géométrie, l’identité de Bianchi $\nabla G = 0$ où $G = \text{Ricci} - \frac{1}{2}rg$ est le tenseur d’Einstein (Ricci est le tenseur de Ricci, r la courbure scalaire, g le tenseur métrique). D’où l’hypothèse que $G = \alpha T$, la limite newtonienne fixant le facteur α . Nous reprenons alors les débats Cartan-Einstein-Weyl sur l’extension de la relativité générale et sur l’unification de la gravitation avec l’électro-magnétisme (équations de Maxwell). Cartan introduisait de la torsion et Einstein pensait à un “parallélisme à distance” intégrable (*Fern-Parallelismus*), un parallélisme infinitésimal intégrable permettant de comparer non seulement les longueurs mais aussi les directions. Le débat est passionnant, Weyl s’opposant au parallélisme à distance et voulant en rester au parallélisme infinitésimal sans postuler qu’il soit intégrable. Mais l’advenue de la mécanique quantique bouleversa ces premières tentatives de théories unitaires.

En conclusion de ce chapitre sur la géométrie des connexions, deux sections introduisent, à partir de l’exemple des métriques de Carnot-Carathéodory sur les groupes de Carnot, la notion de géométrie sous-riemannienne largement approfondie dans les chapitres suivants. La section 7 explique d’abord que, étant donnée une structure de contact (M, \mathcal{K}) il s’agit de métriques définies sur la distribution \mathcal{K} qui permettent de calculer la longueur des courbes “admissibles”, c’est-à-dire des courbes intégrales de \mathcal{K} . Comme la condition d’Hörmander disant que \mathcal{K} et ses commutateurs engendrent le fibré tangent total TM est satisfaite, le théorème de Chow dit que deux points quelconques sont toujours reliables par une courbe régulière “admissible”. On peut donc définir une métrique en prenant pour distance $d(x, y)$ entre deux points

x, y de M la borne inférieure des longueurs des courbes admissibles joignant x à y . Ces métriques sous-riemanniennes de Carnot-Carathéodory sont essentielles pour notre propos puisque nous modélisons la construction par le cortex de contours illusoires modaux à longue portée par des géodésiques de certaines d’entre elles définies sur la structure de contact.

La dernière section 8 du chapitre explicite le fait qu’en dimension 3 le groupe nilpotent de Heisenberg joue le rôle de “structure tangente” universelle pour la géométrie des groupes de Lie munis d’une structure de contact invariante (à gauche). Nous expliquons la structure générale de ces groupes nilpotents stratifiés de Carnot ainsi que leurs propriétés d’homogénéité. Les travaux de Gromov ont été pionniers et centraux. Ces groupes sont extrêmement anisotropes et leurs sous-espaces possèdent différents degrés d’homogénéité (des “poids”). Leur dimension homogène est donc plus grande que leur dimension comme variété. L’exemple le plus simple est le groupe de Heisenberg polarisé qui possède des variables x, p de poids 1 et une variable y de poids 2. Sa structure de contact $\omega = dy - p dx$ est donc homogène de poids 2 et sa dimension homogène, qui est la somme $(2 \times 1) + (1 \times 2)$ des dimensions des sous-espaces homogènes pondérées par leur poids, est 4 alors que sa dimension comme variété est 3.

À cause de l’existence de dilatations caractéristiques de leur nilpotence (analogues pour ces groupes aux homothéties classiques des espaces euclidiens) ces groupes de Carnot ont des propriétés d’*auto-similarité*. En utilisant les dilatations, Pierre Pansu a réussi à définir à leur propos une différentiation et toute une nouvelle généralisation du calcul différentiel en est issue. L’idée de départ est que pour les groupes de Carnot l’équivalent des applications linéaires pour les espaces vectoriels sont les morphismes de groupe commutant avec les dilatations. On peut alors définir les applications différentiables au sens de Pansu entre groupes de Carnot comme celles qui admettent une application linéaire tangente en ce nouveau sens.

9.7. Chapitre 7

Le chapitre 7 approfondit la structure des groupes et algèbres de Lie réelles de dimension finie mixant leurs propriétés algébriques et leurs propriétés différentiables et précise leurs métriques invariantes qu’elles soient riemanniennes ou sous-riemanniennes. Trois niveaux de structure s’y intriquent. Il n’est pas spécifiquement lié à nos modèles neurogéométriques mais il en précise notablement le contexte mathématique général.

Le chapitre commence (section 1) par une définition des algèbres de Lie \mathcal{G} nilpotentes, celles dont la “suite centrale descendante” des commutateurs $\mathcal{G}_{(0)} = \mathcal{G}$, $\mathcal{G}_{(j+1)} = [\mathcal{G}, \mathcal{G}_{(j)}]$ s’arrête et devient nulle à partir d’un certain rang. C’est la plus petite classe d’algèbres de Lie qui est stable par sous-algèbres, algèbres quotient et extensions centrales et qui contient les algèbres abéliennes (dont la suite est nulle

puisque $\mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}] = 0$). Elles possèdent une structure “en drapeau”, ce qui signifie qu’il existe une suite croissante d’idéaux \mathcal{G}_j (ce ne sont pas les $\mathcal{G}_{(j)}$) de dimensions $j = 0, 1, \dots, n$, $n = \dim(\mathcal{G})$, tels que $[\mathcal{G}, \mathcal{G}_{j+1}] \subset \mathcal{G}_j$. Dans une base adaptée à ce drapeau, les matrices de la représentation adjointe sont des matrices triangulaires supérieures strictes et donc nilpotentes.

Les algèbres de Lie nilpotentes se généralisent avec les algèbres de Lie résolubles (section 2) qui s’obtiennent par une suite d’extensions (pas forcément centrales) de sous-algèbres $\mathcal{G} = \mathcal{A}_1 \supseteq \dots \supseteq \mathcal{A}_{k+1} = \{0\}$ telles que \mathcal{A}_{j+1} soit un idéal de \mathcal{A}_j et que les quotients successifs $\mathcal{A}_j/\mathcal{A}_{j+1}$ soient tous abéliens. C’est la plus petite classe stable par sous-algèbres, algèbres quotient et extensions qui contient les algèbres abéliennes. Par exemple l’algèbre de Lie \mathcal{V}_S de notre modèle \mathbb{V}_S n’est pas nilpotente mais elle est résoluble.

La section suivante (3) présente quelques éléments de la théorie des algèbres de Lie générales. Il existe deux grandes classes, les résolubles et les semi-simples, \mathcal{G} étant dite semi-simple si elle n’a pas d’idéal résoluble non trivial, c’est-à-dire si elle est totalement non commutative (puisque alors son centre étant un idéal abélien est nécessairement nul). Le théorème de décomposition de Levi dit que toute algèbre de Lie est un produit semi-direct d’une algèbre de Lie résoluble (son radical, i.e. son idéal résoluble maximal) et d’une algèbre de Lie semi-simple. Les algèbres de Lie simples sont les algèbres non abéliennes n’ayant pas d’idéaux non triviaux. Toute algèbre de Lie semi-simple est une somme directe d’idéaux simples.

Souvent, dans les traités d’introduction aux groupes et algèbres de Lie, on se focalise sur les algèbres semi-simples dont la structure (étudiée en particulier par Cartan et Weyl) est particulièrement intéressante associée qu’elle est aux “systèmes de racines” classifiant les algèbres simples. Dans les modèles neurogéométriques les algèbres résolubles sont en revanche particulièrement pertinentes.

Un outil essentiel pour distinguer les deux grandes classes d’algèbres de Lie est la “forme de Killing”, forme bilinéaire symétrique sur \mathcal{G} définie par $K(X, Y) = \text{Trace}(\text{ad}_X \text{ad}_Y)$, où $\text{ad}_X(Y) = [X, Y]$ est la représentation adjointe. Elle mesure en quelque sorte la non-commutativité. Un critère dû à Cartan dit alors

- (i) que \mathcal{G} est résoluble si et seulement si $K(\mathcal{G}, [\mathcal{G}, \mathcal{G}]) = 0$ (autrement dit, $K(X, Y)$ est nul dès que l’un des termes X ou Y est un commutateur), et
- (ii) que \mathcal{G} est semi-simple si seulement si $K(X, Y)$ est non dégénérée.

Nous nous arrêtons alors sur la classification de Bianchi des algèbres de Lie réelles de dimension 3 puisque nos deux modèles de base sont de ce type. Nous détaillons assez longuement les calculs qui sont très éclairants.

Puis après avoir dit un mot sur les “algèbres enveloppantes” des algèbres de Lie nous revenons sur la formule de Baker-Campbell-Hausdorff concernant l’exponentielle $\exp(X)$ permettant de passer de l’algèbre de Lie \mathcal{G} d’un groupe de Lie G à G

lui-même en reliant la multiplication des $\exp(X)$ dans G aux crochets des X dans \mathcal{G} .

Nous concluons le chapitre par une section (4) sur les métriques invariantes (à gauche) sur les groupes de Lie G . En ce qui concerne les métriques invariantes en dimension 3, nous commentons un intéressant article de John Milnor en reprenant assez en détail ses calculs de courbure. Et en ce qui concerne les métriques sous-riemanniennes invariantes, nous reprenons, toujours avec un certain détail, les travaux d'Andrei Agrachev, Davide Barilari et Ugo Boscain. Leur classification est tout à fait remarquable.

9.8. Chapitre 8

Le chapitre 8 aborde le problème fascinant des contours illusoires à longue portée, c'est-à-dire de portée bien supérieure à la taille des profils récepteurs des neurones corticaux impliqués dans leur construction. Ils sont bien connus et ont été abondamment étudiés en phénoménologie, en psychologie et en psychophysique depuis la Gestalttheorie. Ils manifestent de façon saisissante le fait que le cortex visuel forme géométriquement les stimuli à partir de sa connectivité immanente. En effet, les données sensorielles périphériques (rétiniennes) sont "pixelisées", incomplètes et lacunaires et la perception les traite en quelque sorte comme des conditions aux limites à structurer et à compléter en y projetant sa géométrie corticale interne. Bref, les contours illusoires, en particulier courbes, constituent un phénomène neurogéométrique princeps qu'il s'agit d'expliquer mathématiquement.

La première section (1) est expérimentale et traite de résultats, certains bien connus et d'autres moins connus, concernant les contours subjectifs, qu'ils soient modaux (réellement perçus) ou amodaux (seulement imaginés). Dès la section 2, nous nous focalisons sur les contours modaux. Un intérêt particulier est apporté aux triangles et carrés de Kanizsa, ces célèbres figures ou des "pacmen" (petits disques pleins auxquels on a enlevé un secteur angulaire) situés de façon cohérente aux sommets d'un polygone engendrent des bords illusoires qui reconstruisent le polygone. Un phénomène surprenant, dit "effet néon", est que, si les pacmen sont colorés, la couleur diffuse à l'intérieur du polygone. Mais la diffusion est arrêtée par les bords virtuels et ces derniers sont par conséquent neurophysiologiquement réels. En vision stéréoscopique le système visuel se montre même capable de reconstruire en 3D la surface d'un triangle illusoire (et pas seulement un contour) flottant au-dessus des pacmen.

De très nombreuses expériences neurophysiologiques (enregistrements d'électrodes, EEG, IRMf à partir des années 1995) ont étudié les corrélats neuronaux de ces phénomènes étonnants. Ils sont de bas niveau et bottom-up et remontent loin dans l'évolution biologique comme l'attestent de nombreuses expériences sur différentes

espèces animales (primates, chats, certains oiseaux). Il est donc d'autant plus important d'en proposer des modèles crédibles. Un certain nombre en ont été proposés depuis longtemps, par exemple par Grossberg et Mingolla, Peterhans et van der Heydt ou Heitger.

Nous avons nous-mêmes procédé à une expérience sur les contours illusoires avec Jacques Ninio à l'École Normale Supérieure sur les contours illusoires de Kanizsa courbes. Nous en présentons les résultats. Leur forme exacte renseigne sur leurs mécanismes producteurs qui, bien que tous équivalents pour les contours droits, deviennent néanmoins différents pour les contours courbes.

Beaucoup de données expérimentales de cette section renvoient aux nombreuses expériences exposées dans le Vol I : les mécanismes d'interpolation de Dresch, Grossberg, Mingolla, Shipley, les mécanismes d'extrapolation de Hoffman, Kellman, Shipley, les mécanismes de "feed-back figural" de Hoffman, Gregory, Grossberg, Kourzi, Lee, Mendola, Rubin, Stanley, la théorie du "champ d'association" de Fields, Hayes et Hess qui confirme psychophysiquement l'implémentation neuronale de la structure de contact dans le cortex visuel primaire.

Il existe aussi des contours illusoires en quelque sorte "du second ordre" comme les effets "de peigne" ("comb effect") qui apparaissent lorsque les extrémités d'une famille de segments ("grating") engendrent un contour illusoire orthogonal à la direction des segments ("abutting grating illusion"). Ils ont été étudiés entre autres par Peterhans et von der Heydt et sont engendrés par l'activation de cellules "end-stopped" détectant à l'extrémité d'un segment la direction orthogonale à sa direction. Ils mettent en jeu l'aire $V2$ et en termes neurogéométriques dépendent du fibré des repères mobiles composé des tangentes et des normales le long des courbes.

Il est important d'insister sur le fait que des expériences neurophysiologiques fines montrent que ces contours illusoires modaux à longue portée qui n'existent pas dans les données sensorielles sont toutefois neuralemement réels car les neurones qui les construisent sont bien activés même si ce n'est pas par le stimulus sensoriel qui est lacunaire. Ils résultent d'une intégration globale de bas niveau effectuée par l'architecture fonctionnelle du cortex visuel primaire à partir des quelques neurones effectivement activés par le stimulus. C'est un peu comme en physique lorsqu'un champ électro-magnétique est engendré par quelques charges disséminées. L'activation produite par le stimulus se propage le long des connexions cortico-corticales et c'est parce que cette connectivité implémente neuralemement des conditions d'intégrabilité globale que les contours illusoires à longue portée "sautent aux yeux".

La section 3 commence à investiguer les conditions de possibilité de la modélisation des contours illusoires modaux. Il faut trouver des principes de complétion très puissants qui dérivent des architectures fonctionnelles. L'idée s'impose d'elle-même assez vite qu'il faut avoir recours à des modèles variationnels. Ainsi s'ouvre une

nouvelle relation très profonde entre la neurogéométrie et des formalismes qui sont omniprésents en physique.

Pour les figures comme les triangles de Kanizsa dans \mathbb{R}^2 , des modèles variationnels planaires sans contenu neurophysiologique ont été introduits par Ullman et Horn à partir du milieu des années 1970 en utilisant des principes de minimisation de courbure et des “courbes d’énergie minimale”. Mais le premier modèle vraiment géométrique a été introduit par David Mumford en 1992 dans son célèbre article “Elastica and computer vision”. Ce modèle utilise la théorie des *elasticae* remontant à Euler et interprète un contour illusoire reliant le couple (a, p) d’un point a et d’une direction p à un autre couple (b, q) d’un point b et d’une direction q comme une sorte de lame élastique d’extrémités (a, p) et (b, q) fixées. Le modèle variationnel minimise une énergie du type $E = \int (\alpha\kappa^2 + \beta) ds$ (un mixte de longueur s et de courbure κ) et Mumford (i) en a donné une interprétation stochastique, (ii) en a dérivé une équation différentielle sur la courbure $\kappa(s)$ et (iii) a montré que la longueur s s’exprime au moyen d’une intégrale elliptique sur la courbure. L’intégration du problème des *elasticae* par des fonctions elliptiques se trouve déjà explicitement chez Euler même si la notion de fonction elliptique n’avait pas encore été dégagée en tant que telle.

Toutefois, les modèles purement planaires d’Ullman, Horn et Mumford appliquent un principe variationnel directement aux courbes du plan \mathbb{R}^2 . Ils n’ont pas de contenu neurophysiologique précis car ils ne font pas appel à l’architecture fonctionnelle du cortex visuel. En revanche, dans la mesure où la neurogéométrie repose sur la mathématisation de cette dernière, il est naturel d’y repenser les modèles variationnels des contours illusoires modaux. Pour cela nous introduisons l’hypothèse qu’ils sont les géodésiques d’une métrique sous-riemannienne définie sur la structure de contact des 1-jets.

Le chapitre se conclut donc par une courte section 4 effectuant le passage des modèles planaires de type *elasticae* dans \mathbb{R}^2 aux modèles géodésiques dans $(\mathbb{V}, \mathcal{K}) \simeq \mathbb{R}^3$ et \mathbb{H}_{pol} . Nous avons introduit l’hypothèse géodésique au milieu des années 1990 et nous l’avons développée en deux temps. Un premier modèle, que nous avons appelé celui des “géodésiques legendriennes”, se donnait un lagrangien dans \mathbb{V} et calculait les équations d’Euler-Lagrange. Au début des années 2000 nos prises de contact avec les spécialistes de géométrie sous-riemannienne du groupe d’Andrei Agrachev, en particulier Ugo Boscain et Jean-Paul Gauthier, ainsi qu’avec Alessandro Sarti et Giovanna Citti ont conduit à l’approfondissement du modèle comme modèle géodésique sous-riemannien.

9.9. Chapitre 9

La modélisation neurogéométrique des contours illusoires modaux conduit ainsi à s’aventurer dans l’univers des modèles variationnels classiques, univers tellement

riche historiquement, théoriquement et techniquement, et si universel en physique et dans beaucoup d'autres disciplines (comme par exemple l'économie), que nous lui consacrons le long chapitre 9. Nous y revenons sur son odyssée mathématique depuis Euler, Bernoulli, Lagrange et Hamilton.

Un point essentiel à comprendre est que les phénomènes naturels expliqués par les principes variationnels sont souvent très simples mais que les calculs qui les reconstruisent et en effectuent la "synthèse computationnelle" ne sont en général pas simples du tout. Rien de plus simple physiquement qu'une "chaînette", cette courbe décrite par une chaîne dont les extrémités sont suspendues à deux clous et sur laquelle n'agit que la pesanteur. Et pourtant une telle courbe mécanique "caténaire" est mathématiquement une courbe transcendante, en fait un cosinus hyperbolique. La dialectique entre simplicité physique et complexité mathématique est cruciale. La nature nous parle peut-être le langage mathématique comme le disait Galilée mais son langage mathématique n'est pas forcément celui de structures mathématiques simples comme le pensaient les grecs. Il est aussi celui de structures *sui generis* à découvrir et à explorer.

Insistons-y. Depuis la révolution newtonienne conduisant à la description mathématique des phénomènes naturels au moyen d'équations différentielles, la complexité computationnelle nécessaire pour calculer les phénomènes physiques, même parmi les plus simples, est devenue une évidence. Et l'on ne voit pas pourquoi cela ne concernerait pas aussi la neurophysiologie. Il n'y a donc rien de surprenant à ce qu'un phénomène naturel apparemment aussi simple qu'un contour illusoire courbe exige, comme une chaînette ou une lame élastique en mécanique, une grande complexité mathématique pour être calculé en termes de principes variationnels.

Ce chapitre de retour amont revient brièvement sur l'émergence d'outils mathématiques qui s'avèreront fondamentaux pour nos modèles géodésiques.

Il commence (section 1) par des repères biographiques sur les principaux inventeurs et contributeurs des principes variationnels depuis Descartes : Fermat, Toricelli, Wallis, Huygens, Barrow, Wren, Neil, Newton, Leibniz (une figure particulièrement importante), les Bernoulli, Varignon, de l'Hospital, De Moivre, Fagnano, Taylor, MacLaurin, Euler, D'Alembert, Lagrange, Legendre, Gauss, Cauchy, Abel, Jacobi, Liouville, Weierstrass, Riemann.

Ensuite, la section 2 rappelle que, très tôt, les calculs de l'intégrale de certaines fonctions simples ("quadratures") ou de la longueur de certaines courbes simples géométriques ou mécaniques ("rectification") se révélèrent étonnamment compliqués. Ce fut une grande surprise et déclencha au tournant du XVII^e-XVIII^e siècle une riche période quasi expérimentale de découverte de "l'intelligence de lignes courbes" et de courbes mécaniques "produites par la nature sans aucun artifice" comme le disait si justement Jacques Bernoulli. Cette expression nous ravit car nous parlerions volontiers pour la neurogéométrie de "l'intelligence des contours

subjectifs” “produits par le cortex visuel sans aucun artifice”. Nous pensons que leur analyse théorique et leur synthèse computationnelle correspond pour les neuromathématiques à un moment analogue à celui de la compréhension des principes variationnels en mécanique.

La rectification de l’ellipse (section 3) joua un rôle particulièrement déterminant puisqu’il est à l’origine des intégrales et des fonctions elliptiques. Nous en rappelons l’histoire. Le calcul sur ces fonctions a été élaboré en analogie avec celui des fonctions trigonométriques classiques (dites aussi “circulaires”) et la grande découverte fut qu’il s’agissait en fait de fonctions définies sur \mathbb{C} et doublement périodiques. Elle déclencha une période exceptionnelle de créativité mathématique.

Après avoir indiqué dans la section 4 que, confrontés à la difficulté du problème, les mathématiciens de l’époque cherchèrent au moins des développements en séries permettant d’approximer les intégrales et fonctions elliptiques, nous rappelons scolairement à la section 5 les principales propriétés des fonctions circulaires et des fonctions hyperboliques et leurs liens avec la rectification du cercle $x^2 + y^2 = 1$ et de l’hyperbole équilatère $x^2 - y^2 = 1$ au moyen des intégrales $\int_0^x \frac{1}{\sqrt{1 \pm t^2}} dt$.

La section 6 rappelle ensuite avec un certain détail les travaux sur la courbe surprenante qu’est la lemniscate qui, avec les Bernoulli, fournit le premier exemple explicite de nouvelle trigonométrie généralisant les trigonométries circulaire et hyperbolique classiques. À partir des découvertes remarquables du comte de Fagnano au début du XVIII^e siècle sur les propriétés géométriques exceptionnelles de la lemniscate et sur les relations algébriques d’addition, de duplication et de division entre ses arcs, Euler se convainquit qu’il devait exister une trigonométrie “lemniscatique” qui serait aux intégrales $\int_0^x \frac{1}{\sqrt{1-t^3}} dt$ ou $\int_0^x \frac{1}{\sqrt{1-t^4}} dt$ ce que la trigonométrie classique est aux intégrales $\int_0^x \frac{1}{\sqrt{1-t^2}} dt$. Cette idée fut approfondie par le jeune Gauss qui comprit qu’en étendant les fonctions au corps des nombres complexes \mathbb{C} on aboutissait à des fonctions doublement périodiques. Puis ensuite vint Abel dont les recherches géniales conduisirent à la théorie moderne des intégrales et fonctions elliptiques.

La section 7 rappelle l’autre source principale de cette théorie, à savoir l’étude des mouvements complets du pendule au-delà des petites oscillations. Il s’agissait de comprendre la loi temporelle du pendule sur les arcs de cercle et comment la trigonométrie circulaire classique des arcs se convertit en trigonométrie “chronométrique” des intervalles de temps. C’est de façon naturelle qu’apparurent alors les intégrales elliptiques $\int_0^\varphi \frac{1}{\sqrt{1-k^2 \sin^2(\psi)}} d\psi$ et les fonctions elliptiques qui en sont les fonctions inverses. On obtint ainsi, surtout avec les travaux magistraux de Jacobi, la trigonométrie “elliptique” avec ses trois fonctions de base $\text{sn}(u)$, $\text{cn}(u)$ et $\text{dn}(u)$ paramétrées par leur module k . Nous explicitons en détail leurs graphes, leurs périodes, leurs translations par des demi- ou des quarts de période, leur dérivation, leurs diverses formules remarquables en particulier d’addition, leur extension à \mathbb{C} , leur double périodicité, leur limite pour $k \rightarrow 1$ (elles changent très rapidement au

voisinage de 1). Pour $k = 0$ la trigonométrie elliptique redonne la trigonométrie circulaire. Pour $k = 1$ elle donne une sorte de trigonométrie hyperbolique “inversée”. Elle interpole donc entre ces deux limites.

Après ces premiers éléments de l’histoire des fonctions elliptiques, nous revenons dans la section 8 sur l’extraordinaire *Methodus* d’Euler de 1744 dont la théorie des elasticæ est une Annexe. Il s’agit du traité fondateur du calcul des variations décrivant les équations différentielles (les célèbres équations d’Euler renommées ensuite d’Euler-Lagrange) donnant les solutions des problèmes variationnels. Le texte, rédigé en latin, est méthodologiquement impressionnant. Dans la lignée des principes de moindre action déjà connus en optique et en mécanique il cherche, étant donnée une classe de courbes satisfaisant certaines contraintes, celles qui maximisent ou minimisent des fonctionnelles.

Nous nous arrêtons sur ce texte parce qu’il montre à quel point la neurogéométrie remonte aux principes mêmes du calcul différentiel. En effet, le point le plus remarquable est qu’Euler ne traite pas les dérivées des courbes $y = f(x)$ comme des dérivées mais comme de nouvelles variables indépendantes p, q, r , etc., comme des symboles qui, seulement dans un second temps, seront interprétés comme des dérivées $p = \frac{dy}{dx}$, $q = \frac{d^2y}{dx^2}$, $r = \frac{d^3y}{dx^3}$, etc. Autrement dit, il travaille d’emblée dans ce qui sera plus tard formalisé rigoureusement par les espaces de jets et conçoit de fait une courbe plane comme une courbe dans l’espace de dimension infinie de ses jets, c’est-à-dire comme un développement de Taylor. Le principe sur lequel nous avons tellement insisté, à savoir le remplacement des dérivées par des symboles de “dérivées cachées”, se trouve donc déjà à la base du calcul des variations. Le calcul symbolique prolongeant les x, y avec les p, q, r , etc., permet de n’avoir que dx comme différentielle, les autres différentielles devenant $dy = p dx$, $dp = q dx$, $dq = r dx$, etc. Il y a déjà chez Euler la conscience aiguë que l’on peut considérer une dérivée p non plus comme une limite numérique (une “fluxion” au sens de Newton), mais comme une équation symbolique $dy - p dx = 0$. Nous voyons à quel point nos modèles neurogéométriques s’enracinent dans les origines du calcul différentiel et du calcul variationnel et, en retour, les éclairent sous un nouveau jour.

Le *Methodus* repose sur des fonctionnelles qui sont des fonctions dont les arguments ne sont plus des points mais des courbes et Euler en cherche les extrema en essayant d’y généraliser les formules valables pour les simples fonctions : $F(x)$ est extrémale si $F'(x) = 0$, minimale si $F'(x) = 0$ et $F''(x) > 0$, etc. Il donne une foule d’exemples retrouvant nombre de courbes spéciales déjà connues à son époque (comme la brachistochrone). Nous explicitons certains de ses calculs et montrons comment émerge l’équation d’Euler et la notion de “multiplicateur” lorsqu’il s’agit d’une optimisation sous contrainte.

La section 9 explique brièvement la façon dont on passe d’Euler à Lagrange et comment la physique héritée de Newton a été complètement repensée de façon

systématique. Euler a beaucoup insisté sur le fait que le calcul des variations reposait sur une mutation de la causalité et remplaçait les causes efficientes par des causes finales “à l’aide de la méthode des maxima et minima”. Le calcul des variations est certes un puissant outil mathématique mais il est aussi un principe physique universel possédant un profond sens métaphysique bien explicité par Leibniz.

Nous présentons alors le formalisme lagrangien. Étant donné un lagrangien

$$L(t, x(t), \dot{x}(t)),$$

optimiser l’intégrale d’action $S(\gamma) = \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t)) dt$ revient

- (i) à se placer dans l’espace des 1-jets des courbes $x(t)$ muni de sa structure de contact $\omega = dx - p dt$,
- (ii) à considérer les relevées legendriennes $\gamma(t)$ des trajectoires dans le plan (t, x) satisfaisant certaines conditions aux limites, puis enfin
- (iii) à optimiser $S(\gamma)$.

L’hypothèse géodésique sur les contours illusoires est donc un exemple neuro-physiologique d’optimisation lagrangienne.

La section 10 présente la magistrale théorie des *elasticæ* développée par Euler dans l’Annexe de son *Methodus*. Le problème est celui de savoir quelle forme prend une lame élastique dont les extrémités et les pentes aux extrémités sont fixées. Il fut proposé en 1742 à Euler par Daniel Bernoulli et Euler l’a complètement résolu en le traduisant en un problème variationnel. La fonctionnelle proposée par Bernoulli était $\int \frac{ds}{R^2}$ où R est la courbure et ds l’élément de longueur d’arc. Euler la réécrit $\int \frac{q^2}{(1+p^2)^{\frac{3}{2}}} ds$ avec $dy = p dx$, $dp = q dx$ et $ds = \sqrt{1+p^2} dx$. L’équation d’Euler permet alors d’exprimer y comme une intégrale $\int \frac{P}{\sqrt{Q}}$ où P est un polynôme du second degré en x et Q un polynôme du quatrième degré. Les coefficients de P et Q reflètent les conditions initiales.

Ces intégrales étaient inconnues à l’époque (ce sont les futures intégrales elliptiques) et par des changements de variables appropriés Euler les ramène à des formes normales et en donne la classification complète en 9 classes. Sa démarche est tout à fait admirable.

Nous concluons ce résumé de l’Annexe du *Methodus* par un rappel sur la profonde équivalence établie par Kirchoff en 1859 entre la géométrie des *elasticæ* et la mécanique du pendule. La dynamique des *elasticæ* est simple : les courbes sont paramétrées par leur longueur d’arc et sont donc de vitesse unitaire. Mais leur géométrie est compliquée. À l’inverse la géométrie du pendule est simple : il décrit des arcs de cercle. Mais sa loi temporelle est compliquée. Elle est en fait équivalente à la variation de l’angle de la tangente le long d’un *elastica*.

La dernière section (11) du chapitre est un petit *vademecum* sur les intégrales, les fonctions et les courbes elliptiques. Elle récapitule leur classification due à Legendre et leurs propriétés de base, elle évoque les courbes elliptiques sur \mathbb{C} comme fonctions

doublement périodiques (donc en fait définies sur des tores) et rappelle leur exemple standard dû à Weierstrass.

9.10. Chapitre 10

Le chapitre 10 développe en détail, dans un style eulérien-lagrangien classique, le premier modèle variationnel de $V1$ traitant les contours illusoires comme des géodésiques. Nous l'avons élaboré au milieu des années 1990 sous le titre de "géodésiques legendriennes". Le chapitre initie un long parcours de plusieurs chapitres consacrés à cette idée directrice.

Le modèle part de la structure de contact $\mathbb{V}_J = (\mathbb{V}, \mathcal{K})$ modélisant $V1$ et utilise la métrique euclidienne du \mathbb{R}^3 sous-jacent à \mathbb{V} . Après l'avoir situé dans la classe des modèles variationnels possibles (section 1) et en avoir justifié l'étude, nous en donnons les équations d'Euler-Lagrange (section 2). Il s'agit de minimiser la longueur d'une relevée legendrienne d'extrémités (a_1, p_1) et (a_2, p_2) fixées dans \mathbb{V} en restant dans la classe des intégrales de la structure de contact. Si la courbe plane γ de relevée legendrienne Γ est d'équation $y = f(x)$, la fonctionnelle de longueur correspondant à la longueur d'arc dans \mathbb{V} (et non pas dans la base) a pour lagrangien $L = \sqrt{1 + f'(x)^2 + f''(x)^2} = \sqrt{1 + p^2 + q^2}$. La condition pour une courbe Γ dans \mathbb{V} d'être une intégrale de \mathcal{K} (et donc une relevée) impose la contrainte $\Sigma = p - f'(x) = 0$ et introduit un multiplicateur de Lagrange λ , le lagrangien contraint devenant $L + \lambda\Sigma$.

On en tire les équations d'Euler-Lagrange et une équation différentielle du second ordre pour la fonction $g = f'$ du type $(g')^2 = Q(g)$ où Q est une fraction rationnelle quotient d'un polynôme du quatrième degré par un polynôme du second degré dont les coefficients dépendent de deux constantes d'intégration A et B exprimant les conditions aux limites.

La fonction $g(x)$ est une fonction de x . La fonction réciproque $x(g)$ est une intégrale elliptique et nous en calculons des exemples dans le cas où f est paire, la courbe $y = f(x)$ étant symétrique en forme d'arche. Cela implique $A = 0$ et $B \neq 0$, $k = 1/B$ étant le module de l'intégrale elliptique, qui, dans ce cas, est de première espèce (section 3). Nous calculons numériquement la projection $y = f(x)$ de la géodésique en fonction des angles des pentes aux extrémités et nous la comparons aux arcs de cercle, aux arcs de parabole et aux courbes linéaires par morceaux de mêmes conditions aux limites. Cela montre que des mesures assez fines sur les contours illusoires pourraient permettre de départager entre eux différents modèles.

Après avoir approfondi pour être complet (section 4) l'autre cas extrême $A \neq 0$, $B = 0$ (bien qu'il ne soit pas utile pour la modélisation) nous donnons dans la section 5 l'équation différentielle que satisfait la courbure des géodésiques legendriennes. À titre d'exercice la section 6 reprend les résultats précédents en effectuant les

calculs à partir de la longueur d'arc dans l'espace de base (lagrangien $\sqrt{1 + f'(x)^2} = \sqrt{1 + p^2}$) et de l'angle de la tangente θ .

Puis nous considérons brièvement le modèle du fibré en cercles dans la section 7 et le modèle des courbes de moindre énergie dans la section 8.

Enfin dans une dernière section (9) nous concluons le chapitre sur une reprise des *elasticæ*. Dans le chapitre 9, section 10, le problème était traité de façon eulérienne comme un problème variationnel dans l'espace de base. Mais on peut aussi l'exprimer comme un problème variationnel concernant les relevés legendriennes dans la structure de contact $(\mathbb{V}, \mathcal{K})$ et construire son lagrangien. Étant données les mêmes conditions initiales, on peut alors comparer les trois solutions variationnelles du problème des contours illusoires : les *elasticæ*, les projections de géodésiques legendriennes et les projections des courbes d'énergie minimale. Elles sont assez différentes pour que, comme nous l'avons déjà évoqué plus haut, des mesures expérimentales assez fines puissent les séparer et par conséquent donner des indices sérieux sur le type de principe variationnel commandant effectivement les données empiriques.

9.11. Chapitre 11

Après avoir explicité au cours des précédents chapitres les différents niveaux de structure du modèle du fibré des 1-jets \mathbb{V}_J , sa structure de groupe de Lie (groupe de Heisenberg polarisé), sa structure de contact L -invariante et ses géodésiques legendriennes, nous en enrichissons l'analyse en passant au fibré principal $SE(2)$ de \mathbb{V}_J . $G = SE(2)$ est le groupe des isométries du plan de base produit semi-direct $SE(2) = \mathbb{R}^2 \rtimes SO(2)$ des translations et des rotations. Cela revient à passer des 1-jets aux repères mobiles de Cartan, passage neurologiquement pertinent si l'on tient compte de l'aire $V2$ et du sens kinesthésique du mouvement (section 1).

La section 2 expose la méthode du repère mobile le long d'une courbe γ du plan de base, les repères de Frenet (cf. chapitre 5, section 7) étant constitués en chaque point de la tangente unitaire et de la normale unitaire orientée dans le sens direct. Les repères mobiles sont des éléments du groupe $G = SE(2)$ et relèvent γ en une courbe Γ dans le fibré $SE(2) = \mathbb{R}^2 \rtimes SO(2) \rightarrow \mathbb{R}^2$. C'est ce "relèvement de Frenet" qui correspond à notre second modèle \mathbb{V}_S de fibre $SO(2) \simeq \mathbb{S}^1$.

La section 3 revient alors avec plus de détails sur la structure de groupe de Lie de $G = SE(2)$, détaille son algèbre de Lie, ses représentations adjointe et co-adjointe ainsi que leurs orbites. Puis la section 4 revient quant à elle sur la forme de Maurer-Cartan $\Lambda_G = g^{-1}dg$ (où g est un élément générique de G) satisfaisant l'équation $d\Lambda_G = -[\Lambda_G, \Lambda_G]$.

Cela nous permet de reprendre, en l'approfondissant, le problème des géodésiques legendriennes traité dans le chapitre précédent comme un cas particulier de la théorie

de Robert Bryant et Phillip Griffiths. C'est l'objet de la dernière section 5. Le formalisme du système de Pfaff des relevées de Frenet est plus riche que celui des relevées legendriennes. Bryant et Griffiths traitent de différentes fonctionnelles et explicitent dans leur formalisme les équations d'Euler-Lagrange. Elles sont apparemment compliquées mais éclairent l'essence des problèmes variationnels sur $G = SE(2)$. Nous revenons alors dans ce contexte élargi sur le modèle des géodésique legendriennes, sur celui des fibrés en cercles et sur celui des *elasticae*.

9.12. Chapitre 12

Le chapitre 12 introduit une nouvelle dimension de la neurogéométrie. Les modèles de géodésiques legendriennes reposent sur des lagrangiens. Il s'agit de les reformuler en termes de géométrie sous-riemannienne au moyen d'un double approfondissement (section 1) :

- (i) en passant d'un formalisme lagrangien à un formalisme hamiltonien symplectique,
- (ii) en utilisant le principe du maximum de Pontryagin et la théorie du contrôle optimal.

Cette nouvelle étape a été essentiellement accomplie au début des années 2000 et a profité de notre rencontre avec le groupe d'Andrei Agrachev de la SISSA¹⁹ de Trieste et aux multiples initiatives de Giovanna Citti et Alessandro Sarti manifestant leur intérêt pour la neurogéométrie.

Le formalisme hamiltonien étant universel et, comme le formalisme lagrangien qu'il généralise, hégémonique dans nombre de disciplines, une section (la section 2) en expose pédagogiquement les principaux aspects. Elle part de la notion de transformée de Legendre qui permet de convertir un formalisme dans l'autre et l'illustre à partir d'exemples élémentaires. Elle précise alors ce que l'on appelle les conditions de Legendre et de Weierstrass.

Une propriété essentielle des formalismes lagrangiens et hamiltoniens est le théorème de Noether établissant une correspondance entre les symétries du système considéré (son groupe de relativité) et les lois de conservation de ses grandeurs caractéristiques (les intégrales premières de son évolution). Ce théorème crucial est approfondi par le formalisme de "l'application moment" (développé par Kostant, Souriau et d'autres éminents spécialistes de géométrie symplectique) qui permet d'associer à tout groupe G agissant symplectiquement sur une variété symplectique des intégrales premières de tout hamiltonien G -invariant. On peut même aller encore plus loin en tenant compte du fait que, comme l'a montré Kirillov, les orbites de la représentation co-adjointes de G sont canoniquement munies d'une structure symplectique sur laquelle G opère symplectiquement et qu'on peut donc appliquer

19. Scuola Internazionale Superiore di Studi Avanzati.

le formalisme de l'application moment. On obtient ainsi des intégrales premières universelles pour G et des hamiltoniens en quelque sorte "a priori".

Dans le cadre du contrôle optimal le formalisme hamiltonien permet une vaste généralisation des calculs variationnels classiques. La section 4 introduit à cet immense univers. Il concerne les systèmes dont les évolutions ne sont pas simplement régies par des lagrangiens mais sont en plus contrôlées par des paramètres de contrôle. On passe de systèmes optimisant des lagrangiens $L(q, \dot{q})$ à des systèmes dépendant de contrôles u et optimisant des lagrangiens $L(q, u)$ (un "coût", une "énergie"), la question étant maintenant de trouver le contrôle optimal. Tout un ensemble de situations sont possibles avec des conditions et des contraintes spécifiques. Les équations d'Euler-Lagrange du calcul variationnel classique se trouvent alors généralisées par le principe du maximum de Pontryagin (PMP, section 5) qui transforme le problème des extrémales d'un problème variationnel dans une variété M en un problème hamiltonien dans le fibré cotangent T^*M et retrouve les solutions dans M en projetant sur M les solutions dans T^*M .

On voit que ce formalisme est par essence adapté à l'interprétation neurogéométrique des contours illusoire comme géodésiques legendriennes puisqu'il consiste à relever dans T^*M les trajectoires possibles dans M , à chercher la solution optimale dans T^*M et à la reprojeter sur M . Il a l'avantage sur les approches lagrangiennes des chapitres précédents 10 et 11 de pouvoir être adapté au choix d'une métrique sur $(\mathbb{V}, \mathcal{K})$. Au lieu de choisir une métrique riemannienne sur \mathbb{V} on choisit une métrique sous-riemannienne sur \mathcal{K} ce qui fait que la structure de contact et la métrique deviennent solidaires.

L'idée directrice du contrôle optimal est profonde. On relève les champs de vecteurs X sur une variété M en champs hamiltoniens $\overrightarrow{h_X}$ sur le fibré cotangent T^*M en passant par l'intermédiaire d'hamiltoniens $h_X(p) = \langle p, X \rangle$ sur T^*M (les p sont les covecteurs de T^*M et agissent comme formes linéaires sur les X de TM). Les contrôles imposent de choisir X dans une classe de champs $X(u)$ paramétrés par u et l'on choisit l'hamiltonien $H = \langle p, X_{\text{optimal}} \rangle$ qui optimise les hamiltoniens $h_{X(u)}(p) = \langle p, X(u) \rangle$. Les projections des trajectoires H sont alors les solutions du problème variationnel considéré dans M . Lorsqu'en plus le problème est régi par un lagrangien $L(q, u)$, l'hamiltonien H devient $\langle p, X(u) \rangle + \lambda L(q, u)$ où λ est un multiplicateur de Lagrange. Le PMP exprime que l'on peut obtenir ainsi l'essentiel des extrémales "normales" du problème variationnel considéré. Il est assez technique et nous en explicitons pédagogiquement la démonstration.

Une fois exposés les éléments de la théorie du contrôle optimal et le PMP, nous revenons section 6 aux modèles géodésiques et aux caustiques comme solutions lagrangiennes. Nous pouvons maintenant les reformuler comme des problèmes de géométrie sous-riemannienne avec leurs trajectoires, leurs fronts d'onde, leurs points conjugués et leurs points de coupure. Nous présentons d'abord de façon détaillée les caustiques

en optique géométrique qui offrent un exemple très compréhensible et calculable de façon pas trop compliquée. D'autres exemples classiques et computationnellement bien maîtrisés sont fournis par les déformations de la sphère en ellipsoïdes, qu'ils soient de révolution ou triaxiaux. Un de leurs intérêts est de voir comment se lève les dégénérescences imposées par les symétries de la sphère où le point antipodal \hat{A} d'un point A est à la fois un point conjugué et un point de coupure de tous les méridiens issus de A . Lorsqu'on brise la symétrie et que l'on passe de la sphère à un ellipsoïde de révolution, le point dégénéré \hat{A} associé à A se déploie génériquement (i.e. si A n'est pas un point "ombilic") en une astroïde de points conjugués à quatre cusps et un cut locus qui est un segment joignant deux cusps opposés. Cette belle géométrie remonte à Jacobi (1842).

Techniquement, la résolution peut s'effectuer dans des coordonnées adaptées (dites "elliptiques") introduites par Jacobi et fait intervenir des fonctions et des intégrales elliptiques. Depuis certains travaux de Poincaré, on a généralisé ce genre de résultats à des déformations riemanniennes convexes quelconques de la sphère. Mais la "dernière conjecture géométrique" de Jacobi disant que la situation générique des ellipsoïdes de révolution reste valide dans le cas des ellipsoïdes triaxiaux n'a été démontrée qu'en 2004 par Jin-ichi Itoh et Kazuyoshi Kiyohara.

9.13. Chapitre 13

Une fois que l'on dispose des formalismes de géométrie sous-riemannienne et de la théorie du contrôle optimal, on peut approfondir notablement les modèles neurogéométriques. Les deux chapitres suivants s'y consacrent.

Le chapitre 13 compile et illustre les magnifiques travaux de Yuri Sachkov sur la reformulation de la théorie des *elasticæ* et du modèle de David Mumford des contours illusoires. Les calculs sont techniques mais leur architecture est très bien présentée par l'auteur et particulièrement éclairante. Il est intéressant qu'un problème si ancien puisse être encore si innovant environ deux siècles et demi après sa première résolution par Euler.

Le problème des *elasticæ* était initialement posé dans le plan \mathbb{R}^2 . Nous avons vu dans le chapitre précédent comment il peut être relevé en un problème variationnel dans $\mathbb{V} = \mathbb{V}_J$ (l'espace des 1-jets des courbes planes) et dans $\mathbb{V} = \mathbb{V}_S = SE(2) = \mathbb{R}^2 \rtimes SO(2)$. Sachkov le traduit en un problème de contrôle optimal dans $T^*\mathbb{V}$ et en déduit le système explicite des équations de Hamilton (section 1). Il retrouve les liens avec la mécanique du pendule (section 2) ce qui permet de faire se correspondre la classification des *elasticæ* avec celle des trajectoires du pendule.

Pour retrouver la classification d'Euler, Sachkov utilise les coordonnées introduites par Jacobi (section 3) et les représentations en termes de fonctions elliptiques. Il aboutit ainsi à une stratification des espaces cotangents $T_{q_0}^*\mathbb{V}$ (section 4) et, à partir de là, déploie de vastes calculs pour expliquer structurellement l'origine

des différents types d'elastica. Avec son travail, ce que Bernoulli appelait "l'intelligence des lignes courbes" trouve une expression mathématique particulièrement élégante.

À titre d'exercice, nous calculons numériquement de nombreux exemples en fonction du module k intervenant dans les intégrales elliptiques. Pour $k \in [0, 1]$ (elastica inflexionnels et pendule oscillant, sections 5 et 6) on part de l'elastica sans courbure ($k = 0$) et on suit ses déformations en passant par l'elastica rectangulaire initial de Bernoulli pour $k \sim 0.7$, puis par l'elastica en lemniscate pour $k \sim 0.9$, jusqu'à en arriver à l'elastica à une boucle pour $k = 1$. Lorsque $k > 1$, les boucles prolifèrent et deviennent de plus en plus serrées jusqu'à en arriver au cercle pour k infini (sections 7 et 8).

Sachkov reformule donc en la systématisant la classification d'Euler-Jacobi en termes de contrôle optimal (section 9). Le grand intérêt de son travail est le calcul (très technique) des points conjugués, des points de coupure et des points de Maxwell où différents arcs d'un elastica entrent en compétition. Ces calculs innovants utilisent de façon systématique les symétries des équations des elastica (section 10). La géométrie du cut locus est explicitée, ainsi que celle des points conjugués (section 11) ce qui n'avait pas été fait jusque-là.

9.14. Chapitre 14

Cette reformulation du modèle des elastica permet, en s'en inspirant, de reprendre dans le cadre du contrôle optimal nos modèles géodésiques sous-riemanniens \mathbb{V}_J et \mathbb{V}_S pour les contours illusoires. C'est le but du chapitre 14 qu'on peut considérer comme un point culminant de notre géométrisation de l'architecture fonctionnelle de $V1$. Le lien avec l'analyse fonctionnelle du traitement du signal au moyen d'états cohérents vient plus tard, dans le chapitre 17.

Une première section (1) clarifie le sens et la pertinence du passage à la géométrie sous-riemannienne permettant de définir naturellement les métriques sur les structures de contact. Le modèle \mathbb{V}_J isomorphe au groupe de Heisenberg polarisé a bénéficié des travaux de Beals, Gaveau et Greiner et d'autres spécialistes pionniers des années 1980. Le modèle \mathbb{V}_S que nous avons élaboré avec Giovanna Citti et Alessandro Sarti n'avait pas de précédents et a beaucoup bénéficié, répétons-le, des travaux d'Andrei Agrachev et de son équipe.

Nous commençons donc, dans la section 2, par les géodésiques sous-riemanniennes du groupe de Heisenberg polarisé $(\mathbb{V}_J, \mathcal{K})$ en adaptant les résultats de Beals-Gaveau-Greiner sur le groupe de Heisenberg symétrisé. La métrique sous-riemannienne y est définie en prenant la base naturelle des plans de contact comme base orthonormée. Les géodésiques sont données par la minimisation de la longueur des courbes intégrales de \mathcal{K} . En appliquant le principe du maximum de Pontryagin on obtient

facilement l'hamiltonien et les équations de Hamilton permettant de les calculer explicitement. Leurs projections sur le plan (x, p) muni de sa structure symplectique canonique sont des cercles.

Les formules explicites des géodésiques expliquent leur étonnante multiplicité (inconnue en géométrie riemannienne). Elles font en effet intervenir la fonction impaire $\mu(\varphi) = \frac{\varphi}{\sin^2(\varphi)} - \cot(\varphi)$ qui diverge pour $\varphi = k\pi$ et présente des minima pour $\varphi = \tan(\varphi)$. Mais dans ce dernier cas $\mu(\varphi) = \varphi$ et les minima sont donc sur la diagonale. Comme pour obtenir les géodésiques reliant par exemple $0 \in \mathbb{V}_J$ à un autre point $v = (x, y, p)$ il faut résoudre une condition $\mu(\varphi) =$ une certaine fonction des (x, y, p) donnés on voit que les différentes géodésiques allant de 0 à v sont associées à des solutions de $\mu(\varphi) =$ constante et plus la constante est grande plus celles-ci sont nombreuses.

La sphère sous-riemannienne S centrée sur 0 , lieu des extrémités des géodésiques qui partent de 0 , sont de longueur 1 et sont des minimiseurs globaux, possède une forme caractéristique. C'est une sorte de sphère dont les voisinages des pôles (intersections de S avec l'axe vertical des p) sont comme des surfaces col admettant aux pôles une singularité conique rentrante (qui ressemble à la singularité de l'épicarpe d'une pomme au voisinage du pédoncule). Quant au front d'onde, lieu des extrémités des géodésiques qui partent de 0 , sont de longueur 1 , sont des minimiseurs locaux mais pas forcément globaux, il prolonge S vers l'intérieur en une sorte de pile de disques à bord cuspidal de plus en plus petits qui s'accumulent en convergeant vers 0 . Le cut locus de 0 est l'axe vertical $p - \{0\}$ et 0 fait partie de son adhérence. Tous les points de cet axe sont conjugués de 0 . Ce comportement est très singulier et totalement impossible en géométrie riemannienne.

La section suivante (3) évoque brièvement les espaces de jet d'ordre $k > 1$ en particulier les 2-jets (i.e. la structure d'Engel). Un outil intéressant développé par Richard Montgomery et ses collègues consiste à étudier la façon dont les projections des géodésiques sur le plan (x, p_k) (où p_k est le symbole de la "dérivée cachée" d'ordre k) complexifient les cercles (x, p) du cas $k = 1$. Ardentov et Sachkov ont montré que pour $k = 2$ les projections sur (x, p_2) sont les *elasticae*, résultat qui permet d'ailleurs de retrouver la classification de ces derniers.

La longue section 4 est particulièrement importante pour notre propos car elle concerne l'approche des géodésiques sous-riemanniennes du second modèle \mathbb{V}_S en termes de théorie du contrôle optimal. Elle reprend notre présentation [434] en 2006 à la Société Mathématique de France. Nous nous appuyons sur les travaux d'Agrachev et de Sachkov qui ont compris qu'il y a en arrière-plan, comme pour les *elasticae*, l'équation d'un pendule dont le portrait de phase pilote la classification des géodésiques. La démarche est la même que pour \mathbb{V}_J : hamiltonien sur $T^*\mathbb{V}_S$,

équations de Hamilton, géodésiques comme projections sur \mathbb{V}_S des trajectoires hamiltoniennes. Nous enrichissons leurs calculs par de nombreuses illustrations originales détaillées qui constituent une part importante du chapitre. Les formules sont explicites en termes d'intégrales elliptiques et ont permis à Sachkov et Moiseev d'effectuer la classification des géodésiques. Elles peuvent apparaître compliquées au prime abord mais on y discerne très bien une architecture assez simple. Nous consacrons un assez long développement à leur illustration numérique. Cela permet de visualiser la complexité de la géométrie sous-riemannienne naturelle du groupe $SE(2)$ muni de sa structure de contact, géométrie dont on peut s'étonner qu'il ait fallu attendre si longtemps pour la mettre au jour.

Les sphères et les fronts d'onde de $\mathbb{V}_S = SE(2)$ ressemblent à ceux de \mathbb{V}_J pour un rayon petit, ce qui est normal puisque \mathbb{V}_J est en quelque sorte la structure tangente à \mathbb{V}_S pour des angles θ assez petits. Mais ils deviennent très différents et fort compliqués lorsque le rayon augmente. C'est un véritable tour de force de Sachkov de les avoir explicités et, encore une fois, nous en illustrons très en détail plusieurs exemples.

Après l'analyse approfondie du modèle \mathbb{V}_S , nous concluons le chapitre par quelques investigations supplémentaires. Un problème intéressant vient du fait que, lorsqu'on les projette sur le plan base (x, y) pour obtenir des contours illusoires, les géodésiques peuvent admettre des *singularités cusp*. Or ce phénomène n'est pas observé empiriquement. Ce que l'on observe est plutôt que les contours illusoires que leurs conditions aux limites rendraient trop courbes relaxent leur énergie et deviennent linéaires par morceaux avec des points anguleux. Il en va d'ailleurs de même pour les lames élastiques : lorsqu'on les courbe trop elles se plient.

On peut alors imposer aux projections des géodésiques sur le plan base la contrainte de ne pas présenter de cusp en leur interdisant de posséder des tangentes "verticales", c'est-à-dire d'être tangentes en certains points aux fibres de la projection $\pi : \mathbb{V}_S \rightarrow \mathbb{R}^2$ (section 5). On obtient ainsi une variante de \mathbb{V}_S que nous appelons \mathbb{V}_S^\emptyset et qui a été développée par Sachkov, Boscain, Duits et Rossi. Ces auteurs étudient le domaine \mathcal{R} de $SE(2)$ atteignable (par exemple à partir de l'origine $0 \in \mathbb{V}_S$) par des géodésiques sans cusp. L'intérêt principal de ce modèle \mathbb{V}_S^\emptyset pour la neurogéométrie des contours illusoires est qu'il montre l'existence de "cônes d'accessibilité". Si l'on considère une projection sans cusp de géodésique partant de $0 \in \mathbb{V}_S$ (i.e. de $(x = 0, y = 0) \in \mathbb{R}^2$ avec une pente θ_0 nulle) pour arriver au point $(x_1, y_1) \in \mathbb{R}^2$ avec une pente θ_1 , on montre que toutes les autres géodésiques sans cusp partant de 0 et arrivant au même point (x_1, y_1) y arrivent avec un angle θ appartenant à un cône $[\theta_{\min}, \theta_{\max}]$ encadrant θ_1 . Nous pouvons alors faire l'hypothèse que, lorsqu'un contour illusoire devrait sortir du cône d'accessibilité qui lui correspond, il devient linéaire par morceaux avec un point anguleux à la place du cusp. Il y aurait donc une compétition entre deux stratégies de construction des contours illusoires

suivant les conditions aux limites. C'est bien le cas et elle est même spectaculaire dans l'illusion d'Ehrenstein.

Nous concluons ce long chapitre en indiquant quelques autres résultats (section 6).

- (i) Nous évoquons d'abord la notion de courbe de Jacobi permettant d'analyser la façon dont les géodésiques varient lorsque l'on varie leurs conditions initiales. De nombreux spécialistes l'ont approfondie et nous présentons la version d'Agrachev-Barilari-Boscain.
- (ii) Nous revenons ensuite brièvement sur la possibilité de donner une interprétation stochastique de nos modèles, comme David Mumford l'avait proposé pour les *elasticæ*. On obtient un processus d'advection-diffusion dont Sanguinetti, Citti et Sarti ont analysé l'équation de Fokker-Planck.
- (iii) Les progrès dans la compréhension de la structure sous-riemannienne de $SE(2)$ et de la géométrie globale de son cut locus ont poussé Boscain et Rossi à traiter de la même façon d'autres groupes de Lie, $SU(2)$, $SO(3)$ et $SL(2)$.
- (iv) Nous citons également les résultats d'Agrachev, Gauthier et Zakalyukin sur les caustiques génériques des structures de contact dans \mathbb{R}^3 .

En fait, c'est toute une nouvelle moisson de résultats qui se trouve découler de la compréhension en profondeur de la géométrie sous-riemannienne de groupes de Lie possédant une structure de contact invariante, groupes pourtant connus depuis longtemps.

9.15. Chapitre 15

Les chapitres précédents étaient centrés sur les modèles neurogéométriques appliquant certaines théories (groupes et algèbres de Lie, contrôle optimal, géométrie sous-riemannienne) à la géométrie des espaces de jets et en particulier à des structures de contact. Le chapitre 15 vise quant à lui à recontextualiser ces géométries très particulières dans leur univers mathématique propre. Il relève plus des mathématiques pures que des mathématiques appliquées. Fidèle à notre méthodologie, nous amplifions nos modèles en aval pour introduire aux théories abstraites dont ils sont des cas particuliers. Ce n'est plus l'illustration détaillée de cas concrets qui vient au premier plan mais la présentation de structures générales. En ce sens, on peut dire que ce chapitre constitue un petit *vademecum* introduisant à des théories d'un niveau nettement supérieur à celui de nos modèles.

La première section (1) revient sur la notion générale de distribution \mathcal{D} de type (k, n) de k -plans tangents dans le fibré tangent TM d'une variété M de dimension n . \mathcal{D} est en général définie par un système de Pfaff. Elle possède une suite centrale descendante et une structure en drapeau qui, lorsqu'elles sont régulières (c'est-à-dire lorsque les dimensions des sous-espaces sont les mêmes en tout point) définissent des

“vecteurs de croissance” (“growth vectors”). La structure en drapeau de \mathcal{D} est une filtration à laquelle on associe naturellement une graduation $\text{gr}(\mathcal{D})$.

L’un des premiers objectifs d’une théorie générale est d’obtenir des formes normales locales des distributions \mathcal{D} afin de pouvoir les classifier. Pour cela des outils comme ceux de stratification et de transversalité utilisés massivement depuis Whitney, Thom, Arnold et tant d’autres spécialistes sont indispensables. À titre d’exemple nous donnons la classification des germes de 1-formes sur \mathbb{R}^3 qui nous ont intéressés au premier chef dans les chapitres précédents.

La section 2 unifie et complète les éléments de géométrie de contact et de géométrie symplectique utilisés dans nos modélisations. Nous reprenons la notion de variété de contact et nous redéfinissons sa symplectisée. Les 1-formes de contact et les 2-formes symplectiques possèdent des formes normales de Darboux et nous revenons sur la “trinité” intriquée des trois géométries respectivement symplectiques, riemanniennes et complexes (au sens de définie sur \mathbb{C}).

Nous revenons ensuite de façon détaillée sur la notion de champ de contact. Il s’agit des champs de vecteurs sur M dont le flot préserve la distribution de contact \mathcal{K} (mais pas forcément la 1-forme de contact ω). Nous commençons par l’exemple de \mathbb{V}_J qui, bien qu’élémentaire, est déjà assez sophistiqué. Le champ de Reeb χ est un champ de contact mais il y en a beaucoup d’autres. En effet χ préserve non seulement \mathcal{K} mais aussi ω .

Évidemment, la meilleure façon pour un champ de laisser \mathcal{K} invariante est d’être nul! Là où ils ne sont pas nuls, les champs de contact X sont presque partout transverses à \mathcal{K} . Mais ils peuvent présenter des singularités en des points spéciaux x ou $X_x \in K_x$. Ils sont caractérisés par leur fonction génératrice $f(x) = \omega(X_x)$. Doubrov et Komrakov ont étudié les X qui relèvent des champs de vecteurs de l’espace de base dans la structure de contact.

En dimension 3 les champs de contact permettent de définir la notion de “surface convexe” (relativement à la structure de contact) qui sont les surfaces admettant un champ de contact partout transverse.

Une section particulièrement importante (section 3) est alors consacrée à la géométrie des espaces de jets et aux distributions de Goursat qui les généralisent. On a deux suites croissantes (i) le drapeau de Lie $\mathfrak{D}^0 = \mathfrak{D}$, $\mathfrak{D}^{j+1} = \mathfrak{D}^j + [\mathfrak{D}, \mathfrak{D}^j]$ définissant le “petit” vecteur de croissance, et (ii) le drapeau dérivé $\mathfrak{D}^{(0)} = \mathfrak{D}$, $\mathfrak{D}^{(j+1)} = \mathfrak{D}^{(j)} + [\mathfrak{D}^{(j)}, \mathfrak{D}^{(j)}]$ définissant le “grand” vecteur de croissance. Pour les jets, les deux vecteurs de croissance sont égaux. Les distributions de Goursat sont celles dont le “grand” vecteur de croissance est une suite d’entiers successifs. Elles prolongent le cas de \mathbb{V}_J (Heisenberg polarisé, 1-jets) et d’Engel (2-jets) et font prendre conscience de la complexité géométrique encodée dans la notion traditionnelle de série de Taylor. Elles approfondissent le principe de géométrisation du calcul différentiel sur lequel reposent nos modèles, à savoir que pour pouvoir traiter des

propriétés locales avec des processeurs purement ponctuels il faut introduire des dimensions supplémentaires et ajouter le système de Pfaff imposant comme contrainte à ces nouveaux symboles d'être des "dérivées cachées". Une fois introduite la notion géniale d'élément de contact par Lie et Cartan, les approximations successives par des jets génèrent une géométrie d'une complexité inouïe qui fournit un cadre universel et une voie royale pour la résolution des équations différentielles.

Un théorème fondamental de Cartan dit que les espaces de jets sont les formes normales locales génériques des distributions de Goursat de 2-plans. Mais il existe des distributions de Goursat qui ne se réduisent pas localement à celle des espaces de jets successifs. Qui plus est, dès que la dimension de M est ≥ 5 , les distributions de Goursat ne sont plus génériques, pour de simples raisons dimensionnelles, dans l'espace des distributions de type (k, n) et leur sous-espace devient de codimension croissante, et même infinie à la limite $n \rightarrow \infty$.

En 1978 Giaro, Kumpera et Ruiz ont commencé à étudier en dimension ≥ 5 les distributions de Goursat singulières qui ne sont pas de type jet et n'ont donc pas de forme normale de type Cartan. Il existe une "récurrence" de ces formes normales. Dès la dimension 6 il existe aussi des continua de formes normales (des "modules").

On s'est alors mis à étudier ces phénomènes et Richard Montgomery et Michail Zhitomirskii ont résolu le "hard problem" de leur classification avec les outils de la théorie des stratifications, des théorèmes de transversalité et des feuilletages caractéristiques. Les types de singularités augmentent exponentiellement avec la dimension n et devient infinie pour $n \geq 10$. Nous résumons leurs travaux sur cette "tour" proliférante de formes normales singulières de distributions de Goursat et ce qu'ils appellent le "monster manifold" \mathfrak{M} qui est une tour de hauteur infinie de variétés compactifiées (projectives) $\mathbb{P}_{\mathfrak{M}}^{s=n-2}$ de dimension n compactifiant projectivement les espaces de jets. Ces variétés $\mathbb{P}_{\mathfrak{M}}^{s=n-2}$ sont munies de distributions de Goursat universelles Δ^s qui sont classifiantes au sens où toute distribution de Goursat de dimension n est localement isomorphe à une forme normale représentée par un point de Δ^s . Dans ce "monstre classifiant" il existe des singularités de plus en plus compliquées "verticales" et "tangentielles" analysées en détail par Montgomery et Zhitomirskii. Nous en résumons la géométrie particulièrement riche.

La section 3 s'intéressait à la géométrie locale des espaces de jets et des distributions de Goursat. Mais il est évidemment essentiel de passer du local au global et d'essayer de mieux maîtriser les propriétés globales des structures de contact. Le problème est monumental. Il a commencé à être abordé techniquement dans les années 1970 en dimensions trois et a produit des résultats spectaculaires. La section 4 essaye d'en donner un petit aperçu en compilant quelques travaux particulièrement déterminants. Bien qu'elle n'ait pas d'applications concrètes en neurogéométrie, cette incursion dans de hautes sphères théoriques nous semble quand

même utile pour donner une idée au lecteur de l'univers dans lequel habitent nos modèles.

L'existence d'une structure de contact globale $\mathcal{K} = \ker(\omega)$ sur une variété différentiable M contraint sa topologie. Elle induit en effet une trivialisaton du fibré tangent $TM \simeq \mathcal{K} \oplus \chi$ où χ est le champ de Reeb. ainsi qu'un champ de structures symplectiques sur les hyperplans de contact. Il existe des théorèmes généraux sur les équivalences de structures de contact par homotopie et isotopie sur des variétés ouvertes ou fermées.

Tout un ensemble d'outils puissants permettent d'analyser la géométrie globale des variétés de contact (M^n, ω) . Par exemple les trajectoires fermées du champ de Reeb. Elles sont les points critiques de la "fonctionnelle d'action" $\mathfrak{A}(\gamma) = \int_\gamma \omega$ intégrale de la forme de contact ω le long des courbes fermées γ de M . Sur cette base s'est développée "l'homologie de contact" introduite par Eliashberg.

Un autre outil puissant est celui des courbes pseudo-holomorphes de Gromov. Pour les construire, on symplectise la structure de contact en ajoutant au champ de Reeb χ une variable conjuguée d'échelle et on compose ces plans symplectiques (dimension 2) avec les structures symplectiques induites par ω sur les hyperplans de contact (dimension $(n - 1)$ paire puisque n est impair). Cette structure symplectique de dimension $(n + 1)$ définit une structure presque complexe J sur $M \times \mathbb{R}$ et on peut alors considérer des applications holomorphes de surfaces de Riemann (qui sont des courbes sur \mathbb{C}) dans $(M \times \mathbb{R}, J)$. On peut ainsi utiliser toute la richesse de la théorie des surfaces de Riemann pour explorer la géométrie globale de (M, ω) .

Mais le cœur de cette section 4 est dédiée aux cas de dimension 3 où l'on peut utiliser la théorie des nœuds \mathbf{n} . La théorie de référence est celle de Bennequin (suivi par Eliashberg) qui a révolutionné ce domaine au début des années 1980. Les nœuds bien adaptés à la structure de contact sont soit les nœuds legendriens partout tangents à \mathcal{K} et relevant leur projection dans l'espace de base, soit, à l'inverse, les nœuds partout transverses à \mathcal{K} comme les trajectoires fermées du champ de Reeb.

Une autre particularité de la dimension 3 est que, si l'on considère une surface (régulière) S , comme ω est de contact et donc maximale non intégrable, il n'existe aucun petit domaine ouvert où S soit localement tangente à \mathcal{K} . En $x \in S \cap \mathcal{K}$, $T_x S$ et K_x sont génériquement transverses et se coupent le long d'une droite, ce qui définit le feuilletage caractéristique de S . Ce feuilletage a pour points singuliers les points exceptionnels où, purement ponctuellement, $T_x S = K_x$. Toute une théorie des feuilletages caractéristiques a été développée par Bourgeois, Etnyre, Floer, Giroux, Laudendach et d'autres spécialistes.

Nous présentons avec quelques détails le cas de la fibration de Hopf de \mathbb{S}^3 qui généralise nos deux modèles $\mathbb{V}_J = \mathbb{H}_{pol}$ et $\mathbb{V}_S = SE(2) = \mathbb{R}^2 \times \mathbb{S}^1$. \mathbb{S}^3 y est vue comme la sphère de \mathbb{C}^2 ce qui relie la fibration au corps des quaternions.

Le théorème de Bennequin dit qu’il existe sur \mathbb{S}^3 des structures de contact globales (d’ailleurs faciles à construire) non isotopes (mais quand même homotopes) à la structure standard \mathcal{K}_0 . Cela est dû au fait qu’elles contiennent des “disques vrillés” (“overtwisted”) alors qu’il n’existe pas de disque vrillé dans la structure standard, tous ses disques plongés étant “tendus” (“tight”). Un disque vrillé est un disque plongé Δ (le plongement pouvant être extrêmement compliqué et Δ pouvant être très grand) partout transverse à \mathcal{K} en son intérieur sauf en son centre mais dont tout le bord $\partial\Delta$ est tangent à \mathcal{K} .

Le théorème repose sur un usage sophistiqué de la théorie des nœuds \mathbf{n} et de leurs surfaces de Seifert S (qui sont des surfaces plongées, orientées, compactes, connexes, sans auto-intersection et de bord $\partial S = \mathbf{n}$). De telles surfaces permettent de définir des repères mobiles le long de leur bord \mathbf{n} et on peut analyser la façon dont ceux-ci tournent par rapport à \mathcal{K} le long de \mathbf{n} . Cet enroulement est codé par des nombres que Bennequin calcule pour les nœuds legendriens \mathbf{n}_ℓ et les nœuds \mathbf{n}_τ transverses à \mathcal{K} . Il les relie à la caractéristique d’Euler-Poincaré maximale des surfaces de Seifert, ce qui lui permet de démontrer son théorème.

Nous concluons cette section en évoquant la façon dont Giroux a retrouvé les résultats de Bennequin et d’Eliashberg en utilisant la théorie des surfaces convexes relativement à \mathcal{K} , c’est-à-dire admettant un champ de contact totalement transverse \mathcal{C} . Une telle possibilité fait interférer le champ \mathcal{C} avec avec le feuilletage caractéristiques \mathcal{K}_S induit par \mathcal{K} sur S . Nous résumons la façon dont Giroux utilise la géométrie sophistiquée de la K -convexité afin de définir un critère pour qu’une structure de contact soit tendue et comment, au moyen de méthodes de “chirurgie”, il retrouve les résultats de Bennequin-Eliashberg.

9.16. Chapitre 16

Les chapitres précédents plongent la neurogéométrie de la connectivité des aires visuelles primaires dans la théorie des groupes de Lie, l’univers des structures de contact et la géométrie sous-riemannienne. Mais cette connectivité connecte des neurones qui ont des profils récepteurs et effectuent par convolution une analyse du signal rétinien. La neurogéométrie doit donc articuler sa géométrie des connectivités avec des éléments d’analyse fonctionnelle.

C’est ainsi un nouvel univers que nous devons maintenant explorer et pour cela nous avons besoin d’un certain nombre d’outils classiques d’analyse fonctionnelle afin d’arriver aux notions fondamentales d’état cohérent, d’analyse harmonique non commutative et de diffusion sous-riemannienne qui concluent l’essentiel de notre périple. Cette nouvelle exploration se fait en deux chapitres. Dans le premier (16), nous présentons un petit *vademecum* d’analyse fonctionnelle. Dans le second (17), nous traitons les applications spécifiques à l’analyse du signal optique.

Dans le Vol I nous avons longuement exposé les profils récepteurs des neurones visuels et leur action sur le signal rétinien envoyé par les photorécepteurs le long de la voie rétino-géniculo-corticale. Ils agissent par convolution comme des filtres et effectuent une sorte d'analyse en ondelettes relevant de l'analyse harmonique non commutative. Nous renvoyons donc le lecteur au Vol I pour une reprise en main de toute la richesse de ces données expérimentales.

Nous nous aventurons dans un immense domaine et en particulier dans celui des représentations de nos groupes de Lie \mathbb{V}_J (Heisenberg polarisé) et \mathbb{V}_S ($SE(2)$) dans des espaces de Hilbert appropriés. Mais nous faisons une remarque préalable. Le passage d'une géométrie de l'intégrabilité à une analyse fonctionnelle pilotée par elle est récurrente en physique et nous en avons déjà rencontré un exemple de base à propos des caustiques, celui du passage de l'optique géométrique à l'optique ondulatoire imposé empiriquement par la nature de la lumière comme champ (section 6 du chapitre 12). Nous y avons alors vu comment l'optique géométrique avec toutes ses propriétés géométriques (équations de Hamilton, calcul variationnel, principe de moindre action, etc.) fonctionnait comme le "squelette géométrique" des propriétés fonctionnelles des solutions de l'équation des ondes (les solutions lagrangiennes).

Nous sommes ici confrontés à un problème de même type imposé par la neurophysiologie, à savoir que la géométrie de la connectivité neuronale pilote une analyse fonctionnelle du signal, les images-percepts étant des champs d'activité neuronale de profils récepteurs (des bancs de filtres). Sur le plan formel, dans la mesure où la neurogéométrie utilise des formalismes courants en mécanique classique (groupes de Lie, principes variationnels, équations de Hamilton, géodésiques, etc.), la prise en compte de sa "chair fonctionnelle" ressemble beaucoup au passage de l'optique géométrique à l'optique ondulatoire et de la mécanique classique à la mécanique ondulatoire, c'est-à-dire à la mécanique quantique. Et effectivement, au niveau des outils formels, nous rencontrons beaucoup de formalismes (représentations de groupes de Lie dans des espaces Hilbert, analyse spectrale d'opérateurs dans ces espaces, etc.) développés par des mathématiciens de génie comme Hermann Weyl ou John von Neumann pour formaliser la mécanique quantique après sa découverte par des physiciens de génie comme Heisenberg, Schrödinger, Dirac et plusieurs d'autres.

Fidèle à nos options pédagogiques et méthodologiques nous tenons compte de l'origine historique des formalismes utilisés et le lecteur ne devra pas s'étonner de références récurrentes aux bases de la première mécanique quantique.

Le chapitre 16 est donc propédeutique. C'est un mini-cours d'analyse fonctionnelle partant de la théorie de la mesure et de l'intégration pour définir d'abord les espaces fonctionnels L^p , et en particulier les espaces de Hilbert L^2 , pour en arriver aux opérateurs avec leur décomposition spectrale et même à leur généralisation par les opérateurs pseudo-différentiels. Il commence par une brève section (1) précisant le sens du passage de la géométrie à l'analyse.

La section 2 revient aux bases de la théorie de la mesure et de l'intégration. Un tel retour amont est nécessaire pour pouvoir définir les espaces fonctionnels en jeu. Cela est fait dans la section 3 qui définit les espaces $L^p(X, \nu)$ sur un espace (X, ν) muni d'une mesure ν . Ce sont les espaces des fonctions mesurables sur X

- (i) définies à l'équivalence "être égales ν -presque partout" près,
- (ii) dont la puissance p -ième est de "taille finie" au sens où l'intégrale

$$\int_X |f(x)|^p d\nu(x)$$

est bornée.

Les $L^p(X, \nu)$ sont des espaces métriques pour la norme

$$\|f\|_p = \left(\int_X |f(x)|^p d\nu(x) \right)^{\frac{1}{p}}$$

(l'inégalité triangulaire étant dite inégalité de Minkowski) et complets d'après un théorème de Riesz-Fischer. Ils sont donc des espaces de Banach.

Il existe des liens essentiels de dualité entre L^p et L^q lorsque $\frac{1}{p} + \frac{1}{q} = 1$, $p, q \in]1, \infty[$ étant alors dits "conjugués". La dualité se manifeste entre autres par la célèbre inégalité de Hölder $\|fh\|_1 \leq \|f\|_p \|h\|_q$ qui permet de démontrer l'inégalité de Minkowski. En fait si q est conjugué de p et si l'on associe à $h \in L^q$, la forme linéaire sur L^p $u_h : f \in L^p \rightarrow u_h(f) = \int_X f(x) \overline{h(x)} d\nu(x)$, u est une isométrie de L^q dans le dual topologique de L^p (elle est donc injective mais pas forcément surjective). Les espaces L^p sont réflexifs, c'est-à-dire isomorphes à leur bidual.

Mais si $p \neq 2$ les espaces L^p sont des espaces normés dont la norme ne découle pas d'un produit scalaire. On peut y mesurer les "longueurs" (les normes) mais pas les "angles" $\langle f, g \rangle$ entre deux fonctions f et g de norme 1. D'où un fort supplément de structure lorsque $p = q = 2$ car alors L^2 (l'espace des fonctions de carré intégrable) est isomorphe à son dual (théorème de représentation de Riesz) et sa norme dérive d'un produit scalaire. Il s'agit des fameux espaces de Hilbert qui, du moins lorsqu'on se restreint aux sous-espaces fermés et aux applications continues, généralisent à l'analyse fonctionnelle en dimension infinie les espaces euclidiens de dimension finie (les sous-espaces fermés sont les noyaux des applications linéaires continues et sont définis par leur projecteur orthogonal). La section 4 récapitule leurs propriétés. L'inégalité de Hölder devient l'inégalité dite de Cauchy-Schwarz reliant le produit scalaire aux normes.

Les espaces de Hilbert les plus communément utilisés sont ceux dits "séparables", c'est-à-dire admettant des bases hilbertiennes $\{e_k\}_{k \in \mathbb{N}}$ dénombrables. Avec eux, on se retrouve avec une géométrie (dite "hermitienne" et définie sur le corps de base \mathbb{C}) vraiment analogue à la géométrie euclidienne. Une différence essentielle avec le cas de dimension finie est toutefois que la boule unité n'est plus compacte (théorème

de compacité de Riesz), la non-compacité n'étant pas due au fait que des longueurs deviennent infinies mais précisément à la dimension infinie.

La section 5 aborde la remarquable généralisation de la notion de fonction due à Laurent Schwartz avec les distributions. Lorsque T et f sont des fonctions dont le produit est intégrable $\langle T, f \rangle = \int T(x)f(x)dx$ est une application linéaire continue sur l'espace des fonctions f . Les distributions sont les applications linéaires continues $\langle T, f \rangle$ plus générales. Elles sont indispensables car, dans de très nombreuses situations, l'on doit considérer des "fonctions" qui sont des fonctions "généralisées", comme la "fonction" de Dirac δ satisfaisant $\int \delta(x)f(x)dx = f(0)$, et sont en fait des distributions. C'est en particulier le cas pour les signaux qui, en particulier à cause du bruit sont des "fonctions" assez sauvages.

Nous avons montré dans le Vol I (chapitre III, section 3.3.4.) que, comme y ont insisté Jan Koenderink et Luc Florack, la mesure des signaux optiques $I(x, y)$ par des champs de profils récepteurs de neurones visuels $\varphi_{(x_0, y_0)}(x, y)$ de champ récepteur D centré en (x_0, y_0) est analogue à celle que l'on trouve dans la théorie des distributions : $\langle I, \varphi_{(x_0, y_0)} \rangle = \int I(x, y)\varphi_{(x_0, y_0)}(x, y)dx$. Les profils récepteurs φ sont des fonctions test câblées dans le hardware neuronal, les signaux sont des distributions I externes. Mais ici, contrairement à la situation classique, ce sont les φ qui sont actifs (qui mesurent) et les I qui sont passifs (qui sont mesurés), alors que dans la théorie des distributions c'est l'inverse.

Plus les fonctions test sont contraintes (par exemple C^∞ compactes) plus les distributions peuvent être sauvages ; moins les fonctions test sont contraintes (par exemple seulement C^∞ à décroissance rapide à l'infini) plus les distributions sont "tempérées". Nous précisons les différents types de distributions et définissons leur produit de convolution qui est leur produit naturel.

Une fois esquissé le concept de distribution au sens de Schwartz, nous revenons à l'analyse harmonique. Pour rendre plus concret cet ensemble de notions abstraites, nous détaillons section 6 l'exemple des séries de Fourier, exemple prototypique (et propédeutique) d'analyse harmonique. Son extrême richesse révèle déjà les arcanes de la théorie générale et tous nos autres cas d'analyse harmonique en seront des complexifications progressives.

Les séries de Fourier correspondent à l'analyse harmonique des fonctions périodiques sur le cercle \mathbb{S}^1 de variable angulaire θ et les développe sur une base hilbertienne du Hilbert $L^2(\mathbb{S}^1)$ muni de la mesure $\frac{d\theta}{2\pi}$. Elles ont conduit à l'analyse d'un nombre considérable de problèmes nouveaux, profonds et difficiles concernant le lien entre une fonction périodique f et sa série de Fourier $T(f)$ qui est une série trigonométrique $T(f) = \sum_{k \in \mathbb{Z}} c_k e^{ik\theta}$ la développant dans le système orthonormal des fonctions trigonométriques $e_k = e^{ik\theta}$. Les coefficients c_k définissent la transformée de Fourier $\hat{f}(k)$ comme fonction sur \mathbb{Z} et sont donnés par $c_k = \int_{\mathbb{S}^1} f(\theta) e^{-ik\theta} \frac{d\theta}{2\pi} = \hat{f}(k) = \langle f, e_k \rangle$. La transformation de Fourier $T : f(\theta) \rightarrow \hat{f}(k)$ va de $L^1(\mathbb{S}^1)$

(l'intégrabilité de f est nécessaire pour que les c_k soient bien définis) dans $\mathbb{C}^{\mathbb{Z}}$ et son étude précise pour des fonctions compliquées s'est heurtée à des problèmes inédits qui ont été un formidable moteur d'innovation. Les fonctions quelconques peuvent être très sauvages. Il suffit de penser à la fonction de Dirichlet = 0 sur les rationnels \mathbb{Q} et = 1 sur les irrationnels $\mathbb{R} - \mathbb{Q}$.

La section 6 évoque les théorèmes de base comme le théorème de Riemann-Lebesgue disant que si f est intégrable ses coefficients de Fourier $c_k = \widehat{f}(k)$ tendent vers 0 pour $k \rightarrow \pm\infty$. Cela n'implique pas que $T(f)$ converge et que, même si elle converge, elle converge vers f . D'ailleurs il y a l'exemple célèbre dû à Fatou d'une série convergente de $\mathbb{C}^{\mathbb{Z}}$ qui ne peut être la série de Fourier d'aucune fonction intégrable. Plus la décroissance de $|\widehat{f}(k)|$ pour $k \rightarrow \pm\infty$ est rapide et plus f est régulière.

Il existe de nombreux exemples de fonctions "pathologiques" dont la série de Fourier ne converge pas, même sur un sous-ensemble dense de \mathbb{S}^1 . Le premier, qui fit le tour de l'Europe, est dû à Paul du Bois-Reymond en 1873 et nous l'illustrons en détail. Kolmogorov a même trouvé un exemple d'une f intégrable dont la série de Fourier $T(f)$ diverge partout. Lebesgue trouva des fonctions continues "monstrueuses" dérivables nulle part. Elles étaient les premières courbes fractales. Nous les illustrons également en détail.

Les propriétés fondamentales des espaces de Hilbert L^2 furent découvertes à propos des séries trigonométriques : l'égalité de Parseval-Plancherel, l'isométrie établie par T entre les espaces de Hilbert $L^2(\mathbb{S}^1)$ et $\ell^2(\mathbb{C})$ (les suites $\{c_k\}_{k \in \mathbb{Z}}$ telles que $\sum_{k \in \mathbb{Z}} |c_k|^2$ converge). La convergence L^2 des séries de Fourier $\widehat{f} = T(f)$ des f de $L^2(\mathbb{S}^1)$ n'implique pas forcément la convergence ponctuelle. La conjecture de Lusin (1913) disant qu'elle implique néanmoins la convergence ponctuelle presque partout ne fut démontrée qu'en 1966 par Carleson. Mais si f est continue et C^1 par morceaux, $T(f)$ converge uniformément vers f (Dirichlet 1847). Si f a une "bonne discontinuité" en θ_0 , $T(f)$ converge vers $\frac{1}{2}(f(\theta_0)^- + f(\theta_0)^+)$. Mais la convergence est faible et manifeste le phénomène bien connu dit de Gibbs que nous illustrons sur un exemple.

La démonstration de nombre de ces théorèmes conduisit au XIX^e siècle à l'introduction des méthodes de noyaux de convolution de Dirichlet. L'idée est de considérer les sommes partielles $S_n f$ de $T(f)$ et de les exprimer comme des convolutions $S_n f = D_n * f$, $D_n = \sum_{k=-n}^{k=n} e^{ik\theta}$. À la limite on pourra alors, du moins dans les bons cas, définir $T(f)$ comme la convolution $D_\infty * f$ et les difficultés de sommation et de convergence de f vers $T(f)$ pourront se lire sur la structure de D_∞ . Nous donnons les exemples numériques de plusieurs D_n . Toutefois, ces noyaux de convolution ne sont pas "bons" et c'est pourquoi, comme le note Joël Merker, "les théorèmes de convergence pour les séries de Fourier sont difficiles".

Nous exposons ensuite scolairement les travaux de Féjer (tout début du XX^e siècle) permettant d'améliorer notablement les résultats de convergence en introduisant de "bons" noyaux appelés "résolutions de l'identité". Nous en illustrons quelques exemples. Ils ont permis de comprendre l'exemple de Fatou, la démonstration étant assez subtile.

Après la section assez calculatoire (très XIX^e siècle) sur les séries de Fourier, la section 7 aborde la théorie abstraite et structurale des opérateurs sur les espaces de Hilbert et de leur décomposition spectrale. Il s'agit d'une théorie fondamentale élaborée vers la fin des années 1920 par Banach, Steinhaus, Schauder, Hahn, et qui s'est mise à dominer toute la physique après l'avènement de la mécanique quantique, surtout grâce aux travaux de von Neumann. Il s'agit des opérateurs linéaires continus $T : \mathcal{F} \rightarrow \mathcal{G}$ entre espaces de Banach (et en particulier les Banach L^p et les Hilbert L^2) qui peuvent ne pas être partout définis. Les plus simples sont les opérateurs compacts, puis viennent les opérateurs partout définis et bornés dont on peut définir la norme et qui forment des espaces $\mathcal{L}(\mathcal{F}, \mathcal{G})$ qui sont eux mêmes des Banach, puis les opérateurs non bornés et non partout définis. Nous énonçons les grands théorèmes de base :

1. le théorème de Banach-Schauder : les opérateurs $T \in \mathcal{L}(\mathcal{F}, \mathcal{G})$ sont ouverts et l'image par T de la boule unité de \mathcal{F} est encadrée par deux boules de \mathcal{G} ;
2. le théorème de Banach-Baire : si $T \in \mathcal{L}(\mathcal{F}, \mathcal{G})$ est bijectif alors il est inversible et $T^{-1} \in \mathcal{L}(\mathcal{F}, \mathcal{G})$;
3. le théorème de Banach-Steinhaus : une suite de T_n simplement bornée sur un ensemble dense intersection dénombrable d'ouverts denses est uniformément bornée sur la boule unité de \mathcal{F} .

La plupart du temps, les opérateurs considérés sont des opérateurs différentiels T opérant sur des applications possédant un certain degré de différentiabilité et satisfaisant des contraintes et des conditions aux limites. S'ils ne sont pas définis partout, leur domaine de définition $\mathcal{D}(T)$ fait intrinsèquement partie de leur définition et la spécificité de $\mathcal{D}(T)$ intervient de façon cruciale dans leurs propriétés. Parmi ces opérateurs non bornés les "bons" sont les opérateurs "fermés" dont le domaine $\mathcal{D}(T)$ contient les limites des suites convergentes (au sens d'une notion naturelle de convergence des éléments de $\mathcal{D}(T)$).

La théorie peut se préciser lorsque l'on considère des opérateurs $T \in \mathcal{L}(\mathcal{H})$ sur des Hilbert dont le domaine $\mathcal{D}(T)$ est dense. En effet, on peut alors définir leur adjoint T^* par la formule simple de dualité $\langle T(f), h \rangle = \langle f, T^*(h) \rangle$. Cela conduit à une classe particulièrement importante d'opérateurs sur les Hilbert, celle des opérateurs dits symétriques qui satisfont $\langle T(f), h \rangle = \langle f, T(h) \rangle$ sur $\mathcal{D}(T)$, autrement dit qui sont auto-adjoints sur $\mathcal{D}(T)$. Mais $\mathcal{D}(T^*)$ peut être plus grand que $\mathcal{D}(T)$ et donc T peut-être symétrique sans être véritablement auto-adjoint et posséder plusieurs

extensions auto-adjointes différentes. Lorsqu'en plus $\mathcal{D}(T^*) = \mathcal{D}(T)$, T est vraiment auto-adjoint. Entre les opérateurs symétriques et les opérateurs auto-adjoints il y a les opérateurs essentiellement auto-adjoints qui sont les opérateurs symétriques n'admettant qu'une unique extension auto-adjointe. Nous illustrons ces subtilités par des exemples simples et traditionnels.

Une fois explicitées ces définitions, nous présentons le cœur de la théorie des opérateurs auto-adjoints, à savoir leur décomposition spectrale. Il s'agit de généraliser les notions de valeur propre et de vecteur propre de la dimension finie à la dimension infinie. Mais l'infinitude de la dimension complique énormément la situation en introduisant de nombreux nouveaux phénomènes en particulier lorsqu'on utilise des opérateurs différentiels.

Les opérateurs auto-adjoints sont très vite devenus omniprésents en mécanique quantique où l'on modélise les états d'un système par les vecteurs d'un espace de Hilbert. En effet leurs valeurs propres sont réelles et ils modélisent les mesures qui, d'après le postulat de la "réduction du paquet d'onde", projettent le vecteur du système sur un de leurs vecteurs propres et donnent la valeur propre comme résultat de la mesure.

Il y a beaucoup de subtilité dans la notion générale de décomposition spectrale des opérateurs auto-adjoints A . En effet, très souvent, des vecteurs propres vraisemblables ne sont pas dans le Hilbert du système (ce ne sont pas des fonctions mais des distributions, ou alors ce sont bien des fonctions mais elles ne sont pas de carré intégrable). C'est évident dans le cas d'un système de dimension 1 décrit par une fonction $f(x)$ pour l'opérateur de position $Q(f) = xf(x)$ et pour l'opérateur moment $P(f) = -i\frac{\partial f(x)}{\partial x}$. La distribution de Dirac δ_{x_0} en x_0 est un vecteur propre évident de Q pour la valeur propre x_0 et la fonction de phase e^{ip_0x} est un vecteur propre évident de P pour la valeur propre p_0 . Mais ni δ_{x_0} ni e^{ip_0x} n'appartiennent au Hilbert \mathcal{H} .

Pour résoudre cet épineux problème il faut distinguer les valeurs spectrales λ de A auto-adjoint qui sont telles que, par définition, $(A - \lambda Id)$ n'est pas inversible (elles constituent son "spectre" $\text{Spec}(A)$) et les véritables valeurs propres telles que $Af = \lambda f$ pour une $f \neq 0$. Comme alors $(A - \lambda Id)f = 0$, $(A - \lambda Id)$ n'est pas inversible et λ appartient au spectre de A . Mais en dimension infinie un opérateur comme $(A - \lambda I)$ peut être injectif (i.e. il n'existe pas de $f \neq 0$ telle que $(A - \lambda I)f = 0$ et donc λ n'est pas valeur propre) sans être pour autant inversible (et dans ce cas λ est une valeur spectrale). C'est le cas de $Qf(x) = xf(x)$ qui est bien injectif sur \mathcal{H} (car si $Q(f) = 0$, alors $f(x) = 0$ si $x \neq 0$ et donc $f \equiv 0$ à l'équivalence "être égales presque partout" près) et n'a pas de valeurs propres relativement à \mathcal{H} tandis que, en revanche, tout $\lambda \in \mathbb{R}$ appartient à son spectre.

La stratégie développée par Stone puis von Neumann consiste alors à ne pas introduire des valeurs propres bien définies avec des vecteurs propres "généralisés",

mais au contraire à n'admettre que des vecteurs bien définis de \mathcal{H} “presque” propres avec des valeurs propres seulement “approximatives”. Un peu comme on remplace la distribution de Dirac δ_x par des gaussiennes très piquées centrées sur x . La décomposition spectrale de A devient par suite une intégrale $\int_{\text{Spec}(A)} \lambda d\sigma_A(\lambda)$ où $d\sigma_A(\lambda)$ est une mesure sur les boréliens de \mathbb{R} de support dans $\text{Spec}(A)$ à valeurs dans l'espace des projecteurs de \mathcal{H} (“projection-valued measure”, *PVM*). Nous exposons avec un certain détail le théorème de Stone, d'abord dans sa version “simple” puis dans sa version “générale”. Sa preuve est un fait d'armes de l'analyse fonctionnelle.

La section suivante (8) consacrée à la quantification géométrique est un détour par la mécanique quantique. Il nous semble justifié car, comme nous l'avons vu, les formalismes couplant la neurogéométrie des connectivités avec l'analyse fonctionnelle du signal optique sont très voisins des formalismes utilisés par la mécanique quantique même si les phénomènes considérés n'ont rien à voir entre eux.

Un problème formel fondamental de la mécanique quantique depuis Dirac est de passer systématiquement de la géométrie hamiltonienne des systèmes classiques à leur version quantique. Pour cela il faut pouvoir associer aux grandeurs physiques classiques définies sur un espace des phases classique T^*M (fibré cotangent) des opérateurs auto-adjoints \widehat{f} sur le Hilbert du système quantique quantifiant le système classique. Il s'agit d'une vaste généralisation du passage de l'optique géométrique à l'optique ondulatoire que nous avons déjà traité dans la section 6 du chapitre 12. Il faut pouvoir satisfaire les conditions de Dirac dont la principale est l'égalité $[\widehat{f}, \widehat{g}] = -i\hbar\widehat{\{f, g\}}$ reliant le commutateur (crochet de Lie) des opérateurs au crochet de Poisson des fonctions classiques. La comparaison des cas classiques et quantiques, dont nous rappelons les bases, montre que le problème est hautement non trivial. Les solutions simples du genre $\widehat{f} = -i\hbar X_f$, où X_f est le champ hamiltonien sur T^*M associé à la fonction f , champ considéré comme opérant sur $L^2(T^*M, \mathbb{C})$, sont complètement inadéquates, même si $X_{\{f, g\}} = [X_f, X_g]$.

Les efforts conjugués d'éminents mathématiciens-physiciens comme Weyl, Wigner, Moyal, puis à partir des années 1960 Souriau, Kostant, Arnold, Kirillov, Weinstein et beaucoup d'autres, ont permis de surmonter nombre des difficultés majeures. Ils ont d'abord montré qu'une “préquantification” $\widetilde{f} = -i\hbar X_f - pdq(X_f) + f$ (où pdq est là 1-forme canonique, dite ici “potentiel symplectique”, engendrant la 2-forme canonique $dp \wedge dq$ de T^*M) permet de commencer à satisfaire assez correctement les conditions de Dirac puisque l'harmonie des calculs montre que $[\widetilde{f}, \widetilde{g}] = -i\hbar\widetilde{\{f, g\}}$.

Mais une telle préquantification n'est pas suffisante et, pour aller plus loin, il faut tenir compte du fait que les fonctions d'onde quantiques sont à valeurs dans \mathbb{C} . On introduit donc du côté classique de T^*M un fibré en droites complexe $\delta : L \rightarrow T^*M$ permettant de définir des fonctions d'état classiques à valeurs dans \mathbb{C} en considérant

des sections $s(q, p)$ de L . L'idée de base est alors d'introduire sur L la connexion ∇ définie par $\nabla_\xi s = ds(\xi) - \frac{i}{\hbar} \theta(\xi) s$ où ξ est un champ de vecteurs sur L , s une section de L et θ le potentiel symplectique pdq trivialement étendu à L . ∇ dérive de la 1-forme de connexion à valeurs complexes $\alpha = \frac{1}{i} \frac{dz}{z} - \frac{i}{\hbar} \theta$ sur $L - \{section\ 0\}$ de fibre \mathbb{C}^* au-dessus de T^*M , où z est une coordonnée complexe dans les fibres de L . Si on la restreint au sous-fibré en cercles des $z = e^{i\varphi}$ on a $\alpha = d\varphi - \frac{i}{\hbar} \theta$ et la construction est analogue à ce qui se passe quand on "contactise" T^*M en introduisant sur $T^*M \times \mathbb{R}$ la 1-forme $dy - pdx$ que nous connaissons bien. La 1-forme de connexion α définit (comme dans nos modèles d'espaces de jets) des hyperplans "horizontaux" dans le fibré $T^*M \times \mathbb{C}^*$. On peut alors préciser la préquantification en prenant $\tilde{f}(s) = -i\hbar \{f, s\} - \theta(X_f) s + f(s)$ comme transformation de f en opérateur \tilde{f} opérant sur les sections s du fibré L .

La dernière étape de la quantification géométrique consiste à revenir au fait que les fonctions f sont définies sur T^*M alors que les fonctions d'ondes quantiques sont des fonctions $\psi(q)$ ou $\psi(p)$ définies seulement sur la moitié des variables canoniques. Pour surmonter cet obstacle, les spécialistes ont introduit la notion de polarisation qui revient de la dimension $2n$ de T^*M à la dimension n de façon compatible avec la préquantification. L'idée est de passer des espaces L^2 de sections générales $s(q, p)$ à des espaces L^2 associés aux ensembles maximaux d'observables commutantes (\tilde{f}_i) . Ces ensembles correspondent aux ensembles complets de fonctions classiques indépendantes en involution (i.e. $\{f_i, f_j\} = 0$) et ces derniers définissent des distributions lagrangiennes de dimension n dans T^*M (qui est de dimension $2n$).

La polarisation la plus simple, dite "verticale" et qui conduit à des opérateurs \hat{f} sur le L^2 des fonctions d'ondes $\psi(q)$ de la représentation de Schrödinger, consiste à revenir à la fibration initiale $\pi : T^*M \rightarrow M$. La fibration en droites complexes $\delta : L \rightarrow T^*M$ devient dès lors doublement fibrée par $\delta^\pi = \pi \circ \delta : L \rightarrow T^*M \rightarrow M$. Le cas de base est celui de $M = \mathbb{R}^n$. On identifie les points p_i d'une même fibre T_q^*M par transport parallèle au moyen de la connexion ∇ et l'on obtient des sections s^π de δ^π qui sont ∇ -constantes le long des fibres de π . On peut alors réduire le formalisme de la préquantification à la fibration δ^π en considérant les f telles que X_f laisse la fibration invariante. Elles sont les f linéaires en p , $f(q, p) = r(q) + p.t(q)$ et l'on obtient la quantification $\hat{f}(s^\pi) = -i\hbar \{f, s^\pi\} + r(q) s^\pi$. On généralise ensuite de \mathbb{R}^n à M sous certaines conditions cohomologiques liées à la topologie globale de M .

On peut de même, mais c'est un peu plus compliqué, considérer la polarisation "horizontale" de T^*M .

Une autre possibilité intéressante est celle des polarisations dites "kähleriennes" qui utilisent le fait que la structure symplectique canonique $\Theta(X, Y)$ de T^*M est naturellement associée à une métrique g si l'on prend une base symplectique comme

base orthonormée. On définit alors une structure presque complexe J par $\Theta(X, Y) = g(X, J(Y))$ (cf. la section 2 du chapitre 3). Si g est définie positive, on est dans le cas kählerien. On peut identifier les espaces tangents $T_{(q,p)}(T^*M)$ (qui sont de dimension $2n$) à \mathbb{C}^n au moyen d'un isomorphisme interprétable comme multiplication par i et l'on définit un produit scalaire hermitien $\langle X, Y \rangle$ sur $T(T^*M)$ par la formule $\langle X, Y \rangle = g(X, Y) + i\Theta(X, Y)$. On peut alors construire à partir de ces données une quantification dite "kählerienne".

Mais ces performances formelles ne suffisent encore pas car on ne peut évidemment pas se restreindre à la quantification des fonctions $f(q, p)$ linéaires en p , ne serait-ce que parce que les énergies cinétiques quadratiques $\frac{p^2}{2m}$ (m la masse) quantifiées par des laplaciens sont omniprésentes et ne laissent pas les polarisations invariantes. Pour surmonter cette difficulté, Blattner, Kostant et Sternberg ont développé une technique appelée depuis la méthode des noyaux *BKS*.

Les différentes polarisations sont transformables l'une en l'autre. On passe de la représentation "verticale" (positions q) à la représentation "horizontale" (moments p) par la transformée de Fourier. Le passage à la représentation kählerienne correspond quant à elle à la transformée de Segal-Bargmann-Fock.

Par ailleurs les polarisations soulèvent un problème majeur lié à la définition du produit scalaire de l'espace de Hilbert associé. On part en effet du fibré L muni de la connexion ∇ , puis on introduit une polarisation \mathfrak{P} qui est une fibration lagrangienne $\lambda : T^*M \rightarrow \Lambda$ et enfin l'on construit le Hilbert $\mathcal{H}_{\mathfrak{P}}$ à partir des sections \mathfrak{P} -polarisées de L (qui sont par définition constantes sur les fibres de λ). Mais pour définir le produit scalaire sur $\mathcal{H}_{\mathfrak{P}}$ il faut pouvoir définir correctement la mesure d'intégration et pour cela la mesure de Liouville μ associée à la forme symplectique Θ de T^*M n'est pas utilisable car les sections constantes sur les fibres de λ ne seront pas en général de carré intégrable relativement à μ . Pour résoudre cette difficulté il faut en quelque sorte "couper en deux" la mesure de Liouville μ . Cela conduit à la notion technique de "demi-densité" que nous introduisons mais sans entrer dans les détails techniques.

Les polarisations sont des feuilletages lagrangiens et pour qu'elles soient globales et permettent d'obtenir de bonnes quantification sur T^*M (dont $T^*\mathbb{R}^n$ est le modèle local) il faut que des conditions d'intégrabilité soient satisfaites. Une telle contrainte permet de retrouver sur des bases théoriques solides les premières conditions de quantification introduites par Bohr et Sommerfeld dès 1913 (avant le début de la mécanique quantique proprement dite dans les années 1920) pour expliquer les raies spectrales discrètes des atomes par des sauts d'énergie entre les orbitales électroniques autour des noyaux. Ces conditions d'intégrabilité furent ensuite précisées géométriquement en 1926 par Brillouin, Kramers et Wentzel (méthode *BKW*) qui se fondèrent sur le fait que la mécanique quantique est à la mécanique classique hamiltonienne ce que l'optique ondulatoire est à l'optique géométrique.

Cette section sur la quantification géométrique continue en introduisant alors un certain nombre d'exemples simples et éclairants qui sont en résonance avec des structures déjà rencontrées dans nos modélisations neurogéométriques : le fibré cotangent du cercle, les orbites de la représentation co-adjointe de $SE(2)$, la fibration lagrangienne du plan épointé, l'oscillateur harmonique.

Puis la section se conclut en montrant comment la quantification géométrique acquiert une représentation fonctionnelle, c'est-à-dire comment les opérateurs quantiques peuvent s'écrire sous forme d'intégrales. Il s'agit d'une vaste généralisation de ce que nous avons vu en optique ondulatoire dans la section 6 du chapitre 12 à propos des solutions asymptotiques de l'équation des ondes, à savoir que les solutions lagrangiennes du relèvement de l'équation dans le fibré cotangent de l'espace peuvent acquérir une représentation fonctionnelle au moyen d'intégrales oscillantes. On parle d'approximations "semi-classiques" des systèmes quantiques.

L'une des premières représentations fonctionnelles est celle de d'Hermann Weyl associant à une fonction $a(x, \xi)$ (notations traditionnelles dans cette thématique) sur un fibré cotangent T^*M l'intégrale oscillante

$$a^W(u(x)) = \frac{1}{(2\pi)^n} \int_{T^*M} e^{i(\xi-y)} a\left(\frac{x+y}{2}, \xi\right) u(y) dy d\xi$$

conçue comme opérateur sur les $u(x)$. Quand on introduit le facteur \hbar , les opérateurs a^W satisfont les règles de quantification et leur composition correspondent au produit de Moyal (non commutatif) des fonctions. On obtient au deuxième ordre $[a^W, b^W] = -i\hbar \{a, b\}^W$. D'où la possibilité de développements asymptotiques en \hbar donnant des approximations semi-classiques.

Nous revenons alors sur l'optique ondulatoire et les représentations fonctionnelles en termes d'intégrales oscillantes. Le lien entre les deux niveaux du "squelette" géométrique et de la "chair" fonctionnelle s'effectue à travers le principe de la phase stationnaire disant qu'une intégrale oscillante $I(x, \tau) = \int e^{i\tau\varphi(x, \alpha)} a_\tau(x, \alpha) d\alpha$ se concentre lorsque $\tau \rightarrow \infty$ ($\tau = \frac{1}{\hbar}$ et $\hbar \rightarrow 0$ en mécanique quantique) sur le lieu critique V_φ où la phase $\varphi_x(\alpha) = \varphi(x, \alpha)$ est stationnaire (i.e. $\frac{\partial\varphi_x(\alpha)}{\partial\alpha} = 0$) en tant que fonction de α paramétrée par x . Cela permet de construire des représentations fonctionnelles des singularités des projections des solutions lagrangiennes (c'est-à-dire des caustiques) en prenant pour $\varphi(x, \alpha)$ les déploiements universels au sens de Thom de ces singularités. Les exemples initiaux sont les intégrales d'Airy (1838) et de Pearcey (1948). La théorie générale est due à Maslov et les applications à l'optique en relation avec la théorie des singularités ont été approfondies en particulier par Michael Berry.

Les approximations semi-classiques de la mécanique quantique ont conduit à développer des "calculs" d'opérateurs différentiels comme le calcul de Weyl ou le calcul de Kohn-Nirenberg. Mais pour avoir un vrai calcul permettant de résoudre

des équations aux dérivées partielles (EDP) il faut pouvoir inverser les opérateurs différentiels, ce qui a conduit à la profonde théorie des opérateurs pseudo-différentiels ($O\Psi D$). La dernière section 9 du chapitre leur consacre un petit *vademecum*.

Les opérateurs différentiels linéaires d'ordre m sur une variété M de coordonnées (locales) $x = (x_1, \dots, x_n)$ sont du type $P = \sum_{0 \leq |\alpha| \leq m} a_\alpha(x) \frac{\partial^\alpha}{\partial x^\alpha}$ ($\alpha = (\alpha_1, \dots, \alpha_n)$ et $|\alpha| = (\alpha_1 + \dots + \alpha_n)$) et l'on cherche à résoudre des EDP $Pf(x) = u(x)$, u donnée, satisfaisant certaines conditions initiales et certaines conditions aux limites. On voudrait pouvoir écrire $f = P^{-1}u$ avec un inverse P^{-1} bien défini. Pour ce faire, l'idée générale est d'utiliser la transformée de Fourier $f(x) \rightsquigarrow \widehat{f}(\xi)$ qui convertit les dérivées partielles $\frac{\partial}{\partial x_k}$ en multiplication par $i\xi_k$. En tenant compte des $i^{|\alpha|}$ pour modifier les coefficients, on se retrouve avec des intégrales (du moins dans le cas de $T^*\mathbb{R}^n$) $Pf(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{i\xi \cdot (x-y)} \sigma(x, \xi) f(y) dy d\xi$ où $\sigma(x, \xi) = \sum_{0 \leq |\alpha| \leq m} a_\alpha(x) \xi^\alpha$ est un polynôme de degré m égal à l'ordre de P appelé le symbole de P . Le symbole est une fonction définie sur le fibré cotangent T^*M et sa partie homogène de degré maximal m , $\sigma_m(x, \xi) = \sum_{|\alpha|=m} a_\alpha(x) \xi^\alpha$, s'appelle le symbole principal de P . D'où l'idée de généraliser ces intégrales au cas où $\sigma(x, \xi)$ est une série et pas seulement un polynôme. On pourra alors inverser les $\sigma(x, \xi)$ au moyen de développements en série en généralisant des formules standard. On obtient ainsi des $O\Psi D$ et, si le calcul est réalisable, on peut effectuer une transformée de Fourier inverse pour définir correctement P^{-1} . En partant (dans le cas de $T^*\mathbb{R}^n$) de la représentation de $u(x)$ comme $u(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{i\xi \cdot (x-y)} u(y) dy d\xi$, on obtient $f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{i\xi \cdot (x-y)} \frac{1}{\sigma(x, \xi)} u(y) dy d\xi$.

Évidemment, de nombreuses difficultés techniques sont apparues et, sur des bases établies par Calderón et Zygmund en 1957, de nombreux auteurs ont édifié la théorie : Hörmander, Nirenberg, Kohn, Stein, Folland, Rothschild, Beals, Greiner, Maslov, Guillemin, Sternberg et plusieurs autres. Les symboles étant des fonctions sur les espaces de phases T^*M (fibrés cotangents munis de leur structure symplectique canonique), ce sont des hamiltoniens et ce que l'on appelle l'analyse "microlocale" est l'analyse locale des EDP non seulement au voisinage d'un point x , mais aussi autour d'un covecteur ξ .

Pour développer une "bonne" théorie il faut satisfaire de nombreuses contraintes sur les symboles $\sigma(x, \xi)$: être C^∞ , être à croissance "tempérée" en ξ , être susceptible d'avoir un certain "ordre" éventuellement négatif car si P est d'ordre m son inverse $Q = P^{-1}$ devrait être en gros d'ordre $-m$ et il faudrait, puisque $QP = PQ = Id$ que $\sigma_{QP,0}(x, \xi) = 1 = \sigma_{P,m}(x, \xi) \sigma_{Q,-m}(x, \xi)$. Les zéros $\xi \neq 0$ du symbole principal $\sigma_m(x, \xi)$ en ξ jouent un rôle déterminant pour son inversibilité. Si $\sigma_m(x, \xi)$ est sans zéro $\xi \neq 0$, P est dit elliptique (cas du laplacien) et sinon parabolique (cas de l'équation de la chaleur) ou hyperbolique (cas de l'équation des ondes).

Dans le cas des opérateurs elliptiques on peut inverser P au moyen de développements en séries et l'on obtient ce qu'on appelle une "parametrix" qui donne une bonne approximation $\Pi(x)$ de la solution fondamentale de $Pf(x) = \delta(x)$ (δ est la distribution de Dirac) et permet de calculer les solutions de $Pf(x) = u(x)$ par convolution. On a $P\Pi(x) = \delta(x) + w(x)$ où $w(x)$ est une fonction C^∞ à support compact concentrée sur un petit voisinage de l'origine.

Dans le cas des opérateurs hyperboliques, le symbole principal $\sigma_m(x, \xi)$ possède des zéros et $\sigma_{P,m}(x, \xi) = 0$ définit la variété caractéristique Σ de P . On peut chercher à inverser P en dehors de Σ . On se focalise sur les singularités à la fois de f et de Pf et en particulier sur ce qu'on appelle les "fronts d'onde" ("wave fronts") $WF(f)$ et $WF(P)$ et l'on étudie leurs relations. Autrement dit, on étudie la propagation des singularités comme on le fait pour la propagation le long des rayons dans le cas de l'optique.

Nous concluons alors cette dernière section du chapitre en disant un mot de la théorie des $O\Psi D$ sur les variétés de Heisenberg (M, \mathcal{K}) . Le prototype en est l'opérateur sous-laplacien $P = \sum_{j=1}^{j=n-1} X_j^2$ où les $(n-1)$ X_j engendrent la distribution \mathcal{K} . On doit tenir compte du fait que les espaces tangents sont maintenant remplacés par les groupes de Carnot de la nilpotentisation. Stern, Beals, Greiner, Taylor, Ponge ont montré comment on pouvait adapter le formalisme des $O\Psi D$ à cette nouvelle situation.

9.17. Chapitre 17

Après ce petit *vademecum* d'analyse fonctionnelle, le chapitre 17 revient sur nos modèles $\mathbb{V}_J = \mathbb{H}_{pol}$ et $\mathbb{V}_S = SE(2)$ pour approfondir les représentations fonctionnelles de la géométrie intrinsèque de la connectivité des aires visuelles primaires. Il s'agit d'explicitier l'analyse harmonique non commutative de ces structures, leurs "états cohérents" et leurs processus de diffusion. Cette approche neurogéométrique de l'analyse du signal optique permet aussi de développer, au moyen de méthodes neuromimétiques, de très puissants algorithmes d'analyse d'images et d'inpainting.

Le sujet est compliqué. Il suffit de penser au fait que l'analyse harmonique classique, déjà tellement sophistiquée, ne concerne que \mathbb{R} et \mathbb{S}^1 , pour se convaincre que l'analyse harmonique sur \mathbb{H}_{pol} ou $SE(2)$ ne peut être que hautement non triviale.

Le chapitre commence par une section (1) introduisant la notion fondamentale de représentation de groupe. Après en avoir rappelé l'histoire et les bases, elle dit un mot des représentations de dimension finie sur \mathbb{R} ou sur \mathbb{C} des groupes finis G , c'est-à-dire des morphismes de groupe $\rho : G \rightarrow GL(V)$ de G dans le groupe des automorphismes linéaires d'un espace vectoriel V de dimension finie sur \mathbb{R} ou sur \mathbb{C} . Même si G peut être horriblement compliqué, sa finitude permet de moyenniser la représentation sur G ce qui donne un projecteur sur le sous-espace de V fixé par G . On peut alors définir facilement le dual de Pontryagin \widehat{G} de G , à savoir l'ensemble des

classes d'équivalence des représentations irréductibles de G (i.e. telles que les seuls sous-espaces globalement G -invariants de V soient $\{0\}$ et $V \neq \{0\}$ lui-même). Elles sont en nombre fini ρ_j , $j = 1, \dots, k$ et de dimension finie d_j , k étant le nombre de classes de conjugaison de G . Les d_j divisent le cardinal $|G|$ de G et satisfont l'identité de Burnside $\sum_{j=1}^k d_j^2 = |G|$. Ce théorème se démontre en regardant la représentation régulière R de G dans $L^2(G)$ muni du produit scalaire $\langle \varphi, \psi \rangle = \sum_{g \in G} \varphi(g) \overline{\psi(g)}$. Par définition $(R(h)\varphi)(g) = \varphi(gh)$. Les caractères des représentations sont les $\chi_\rho(g) = \text{Trace}(\rho(g)) : G \rightarrow \mathbb{R}^*$ ou \mathbb{C}^* qui sont des fonctions "centrales", c'est-à-dire constantes sur les classes de conjugaison.

Quand on passe aux représentations linéaires et continues des groupes de Lie $\rho : G \rightarrow GL(V)$ et d'algèbres de Lie $\rho : \mathcal{G} \rightarrow \mathfrak{gl}(V)$, la situation devient beaucoup plus compliquée. Souvent les représentations sont de dimension infinie, par exemple dans des espaces de Hilbert $\mathcal{H} = L^2(X, \nu)$. Elles sont alors des morphismes continus de groupes $\rho : G \rightarrow \mathcal{L}(\mathcal{H})$ où $\mathcal{L}(\mathcal{H})$ est l'algèbre des opérateurs continus et bornés sur \mathcal{H} . Les cas les plus intéressants sont ceux où la mesure ν est G -invariante et où les représentations sont de carré intégrable unitaires (les $\rho(g)$ préservent le produit scalaire de \mathcal{H}). On étudie par conséquent en priorité les représentations unitaires irréductibles ("unirreps").

L'exemple le plus simple de représentation unitaire est celui des représentations unitaires continues de \mathbb{R} , c'est-à-dire celui des sous-groupes à un paramètre $U(t)$ de \mathcal{H} bornés en norme et satisfaisant $U(t) \circ U(t') = U(t') \circ U(t) = U(t+t')$, $U(0) = Id_{\mathcal{H}}$, $U(-t) = U(t)^{-1} = U(t)^*$. Un tel sous-groupe à un paramètre possède un générateur infinitésimal \mathbf{u} antihermitien ($\mathbf{u}^* = -\mathbf{u}$) tel que $U(t) = e^{t\mathbf{u}}$. Si $\mathbf{u} = -iA$ avec A auto-adjoint, $U(t) = e^{-itA}$. Les translations $T(t) = f(x-t)$ ont pour générateur infinitésimal $A = -i\partial_x$ et sont donc de la forme $e^{-t\partial_x}$, $\mathbf{u} = -\partial_x$ étant un opérateur fermé de domaine dense.

On peut généraliser cette situation en prenant dans un groupe de Lie G de dimension n , une base X_j de son algèbre de Lie \mathcal{G} et une représentation ρ de \mathcal{G} dans \mathcal{H} par des opérateurs antihermitiens $\mathbf{u}_j = \rho(X_j)$ de domaine $\mathcal{D}(\mathbf{u}_j)$ dense satisfaisant les relations de commutation de \mathcal{G} . La question est de savoir si cette représentation infinitésimale donnant les groupes à un paramètre $e^{t\mathbf{u}_j}$ peut être intégrée en une représentation unitaire de G . Cela est possible mais le problème est très difficile.

La section 2 définit le concept fondamental "d'état cohérent" des représentations de carré intégrable et établit son lien avec les techniques d'analyse du signal par des champs récepteurs pilotés par la géométrie de contact sous-riemannienne de nos modèles \mathbb{V}_J et \mathbb{V}_S . Elle constitue la véritable motivation du précédent *vademecum* d'analyse fonctionnelle du chapitre 16 ainsi que de l'usage de la théorie des représentations des groupes et algèbres de Lie.

Nous commençons pas revenir sur les exemples du Vol I. Il s'agit de reprendre les transformées de Fourier, de Gabor et par ondelettes en les interprétant maintenant dans le contexte des représentations de groupes. D'abord, à tout seigneur tout honneur, la transformée de Fourier sur $L^2(\mathbb{R})$ (espace des signaux d'énergie finie) qui est plus subtile que la théorie des séries de Fourier sur \mathbb{S}^1 parce que \mathbb{R} est non compact. L'analyse positionnelle du signal comme fonction $f(x)$ de la position x est transformée en son analyse fréquentielle comme fonction $\hat{f}(\omega)$ de la fréquence ω . Nous rappelons les relations d'orthogonalité des ondes planes $e^{i\omega x}$, le théorème de Parseval-Plancherel, l'isométrie entre $L^2(\mathbb{R}_{(x)})$ et $L^2(\mathbb{R}_{(\omega)})$. Cette dernière exprime la dualité entre $G = (\mathbb{R}_{(x)}, +)$ et son dual $\hat{G} = (\mathbb{R}_{(\omega)}, +)$ qui est le groupe des caractères (ondes planes) $\psi_\omega(x) = e^{i\omega x}$ agissant sur les $h(x) \in L^2(\mathbb{R}_{(x)})$ par $\pi(\omega)h(x) = \psi_{-\omega}(x)h(x) = \langle h(x), \psi_\omega(x) \rangle$ ce qui donne une unirrep de G . Il est exceptionnel que \hat{G} soit un groupe, et en plus un groupe isomorphe à G , et nous rappelons les propriétés structurelles exceptionnelles de la transformée de Fourier qui l'explique. Les $\psi_\omega(x) = e^{i\omega x}$ sont les éléments de l'"état cohérent" associé.

Nous passons ensuite à la transformée de Gabor qui a joué un rôle historique fondamental dans la définition des états cohérents. C'est elle qui établit véritablement le lien avec les profils récepteurs des neurones visuels qui sont prototypiquement des fonctions de Gabor. Elle a été introduite par Gabor au début de la mécanique quantique pour adapter les analyses de type Fourier aux relations d'incertitude de Heisenberg (1927) $\Delta x \Delta p \geq \frac{\hbar}{2}$ (Δx et Δp étant les incertitudes sur les mesures de la position x et du moment p des particules). Dans la transformée de Fourier, la formule $\delta(x) = \frac{1}{2\pi} \int e^{i\omega x} d\omega$ montre que la localisation spatiale parfaite correspond à une indétermination totale de la fréquence et que, réciproquement, l'onde plane $e^{i\omega x}$ avec une fréquence bien définie est totalement délocalisée spatialement. La relation d'incertitude de Heisenberg sur $\Delta x \Delta \omega$ est par suite dégénérée, de la forme $0 \cdot \infty$ ou $\infty \cdot 0$. La transformée de Gabor introduit des "atomes" localisés, mais imparfaitement, à la fois en position et en fréquence, atomes constituant des sortes d'intermédiaires dans la représentation d'une observable f entre la représentation positionnelle $f(x) = \langle f, \delta_x \rangle$ et la représentation fréquentielle $\hat{f}(\omega) = \langle f, e^{i\omega x} \rangle$. La transformée de Gabor est donc à la fois fréquentielle et positionnelle. Elle décompose un signal f en "atomes" espace-fréquence qui sont des fonctions $g_{(a,\omega)}(x) = e^{i\omega x} g(x - a)$ localisée à la fois dans l'espace $\mathbb{R}_{(x)}$ des positions x et dans l'espace $\mathbb{R}_{(\omega)}$ des fréquences ω . La fenêtre $g \in L^2(\mathbb{R})$ engendrant l'état cohérent des $g_{(a,\omega)}$ est une fenêtre régulière localisée autour de l'origine que l'on déplace le long de l'axe des positions. La transformée de Gabor $G_f(a, \omega) = \langle f, g_{(a,\omega)} \rangle_{\mathcal{H}}$ est par conséquent bidimensionnelle. On montre qu'elle est néanmoins une "bonne" transformée possédant une transformée inverse, satisfaisant l'identité de Parseval-Plancherel et établissant une isométrie entre $\mathbb{R}_{(x)}$ et $\mathbb{R}_{(a,\omega)}^2$.

Il est alors intéressant de savoir comment se comporte la transformée de Gabor lorsqu'on essaye de l'interpréter comme une représentation du groupe additif $(\mathbb{R}_{(a,\omega)}^2, +)$ (de même qu'on représente la transformée de Fourier comme une représentation du groupe additif $(\mathbb{R}, +)$), et cela même si les variables a et ω ne sont pas de même statut puisqu'elles sont des variables conjuguées. Le constat remarquable est qu'il faut passer du groupe $\mathbb{R}_{(a,\omega)}^2$ au groupe de Heisenberg \mathbb{H} en rajoutant une dimension et nous assistons ainsi à une véritable "genèse structurale" de \mathbb{H} .

En effet si l'on veut que $(\pi_{(a,\omega)}h)(x) = e^{i\omega x}h(x-a)$ soit une représentation du groupe additif $\mathbb{R}_{(a,\omega)}^2 = G_0$, un calcul immédiat montre qu'il faut pouvoir éliminer un facteur de phase $e^{i(\omega+\omega')x}$ pour obtenir l'égalité $\pi_{(a',\omega')}\circ\pi_{(a,\omega)} = \pi_{(a'+a,\omega'+\omega)}$. On doit donc introduire un facteur de phase $e^{-i\phi}$ dépendant d'une troisième variable ϕ et utiliser la loi de groupe $(a',\omega',\phi')(a,\omega,\phi) = (a'+a,\omega'+\omega,\phi'+\phi+\omega a')$ qui est précisément la loi du groupe de Heisenberg polarisé donnant notre premier modèle neurogéométrique $\mathbb{V}_J \simeq \mathbb{H}_{pol}$. Ce passage de \mathbb{R}^2 à \mathbb{H}_{pol} est nécessaire car les actions de translation et de modulation fréquentielle de a et de ω ne commutent pas. Ainsi la transformée de Gabor apparaît en fait comme la représentation $\pi_{(a,\omega,\phi)}$ de \mathbb{H}_{pol} modulo son centre. Le \mathbb{H}_{pol} que l'on obtient est isomorphe à notre modèle \mathbb{V}_J mais a une autre interprétation. Dans \mathbb{V}_J , les (a,ω) correspondent aux variables conjuguées (x,p) et ϕ correspond à y . Donc x et y sont les variables d'un plan naturel (le plan rétinien) ce qui n'est pas le cas de (a,ϕ) .

La reprise des modes d'analyse du signal exposés dans le Vol I se continue avec le retour sur l'analyse en ondelettes en tant qu'analyse espace-échelle (a,σ) et non plus espace-fréquence (a,ω) . Elle repose sur l'état cohérent des $\varphi_{(a,\sigma)} = \frac{1}{\sqrt{\sigma}}\varphi_0\left(\frac{x-a}{\sigma}\right)$. On doit introduire une condition d'admissibilité sur la transformée de Fourier $\widehat{\varphi_0}(\xi)$ de l'ondelette mère $\varphi_0(x)$, $C_{\varphi_0} = \int_{\mathbb{R}^+} |\widehat{\varphi_0}(\xi)|^2 \frac{d\xi}{\xi} < \infty$ afin de pouvoir définir la transformée inverse qui comprend le facteur $\frac{1}{C_{\varphi_0}}$. On obtient ainsi une "bonne" transformation. Le groupe associé est cette fois le groupe affine $(a',\sigma')(a,\sigma) = (a'+\sigma'a,\sigma'\sigma)$ produit semi-direct du groupe additif $(\mathbb{R}_{(a)}, +)$ des translations spatiales et du groupe multiplicatif $(\mathbb{R}_{(\sigma)}^+, \cdot)$ des échelles. La représentation par ondelettes $\pi_{(a,\sigma)}$ est une représentation irréductible du groupe affine dans l'espace de Hardy \mathfrak{H} des $f \in L^2(\mathbb{R}_{(x)})$ de transformée de Fourier $\widehat{f}(\xi)$ nulle pour $\xi \leq 0$.

Avant de synthétiser ces trois transformées (Fourier, Gabor, ondelettes) dans une théorie générale des états cohérents, nous nous arrêtons sur le problème de leur discrétisation, le problème étant que celle-ci doit préserver de bonnes propriétés d'analyse et de synthèse. Les méthodes de discrétisation sont à la fois computationnellement essentielles et pertinentes pour notre propos puisque les modèles \mathbb{V}_J et \mathbb{V}_S sont des limites continues d'architectures fonctionnelles neuronales discrètes (cartes

d'orientations avec pinwheels, etc.). Pour la transformée de Gabor en dimension 1, de nombreux résultats ont été obtenus. En dimension 2 des résultats nouveaux et particulièrement intéressants ont été démontrés par Mathilde Marcolli et Vasiliki Liontou à partir de la structure de contact du modèle \mathbb{V}_S . D'autres résultats spectaculaires d'inpainting ont été obtenus par Jean-Paul Gauthier sur ce même modèle. Nous y revenons à la section 11.

En général on approxime les éléments d'un Hilbert \mathcal{H} en les développant sur un système générateur de \mathcal{H} adapté au problème considéré. Ce peut être une base hilbertienne orthonormée, une base de Riesz non orthonormée, un système générateur redondant. Pour les transformées de Gabor en dimension 1 on cherche quelles fenêtres $g(x)$ et quelles discrétisations du plan $\mathbb{R}_{(a,\omega)}^2$ par un réseau Λ sont les meilleures pour obtenir une synthèse correcte. Ainsi s'introduit la question des "trames" ou des "repères" de Gabor ("Gabor frames"). Notons λ les $(a, \omega) \in \Lambda$. L'idéal serait d'arriver à satisfaire l'égalité $\|f\|_{L^2(\mathbb{R})}^2 = \sum_{\lambda \in \Lambda} |\langle f, g_\lambda \rangle|^2 = \sum_{\lambda \in \Lambda} |c_\lambda|^2 = \|Tf\|_{\ell^2}^2$ où Tf est la transformée $\{\langle f, g_\lambda \rangle\}_{\lambda \in \Lambda}$ de f . L'approximation de f par Tf serait alors la meilleure possible. Mais c'est un peu trop demander et l'on impose seulement une "condition de trame", à savoir la contrainte $A\|f\|_{L^2(\mathbb{R})}^2 \leq \sum_{\lambda \in \Lambda} |c_\lambda|^2 \leq B\|f\|_{L^2(\mathbb{R})}^2$ pour toute $f \in L^2(\mathbb{R})$, les constantes positives A et B étant indépendantes de f . Quand la fenêtre g est donnée, la recherche des bons réseaux Λ est un problème difficile résolu pour les gaussiennes grâce à la transformée de Bargmann. La condition de trame est essentiellement $a_0\omega_0 < 1$ pour des générateurs $(a_0, \omega_0) \in \Lambda$ de Λ sur \mathbb{Z} . L'intérêt de la transformée de Bargmann est de traduire les conditions de trame en ce qu'on appelle des conditions d'ensembles d'échantillonnage et d'ensembles d'unicité pour les fonctions holomorphes, problèmes que l'on sait résoudre à cause de la solidarité local-global caractéristique des fonctions analytiques.

Mais en dimension > 1 presque rien n'est connu. Comme le disait Gröchenig en 2011, "[it is] a completely uncharted territory". D'où l'intérêt des résultats de Marcolli et Liontou (2022) sur le modèle \mathbb{V}_S qui concerne les signaux en dimension 2. Les auteurs construisent une trame de Gabor à partir de la structure de contact ω de \mathbb{V}_S . En dotant la surface rétinienne de base S_r (coordonnées (x, y)) de sa structure complexe $z = x + iy$ (de surface de Riemann si le plan rétinien est considéré comme un morceau de sphère) ils construisent des discrétisations satisfaisant les conditions de trame. \mathbb{V}_S devient interprétable comme le fibré en sphères $M = \mathbb{S}(T^*S_r)$ du fibré cotangent T^*S_r (la fibre $T_z^*S_r$ est restreinte aux covecteurs de norme 1 d'orientation θ) et la structure complexe de la base S_r permet d'associer à ω et à son champ de Reeb R , une autre forme ω^J avec son champ de Reeb R^J où J est l'opérateur satisfaisant $J^2 = -Id$ (multiplication par i) associé à la structure complexe. Le point essentiel est que R et R^J avec ∂_θ forment une base naturelle de TM . Les auteurs montrent que les filtres de Gabor introduits dans notre Vol 1 peuvent conduire à de bonnes trames de Gabor sur la base S_r de dimension 2. Ces filtres sont maintenant

paramétrés par $(a, \xi) \in \mathbb{R}_{(a)}^2 \times \mathbb{R}_{(\xi)}^2 \simeq \mathbb{R}^4$ (a et ξ sont des bi-variables) et c'est dans ce \mathbb{R}^4 que l'on considère le réseau Λ composé de $\mathbb{Z}R + \mathbb{Z}\omega$ (dimension 2) et $\mathbb{Z}R^J + \mathbb{Z}\omega^J$ (dimension 2). En restreignant le \mathbb{R}_{ξ}^2 à \mathbb{S}_{ξ}^1 on passe à la structure de contact de dimension 3 sous-jacente et au champ de profils récepteurs associé. Les auteurs vérifient alors les conditions de trame en repassant, comme en dimension 1, par la transformée de Bargmann.

Après cette section sur la discrétisation et avant que d'en venir à l'analyse harmonique proprement dite sur les modèles neurogéométriques \mathbb{V}_J et \mathbb{V}_S , la section 3 formule le cadre général de l'analyse harmonique et de la théorie des représentations d'un groupe de Lie G en partant des exemples précédents. Pour les groupes compacts, la théorie peut s'inspirer directement de celle des groupes finis car la compacité permet de moyennner et de construire pour toute représentation $\rho : G \rightarrow GL(V)$ une forme hermitienne G -invariante sur V qui rend ρ unitaire. Les représentations sont complètement réductibles et le dual \widehat{G} de G est l'ensemble des unirreps, unirreps que la compacité force de plus à être de dimension finie d_ρ . \widehat{G} est discret et la mesure de Plancherel est composée des masses de Dirac d_ρ associées à ses éléments ρ . Nous exposons à leur propos le théorème fondamental de Peter-Weyl.

Pour les groupes non compacts, la théorie est aussi assez simple s'ils sont abéliens (commutatifs). Ce cas généralise directement la transformée de Fourier. Toutes les unirreps sont de dimension 1 et sont des caractères $\chi(g)$ de G . \widehat{G} possède alors une structure de groupe, ce qui est une propriété exceptionnelle. La transformée de Fourier $\widehat{f}(\chi) = \int_G f(g) \overline{\chi(g)} dg$ est une isométrie entre $L^2(G)$ et $L^2(\widehat{G})$. Si G est à la fois compact et abélien, on retrouve alors une théorie tout à fait analogue à celle des séries de Fourier pour \mathbb{S}^1 .

Mais pour les groupes non compacts et non abéliens, cette belle simplicité s'évanouit et la théorie devient beaucoup plus compliquée. Or c'est justement d'elle dont nous avons besoin pour nos modèles.

La section 4 se focalise sur la structure générale des états cohérents. Elle est particulièrement importante pour notre propos car elle formalise la "dialectique" qui relie une géométrie de connectivité neuronale à une analyse fonctionnelle au moyen de profils récepteurs. On y analyse un Hilbert \mathcal{H} en utilisant un unirrep $\pi(G)$ d'un groupe G (séparable, localement compact, avec une mesure de Haar $d\mu(g)$ invariante par translations à gauche) dans \mathcal{H} . Supposons que $\mathcal{H} = L^2(X)$. On part d'un profil récepteur de base $\varphi_0(x) \in \mathcal{H}$ (une "ondelette mère") bien localisé sur X et sur l'espace de Fourier \widehat{X} et satisfaisant les conditions d'admissibilité $C_{\varphi_0} = \int_G |\langle \varphi_0, \varphi_g \rangle_{\mathcal{H}}|^2 d\mu(g) < \infty$ faisant de l'unirrep π une représentation de carré intégrable et permettant d'inverser l'analyse des f en synthèse. Bref, on essaye d'approximer les $f \in \mathcal{H}$ par les $\varphi_g(x) = (\pi(g)(\varphi_0))(x)$ définis par la représentation π et

le profil φ_0 . $\Phi = \{\varphi_g\}_{g \in G}$ est un état cohérent si l'analyse-synthèse des f au moyen de Φ satisfait les bonnes propriétés que l'on en attend.

Ces représentations possèdent un noyau reproduisant $K(g, g') = \langle \varphi_{g'}, \varphi_g \rangle_{\mathcal{H}}$ satisfaisant $K(g', g) = K(g, g')^*$ et $\int_G K(g, g'') K(g'', g') d\mu(g) = K(g, g')$. C'est un noyau de convolution car $K(g, g') = \langle \varphi_0, \pi(g'^{-1}g)(\varphi_0) \rangle_{\mathcal{H}}$. Comme dans nos exemples, l'état cohérent Φ permet de représenter les $f \in \mathcal{H}$ par les mesures $\langle f, \varphi_g \rangle_{\mathcal{H}} = T_f(g)$ qui appartiennent à $L^2(G)$. Si f est donnée, on obtient ainsi une transformation $f \in \mathcal{H} \mapsto T_f \in L^2(G)$ et l'état cohérent Φ entrelace la représentation π avec la représentation régulière Λ de G sur $L^2(G)$ car $T \circ \pi = \Lambda \circ T$. Rappelons que $\Lambda(g)$ agit sur $T(h) \in L^2(G)$ par $\Lambda(g)(T(h)) = T(g^{-1}h)$.

Cette section 4 consacrée à la structure générale des états cohérents aborde également des problèmes un peu plus techniques faisant intervenir des passages au quotient, des orbites de la représentation coadjointe de G munies de leur structure symplectique canonique de Kirillov, la décomposition possible de G en produit semi-direct et sa représentation "quasi-régulière". Elle les illustre en revenant sur nos exemples de base.

La section 5 commence à appliquer spécifiquement ces résultats généraux à la neurogéométrie. Elle traite des représentations du groupe de Heisenberg \mathbb{H} qui permettent de formaliser l'analyse harmonique (non commutative) sur notre premier modèle (nilpotent) \mathbb{V}_J qui n'est ni compact ni abélien. Ces représentations sont omniprésentes en physique quantique. Le théorème central à leur propos, l'un des grands théorèmes sur les rapports entre la géométrie des groupes de Lie et l'analyse fonctionnelle, est le célèbre théorème de Stone-von Neumann. Il dit que toute unirrep ρ de \mathbb{H} dans un Hilbert \mathcal{H} qui est non triviale sur le centre Z de \mathbb{H} est unitairement équivalente à la représentation de Schrödinger sur les fonctions d'ondes $\psi(q) \in L^2(\mathbb{R}, \mathbb{C})$ (q la position). Cette dernière est donnée sur \mathbb{H}_{pol} par $\rho_\lambda(x, y, p)(\psi(q)) = e^{i\lambda(y+xq)}\psi(q+p)$ où λ est un paramètre (la constante de Planck en mécanique quantique). Il existe plusieurs variantes de la formule. La variable y opère comme un facteur de phase et est nécessaire car les variables conjuguées x et p ne commutent pas. Les unirreps triviales sur le centre Z sont de dimension 1 et sont les caractères $\rho_{\mu,\nu}(x, y, p)(\psi(q)) = e^{i(\mu x + \nu p)}\psi(q)$. Le dual de Pontryagin $\widehat{\mathbb{H}_{pol}}$ est donc l'union de la famille à un paramètre des ρ_λ , $\lambda \neq 0$, de dimension infinie et de la famille à deux paramètres $\rho_{\mu,\nu}$ de dimension 1. Les $\rho_{\mu,\nu}$ sont une sorte de limite des ρ_λ pour $\lambda \rightarrow 0$ car lorsque $\lambda \rightarrow 0$ l'unirrep ρ_λ tend vers une représentation très réductible qui se décompose en les unirreps $\rho_{\mu,\nu}$ de dimensions 1. Cela est lié au fait que les orbites coadjointes de \mathbb{H}_{pol} qui sont des plans pour $\lambda \neq 0$ se décomposent en points (μ, ν) pour $\lambda = 0$.

L'esquisse de la preuve du théorème de Stone-von Neumann occupe l'essentiel de la section. On vérifie d'abord que les ρ_λ , $\lambda \neq 0$, sont bien des unirreps

en utilisant le lemme de Schur et on décrit l'utilisation des systèmes d'imprimitivité et des représentations induites (Stone, von Neumann, Weyl, Mackey). Nous faisons ensuite explicitement le lien avec les travaux de Kirillov avant de passer aux représentations de Segal-Bargmann (ou de Fock). La transformée de Bargmann associe à une $f \in L^2(\mathbb{R})$ une fonction holomorphe entière (i.e. holomorphe sur tout \mathbb{C}) $(Bf)(z) \in L^2(\mathbb{C})$ de façon à obtenir une isométrie de $L^2(\mathbb{R})$ sur un sous-espace fermé \mathcal{B} de $L^2(\mathbb{C})$, dit de Segal-Bargmann. \mathcal{B} est l'espace des fonctions holomorphes de carré intégrable pour la mesure gaussienne $e^{-\pi|z|^2}$ (c'est bien un espace de Hilbert). $(Bf)(z)$ est descriptible à partir d'un noyau $B(z, t)$, dit noyau de Bargmann, par la formule $(Bf)(z) = \langle f(t), B_{\bar{z}}(t) \rangle$ ($B_{\bar{z}}(t)$ est $B(\bar{z}, t)$ considérée comme fonction de t paramétrée par \bar{z}). Ce noyau est intimement lié au noyau de la chaleur sur \mathbb{R} . Il est aussi intimement lié à la théorie quantique de l'oscillateur harmonique (l'un des exemples les plus pédagogiques de quantification géométrique) avec ses opérateurs de création et d'annihilation a et a^\dagger satisfaisant la relation de commutation $[a, a^\dagger] = Id$. Nous en rappelons la dérivation bien connue. Le noyau de Bargmann $B(z, t)$ définit un noyau reproduisant $K(z, w)$ donné par $K(z, w) = \langle B_z(t), B_w(t) \rangle_{\mathcal{H}} = e^{\pi z \bar{w}}$. On vérifie que $K(w, z) = \overline{K(z, w)}$ et surtout que $\int_{\mathbb{C}} K(z, w) K(w, z') d\mu(w) = K(z, z')$.

Dans l'isométrie entre \mathcal{B} et $L^2(\mathbb{R})$, les z^k engendrant les fonctions entières correspondent essentiellement aux polynômes d'Hermite et la représentation de Schrödinger de \mathbb{H} dans $L^2(\mathbb{R})$ se transforme par entrelacement en une représentation de \mathbb{H} dans \mathcal{B} que nous explicitons.

Cette section d'application particulièrement importante se conclut par le formalisme de Bargmann étendu par Hall aux groupes de Lie connexes compacts (ce qui n'est pas le cas de \mathbb{H}), l'exemple le plus élémentaire étant celui du cercle \mathbb{S}^1 .

La section suivante (6) généralise la section sur \mathbb{H} aux groupes de Lie nilpotents, d'abord ceux de niveaux 2 (i.e. dont la suite centrale descendante s'arrête à $\mathcal{G}_{(2)}$), puis à ceux de niveau ≥ 2 . Il s'agit d'une section technique, plus abstraite et plus structurale, qui offre au lecteur un petit aperçu de la difficulté d'une théorie un tant soit peu générale. Elle mène au théorème de Kirillov caractérisant les unirreps de G comme représentations induites. Assez compliquée, la théorie conduit à identifier le dual de Pontryagin \widehat{G} de G à l'espace \mathfrak{D}_G des orbites coadjointes \mathcal{O} munies de la topologie quotient, la mesure de Plancherel venant alors de la mesure sur \mathfrak{D}_G qui vient elle-même de la décomposition de la mesure de Lebesgue sur G en mesures sur les \mathcal{O} .

La section suivante (7) est très courte. Elle se borne à évoquer les représentations des groupes de Lie qui sont à l'opposé des groupes nilpotents, à savoir les groupes semi-simples. La théorie des unirreps et de la structure du dual \widehat{G} est d'une difficulté "himalayesque" et elle est dominée (fin années 1940, début années 1950) par les travaux de Gelfand, Naimark, Bargmann et surtout Harish-Chandra. L'exemple

typique est celui de $SL(2, \mathbb{R})$ (groupe simple non compact) ayant fait l'objet d'un traité célèbre de Serge Lang.

La section suivante (8), pour nous particulièrement importante, revient aux applications neurogéométriques et traite de notre second modèle $\mathbb{V}_S = SE(2) = \mathbb{R}^2 \rtimes SO(2)$ qui est un groupe résoluble non nilpotent. Les unirreps (i) sont définies sur le $L^2(\mathbb{S}^1, \mathbb{C}) = \mathcal{H}$ des $\psi(\theta)$, (ii) sont paramétrées par un réel $\lambda > 0$ et (iii) sont du type $\rho_\lambda : \mathbb{V}_S \rightarrow \mathcal{U}(\mathcal{H})$, $\rho_\lambda(x, y, \alpha)\psi(\theta) = e^{i\lambda(x\sin(\theta)+y\cos(\theta))}\psi(\theta + \alpha)$. Pour les petits angles θ on retrouve évidemment les représentations de \mathbb{H}_{pol} puisque \mathbb{H}_{pol} est la nilpotentisation tangente à \mathbb{V}_S à l'origine. Cette forme des unirreps de \mathbb{V}_S découle de la structure des orbites de la représentation coadjointe de $SE(2)$ muni de sa structure de produit semi-direct.

On peut alors, comme pour \mathbb{H} , passer de ces représentations à des représentations sur un espace de fonctions holomorphes entières au moyen d'une transformée de Bargmann appropriée. Ce passage a été introduit et étudié par Citti, Sarti, Barbieri et Sanguinetti. La différentielle $d\rho_\lambda$ de la représentation ρ_λ induit une représentation de l'algèbre de Lie \mathcal{G} de $G = \mathbb{V}_S$ dans $\mathcal{H} = L^2(\mathbb{S}^1, \mathbb{C})$ définie par $d\rho_\lambda(X) = \left. \frac{d}{dt} \right|_{t=0} \rho_\lambda(e^{tX})$. On montre que les générateurs X_1 et X_2 des plans de contact de \mathbb{V}_S définissent les opérateurs sur $\widehat{X}_1^\lambda \psi(\theta) = i\lambda \sin(\theta) \psi(\theta)$ et $\widehat{X}_2^\lambda \psi(\theta) = \frac{d\psi(\theta)}{d\theta}$. Les auteurs les utilisent pour définir un état cohérent $\Phi_{(x,y,\alpha)}(\theta) = \rho_\lambda(x, y, \alpha) \Phi_0(\theta)$ sur $\mathcal{H} = L^2(\mathbb{S}^1_{(\theta)}, \mathbb{C})$. La transformée de Bargmann de $\psi(\theta) \in \mathcal{H}$ sur \mathbb{S}^1 est alors la fonction $(B_\lambda \psi)(x, y, \alpha) = \Psi(x, y, \alpha)$ sur \mathbb{V}_S définie par $\Psi(x, y, \alpha) = \langle \Phi_{(x,y,\alpha)}(\theta), \psi(\theta) \rangle_{\mathcal{H}}$. La partie technique de leur travail consiste à montrer que l'on obtient ainsi une "bonne" transformation et leur théorème de base dit que B_λ est une isométrie de $\mathcal{H} = L^2(\mathbb{S}^1_{(\theta)}, \mathbb{C})$ sur l'espace $\mathcal{H} = L^2(\mathbb{S}^1_{(\alpha)}, \mathfrak{H})$. L'espace \mathfrak{H} est la clé du problème. Il est celui dont la transformée de Fourier est concentrée sur le cercle C_λ de rayon λ dans l'espace de Fourier $\widehat{\mathbb{R}^2}$ et normalisée par $\frac{1}{\lambda}$ ($\lambda \neq 0$).

La section 9 aborde le dernier grand thème du chapitre, celui de la diffusion sous-riemannienne. Rappelons qu'une de nos hypothèses centrales est que les contours subjectifs illusoire sont des géodésiques pour la géométrie sous-riemannienne naturelle des modèles \mathbb{V}_J et \mathbb{V}_S . Or dans la "dialectique" entre géométrie et analyse expliquant comment les "squelettes" géométriques sont revêtus d'une "chair" fonctionnelle, les corrélats fonctionnels des géodésiques se trouvent, entre autres, du côté des processus de diffusion pilotés par l'équation et le noyau de la chaleur. En effet ces processus sont des dynamiques de transfert le long des géodésiques. Les travaux pionniers de Beals, Gaveau et Greiner ont résolu le problème de la diffusion dans le groupe de Heisenberg. Ils sont remarquables parce que, en géométrie sous-riemannienne, les singularités de la distance et de la fonction exponentielle font que la théorie riemannienne généralisant la théorie euclidienne classique n'est plus du tout applicable.

Nous commençons par quelques rappels pédagogiques (très scolaires) sur la diffusion euclidienne et l'équation de diffusion (équation de la chaleur) $\frac{\partial f(x, \mathbf{r})}{\partial \mathbf{r}} = \Delta f(x, \mathbf{r})$ sur $\mathbb{R}_{(x)}^3$ avec la condition initiale $f(x, 0) = u(x)$, où Δ est le laplacien sur la 3-variable x . On rappelle d'abord que la solution élémentaire (la "fonction de Green") du laplacien, solution de $\Delta G(x) = \delta(x)$ ($\delta(x)$ étant la distribution de Dirac), est le potentiel de Newton $-\frac{1}{4\pi\|x\|}$. Les géodésiques de la métrique euclidienne de \mathbb{R}^3 sont les droites et leur hamiltonien $H(x, \xi)$ sur le fibré cotangent $T^*\mathbb{R}^3$ est (au signe près) le symbole $\|\xi\|^2$ du laplacien Δ . L'action est $S(x, \mathbf{r}) = \frac{\|x\|^2}{4\mathbf{r}}$ et la solution élémentaire de l'équation de diffusion est $P(x, \mathbf{r}) = \frac{1}{(4\pi\mathbf{r})^{\frac{3}{2}}} e^{-S(x, \mathbf{r})}$. La solution pour un second membre $u(x)$ quelconque est donc la convolution $f(x, \mathbf{r}) = (P_{\mathbf{r}} * u)(x)$ avec le noyau de la chaleur $P_{\mathbf{r}}(x, y) = \frac{1}{(4\pi\mathbf{r})^{\frac{3}{2}}} e^{-\frac{\|x-y\|^2}{4\mathbf{r}}}$. Le lien entre le noyau de la chaleur et la fonction de Green du laplacien est $G(x) = \int_0^\infty P(x, \mathbf{r}) d\mathbf{r}$.

Nous rappelons alors les liens entre le processus de diffusion et le calcul stochastique des marches aléatoires. C'est une occasion de rappeler les bases du mouvement brownien et des processus de Wiener (Bachelier, Einstein, Perrin, Langevin, Fokker, Planck, Uhlenbeck, Ornstein, Lévy, Wiener, Doob, Ito).

La généralisation du cas euclidien au cas riemannien est assez directe, les exponentielles $e^{-\frac{\|x-y\|^2}{4\mathbf{r}}}$ y devenant (en dimension 3) $e^{-\frac{d(x,y)^2}{4\mathbf{r}}}$. C'est la diffusion sous-riemannienne qui présente des propriétés radicalement nouvelles. La section 10 l'étudie pour les deux modèles \mathbb{V}_J et \mathbb{V}_S . Le cas du groupe the Heisenberg \mathbb{H} (non polarisé) a été magistralement résolu à partir de papiers fondateurs remontant à 1977 par Beals, Gaveau, Greiner, Chang et Hulanicki. Il montre la difficulté majeure du problème. Notons (z, t) les points de \mathbb{H} , $z = x_1 + ix_2$, \mathcal{K} la distribution des plans de contact et X_1, X_2 les générateurs de \mathcal{K} . Le laplacien hypoelliptique sous-riemannien est $\Delta_{\mathcal{K}} = (X_1)^2 + (X_2)^2$. Un théorème de Folland (1973) dit que la solution fondamentale de $\Delta_{\mathcal{K}}$ est $C \frac{1}{\sqrt{|z|^4 + 16t^2}}$ (à comparer au cas euclidien $C \frac{1}{\|x\|}$). On note les poids différents des variables z et t . La formule de Gaveau pour la solution fondamentale de l'équation de diffusion $\frac{\partial f(z, t, \mathbf{r})}{\partial \mathbf{r}} = \frac{1}{2} \Delta_{\mathcal{K}} f(z, t, \mathbf{r})$ (le $\frac{1}{2}$ est d'origine physique) calculée en (z, t) au temps \mathbf{r} est $P(z, t, \mathbf{r}) (= P_{\mathbf{r}}(z, t)) = \frac{1}{(2\pi\mathbf{r})^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} e^{i\frac{\tau t}{2} - \frac{|z|^2}{2\mathbf{r}} \frac{2\tau}{\sinh(2\tau)}} d\tau$. On remarque que le facteur en $\mathbf{r}^{-\frac{3}{2}}$ des cas euclidien et riemannien devient un facteur $\mathbf{r}^{-2} = \mathbf{r}^{-\frac{4}{2}}$, ce qui manifeste que, bien que topologiquement de dimensions 3, \mathbb{H} est de dimension de Hausdorff 4 comme espace métrique sous-riemannien. Le lien avec la fonction de Green du laplacien hypoelliptique est $G_{\mathbb{H}}(z, t) = \int_0^\infty P(z, t, \mathbf{r}) d\mathbf{r}$ et se calcule par la méthode des résidus.

La formule de Gaveau se démontre à partir des théorèmes de Lévy sur le mouvement brownien. Elle est une intégrale oscillante difficile à évaluer et on doit utiliser le principe de la phase stationnaire. On cherche en priorité ses approximations semi-classiques pour $\tau \rightarrow 0^+$, c'est-à-dire ses asymptotiques pour les temps courts. Le sujet est technique mais certains cas particuliers sont facilement calculables. Si $t = 0$, $P_\tau(z \neq 0, 0)$ pour $\tau \rightarrow 0^+$ se calcule par la méthode de Laplace et donne $P_\tau(z \neq 0, 0) \underset{\tau \rightarrow 0^+}{\sim} \frac{\sqrt{3}}{2} \frac{1}{(2\pi\tau)^{\frac{3}{2}}} \frac{1}{|z|} e^{-\frac{|z|^2}{2\tau}}$. Par ailleurs, pour $z = 0$ on obtient $P_\tau(0, t \neq 0) \underset{\tau \rightarrow 0^+}{\sim} \frac{1}{(2\tau)^2} \frac{1}{4 \cosh^2(\frac{\pi t}{4\tau})}$.

Nous disons alors un mot de la façon dont les points conjugués et les points de coupure (cut locus) influent sur la diffusion. Dans le cas riemannien (par exemple des sphères ou des ellipsoïdes) on connaît certaines formules. Sur la sphère, si $c(a)$ est le point antipodal de a , on montre que le facteur de l'exponentielle $e^{-\frac{d(a,b)^2}{4\tau}}$ du noyau $P_\tau(a, b \neq c(a))$ qui est en $\frac{1}{\tau}$ devient en $\frac{1}{\tau^{\frac{3}{2}}}$ pour $b = c(a)$. Pour les ellipsoïdes de révolution, si a est sur l'équateur et si b_c est une extrémité de type cusp du cut locus de a (qui est une astroïde), alors pour $\tau \rightarrow 0^+$ ce facteur est en $\frac{1}{\tau^{\frac{5}{2}}}$ (Barilari-Jendrej, 2013). En fait, dès la fin des années 1960 on s'est mis à étudier systématiquement ces asymptotiques en temps court en géométrie riemannienne et sous-riemannienne. Nous citons entre autres les résultats de Ben Arous, Barilari, Boscain, Neel.

Avant que d'en venir au noyau de la chaleur du second modèle \mathbb{V}_S , nous rappelons qu'à peu près en même temps que Gaveau, Hulanicki a calculé le noyau de la chaleur du cas Heisenberg polarisé \mathbb{H}_{pol} . Nous signalons également la généralisation de la diffusion sous-riemannienne de \mathbb{H} aux groupes de Lie de dimensions 3 unimodulaires par Agrachev et ses collaborateurs en termes d'analyse spectrale du laplacien sous-riemannien. Cela est essentiel car on ne connaît pas en général d'équivalent de la formule de Gaveau. C'est en particulier le cas pour $SE(2)$! Pour surmonter l'obstacle on repasse par la différentielle des unirreps sur les champs de vecteurs invariants à gauche (i.e. sur l'algèbre de Lie). Pour la base standard (X_1, X_2) de la distribution de contact \mathcal{K} on regarde \widehat{X}_1^λ et \widehat{X}_2^λ .

Dans le cas de \mathbb{H}_{pol} on a $\widehat{X}_1^\lambda u(s) = i\lambda u(s)$ et $\widehat{X}_2^\lambda u(s) = \frac{du(s)}{ds}$ et donc

$$\widehat{\Delta}_{\mathcal{K}}^\lambda u(s) = \left(\left(\widehat{X}_1^\lambda \right)^2 + \left(\widehat{X}_2^\lambda \right)^2 \right) u(s) = \frac{d^2 u(s)}{ds^2} - \lambda^2 u(s)$$

qui est l'équation de l'oscillateur harmonique. Le spectre est discret et on peut développer le noyau de la chaleur sur un ensemble de fonctions propres normalisées. Les valeurs propres sont $\alpha_n^\lambda = \frac{2n+1}{\lambda}$ et les fonctions propres sont les fonctions d'Hermite rescalées

$$u_n^\lambda(s) = (2^n n! \sqrt{\pi})^{-\frac{1}{2}} \lambda^{\frac{1}{4}} e^{-\lambda \frac{s^2}{2}} H_n(\sqrt{\lambda} s),$$

H_n étant le n -ème polynôme d'Hermite. On obtient ainsi le noyau dit de Mehler.

Dans le cas de \mathbb{V}_S , on a $\widehat{X}_1^\lambda \psi(\theta) = i\lambda \sin(\theta) \psi(\theta)$ et $\widehat{X}_2^\lambda \psi(\theta) = \frac{d\psi(\theta)}{d\theta}$ et l'on obtient $\widehat{\Delta}_K^\lambda = \frac{d^2\psi(\theta)}{d\theta^2} - \lambda^2 \sin^2(\theta) \psi(\theta)$ qui est l'équation de Mathieu. Les valeurs propres sont connues et les fonctions propres sont déductibles des fonctions spéciales ce_n et se_n . Enfin nous terminons par la "confluence" entre les modèles \mathbb{V}_J et \mathbb{V}_S au moyen d'une interpolation due à Brahim Zahaf et Manchon.

Pour conclure le chapitre, nous présentons dans une dernière section 11 la façon dont, en discrétisant la diffusion sous-riemannienne du modèle \mathbb{V}_S , Jean-Paul Gauthier a obtenu des résultats spectaculaires d'inpainting. Ces algorithmes neuro-mimétiques remarquables ont des capacités de complétion et de reconstruction d'images lacunaires très supérieures à celles du système visuel. Il n'y a pas à s'en étonner car lorsqu'on a bien compris la structure d'une réalité naturelle R on peut en général en dériver des réalités virtuelles R -mimétiques plus puissantes. Mais cela reste néanmoins saisissant.

L'idée développée initialement par Citti, Sarti, Duits et Franken est, étant donnée une image lacunaire $f_\Omega(x, y)$ définie sur un domaine très fragmenté Ω du champ visuel, de relever ses lignes de niveau dans le fibré $\mathbb{V}_S = \mathbb{R}^2 \rtimes \mathbb{S}^1$ au-dessus de Ω . Les relevées legendriennes définissent ainsi une surface S_Ω de \mathbb{V}_S au-dessus de Ω . On effectue alors une diffusion sous-riemannienne pendant un certain laps de temps pour compléter de façon géodésique ces \mathcal{K} -courbes intégrales lacunaires de S_Ω au-delà de S_Ω puis on reprojette sur la base \mathbb{R}^2 de \mathbb{V}_S en moyennant sur les fibres afin d'obtenir une image $f_{\Omega'}(x, y)$ qui complète $f_\Omega(x, y)$ sur $\Omega' \supset \Omega$. On itère le processus. Évidemment le problème est que lorsqu'on part d'un point du bord $\partial\Omega$ de Ω on ne sait pas à quel point situé de l'autre côté de la lacune doit aboutir la géodésique. On peut alors introduire des processus stochastiques de façon à trouver la géodésique la plus probable.

Les calculs étant extrêmement coûteux, Gauthier et Prandi ont eu l'idée de discrétiser les orientations θ de la fibre \mathbb{S}^1 (i.e. $SO(2)$) de \mathbb{V}_S (la discrétisation du plan de base \mathbb{R}^2 conduit aux trames de Gabor analysées par Marcolli et Liontoux, section 2.2.2). L'avantage de la discrétisation de $SO(2)$ en N orientations est que le groupe $\mathbb{V}_S \simeq SE(2)$ devient un groupe de Moore $SE(2, N)$, groupe localement compact non compact et non commutatif dont toutes les unirreps sont de dimension finie. Qui plus est $SE(2, N)$ est aussi un groupe "maximalement presque périodique" (MAP) qui possède assez d'unirreps (de dimension finie, donc) pour "séparer" les points. Le laplacien hypoelliptique sous-riemannien peut alors être "désintégré" en une intégrale d'opérateurs de type Mathieu de dimension finie ce qui permet d'expliciter le noyau de la chaleur de façon pas trop compliquée. On peut ainsi approximer de façon très accélérée les fonctions sur le groupe par des fonctions presque périodiques au sens de Harald Bohr. On obtient alors une formule explicite du noyau de la chaleur pour $SE(2, N)$, ce qui n'est pas le cas pour $SE(2)$.

9.18. Chapitre 18

Dans un dernier chapitre 18 nous regroupons quelques brèves indications sur d'autres aspects de la neurogéométrie que nous ne traitons pas de façon détaillée.

Nous revenons d'abord sur des phénomènes de complétion, plus spectaculaires que ceux des contours illusoires et de l'inpainting, qui concernent la construction de structures géométriques du percept qui ne sont pas du tout contenues, même à l'état très lacunaire, dans les données sensorielles.

Le premier phénomène, traité dans la section 1 est celui du cut-locus (ou "squelette", ou "medial axis") des contours fermés C d'une forme E dans un plan. Nous l'avons déjà évoqué dans l'*Introduction* du Vol I. Le spécialiste de la perception Harry Blum et René Thom en ont souligné l'importance et Tai Sing Lee et David Mumford en ont montré la réalité neurophysiologique. Le cut-locus est le lieu singulier d'une propagation de fronts d'ondes activée par le contour C . Nous en rappelons la définition géométrique comme ensemble de Maxwell et expliquons son utilité pour la décomposition méréologique de la forme E en parties.

Le second phénomène, traité section 2, est celui des patterns de vision entoptique, bien connu des spécialistes de la perception visuelle depuis les travaux de Heinrich Klüver. Il s'agit d'hallucinations très géométriques dues à une activité endogène, spontanée ou provoquée, du cortex visuel. Nous en avons également parlé dans l'*Introduction* du Vol I. Paul Bressloff, Jack Cowan, Bard Ermentrout et Martin Golubitsky en ont donné des modèles neuronaux remarquables à partir des équations de réseaux de neurones de Wilson-Cowan-Hopfield. En fait, leur travail revient à implémenter l'architecture fonctionnelle de $V1$ dans les poids synaptiques des neurones du réseau et à montrer que la solution homogène initiale (champ visuel homogène) devient structurellement instable et bifurque spontanément lorsque le paramètre "d'excitabilité" de $V1$ dépasse une certaine valeur critique. Il existe alors des "modes propres" (des "eigenforms") engendrés par les bifurcations et correspondant très bien aux données empiriques.

Dans une troisième section (3), nous évoquons très brièvement quelques aspects supplémentaires de la neurogéométrie, certains liés à la phénoménologie de la perception et déjà décrits (mais évidemment non modélisés) par Husserl et Merleau-Ponty, d'autres dus à l'équipe de Giovanna Citti et Alessandro Sarti.

1. Le passage $2D \rightarrow 3D$ ($D =$ "dimension") et la "perception par esquisses" qui concernent la reconstruction d'objets $3D$ dans \mathbb{R}^3 à partir de l'espace fonctionnel de leurs contours apparents $2D$ sur les plans de projection.

2. Le passage $2D \rightarrow 3D$ fait intervenir la stéréopsie. En $3D$, le modèle des relevées legendriennes des courbes planes de \mathbb{R}^2 dans le fibré de dimension 3 $\mathbb{R}^2 \times \mathbb{S}^1$ doit être notablement complexifié puisqu'il faut considérer les relevées legendriennes des courbes gauche de \mathbb{R}^3 dans le fibré de dimension 5 $\mathbb{R}^3 \times \mathbb{S}^2$. Il s'agit alors de savoir si la vision binoculaire peut effectuer une telle complexification. Nous résumons un

travail qui montre que les relevées legendriennes des courbes gauche de \mathbb{R}^3 peuvent effectivement être reconstruites à partir de deux modèles monoculaires \mathbb{V}_S .

3. Le contrôle kinesthésique du regard et le lien entre vision et motricité a été beaucoup étudié. Nous évoquons à cet égard d’une part les travaux bien connus d’Alain Berthoz sur la sensori-motricité et, d’autre part, les liens avec la phénoménologie husserlienne de la kinesthésie approfondis par Jean-Luc Petit.

4. À propos de la motricité nous évoquons aussi une modélisation neurogéométrique sous-riemannienne de l’organisation columnaire des mouvements de la main dans le cortex moteur primaire.

Dans une quatrième section (4), nous nous arrêtons un instant sur certaines extensions qu’il serait nécessaire d’apporter à nos modèles. Bien que déjà mathématiquement assez sophistiqués, ceux-ci ne modélisent en effet qu’une toute petite partie des phénomènes neurophysiologiques du cortex visuel primaire. Une foule de problèmes restent en suspens.

1. D’abord il faudrait tenir compte du fait que la réponse des neurones aux stimuli naturels est plus compliquée que celle à des stimuli simplifiés comme des barres ou des grattings utilisés dans les expériences ayant révélé l’organisation en pinwheels de $V1$.

2. Ensuite, il existe de nombreux autres paramètres (“engrafted variables”) que l’orientation, la courbure ou l’échelle (qui sont à la base de nos modélisations par des structures de contact et des structures symplectiques). C’est par exemple le cas de la phase, de la fréquence spatiale ou de la dominance oculaire que nous avons traitées dans le Vol I. Il y a donc plusieurs “cartes” “d’engrafted variables”, c’est-à-dire plusieurs fibrations, en interaction.

3. Enfin, notre niveau d’analyse est un niveau mésoscopique “coarse grained”. Un élément de contact (a, p) correspond en fait à une colonne au niveau des neurones individuels, et représente donc un petit module de plusieurs milliers de neurones dont la dynamique neurophysique “fine grained” est d’une grande complexité. Le lien entre neurogéométrie et micro physique neuronale reste entièrement à développer.

9.19. Chapitre 19 : Conclusion

Après ce long parcours de modélisation nous présentons quelques éléments de conclusion. Ils sont focalisés sur des problèmes épistémologiques liés à la philosophie de l’espace et se proposent de montrer en quoi la neurogéométrie n’est pas seulement un outil de modélisation mais également une nouvelle perspective fondationnelle.

Dans une première section (1), nous montrons comment la neurogéométrie permet de notablement approfondir le problème des corrélats neuronaux de la conscience visuelle, “hard problem” qui est au centre des débats sur les rapports entre neurosciences et philosophie de l’esprit. Épistémologiquement, la neurogéométrie n’est ni un monisme réductionniste (réductibilité de la conscience à ses corrélats neuronaux),

ni un dualisme (irréductibilité de la conscience à ses corrélats neuronaux). Elle favorise une conception “émergentielle” de la conscience.

Pour préciser exactement dans quel sens, nous commentons un peu en détail un article d’Alva Noe et Evan Thomson “Are there Neural Correlates of Consciousness?” (2004) et nous montrons, en nous aidant des résultats sur la vision entoptique, comment nos modèles permettent de dépasser tout un ensemble d’antinomies conceptuelles classiques. Cette capacité vient de ce qu’ils expliquent comment les architectures fonctionnelles intègrent en structures globales les détections locales effectuées par les neurones individuels.

Puis nous concluons l’ouvrage sur une section purement philosophique (2) mettant en relation la neurogéométrie avec le problème ancien et critique de l’esthétique transcendantale kantienne comme philosophie de l’espace et du temps. Nous argumentons en faveur d’une solidarité architectonique entre le matérialisme neuronal et l’idéalisme transcendantal. En effet, ainsi que l’ont affirmé plusieurs éminents spécialistes (tels John O’Keefe ou Stanislas Dehaene que nous avons cités en ouverture de notre Vol I) les neurosciences cognitives donnent un sens précis et opérationnel au “synthétique a priori” comme formatage géométrique des données sensorielles (les “formes de l’intuition”). Nos modèles confirment cette affirmation et l’on peut dire que la neurogéométrie est la géométrie des structures synthétiques a priori de la perception.

Si l’on y réfléchit, cela est d’ailleurs assez normal. Entreprendre, comme Kant l’a fait, une réflexion fondationnelle sur les conditions de possibilité de la connaissance, c’est en fait, comme il le disait lui-même à propos du schématisme des concepts empiriques, réfléchir sur

“un art caché dans les profondeurs de l’âme humaine, et dont nous aurons de la peine à arracher à la nature les secrets du fonctionnement pour les mettre à découvert sous les yeux”.

Mais ces “profondeurs de l’âme” sont précisément les architectures fonctionnelles neuronales restées effectivement “cachées” tant que le cerveau était une “boîte noire”, malgré des intuitions inspirées comme celles de Descartes sur la “glande pinéale” (épiphyse) dans son traité sur “Les passions de l’âme” (1649). Mais à partir du moment où le cerveau devient, serait-ce partiellement, une boîte “transparente” grâce à des outils révolutionnaires d’observation, les “conditions de possibilité” se convertissent en un ensemble de phénomènes neurophysiologiques expérimentalement accessibles et mathématiquement modélisables. Leur réalité profonde commence à être “mise à découvert sous les yeux” grâce à l’imagerie *in vivo* (notre volume I) et mathématisée grâce à la neurogéométrie immanente des architectures fonctionnelles corticales (ce volume-ci).

Partie I : Géométrie à la Lie-Cartan

La géométrie différentielle de $V1$ comme fibration

1. Brefs rappels expérimentaux (Vol I)

Commençons pour fixer les idées par quelques très brefs rappels expérimentaux largement détaillés dans le Vol I.

1.1. Champs récepteurs et analyse en ondelettes

Le premier concept neurophysiologique de base dont on a besoin pour comprendre la façon dont, dès les plus bas niveaux, le système visuel représente le signal optique est celui de *champ récepteur* (CR) d'un neurone visuel. Dès le bas niveau précoce, sensoriel et périphérique des cellules ganglionnaires de la rétine qui effectuent ce que l'on appelle la transduction (le codage neuronal) du signal (i.e. le mesurent et le transforment en une information neuronale exploitable par le système nerveux central), le filtrage induit un formatage géométrique.

La définition la plus simple du CR d'un neurone visuel est la zone de la rétine à laquelle il répond parce qu'il s'y trouve relié à travers la connectivité compliquée des voies rétino-géniculocorticales menant de la rétine au cortex à travers le relais thalamique du corps genouillé latéral. On montre qu'il existe des zones – dites ON – du CR qui répondent de façon positive et excitatrice à des stimuli lumineux ponctuels (c'est un problème de réponse impulsionnelle). D'autres zones – dites OFF – répondent de façon négative et inhibitrice. D'où le concept de profil récepteur (PR) d'un champ récepteur. Le PR est une fonction $\varphi : D \rightarrow \mathbb{R}$ qui est définie sur le domaine D du CR et s'identifie à la fonction de transfert du neurone considéré comme filtre.

En fait, la définition du CR peut varier considérablement suivant la façon dont on définit la réponse d'un neurone visuel. Les réponses impulsionnelles en termes de trains de potentiels d'action (spikes) définissent un concept classique étroit de CR dit "champ minimal de décharge". Mais dans la mesure où le déclenchement d'un spike exige le franchissement d'un seuil pour le potentiel de membrane (les neurones sont des sortes d'automates à seuil), il peut y avoir de nombreuses réponses sous-liminaires d'un neurone. Comme l'a montré Yves Frégnac (Frégnac, Shulz [194]; voir aussi Maffei [344] et Lamme [313]), cela permet de raffiner considérablement le concept classique de CR.

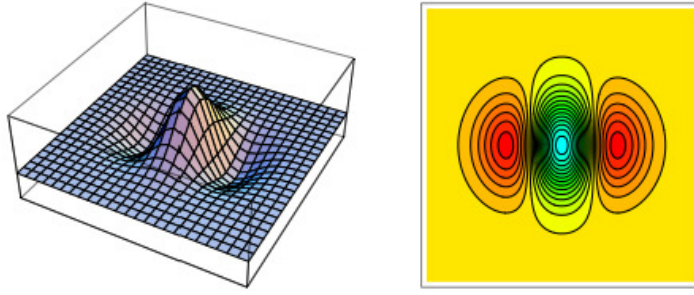


FIGURE 1. Profil récepteur standard d'un neurone simple de V1.

Des méthodes raffinées d'électrophysiologie (cf. par ex. DeAngelis [137]) ont permis de mesurer les lignes de niveau des PRs de différents neurones visuels. On commence par utiliser des gratings sinusoïdaux de bandes claires et sombres et, en variant les largeurs des bandes, leur fréquence spatiale, leur orientation, leur vitesse de déplacement, leur fréquence temporelle et leur taille, on définit avec précision la localisation, la taille et l'orientation préférentielle du PR. Puis on utilise des séquences de flashes (par ex. 50ms) rapides et aléatoires, spots ponctuels et petites barres, placés sur une grille (par ex. 20×20) séparés par des intervalles d'environ 100ms-1s. On obtient ainsi quelques milliers de spikes. La corrélation entre l'input (les flashes) et l'output (les potentiels d'action, les spikes), fournit la fonction de transfert du neurone. Il s'agit là d'un véritable tour de force expérimental et les résultats en sont tout à fait remarquables.

Les profils récepteurs de neurones simples de V1 les plus communément utilisés dans la littérature sont constitués d'un domaine ON allongé dans une direction préférentielle flanqué de deux domaines OFF plus petits. On peut en donner un modèle idéalisé en dérivée seconde de Gaussienne $\frac{\partial^2 G}{\partial x^2}$ (cf. figure 1).

Soit alors $I(x, y)$ l'intensité du signal optique défini sur le domaine R de la rétine. Soit $\varphi(x, y)$ le PR d'un CR centré sur 0 (le centre de R) d'un certain type de neurone visuel. Si le CR est centré en (x_0, y_0) , le PR est $\varphi(x - x_0, y - y_0)$. Un neurone visuel agissant comme un filtre sur le signal optique, sa réponse a pour valeur la moyenne du signal pondérée par φ soit l'intégrale :

$$I_\varphi(x_0, y_0) = \int_D I(x', y') \varphi(x' - x_0, y' - y_0) dx' dy' . \quad (1)$$

C'est la mesure du signal I en (x_0, y_0) . Si un champ de CRs de profils identiques recouvre R , alors la réponse est la *convolution* de I par φ :

$$I_\varphi(x, y) = \int_D I(x', y') \varphi(x' - x, y' - y) dx' dy' = (I * \varphi)(x, y) . \quad (2)$$

Comme l'a proposé Luc Florack [186], un disciple de Jan Koenderink, une bonne façon de voir les choses est de traiter le signal I (qui est une "mauvaise" fonction, très bruitée) comme une *distribution* au sens de Schwartz, c'est-à-dire une fonctionnelle linéaire continue sur un espace de fonctions test (fonctions C^∞ à support compact, ou à décroissance rapide pour les distributions tempérées) et de traiter les PRs $\varphi(x - x_0, y - y_0)$ – qui sont des fonctions bien régulières et bien localisées – comme des classes de fonctions test *neuralement câblées* dans le système visuel.

Si maintenant le PR $\varphi(x, y)$ est une dérivée de Gaussienne DG , la formule classique $I * DG = D(I * G)$ fournit une interprétation fonctionnelle du filtrage $I * DG$: le filtrage est équivalent à $D(I * G)$, c'est-à-dire à l'application de l'opérateur différentiel D au signal I lissé à l'échelle de G . Par exemple, les cellules ganglionnaires de la rétine ou celles du corps genouillé latéral dont les PRs sont en laplacien de Gaussienne ΔG calculent le laplacien du signal régularisé $\Delta(I * G) = I * \Delta G$. Il est bien connu depuis longtemps par les neurophysiologistes qu'elles détectent des contrastes spatiaux (cf. Buser, Imbert [84] et Marr [354]). En fait, elles effectuent une *analyse en ondelettes* permettant d'extraire les discontinuités encodées dans le signal (Mallat [347]).

1.2. La structure hypercolumnaire de l'aire V1

Les études neurophysiologiques ont permis de distinguer trois types de structures de V1, respectivement laminaire, rétinotopique et (hyper)columnaire.

(i) La structure laminaire (d'épaisseur environ 1,8 mm) est constituée de 6 couches "horizontales" (i.e. parallèles à la surface du cortex), la plus importante pour notre propos étant la couche 4. Cette structure varie suivant les espèces.

(ii) La rétinotopie signifie que les projections (au sens neurophysiologique) de la rétine sur les couches corticales sont des applications préservant la topographie rétinienne. Un exemple typique en est la représentation conforme logarithmique existant entre la rétine et la sous-couche de la couche 4 où se projettent majoritairement les fibres issues du corps genouillé latéral.

(iii) La structure columnaire et hypercolumnaire est la grande découverte des prix Nobel Hubel et Wiesel au début des années 1960. Il existe dans l'aire V1 des neurones sensibles à l'orientation (cellules dites "simples" par opposition aux cellules "complexes" ne détectant pas d'orientation), à la dominance oculaire et à la couleur. Si l'on se restreint à l'orientation, on peut dire que ces neurones détectent des couples (a, p) d'une position rétinienne a et d'une orientation p en a , c'est-à-dire, géométriquement parlant, des "éléments de contact" du plan visuel.

Par les méthodes d'enregistrement de réponse à des stimuli appropriés (par exemple des barres orientées traversant le champ récepteur des neurones), on a pu montrer que, perpendiculairement à la surface du cortex, la position rétinienne a et

l'orientation préférentielle p restent à peu près constantes. Cette redondance “verticale” (ce que les neurophysiologistes appellent un codage par population) définit les “colonnes d'orientation”. Comme l'a montré DeAngelis [136] c'est la variation de la phase qui domine dans les colonnes :

“spatial phase is the single parameter that accounts for most of the difference between receptive fields of nearby neurons”.

Par ailleurs le codage par population permet au système d'avoir une résolution supérieure à celle des neurones individuels (Snippe-Koenderink [506]).

En revanche, parallèlement à la surface du cortex, l'orientation préférentielle p varie de 0° à 180° par pas d'environ 10° tous les $50 - 100\mu$. Ce regroupement “horizontal” de colonnes définit une “hypercolonne d'orientation” qui est un micromodule neuronal d'environ $500\mu - 1\text{mm}$.

À travers cette architecture fonctionnelle hypercolumnnaire, à chaque position rétinienne a se trouve associé de façon rétinotopique un exemplaire (discrétisé) de l'espace P des directions p du plan (droite projective \mathbb{P}^1). Il existe par conséquent une implémentation neuronale de la fibration (triviale) $\pi : R \times P \rightarrow R$ ayant pour base l'espace rétinien R et pour fibre la variété P .

La structure de fibration avec R comme espace de base et un espace de variables “secondaires” (orientation, dominance oculaire, couleur, direction du mouvement, texture, etc.) comme fibre a été bien formulée par Hubel (de façon évidemment neurophysiologique et non pas mathématique) au moyen du concept d’“engrafted variables” :

“What the cortex does is map not just two but many variables on its two-dimensional surface. It does so by selecting as the basic parameters the two variables that specify the visual field coordinates (distance out and up or down from the fovea), and on this map it engrafts other variables, such as orientation and eye preference, by finer subdivisions.” (Hubel [259], p. 131)

1.3. L'organisation en pinwheels de V1

À partir des années 1990, des expériences rendues possibles par les progrès de l'imagerie cérébrale (“in vivo optical imaging based on activity-dependent intrinsic signals”) ont montré que les hypercolonnes d'orientation sont en fait géométriquement organisées en “roues d'orientation” ou pinwheels. La couche corticale (ou les couches) est réticulée par un réseau de points singuliers qui sont les centres de pinwheels locaux qui se recollent en une structure globale.

La méthode d'imagerie permet d'acquérir des images de l'activité des couches corticales superficielles. On présente à l'animal plusieurs dizaines de fois (de 20 à 80 fois) des grilles de fort contraste constituées de bandes noires (par ex. de 6.25°) alternées avec des bandes blanches (par ex. de 1.25°), avec plusieurs (par ex.

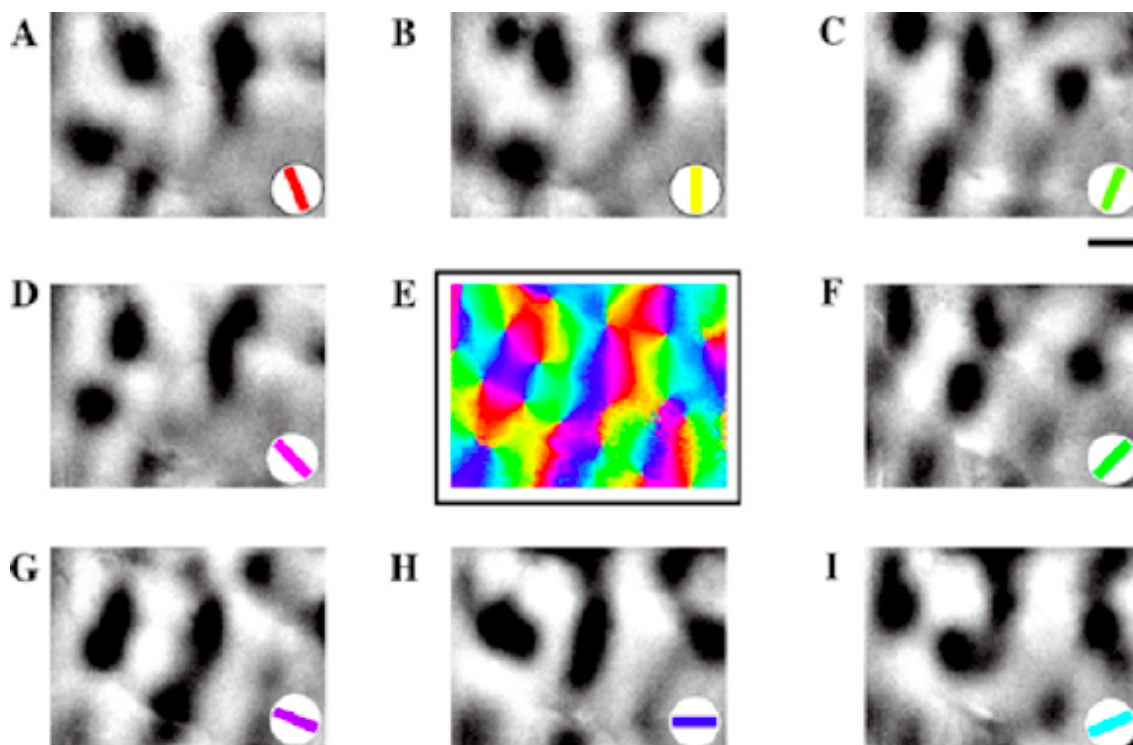


FIGURE 2. Méthode d'établissement d'une carte d'orientation de $V1$ (les orientations préférentielles sont codées par des couleurs).

8) orientations différentes et une vitesse par ex. de $22.5^\circ/s$. On ouvre une fenêtre crânienne au-dessus de $V1$ et on illumine le cortex en lumière orange. On somme ensuite les images de l'activité de $V1$ obtenue pour différentes grilles et l'on construit des cartes différentielles que l'on normalise (en divisant la déviation relative à la moyenne de chaque pixel par la déviation moyenne globale) et dont on élimine le bruit de basse fréquence. On obtient ainsi des cartes comme celles de la figure 2 et de la figure 3 dues à William Bosking [71] (laboratoire de David Fitzpatrick) où les orientations sont codées par des couleurs et où les lignes d'iso-orientation sont donc les lignes monochromatiques.

On remarque qu'il existe trois classes de points :

(i) des points réguliers où le champ d'orientation est localement trivial au sens où les lignes d'iso-orientation y sont approximativement parallèles ;

(ii) des points singuliers au centre des pinwheels où convergent toutes les orientations ; ils sont de chiralités opposées (lévogyres ou dextrogyres) lorsqu'ils sont adjacents ;

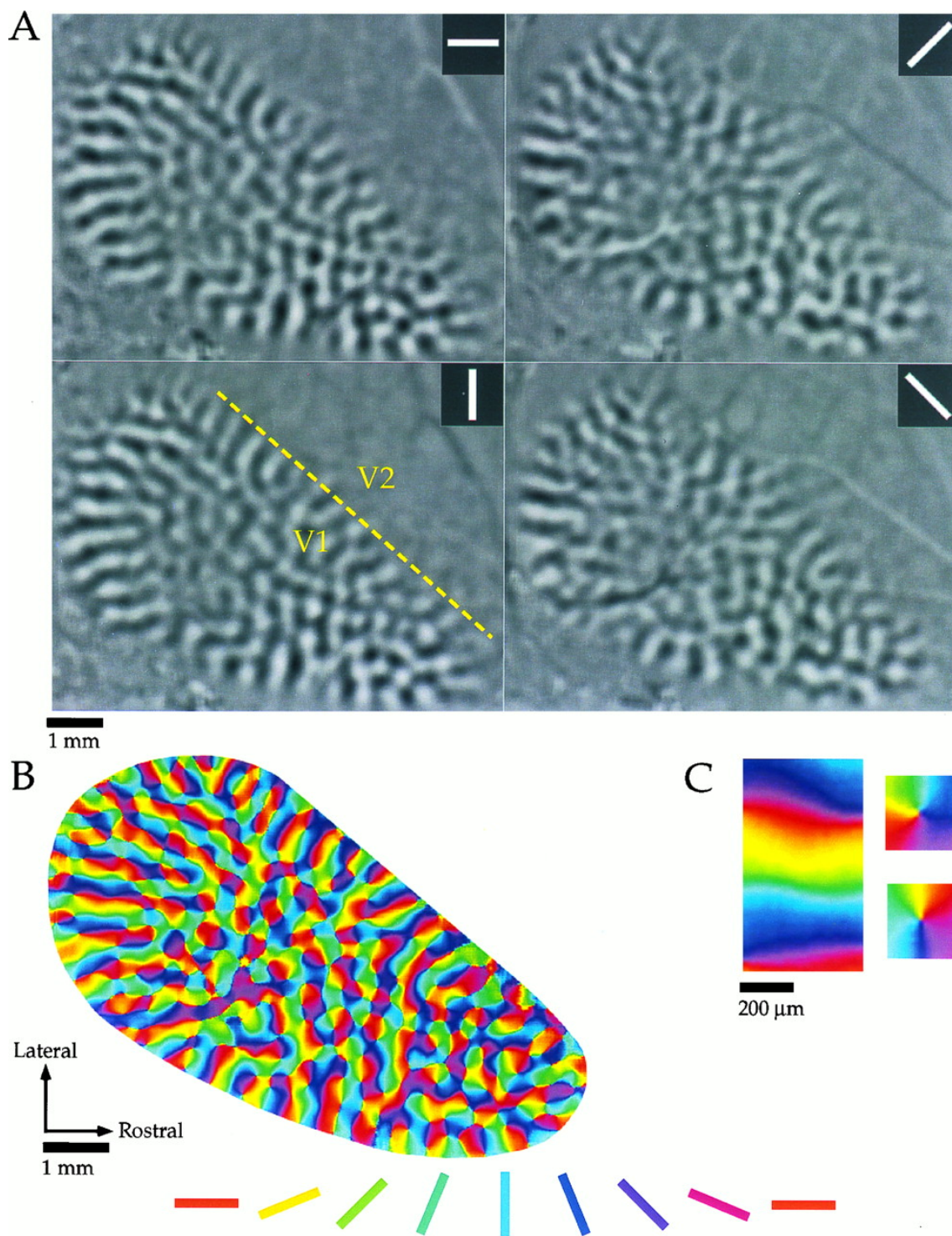


FIGURE 3. Les différentes orientations sont codées par des couleurs. À droite, zoom sur des exemples de points réguliers et de points singuliers de chiralités opposées. (D'après Bosking [71])

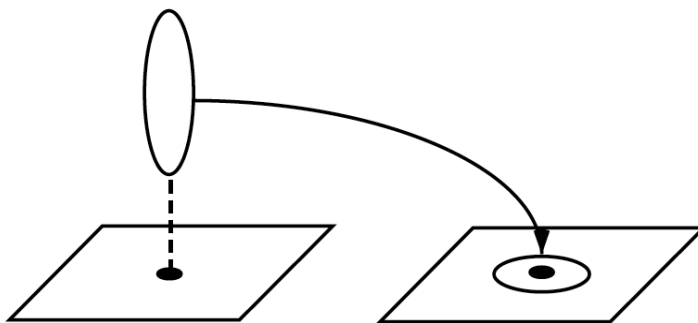


FIGURE 4. Une fibre P “contractée” et “rabattue” sur le plan base.

(iii) des points col au centre des mailles du réseau, points où les lignes d’iso-orientation bifurquent : deux lignes d’iso-orientations voisines partent du même point singulier mais aboutissent à deux points singuliers opposés.

Lorsque l’on pénètre le cortex “verticalement” en un point régulier, on retrouve les colonnes de Hubel et Wiesel avec leur redondance et leur codage par population. En revanche lorsque l’on pénètre en un point singulier on rencontre des neurones de toutes les orientations. P. E. Maldonado, I. Gödecke, C. M. Gray et T. Bonhöffer [346] ont analysé la structure fine des cartes d’orientation aux singularités. Ils ont constaté que

“orientation columns contain sharply tuned neurons of different orientation preference lying in close proximity”.

Autrement dit, la redondance columnaire disparaît aux points singuliers. James Schummers [492] a montré de même que

“neurons near pinwheel centers have subthreshold responses to all stimulus orientations but spike responses to only a narrow range of orientations”.

Cela indique qu’aux points singuliers a_i toutes les orientations sont en quelque sorte présentes. C’est comme si les pinwheels étaient des “éclatements” de points singuliers avec ce que l’on appelle en géométrie algébrique un “diviseur exceptionnel” égal à $\mathbb{P}_{a_i}^1$ comme fibre au-dessus du point de base a_i . L’image intuitive (cf. figure 4) est que la fibre exceptionnelle $\mathbb{P}_{a_i}^1$ est “contractée” et “rabattue” sur un petit voisinage de a_i . Il y a donc un collapse dimensionnel $3D \rightarrow 2D$ et une carte d’orientation est une structure géométrique de dimension “intermédiaire” entre 2 et 3.

Un réseau de pinwheels peut alors être conçu comme un recollement de tels éclatement locaux et fournit une approximation *discrète* de la projection $\pi : R \times \mathbb{P}^1 \rightarrow R$, cette dernière correspondant à une situation limite où, en quelque sorte,

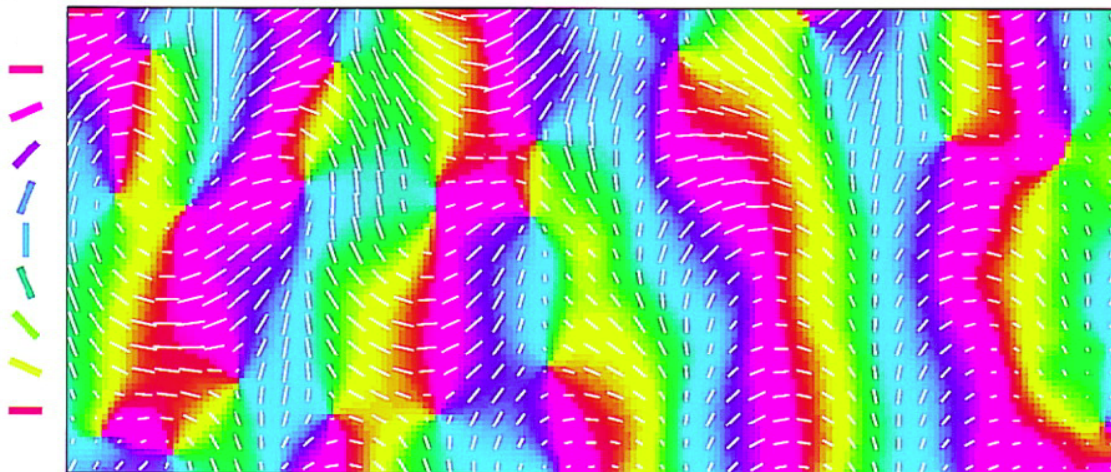


FIGURE 5. Cartes des orientations et des pinwheels du $V1$ d'un tupaya. On y observe la relation entre les pinwheels (couleurs) et les orientations préférentielles. (D'après Shmuel [500])

tous les points de R sont éclatés en parallèle. Les diviseurs exceptionnels deviennent alors les fibres de π .

Il est intéressant de regarder la façon dont varie l'orientation en relation avec la structure en pinwheels de $V1$. La figure 5 est due à Shmuel.

On voit que les pinwheels respectivement dextrogyres et lévogyres sont associés aux deux types de singularités génériques des champs de directions dans le plan. Cela est dû au fait que quand le rayon tourne autour du centre du pinwheel d'un angle θ , l'orientation associée tourne de $\theta/2$. Donc deux rayons diamétralement opposés correspondent à des orientations orthogonales.

(i) Si l'orientation p_θ associée avec le rayon d'angle θ est $p_\theta = \alpha + \theta/2$, les deux directions seront les mêmes pour $p_\theta = \alpha + \theta/2 = \theta$, c.a.d. pour $\theta = 2\alpha$. Comme α est défini modulo π , il n'y a qu'une solution et l'on obtient le modèle local de la figure 6 (end point).

(ii) Si l'orientation p_θ associée avec le rayon d'angle θ est $p_\theta = \alpha - \theta/2$, les deux directions seront les mêmes pour $p_\theta = \alpha - \theta/2 = \theta$, c.a.d. pour $\theta = 2\alpha/3$. Il y a trois solutions et l'on obtient le modèle local de la figure 7 (point triple).

1.4. Connexions horizontales et structure de contact

Nous avons longuement expliqué dans le Vol I que pour passer du local au global et, en particulier, intégrer des contours, il faut bien plus que la structure rétinotopique "verticale" précédente. Il faut aussi pouvoir comparer entre elles des orientations (et donc des fibres P_a et P_b) au-dessus de points différents a et b du plan base R . Il s'agit d'un problème de *transport parallèle* qui est implémenté à travers ce

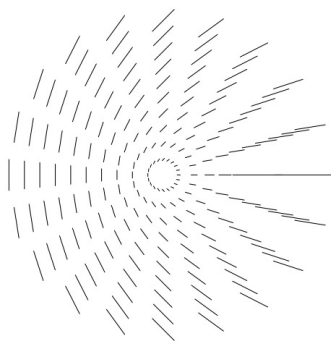


FIGURE 6. La singularité “end point”.

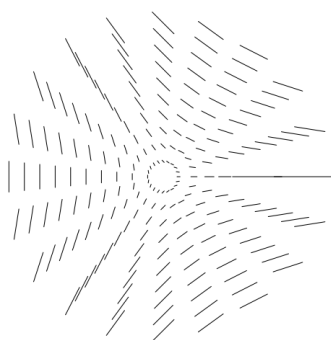


FIGURE 7. La singularité “point triple”.

que les neurophysiologistes appellent les connexions “horizontales” cortico-corticales, l’une des grandes découvertes expérimentales des années 1980.

Ces connexions horizontales sont à longue portée (jusqu’à 6-7mm) et leur propriété fondamentale est de relier des cellules de *même* orientation dans des hypercolonnes *éloignées*. Pour les détecter, on peut mesurer les corrélations entre cellules appartenant à des hypercolonnes différentes : on compare les orientations des cellules rencontrées lors d’une pénétration corticale avec celle d’une même cellule de référence ; en établissant des cross-corrélogrammes, on constate alors que les cellules d’orientations voisines sont fortement corrélées (existence d’un pic dans le corrélogramme) et seulement elles (cf. par exemple Ts’o, Gilbert, Wiesel [545]).

On peut aussi utiliser les méthodes d’imagerie optique *in vivo*. La figure 8, due à William Bosking [71] montre comment un marqueur (de la biocytine) injecté localement dans une zone de V1 d’orientation donnée (codée par du bleu-vert) diffuse le long des connexions horizontales. On constate que la diffusion à courte portée est isotrope alors que la diffusion à longue portée est au contraire hautement *anisotrope*

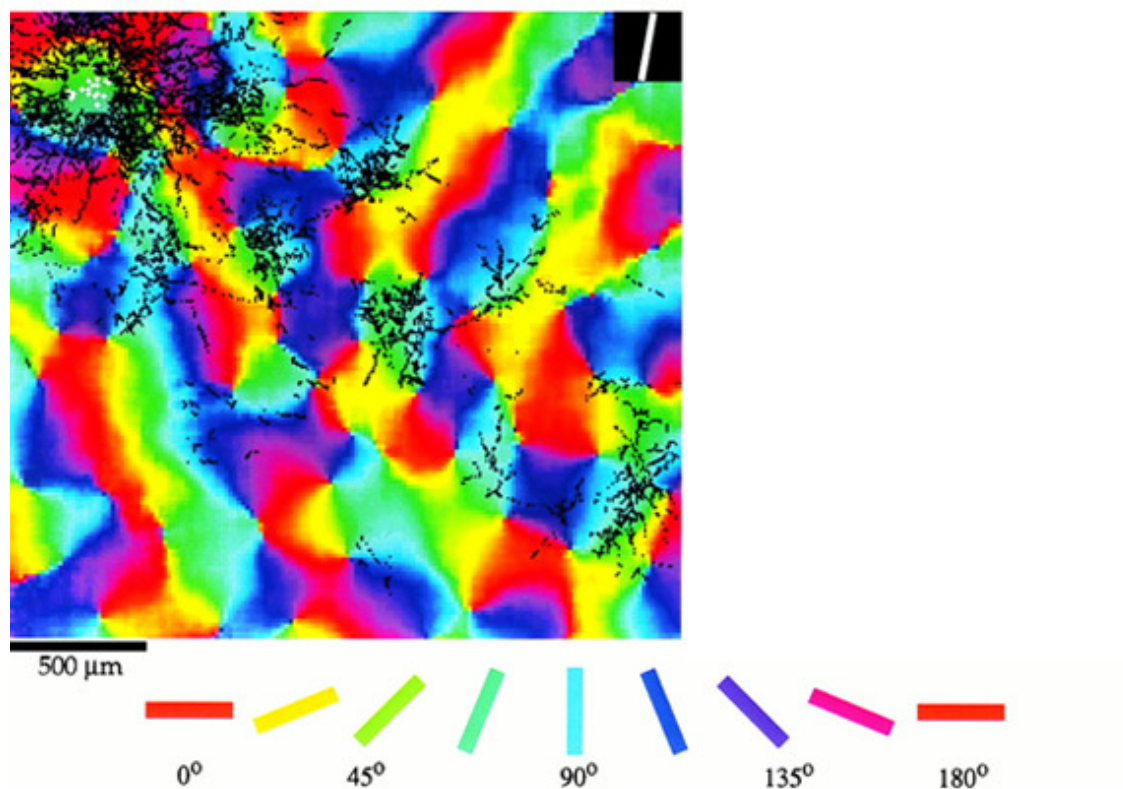


FIGURE 8. La diffusion du marqueur de la biocytine le long des connexions horizontales. (D’après Bosking [71]).

et restreinte à des domaines essentiellement de *même* orientation (de même couleur) que celle du site d’injection.

Qui plus est, on peut montrer que les connexions cortico-corticales connectent de façon préférentielle non seulement des éléments de contact parallèles (a, p) et (b, p) mais surtout des éléments coaxiaux, c’est-à-dire. des paires telles que p soit l’orientation de l’axe ab . En résumé, comme l’explique William Bosking [71] :

“The system of long-range horizontal connections can be summarized as preferentially linking neurons with co-oriented, co-axially aligned receptive fields”.

1.5. Intégration des contours et champ d’association

L’architecture fonctionnelle très caractéristique de V1 permet de comprendre les mécanismes corticaux d’intégration des contours. Cela est confirmé sur le plan psychophysique (et non pas neurophysiologique), par les expériences de David Field,

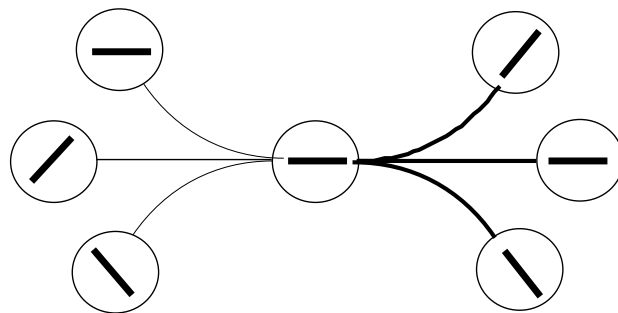


FIGURE 9. Schéma du champ d'association de Field, Hayes et Hess. Deux éléments de contact (a_1, p_1) et (a_2, p_2) sont connectés (lignes épaisses à droite) seulement s'il est possible d'interpoler entre les positions a_1 et a_2 au moyen d'une courbe γ de faible courbure, tangente à p_1 en a_1 et à p_2 en a_2 . Autrement les deux éléments ne peuvent pas être connectés (lignes fines à gauche).

Anthony Hayes et Robert Hess dans leur article de référence de 1993 (Field *et al.* [182]) où ils introduisent le concept d'“association field”. À partir de protocoles expérimentaux originaux ils montrent de façon précise que la reconnaissance d'un alignement dans une grille de patches de Gabor orientés aléatoirement vient d'un “pop out” d'une suite d'éléments (a_i, p_i) dont les centres a_i sont disposés le long d'un chemin lisse γ dont la courbure est faible et dont les orientations p_i sont tangentes à γ . La courbe γ interpole ainsi entre les (a_i, p_i) en *couplant* les positions et les orientations. Si une telle interpolation n'existe pas, les (a_i, p_i) ne sont pas connectés entre eux. La figure 9 schématise le champ d'association.

Le pop-out est dû au fait que, à cause des connexions horizontales, l'activation d'un neurone détectant un élément de contact (a, p) “pré-active” des neurones (b, q) avec b à peu près aligné avec a dans la direction p et q à peu près parallèle à p . On peut donc dire que les connexions horizontales cortico-corticales implément neuralemement le phénomène gestaltiste de “bonne continuation” (“Gesetz der guten Fortsetzung” ou “Gesetz der durchgehenden Linie”).

Jean Lorenceau a fourni une confirmation assez spectaculaire du lien entre les manifestations psychophysiques du champ d'association et les connexions cortico-corticales qui les implément neurophysiologiquement (cf. Georges *et al.* [209]). Il a utilisé pour ce faire une technique de vitesse apparente de séquences rapides d'éléments orientés. Cette vitesse apparente est plus rapide (sur-estimée) lorsque les éléments sont alignés avec la direction du mouvement et en revanche plus lente (sous-estimée) lorsque le mouvement se fait dans une direction orthogonale à celle des éléments. Qui plus est, dans le premier cas, l'augmentation de vitesse mesurée par des méthodes psychophysiques est la même que la vitesse de propagation de l'activation horizontale dans les connexions cortico-corticales mesurée électrophysiologiquement (de l'ordre de 0.2m/s). Ce résultat remarquable montre qu'il existe bien un effet

de facilitation et de pré-activation de certains neurones à travers les connexions horizontales.

En fait, pour bien comprendre les phénomènes de pop out perceptif il faut faire appel aux phénomènes dits de *liage* (de “binding”). L’idée – qui remonte à des travaux de Christoph von der Malsburg – est que la cohérence globale des constituants des percepts est la manifestation de la *synchronisation* de réponses neuronales. C’est ce qui se passe ici. Les connexions horizontales induisent une synchronisation qui rend perceptivement manifestes les effets de l’architecture fonctionnelle.

Field, Hayes et Hess [182] formulent très bien le mécanisme d’intégration du champ d’association :

“The points along the length of a curved edge can be linked together according to a set of local rules that allow the edge to be seen as a whole, even though different components of the edge are detected by independent mechanisms.” (p. 174)

Le point clé est la corrélation entre position et orientation :

“Elements are associated according to joint constraints of position and orientation.” (p. 187)

“There is a unique link between the relative positions of the elements and their relative orientations. (...) The orientation of the elements is locked to the orientation of the path.” (p. 181)

1.6. Un double vocabulaire

Dans une certaine mesure tout ce volume introduit à des formalismes qui permettent de faire se correspondre deux vocabulaires, l’un neurophysiologique et l’autre mathématique.

neurones simples	éléments de contact (a, p)
rétinotopie “rétine $\rightarrow V1$ ”	application conforme de \mathbb{R}^2 dans \mathbb{V}
variables de base / variables “engrafted”	fibration $\mathbb{V} : \mathbb{R}^2 \times P \rightarrow \mathbb{R}^2$
hypercolonnes d’orientation et pinwheels	espace des 1-jets $J^1 \not\subseteq \mathbb{R}^2 \times P \rightarrow \mathbb{R}^2$
connexions horizontales	structure de contact
– “co-oriented, co-axially aligned receptive fields” – “joint constraints on positions and orientations” – “good continuation”	relevées legendriennes

2. $V1$ comme fibration

Commençons maintenant notre périple.

Nous avons vu dans Vol I. 4.8. que l'on peut considérer l'organisation en pinwheels de $V1$ de trois façons complémentaires. Soit M la surface, identifiée à un plan pour simplifier, de la couche corticale de $V1$ considérée.

1. On peut d'abord traiter les centres des pinwheels comme un réseau de singularités discrètes a_i centres de petites roues d'orientation qui se recollent dans M . Dans le Vol I, nous avons longuement traité de ces modèles discrets dont nous venons de rappeler les grandes lignes.
2. Mais, comme nous venons de le voir avec la figure 9, on peut aussi, en utilisant la notion d'éclatement ("blow-up"), considérer chaque pinwheel centré en $a \in M$ comme le rabattement, dans un voisinage de a , de l'espace P des orientations, ce dernier étant considéré comme une "fibre" de dimension 1 "au-dessus" de a déployée dans une troisième dimension. Dans ce cas, le réseau discret de pinwheels s'interprète comme l'éclatement en parallèle des centres a_i .
3. On peut alors passer à la limite et considérer un *continuum* de pinwheels infiniment rapprochés éclatés en parallèle, ce qui conduit à une fibration $\pi : \mathbb{V} = M \times P \rightarrow M$. Les concepts de fibration et de section sont explicités dans Vol I. 4.3.

C'est dans cette dernière perspective, celle des modèles continus, que nous nous situerons ici. Nous traiterons donc $V1$ comme une fibration $\pi : \mathbb{V} = M \times P \rightarrow M$ en prenant en général pour variété de base M le plan euclidien \mathbb{R}^2 et pour fibre P par exemple la droite projective \mathbb{P}^1 des orientations du plan (identifiable au cercle \mathbb{S}^1 des angles θ modulo π), ou le cercle \mathbb{S}^1 des angles θ modulo 2π , ou encore la droite \mathbb{R} des $p = \tan(\theta)$.

Nous présenterons deux modèles de base faisant de $V1$ une implémentation d'une structure de contact tridimensionnelle au sens défini dans Vol I. 5.4. et repris plus bas dans la section 2.1. Le premier modèle, noté \mathbb{V}_J , repose sur la structure de contact canonique de l'espace des 1-jets et le second, noté \mathbb{V}_S , sur la structure de contact du groupe euclidien $SE(2)$ qui opère sur \mathbb{V}_J . Pour faire la transition, nous interpréterons d'abord le premier modèle en termes d'une structure de groupe sur \mathbb{V}_J isomorphe au groupe de Heisenberg. Mais commençons par quelques brèves remarques méthodologiques.

3. Méthodologie des modèles géométriques

Il existe plusieurs modèles fort intéressants de formation de pinwheels à partir de réseaux de neurones susceptibles d'apprentissage hebbien (cf. par exemple les propositions de Haim Sompolinsky, Michael Shelley et Kukjin Kang [278], cf. également le modèle de Mäkkäläinen-Sirosh [380]). Nous renvoyons à la partie 3 de *Neurogéométrie de la Vision* [435], en particulier au chapitre 8 "Réseaux de neurones

et champs d'oscillateurs", pour les liens avec la neurogéométrie. Mais nous allons nous focaliser ici plutôt sur des modèles purement géométriques mésoscopiques.

Nous adopterons trois principes.

- (i) Nous suivrons la méthodologie standard de toute modélisation :
 - simplifier les phénomènes empiriques ;
 - se focaliser sur un nombre restreint de propriétés pertinentes ;
 - construire un premier modèle sur ces bases ;
 - revenir aux données empiriques ;
 - complexifier le modèle de départ.
- (ii) Nous ne chercherons à élaborer que des premiers modèles élémentaires des phénomènes de base que sont l'intégration des contours et l'existence de contours illusoires, phénomènes évoqués dès l'*Introduction* du Vol I.
- (iii) Nous simplifierons drastiquement les données neurophysiologiques exposées en détail dans le Vol I. Les restrictions seront :
 - seulement le "minimal discharge field" ("spiking response") ;
 - pas d'échelle (résolution infinie) ;
 - seulement la variable d'orientation : pas de dominance oculaire, pas de phase spatiale, pas de fréquence spatiale ;
 - pas de contrôles moteurs : yeux, tête, mouvements du corps ;
 - pas de focalisation attentionnelle ;
 - seulement $V1$, pas d'autres aires comme $V2$, $V4$ ou $MT/V5$ avec des feedbacks descendants sur $V1$ et le LGN .

Mais, même avec ces simplifications drastiques, la reconstruction mathématique des phénomènes et l'explication géométrique des architectures fonctionnelles sont non triviales et utilisent, comme nous l'avons anticipé dans le Vol I, des concepts sophistiqués comme ceux de fibration, de transport parallèle, de structure de contact, de condition d'intégration de Frobenius, de géodésiques¹ sous-riemanniennes. Cela n'est pas aussi étonnant que cela pourrait paraître au prime abord dans la mesure où ces concepts ont été introduits par certains des plus grands géomètres après Riemann (Sophus Lie, Élie Cartan, entre autres) précisément pour comprendre comment une géométrie globale pouvait émerger d'une intégration cohérente de données locales infinitésimales.

4. Neurotopologie

Pour pouvoir présenter les modèles, nous avons besoin des concepts basiques d'espace topologique et de variété différentiable ainsi que d'une compréhension de la façon dont de telles structures géométriques peuvent être neuralemement implémentées.

1. Le terme de "géodésique" a été introduit par Laplace en relation avec les mesures sur le globe terrestre.

4.1. Le problème d'un espace immanent global

Le système visuel construit plusieurs représentations successives de l'espace visuel. Nous en avons étudié dans le Vol I la structure locale à travers les champs et profils récepteurs (CRs et PRs) ainsi que des éléments de structure globale à travers l'architecture fonctionnelle des aires visuelles primaires. Cette question du passage du local au global se révèle être particulièrement problématique. En effet, quand deux cellules ganglionnaires de la rétine sont assez voisines pour que leurs CRs se recouvrent, elles sont connectées à des photorécepteurs communs à travers des cellules intermédiaires (horizontales et amacrines) ce qui *coordonne* leur activité. Mais en revanche quand leurs CRs sont disjoints, leurs activités deviennent apparemment décorréélées. À ce niveau périphérique immanent il n'existerait donc pas, semble-t-il, de structure géométrique globale. Alors que l'espace euclidien externe possède une structure globale exprimée par son groupe d'isométries (qui opère transitivement et relie donc entre eux n'importe quels domaines locaux, aussi distants soient-ils l'un de l'autre), cela ne serait pas le cas pour l'espace visuel interne immanent qui ne serait défini que localement.

Mais il ne s'agit là que d'une apparence. En fait un passage du local au global se réalise au moyen d'opérations de *recollement* des domaines locaux que sont les CRs. En ce sens, au niveau immanent, la globalité de l'espace n'est pas donnée mais bien *constituée*.

La même situation se rencontre à chacun des niveaux successifs de représentation du champ visuel : deux cellules dont les CRs sont disjoints et qui ne sont reliées entre elles par aucune interconnexion risquent de s'ignorer l'une l'autre. Les connexions "verticales" à longue portée relient des niveaux successifs de traitement, mais s'il n'y avait pas de recouvrements partiels des CRs et si les connexions "horizontales" étaient toutes à courte portée, le passage du local au global serait impossible. Nous devons donc distinguer trois types de structures :

- (i) les structures purement locales,
- (ii) les structures semi-globales ou globales "faibles" obtenues par recollement de cartes locales : elles définissent des niveaux topologiques et différentiables,
- (iii) les structures globales "fortes" permettant des opérations comme le transport parallèle et la comparaison de ce qui se passe en des points distants.

Nous parlerons longuement des structures globales "fortes" dans la suite. Nous voudrions ici dire quelques mots sur l'implémentation neuronale des structures globales "faibles" topologiques et différentiables.

4.2. Un minimum de topologie

L'idée de passer du local au global par recollement est devenue fondamentale dans la géométrie moderne : depuis longtemps² les mathématiciens ont été intéressés par les propriétés des espaces obtenus en recollant des domaines locaux et ils ont généralisé les outils du calcul différentiel. Cela les a conduit au concept de variété différentiable définie sur un espace topologique.

Donnons quelques très brèves indications initiales sur cette notion. Dans la suite de l'ouvrage nous donnerons quelques précisions supplémentaires lorsque nous en aurons besoin. Intuitivement, un espace topologique est un ensemble M dont les éléments sont positionnés les uns par rapport aux autres par des rapports de proximité, c'est-à-dire un ensemble où l'on sait définir le fait qu'un point x est dans un voisinage d'un autre point y . Il y a plusieurs façons d'axiomatiser une telle notion. Une fois celle-ci définie, tout un ensemble d'autres notions en découlent immédiatement.

1. Un sous-ensemble $U \subset M$ est dit *ouvert* s'il est un voisinage de chacun de ses points.
2. Un sous-ensemble $F \subset M$ est dit *fermé* si son complémentaire $M - F$ est ouvert.
3. Un point $x \in M$ est dit *adhérent* à un sous-ensemble (quelconque) $A \subset M$ si tout voisinage de x contient au moins un élément de A . Un fermé contient tous ses points adhérents.
4. Si A est un sous-ensemble (quelconque) $A \subset M$, l'*intérieur* de A , noté $\overset{\circ}{A}$, est le plus grand ouvert contenu dans A (i.e. la réunion de tous les ouverts contenus dans A) et la *fermeture* de A (ou son *adhérence*), notée \bar{A} , est le plus petit fermé contenant A (i.e. l'intersection de tous les fermés contenant A). $\bar{A} - \overset{\circ}{A}$ s'appelle la *frontière* de A .
5. Les morphismes de structure entre espaces topologiques sont les applications *continues*. Si $f : M \rightarrow N$ est une application ensembliste entre espaces topologiques, elle est dite continue si l'image inverse $U = f^{-1}(V)$ de tout ouvert V de N est un ouvert U de M . Les bijections bicontinues entre deux espaces topologiques sont les isomorphismes de structure et sont appelées *homéomorphismes*.

Parmi les axiomatisations possibles rendant ces définitions cohérentes, celle des ouverts est la plus connue. Les ouverts U de M sont caractérisés par les axiomes :

2. L'idée remonte à Gauss au début du XIX^e siècle (\sim 1820-1830). Riemann introduisit ensuite le concept de variété (*Mannigfaltigkeit*) possédant une métrique dans son *Habilitationsarbeit* de 1854 "Über die Hypothesen, welche der Geometrie zu Grunde liegen" [464], puis Hermann Weyl introduisit le concept de variété différentiable abstraite au début du XX^e siècle (1913) dans son ouvrage *Die Idee der Riemannschen Fläche* [561].

- (i) M et \emptyset sont ouverts,
- (ii) une réunion *quelconque* d'ouverts est un ouvert,
- (iii) une intersection *finie* d'ouverts est un ouvert.

Un voisinage de $x \in M$ est alors un sous-ensemble V de M contenant non seulement x mais tout un ouvert U contenant x . Si l'on note $\mathcal{V}(x)$ l'ensemble des voisinages de x , on vérifie alors aussitôt que

- (i) si $V \in \mathcal{V}(x)$ et si $V \subset W$ alors $W \in \mathcal{V}(x)$ ($\mathcal{V}(x)$ est stable par extension),
- (ii) si $V, W \in \mathcal{V}(x)$ alors $V \cap W \in \mathcal{V}(x)$ ($\mathcal{V}(x)$ est stable par intersection finie),
- (iii) M est un voisinage de chacun de ses points et l'ensemble vide \emptyset n'est jamais un voisinage.

On peut donc axiomatiser la notion de voisinage en prenant ces propriétés comme axiomes.

Une façon intuitive de définir un ouvert est d'utiliser une notion de *stabilité*. Disons qu'une propriété $P(x)$ des éléments $x \in M$ est stable si lorsque $P(x_0)$ est satisfaite pour x_0 elle est satisfaite pour tous les éléments $y \in V$ d'un voisinage $V \in \mathcal{V}(x_0)$. Alors dire que $U \subset M$ est ouvert équivaut à dire que la propriété d'appartenance $x \in U$ est stable.

Les topologies susceptibles d'être définies sur un espace M peuvent être extrêmement variées car elles décrivent en quelque sorte des degrés de "cohésivité". À un extrême, on trouve la topologie "discrète" pour laquelle tous les éléments x de M sont à la fois ouverts et fermés et donc également tout sous-ensemble A de M . Il n'y a aucune cohésion et M est une poussière de points : la topologie n'impose aucune contrainte de cohésivité à l'ensemble sous-jacent. À l'autre extrême, on trouve la topologie "grossière" pour laquelle M est le seul ouvert non vide. La cohésion de M est maximale : tous les points ont le même degré de voisinage et sont unifiés en un bloc indécomposable. Mais en général il existe différents niveaux de cohésion dès que les éléments de M ont une structure interne, chaque niveau étant lui-même susceptible de degrés. Nous en rencontrerons de nombreux exemples. Considérons par exemple des fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$. Utilisons la métrique standard de \mathbb{R} pour mesurer la proximité de deux valeurs numériques. On pourra alors décider que f et g sont plus ou moins voisines suivant que leurs valeurs le sont soit localement soit globalement. Dans ce dernier cas, il faudra décider comment on évalue l'écart global entre f et g . On pourrait prendre la valeur de l'intégrale $\int_{\mathbb{R}} (f - g) dx$ mais ce n'est pas très contraignant car une intégrale $\int_{\mathbb{R}}$ peut être nulle tout en ayant de très grandes parties positives et négatives : il suffit que celles-ci s'équilibrent. On pourra alors utiliser $\int_{\mathbb{R}} |f - g| dx$ qui ne sera petite que si les valeurs de f et g sont partout assez voisines. On peut aussi utiliser des puissances $|f - g|^n$, etc. Dans le cas de l'espace fonctionnel des fonctions continues, on définit ainsi la topologie C^0 . Mais il

faut bien voir que celle-ci ne concerne que les valeurs des fonctions. Si l'on travaille dans l'espace fonctionnel des fonctions dérivables, g est voisine de f si ses valeurs sont incluses dans un petit voisinage autour des valeurs de f . Mais, à l'intérieur d'un tel voisinage ces valeurs de g peuvent varier comme elles le veulent et, par exemple, présenter des oscillations très fortes et très rapprochées par rapport à celles de f . On peut alors introduire la topologie C^1 beaucoup plus contraignante imposant que les dérivées f' et g' soient également voisines. Si f et g sont 2 fois différentiables, cela permettra toutefois à g d'avoir des ondulations très fortes et très rapprochées changeant constamment la courbure (phénomène de tôle ondulée). On pourra alors introduire la topologie C^2 encore plus contraignante imposant que les dérivées f'' et g'' définissant la courbure soient également voisines. Etc.

Remarque. Profitons de cette évocation des topologies C^0 et C^1 pour insister sur l'abîme qui sépare la simple continuité de la différentiabilité. Comme le montrent les *fractales*, des courbes continues peuvent ne posséder nulle part de tangentes. Il est très facile d'en construire. L'exemple le plus connu est la courbe de Koch. On part d'un segment $S = AE$, on le divise en trois segments AB , BD , DE et on remplace le tiers BD du milieu par une dent équilatérale BCD . Les trois points B , C , D sont devenus anguleux et la courbe n'y a plus de tangente. On a ainsi une façon de transformer un segment $S = AE$ en un pattern $P = ABCDE$. On part alors d'un segment K_0 et on itère le processus en transformant à chaque étape chacun des segments S de K_n en un pattern P . À la limite, la courbe K_∞ n'a plus que des points anguleux. □

Les morceaux d'un espace topologique M s'appellent ses *composantes connexes*. Ce sont les sous-ensembles qui sont à la fois ouverts *et* fermés. Un sous-ensemble $A \subset M$ est muni de la topologie "induite" qui est la trace sur A de celle de M (les ouverts de A sont les $A \cap U$, U ouvert de M). On dit alors que A est connexe s'il est sa seule composante connexe. La composante connexe de M contenant x est l'union des A connexes contenant x , c'est-à-dire le $A \ni x$ connexe maximal.

Les espaces topologiques généraux peuvent être extraordinairement compliqués, par exemple lorsqu'il s'agit d'espaces fonctionnels qui sont des espaces vectoriels de dimension infinie. Il peut exister des topologies spécifiques (comme la topologie de Zariski qui est la topologie de base en géométrie algébrique) avec des propriétés spéciales, comme par exemple ne pas séparer les points, sans être pour autant banalement "grossières". On dit qu'un espace topologique est *séparé* lorsque deux points différents x et y admettent des voisinages *disjoints*. La séparation signifie que la topologie localise suffisamment les points. Les topologies discrètes sont les topologies séparées les plus fines possibles.

Une notion importante est celle de "compacité" qui est une propriété de *finitude globale* d'un espace topologique : M (séparé) est dit *compact* si, de tout recouvrement

ouvert, on peut extraire un recouvrement *fini*. En particulier M est la réunion d'un nombre fini d'ouverts.

La compréhension des espaces topologiques devient beaucoup plus intuitive lorsque leur topologie est *métrisable*, i.e. induite par une structure métrique.³ En effet la relation de voisinage entre deux points x et y devient alors quantifiable et s'exprime simplement par le fait que la distance $d(x, y)$ est petite. Les intervalles ouverts et fermés de \mathbb{R} servent dans ce cas de modèles. On définit immédiatement les boules ouvertes de centre x et de rayon ε comme les sous-ensembles $B(x, \varepsilon) = \{y \mid d(x, y) < \varepsilon\}$ et les boules fermées par $\overline{B}(x, \varepsilon) = \{y \mid d(x, y) \leq \varepsilon\}$. Les ouverts sont les réunions quelconques de boules ouvertes et tout le reste s'ensuit. Par exemple, on voit immédiatement que \mathbb{R}^n n'est pas compact car il n'est pas la réunion d'un nombre fini de boules ouvertes. De même, on voit pourquoi une intersection quelconque d'ouverts ne peut pas être ouverte. En effet les points x sont fermés puisque $x = \overline{B}(x, 0)$ alors que $x = \bigcap_{\varepsilon > 0} B(x, \varepsilon)$ est une intersection (infinie) de boules ouvertes. On voit aussi pourquoi si $f : M \rightarrow N$ est une application continue il n'y a aucune raison pour que l'image d'un ouvert soit ouverte. Il suffit de prendre la surjection $M \rightarrow \{y\}$ avec y un point fermé non ouvert de N : M est ouvert alors que $\{y\}$ est fermé.

Ceci dit, les espaces métriques peuvent être eux-mêmes d'une complexité inouïe.

4.3. Topologie et Gestalttheorie : ouvert/fermé et fond/forme

La topologie est *constitutive* du champ visuel et se manifeste de façon spectaculaire dans certains mécanismes perceptifs subtils qui reposent sur la différence topologique entre ouvert et fermé. Tel est le cas de l'opposition fondamentale *fond / forme* (ou figure / fond) étudiée de façon très approfondie, au moins depuis Brentano, par les phénoménologues et les psychologues de la théorie de la Gestalt (fondée en 1912 par Max Wertheimer, Wolfgang Köhler et Kurt Koffka). Lorsqu'une figure se détache sur un fond servant d'arrière-plan, le bord de la figure appartient à la figure et c'est le processus "d'affectation de bord" ("edge-assignment") qui caractérise le détachement. Évidemment cela est dû au fait que les objets $3D$ sont mobiles dans l'espace et manipulables, leur perception étant un processus sensori-moteur et cognitif compliqué. De nombreux indices visuels permettent de détacher l'objet du fond : profondeur, mouvement, illumination, forme, taille, couleur, texture, etc. Mais, au niveau de l'image visuelle purement $2D$ la détection des contours et leur assignation faisant de l'objet un fermé topologique séparé du fond qu'est l'ouvert complémentaire est un phénomène de saillance fondamentale. Il peut induire pour des

3. Rappelons qu'un espace métrique est un ensemble dans lequel on sait mesurer la distance $d(x, y)$ entre deux points (x, y) , $d(x, y)$ étant un nombre réel ≥ 0 satisfaisant (i) $d(x, y) = 0$ si et seulement si $x = y$ (axiome de séparation), (ii) $d(x, y) = d(y, x)$ (axiome de symétrie) et (iii) $d(x, z) \leq d(x, y) + d(y, z)$ (inégalité triangulaire).



FIGURE 10. Exemple d'illusion de Rubin. (Phæno Science Center, Wolfsburg).

images $2D$ artefactuelles bien construites la célèbre bimodalité “vase/profils” d’Edgar Rubin. Si le fond est lui aussi une forme reconnaissable, l’assignation de bord au fond transforme ce dernier en un objet qui se détache maintenant sur un fond qui est l’objet de départ décomplété de son bord. La figure 10 en donne un exemple artistique exposé au “Phæno Science Center” de Wolfsburg (2005, architecte Zaha Hadid). On y voit quatre objets noirs sur un fond beige (deux paires entrelacées de sortes de pièces d’échec) ; mais l’inversion figure / fond fait apparaître deux couples de personnages face à face à la tête légèrement inclinée.

M.C. (Maurits Cornelis) Escher s’est beaucoup amusé avec ce phénomène en construisant des inversions progressives figure / fond. Un des exemples les plus connus est son dessin “Métamorphose : l’air et l’eau” de la figure 11 où des poissons se métamorphosent en oiseaux.

On a étudié très à fond cette bimodalité et les instabilités qui l’engendrent (par exemple la série temporelle des inversions figure / fond). On connaît la période critique d’apparition de cette faculté (“skill”) visuo-cognitive chez les enfants (entre 3 et 5 ans) qui permet de se *focaliser* sur des objets. Ce qui est essentiel est que le processus neuro-topologique de bas niveau d’assignation des bords est inséparable du processus neuro-cognitif de plus haut niveau de reconnaissance des formes. Si le fond n’est pas reconnaissable comme une forme répertoriée dans une mémoire de formes labellisées, l’inversion figure / fond ne se produit plus. Et quand les deux domaines complémentaires du champ visuel sont tous deux des formes reconnaissables, la reconnaissance de l’une inhibe la reconnaissance de l’autre.

Au niveau des corrélats neuronaux, le processus de bas niveau de détection de bords effectué par les aires primaires rétinotopiques est donc associé à un processus visuo-cognitif de reconnaissance des formes qui est post-rétinotopique et fait intervenir en particulier le cortex latéral occipital (LOC) situé entre l’aire V_3 et l’aire MT/V_5 (cf. Vol I, figures 4.3 et 4.5). Le LOC construit des représentations

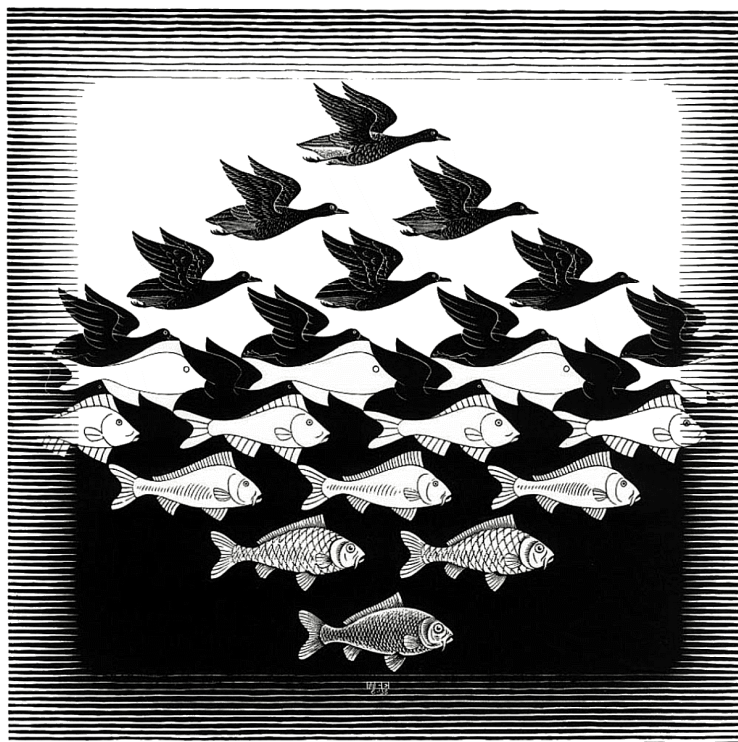


FIGURE 11. “Métamorphose : l’air et l’eau” : des poissons se transforment en oiseaux par inversion progressive figure / fond. (Xylogravure de M.C. Escher, 1938).

globales d’objets qui sont largement indépendantes de l’infinité des images $2D$ possibles de l’objet (taille, couleur, illumination, contours apparents, points de vue, postures des corps articulés, etc.). Il possède donc de très fortes capacités d’invariance. Mais il reste malgré tout très sensible à l’attribution de contours puisque lorsque le système visuel traite une illusion de Rubin vase / visages, la partie du LOC traitant le vase (resp. les visages) n’est active que lorsque le sujet dit qu’il voit un vase (resp. des visages). Le lecteur intéressé pourra consulter par exemple l’article [223] de Grill-Spector, Kourtzi et Kanwisher qui fait le point sur la question en 2001 avec les techniques de *IRMf* (imagerie par résonance magnétique fonctionnelle). Aujourd’hui les résultats sont raffinés par l’utilisation de techniques d’*ISPIf* (imagerie spectroscopique proche infrarouge fonctionnelle).

4.4. L'implémentation neuronale de la topologie et son immanence

Revenons au niveau le plus primitif du champ visuel, celui de la rétine. Comment comprendre sa topologie globale “immanente” *de l'intérieur*⁴ ? Les photorécepteurs pixelisent l'input lumineux, ce qui fait qu'il n'y a pas de continuum sous-jacent ultime. Mais l'on peut considérer que les champs récepteurs (CRs) des cellules ganglionnaires, puis des neurones du corps genouillé latéral, puis des neurones des différentes aires corticales visuelles, sont de petites cartes locales (mais de plus en plus grosses), des “points épais” perceptifs (de plus en plus épais) qui sont définis à une certaine méso-échelle et qui, par rapport aux photorécepteurs pixelisant, forment des petits voisinages *câblés*. On peut alors les traiter comme des petits ouverts locaux connexes et convexes recollant des points idéaux qui correspondraient à une résolution infinie. Ces CRs locaux ouverts “point-like” des cellules ganglionnaires sont indécomposables dans la mesure où le nerf optique est constitué des axones de ces dernières. Ils effectuent en quelque sorte un “cut-off” de la résolution.

Comment dès lors définir sur cette base la topologie globale du champ visuel d'une façon qui soit neuralelement pertinente “de l'intérieur”. Dès les années 1980, Jan Koenderink [297] a montré que cela est possible en utilisant les *relations fonctionnelles* auxquelles le système neuronal a accès de façon immanente. Le recollement des CRs se recouvrant partiellement est codé par les *corrélations* temporelles des signaux circulant le long des fibres nerveuses. Ces corrélations *temporelles*, et en particulier la synchronisation, sont la trace des recollements *spatiaux* des CRs. Leur information temporelle fine – la seule accessible au système de façon immanente – encode en partie la structure topologique constituée par recollement. En effet, la structure d'intersection de tout recouvrement d'un espace topologique M par des ouverts caractérise dans certains cas, nous allons le voir, son “homotopie” et même, lorsque l'on sait définir le fait que les ouverts sont convexes⁵, des contraintes dimensionnelles. Cela permet par conséquent de reconstruire en partie sa topologie. L'immanence de la topologie consiste à la reconstruire comme un *problème inverse* à partir des corrélations manifestées par le flux de l'activité neuronale.

Plus précisément, soit $\mathcal{U} = (U_i)_{i \in I}$ un recouvrement ouvert d'un espace topologique M .⁶ On lui associe une structure combinatoire $\mathcal{N}(\mathcal{U})$, appelée le “nerf” de \mathcal{U} , en considérant les intersections $U_i \cap U_j$, $U_i \cap U_j \cap U_k$, etc., qui sont *non vides* (cf. figure 12).

4. Il s'agit d'un point méthodologique crucial. Le fait que nous observions de l'extérieur en laboratoire des rétines et des couches corticales qui sont des continus de dimension 2 n'explique rien puisque cela résulte de propriétés immanentes de la perception qu'il s'agit justement d'expliquer.

5. $U \subset M$ est convexe si, pour tout $x, y \in U$, le segment $[x, y]$ est entièrement contenu dans U . Il faut donc que la notion de segment ait un sens dans M comme c'est le cas dans \mathbb{R}^n . La convexité implique la connexité.

6. Dans les cas qui nous intéressent, on supposera M et les U_i connexes.

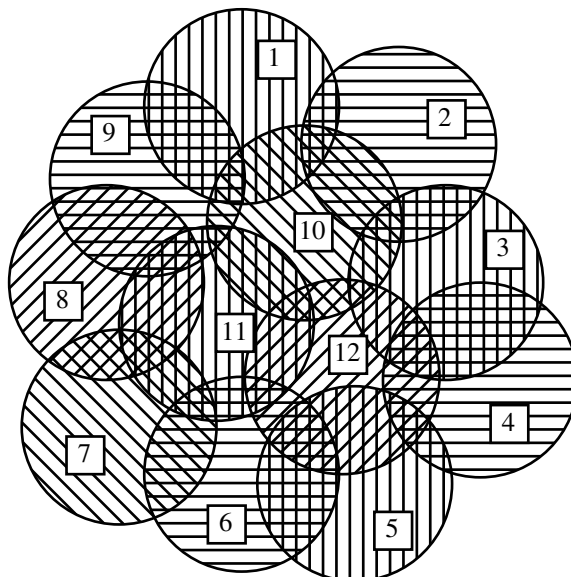


FIGURE 12. Le recouvrement d'un espace topologique par un ensemble d'ouverts. Le nerf du recouvrement est le complexe simplicial défini par les intersections non vides.

Le nerf $\mathcal{N}(\mathcal{U})$ est ce que l'on appelle techniquement un *complexe simplicial*. Pour simplifier et rester très intuitif, plaçons-nous dans le cas fini, $I = \{1, \dots, N\}$. Un complexe simplicial (CS) est alors simplement un ensemble Ξ de sous-ensembles de I tel que si $\sigma \in \Xi$ et $\tau \subseteq \sigma$ alors $\tau \in \Xi$. Si $|\sigma| = s$ est le cardinal de σ , σ est appelée une face de dimension $s - 1$. La condition est donc que toutes les sous-faces des faces de Ξ sont dans Ξ . Si $I \in \Xi$, Ξ est appelé un *simplexe* de dimension $N - 1$. Si l'on utilise une intuition géométrique on peut dire que $\sigma = \{i\} \in \Xi$, est un "point" (un "sommets", un "vertex" de dimension $1 - 1 = 0$), que $\sigma = \{i, j \neq i\} \in \Xi$ est un "segment" (une "arête" de dimension $2 - 1 = 1$) dont les extrémités (de dimension 0) $\{i\}, \{j\} \in \Xi$, que $\sigma = \{i, j \neq i, k \neq i, j\} \in \Xi$ est une face (un "triangle" de dimension $3 - 1 = 2$) dont les côtés (les arêtes de dimension 1) $\{i, j\}, \{j, k\}, \{k, i\} \in \Xi$ ainsi que les sommets $\{i\}, \{j\}, \{k\}$, etc. La dimension de $\mathcal{N}(\mathcal{U})$ peut donc être plus grande que la dimension de M s'il existe des intersections non vides de plus de $d + 1$ ouverts U_i .

Les complexes simpliciaux peuvent être très compliqués. Ils peuvent être composés d'un très grand nombre de faces et recoller des faces de dimensions très différentes.⁷ On peut par exemple rattacher un triangle à un tétraèdre par une de ses arêtes et y adjoindre une autre arête par un de ses sommets et aussi un point isolé. Un CS peut contenir toutes les faces de dimension $< d$ d'un d -simplexe sans contenir son intérieur, etc.

7. Les méthodes de triangulation de toute sorte de surfaces compliquées (visages, avions, outils, etc.) en CAO ont rendu ces CSs familiers.

Dans la figure 12, il y a 12 ouverts U_i . Pour les $U_i \cap U_j$ la matrice d'intersection est la suivante (1 si $U_i \cap U_j \neq \emptyset$, 0 si $U_i \cap U_j = \emptyset$) :

	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	0	0	0	0	0	0	1	1	0	0
2	1	1	1	0	0	0	0	0	0	1	0	0
3	0	1	1	1	0	0	0	0	0	1	0	1
4	0	0	1	1	1	0	0	0	0	0	0	1
5	0	0	0	1	1	1	0	0	0	0	0	1
6	0	0	0	0	1	1	1	0	0	0	1	1
7	0	0	0	0	0	1	1	1	0	0	1	0
8	0	0	0	0	0	0	1	1	1	0	1	0
9	1	0	0	0	0	0	0	1	1	1	1	0
10	1	1	1	0	0	0	0	0	1	1	1	1
11	0	0	0	0	0	1	1	1	1	1	1	1
12	0	0	1	1	1	1	0	0	0	1	1	1

Pour les intersections de 3 ouverts $U_i \cap U_j \cap U_k$ on obtient 13 triplets : (1, 2, 10), (1, 10, 9), (2, 3, 10), (3, 4, 12), (3, 10, 12), (4, 5, 12), (5, 6, 12), (6, 11, 12), (6, 7, 11), (7, 8, 11), (8, 9, 11), (9, 10, 11), (10, 11, 12). Toutes les intersections à 4 ouverts sont vides. La figure 13 montre le complexe simplicial $\mathcal{N}(\mathcal{U})$ associé au recouvrement \mathcal{U} de $M = \bigcup_{i \in I} U_i$ de la figure 12. On voit qu'il constitue bien une triangulation de dimension 2, ce qui est une façon de dire que la combinatoire de $\mathcal{N}(\mathcal{U})$ comme graphe est "planaire" i.e. réalisable par des ouverts *convexes* d'une surface (ici le domaine M du plan). Comme complexe simplicial, $\mathcal{N}(\mathcal{U})$ est un recollement "plan" de triangles dont les arêtes ne se rencontrent qu'aux sommets et ne se croisent pas en d'autres points.

La figure 14 (a) montre en revanche un recouvrement de 4 ouverts avec des intersections non vides de 4 ouverts puisque chaque ouvert rencontre les 3 autres. On voit que le complexe simplicial n'est plus associé à une triangulation plane mais à la projection plane des arêtes d'un tétraèdre (le simplexe de dimension 3). Cette projection est le graphe complet sur 4 sommets et comporte des croisements d'arêtes (figure (b)). Bien sûr, il suffit de déformer une arête pour retrouver un graphe planaire, mais ce n'est plus une triangulation (figure (c)). Cela montre que ce recouvrement en quelque sorte redondant d'un domaine du plan est combinatoirement de dimension 3.

Dans la mesure où $\mathcal{N}(\mathcal{U})$ caractérise dans une certaine mesure des propriétés topologiques de M , on peut développer à partir de ces intuitions élémentaires une théorie géométrico-combinatoire du "codage neuronal" qui précise l'hypothèse de Koenderink, à savoir que, dans le cas où M est le champ visuel et où \mathcal{U} est le recouvrement de M par les CRs des cellules ganglionnaires, le nerf $\mathcal{N}(\mathcal{U})$ se trouve

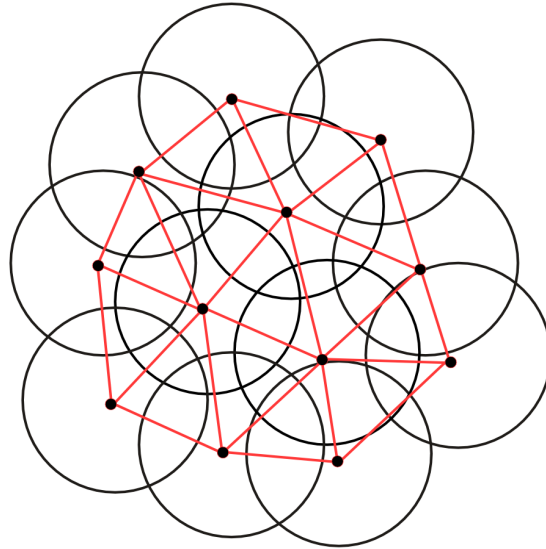


FIGURE 13. Le nerf $\mathcal{N}(\mathcal{U})$ du recouvrement \mathcal{U} de la figure 12. Ce complexe simplicial est associé à une triangulation de dimension 2, c'est-à-dire un recollement "plan" de triangles dont les arêtes ne se rencontrent qu'aux sommets et ne se croisent pas en d'autres points.

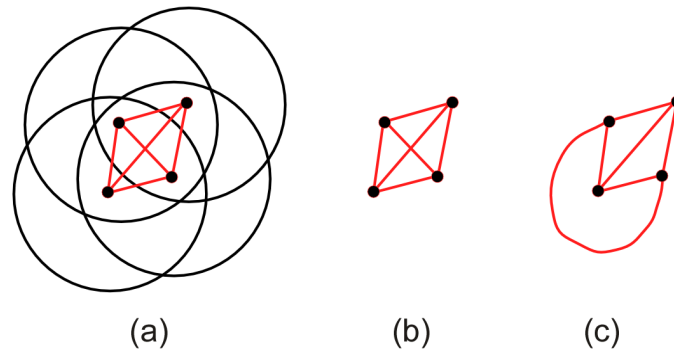


FIGURE 14. (a) Un recouvrement de 4 ouverts où chaque ouvert rencontre les 3 autres. Le complexe simplicial est le simplexe de dimension 3. (b) Sa représentation 2D naturelle est la projection d'un tétraèdre et comporte des croisements d'arêtes. (c) En déformant une arête on retrouve un graphe planaire, mais ce n'est plus une triangulation.

encodé dans les corrélations temporelles des signaux se propageant dans les fibres du nerf optique.⁸ Avec ses collaborateurs, Koenderink a montré comment une telle

8. Le terme "nerf" est un nouvel exemple (amusant) d'ambiguïté terminologique entre géométrie et neurosciences.

structure peut s'établir lors du développement du système visuel (cf. Toet *et al.* [536]).

Une idée, introduite par Carina Curto, Vladimir Itskov et leurs collaborateurs Alan Veliz-Cuba et Nora Youngs [129], est d'associer à un recouvrement \mathcal{U} un "code neural" purement binaire $\mathcal{C}(\mathcal{U})$ et de chercher ensuite à résoudre le *problème inverse* de la relation donnant $\mathcal{C}(\mathcal{U})$ à partir de \mathcal{U} . Ce problème inverse est, étant donné un code \mathcal{C} dont les éléments sont des patterns de co-activations synchronisées de neurones, de savoir dans quelle mesure l'on peut reconstruire à partir de \mathcal{C} un système de CRs \mathcal{U} tel que $\mathcal{C} = \mathcal{C}(\mathcal{U})$.

"What can be learned about the underlying stimulus space from knowledge only of \mathcal{C} , and not of \mathcal{U} ? The answer to this question will depend critically on whether or not we can assume that the RF [receptive fields] code is convex." ([129], p.6)

Ce problème inverse généralise bien celui de l'immanence de la topologie rétinienne $2D$ à l'activité neuronale décrite dans un espace de dimension considérable. Ces auteurs parlent aussi de géométrie "latent" ou "hidden".⁹

Remarque 1. On peut intuitionner la situation concrète de la façon simple suivante pour la voie rétino-géniculo-corticale. Les stimuli sont des impulsions lumineuses de Dirac δ_a envoyées aux différentes positions a sur la rétine. Toutes les cellules ganglionnaires (CGs) sont considérées comme ayant un CR convexe purement ON. Toutes les CGs tels que a soit dans leur CR sont activées en parallèle et leurs activations sont transmises de façon synchronisée au cortex par les fibres du nerf optique. Il s'agit de comprendre jusqu'à quel point on peut reconstruire la géométrie du champ visuel comme domaine de \mathbb{R}^2 à partir de la synchronisation temporelle des activités corticales. Cette géométrie est *intrinsèquement* encodée dans cette activité. Il faut bien comprendre qu'il ne s'agit pas tant de considérer l'espace fonctionnel des activités A codant la quasi infinité des images rétinienne possibles (les images binaires pixelisées sont des juxtapositions de δ_a correspondant à leur fonction caractéristique). mais de comprendre de façon immanente comment ces activités sont des fonctions $A(a)$ sur un espace de base M de *faible* dimension et quelle est la géométrie de cet espace M sur lequel ces images sont définies. Pour V1 dont les neurones simples détectent des éléments de contact (a, p) d'une position rétinienne

9. Au niveau de l'histoire de la pensée, il est fascinant de voir comment des problèmes dont l'histoire métaphysique est très lourde peuvent être scientifiquement réactivés sans aucune référence philosophique alors même qu'elles donnent pourtant raison à des thèses qui ont été farouchement critiquées. Le problème de l'immanence d'une "background structure" géométrique qui soit implicite et latente dans les activités perceptives est celui de "l'intuition pure" kantienne et donc de l'esthétique transcendantale et du synthétique a priori de Kant à Husserl. Dans le domaine perceptif, le transcendantal ne signifie rien d'autre que l'existence d'architectures fonctionnelles neuronales.

a et d'une orientation p en a , c'est la fibration $\mathbb{R}^2 \times \mathbb{P}^1 \rightarrow \mathbb{R}^2$ qui compte. C'est la *dualité spectrale* fondamentale entre un espace de base et l'algèbre des fonctions définies sur lui qui ouvre une voie de solution de ce problème inverse. Il s'agit là d'une dualité cruciale sur laquelle nous reviendrons plusieurs fois (dès la section 6.3), et qui opère au cœur de la conceptualité géométrique moderne. \square

Remarque 2. On devrait tenir compte des trains de spikes de chaque neurone, c'est-à-dire considérer le décours temporel du réseau et à chaque instant d'un temps discrétisé considérer les co-activations neuronales. Nous y reviendrons plus bas. Mais dans un premier temps on considèrera d'abord, pour simplifier, les co-activations possibles à un moment donné de façon à construire, comme en physique, un "espace de configurations" ou un "espace des phases" définissant la "cinématique" du réseau. La géométrie-topologie-homotopie de cet espace fonctionnera comme une "background structure" synthétique a priori (nous reviendrons longuement sur ce point historique et épistémologique crucial dans le chapitre 4, section 1). On reviendra ensuite dans un second temps sur les dynamiques possibles du réseau. \square

Remarque 3. Les taux de décharge (firing rates) des neurones permettent de coder des valeurs numériques. Toujours pour simplifier et se focaliser sur le cœur du problème, on se bornera dans un premier temps à des codages binaires qui se ramènent à l'alternative actif/non actif et mettent donc entre parenthèses les taux de décharge. \square

Soit N le nombre de neurones. On identifie chaque neurone à son label $i \in \{1, \dots, N\}$ et on suppose pour simplifier qu'il est susceptible de deux états $0 =$ non actif et $1 =$ actif ou "actif" signifie que le stimulus externe rencontre son CR U_i .¹⁰ Un N -vecteur $v = \{v_i\} \in \{0, 1\}^N$ représente donc un pattern d'activité. Le vecteur v_0 constitué uniquement de 0 constitue "l'état de base" (non activé par des stimuli externes) par rapport auquel l'activité induite par des stimuli externes est mesurée. Évidemment, pour les réseaux de neurones biologiques, cette non-activation ne signifie pas la non-activité absolue qu'est la mort cérébrale. Le réseau a toujours une activité interne endogène.¹¹

Le code $\mathcal{C}(\mathcal{U})$ associé à \mathcal{U} est un ensemble de tels patterns qui reflète de façon "appropriée" les relations d'intersections des U_i . Explicitons ce que signifie ici "appropriée". Soit $U_i \in \mathcal{U}$. On voudrait que le vecteur $v_{\{i\}} = \{1_i\}$ composé de 0 avec seulement un 1 en i -ème position, vecteur qui correspond à l'activité de seulement

10. Il s'agit d'une grande simplification. Traiter le CR comme un ouvert (y compris un ouvert convexe) du plan rétinien ne fait pas problème car les profils récepteurs sont des dérivées de gaussiennes ou des fonctions de Gabor qui sont à décroissance rapide au bord. Considérer en revanche que le CR est actif dès qu'un stimulus le rencontre ne tient pas compte de sa décomposition en zones ON (activatrices) et OFF (inhibitrices) expliquées Vol I, chapitre 3. Physiologiquement, l'inhibition est aussi importante que l'activation et il faudrait considérer des codes ternaires à valeur $\{-1, 0, 1\}$ plutôt que des codes binaires.

11. Dans le cas de la vision il s'agit de l'état du cortex visuel quand le sujet est dans le noir.

U_i , soit admissible. Mais on veut aussi éviter deux choses. Soit que U_i soit disjoint de l'union des autres U_j car cela romprait la connexité du recouvrement, soit que U_i soit inclus dans l'union des autres U_j car cela empêcherait U_i d'être actif alors que tous les autres U_j seraient non actifs. On n'admettra donc le vecteur $v_{\{i\}}$ que si $U_i \setminus \bigcup_{j \neq i} U_j \neq \emptyset$.¹² De même, considérons deux ouverts s'intersectant $U_i \cap U_j \neq \emptyset$ ($j \neq i$, $\{i, j\} \in \mathcal{N}(\mathcal{U})$). Le vecteur $v_{\{i, j\}}$ constitué de 1 aux positions i et j et de 0 partout ailleurs sera admis si l'intersection $U_i \cap U_j$ n'est ni vide ni incluse dans l'union $\bigcup_{k \neq i, j} U_k$ des autres U_k de façon à permettre à U_i et U_j d'être actifs alors que tous les autres U_k sont non actifs. De façon générale, si $v \in \{0, 1\}^N$ est un N -vecteur, soit 1_v (resp. 0_v) l'ensemble des i tels que $v_i = 1$ (resp. $v_i = 0$). On admettra v dans le code $\mathcal{C}(\mathcal{U})$ si et seulement si $\bigcap_{i \in 1_v} U_i \neq \emptyset$ et si de plus la condition (C) est satisfaite :

$$\left(\bigcap_{i \in 1_v} U_i \right) \setminus \left(\bigcup_{k \in 0_v} U_k \right) \neq \emptyset. \quad ((C))$$

On veut donc éviter les “mauvaises” relations d'inclusion $\bigcap_{i \in \sigma} U_i \subseteq \bigcup_{k \in \tau = I - \sigma} U_k$. On peut généraliser aux relations $\bigcap_{i \in \sigma} U_i \subseteq \bigcup_{k \in \tau} U_k$ pour $\sigma \cap \tau = \emptyset$. Supposons par exemple que $\mathcal{C} = \{(1, 0, 1), (0, 1, 1)\}$ soit convexe. Par hypothèse, $U_1 \cap U_3 \neq \emptyset$ et $U_2 \cap U_3 \neq \emptyset$. Si on avait aussi $U_1 \cap U_2 \neq \emptyset$ alors on aurait $U_1 \cap U_2 \cap U_3 \neq \emptyset$ et donc $(1, 1, 1) \in \mathcal{C}$ ce qui n'est pas le cas. Donc $U_1 \cap U_2 = \emptyset$. Mais comme $U_3 = (0, 0, 1) \notin \mathcal{C}$, il faut $U_3 \setminus (U_1 \cup U_2) = \emptyset$ et donc $U_3 = U_1 \cup U_2$. Mais alors U_3 n'étant pas connexe (puisque l'union de deux ouverts disjoints), il ne peut pas être convexe.

On notera que, d'après la condition (C), le vecteur nul $v_0 = \{0, \dots, 0\}$ qui correspond à l'absence complète d'activité ne peut être un vecteur du code que si $\bigcup_{k \in I} U_k \subsetneq M$. En effet, si $\bigcup_{k \in I} U_k = M$, alors $\bigcap_{i \in 1_v} U_i = \bigcap_{i \in \emptyset} U_i = M$ et $\bigcup_{k \in 0_v} U_k = \bigcup_{k \in I} U_k = M$ et donc, puisque $M \setminus M = \emptyset$, la condition (C) n'est pas satisfaite. Mais on peut aussi poser par convention que le vecteur nul appartient toujours au code même s'il ne satisfait pas (C). L'adjonction systématique de v_0 fait du code un code “pointé” avec v_0 comme élément distingué.

La condition (C) est plus forte que $\bigcap_{i \in 1_v} U_i \neq \emptyset$, c'est-à-dire $1_v \in \mathcal{N}(\mathcal{U})$ car elle reflète des contraintes plus globales sur les recouvrements partiels. On notera que $\mathcal{C}(\mathcal{U})$ est quand même par définition “complet” pour la condition (C), i.e. si v satisfait (C) alors $v \in \mathcal{C}(\mathcal{U})$. Par construction, on aura toujours $\mathcal{C}(\mathcal{U}) \subseteq \mathcal{N}(\mathcal{U})$ mais on aura aussi en général $\mathcal{C}(\mathcal{U}) \subsetneq \mathcal{N}(\mathcal{U})$. Toutefois, on peut montrer le résultat suivant (on suppose $N > \dim M$) :

Théorème. Si \mathcal{U} est *convexe*, $\mathcal{N}(\mathcal{U})$ est le complexe simplicial engendré par $\mathcal{C}(\mathcal{U})$. ◇

12. Si A et B sont deux ensembles, $A \setminus B$ dénote le complément de B dans A .

Si $\mathcal{C} = \mathcal{C}(\mathcal{U}) \subsetneq \mathcal{N}(\mathcal{U})$, on peut considérer pour tout $v \in \mathcal{C}$ l'ensemble $\Delta(v)$ des $w \in \mathcal{C}$ tels que $1_w \subset 1_v$. Ce n'est pas forcément l'ensemble des parties de 1_v puisque \mathcal{C} n'est pas un complexe simplicial.

Dans l'exemple de la figure 13, les v de $\mathcal{C}(\mathcal{U})$ correspondent aux U_i et aux $U_i \cap U_j \neq \emptyset$ ($j \neq i$). Les $\{i, j, k\}$ des intersections à 3 $U_i \cap U_j \cap U_k \neq \emptyset$ (i, j, k différents) sont dans $\mathcal{N}(\mathcal{U})$ mais ne sont pas dans $\mathcal{C}(\mathcal{U})$ car toutes les intersections à 4 sont vides.

Le problème inverse, celui de l'immanence, est alors le suivant. On se donne un ensemble \mathcal{C} de N -vecteurs $v = \{v_i\} \in \{0, 1\}^N$ considéré comme les patterns de co-activation de N neurones dans un réseau et on suppose qu'il constitue l'ensemble complet des réponses du réseau aux stimuli que le réseau a pour fonction d'encoder. Il s'agit de reconstruire la géométrie de l'espace *sous-jacent* aux patterns d'activité à partir des N -vecteurs v qui sont dans \mathcal{C} et aussi de ceux qui *n'y sont pas* (complétude de \mathcal{C} par rapport aux activités possibles du réseau).

On se demande donc comment reconstruire un ensemble de CRs $\mathcal{U} = (U_i)_{i \in I}$ recouvrant un espace M de dimension d (la plus faible possible) tels que $\mathcal{C} = \mathcal{C}(\mathcal{U})$. Cela n'est intéressant qu'avec des U_i convexes car sinon c'est toujours possible de multiples façons et en toute dimension. La convexité (qui implique la connexité) est une forte contrainte.¹³ Considérons par exemple trois ouverts convexes de \mathbb{R}^2 . S'ils s'intersectent deux à deux alors ils ont une intersection non vide. Pour que par exemple U_3 puisse intersecter U_1 et U_2 en contournant $U_1 \cap U_2$ il devrait avoir une forme de "haricot" et ne serait donc plus convexe. Toutefois, si l'on considère des U_i convexes et si $d < N$, on peut utiliser le théorème ci-dessus disant que, puisque les U_i sont convexes, le complexe simplicial $\Xi(\mathcal{C}(\mathcal{U}))$ engendré par $\mathcal{C}(\mathcal{U})$ est exactement le nerf $\mathcal{N}(\mathcal{U})$.

Mais il faut cependant bien voir que l'égalité de complexes simpliciaux $\Xi(\mathcal{C}(\mathcal{U})) = \mathcal{N}(\mathcal{U})$ peut être réalisée par des configurations U_i de même nerf mais *très différentes*. Par exemple si \mathcal{U} est la configuration $\{U_1, U_2, U_{10}\}$ de la figure 12, $\mathcal{N}(\mathcal{U})$ est le 3-simplexe. Mais on obtient le même nerf avec la configuration $\mathcal{U}' = \{U'_1 \subsetneq U'_2 \subsetneq U'_3\}$. La différence entre les deux est que pour \mathcal{U}' les relations d'intersection découlent trivialement des relations d'inclusion qui sont plus fortes. Mais la condition (C) différencie les deux cas. En effet pour \mathcal{U} , $\mathcal{C}(\mathcal{U})$ est lui-même le 3-simplexe $\Xi(\mathcal{C}(\mathcal{U}))$. En revanche, dans le cas de \mathcal{U}' , $\mathcal{C}(\mathcal{U}')$ n'est pas un complexe simplicial car $U_1 \cap U_2 \cap U_3 \setminus \emptyset = U_3 \neq \emptyset$ mais $U_1 \cap U_2 \setminus U_3 = \emptyset$ et donc $\{1, 2\} \notin \mathcal{C}(\mathcal{U}')$.

La géométrisation des codes \mathcal{C} par des recouvrements \mathcal{U} est un problème très difficile ("NP-difficile" comme on dit). Les travaux de Carina Curto *et al.* dans [129] consistent à utiliser des outils de géométrie algébrique pour résoudre algorithmiquement ce problème inverse au moyen de la dualité spectrale entre un espace et

13. La géométrie convexe est un sujet riche et difficile qui s'est beaucoup développé à partir de la fin du XIX^e siècle à travers de célèbres théorèmes comme ceux de Hermann Minkowski (1891), Constantin Carathéodory (1907), Eduard Helly (1913) ou Johann Radon (1921).

l'algèbre des fonctions définies sur cet espace. Comme l'activité des neurones est $\{0, 1\}$ on travaille sur le corps premier à 2 éléments \mathbb{F}_2 . À chaque neurone, on associe une variable x_i , on considère l'algèbre de polynômes $\mathbb{F}_2[x] = \mathbb{F}_2[x_1, \dots, x_N]$ à N variables et on considère le code \mathcal{C} comme une sous-variété algébrique de l'hypercube $\{0, 1\}^N$ qui est l'espace associé à $\mathbb{F}_2[x]$ par dualité spectrale. Pour ce faire, il faut récupérer \mathcal{C} comme la sous-variété algébrique associée au *quotient* de l'algèbre $\mathbb{F}_2[x]$ par l'*idéal* $I_{\mathcal{C}}$ des $h(x) \in \mathbb{F}_2[x]$ qui s'annulent sur \mathcal{C} . Les relations obligatoires $x_i(1 - x_i) = 0$ dues à $x_i = 0$ ou 1, sont évidemment dans $I_{\mathcal{C}}$ et y engendrent un idéal "booléen" \mathcal{B} . Les autres reflètent la géométrie immanente de l'espace M que l'on cherche.

Le quotient $\mathbb{F}_2[x_1, \dots, x_N]/I_{\mathcal{C}} = R_{\mathcal{C}}$ est une algèbre sur \mathbb{F}_2 que les auteurs appellent "neural ring". Cela revient à ne retenir des $f(x) \in \mathbb{F}_2[x]$ que leurs valeurs sur \mathcal{C} .¹⁴ Le reste de $I_{\mathcal{C}}$ par rapport à \mathcal{B} peut facilement s'obtenir. Si v est un N -vecteur, il est facile de construire un polynôme $\chi_v(x)$ qui soit = 1 sur v et = 0 pour tous les autres N -vecteurs. Il suffit de prendre $\chi_v(x) = \prod_{i \in 1_v} x_i \prod_{j \in 0_v} (1 - x_j)$. Soit $J_{\mathcal{C}}$ l'idéal de $\mathbb{F}_2[x]$ engendré par les $\chi_v(x)$ pour $v \notin \mathcal{C}$. On montre alors que $I_{\mathcal{C}} = J_{\mathcal{C}} + \mathcal{B}$.

On peut ensuite, au moyen d'une version sur le corps \mathbb{F}_2 de la "dualité spectrale" sur laquelle nous allons revenir dans la section 6.3, retrouver \mathcal{C} comme le "spectre" de $R_{\mathcal{C}}$, c'est-à-dire comme l'ensemble de ses idéaux premiers. Ces idéaux premiers sont aussi maximaux et tous les idéaux maximaux sont les \mathfrak{m}_v engendrés par les $\{x_i - v_i\}$. Ce résultat fondamental exige d'adapter à la situation (corps de base \mathbb{F}_2) des résultats classiques et fondateurs de la géométrie algébrique sur le corps de base \mathbb{C} (algébriquement clos de caractéristique 0) comme le célèbre Nullstellensatz de Hilbert.

Maintenant, dans le cas où $\mathcal{C} = \mathcal{C}(\mathcal{U})$ est réalisé au moyen d'un recouvrement ouvert \mathcal{U} d'un espace topologique M , on peut *évaluer* les polynômes $f(x) \in \mathbb{F}_2[x]$ en tout point p de M en posant $x_i(p) = 1$ si $p \in U_i$ et $x_i(p) = 0$ sinon. Le résultat principal, justifiant le formalisme, et moins évident qu'il ne peut paraître, est le théorème suivant :

Théorème. L'idéal $I_{\mathcal{C}(\mathcal{U})}$ est égal à l'idéal $I_{\mathcal{U}}$ engendré par les relations suivantes où, si $\sigma \in \mathcal{C}$, on note $U_{\sigma} = \bigcap_{i \in \sigma} U_i$ et $x_{\sigma} = \prod_{i \in \sigma} x_i$:

$$I_{\mathcal{U}} = \left\{ x_{\sigma} \prod_{k \in \tau} (1 - x_k) \mid U_{\sigma} \subseteq \bigcup_{k \in \tau} U_k \right\} .$$

On notera que l'inclusion triviale $U_i \subseteq U_i$ correspond à la relation booléenne

$$x_i(1 - x_i) = 0$$

14. En effet $f = g$ dans l'anneau $R_{\mathcal{C}} \iff f - g \in I_{\mathcal{C}} \iff f(v) = g(v)$ pour tout $v \in \mathcal{C}$.

qui est toujours dans I_C . Quant à la relation x_σ elle correspond à $U_\sigma = \emptyset$, et la relation $\prod_{k \in \tau} (1 - x_k)$ à $M \subseteq \bigcup_{k \in \tau} U_k$. $\text{-----} \diamond$
 Selon les auteurs de [129], l'anneau R_C .

“retains the full combinatorial data of a neural code” ([129], p.3).

et permet de “lire” la structure de CRs associée (p.4).

4.5. Topologie et homotopie

Le théorème de base reliant la géométrie d'un espace topologique M à un de ses recouvrements ouverts $\mathcal{U} = \{U_i\}_{i \in I}$ de nerf $\mathcal{N}(\mathcal{U})$ dit que si les U_i et leurs intersections sont *contractiles* (on dit alors que \mathcal{U} est un “bon” recouvrement, c'est toujours le cas pour des U_i convexes), alors le *type d'homotopie* de M est égal à celui du nerf $\mathcal{N}(\mathcal{U})$.

Théorème. Si \mathcal{U} est un “bon” recouvrement de M alors M a le type d'homotopie du nerf $\mathcal{N}(\mathcal{U})$. $\text{-----} \diamond$

Précisons les termes employés. Étant donnés deux espaces topologiques M et N et deux applications continues $f, g : M \rightarrow N$, une homotopie entre f et g est tout simplement une déformation continue de f en g , c'est-à-dire une famille d'applications continues $h_t : M \rightarrow N$ paramétrées de façon continue par $t \in [0, 1]$ et telle que $h_0 = f$ et $h_1 = g$. Deux espaces M et N sont alors dits de même type d'homotopie s'il existe $f : M \rightarrow N$ et, dans l'autre sens, $g : N \rightarrow M$ continues telles que $g \circ f$ soit homotope à Id_M et $f \circ g$ homotope à Id_N . Un espace est dit contractile s'il a le type d'homotopie d'un point. Le théorème est alors assez intuitif : en contractant les U_i on peut contracter le recouvrement \mathcal{U} sur son nerf $\mathcal{N}(\mathcal{U})$.

Un convexe $U \neq \emptyset$ est toujours contractile. On prend un point base $*$ et si $x \neq *$ on considère le segment $x_t = t * + (1 - t)x$ paramétré par $t \in [0, 1]$ qui joint x à $*$. Les $x \rightarrow x_t$ contractent U sur $*$. Mais l'inverse est trivialement faux : si U est contractile il n'a aucune raison d'être convexe. Il suffit de considérer une étoile.

Ce lien entre le codage neuronal et l'homotopie de l'espace M – type d'homotopie qu'il contient de façon implicite et immanente – est fondamental. Comme le notent Yuri Manin et Matilde Marcolli¹⁵ dans [349],

“the neural code generate a representation of the stimulus space in the form of a homotopy type” (p.4)

et, dans certains cas, la non trivialité de l'homotopie possède un rôle *fonctionnel* pour garantir la cohérence du traitement informationnel distribué dans le réseau neuronal. Elle permet en effet de lever des “obstructions topologiques” à l'obtention de cette cohérence. Ainsi que l'affirment ces auteurs

15. Ces auteurs sont tous deux d'éminents spécialistes de géométrie algébrique non commutative, de théorie des nombres et de physique quantique.

“A good mathematical modeling of network architecture in the brain should also include a mechanism that generates homotopy types, through the information carried by the network via neural codes.” ([349] p. 5)

Ces problématiques se sont considérablement développées ces dernières années et sont devenues assez techniques. Le lecteur intéressé pourra consulter par exemple [349]. Elles se sont en particulier enrichies de liens avec des méthodes bien connues comme, par exemple, celles de l'*inférence bayésienne*. Dans une optique bayésienne, on calcule des probabilités. On suppose que l'on a une probabilité $P(x)$ pour la position x dans un espace M qui fonctionne comme un “prior” bayésien.¹⁶ On observe des données empiriques v qui sont des effets des causes x et on obtient une probabilité a posteriori $P(v) \neq 0$. Les probabilités conditionnelles $P(v|x)$ reflètent le lien causal entre $P(x)$ et $P(v)$. Si l'on connaît $P(x)$ et $P(v|x)$, on peut en effet calculer $P(v)$ au moyen de la formule de Bayes $P(v) = \int P(v|x) P(x) dx$.¹⁷ C'est le problème direct. Le problème inverse (“reverse”, “top-down”) consiste alors à remonter de la probabilité des effets à la probabilité des causes, autrement dit à calculer $P(x|v)$ au moyen de la formule

$$P(x|v) = \frac{1}{P(v)} P(v|x) P(x) = \frac{P(v|x) P(x)}{\int P(v|x) P(x) dx} .$$

Il est en général extrêmement difficile à résoudre mais peut parfois l'être approximativement au moyen d'hypothèses sur le type de probabilités que sont $P(x)$ et $P(v|x)$.

Pour les codes neuronaux on peut spécifier ce modèle très général en utilisant des processus gaussiens dont nous avons parlé à propos de la statistique des pinwheels dans le Vol I à la section 4.6.12 du chapitre 4. L'idée est la suivante. On suppose qu'il existe une géométrie “latente”, “cachée” et “inobservable” de l'espace sous-jacent M (de multivariable x de basse dimension) et on considère des processus “latents” $x(t)$ choisissant aléatoirement un point x de M pour chaque valeur de t avec une certaine loi $P_t(x)$. On obtient un chemin aléatoire de probabilité $P(x(t))$ qui est une réalisation de ce champ paramétré par t . Du côté des spikes des N neurones on a les vecteurs v (de très grande dimension N) qui décrivent les réponses observées. Pour raffiner le cas purement binaire et introduire des taux de décharge, on peut discrétiser le temps en considérant T petits intervalles paramétrés par $t \in \{1, \dots, T\}$ (ce que l'on appelle un “binning” en statistique) et poser que v compte alors des nombres de spikes dans un intervalle. Le chemin aléatoire $x(t)$ dans M engendre une réponse $v(t)$ dans l'espace des états du réseau v (où l'état d'activité est maintenant le vecteur des taux de décharge des neurones).

16. Autrement dit, une connaissance a priori, un “synthétique a priori” probabiliste.

17. Formule due à Bayes (1763) et Laplace (1774).

Une foule de modèles de ce type existent. Citons par exemple le “latent manifold tuning model” d’Anqui Wu¹⁸ [570] faisant des hypothèses probabilistes gaussiennes sur les $P_t(x)$ et les $P(v|x)$ supposés traduire la forme et les propriétés de recollage des profils récepteurs $R_i(x)$ (PRs, “tuning curves”) définis sur les CRs U_i des neurones. On suppose que le processus $P_t(x)$ est *gaussien*. Cela signifie que pour chaque date t on a une loi gaussienne $P_t(x)$ de moyenne $m(t)$ et de variance $\sigma(t)$. Une réalisation de ce champ est donc une marche aléatoire gaussienne $x(t)$ dans M . Nous avons vu Vol I, chapitre 4, section 4.6.12 que de tels champs sont définis par leur *fonction de corrélation* qui est l’espérance $\mathbb{E}[(P_t(x) - m(t))(P_{t'}(x) - m(t'))]$ et garantit un minimum de géométrie (au moins la continuité) de M et des $x(t)$. On suppose par ailleurs que les PRs $R_i(x)$ sont aussi donnés par des processus gaussiens paramétrés cette fois par M . Les taux de décharge sont alors donnés par $\tau_i(t) = R_i(x(t))$. On introduit ensuite comme dernière hypothèse que l’émission des spikes est un *processus de Poisson* d’intensité $\tau_i(t)$ dans l’intervalle t , autrement dit que la probabilité qu’il y ait s spikes dans l’intervalle t est $e^{-\tau_i(t)} \frac{\tau_i(t)^s}{s!}$. Il s’agit d’une hypothèse tout à fait standard dans ce genre de situation. Comme nous l’avons dit, le problème inverse est extrêmement difficile à résoudre mais peut l’être approximativement.

5. Interlude philosophique : vers un transcendantalisme neuronal

En accord avec nos choix méthodologiques, nous allons faire un retour amont de nature plus philosophique.

5.1. Problèmes inverses et “révolution copernicienne”

Nous venons de voir à quel point la possibilité de résoudre des problèmes inverses est devenue centrale dans les neurosciences. Cette question est en fait très ancienne et possède une puissante généalogie philosophique. Depuis longtemps nous insistons sur le fait qu’il s’agit d’une *version neuronale du transcendantalisme* comme recherche des conditions de possibilité des observations empiriques. Dans le domaine de la perception, de l’action et, plus généralement, de la cognition, le “synthétique a priori” correspond aux architectures fonctionnelles neuronales. Bien que résultant à la fois de l’évolution au niveau phylogénétique des espèces et de l’apprentissage au niveau ontogénétique et épigénétique des individus, celles-ci fonctionnent pourtant bien comme des a priori pour les propriétés empiriques de la cognition. Des philosophes de plus en plus nombreux en conviennent. On parle souvent à ce propos de “reverse-engineering”, de “predictive processing” et de “top-down analysis”.

18. Thèse du Neuroscience Institute de l’Université de Princeton sous la direction de Jonathan Pillow en 2019.

En ce qui concerne la perception, la “révolution copernicienne” de Kant est la première thématization philosophique du passage d’un problème direct à un problème inverse. Le problème direct correspond à l’idée commune que les perceptions sont engendrées passivement par les sensations et nous donnent directement accès à l’ontologie du monde externe qui en est la cause. L’empirisme et l’associationnisme (Locke, Hume, et leurs continuateurs) reposent sur cette évidence. Mais les difficultés insurmontables rencontrées dans la reconstruction des causes externes des sensations a conduit Kant à inverser le point de vue et à partir à la recherche, comme l’explorateur d’un nouveau monde, des *conditions de possibilité*, inconnues et cachées “dans les profondeurs de l’âme”, des structures que l’on observe dans la perception. Il s’agit de la première recherche de “priors”, non pas au sens probabiliste de Bayes mais en un sens eidétique.

Si nous appliquons cette “révolution” à l’espace immanent et interne dont nous avons parlé, nous voyons apparaître un lien frappant *entre la neurogéométrie et l’esthétique transcendantale*. Certes, l’esthétique transcendantale kantienne semble fort loin de nous, d’autant plus qu’elle a été discréditée pendant plus de deux siècles. Mais elle garde pourtant une pertinence remarquable dans ce nouveau contexte et se trouve en quelque sorte revivifiée et même “vindicated”. Nous ne cessons de l’affirmer depuis longtemps comme un leitmotiv.¹⁹

La neurogéométrie des architectures fonctionnelles que nous allons développer en détail dans les chapitres suivants est une géométrie fonctionnant comme un *a priori* au sens d’un “reverse engineering”. L’espace apparaît ainsi comme un *format* pour le traitement des informations sensorielles véhiculées par les fibres du nerf optique.

Or il suffit de comprendre que l’espace est un format neurophysiologique défini par l’architecture fonctionnelle du système visuel pour valider immédiatement la plupart des thèses kantiennees sur l’espace sensible. Ce formatage est en effet par définition une “forme a priori” des contenus sensoriels et n’appartient donc pas à ces contenus eux-mêmes puisque ceux-ci ne sont que des inputs qui, ainsi formatés, structurés et traités, se transforment en percepts. Il est “synthétique” et non conceptuel (ante-prédictif et pré-judicatif) et le fait qu’il soit un résultat de l’évolution biologique ne remet pas en cause son statut a priori car les a posteriori de la phylogénèse sont des a priori de l’ontogénèse.

5.2. Intuition, schème, concept

Les jugements synthétiques a priori géométriques peuvent ainsi, selon nous, être interprétés comme la “réflexion” de structures neurophysiologiques fondamentales. On ne saurait donc arguer – et ce point est vraiment pour nous crucial – du fait que toute géométrie se développe nécessairement comme un système d’énoncés en tant que théorie mathématique, pour conclure que les formes de l’intuition sont

19. Cf. par exemple [419], [420], [427], [436], [437].

“analytiques”. En effet, la géométrie intuitive que nous éprouvons constamment dans la perception visuelle provient de l’architecture fonctionnelle des aires visuelles.

Nous allons traiter techniquement dans les chapitres suivants plusieurs exemples :

1. le transport parallèle d’orientations (reconnaître que deux orientations en deux points différents du champ visuel sont les mêmes),
2. le principe gestaltiste de “bonne continuation” (le système visuel a tendance à prolonger coaxialement les segments orientés),
3. les mécanismes d’intégration des bords (des détections locales de bords d’objets sont intégrées en bords globaux),
4. les contours illusoires comme géodésiques de structures sous-riemanniennes, ces contours illusoires étant la manifestation du fait que la perception est active et structure et complète les sensations à partir de ses structures propres immanentes et endogènes,
5. de façon plus générale les illusions visuelles,
6. les hallucinations visuelles purement géométriques apparaissant dans certains cas de “vision entoptique”.

Toute cette neurogéométrie fonctionnelle qui structure géométriquement le monde sensible est une conséquence des architectures fonctionnelles visuelles devenues aujourd’hui en partie expérimentalement accessibles grâce à l’imagerie mentale. Elle explicite la thèse très kantienne que, contrairement aux sensations, les perceptions sont actives et inséparables de l’“imagination”. Les contenus sensoriels correspondent à des inputs neuronaux qui se propagent dans les réseaux corticaux, mais c’est le “design” de ces derniers qui les forme et les structure et fonctionne donc comme une forme a priori de la sensibilité.

Cela signifie que la structuration active des sensations en percepts par les architectures fonctionnelles fait de la reconnaissance d’un objet une reconstruction “top-down” de l’objet à partir de règles de construction et non pas une identification à une image stockée dans une immense base de données. On retrouve ici la théorie kantienne du *schématisme*. Reconnaître c’est reconstruire au moyen d’un schème, le schème étant un concept transformé en modèle générateur de ses référents. Il y a là une complète inversion du statut traditionnel des concepts. Les concepts ne sont plus obtenus par abstraction (ce qu’on appelait à l’époque de Kant la “subsumption” sous le concept) mais, une fois schématisés, des modèles générateurs.

Nous reviendrons plus en détail sur le “kantisme neuronal” dans la section 2 du chapitre 19 de Conclusion.

5.3. Kant \implies Helmholtz \implies aujourd’hui

Une bonne partie des neurosciences cognitives d’aujourd’hui est donc “kantienne” même si elle l’est la plupart du temps de façon implicite. Souvent, c’est plutôt une

référence à Helmholtz qui est explicite et ce dernier joue donc un rôle intermédiaire essentiel dans la mesure où, d'un côté, il cherchait à donner une interprétation scientifique (physiologique, psycho-physique et psychologique) de l'épistémologie kantienne alors que, d'un autre côté, il est resté une référence fondatrice consensuelle pour l'actualité. Nous en avons plusieurs fois parlé dans le Vol I, en particulier au chapitre 2, section 1 (le problème de Riemann-Helmholtz-Lie), aux chapitres 3, section 3.2.5 et 4, section 4.12.2 (théorie du trichromatisme de Young-Helmholtz), au chapitre 4, section 4.6.6 (équation d'Helmholtz pour les pinwheels), ainsi qu'au chapitre 5, section 5.2.2 (principe d'Helmholtz disant que pour la perception les configurations rares sont saillantes). Nous en reparlerons dans ce volume, par exemple au chapitre 6 de nouveau à propos du "Raumproblem" et au chapitre 8 à propos des contours illusoire.

Pour un aperçu sur la façon dont le milieu philosophique-cognitivist anglo-américain se positionne par rapport à cette généalogie transcendantale, on pourra se référer par exemple à l'article de synthèse de Link Swanson [518] qui explique que le

"predictive processing's 'reversal' is Kant's 'Copernican revolution'".

L'avantage de ce texte est qu'il articule bien l'approche kantienne avec les travaux actuels sur le reverse engineering, la génération "top-down" des percepts, les "hyperpriors", les processus d'analyse-par-synthèse, le rôle de l'imagination et cela en le contextualisant par rapport au contexte anglo-américain de "revisitation" de l'a priori (Michael Friedman, Paul Guyer, Philip Kitcher, etc.) et par rapport à de nombreuses réflexions épistémologiques assez récentes sur la cognition comme celles d'Andy Clark (la perception comme prédiction), de Karl Friston (le cerveau bayésien), de Jakob Hohwy (l'esprit prédictif) ou de Daniel Kersten et Alan Yuille (inférences bayésiennes).

6. De la topologie à la géométrie différentielle

Après cet interlude philosophique, nous allons maintenant passer de la neuro-topologie aux structures *différentielles* qui sont au cœur de la neurogéométrie. Il s'agit d'enrichir les "hyperpriors" géométriques de la perception – son "esthétique transcendantale" – et de reconstruire les propriétés des percepts à partir de cette géométrie immanente.

6.1. Variétés différentiables et fibrés tangents

6.1.1. Variétés et cartes locales.

De nombreux espaces topologiques que nous rencontrerons dans cet ouvrage seront munis d'une structure de variété différentiable. Intuitivement, une variété différentiable est un espace topologique localement identifiable à un espace standard

\mathbb{R}^n à travers des systèmes de coordonnées locales. Cela signifie qu'on s'est donné un ensemble d'ouverts U_α , appelés *cartes locales*, homéomorphes à des ouverts U'_α de \mathbb{R}^n par des homéomorphismes locaux $\varphi_\alpha : U'_\alpha \rightarrow U_\alpha$ et tels que les changements de coordonnées locales (ou changements de cartes) sur les intersections $U_\alpha \cap U_\beta$ de cartes U_α et U_β se recouvrant partiellement s'obtiennent au moyen de difféomorphismes classiques (homéomorphismes différentiables dans les deux sens) $\varphi_{\alpha\beta}$ entre les domaines $\varphi_\alpha^{-1}(U_\alpha \cap U_\beta) \subset U'_\alpha$ et $\varphi_\beta^{-1}(U_\alpha \cap U_\beta) \subset U'_\beta$ de \mathbb{R}^n .²⁰ Un tel système s'appelle un "atlas" et il peut évidemment y avoir plusieurs atlas compatibles définissant la même structure différentiable.

Il est trivial de définir les degrés ou les classes de différentiabilité C^k à partir de ceux des fonctions sur \mathbb{R} . En général "différentiable" signifie indéfiniment différentiable (classe C^∞).

Le concept de variété différentiable correspond à un niveau de structure plus riche mais aussi plus contraint que celui d'espace topologique. Il a été introduit pour disposer d'une classe d'espaces généralisés auxquels il soit possible d'appliquer les techniques de géométrie différentielle. Dans la mesure où chaque point est entouré par un voisinage difféomorphe à un domaine (ouvert) d'un espace classique, toutes les opérations différentielles *locales* peuvent être transférées aux variétés. Jan Koenderink [298] a proposé l'idée que l'espace tangent en un point du champ visuel considéré comme variété pouvait être implémenté par une hypercolonne au-dessus de ce point. Pour préciser cette hypothèse, qui rejoint en partie la notre, nous devons définir ces notions plus précisément.

6.1.2. Vecteurs tangents.

Les notions duales classiques de vecteur tangent et de vecteur cotangent de \mathbb{R}^n se généralisent immédiatement aux variétés différentiables puisqu'il s'agit de notions locales et que les cartes locales d'une variété M de dimension n sont difféomorphes à \mathbb{R}^n . Si une équation paramétrique locale d'une courbe différentiable γ dans M au voisinage de $x(t_0) = x$ est $\gamma(t) = x(t) = (x_1(t), \dots, x_n(t))$ où les coordonnées $x_i(t) \in \mathbb{R}$ sont des fonctions différentiables classiques de t , alors son vecteur tangent X en $x(t_0) = x$ sera son vecteur "vitesse"

$$X = (X_1, \dots, X_n) = \left(\frac{dx_1(t)}{dt}, \dots, \frac{dx_n(t)}{dt} \right) \Big|_{t=t_0} = \frac{dx(t)}{dt} \Big|_{t=t_0}.$$

Il est évident que les vecteurs tangents en x constituent un espace vectoriel sur \mathbb{R} .²¹ Celui-ci est noté $T_x M$ et appelé l'*espace tangent* à M en x . Les différents

20. Pour une définition plus technique des variétés différentiables, cf. par exemple le classique Spivak [511] ou n'importe quelle encyclopédie.

21. On considère des coordonnées locales centrées en x (i.e. $x = 0$ dans la carte) et des courbes $\gamma(t)$ passant en $x = 0$ pour $t = 0$. λX s'obtient en considérant la courbe $\lambda x(t)$ et $X + Y$ en considérant la courbe $x(t) + y(t)$.

$T_x M$ se recollent entre eux de façon différentiable pour constituer un espace fibré $\pi : TM \rightarrow M$ de base M et de fibres $T_x M$ isomorphes à \mathbb{R}^n . Il est de dimension $2n$ et s'appelle le *fibré tangent* de M .

Nous renvoyons au Vol I, chapitre 4, section 4.3. (“ $V1$ as a mesoscopic fibration”) pour des précisions sur le concept de fibration. Insistons encore une fois sur son importance. Le fibré tangent est un exemple typique de fibré vectoriel différentiable. On a un espace de base qui est une variété différentiable M avec en chaque point x de M des coordonnées locales $x = (x_1, \dots, x_n)$ de “position” et un espace $T_x M$ d’un certain *type* correspondant à des variables supplémentaires qui, pour reprendre la belle expression de David Hubel (Vol 1, chapitre 4, section 4.3.2.) rappelée plus haut, sont des “engrafted variables”. Les fibres vectorielles $T_x M$ sont des sous-variétés différentiables munies en général de structures supplémentaires (souvent algébriques) et elles varient différentiablement avec x . En quelque sorte, les x correspondent à des variables de position “externes” et les $T_x M$ à des variables “internes” paramétrées par ces variables “externes”. On peut donc considérer des champs différentiables de $X_x \in T_x M$, autrement dit des sections différentiables (voir Vol 1, 4.3.) de la projection canonique $\pi : TM \rightarrow M$. On comprend donc pourquoi ce formalisme est omniprésent dans les diverses théories des champs rencontrées en physique. Il est adapté aux situations où des entités d’un certain type (vecteurs, repères, tenseurs, spineurs, nombres quantiques, etc.) sont paramétrées par les positions dans l’espace-temps. C’est pour la même raison qu’il est omniprésent en neurogéométrie où des variables traitées corticalement sont fibrées rétinotopiquement (“engrafted”) “au-dessus” de positions rétinienne.

Dans notre cas où la structure algébrique des fibres est celle d’espace vectoriel de dimension n , les sections de $\pi : TM \rightarrow M$ sont les champs de vecteurs tangents sur la variété M . Il faut noter que cet espace de sections, traditionnellement noté $\Gamma(TM)$ est non seulement un \mathbb{R} -espace vectoriel mais un *module* sur l’algèbre $C^\infty(M)$ des fonctions différentiables réelles $\varphi(x)$ sur M .²² En effet chaque X_x peut être multiplié dans sa fibre $T_x M$ par un scalaire réel $\varphi(x)$ et la C^∞ -différentiabilité de X impose simplement que $\varphi(x)$ soit C^∞ par rapport à la variation de x dans la base.

6.1.3. *Ponctuel, infinitésimal, local, global.*

Insistons à propos de la notion de champ de vecteurs sur l’importance qu’il y a à être terminologiquement précis. C’est la rançon de l’utilisation du concept fondamental de fibration. Et rappelons encore une fois que si, mathématiquement parlant, une fibration peut n’être qu’une simple façon de formuler des structures locales de produits cartésiens²³, il n’en va pas du tout de même dans le cas des

22. La structure de module sur une algèbre est la généralisation directe de celle d’espace vectoriel sur un corps.

23. Par exemple on peut voir un produit direct $A \times B$ comme la fibration $\pi_A : A \times B \rightarrow A$ ou comme la fibration $\pi_B : A \times B \rightarrow B$.

modèles neurogéométriques où les fibrations sont *neurophysiologiquement* incarnées dans la connectivité neurale, en particulier par la rétinotopie des aires visuelles. Dans une fibration, ce qui se passe dans les fibres est *ponctuel* relativement aux positions dans la base qui paramétrise ces fibres, mais en même temps varie différemment lorsque cette position varie. Précisons un peu cette dialectique “valeurs ponctuelles dans les fibres” VS “variation différentiable relativement à la base” :

1. Comme section du fibré tangent $\pi : TM \rightarrow M$, un champ de vecteurs X est une structure *globale* définie sur la variété M tout entière.
2. On peut la restreindre à des ouverts de M , c'est-à-dire la localiser plus ou moins. L'opposition local/global reste assez relative car en prenant un ouvert V suffisamment plus petit qu'un ouvert U , on peut considérer que la restriction à U est en fait globale par rapport à la restriction à V .
3. La vraie localisation correspond aux restrictions à des cartes locales car, dans ce cas, on peut utiliser des coordonnées locales pour calculer. Sur des ouverts U plus grands que des cartes locales, il faut utiliser des changements de coordonnées pour recoller les restrictions sur les cartes locales et cela complique notablement les calculs.
4. Un champ X possède des valeurs ponctuelles X_x ou $X(x)$ (variante notationale) aux points x de M , valeurs qui sont des vecteurs de $T_x M$ définissables en termes de coordonnées locales.
5. Mais ces valeurs sont des valeurs de dérivées et donc des valeurs ponctuelles de variations *infinitésimales*.

6.2. Le calcul différentiel classique

Nous devons donc approfondir ces notions primitives élémentaires de façon à pouvoir comprendre comment s'est opérée progressivement la transformation complète du calcul différentiel classique qui permet d'introduire des modèles de géométrie différentielle en neurogéométrie. Il s'agit d'une très longue histoire qui nous force en fait à remonter jusqu'aux sources *leibniziennes* du calcul différentiel. Nous aurons l'occasion d'y revenir de façon récurrente au cours des prochains chapitres.

6.2.1. Les infinitésimales leibniziennes et leur paradoxe.

Dès l'origine, les symboles leibniziens dx pour désigner des différentielles infinitésimales ont eu *un double statut*. D'un côté, on interprétait intuitivement les dx_i en un point $x = (x_1, \dots, x_n)$ comme des incréments infinitésimales δx_i des coordonnées x_i . Le lien avec ce qui précède se fait en considérant des $\delta x_i = \varepsilon X_i$ où ε est une infinitésimale. Si $\varepsilon = dt$, alors

$$\delta x_i = \varepsilon X_i = dt \frac{dx_i}{dt} = dx_i .$$

Dans cette perspective, les δx sont définis *dans l'espace M* et concernent des variations infinitésimales de position dans l'espace.

Mais d'un autre côté, chez Leibniz, les symboles dx sont bien des *symboles*, et des symboles faisant l'objet d'un *calcul symbolique*. On en connaît les règles :

1. $da = 0$ si a est une constante,
2. $d(x + y) = dx + dy$ (linéarité par rapport à l'addition),
3. $d(xy) = xdy + ydx$ (règle fondamentale de Leibniz pour le produit).

La règle du produit implique que, si a est constante, $d(ax) = adx + xda = adx$ (linéarité par rapport à la multiplication par un scalaire). Elle implique aussi, trivialement, $d(x^n) = nx^{n-1}dx$ et donc, pour des fonctions $f(x) = \sum_{n=0}^{n=\infty} a_n x^n$ exprimables par des séries convenablement convergentes, la dérivation terme à terme $f'(x) = \frac{df(x)}{dx} = \sum_{n=1}^{n=\infty} na_n x^{n-1}$ si la série dérivée est elle aussi convenablement convergente.²⁴

L'articulation entre le statut *symbolique* et le statut *référentiel* et *numérique* (i.e. l'implémentation numérique de symboles) des infinitésimales est à l'origine de controverses métaphysiques et d'innovations théoriques parmi les plus récurrentes et les plus importantes de l'histoire des mathématiques. La question est de savoir si un symbole dx peut référer à une grandeur numérique infinitésimale ε ?

Chez Leibniz, les infinitésimales sont des "fictions bien fondées" ou des "fictions fondées en réalité" (l'expression est admirable), un moyen utile "pour abrégé et pour parler universellement". Les symboles dx ne peuvent pas avoir d'implémentation numérique car le corps \mathbb{R} est "archimédien", ce qui signifie qu'on peut atteindre tout nombre réel y aussi grand que l'on veut à partir d'un nombre réel $x \neq 0$ aussi petit que l'on veut au moyen de multiples Nx de x si N est assez grand. Autrement dit, tous les nombres réels sont commensurables et il n'y a pas plusieurs échelles dans \mathbb{R} . Il ne peut donc pas exister de nombre réel infinitésimal $\varepsilon \neq 0$ plus petit que tous les $\frac{1}{n}$ car son inverse $\Omega = \frac{1}{\varepsilon}$ (qui existerait puisque $\varepsilon \neq 0$) serait plus grand que tous les entiers n . D'ailleurs, si ε existait, $\frac{\varepsilon}{\Omega}$ serait une infinitésimale d'infinitésimale, autrement dit une infinitésimale du deuxième ordre, et il y aurait une infinité d'échelles incommensurables dans \mathbb{R} . Bref, les symboles dx sont sans référents numériques.

6.2.2. Quelques réponses au paradoxe.

C'est pourquoi l'interprétation intuitive des symboles dx comme dénotant des réels infiniment petits qui seraient des entités statiques en quelque sorte indivisibles était considérée comme paradoxale et a déchaîné des controverses terribles (en particulier de Berkeley contre Leibniz). Les tentatives pour surmonter le paradoxe ont été nombreuses.

24. Nous ne précisons pas les critères de convergence comme la convergence absolue ou la convergence uniforme.

Une des plus anciennes et des plus intéressantes, mais qui n'a eu aucun succès à l'époque, a été celle de Bernard Nieuwentijt (1654-1718), un contemporain de Leibniz, qui supposa que si dx dénotait des unités indivisibles alors dx devait être un symbole d'infinitésimale *nilpotente* : le symbole dx dénote un "nombre" $\varepsilon \neq 0$ satisfaisant la propriété purement algébrique $\varepsilon^2 = 0$.²⁵ Bien que cette approche expliquât fort bien les règles du calcul symbolique leibnizien et correspondît à la pratique (on néglige au premier ordre les infinitésimales d'ordre supérieur) elle fut violemment rejetée pour raison d'inconsistance par Leibniz, Johann Bernoulli et Jacob Hermann dans une dure controverse entre 1694 et 1696.²⁶ En fait, elle est algébriquement parfaitement cohérente mais brise la structure de corps car si $\varepsilon^2 = 0$ alors on a $\varepsilon \neq 0$ sans que ε soit pour autant inversible (ε est un diviseur de 0 et un corps n'a pas de diviseurs de 0²⁷). Elle introduit l'extension de \mathbb{R} qu'est l'algèbre $\mathbb{R}[\varepsilon] = \mathbb{R}[X]/(X^2)$ quotient de l'algèbre des polynômes $\mathbb{R}[X]$ par l'idéal (X^2) et $\mathbb{R}[\varepsilon]$ n'est pas un corps. C'est cette définition algébrique qui est aujourd'hui universellement utilisée en géométrie algébrique. Nous y reviendrons dans la section 1 du chapitre 9 où nous donnerons quelques repères historiques et biographiques sur le calcul des variations.

Le débat est particulièrement intéressant. Leibniz acceptait l'idée que l'algèbre des *grandeurs* normales (archimédiennes) puisse être augmentée, d'un côté, par des grandeurs infinitésimales et, d'un autre côté, par des grandeurs infinies qui en seraient les inverses. Mais il considérait que, même ainsi étendue, l'algèbre devait garder la même structure algébrique que \mathbb{R} (i.e. rester un corps comme on le dirait maintenant) en vertu du "principe de permanence" des propriétés algébriques des opérations dans le passage du fini à l'infini. Car, comme il l'affirmait en 1701 dans [324] :

"Il se trouve que les règles du fini réussissent dans l'infini comme s'il y avait des atomes"

et que réciproquement

"les règles de l'infini réussissent dans le fini comme s'il y avait des infiniment petits métaphysiques."

Il faudra attendre l'Analyse non standard (cf. ci-dessous) pour formaliser le statut d'extensions non archimédiennes de \mathbb{R} par des corps \mathbb{R}^* à plusieurs échelles incommensurables contenant des infinitésimales et leurs inverses.

25. Cf. l'ouvrage de Giulio Giorello *Lo Spettro e il Libertino* [213] ainsi que notre étude [430].

26. Leibniz et ses alliés ironisaient sur le "tissus d'absurdités" du pauvre théologien calviniste hollandais en le traitant de façon méprisante de "Sapientissimo Autore".

27. Si $x, y \neq 0$ alors $xy \neq 0$ car, comme x est inversible puisque non nul et que $a0 = 0$ pour tout a , si $xy = 0$ alors $x^{-1}(xy) = (x^{-1}x)y = y = 0$.

Mais l'interprétation devenue classique dans l'analyse dès d'Alembert (cf. ses articles "Différentiel" et "Limite" [131] dans l'*Encyclopédie*) a consisté à simplement éliminer les infinitésimales comme entités, c'est-à-dire comme symboles pouvant dénoter un type de nombres statiques (même idéaux et différents des nombres réels), et à les interpréter comme une façon de parler de *processus* dynamiques de passage à la limite de quantités finies. Comme le dit d'Alembert,

"la théorie des limites est à la base de la vraie métaphysique du calcul".

Le symbole ε n'a plus dès lors de sens en tant que tel car tout référent de dx serait une hypostase et seule la notation $\varepsilon \rightarrow 0$ possède un sens. Cette interprétation, qui correspond également très bien à la pratique, a été rendue rigoureuse par Weierstrass et c'est celle que nous avons tous appris à l'école. Le problème est qu'elle repose sur la structure très particulière de \mathbb{R} qui sert de base à toute l'analyse classique et n'est pas généralisable à des corps \mathbb{K} ne possédant pas de structure de ce type (par exemple les corps finis de caractéristique un nombre premier p). Au contraire, l'interprétation nilpotente $\mathbb{K}[\varepsilon] = \mathbb{K}[X]/(X^2)$ continue à garder tout son sens.

Une autre façon de légitimer les infinitésimales, cette fois en restant fidèle à l'esprit leibnizien, est d'introduire des corps non archimédiens possédant plusieurs échelles incommensurables. Un exemple très intuitif, bien expliqué par le grand spécialiste d'analyse fonctionnelle Paul du Bois-Reymond (1831-1869), est celui du degré de croissance des fonctions $f(x)$ strictement croissantes sur \mathbb{R} . Si l'on assigne à x^n le degré de croissance n alors le degré de croissance de l'exponentielle e^x est infini car e^x croît plus vite que x^n pour tout n et celui du logarithme $\text{Log}(x)$ est infinitésimal car $\text{Log}(x)$ croît moins vite que $x^{\frac{1}{n}}$ pour tout n . On peut à partir de là construire les degrés de croissance positifs intermédiaires et négatifs (en prenant les inverses des fonctions) de façon à obtenir un continuum qui est une extension non archimédienne de \mathbb{R} .

Mais c'est Giuseppe Veronese (1854-1917) qui a été le premier à introduire un continuum *multi-échelle*. Pour lui, le continu était une "intuition pure" et quand on l'arithmétisait à la Cantor-Dedekind il n'y avait aucune raison d'imposer l'hypothèse archimédienne, la similarité leibnizienne du fini et de l'infini pouvant opérer à une infinité d'échelles différentes. Ses constructions n'étaient toutefois pas suffisamment rigoureuses et furent critiquées entre autres par Cantor, Peano et Fraenkel.

C'est l'*analyse non standard* (ANS) d'Abraham Robinson et de Wilhelmus Luxemburg qui arriva à résoudre le paradoxe des infinitésimales comme entités numériques possédant une identité et pouvant être dénotées par des symboles. Pour ce faire, elle dût utiliser des possibilités logiques très spéciales de la théorie des ensembles (cf. l'ouvrage de référence de Robinson [466]).²⁸ Disons simplement quelques mots. On se

28. Pour des précisions, le lecteur intéressé pourra consulter nos études [410] et [416] ainsi que leurs bibliographies.

situé dans un univers \mathcal{U} de théorie des ensembles et on montre à partir du théorème fondamental dit de Löwenheim-Skolem qu'il existe des *extensions strictes* $\mathbb{R} \prec \mathbb{R}^*$ où le cardinal de \mathbb{R}^* est strictement supérieur à celui de \mathbb{R} et qui sont pourtant *élémentaires* au sens suivant, dû à Tarski et Vaught : \mathbb{R}^* a la même théorie que \mathbb{R} dans la logique du premier ordre²⁹ où il existe des symboles de constantes pour tous les éléments de \mathbb{R} . Cela signifie que, quant à sa théorie, c'est-à-dire quant à sa maîtrise "discursive" au moyen d'un langage formel, \mathbb{R}^* est *indiscernable* de \mathbb{R} au premier ordre. Tout énoncé du premier ordre valide dans \mathbb{R} l'est automatiquement dans \mathbb{R}^* et aucun élément de \mathbb{R} ne peut se trouver "supplanté" dans une propriété du premier ordre par un élément de \mathbb{R}^* .

Les éléments de \mathbb{R} sont appelés les éléments standard de \mathbb{R}^* . Soit \mathbb{N} l'ensemble des entiers naturels et $n \in \mathbb{N}$. Comme l'énoncé affirmant qu'il n'existe aucun entier strictement compris entre n et $n+1$ est un énoncé du premier ordre, il est valide dans \mathbb{R}^* . Si \mathbb{N}^* est l'ensemble des entiers de \mathbb{R}^* (que l'on peut définir rigoureusement), \mathbb{N} est donc un *segment initial* de son extension élémentaire \mathbb{N}^* . Ainsi \mathbb{N}^* est composé de \mathbb{N} et de nombres entiers "infinis". \mathbb{R}^* comprend par conséquent des nombres "infinis" plus grands que tout entier de \mathbb{R} puisque \mathbb{R} est archimédien par rapport à \mathbb{N} . Et comme \mathbb{R}^* est un corps, il contient les inverses de ces nombres infinis, et donc des infinitésimales relativement à \mathbb{R} .

Comportant des infinitésimales, \mathbb{R}^* n'est pas archimédien par rapport à \mathbb{N} . Pourtant l'axiome d'Archimède étant du premier ordre, il est valide dans \mathbb{R}^* . Il n'y a là nulle contradiction car \mathbb{R}^* est bien archimédien par rapport à \mathbb{N}^* . On voit la subtilité de la solution du paradoxe des infinitésimales. Elle aurait enchanté Leibniz car elle repose sur des propriétés subtiles des relations entre syntaxe et sémantique dans les théories logiques. Les infinitésimales de \mathbb{R} sont dans \mathbb{R}^* et pas dans \mathbb{R} et \mathbb{R} ne possède donc pas d'infinitésimales. Mais comme $\mathbb{R} \prec \mathbb{R}^*$ est une extension élémentaire, \mathbb{R}^* est indiscernable de \mathbb{R} et tout se passe "comme si" les infinitésimales dans \mathbb{R}^* appartaient à \mathbb{R} . Le symbole dx peut maintenant dénoter quelque chose de bien défini à savoir le sous-ensemble μ de \mathbb{R}^* , $\mu = \{\varepsilon \neq 0 \in \mathbb{R}^* \mid |\varepsilon| < x, \forall x > 0 \in \mathbb{R}\}$. En hommage à Leibniz, Luxemburg a proposé d'appeler μ la "monade" de 0. Par invariance par translation, si $x \in \mathbb{R}$, $x + dx$ est la monade de x .

29. On appelle logique du premier ordre, l'application du calcul des prédicats à une structure où l'on restreint la quantification aux *éléments* de l'ensemble sous-jacent à cette structure. En particulier on ne peut pas quantifier sur les sous-ensembles si ceux-ci ne sont pas tous caractérisables par un ensemble fini d'énoncés du premier ordre. La théorie du premier ordre d'une structure infinie est faible et il en existe une infinité d'extensions élémentaires. Par exemple, au premier ordre, l'extension $\langle \mathbb{Q}, < \rangle \prec \langle \mathbb{R}, < \rangle$ de l'ensemble dénombrable densément ordonné \mathbb{Q} des rationnels par l'ensemble ordonné non dénombrable et complet des réels est une extension élémentaire. La propriété fondamentale de complétude des réels est une propriété du deuxième ordre.

Pour être correctement développée sans contradictions l'ANS exige de se placer dans une extension élémentaire \mathcal{U}^* de l'univers des ensembles \mathcal{U} de départ, ce qui est possible puisque la théorie des ensembles concernant \mathcal{U} est une théorie axiomatisée du premier ordre de la relation primitive d'appartenance \in .³⁰ Toutes les constructions faites dans \mathcal{U} s'étendent automatiquement à \mathcal{U}^* . Mais le point crucial est que lorsqu'on étend la relation $a \in A$ de \mathcal{U} à la relation $a^* \in^* A^*$ de \mathcal{U}^* , \in^* ne donne pas tous les sous-ensembles de A^* dans \mathcal{U}^* mais seulement les sous-ensembles dits "internes". Autrement dit, alors que $\in = \in_{\mathcal{U}}$ est la relation d'appartenance complète de \mathcal{U} , \in^* n'est qu'une sous-relation incomplète de la relation d'appartenance complète $\in_{\mathcal{U}^*}$ de \mathcal{U}^* . La différence entre sous-ensembles internes et externes (non internes) permet d'introduire le principe que les énoncés valides dans \mathcal{U} restent valides par extension à \mathcal{U}^* à condition d'y restreindre les quantifications aux sous-ensembles *internes*. Sans ce principe, on se heurterait immédiatement à des inconsistances. Par exemple l'énoncé (E) "tout sous-ensemble de \mathbb{N} admet un plus petit élément" est valide dans \mathcal{U} . Par transfert à \mathcal{U}^* , cet énoncé y devient valide. Pourtant il est clair que l'ensemble $\mathbb{N}_{\infty}^* = \mathbb{N}^* - \mathbb{N}$ des entiers "infinis" de \mathbb{N}^* ne saurait posséder de plus petit élément. Il n'y a là aucun paradoxe car, interprété dans \mathcal{U}^* , l'énoncé (E) signifie "tout sous-ensemble interne de \mathbb{N}^* admet un plus petit élément". Mais \mathbb{N}_{∞}^* est un sous-ensemble *externe*.

Évidemment, toutes les monades μ sont des sous-ensembles *externes* de \mathbb{R}^* et, dans l'après-coup de l'ANS, on peut dire que c'est leur traitement comme sous-ensembles internes qui est à l'origine du paradoxe des infinitésimales.

L'approche de Robinson de l'ANS est sémantique. Elle repose sur la théorie des modèles. Il existe aussi des versions purement syntaxiques de l'ANS comme l'*Internal Set Theory* proposée par Edward Nelson et aussi des versions étrangement plus "finitistes" comme celle, fort intéressante, proposée par Georges Reeb, Jacques Hartong et leurs collègues. Il existe aussi des versions simplifiées de l'ANS comme celle proposée par Vieri Benci et Mauro Di Nasso [39]. Mais nous arrêterons ici cet aperçu sur les interprétations des infinitésimales leibniziennes dans le cadre de la théorie des ensembles. Le lecteur intéressé pourra trouver de nombreuses précisions dans l'ouvrage de référence de Jean-Michel Salanskis *Le constructivisme non standard* [482] ou dans *Le Labyrinthe du Continu* [317], Actes d'un Colloque de Cerisy organisé en 1990 par Jean-Michel Salanskis et Hourya Sinaceur.

30. Il y a ici quelques subtilités. D'abord, la théorie $\langle \mathcal{U}, \in \rangle$ de l'univers \mathcal{U} est du premier ordre mais on peut y formaliser les théories d'ordre supérieur de toutes les structures définies sur des ensembles. Ensuite, il faut montrer que la notion de modèle non standard d'une structure définie sur un ensemble est généralisable sans contradictions à l'univers des ensembles lui-même et que \mathcal{U}^* a bien un sens.

6.3. Opérateurs de dérivation et formes différentielles

Ce qui nous intéresse ici au premier chef est une interprétation du symbole dx qui est complètement différente de l'interprétation naïve et qui le traduit en une *forme différentielle*, c'est-à-dire en une entité concernant non plus des petites variations de position mais des *opérateurs* opérant sur des fonctions sur l'espace M . Cette réinterprétation, due en grande partie à Élie Cartan, est essentielle pour notre propos et il nous faut donc nous y arrêter un peu en détail.

6.3.1. Dualité "spectrale" entre espace et fonctions.

Pour comprendre ce qui est en jeu, il faut comprendre la dualité fondamentale qui existe entre un espace M et l'algèbre commutative \mathcal{A} des fonctions à valeurs réelles $f : M \rightarrow \mathbb{R}$, $x \mapsto f(x)$, définies sur cet espace (on peut généraliser de \mathbb{R} à \mathbb{C} si besoin est). Cette dualité est à la racine de toutes les interprétations "spectrales" de la géométrie qui dominent

- (i) la géométrie algébrique depuis Serre et Grothendieck,
- (ii) la physique mathématique depuis l'avènement de la mécanique quantique et
- (iii) la géométrie non commutative d'Alain Connes.

Les $f : M \rightarrow \mathbb{R}$ sont des "observables" qui mesurent au moyen d'un nombre réel une grandeur paramétrée par les positions x dans M . Par exemple, dans le contexte neurogéométrique, une image est une fonction d'intensité $I(x)$ sur le plan rétinien $M = \mathbb{R}^2$. Les champs de profils récepteurs (cf. Vol 1, chapitre 3) sont des observables.

La structure d'algèbre de \mathcal{A} est héritée de la structure de corps de \mathbb{R} , les opérations algébriques sur les f s'effectuant point par point : $(f + g)(x) = f(x) + g(x)$, etc. \mathcal{A} est seulement une algèbre et non pas un corps car le 0 de \mathcal{A} est la fonction identiquement nulle 0_M et si $f \neq 0_M$, f peut néanmoins s'annuler quelque part en des $x \in M$ et alors f ne peut pas s'inverser car $\frac{1}{f(x)} = \frac{1}{0} = \infty \notin \mathbb{R}$.

La dualité espace / fonction devient évidente dès que l'on remarque que si une fonction particulière f associe une valeur $f(x)$ à chaque point x de M , un point particulier x de M associe dualement une valeur $f(x)$ à chaque fonction f , ce que l'on appelle l'évaluation de f en x . On peut donc considérer l'application $\delta_x : \mathcal{A} \rightarrow \mathbb{R}$, $f \mapsto \delta_x(f) = f(x)$ (la notation δ vient de la distribution de Dirac δ) qui est par construction un morphisme d'algèbres. Ce faisant, x devient un *opérateur* sur les f . Il est devenu traditionnel de noter $\langle f, x \rangle$ la valeur $f(x) = \delta_x(f)$ de façon à faire ressortir la dualité.

La considération de cette dualité a ouvert un immense champ de réflexions consistant à coder les propriétés topologico-géométriques des espaces par les propriétés topologico-algébriques de leurs algèbres de fonctions. Par exemple, si $x \in M$, l'ensemble \mathfrak{m}_x des f qui s'annulent en x est trivialement un *idéal* de \mathcal{A} , c'est-à-dire un sous-ensemble stable par addition (un sous-groupe additif) et par multiplication par

tout élément de \mathcal{A} . Cela ne fait qu'exprimer le fait que 0 est l'élément neutre de l'addition de \mathbb{R} et qu'il est absorbant pour la multiplication ($0a = 0$ pour tout $a \in \mathbb{R}$). Cet idéal est *premier* (i.e. si $f.g \in \mathfrak{m}_x$ alors f ou $g \in \mathfrak{m}_x$ puisque si $f(x)g(x) = 0$ alors $f(x)$ ou $g(x) = 0$). Qui plus est, sous certaines conditions (en particulier si l'on travaille avec le corps \mathbb{C} qui est *algébriquement clos*), l'idéal \mathfrak{m}_x sera *maximal*.

Les idéaux premiers (resp. maximaux) constituent ce que l'on appelle le *spectre* (resp. le spectre maximal) de \mathcal{A} et l'on associe ainsi dualement à \mathcal{A} un espace géométrique $\text{Spec}(\mathcal{A})$ que l'on peut comparer à M . Les approches spectrales de la géométrie se sont vraiment développées à partir du résultat magistral d'Israël Gelfand (1913-2009) en 1943 disant que si M est un espace topologique séparé compact et si l'on considère l'algèbre $\mathcal{A} = \mathcal{C}^0(M)$ des fonctions continues à valeurs dans \mathbb{C} , alors les idéaux maximaux \mathfrak{m} de \mathcal{A} sont les noyaux des "caractères" $\chi : \mathcal{A} \rightarrow \mathbb{C}$ de \mathcal{A} (i.e. des formes linéaires multiplicatives sur \mathcal{A}) et que, si l'on munit $\text{Spec}(\mathcal{A})$ de la topologie de la convergence simple des caractères (i.e. $\chi_n \rightarrow \chi$ si et seulement si $\chi_n(f) \rightarrow \chi(f)$ pour tout $f \in \mathcal{A}$), alors l'algèbre \mathcal{A} est isomorphe à l'algèbre $\mathcal{C}^0(\text{Spec}(\mathcal{A}))$ des fonctions continues sur son spectre $\text{Spec}(\mathcal{A})$, ce dernier étant homéomorphe (i.e. topologiquement isomorphe) à M . Les caractères d'évaluation δ_x donnent tous les caractères χ .³¹ Bref, les observables sont "localisables" et l'espace de départ M est reconstituable à partir de la pure structure algébrique de l'algèbre de ses fonctions.

6.3.2. Les vecteurs tangents comme opérateurs de dérivation.

La réinterprétation du symbole leibnizien dx comme forme différentielle consiste alors à réinterpréter tout le calcul différentiel classique non plus à partir des positions x dans M mais à partir des observables f .

Pour cela, on commence par réinterpréter les vecteurs tangents à une courbe γ comme des opérateurs de *dérivation* des observables le long de γ . Plus précisément, si f est une observable régulière, i.e. une fonction suffisamment dérivable $f : M \rightarrow \mathbb{R}$, et si l'on considère une courbe différentiable γ d'équation $\gamma(t)$, f devient *le long de* γ une fonction différentiable de t à travers les coordonnées (locales) x_i des x , c'est-à-dire une fonction composée $f(\gamma(t))$. Soient alors X_i les composantes du vecteur tangent à γ en $x \in \gamma$. La dérivée *totale* de f par rapport à t le long de γ est alors donnée, d'après la règle de dérivation des fonctions composées, par :

$$\frac{df(\gamma(t))}{dt} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} = \sum_{i=1}^n X_i \frac{\partial f}{\partial x_i} = \left(\sum_{i=1}^n X_i \frac{\partial}{\partial x_i} \right) (f) = X(f)$$

31. Dans la suite de cet ouvrage, nous noterons C^k la propriété de différentiabilité à l'ordre k d'une application ($k = 0$ signifiant la simple continuité) et \mathcal{C}^k les espaces fonctionnels d'applications C^k .

où X est l'opérateur de dérivation $X = \sum_{i=1}^n X_i \frac{\partial}{\partial x_i}$ de composantes X_i relativement aux $\frac{\partial}{\partial x_i}$. Les vecteurs tangents à γ peuvent par conséquent être interprétés comme des opérateurs de dérivation sur les observables, les dérivations particulières $\frac{\partial}{\partial x_i}$ associées aux coordonnées locales x_i constituant une base du vectoriel $T_x M$.

Pour être rigoureux, il faut tenir compte de ce que nous avons dit plus haut des relations entre ponctuel, infinitésimal, local et global. Dans une carte locale U de coordonnées locales (x_i) , un champ de vecteur X a des composantes $X_i(x)$ relativement à la base $\frac{\partial}{\partial x_i}$ des espaces tangents $T_x M$, base qui est utilisable en tout point $x \in U$. Les formules ci-dessus sont ponctuellement valables dans les fibres en tout $x_0 \in U$ avec $X_i \rightsquigarrow X_i(x_0)$, $\gamma(t) \rightsquigarrow \gamma(t_0) = x_0$, $\left. \frac{df(\gamma(t))}{dt} \rightsquigarrow \frac{df(\gamma(t))}{dt} \right|_{t=t_0}$ et $\frac{\partial f}{\partial x_i} \rightsquigarrow \frac{\partial f}{\partial x_i}(x_0)$ et expriment que ces valeurs ponctuelles varient différemment avec x .

On voit que l'on peut traiter $X = (X_i)$ soit comme un simple vecteur dans la fibre $T_x M \simeq \mathbb{R}^n$, soit comme un opérateur de dérivation. Dans le premier cas les X_i sont les composantes de X dans une base e_i de \mathbb{R}^n alors que dans le second cas ce sont les composantes ponctuelles en x d'un champ X dans le *champ* de bases $\left(\frac{\partial}{\partial x_i} \right)$ dans un voisinage U de x .

Nous reviendrons plusieurs fois, entre autres à la section 3.1 du chapitre 8 et à la section 8 du chapitre 9 (où nous commenterons l'extraordinaire *Methodus* d'Euler introduisant les modèles variationnels), sur ce double aspect des dérivations des observables f définies sur M lorsqu'elles deviennent des fonctions composées à travers des trajectoires $\gamma(t)$ dans M . On peut les voir d'une part en termes de dérivées partielles $\frac{\partial f}{\partial x_i}$ et d'autre part en termes de dérivées totales $\frac{df(\gamma(t))}{dt}$. Ce double statut mathématiquement crucial se trouve en effet *implémenté* dans les architectures fonctionnelles du hardware neuronal.

6.3.3. L'algèbre de Lie des champs de vecteurs.

Une variété différentiable M possède donc un fibré tangent TM dont les sections sont les champs de vecteurs différentiables X qui associent à tout $x \in M$ un vecteur tangent $X(x) \in T_x M$. Ces champs forment un module $\mathfrak{X}(M)$ sur l'algèbre $\mathcal{C}^\infty(M)$ et nous venons de voir qu'ils peuvent être interprétés comme des opérateurs de dérivation $X = \sum_i X_i \frac{\partial}{\partial x_i}$ sur les fonctions différentiables (à valeurs réelles) f sur M . Cette action de X sur les fonctions f s'appelle la *dérivée de Lie* de f par rapport à X et se note aussi $\mathcal{L}_X f$. Elle est linéaire par rapport aux $X \in \mathfrak{X}(M)$, autrement dit $\mathcal{L}_{hX} f = h \mathcal{L}_X f$ et $\mathcal{L}_{X+Y} f = \mathcal{L}_X f + \mathcal{L}_Y f$ et est une dérivation de l'algèbre $\mathcal{C}^\infty(M)$ au sens structurel, ce qui signifie qu'elle est linéaire par rapport aux f et satisfait la *règle de Leibniz*

$$\mathcal{L}_X (fg) = f \mathcal{L}_X (g) + \mathcal{L}_X (f) g .$$

On peut démontrer que l'on obtient ainsi *toutes* les dérivations de l'algèbre $\mathcal{C}^\infty(M)$.

Une construction fondamentale est alors celle du *crochet de Lie* $[X, Y]$ de deux champs X et Y , crochet défini comme $X \circ Y - Y \circ X$,³² soit

$$\mathcal{L}_{[X, Y]} = \mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X = [\mathcal{L}_X, \mathcal{L}_Y] .$$

Si $Y = \sum_j Y_j \frac{\partial}{\partial x_j}$, on a pour tout f

$$\begin{aligned} [X, Y](f) &= \sum_i X_i \frac{\partial}{\partial x_i} \left(\sum_j Y_j \frac{\partial f}{\partial x_j} \right) - \sum_j Y_j \frac{\partial}{\partial x_j} \left(\sum_i X_i \frac{\partial f}{\partial x_i} \right) \\ &= \sum_i X_i \left(\sum_j Y_j \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_j \frac{\partial Y_j}{\partial x_i} \frac{\partial f}{\partial x_j} \right) \\ &\quad - \sum_j Y_j \left(\sum_i X_i \frac{\partial^2 f}{\partial x_j \partial x_i} + \sum_i \frac{\partial X_i}{\partial x_j} \frac{\partial f}{\partial x_i} \right) . \end{aligned}$$

Le point crucial est que la *commutativité* des dérivations partielles $\frac{\partial}{\partial x_i}$ et $\frac{\partial}{\partial x_j}$ conduit à l'élimination des termes du second ordre : $[X, Y]$ est donc lui aussi un champ de vecteurs, un champ de composantes

$$[X, Y]_i = \sum_j \left(X_j \frac{\partial Y_i}{\partial x_j} - Y_j \frac{\partial X_i}{\partial x_j} \right) .$$

On peut démontrer le même résultat plus structurellement en montrant (c'est un petit exercice symbolique) que $\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X$ est une dérivation parce que \mathcal{L}_X et \mathcal{L}_Y le sont et définir $[X, Y]$ comme le champ qui lui est associé. En effet la linéarité est évidente et on a bien la règle de Leibniz

$$\begin{aligned} [\mathcal{L}_X, \mathcal{L}_Y](fg) &= \mathcal{L}_X(\mathcal{L}_Y(fg)) - \mathcal{L}_Y(\mathcal{L}_X(fg)) \\ &= \mathcal{L}_X(\mathcal{L}_Y(f)g + f\mathcal{L}_Y(g)) - \mathcal{L}_Y(\mathcal{L}_X(f)g + f\mathcal{L}_X(g)) \\ &= \mathcal{L}_X(\mathcal{L}_Y(f)g) + \mathcal{L}_X(f\mathcal{L}_Y(g)) - \mathcal{L}_Y(\mathcal{L}_X(f)g) - \mathcal{L}_Y(f\mathcal{L}_X(g)) \\ &= \mathcal{L}_X(\mathcal{L}_Y(f))g + \mathcal{L}_Y(f)\mathcal{L}_X(g) + \mathcal{L}_X(f)\mathcal{L}_Y(g) + f\mathcal{L}_X(\mathcal{L}_Y(g)) \\ &\quad - \mathcal{L}_Y(\mathcal{L}_X(f))g - \mathcal{L}_X(f)\mathcal{L}_Y(g) - \mathcal{L}_Y(f)\mathcal{L}_X(g) - f\mathcal{L}_Y(\mathcal{L}_X(g)) \\ &= \mathcal{L}_X(\mathcal{L}_Y(f))g + f\mathcal{L}_X(\mathcal{L}_Y(g)) - \mathcal{L}_Y(\mathcal{L}_X(f))g - f\mathcal{L}_Y(\mathcal{L}_X(g)) \\ &= [\mathcal{L}_X, \mathcal{L}_Y](f)g + f[\mathcal{L}_X, \mathcal{L}_Y](g) . \end{aligned}$$

32. Rappelons que, de façon générale, le symbole \circ désigne la composition de deux applications. Parfois, pour simplifier, nous utiliserons l'abus de notation XY au lieu de $X \circ Y$, ce qui ne veut pas dire que l'on sait multiplier les vecteurs mais simplement qu'on sait les composer en tant qu'opérateurs.

L'espace fonctionnel $\mathfrak{X}(M)$ des champs de vecteurs est ainsi naturellement muni de commutateurs $[X, Y]$ qui sont linéaires et antisymétriques (il est trivial de le vérifier). Ceux-ci satisfont la relation caractéristique, dite *identité de Jacobi* :

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$$

car

$$\begin{aligned} [X, [Y, Z]] &= XYZ - XZY - YZX + ZYX \\ [Y, [Z, X]] &= YZX - YXZ - ZXY + XZY \\ [Z, [X, Y]] &= ZXY - ZYX - XYZ + YXZ \end{aligned}$$

et quand on somme, les termes s'éliminent deux à deux.

Cette structure fondamentale des champs de vecteurs est la structure *d'algèbre de Lie* sur laquelle nous reviendrons constamment.

Comme nous le verrons à la section 7.2.12, l'opérateur de dérivée de Lie \mathcal{L}_X est généralisable aux entités tensorielles et aux formes différentielles que nous allons maintenant aborder.

6.3.4. Les 1-formes différentielles.

Les 1-formes différentielles ω sont les entités *duales* des champs de vecteurs tangents conçus comme opérateurs de dérivation et sont aussi appelées des *champs de vecteurs cotangents* : ce sont des formes linéaires sur les vecteurs tangents. Conformément à la dialectique "valeurs ponctuelles dans les fibres" VS "variation différentiable relativement à la base", elles ont des valeurs ω_x ou $\omega(x)$ (variante nototationnelle) en chaque point x de la base M .

Si $X = \sum_{i=1}^n X_i \frac{\partial}{\partial x_i}$ est un vecteur tangent, alors $\omega(X)$ est un nombre dépendant linéairement à la fois de X et de ω . Si $\left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right)$ est la base de l'espace tangent $T_x M$ de M en x associée aux coordonnées locales (x_1, \dots, x_n) , alors la base de l'espace cotangent dual $T_x^* M$ est la base duale (dx_1, \dots, dx_n) définie par $dx_j \left(\frac{\partial}{\partial x_i}\right) = \delta_{ji}$ (où δ_{ji} est le symbole de Kronecker $\delta_{ji} = 1$ si $i = j$ et $\delta_{ji} = 0$ sinon), et si $\omega = \sum_{j=1}^n \omega_j dx_j$ on a la formule (déjà utilisée dans un cas élémentaire dans le Vol I, chapitre 5, section 5.4.) :

$$\omega(X) = \sum_{i,j=1}^n \omega_j X_i \delta_{ji} = \sum_{i=1}^n \omega_i X_i = \langle \omega, X \rangle$$

$\langle \cdot, \cdot \rangle$ étant le couplage naturel entre covecteurs et vecteurs.

De façon plus précise, $\omega = \sum_{j=1}^n \omega_j dx_j$ est définie par les coefficients ω_j qui sont des fonctions différentiables $\omega_j(x)$ des coordonnées $x = (x_1, \dots, x_n)$ et, en chaque

point x , $\omega_x = \sum_{j=1}^n \omega_j(x) dx_j \in T_x^*M$ est une forme linéaire sur $T_x M$.³³ Les $T_x M$ se recollent en un fibré, le fibré cotangent de M , $\pi^* : T^*M \rightarrow M$ et ω est un champ, différentiable par rapport à x , de 1-formes $\omega_x \in T_x^*M$, autrement dit une section de la fibration π^* .

Si f est une observable, on peut lui associer une 1-forme différentielle $df = \sum_{j=1}^n \frac{\partial f}{\partial x_j} dx_j$ qui retrouve la formule leibnizienne classique. On a alors la dualité fondamentale :

$$df(X) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} X_i = X(f).$$

Remarque sur les notations. Dans les textes mathématiques vraiment rigoureux, on tient compte du type tensoriel, covariant ou contravariant, des entités tensorielles considérées et l'on indexe les composantes par des indices ou des exposants pour en tenir compte. Si par exemple les composantes des vecteurs tangents X_i sont indexées par des indices, les vecteurs tangents de base $e^i = \frac{\partial}{\partial x_i}$ doivent être indexés par des exposants et, en tant qu'entité tensorielle, un vecteur tangent s'écrira $X = \sum_{i=1}^n X_i e^i$. On pourra lui appliquer la règle de sommation (dite d'Einstein parce qu'elle a été abondamment utilisée en relativité générale) selon laquelle dans une expression $X_i e^i$ où le même index apparaît en indice et en exposant la somme \sum_i est implicite. Les dx_i duales des e^i sont des entités tensorielles e_i^* indexées par des indices et, en tant qu'entité tensorielle, une 1-forme $\omega = \sum_j \omega_j dx_j$ devrait donc s'écrire $\omega = \sum_j \omega^j e_j^* = \omega^j e_j^*$ avec des composantes ω^j indexées par des exposants. Nous serons moins rigoureux et ne tiendrons en général pas compte de ces conventions notationnelles. □

6.4. L'implémentation neuronale de la structure différentiable

Considérons maintenant un neurone dont le profil récepteur est, du moins approximativement, une dérivée de gaussienne DG . Comme nous l'avons vu dans le Vol I, chapitre 3., section 3.3. et rappelé dans l'Introduction, DG opère sur le signal optique I par la convolution $DG * I$. Rappelons que f et g étant deux fonctions sur \mathbb{R}^n , leur produit de convolution $f * g$ est l'intégrale $f * g(x) = \int_{\mathbb{R}^n} f(y) g(x - y) dy = \int_{\mathbb{R}^n} f(x - y) g(y) dy$ lorsque celle-ci est définie. Nous avons vu également qu'une

33. Si l'on traite X ponctuellement simplement comme un élément de l'espace tangent fibre $T_x M$ de base (e_i) , alors la base de l'espace cotangent dual T_x^*M est la base duale (ε_j) définie par $\varepsilon_j(e_i) = \delta_{ji}$ et $\omega_x = \sum_{j=1}^n \omega_j(x) \varepsilon_j$.

propriété fondamentale de la convolution est son comportement relativement aux dérivations :

$$\frac{\partial}{\partial x}(\varphi * \psi) = \frac{\partial \varphi}{\partial x} * \psi = \varphi * \frac{\partial \psi}{\partial x} .$$

Nous y reviendrons à la section 5 du chapitre 16.

Soit alors $A = G * I$ le signal I lissé à l'échelle définie par G . Il s'ensuit qu'un neurone de profil récepteur par exemple $\frac{\partial G}{\partial x}$ opère via :

$$A' = \frac{\partial G}{\partial x} * I = G * \frac{\partial I}{\partial x} = \frac{\partial}{\partial x}(G * I) = \frac{\partial A}{\partial x} .$$

Le profil $\frac{\partial G}{\partial x}$ opérant sur I fournit par conséquent la dérivée directionnelle $X(A)$ (avec $X = \frac{\partial}{\partial x}$).

Supposons pour simplifier que les colonnes d'orientation correspondent à cela (bien que nous sachions que les profils récepteurs en jeu sont des dérivées de gaussiennes plus compliquées). Les données neurophysiologiques ne semblent pas permettre de faire l'hypothèse que les structures *vectérielles* des espaces tangents et cotangents sont complètement neuralemement implémentées. En effet :

- (i) Bien que toutes les directions soient représentées dans une hypercolonne, il ne semble pas que ce soit le cas pour les vecteurs eux-mêmes.
- (ii) Même si l'implémentation de la structure vectorielle (addition et multiplication par des scalaires) est théoriquement possible, elle exigerait une masse de connexions qui ne semblent pas avoir été observées.
- (iii) La fonction principale des colonnes d'orientation est de représenter les contours. Or le vecteur tangent à une courbe en un point dépend de la paramétrisation et ne possède pas de signification géométrique intrinsèque alors que la direction tangente est, elle, une entité géométrique intrinsèque.

Il semble donc en définitive que ce qui est représenté par un pinwheel "au-dessus" de chaque point a du plan rétinien R ne soit pas le plan tangent $T_a R$ lui-même, mais bien plutôt sa *projectivisation*, que celle-ci soit modélisée par la droite projective \mathbb{P}^1 , ou par les orientations θ , ou par $\tan(\theta)$.

7. Élie Cartan et le calcul des formes différentielles

7.1. Une vision exceptionnelle

Approfondissons maintenant cette conception de la géométrie différentielle qui nous semble être la plus idoine pour la neurogéométrie. Sophus Lie (1842-1899) et Élie Cartan (1869-1951) sont centraux, non seulement pour eux-mêmes mais aussi

pour le contexte qu'ils ont partagé avec d'autres grands mathématiciens comme Ferdinand Frobenius (1849-1917), Friedrich Engel (1861-1941), Gaston Darboux (1842-1917), ou Édouard Goursat (1858-1936). Nous reviendrons brièvement à la section 2.2 du chapitre 3 sur quelques éléments biographiques de ces grands savants, et en particulier Élie Cartan.

Il y a chez Cartan une vision tout à fait exceptionnelle de ce que signifie l'intégration de données différentielles et une conception de l'espace dont la profondeur ne le cède qu'à celle de Riemann. Son œuvre est immense et pour s'en faire une idée, le mieux est de consulter la "Notice sur les travaux scientifiques" [98] rédigée par Cartan lui-même en 1939. On peut aussi se référer à l'hommage [114] publié en 1952 (l'année après sa disparition) par Shiing-Shen Chern et Claude Chevalley ainsi qu'au texte de Paulette Libermann [332]. Le lecteur intéressé pourra également consulter le Fonds Élie Cartan à l'Académie des Sciences, en particulier les Cahiers de Cartan³⁴.

Dans la mesure où la neurogéométrie repose sur l'hypothèse que certains algorithmes de géométrie différentielle sont neuralemement implémentés, elle présuppose une enquête approfondie sur *les conditions de possibilité* d'une telle hypothèse car celle-ci ne va pas du tout de soi. Une implémentation neuronale d'algorithmes différentiels soulève aujourd'hui autant de problèmes métaphysiques (en fait transcendants) qu'une implémentation mécanique d'algorithmes différentiels a pu le faire à l'époque de Newton et Leibniz. Elle demande de remonter jusqu'aux *sources* du calcul différentiel et d'en redérouler le cours jusqu'à en trouver des formulations alternatives qui soient implémentables dans des architectures fonctionnelles neuronales. Et c'est la théorie des formes différentielles de Cartan qui nous servira en quelque sorte de levier d'Archimède.

7.2. La notion de forme différentielle

Dans cette section, nous allons exposer quelques rudiments de la théorie avant de revenir au texte fondateur de Cartan de 1899. La littérature sur ce sujet est immense, aussi bien en mathématiques qu'en physique, et il est très facile de se documenter. Pour notre part, nous avons depuis le début utilisé les "bibles" que sont le Vladimir Arnold [20], le Ralph Abraham & Jerrold Marsden [2] et le Michael Spivak [511]. Parmi le nombre considérable de "Graduate texts" assez récents, le lecteur intéressé pourra consulter par exemple le Victor Guillemin [236] ou le David Bachman [23].

7.2.1. Formes multilinéaires antisymétriques.

La notion de forme différentielle généralise celle de 1-forme en introduisant une notion de *degré*. On note traditionnellement $\Omega^k M$ ou $\Lambda^k M$ les formes différentielles de degré k . Une k -forme différentielle $\omega \in \Omega^k M$ est d'abord un champ différentiable

34. Cf. le site <http://eliecartanpapers.ahp-numerique.fr> ainsi que le dossier d'Emmylou Haffner [239].

de k -formes *multilinéaires* sur les espaces tangents $T_x M \simeq \mathbb{R}^n$. Cela signifie que si ω_x est la valeur ponctuelle de ω en x , ω_x associe à tout k -uplet (X^1, \dots, X^k) de vecteurs tangents $X^i \in T_x M$ un réel $\omega_x(X^1, \dots, X^k)$ qui est linéaire en chacun des X^i .³⁵

Mais la caractéristique des k -formes différentielles est un comportement particulier par rapport aux *permutations* des X^i : on exige que les ω_x soient des formes multilinéaires *antisymétriques* (aussi dites *alternées*³⁶). Cela signifie que si σ est une permutation de $\{1, \dots, k\}$ de parité $p(\sigma)$, on a

$$\omega_x(X^{\sigma(1)}, \dots, X^{\sigma(k)}) = (-1)^{p(\sigma)} \omega_x(X^1, \dots, X^k) .$$

En particulier, si $X^i = X^j$ ($j \neq i$) alors $\omega_x(X^1, \dots, X^k) = 0$ puisque

$$\begin{aligned} \omega_x(X^1, \dots, X^i, \dots, X^j, \dots, X^k) &= -\omega_x(X^1, \dots, X^j, \dots, X^i, \dots, X^k) \\ &= -\omega_x(X^1, \dots, X^i, \dots, X^j, \dots, X^k) . \end{aligned}$$

Cette condition nécessaire est également suffisante.

7.2.2. Intégration des 1-formes.

Pour comprendre la pertinence de la notion d'antisymétrie dans le contexte du calcul différentiel il faut comprendre que les formes différentielles sont faites pour être intégrées et Cartan les a précisément introduites pour pouvoir développer une théorie de l'intégration géométrique, c'est-à-dire possédant un sens *intrinsèque*, indépendant du choix de coordonnées locales.

Revenons au cas le plus simple, à savoir celui de \mathbb{R} et reconsidérons ce que signifie l'intégration $\int f(x) dx$ d'une fonction différentiable $f(x)$.³⁷ \mathbb{R} est à lui-même une carte locale qui est en fait globale et il s'identifie par translation à son espace tangent $T_x \mathbb{R}$ en chacun de ses points x . Donc dx est la 1-forme de base et $f(x) dx$ est une 1-forme ω . Et dans son ancienne version intuitive (celle que nous avons apprise à l'école), $\int_a^b f(x) dx$ consiste à découper l'intervalle $[a, b]$ en un grand nombre N de petits intervalles fermés $\Delta_\ell = [x_\ell, x_{\ell+1}]$, à prendre des a_ℓ à l'intérieur de chaque Δ_ℓ , à considérer la somme $\sum_{\ell=1}^N f(a_\ell) \Delta_\ell$ (ce qui revient à approximer f par des fonctions en escalier), puis à faire tendre la taille maximale Δ des Δ_ℓ vers 0 et donc N vers l'infini. Une fois obtenue les $\int_a^b f(x) dx$ pour les fonctions C^∞ , on peut ensuite considérablement enrichir la classe des fonctions intégrables par diverses méthodes d'approximation. On en arrive ainsi à la mesure de Lebesgue dx .³⁸

35. Les X^i , $i = 1, \dots, k$, sont des vecteurs tangents et non pas des composantes X_j , $j = 1, \dots, n$, d'un vecteur tangent X . Les X^i sont de composantes X_j^i .

36. Les deux notions sont équivalentes dans notre contexte. Sur des corps de scalaires de caractéristique $\neq 0$ on est amené à introduire deux notions non équivalentes.

37. Pour des précisions, cf. par exemple Arnold [20] ou n'importe quelle encyclopédie.

38. Nous reviendrons sur la théorie de la mesure, la mesure de Lebesgue et l'intégration à la section 2 du chapitre 16. Disons ici simplement les choses suivantes. (i) sur \mathbb{R} , la mesure de

Mais Δ_ℓ peut être considéré comme la composante d'un vecteur tangent X_ℓ à \mathbb{R} en a_ℓ . Si ∂_x est la base de $T_x\mathbb{R}$, $X_\ell = \Delta_\ell\partial_x$ et donc, en considérant maintenant dx comme une 1-forme et non plus comme un accroissement infinitésimal, $dx(X_\ell) = \Delta_\ell dx(\partial_x) = \Delta_\ell$. On a par conséquent

$$\begin{aligned} \int_a^b f(x) dx &= \lim_{\Delta \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^{\ell:=N} f(a_\ell) \Delta_\ell = \lim_{\Delta \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^{\ell:=N} f(a_\ell) dx(X_\ell) \\ &= \lim_{\Delta \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^{\ell:=N} \omega(X_\ell) \end{aligned}$$

où ω est la 1-forme $f(x) dx$. On voit ainsi la relation entre le dx comme symbole d'un accroissement infinitésimal (mesure de Lebesgue) et le dx comme symbole de différentielle. L'intégrale $\int_a^b f(x) dx$ est interprétable non seulement comme l'intégrale d'une fonction par rapport à la mesure de Lebesgue dx mais également comme l'intégrale $\int_{[a,b]} \omega$ d'une 1-forme sur le segment $[a, b]$ de dimension 1.

Supposons maintenant que l'on parcourt l'intervalle $[a, b]$ avec une loi temporelle différentiable $x(t)$, $t \in [t_a, t_b]$, ce qui revient à faire un changement de variable. Alors $\int_a^b f(x) dx$ devient $\int_{t_a}^{t_b} f(x(t)) x'(t) dt$. C'est maintenant $[t_a, t_b]$ que l'on découpe en intervalles δ_ℓ correspondant à des vecteurs tangents $T_\ell = \delta_\ell \partial_t$ en des points t_ℓ situés dans les intervalles δ_ℓ et l'on obtient la même formule que précédemment mais avec maintenant $X_\ell = x'(t_\ell) T_\ell$, $x'(t)$ étant considéré comme l'application linéaire tangente de $x(t)$.

Il est immédiat de généraliser cette définition de l'intégrale d'une 1-forme à des 1-formes ω sur une variété M de dimension n . On considère des courbes différentiables paramétrées $\gamma(t)$, $t \in [t_a, t_b]$, dans M et on définit $\int_\gamma \omega$ de la même façon : $\int_\gamma \omega = \lim_{\Delta \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^{\ell:=N} \omega(X_\ell)$ où les X_ℓ sont les $\gamma'(t_\ell) T_\ell$. En termes de coordonnées locales on a $\omega = \sum_{j=1}^n \omega_j dx_j$ et l'on voit que lorsque l'on découpe γ en petits segments Δ_ℓ on doit considérer les *projections* des Δ_ℓ sur les axes des coordonnées et appliquer à chaque axe la formule valable en dimension 1.

Lebesgue d'un intervalle $P = [a, b]$ est sa longueur $\ell(P) = b - a$; (ii) sur \mathbb{R}^n qui est le produit direct de n exemplaires \mathbb{R}_i de \mathbb{R} , la mesure de Lebesgue d'un pavé P produit de n intervalles P_i des \mathbb{R}_i est le produit des longueurs $\ell(P_i)$ (i.e. le volume de P); (iii) l'intégrale de la fonction caractéristique $\chi(P)$ est la mesure de P ; (iv) l'intégrale d'une fonction que l'on peut obtenir par limite de combinaisons linéaires de $\chi(P)$ s'obtient par linéarité de l'intégrale et c'est là qu'interviennent les problèmes techniques dont nous reparlerons plus tard; (v) si M est une variété différentiable de dimension n (cf. section 4) recollant des cartes U_α , on généralise les notions de mesure de Lebesgue et d'intégration de fonctions mesurables en recollant les valeurs locales au moyen des difféomorphismes de changement de cartes.

7.2.3. *Intégration des k -formes et déterminants.*

En dimension 1 la notion d'antisymétrie n'intervient pas. Pour comprendre pourquoi et comment elle intervient à partir de la dimension $n = 2$, considérons le cas d'une 2-forme ω dans \mathbb{R}^2 . Quand on voudra l'intégrer sur un domaine $D \subset \mathbb{R}^2$ on décomposera D en N petits rectangles fermés D_ℓ d'aire $S_\ell \leq S$, on choisira des points a_ℓ à l'intérieur des D_ℓ et on définira $\int_D \omega = \lim_{S \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^{\ell:=N} f(a_\ell) S_\ell$, les $f(a_\ell)$ étant des coefficients associés aux valeurs ponctuelles de ω en les a_ℓ . Mais les D_ℓ sont définis par deux vecteurs tangents $X^\ell, Y^\ell \in T_{a_\ell} \mathbb{R}^2$ et il faut donc que l'aire S_ℓ puisse être considérée comme la valeur d'une 2-forme de base ω_0 qui mesure les aires S des parallélogrammes (X, Y) dans les $T_x \mathbb{R}^2$ au moyen de la formule $\omega_0(X, Y) = S$. En effet, dans ce cas, si $\omega = f(x) \omega_0$, on aura bien

$$\begin{aligned} \int_D \omega &= \lim_{S \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^{\ell:=N} f(a_\ell) S_\ell = \lim_{s \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^{\ell:=N} f(a_\ell) \omega_0(X^\ell, Y^\ell) \\ &= \lim_{S \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^{\ell:=N} \omega(X^\ell, Y^\ell) . \end{aligned}$$

Mais dans le plan $T_x \mathbb{R}^2$ de coordonnées (ξ, η) quelle est l'aire (orientée) $\omega_0(X, Y)$ du parallélogramme (X, Y) défini par les deux vecteurs $X = (X_\xi, X_\eta)$ et $Y = (Y_\xi, Y_\eta)$? Si $X = (X_\xi, 0)$ et $Y = (0, Y_\eta)$ c'est tout simplement $X_\xi Y_\eta$. C'est bien une forme bilinéaire : la linéarité par rapport à la multiplication est évidente et celle par rapport à l'addition est illustrée à la figure 15. Si $X = Y$, le parallélogramme dégénère et sa surface s'annule et $\omega_0(X, X) = 0$. Par conséquent, $\omega_0(X, X)$ est *antisymétrique*.

En fait, $\omega_0(X, Y)$ est le *déterminant* de la matrice composée des vecteurs lignes X et Y :

$$\omega_0(X, Y) = \det \begin{pmatrix} X_\xi & X_\eta \\ Y_\xi & Y_\eta \end{pmatrix} = X_\xi Y_\eta - X_\eta Y_\xi .$$

De façon générale, le volume orienté d'un parallélépipède construit sur n vecteurs X^1, \dots, X^n de \mathbb{R}^n est la n -forme multilinéaire antisymétrique égale au déterminant de la matrice des X^k .³⁹ C'est à cause de ce rôle des déterminants que l'antisymétrie intervient de façon constituante dans la définition des formes différentielles. Elle est imposée par le fait que lorsqu'une k -forme ω est intégrée sur un domaine D^k de dimension k il faut que l'intégration sur des sous-domaines $D^{<k}$ de dimension $< k$ s'annule car la mesure de Lebesgue $dx_1 \dots dx_k$ mesure des k -volumes et le k -volume des $D^{<k}$ est nul.

39. Rappelons que les X_i sont les composantes d'un même X alors que les X^k sont des X différents.

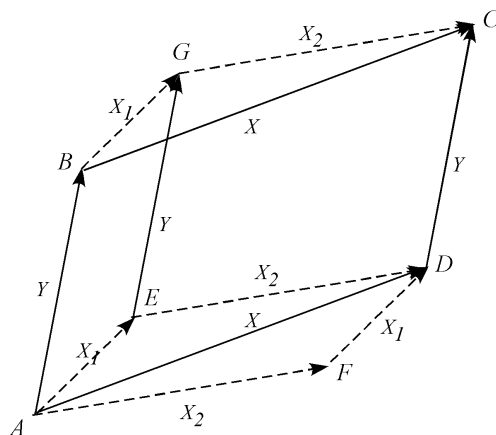


FIGURE 15. La linéarité par rapport à l'addition de l'aire (orientée) $\omega_0(X, Y)$ des parallélogrammes du plan. L'aire $\omega_0(X, Y)$ est celle du parallélogramme ADCBA. Si $X = X_1 + X_2$, cette aire est celle du polygone ADCGBA moins celle du triangle ADEA. Mais cette différence est la somme des aires $\omega_0(X_1, Y)$ du parallélogramme AEGBA et $\omega_0(X_2, Y)$ du parallélogramme EDCGE.

Ceci dit, la grande découverte de Cartan fut que le *calcul* différentiel peut se prolonger en un *calcul* sur les formes différentielles. Pour en préciser les règles, il a commencé par introduire le *produit extérieur*.

7.2.4. Le produit extérieur.

Le produit extérieur, noté $\omega \wedge \rho$, de deux formes $\omega \in \Omega^k$ et $\rho \in \Omega^r$ doit d'abord être additif pour les degrés, i.e. on veut que $\omega \wedge \rho \in \Omega^{k+r}$. Commençons par le produit extérieur de deux 1-formes $\omega, \rho \in \Omega^1$. $\omega \wedge \rho$ doit être une 2-forme, i.e. si X, Y sont deux vecteurs de $T_x M \simeq \mathbb{R}^n$, $(\omega \wedge \rho)(X, Y)$ doit être une forme bilinéaire antisymétrique. Étant donné ce qui vient d'être exposé, on pose par définition :

$(\omega \wedge \rho)(X, Y) = \det \begin{pmatrix} \omega(X) & \rho(X) \\ \omega(Y) & \rho(Y) \end{pmatrix}$, c'est-à-dire l'aire du parallélogramme de \mathbb{R}^2 défini par les vecteurs $(\omega(X), \rho(X))$ et $(\omega(Y), \rho(Y))$. Il s'agit bien d'une 2-forme et l'application $(\omega, \rho) \mapsto \omega \wedge \rho$ de $\Omega^1 \times \Omega^1 \rightarrow \Omega^2$ est elle-même bilinéaire et antisymétrique.

Si $\omega = \sum_{j=1}^n \omega_j dx_j$ et $\rho = \sum_{l=1}^n \rho_l dx_l$ alors, puisque $dx_j \wedge dx_j = 0$ et $dx_j \wedge dx_l = -dx_l \wedge dx_j$ par antisymétrie, $\omega \wedge \rho = \sum_{j < l} (\omega_j \rho_l - \omega_l \rho_j) dx_j \wedge dx_l$. Les 2-formes $dx_j \wedge dx_l$

($j < l$) forment une base de Ω^2 (dans la carte locale des x_j). Elles mesurent l'aire de la projection d'un parallélogramme (X, Y) de $T_x M \simeq \mathbb{R}^n$ sur les différents plans de coordonnées puisque $(dx_j \wedge dx_l)(X, Y) = \det \begin{pmatrix} X_j & X_l \\ Y_j & Y_l \end{pmatrix}$. La 2-forme $\omega_0(X, Y)$

introduite ci-dessus était donc tout simplement $dx \wedge dy$ dans le plan de coordonnées (x, y) .

Il est trivial de généraliser ces définitions et ces calculs à $(\omega, \rho) \mapsto \omega \wedge \rho$ de $\Omega^k \times \Omega^r \rightarrow \Omega^{k+r}$. Les bases sont maintenant les $dx_{j_1} \wedge \cdots \wedge dx_{j_k}$ pour $j_1 < \cdots < j_k \in \{1, \dots, k\}$ pour Ω^k et de même pour Ω^r et Ω^{k+r} . Elles forment bien une base car elles sont linéairement indépendantes et engendrent tout l'espace Ω . Toutefois, quand on fait un produit extérieur $\omega \wedge \rho$ on trouvera des $(dx_{j_1} \wedge \cdots \wedge dx_{j_k}) \wedge (dx_{l_1} \wedge \cdots \wedge dx_{l_r})$ qu'il faudra annuler si l'un des dx_j est égal à l'un des dx_l et réorganiser au moyen de permutations de façon à ce que les dx_j et les dx_l se retrouvent en ordre croissant. Cela introduit des facteurs $(-1)^p$ où p est la parité des permutations utilisées. Le produit extérieur est par conséquent *anticommutatif* : $\rho \wedge \omega = (-1)^{kr} \omega \wedge \rho$ et donc antisymétrique si k et r sont *impairs* et symétrique si k ou r est *pair*. En regardant bien comment fonctionnent les signes introduits par la parité des permutations utilisées, on montre qu'il est associatif et distributif.

Remarque. On notera que si k est impair, $\omega^2 = \omega \wedge \omega = -\omega \wedge \omega$ et donc $\omega^2 \equiv 0$. Mais si k est pair l'anticommutativité n'implique que l'identité $\omega \wedge \omega = \omega \wedge \omega$. et il faut calculer.⁴⁰ Si par exemple $\omega = dx \wedge dy + dz \wedge dt$ dans \mathbb{R}^4 , alors $\omega^2 = 2dx \wedge dy \wedge dz \wedge dt$ est le double de la forme volume (voir plus bas le point 3). \square

Remarque. Encore une fois, si l'on se restreint exclusivement à la fibre $T_x M$, on peut se placer dans \mathbb{R}^n avec une base (e_1, \dots, e_n) remplaçant la base $(\partial_{x_1}, \dots, \partial_{x_n})$ de $T_x M$ et la base duale (e_1^*, \dots, e_n^*) de $(\mathbb{R}^n)^*$ remplaçant la base (dx_1, \dots, dx_n) de $T_x^* M$. \square

Les k -formes de base $dx_{j_1} \wedge \cdots \wedge dx_{j_k}$ mesurent les volumes des projections sur les différents sous-espaces $(\partial_{x_{j_1}}, \dots, \partial_{x_{j_k}})$ de $T_x M \simeq \mathbb{R}^n$ des parallélépipèdes définis par k vecteurs tangents. Il y en a $C_n^k = \frac{n!}{k!(n-k)!}$ (coefficients du binôme) et donc en particulier n pour $k = 1$, $\frac{n(n-1)}{2}$ pour $k = 2$, 1 pour $k = n$.

On vérifie trivialement toutes les propriétés qu'on peut attendre d'un calcul multilinéaire antisymétrique. Citons-en quelques-unes.

1. D'abord on peut considérer les fonction $f(x) \in \mathcal{C}^\infty(M)$ comme des formes de degré 0, leur différentielle $df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i$ étant une 1-forme.
2. Si $k > n$ toute k -forme est identiquement nulle car tous les monômes de base contiennent deux fois un même dx_j .
3. Si $k = n$, il n'y a qu'un monôme de base $dx_1 \wedge \cdots \wedge dx_n$ et donc toute n -forme s'écrit en coordonnées locales $f(x) dx_1 \wedge \cdots \wedge dx_n$ avec $f(x) \in \mathcal{C}^\infty(M)$. C'est ce que nous avons utilisé plus haut avec $\omega = f(x) \omega_0 = f(x) dx \wedge dy$. La n -forme $dx_1 \wedge \cdots \wedge dx_n$ est appelée la "forme volume" car elle permet de mesurer le volume d'un parallélépipède de dimension n de \mathbb{R}^n .

40. Comme ω^2 est de degré $2k$, $\omega^2 \equiv 0$ si $2k > n$. Voir plus bas le point 2.

4. Si la k -forme ω s'écrit $\omega = \sum \omega_{j_1 \dots j_k} dx_{j_1} \wedge \dots \wedge dx_{j_k}$ alors ses coefficients sont $\omega_{j_1 \dots j_k} = \omega(e_{j_1}, \dots, e_{j_k})$.
5. Pour calculer la valeur de $\omega \wedge \rho$ sur une suite ordonnée \mathcal{X} de $(k+r)$ vecteurs (X^1, \dots, X^{k+r}) , on décompose de toutes les façons possibles \mathcal{X} en paires de sous-suites \mathcal{X}_k de k vecteurs et \mathcal{X}_r de r vecteurs, et on somme les produits $\omega(\mathcal{X}_k) \rho(\mathcal{X}_r)$ avec un facteur $(-1)^p$ où p est la parité de la permutation ramenant la suite ordonnée $(\mathcal{X}_k, \mathcal{X}_r)$ à la suite \mathcal{X} .
6. Les k -formes se comportent bien par applications linéaires. Si $L : \mathbb{R}^m \rightarrow \mathbb{R}^n$ est linéaire et si ω est une k -forme sur \mathbb{R}^n , son image inverse $L^*\omega$ sur \mathbb{R}^m est définie par $L^*\omega(X^1, \dots, X^k) = \omega(L(X^1), \dots, L(X^k))$. $L^*\omega$ est bien une k -forme. $L^* : \Omega^k(\mathbb{R}^n) \rightarrow \Omega^k(\mathbb{R}^m)$ est linéaire, préserve les produits extérieurs (i.e. $L^*(\omega \wedge \rho) = L^*\omega \wedge L^*\rho$) et est compatible à la composition au sens où, si $\mathbb{R}^m \xrightarrow{L_1} \mathbb{R}^n \xrightarrow{L_2} \mathbb{R}^p$ sont deux applications linéaires composables, $(L_2 \circ L_1)^* = L_1^* \circ L_2^* : \Omega^k(\mathbb{R}^p) \rightarrow \Omega^k(\mathbb{R}^m)$.⁴¹
7. Les k -formes se comportent également bien par applications différentiables $f : M \rightarrow N$ entre variétés différentiables. Si η est une k -forme différentielle sur N , son image réciproque par f est la k -forme $f^*\eta$ sur M définie par $(f^*\eta)_x(X^1, \dots, X^k) = \eta_{f(x)}(D_x f(X^1), \dots, D_x f(X^k))$ où $D_x f$ est l'application linéaire tangente de f en x .⁴²

7.2.5. Le produit intérieur.

À côté du produit extérieur, il existe aussi un “produit intérieur” naturel, aussi appelé “contraction”, entre k -formes ω et vecteurs tangents X . Il consiste à fixer l'un des arguments X de la k -forme pour obtenir une $(k-1)$ -forme $i_X \omega$. Plus précisément, $i_X \omega(X^1, \dots, X^{k-1}) = \omega(X, X^1, \dots, X^{k-1})$.⁴³ On notera

- (i) que $i_X df = df(X) = X(f) = \mathcal{L}_X f$,
- (ii) que i_X est linéaire pour la structure de $\mathcal{C}^\infty(M)$ -module de $\mathfrak{X}(M)$ car $i_{X+Y}\omega = i_X \omega + i_Y \omega$ et $i_{fX}\omega = f i_X \omega = i_X(f\omega)$,
- (iii) que $i_Y i_X \omega = -i_X i_Y \omega$ (et donc évidemment $i_X i_X \omega = 0$)
- (iv) et que $i_X(\omega \wedge \rho) = (i_X \omega) \wedge \rho + (-1)^k \omega \wedge (i_X \rho)$ si ω est une k -forme.

L'opération $i_X \omega$ est une dérivation *graduée*⁴⁴ de l'algèbre graduée des formes différentielles, ce qui signifie qu'elle ne satisfait la formule de Leibniz qu'à un signe près lié au degré.

41. En termes de théorie des catégories ce “bon” comportement s'appelle “fonctoriel”.

42. Rappelons que par rapport à des coordonnées locales (x_1, \dots, x_m) de M en x et (y_1, \dots, y_n) de N en $f(x)$ où f s'écrit $f(x_1, \dots, x_m) = (f_1(x_1, \dots, x_m), \dots, f_n(x_1, \dots, x_m))$ et par rapport à des bases de $T_x M$ et de $T_{f(x)} N$, $D_x f$ est l'application linéaire de matrice $\left(\frac{\partial f_i}{\partial x_j} \right)$.

43. Pour une 0-forme f , on pose $i_X f = 0$.

44. On dit aussi “antidérivation”.

7.2.6. *Intégration invariante et signification géométrique intrinsèque.*

Revenons alors sur le problème de l'intégration des k -formes sur une variété M de dimension n (avec $k \leq n$). Si $\omega = \sum \omega_{j_1 \dots j_k} dx_{j_1} \wedge \dots \wedge dx_{j_k}$ et si D est un domaine de M de dimension k , alors, comme nous l'avons expliqué précédemment, $\int_D \omega$ s'écrira comme l'intégrale multiple $\int_D \omega_{j_1 \dots j_k}(x) dx_{j_1} \dots dx_{j_k}$ et s'obtiendra en décomposant D en un grand nombre N de petits sous-domaines D_ℓ de k -volume $V_\ell \leq V$, en choisissant des points a_ℓ à l'intérieur des D_ℓ et en définissant $\int_D \omega$ comme la limite $\int_D \omega = \lim_{V \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^N f(a_\ell) V_\ell$, les coefficients $f(a_\ell)$ étant dérivés des valeurs ponctuelles de ω en les a_ℓ . Les D_ℓ sont définis par k vecteurs tangents $X^{\ell,1}, \dots, X^{\ell,k} \in T_{a_\ell} M$ et l'on a

$$\int_D \omega = \lim_{V \rightarrow 0, N \rightarrow \infty} \sum_{\ell=1}^N \omega(X^{\ell,1}, \dots, X^{\ell,k}) .$$

Mais faisons maintenant un changement de coordonnées locales, les coordonnées x_i devenant des fonctions C^∞ des nouvelles coordonnées y_j . Explicitons ce qui se passe pour une 2-forme $\omega = f(x_1, x_2) dx_1 \wedge dx_2$ sur \mathbb{R}^2 . Dans le calcul intégral classique, la formule de changement de variables pour l'intégrale double

$$\int_D f(x_1, x_2) dx_1 dx_2$$

est $\int_D f(x_1(y_1, y_2), x_2(y_1, y_2)) J(x_1, x_2) dy_1 dy_2$ où $J(x_1, x_2)$ est le *jacobien* du changement de variable, c'est-à-dire le déterminant de la matrice jacobienne

$$J = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \frac{\partial x_1}{\partial y_1} \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \frac{\partial x_2}{\partial y_1}$$

D'un autre côté, pour ω , le changement de variables donne

$$\begin{aligned} dx_1 \wedge dx_2 &= \left(\frac{\partial x_1}{\partial y_1} dy_1 + \frac{\partial x_1}{\partial y_2} dy_2 \right) \wedge \left(\frac{\partial x_2}{\partial y_1} dy_1 + \frac{\partial x_2}{\partial y_2} dy_2 \right) \\ &= \frac{\partial x_1}{\partial y_1} \frac{\partial x_2}{\partial y_2} dy_1 \wedge dy_2 + \frac{\partial x_1}{\partial y_2} \frac{\partial x_2}{\partial y_1} dy_2 \wedge dy_1 \\ &= \left(\frac{\partial x_1}{\partial y_1} \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \frac{\partial x_2}{\partial y_1} \right) dy_1 \wedge dy_2 \\ &= J dy_1 \wedge dy_2 \end{aligned}$$

et, exprimée dans les nouvelles coordonnées, l'intégrale $\int_D \omega = \int_D f(x_1, x_2) dx_1 dx_2$ devient $\int_D \omega = \int_D f(x_1(y_1, y_2), x_2(y_1, y_2)) J(x_1, x_2) dy_1 dy_2$. Autrement dit, les deux façons de changer de variables donnent exactement le *même* résultat.

Il est facile de généraliser aux k -formes sur M . Les opérations d'intégration et de changement de variables *commutent* : intégrer ω puis changer de variables dans

l'intégrale donne le même résultat que changer d'abord de variables dans ω puis intégrer. Grâce au lien étroit de leur propriétés d'antisymétrie et de leurs produits extérieurs avec les déterminants jacobiens, les formes différentielles peuvent ainsi être intégrées de façon *invariante*. C'est pourquoi elles confèrent une *signification géométrique intrinsèque* au calcul différentiel. Ce caractère intrinsèque est la raison de leur rôle dans les applications à la physique et, ici, à la neurogéométrie.

7.2.7. La dérivation extérieure.

Mais ce qui confère toute leur puissance aux formes différentielles est l'opération dite de *dérivation extérieure*. Nous en avons rencontré plus haut l'exemple le plus simple quand nous avons traité les fonctions $f(x) \in \mathcal{C}^\infty(M)$ comme des formes de degré 0 et avons noté que leurs différentielles $df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i$ sont des 1-formes.

Le point essentiel et fondamentalement nouveau est que nous ne considérons plus désormais les k -formes $\omega = \sum \omega_{j_1 \dots j_k}(x) dx_{j_1} \wedge \dots \wedge dx_{j_k}$ ponctuellement dans les fibres avec des coefficients $\omega_{j_1 \dots j_k}(x)$ qui sont des valeurs ponctuelles en x , mais *localement*, c'est-à-dire en tenant compte du fait que les coefficients $\omega_{j_1 \dots j_k}(x)$ sont des fonctions $\in \mathcal{C}^\infty(M)$ possédant des différentielles $d\omega_{j_1 \dots j_k}(x)$. On peut donc définir tout naturellement la dérivée extérieure $d\omega$ de ω par la formule

$$d\omega = \sum d\omega_{j_1 \dots j_k}(x) \wedge dx_{j_1} \wedge \dots \wedge dx_{j_k} .$$

Dans chaque composante, les dérivées partielles $\frac{\partial \omega_{j_1 \dots j_k}}{\partial x_{j_i}}$ ($i = 1, \dots, k$) donnent des termes $\frac{\partial \omega_{j_1 \dots j_k}}{\partial x_{j_i}} dx_{j_i} \wedge dx_{j_1} \wedge \dots \wedge dx_{j_k}$ qui s'annulent puisque dx_{j_i} est l'un des $dx_{j_1}, \dots, dx_{j_k}$. Les autres termes doivent être réorganisés au moyen de permutations appropriées et l'on obtient une $(k+1)$ -forme. Ainsi, pour tout k , on obtient une application $d : \Omega^k \rightarrow \Omega^{k+1}$.

Par exemple si $\omega = \omega_1(x) dx_1 + \omega_2(x) dx_2 + \omega_3(x) dx_3$ est une 1-forme sur \mathbb{R}^3 (avec $x = (x_1, x_2, x_3)$), on trouve

$$\begin{aligned} d\omega &= \begin{cases} \left(\frac{\partial\omega_1(x)}{\partial x_1} dx_1 + \frac{\partial\omega_1(x)}{\partial x_2} dx_2 + \frac{\partial\omega_1(x)}{\partial x_3} dx_3 \right) \wedge dx_1 \\ + \\ \left(\frac{\partial\omega_2(x)}{\partial x_1} dx_1 + \frac{\partial\omega_2(x)}{\partial x_2} dx_2 + \frac{\partial\omega_2(x)}{\partial x_3} dx_3 \right) \wedge dx_2 \\ + \\ \left(\frac{\partial\omega_3(x)}{\partial x_1} dx_1 + \frac{\partial\omega_3(x)}{\partial x_2} dx_2 + \frac{\partial\omega_3(x)}{\partial x_3} dx_3 \right) \wedge dx_3 \end{cases} \\ &= \frac{\partial\omega_1(x)}{\partial x_2} dx_2 \wedge dx_1 + \frac{\partial\omega_1(x)}{\partial x_3} dx_3 \wedge dx_1 + \frac{\partial\omega_2(x)}{\partial x_1} dx_1 \wedge dx_2 + \frac{\partial\omega_2(x)}{\partial x_3} dx_3 \wedge dx_2 \\ &+ \frac{\partial\omega_3(x)}{\partial x_1} dx_1 \wedge dx_3 + \frac{\partial\omega_3(x)}{\partial x_2} dx_2 \wedge dx_3 \\ &= \left(\frac{\partial\omega_2(x)}{\partial x_1} - \frac{\partial\omega_1(x)}{\partial x_2} \right) dx_1 \wedge dx_2 + \left(\frac{\partial\omega_3(x)}{\partial x_2} - \frac{\partial\omega_2(x)}{\partial x_3} \right) dx_2 \wedge dx_3 \\ &+ \left(\frac{\partial\omega_1(x)}{\partial x_3} - \frac{\partial\omega_3(x)}{\partial x_1} \right) dx_3 \wedge dx_1 . \end{aligned}$$

On vérifie que d possède les propriétés que l'on attend d'une dérivation : la \mathbb{R} -linéarité et une formule de Leibniz pour le produit : $d(\omega \wedge \rho) = d\omega \wedge \rho + (-1)^k \omega \wedge d\rho$ pour $\omega \in \Omega^k$ et $\rho \in \Omega^r$. Mais comme un signe $(-1)^k$ lié au degré de ω intervient, l'opérateur d est une dérivation *graduée* (une antidérivation) de l'algèbre graduée des formes différentielles. On vérifie également que d se comporte bien relativement aux applications différentiables $f : M \rightarrow N$ entre variétés, au sens où si $\omega \in \Omega^k(N)$, $d(f^*(\omega)) = f^*(d\omega)$.

Il existe tout un ensemble de formules utilisables dans les calculs. Citons par exemple cette relation fondamentale entre la 2-forme $d\omega$ d'une 1-forme ω et l'interprétation en termes d'opérateurs de dérivation des vecteurs tangents X, Y qui en sont les arguments :

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]) . \quad ((D))$$

En effet, si $\omega = \sum_j \omega_j(x) dx_j$, on a $d\omega = \sum_{i < j} \left(\frac{\partial\omega_j}{\partial x_i} - \frac{\partial\omega_i}{\partial x_j} \right) dx_i \wedge dx_j$. Si par ailleurs $X = \sum_k X_k \frac{\partial}{\partial x_k}$ et $Y = \sum_l Y_l \frac{\partial}{\partial x_l}$, on a $dx_i \wedge dx_j(X, Y) = X_i Y_j - X_j Y_i$ et $[X, Y]_j = \sum_k \left(X_k \frac{\partial Y_j}{\partial x_k} - Y_k \frac{\partial X_j}{\partial x_k} \right)$. On obtient donc d'un côté (en intervertissant les indices dans les deux derniers termes)

$$\begin{aligned}
d\omega(X, Y) &= \sum_{i < j} \left(\frac{\partial \omega_j}{\partial x_i} - \frac{\partial \omega_i}{\partial x_j} \right) (X_i Y_j - X_j Y_i) \\
&= \sum_{i < j} \frac{\partial \omega_j}{\partial x_i} (X_i Y_j) - \sum_{i < j} \frac{\partial \omega_j}{\partial x_i} (X_j Y_i) - \sum_{i < j} \frac{\partial \omega_i}{\partial x_j} (X_i Y_j) + \sum_{i < j} \frac{\partial \omega_i}{\partial x_j} (X_j Y_i) \\
&= \sum_{i, j} \frac{\partial \omega_j}{\partial x_i} (X_i Y_j) - \sum_{i, j} \frac{\partial \omega_j}{\partial x_i} (X_j Y_i)
\end{aligned}$$

et d'un autre côté

$$\begin{aligned}
X(\omega(Y)) &= \sum_k X_k \frac{\partial}{\partial x_k} \left(\sum_j \omega_j Y_j \right) = \sum_{k, j} X_k \left(\frac{\partial \omega_j}{\partial x_k} Y_j + \omega_j \frac{\partial Y_j}{\partial x_k} \right) \\
&= \sum_{k, j} \frac{\partial \omega_j}{\partial x_k} (X_k Y_j) + \sum_{k, j} \omega_j X_k \frac{\partial Y_j}{\partial x_k} \\
Y(\omega(X)) &= \sum_{l, j} Y_l \frac{\partial}{\partial x_l} \left(\sum_j \omega_j X_j \right) = \sum_{l, j} Y_l \left(\frac{\partial \omega_j}{\partial x_l} X_j + \omega_j \frac{\partial X_j}{\partial x_l} \right) \\
&= \sum_{l, j} \frac{\partial \omega_j}{\partial x_l} (Y_l X_j) + \sum_{l, j} \omega_j Y_l \frac{\partial X_j}{\partial x_l} \\
\omega([X, Y]) &= \sum_j \omega_j \left(\sum_k \left(X_k \frac{\partial Y_j}{\partial x_k} - Y_k \frac{\partial X_j}{\partial x_k} \right) \right) \\
&= \sum_{j, k} \omega_j X_k \frac{\partial Y_j}{\partial x_k} - \sum_{j, k} \omega_j Y_k \frac{\partial X_j}{\partial x_k}
\end{aligned}$$

et l'on vérifie immédiatement la formule.

Pour une 2-forme ρ , on obtient de même la formule (un peu plus compliquée)

$$\begin{aligned}
d\rho(X, Y, Z) &= X\rho(Y, Z) + Y\rho(Z, X) + Z\rho(X, Y) \\
&\quad - \rho([X, Y], Z) - \rho([Y, Z], X) - \rho([Z, X], Y)
\end{aligned}$$

et pour une k -forme

$$\begin{aligned}
d\omega(X^0, X^1, \dots, X^k) &= \sum_{i=0}^{i=k} (-1)^i \mathcal{L}_{X^i} \left(\omega \left(X^0, \dots, \widehat{X^i}, \dots, X^k \right) \right) \\
&\quad + \sum_{0 \leq i < j \leq k} (-1)^{i+j} \omega \left([X^i, X^j], X^0, \dots, \widehat{X^i}, \dots, \widehat{X^j}, \dots, X^k \right).
\end{aligned}$$

Cette définition de la dérivée extérieure $d\omega$ repose sur les propriétés *symboliques* du calcul différentiel. Dans [20], Vladimir Arnold en donne une version “numérique” en termes d’infinitésimales classiques $\varepsilon \rightarrow 0$ (cf. plus haut section 6.2.1 $\delta x = \varepsilon X$).⁴⁵ Elle revient à montrer, pour des domaines P *infinitésimaux*, la formule $\int_{\partial P} \omega = \int_P d\omega$.

Soit $\varphi : U' \rightarrow U$ une carte locale de M au voisinage de x , i.e. un difféomorphisme d’un voisinage U' de 0 dans \mathbb{R}^n avec un voisinage U de x dans M . Soit ω une k -forme et (X^1, \dots, X^{k+1}) $k+1$ vecteurs tangents de $T_x M$. On veut définir $d\omega(X^1, \dots, X^{k+1})$. Pour cela on dispose de deux choses. D’abord, l’application linéaire tangente $D\varphi$ de φ en 0 étant un isomorphisme entre les espaces vectoriels $T_0\mathbb{R}^n \simeq \mathbb{R}^n$ et $T_x M$, l’inverse $(D\varphi)^{-1}$ transforme les $X^i \in T_x M$ en $\tilde{X}^i \in T_0\mathbb{R}^n \simeq \mathbb{R}^n$. Quitte à multiplier les X^i par un scalaire ε assez petit, on peut supposer que, *considérés comme vecteurs de \mathbb{R}^n* au moyen de l’identification canonique $T_0\mathbb{R}^n = \mathbb{R}^n$, les \tilde{X}^i sont dans U' et y définissent un parallélépipède *linéaire* P' . En appliquant φ on obtient alors un parallélépipède différentiable P dans U (qui peut être très tordu). Le bord ∂P de P (nous allons préciser cette notion plus bas à la section 7.2.10) est le k -domaine de U recollant les sous-variétés images des faces de P' et on peut y intégrer ω qui est une k -forme. Soit $\Delta(X^1, \dots, X^{k+1}) = \int_{\partial P} \omega$ l’“accroissement” ainsi défini. Il est facile de voir qu’il est antisymétrique. Les $\Delta(\varepsilon X^1, \dots, \varepsilon X^{k+1})$ pour $\varepsilon \rightarrow 0$ sont des infiniment petits d’ordre $k+1$ et l’on montre que $d\omega(X^1, \dots, X^{k+1})$ est la limite des $\Delta(\varepsilon X^1, \dots, \varepsilon X^{k+1})$ pour $\varepsilon \rightarrow 0$ au sens où $\Delta(\varepsilon X^1, \dots, \varepsilon X^{k+1}) \sim \varepsilon^{k+1} d\omega(X^1, \dots, X^{k+1})$.

Pour bien comprendre ce qui se passe reprenons (en le complétant) un exemple très simple d’Arnold. Soit $\omega = \omega_1(x_1, x_2) dx_1 + \omega_2(x_1, x_2) dx_2$ une 1-forme sur \mathbb{R}^2 et $X = (X_1, X_2), Y = (Y_1, Y_2) \in T_0\mathbb{R}^2 = \mathbb{R}^2$ assez petits (i.e. de la forme εX^0 et εY^0 avec $\varepsilon \rightarrow 0$) et de composantes $X_1 = dx_1(X), X_2 = dx_2(X), Y_1 = dx_1(Y), Y_2 = dx_2(Y)$. On a par définition $d\omega = \left(\frac{\partial \omega_2(x_1, x_2)}{\partial x_1} - \frac{\partial \omega_1(x_1, x_2)}{\partial x_2} \right) dx_1 \wedge dx_2$. Si, considérés comme vecteurs de \mathbb{R}^2 d’origine O , X et Y sont les segments orientés $X = OA, Y = OC$, le parallélogramme $P = OABCO$ construit sur X et Y a ses côtés paramétrés par $OA = tX, AB = X + tY, CB = Y + tX, OC = tY, t \in [0, 1]$. Intégrons alors ω sur ∂P . Sur OA, AB, BC, CD , on a respectivement $dx_1 = X_1 dt,$

45. Rappelons (section 6.2.2) qu’il n’existe pas d’infinitésimales numériques individuels dans \mathbb{R} puisque \mathbb{R} est archimédien. Nous appelons ici infinitésimale numérique un petit nombre $\varepsilon \rightarrow 0$.

$Y_1 dt, -X_1 dt, -Y_1 dt, dx_2 = X_2 dt, Y_2 dt, -X_2 dt, -Y_2 dt$ et donc

$$\begin{aligned} \Delta(X, Y) &= \int_{\partial P} \omega = \int_0^1 (\omega_1(tX_1, tX_2) X_1 + \omega_2(tX_1, tX_2) X_2) dt + \\ &\quad \int_0^1 (\omega_1(X_1 + tY_1, X_2 + tY_2) Y_1 + \omega_2(X_1 + tY_1, X_2 + tY_2) Y_2) dt \\ &\quad - \int_0^1 (\omega_1(Y_1 + tX_1, Y_2 + tX_2) X_1 + \omega_2(Y_1 + tX_1, Y_2 + tX_2) X_2) dt \\ &\quad - \int_0^1 (\omega_1(tY_1, tY_2) Y_1 + \omega_2(tY_1, tY_2) Y_2) dt . \end{aligned}$$

En développant au premier ordre les différences

$$\begin{cases} \omega_1(X_1 + tY_1, X_2 + tY_2) - \omega_1(tY_1, tY_2), \\ \omega_2(X_1 + tY_1, X_2 + tY_2) - \omega_2(tY_1, tY_2), \\ \omega_1(Y_1 + tX_1, Y_2 + tX_2) - \omega_1(tX_1, tX_2), \\ \omega_2(Y_1 + tX_1, Y_2 + tX_2) - \omega_2(tX_1, tX_2), \end{cases}$$

on trouve (avec des dérivées partielles prises en 0)

$$\omega_1(X_1 + tY_1, X_2 + tY_2) - \omega_1(tY_1, tY_2) \sim \frac{\partial \omega_1}{\partial x_1}(0) X_1 + \frac{\partial \omega_1}{\partial x_2}(0) X_2 = X \frac{\partial \omega_1}{\partial X}(0)$$

$$\omega_2(X_1 + tY_1, X_2 + tY_2) - \omega_2(tY_1, tY_2) \sim \frac{\partial \omega_2}{\partial x_1}(0) X_1 + \frac{\partial \omega_2}{\partial x_2}(0) X_2 = X \frac{\partial \omega_2}{\partial X}(0)$$

$$\omega_1(Y_1 + tX_1, Y_2 + tX_2) - \omega_1(tX_1, tX_2) X_1 \sim \frac{\partial \omega_1}{\partial x_1}(0) Y_1 + \frac{\partial \omega_1}{\partial x_2}(0) Y_2 = Y \frac{\partial \omega_1}{\partial Y}(0)$$

$$\omega_2(Y_1 + tX_1, Y_2 + tX_2) - \omega_2(tX_1, tX_2) X_1 \sim \frac{\partial \omega_2}{\partial x_1}(0) Y_1 + \frac{\partial \omega_2}{\partial x_2}(0) Y_2 = Y \frac{\partial \omega_2}{\partial Y}(0)$$

et par conséquent

$$\begin{aligned} \int_{\partial P} \omega &\sim \int_0^1 \left(\frac{\partial \omega_1}{\partial x_1}(0) X_1 + \frac{\partial \omega_1}{\partial x_2}(0) X_2 \right) Y_1 dt + \int_0^1 \left(\frac{\partial \omega_2}{\partial x_1}(0) X_1 + \frac{\partial \omega_2}{\partial x_2}(0) X_2 \right) Y_2 dt \\ &\quad - \int_0^1 \left(\frac{\partial \omega_1}{\partial x_1}(0) Y_1 + \frac{\partial \omega_1}{\partial x_2}(0) Y_2 \right) X_1 dt - \int_0^1 \left(\frac{\partial \omega_2}{\partial x_1}(0) Y_1 + \frac{\partial \omega_2}{\partial x_2}(0) Y_2 \right) X_2 dt \\ &= \int_0^1 \left(\frac{\partial \omega_1}{\partial x_2}(0) - \frac{\partial \omega_2}{\partial x_1}(0) \right) (X_2 Y_1 - Y_2 X_1) dt . \end{aligned}$$

Au premier ordre, on peut traiter l'intégrande comme constant et le sortir de l'intégrale et, comme $\int_0^1 dt = 1$, on obtient $\int_{\partial P} \omega \sim \left(\frac{\partial \omega_1}{\partial x_2}(0) - \frac{\partial \omega_2}{\partial x_1}(0) \right) (X_2 Y_1 - Y_2 X_1)$.

Mais $X_2Y_1 - Y_2X_1 = (dx_2 \wedge dx_1)(X, Y)$ et donc

$$\Delta(X, Y) = \int_{\partial P} \omega \sim - \left(\frac{\partial \omega_1}{\partial x_2}(0) - \frac{\partial \omega_2}{\partial x_1}(0) \right) (dx_1 \wedge dx_2)(X, Y) = d\omega(X, Y)$$

au point 0.

On voit bien ce qui se passe intuitivement. P est un $(k+1)$ -parallélépipède infinitésimal de côtés $\delta^i = \varepsilon X^i$ ($i = 1, \dots, k+1$) dont le $(k+1)$ -volume $\text{Vol}_k(P) = \varepsilon^{k+1} \prod_i \|X^i\|$ est un infinitésimal d'ordre $k+1$. Ses faces F_m sont des k -parallélépipèdes infinitésimaux de k -volume d'ordre k . Pour intégrer ω sur ∂P on choisit la valeur de ω en des points a_m intérieurs aux F_m (par exemple les centres) et on calcule la somme algébrique (algébrique à cause des signes \pm imposés par la cohérence des orientations) des $\omega(a_m) \text{Vol}_k(F_m)$.⁴⁶ Mais, précisément pour des raisons de cohérence d'orientation, les faces opposées F_m et $F_{m'}$ donnent

$$(\omega(a_m) - \omega(a_{m'})) \text{Vol}(F_m)$$

car $\text{Vol}_k(F_{m'}) = \text{Vol}_k(F_m)$. Le développement au premier ordre de $\omega(a_m) - \omega(a_{m'})$ introduit une dérivée partielle de ω dans la direction de P qui manque à ∂P , dérivée partielle dont la valeur est prise au centre du segment $a_m a_{m'}$. En sommant sur les paires $(F_m, F_{m'})$ on obtient $\Omega \text{Vol}_k(P)$ où Ω est un facteur antisymétrique construit à partir des dérivées partielles de ω dans les directions X^i . Ce qui donne $\int_P d\omega$.

Cette définition plus classique et "numérique" de $d\omega$ équivaut donc, au niveau infinitésimal qui est le sien, à la formule $\int_{\partial P} \omega = \int_P d\omega$ dont la version *globale* est le célèbre *théorème de Stokes* qui sera explicité plus bas à la section 7.2.11.

7.2.8. La nilpotence $d^2 = 0$.

Abordons maintenant la propriété la plus fondamentale de la dérivation extérieure qui est sa *nilpotence* : $d^2\omega = 0$. Elle est une conséquence directe de l'antisymétrie des formes couplée à la symétrie des dérivées partielles secondes des fonctions coefficients⁴⁷. Vérifions-la pour la 1-forme $\omega = \omega_1(x) dx_1 + \omega_2(x) dx_2 + \omega_3(x) dx_3$ considérée plus haut.

46. $\text{Vol}_k =$ volume d'un domaine de dimension k .

47. i.e. $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$.

$$\begin{aligned}
d^2\omega &= d \left(\begin{aligned} &\left(\frac{\partial\omega_2(x)}{\partial x_1} - \frac{\partial\omega_1(x)}{\partial x_2} \right) dx_1 \wedge dx_2 + \left(\frac{\partial\omega_3(x)}{\partial x_2} - \frac{\partial\omega_2(x)}{\partial x_3} \right) dx_2 \wedge dx_3 \\ &+ \left(\frac{\partial\omega_1(x)}{\partial x_3} - \frac{\partial\omega_3(x)}{\partial x_1} \right) dx_3 \wedge dx_1 \end{aligned} \right) \\
&= \frac{\partial}{\partial x_3} \left(\frac{\partial\omega_2(x)}{\partial x_1} - \frac{\partial\omega_1(x)}{\partial x_2} \right) dx_3 \wedge dx_1 \wedge dx_2 \\
&\quad + \frac{\partial}{\partial x_1} \left(\frac{\partial\omega_3(x)}{\partial x_2} - \frac{\partial\omega_2(x)}{\partial x_3} \right) dx_1 \wedge dx_2 \wedge dx_3 \\
&\quad + \frac{\partial}{\partial x_2} \left(\frac{\partial\omega_1(x)}{\partial x_3} - \frac{\partial\omega_3(x)}{\partial x_1} \right) dx_2 \wedge dx_3 \wedge dx_1 \\
&= \left(\begin{aligned} &\frac{\partial}{\partial x_3} \left(\frac{\partial\omega_2(x)}{\partial x_1} \right) - \frac{\partial}{\partial x_1} \left(\frac{\partial\omega_2(x)}{\partial x_3} \right) + \frac{\partial}{\partial x_2} \left(\frac{\partial\omega_1(x)}{\partial x_3} \right) - \frac{\partial}{\partial x_3} \left(\frac{\partial\omega_1(x)}{\partial x_2} \right) \\ &+ \frac{\partial}{\partial x_1} \left(\frac{\partial\omega_3(x)}{\partial x_2} \right) - \frac{\partial}{\partial x_2} \left(\frac{\partial\omega_3(x)}{\partial x_1} \right) \end{aligned} \right) dx_1 \wedge dx_2 \wedge dx_3 \\
&= 0 .
\end{aligned}$$

7.2.9. Formes et cohomologie.

Notons $d^{(k)} : \Omega^k \rightarrow \Omega^{k+1}$, l'opérateur d en degré k . On obtient donc la suite

$$\Omega^0 \xrightarrow{d^{(0)}} \Omega^1 \longrightarrow \dots \longrightarrow \Omega^{k-1} \xrightarrow{d^{(k-1)}} \Omega^k \xrightarrow{d^{(k)}} \Omega^{k+1} \longrightarrow \dots \longrightarrow \Omega^{n-1} \xrightarrow{d^{(n-1)}} \Omega^n .$$

Les k -formes $\omega \in \ker(d^{(k)})$ (le noyau de $d^{(k)}$) sont les formes appelées *fermées* $d\omega = 0$ et les k -formes $\omega \in \text{im}(d^{(k-1)})$ (l'image de $d^{(k-1)}$) sont les formes appelées *exactes* $\omega = d\alpha$ ($\alpha \in \Omega^{k-1}$). La nilpotence $d^2 = 0$ signifie que $\text{im}(d^{(k-1)}) \subset \ker(d^{(k)})$, autrement dit que toute forme exacte est fermée. Le quotient $H^k = \ker(d^{(k)}) / \text{im}(d^{(k-1)})$ s'appelle le k -ème groupe de *cohomologie* de M .

Remarque. Il faut faire attention à des erreurs que peuvent induire des abus de notation lorsqu'on traite une fonction multivaluée comme une vraie fonction univaluée. L'exemple scolaire le plus élémentaire est celui de l'angle θ sur le cercle \mathbb{S}^1 . θ n'est pas périodique et n'est définie que modulo 2π si on la traite comme fonction sur \mathbb{R} le revêtement universel de \mathbb{S}^1 . Si on veut la traiter comme fonction univaluée sur \mathbb{S}^1 alors il faut recouvrir \mathbb{S}^1 par (au moins) deux cartes locales et recoller les deux restrictions de θ sur ces deux cartes. Si on veut utiliser θ comme fonction de base sur \mathbb{S}^1 alors il faut prendre les *deux* fonctions trigonométriques $\sin(\theta)$ et $\cos(\theta)$. Bref, malgré la notation, $d\theta$ n'est pas une différentielle exacte. Évidemment, elle est fermée, et donc localement exacte, car $d^2\theta = 0$ (puisque $d(2\pi) = 0$), mais elle n'est pas globalement exacte. En revanche $d(\sin(\theta)) = \cos(\theta)d\theta$ et $d(\cos(\theta)) = -\sin(\theta)d\theta$ sont bien globalement exactes. \square

Nous allons maintenant voir pourquoi les formes différentielles sont aussi appelées "cochaînes", l'opérateur de différentiation d l'opérateur "cobord", les k -formes

fermées des “cocycles” et les k -formes exactes des “cobords”. Cette terminologie provient d’une *dualité* avec les chaînes, les cycles, les bords et l’homologie qui ont trait à des propriétés *géométriques* de M relevant de ce qu’on appelle la “*topologie algébrique*”.

C’est la théorie de l’intégration qui induit cette dualité fondamentale.

7.2.10. Chaînes, cycles, bords et homologie.

Revenons donc encore une fois sur l’intégration des k -formes. Lorsque nous avons défini plus haut $\int_D \omega$, nous n’avons pas été très précis sur la nature du domaine D de dimension k dans M . Précisons donc un peu les choses tout en restant malgré tout très élémentaire et intuitif. Les théories qui sont en jeu sont la topologie dite d’abord “combinatoire” puis “algébrique” et l’homologie développées à partir de la fin du XIX^e siècle par Henri Poincaré, puis, à partir des années 1920, par James W. Alexander (1888-1971), Leopold Vietoris (1891-2002!, proche de Brouwer), Heinz Hopf (1894-1971), etc.

1. On part de k -polyèdres convexes bornés et orientés P de \mathbb{R}^k et on les envoie dans M par des applications différentiables f , obtenant ainsi des domaines $D = f(P)$. Comme les k -formes se comportent bien par rapport aux f , on peut alors définir $\int_D \omega = \int_P f^*(\omega)$ et calculer cette dernière intégrale comme ci-dessus. L’orientation positive de P est celle de \mathbb{R}^k et, si on la change, l’intégrale change de signe.
2. Si les P contiennent leurs faces, celles-ci sont des $(k-1)$ -polyèdres qui se recollent eux-mêmes le long de leurs $(k-1)$ -faces. On a donc des relations d’*incidence* possibles entre polyèdres de différentes dimensions. Ici s’introduit une relation fondamentale entre une intuition géométrique (incidence) et une opération algébrique (somme). Considérons par exemple le 2-polyèdre donné par un quadrilatère orienté $P = ABCDA$ dans \mathbb{R}^2 . Il a 4 faces incidentes qui sont les segments orientés AB, BC, CD, DA . Ces segments sont des 1-polyèdres, mais ils ne sont pas indépendants : ils se concatènent en un polygone qui est le périmètre de P , c’est-à-dire son *bord* ∂P . ∂P est plus que le simple ensemble $\{AB, BC, CD, DA\}$ car ses éléments sont reliés par des relations. Cette relation est *à la fois* une relation topologique d’adhérence (les segments incidents appartiennent à la frontière de P) et une relation algébrique de somme. On peut écrire $\partial P = AB + BC + CD + DA$ en tenant compte en plus de l’incidence. La somme algébrique ici une union ensembliste qui est l’union disjointe des intérieurs des segments à laquelle on rajoute les 0-polyèdres (points) que sont les extrémités des segments et qui articulent ceux-ci entre eux. À partir d’Alexander, il faudra un certain temps pour bien comprendre la structure de groupe additif intriquée ici avec la géométrie.

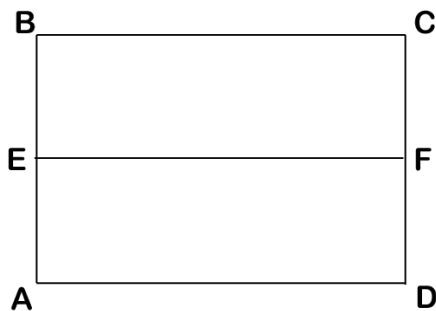


FIGURE 16. Le bord d'une chaîne.

3. On peut ainsi décomposer les k -polyèdres en sous k -polyèdres jusqu'à en arriver à des k -simplexes qui sont les fermetures convexes de $k + 1$ points linéairement indépendants et les plus simples des k -polyèdres (la généralisation en dimension k des triangles en dimension 2 et des tétraèdres en dimension 3). Si l'on n'utilise que des sommets de P la combinatoire est pauvre (par exemple on ne peut décomposer un quadrilatère qu'en deux triangles en utilisant une diagonale). Mais si l'on introduit des points de P comme nouveaux sommets possibles alors la combinatoire devient très riche (par exemple en reliant un point interne aux 4 sommets d'un quadrilatère on le décompose en 4 triangles). Bref, on peut décomposer de façon aussi compliquée que l'on veut un P en un "pavage", une "tessellation", une "mosaïque", une "marqueterie", de $P_\ell : P = \sum_\ell P_\ell$. Ce qui est essentiel est la "linéarité" de ∂ par rapport à $+$: $\sum_\ell \partial P_\ell = \partial(\sum_\ell P_\ell) = \partial P$. Cette commutation de ∂ et \sum relie somme algébrique et incidence géométrique. Elle vient du fait que toutes les composantes des ∂P_ℓ qui ne sont pas dans ∂P interviennent deux fois avec une orientation opposée et s'éliminent donc. En revanche les composantes des ∂P_ℓ qui sont dans ∂P n'interviennent qu'une fois et avec l'orientation de ∂P . Par exemple (cf. figure 16) si l'on subdivise un quadrilatère $ABCD$ en deux quadrilatères $AEFDA$ et $EBCFE$ en introduisant une arête interne EF , on obtient

$$\begin{aligned}
 & \partial(AEFDA + EBCFE) \\
 &= (AE + EF + FD + DA) + (EB + BC + CF + FE) \\
 &= (AE + EB + BC + CF + FD + DA) + (EF + FE) \\
 &= AB + BC + CD + DA + 0 \\
 &= \partial(ABCD) .
 \end{aligned}$$

4. Cela a conduit à l'introduction du concept de *k*-chaîne C comme combinaison linéaire $C = \sum_{i=1}^{i=s} m_i D_i$ de domaines de type $D_i = f_i(P_i)$ avec des coefficients m_i , appelés "multiplicités", qui sont des entiers relatifs. Les *k*-chaînes forment un groupe commutatif Γ_k . L'intégrale $\int_C \omega$ est alors définie par simple linéarité comme $\int_C \omega = \sum_{i=1}^{i=s} m_i \int_{D_i} \omega$.
5. Ce qui est essentiel est que, en vertu de ce qui précède, les *k*-chaînes C peuvent posséder des *bords* ∂C qui sont des $(k-1)$ -chaînes. Pour les polyèdres fermés P de \mathbb{R}^k , ∂P est simplement la somme des faces orientées de façon à ce que la normale externe suivie de l'orientation de la face donne l'orientation de \mathbb{R}^k . En dimension 1, le bord d'un segment orienté AB est la 0-chaîne $B - A$.
6. On peut ainsi construire des chaînes C qui sont des combinaisons linéaires de sous-variétés à bord de M . En effet une variété à bord peut se décomposer en images de polyèdres inclus dans des cartes locales et, comme nous l'avons vu, les morceaux de bords non dans ∂M s'éliminent.
7. La propriété fondamentale de l'opérateur bord $\partial : \Gamma_k \rightarrow \Gamma_{k-1}$ est sa *nilpotence* : le bord d'un bord est nul. Dans le cas d'un *k*-polyèdre P , cela vient du fait que les $(k-2)$ -polyèdres qui sont les bords des faces interviennent chacun deux fois avec des signes opposés ce qui donne une somme nulle. Considérons par exemple un triangle orienté $\Delta : A \rightarrow B \rightarrow C \rightarrow A$. Son bord est $\partial\Delta = AB + BC + CA$ et donc

$$\partial^2\Delta = (B - A) + (C - B) + (A - C) = 0 .$$

De même, si $ABCD$ est un tétraèdre, ses faces convenablement orientées sont les triangles orientés $ABCA$, $ADBA$, $BDCB$, $ACDA$ et donc

$$\begin{aligned} \partial^2\Delta &= (AB + BC + CA) + (AD + DB + BA) \\ &\quad + (BD + DC + CB) + (AC + CD + DA) \\ &= (AB + BA) + (BC + CB) + (CA + AC) \\ &\quad + (AD + DA) + (DB + BD) + (DC + CD) \\ &= 0 . \end{aligned}$$

Les groupes Γ_k de *k*-chaînes ($0 \leq k \leq \dim M$) sont donc reliés par l'opérateur "bord" $\partial_{(k)} : \Gamma_k \rightarrow \Gamma_{k-1}$, $C \mapsto \partial C$ avec $\partial^2 C = 0$. On a à chaque stade de la suite

$$\longrightarrow \Gamma_{k+1} \xrightarrow{\partial_{(k+1)}} \Gamma_k \xrightarrow{\partial_{(k)}} \Gamma_{k-1} \longrightarrow$$

l'inclusion $\text{im}(\partial_{(k+1)}) \subset \ker(\partial_{(k)})$. Les chaînes $C \in \Gamma_k$ telles que $\partial C = 0$ sont appelées des *cycles* et celles qui sont des bords ∂D de $D \in \Gamma_{k+1}$ sont appelées des *bords*. Tout bord est un cycle mais la réciproque est en général complètement fautive pour des raisons relevant de la topologie *globale* de M . En effet, dire par exemple que tout 1-cycle est un bord signifie intuitivement que tout cercle plongé dans M borde

un disque plongé dans M . Cette propriété, dite de *simple connexité*, est vérifiée dans \mathbb{R}^n et donc *localement* dans M . Mais si, par exemple M est un tore, M aura deux classes de cycles (les parallèles et les méridiens) qui entoureront des “trous” et ne seront pas des bords.

Le quotient $H_k = \ker(\partial_{(k)}) / \text{im}(\partial_{(k+1)})$ s'appelle le k -ème groupe d'*homologie* de M . La suite des groupes d'homologie est un invariant essentiel de la topologie globale de M .

7.2.11. *Le théorème de Stokes : dualité homologie/cohomologie.*

Explicitons maintenant la dualité entre chaînes et formes différentielles et entre homologie et cohomologie. Elle repose sur un théorème célèbre, dit *théorème de Stokes*, qui unifie et généralise une foule de théorèmes clés partout utilisés en physique depuis le XVIII^e siècle. Ayant des prémices remontant à Newton et Leibniz, généralisant des premières formules remontant à Lagrange (1762), redécouvert par Carl Friedrich Gauss (1777-1855) en 1813 et George Green (1793-1841) en 1825, démontré en 1831 par Mikhail Ostrogradski (1801-1862), approfondi par William Thompson (Lord Kelvin, 1824-1907), George Stokes (1819-1903) puis Henri Poincaré et Élie Cartan, ce théorème est l'un des plus importants de toute la géométrie différentielle et de toute la physique mathématique. Nous en avons vu plus haut la version *infinitésimale*. Sa version globale généralise en dimension quelconque le résultat scolaire bien connu disant que si $f(x)$ est une fonction différentiable sur l'intervalle $[a, b]$ de \mathbb{R} alors

$$\int_a^b f'(x) dx = f(b) - f(a).$$

Si f est considérée comme une 0-forme, $f'(x) dx$ est la 1-forme df , et si $[a, b]$ est considéré comme une 1-chaîne, son bord est la 0-chaîne $b - a$. Comme l'intégrale de f sur un point x est tout simplement la valeur $f(x)$, la formule s'écrit

$$\int_{\partial[a,b]} f = \int_{[a,b]} df.$$

Soit M une variété de dimension n , ω une k -forme ($k \leq n-1$) à support compact de dérivée extérieure $d\omega$ et C une $(k+1)$ -chaîne de bord ∂C . Alors, ω étant une k -forme, elle peut être intégrée sur ∂C qui est de dimension k , et $d\omega$ étant une $(k+1)$ -forme, elle peut être intégrée sur C qui est de dimension $k+1$.

Théorème de Stokes. $\int_{\partial C} \omega = \int_C d\omega$. ◇

En particulier, si $\partial C = 0$ (i.e. si la chaîne C est un $(k+1)$ -cycle), $\int_C d\omega = 0$: l'intégrale d'une forme exacte (i.e. d'un “cobord”) sur un cycle est nul. De même, dualement, si $d\omega = 0$, $\int_{\partial C} \omega = 0$: l'intégrale d'une forme fermée (i.e. d'un “cocycle”) sur un bord est nulle.

Remarque. Pour $\omega = f(x, y) dx + g(x, y) dy$ une 1-forme sur \mathbb{R}^2 et K un compact, le théorème de Stokes $\int_{\partial K} \omega = \int_K \left(\frac{\partial g(x, y)}{\partial x} - \frac{\partial f(x, y)}{\partial y} \right) dx dy$ est connu sous le nom de formule de Green-Riemann. \square

En utilisant des cartes locales et les propriétés d'additivité exposées plus haut, la démonstration se ramène facilement au cas où $M = \mathbb{R}^+ \times \mathbb{R}^{n-1}$ est de coordonnées $(x_1 \geq 0, x_2, \dots, x_n)$, $\partial M = \{0\} \times \mathbb{R}^{n-1}$, où $k = n - 1$ et $\omega = \sum_{i=1}^{i=n} \omega_i(x) dx_1 \wedge \dots \wedge \widehat{dx}_i \wedge \dots \wedge dx_n$ (où \widehat{dx}_i signifie que dx_i est absent dans le produit extérieur). On a donc

$$\begin{aligned} d\omega &= \sum_{i=1}^{i=n} \frac{\partial \omega_i(x)}{\partial x_i} dx_i \wedge dx_1 \wedge \dots \wedge \widehat{dx}_i \wedge \dots \wedge dx_n \\ &= \sum_{i=1}^{i=n} (-1)^{i-1} \frac{\partial \omega_i(x)}{\partial x_i} dV \end{aligned}$$

où dV est la n -forme volume $dx_1 \wedge \dots \wedge dx_n$. Le théorème résulte alors d'un calcul élémentaire. Considérons pour simplifier le cas $n = 2$, $\omega = \omega_2(x) dx_1 + \omega_1(x) dx_2$, $d\omega = \left(\frac{\partial \omega_1(x)}{\partial x_1} - \frac{\partial \omega_2(x)}{\partial x_2} \right) dx_1 \wedge dx_2$.⁴⁸ On a $\int_{\partial M} \omega = \int_{\partial M} \omega_2(x) dx_1 + \int_{\partial M} \omega_1(x) dx_2$. Mais $dx_1 = 0$ sur $\partial M = \{0\} \times \mathbb{R}_{(x_2)}$ et il reste donc $\int_{\partial M} \omega = \int_{\mathbb{R}_{(x_2)}} \omega_1(0, x_2) dx_2$. Par ailleurs,

$$\begin{aligned} \int_M d\omega &= \int_{\mathbb{R}_{(x_1)}^+ \times \mathbb{R}_{(x_2)}} \left(\frac{\partial \omega_1(x_1, x_2)}{\partial x_1} - \frac{\partial \omega_2(x_1, x_2)}{\partial x_2} \right) dx_1 dx_2 \\ &= \int_{\mathbb{R}_{(x_2)}} \left(\int_{\mathbb{R}_{(x_1)}^+} \frac{\partial \omega_1(x_1, x_2)}{\partial x_1} dx_1 \right) dx_2 - \int_{\mathbb{R}_{(x_1)}^+} \left(\int_{\mathbb{R}_{(x_2)}} \frac{\partial \omega_2(x_1, x_2)}{\partial x_2} dx_2 \right) dx_1 . \end{aligned}$$

Mais l'intégrale $\int_{\mathbb{R}_{(x_2)}} \frac{\partial \omega_2(x_1, x_2)}{\partial x_2} dx_2 = \omega_2(x_1, +\infty) - \omega_2(x_1, -\infty) = 0$ pour tout x_1 car ω est à support compact par hypothèse et donc $\omega_2(x_1, +\infty) = \omega_2(x_1, -\infty) = 0$ et $\int_{\mathbb{R}_{(x_1)}^+} \frac{\partial \omega_1(x_1, x_2)}{\partial x_1} dx_1 = \omega_1(0, x_2) - \omega_1(+\infty, x_2) = \omega_1(0, x_2)$ car $\omega_1(+\infty, x_2) = 0$ (ω est à support compact) et il faut prendre $\int_{\mathbb{R}_{(x_1)}^+} = \int_{+\infty}^0$ pour que l'orientation soit correcte. On obtient donc bien $\int_{\partial M} \omega = \int_M d\omega = \int_{\mathbb{R}_{(x_2)}} \omega_1(0, x_2) dx_2$.

48. On note ici $\omega = \omega_2(x) dx_1 + \omega_1(x) dx_2$ au lieu de $\omega = \omega_1(x) dx_1 + \omega_2(x) dx_2$ parce que ω n'est pas définie directement comme une 1-forme mais comme une $(n - 1)$ -forme pour $n = 2$ et que l'on applique la notation $\omega_i dx_1 \wedge \dots \wedge \widehat{dx}_i \wedge \dots \wedge dx_n$.

On voit que l'intégration définit effectivement une dualité entre les deux suites d'applications linéaires

$$\begin{array}{ccccccc}
 \longrightarrow & \Omega^{k-1} & \xrightarrow{d^{(k-1)}} & \Omega^k \ni \omega & \xrightarrow{d^{(k)}} & \Omega^{k+1} \ni d\omega & \longrightarrow \\
 & & & \downarrow \int_{\partial C} \omega & & \downarrow \int_C d\omega & \\
 \longleftarrow & \Gamma_{k-1} & \xleftarrow{\partial^{(k)}} & \Gamma_k \ni \partial C & \xleftarrow{\partial^{(k+1)}} & \Gamma_{k+1} \ni C & \longleftarrow
 \end{array}$$

L'intégrale $\int_C \omega$ d'un k -cocycle ω sur un k -cycle C ne dépend que de la classe d'homologie de C et de la classe de cohomologie de ω et peut s'écrire comme un produit $\langle \omega, C \rangle$ entre les H^k et les H_k . Le théorème de Stokes dit que d et ∂ sont des opérateurs "adjoints" au sens où $\langle d\omega, C \rangle = \langle \omega, \partial C \rangle$.

Remarque. En appliquant le théorème de Stokes à l'exemple de l'angle θ du cercle \mathbb{S}^1 traité plus haut dans la remarque de la section 7.2.9, on voit que la 1-forme $d\theta$ n'est pas exacte. $\mathbb{S}^1 = \partial\Delta$ est le bord du disque unité. Comme $d\theta$ est fermée, $\int_{\partial\Delta} d\theta = \int_C d^2\theta = 0$. Mais si θ était une vraie fonction et donc si $d\theta$ était exacte, une seconde application de Stokes donnerait $\int_{\mathbb{S}^1} d\theta = \int_{\partial\mathbb{S}^1} \theta = \int_{\emptyset} \theta = 0$. Et pourtant $\int_0^{2\pi} d\theta = 2\pi$. La différence vient du fait que \mathbb{S}^1 est l'intervalle $[0, 2\pi]$ avec les extrémités identifiées et que θ n'a pas la même valeur aux extrémités (multi-valuation). La fonction unique θ n'est pas adaptée à \mathbb{S}^1 car elle n'est pas périodique. Les fonctions de base sur \mathbb{S}^1 sont les fonctions trigonométriques $\sin(\theta)$ et $\cos(\theta)$ et il faut les deux pour engendrer les fonctions sur \mathbb{S}^1 . En revanche on a bien $\int_0^{2\pi} d(\sin(\theta)) d\theta = \int_0^{2\pi} \cos(\theta) d\theta = 0$ et $\int_0^{2\pi} d(\cos(\theta)) d\theta = -\int_0^{2\pi} \sin(\theta) d\theta = 0$. \square

7.2.12. *Dérivées de Lie.*

Nous avons vu dans la section 6.3.2 que les champs de vecteurs tangents X sont interprétables comme des opérateurs de dérivation $X(f)$ sur les fonctions dérivables f ("dériver f dans la direction de X "). Cette dérivation se généralise aux champs de vecteurs Y et aux formes différentielles ω . On parle alors de *dérivée de Lie*, notée \mathcal{L}_X . Par définition, $\mathcal{L}_X(f) = X(f)$ et $\mathcal{L}_X Y = [X, Y]$. On a évidemment $\mathcal{L}_{[X, Y]}(f) = [\mathcal{L}_X, \mathcal{L}_Y](f) = [X, Y](f)$ et l'identité de Jacobi dit exactement que $\mathcal{L}_{[X, Y]}(Z) = [\mathcal{L}_X, \mathcal{L}_Y](Z)$.

Pour les 1-formes, on définit $\mathcal{L}_X \omega$ comme la 1-forme (attention aux parenthèses)

$$(\mathcal{L}_X \omega)(Y) = \mathcal{L}_X(\omega(Y)) - \omega([X, Y]) .$$

Il ne faut donc pas confondre la fonction qu'est la valeur de $\mathcal{L}_X \omega$ sur Y et la fonction qu'est la dérivation le long de X de la fonction $\omega(Y)$. On vérifie alors la célèbre *formule de Cartan*, dite "formule d'homotopie" :

$$\mathcal{L}_X \omega = i_X d\omega + d(i_X \omega) .$$

En effet, nous avons vu à la section 7.2.7 que $d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y])$. On a donc bien

$$\begin{aligned}
(\mathcal{L}_X \omega)(Y) &\stackrel{?}{=} i_X d\omega(Y) + d(i_X \omega)(Y) \\
&= d\omega(X, Y) + d(\omega(X))(Y) \\
&= X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]) + Y(\omega(X)) \\
&= X(\omega(Y)) - \omega([X, Y]) \\
&= \mathcal{L}_X(\omega(Y)) - \omega([X, Y]) \\
&= (\mathcal{L}_X \omega)(Y) .
\end{aligned}$$

Si ω est une k -forme, on peut définir $\mathcal{L}_X \omega$ par la formule de Cartan. Le produit intérieur $i_X d\omega$ est une k -forme puisque $(k+1)-1 = k$ et $d(i_X \omega)$ est également une k -forme puisque $(k-1)+1 = k$. La dérivée de Lie $\mathcal{L}_X \omega$ respecte par conséquent le type des entités sur lesquelles elle opère, en particulier le degré des formes différentielles. Il s'agit d'une véritable dérivation, et non pas d'une dérivation graduée comme la différentielle extérieure ou le produit extérieur. On a en effet

$$\mathcal{L}_X(\omega \wedge \rho) = \mathcal{L}_X(\omega) \wedge \rho + \omega \wedge \mathcal{L}_X(\rho) .$$

On montre que la formule pour les 1-formes se généralise bien aux k -formes :

$$\mathcal{L}_X \omega(X^1, \dots, X^k) = \mathcal{L}_X(\omega(X^1, \dots, X^k)) - \sum_{i=1}^{i=k} \omega(X^1, \dots, [X, X^i], \dots, X^k) .$$

La formule de Cartan permet de calculer l'action de \mathcal{L}_{fX} sur une k -forme. On trouve la k -forme

$$\mathcal{L}_{fX} \omega = f \mathcal{L}_X \omega + df \wedge i_X \omega .$$

Il s'agit d'un calcul symbolique facile. En effet

$$\begin{aligned}
\mathcal{L}_{fX} \omega &= i_{fX} d\omega + d(i_{fX} \omega) \\
&= f i_X d\omega + d(f i_X \omega) \\
&= f i_X d\omega + df \wedge i_X \omega + f d(i_X \omega) \\
&= f(i_X d\omega + d(i_X \omega)) + df \wedge i_X \omega \\
&= f \mathcal{L}_X \omega + df \wedge i_X \omega .
\end{aligned}$$

Ensuite, l'égalité $\mathcal{L}_{[X, Y]} = [\mathcal{L}_X, \mathcal{L}_Y]$ vérifiée plus haut pour les fonctions (où c'est la définition même de $[X, Y]$) et les champs de vecteurs se généralise facilement. Pour les 1-formes, elle s'écrit (exercice scolaire, en faisant attention aux parenthèses distinguant \mathcal{L}_X comme opérant sur la 1-forme ω ou sur la fonction $\omega(Z)$)

$$\begin{aligned}
(\mathcal{L}_{[X,Y]}\omega)(Z) &= \mathcal{L}_{[X,Y]}(\omega(Z)) - \omega([X, Y], Z) \\
([\mathcal{L}_X, \mathcal{L}_Y]\omega)(Z) &= (\mathcal{L}_X(\mathcal{L}_Y\omega))(Z) - (\mathcal{L}_Y(\mathcal{L}_X\omega))(Z) \\
&= \mathcal{L}_X(\mathcal{L}_Y\omega(Z)) - (\mathcal{L}_Y\omega)([X, Z]) - \mathcal{L}_Y(\mathcal{L}_X\omega(Z)) \\
&\quad + (\mathcal{L}_X\omega)([Y, Z]) \\
&= \mathcal{L}_X(\mathcal{L}_Y\omega(Z)) - \mathcal{L}_X(\omega([Y, Z])) - \mathcal{L}_Y(\omega([X, Z])) \\
&\quad - \omega([X, Z], Y) - \left(\begin{array}{l} \mathcal{L}_Y(\mathcal{L}_X\omega(Z)) - \mathcal{L}_Y(\omega([X, Z])) \\ -\mathcal{L}_X(\omega([Y, Z])) - \omega([Y, Z], X) \end{array} \right) \\
&= \mathcal{L}_{[X,Y]}(\omega(Z)) - \omega([X, Z], Y) + \omega([Y, Z], X) .
\end{aligned}$$

L'égalité vient, encore une fois, de l'identité de Jacobi. On a bien

$$-\omega([X, Y], Z) = -\omega([X, Z], Y) + \omega([Y, Z], X)$$

car

$$\begin{aligned}
-[[X, Y], Z] &= [Z, [X, Y]] = -[X, [Y, Z]] - [Y, [Z, X]] \\
&= -[[X, Z], Y] + [[Y, Z], X] .
\end{aligned}$$

Il existe de nombreuses formules reliant les dérivées de Lie, les crochets de Lie, les différentielles et les produits intérieurs. D'abord, \mathcal{L}_X commute avec d :

$$\mathcal{L}_X d\omega = d\mathcal{L}_X\omega$$

car $d^2 = 0$. De même, \mathcal{L}_X commute avec le produit intérieur i_X :

$$\mathcal{L}_X(i_X\omega) = i_X(\mathcal{L}_X\omega) .$$

On a également cette interprétation du lien entre produit intérieur et crochet de Lie :

$$i_{[X,Y]}\omega = \mathcal{L}_X(i_Y\omega) - i_Y(\mathcal{L}_X\omega) .$$

Il est très instructif de scruter ces formules qui relèvent d'un calcul purement symbolique.

7.3. Le texte fondamental d'Élie Cartan de 1899

7.3.1. Un retournement structural du calcul intégral.

Après ce petit "vademecum" sur les formes différentielles, il est intéressant de revenir au grand texte d'Élie Cartan *Sur certaines expressions différentielles et le problème de Pfaff* de 1899 où se trouve introduits de façon systématique les formes différentielles et leur calcul.

Le contexte de ce mémoire fondamental était celui de l'intégration de ce que l'on appelle des *équations de Pfaff* $\omega = 0$ où ω est une 1-forme. Ce problème particulièrement difficile avait remis en mouvement au début du XIX^e siècle la

réflexion sur les bases du calcul différentiel classique (disons à la Euler) tel qu'il s'était développé au XVIII^e siècle. Nous reviendrons plus bas à la section 2 du chapitre 4 sur la technicité du problème. Bornons-nous ici à souligner que c'est pour *unifier* et systématiser tout un ensemble de méthodes d'intégration introduites et développées par Monge, Pfaff, Grassmann, Natani, Clebsch, Lie, Frobenius ou Darboux que Cartan introduisit la notion *abstraite* de forme différentielle.

Avec Cartan, c'est un *tournant structural* dans le problème général de l'intégration des équations différentielles qui se parachève, après s'être progressivement développé et imposé au cours du XIX^e siècle. Ce tournant n'est pas sans rappeler celui accompli par Galois pour la résolution des équations algébriques. Jusqu'à Galois, le problème des équations algébriques était de trouver des formules explicites pour les solutions (les "racines") généralisant pour tout degré celles de l'équation du second degré connues depuis l'antiquité et celles de l'équation du troisième degré trouvées par Tartaglia et Cardan au milieu du XIV^e siècle. Ces formules donnent les solutions au moyen de polynômes et de racines $\sqrt[n]{}$ des coefficients de l'équation.⁴⁹ Devant l'impossibilité de généraliser ces formules de "résolubilité par radicaux" aux degrés ≥ 5 , Galois inversa la perspective, analysa la structure des équations en termes de groupes de permutation des racines et montra que la (très riche) structure des groupes imposait des contraintes drastiques à la possibilité d'existence de telles formules. Ces conditions ne sont pas satisfaites pour les groupes de permutations de n éléments pour $n \geq 5$, ce qui explique pourquoi l'équation générale du 5^{ème} degré n'est pas résoluble par radicaux. On passe donc, dans les cas où l'on ne connaît pas de formules générales, de la recherche "calculatoire" de solutions au moyen de méthodes *ad hoc* plus ou moins astucieuses à une analyse "structurale" des *contraintes* que doivent satisfaire les solutions.

Ce qui est remarquable est que la nature même des contraintes devient alors un fil directeur pour la résolubilité parce que les objets mathématiques qui l'expriment sont des objets mathématiques *eux-mêmes* dotés de structure (groupes et sous-groupes de permutation). Philosophiquement parlant, on passe de la quête directe de solutions à l'explicitation de *conditions de possibilité de résolubilité*. Ce retournement galoisien du problème est en fait un tournant "transcendantal" très analogue à celui opéré par Kant à propos des conditions de possibilité de l'expérience en physique.

Nous reviendrons plus en détail dans la section 1 du chapitre 4 sur les "révolutions coperniciennes" que constituent ces retournements transcendants et galoisiens.

Il en va de même en théorie de l'intégration des équations différentielles où, dans les nombreux cas où les méthodes classiques d'intégration ne sont pas utilisables, au lieu de chercher à imaginer des procédures *ad hoc*, on analyse structurellement les conditions de possibilité de l'intégrabilité, les contraintes s'exprimant, comme nous

49. Rappel scolaire pour mémoire : pour $ax^2 + bx + c = 0$, la formule est $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

le verrons de façon récurrente dans cet ouvrage, en termes de formes différentielles et de leurs dérivées extérieures.

Nous reviendrons de façon plus approfondie sur ce “retournement structural” du calcul différentiel à la section 2.1 du chapitre 4.

7.3.2. *Un calcul symbolique.*

La première partie du mémoire de Cartan introduit les formes ω et leur calcul *symbolique*. D'emblée, Cartan souligne fortement qu'il s'agit d'“expressions différentielles” “purements symboliques” (p. 244)⁵⁰ de type $\sum_i A_i dx_i$, $\sum_{j,k} B_{jk} dx_j dx_k$, etc. La nature *symbolique* de ces expressions et des règles qui les régissent est absolument essentielle. Elle remonte à Grassmann pour l'antisymétrie et prolonge et généralise de façon magistrale cette acuité métaphysique de Leibniz et de ses successeurs (les Bernouilli, Euler, etc.) affirmant que, même si les entités différentielles sont bien *concrètement* une affaire d'approximations par des infinitésimales tendant vers 0 (la tradition que l'on peut suivre jusqu'à Weierstrass en passant par Newton), elles relèvent néanmoins, *idéalement* et *abstraitement*, d'un calcul symbolique essentiellement caractérisé par un *produit* différent du produit classique.⁵¹

Ces formes abstraites sont

“des expressions différentielles symboliques, entières et homogènes par rapport aux différentielles de n variables, les coefficients étant des fonctions quelconques de ces variables.” (p. 241)

Plus précisément, elles sont constituées

1. de symboles dx_i de différentielles de variables x_1, \dots, x_n ;
2. de produits symboliques *non commutatifs* $dx_{i_1} \cdots dx_{i_k}$ de degré k de tels symboles ; à cause de la non commutativité, la $l^{\text{ème}}$ place dx_{i_l} dans un produit symbolique $dx_{i_1} \cdots dx_{i_k}$ est bien définie ; il l'appelle le “rang” de dx_{i_l} ;
3. de combinaisons linéaires *homogènes* (comme $\sum_{j,k} B_{jk} dx_j dx_k$) de tels produits avec des coefficients qui sont des fonctions différentiables des variables x_1, \dots, x_n (le degré est celui des termes, qui sont tous de même degré à cause de l'homogénéité) ;
4. de combinaisons auxquelles on applique “les règles ordinaires du calcul” : addition, multiplication, distributivité de la multiplication par rapport à l'addition, à condition de respecter la non-commutativité et l'homogénéité.

50. Les numéros de pages du mémoire seront indiquées dans le corps du texte.

51. La formule de Leibniz pour la différentielle d'un produit de fonctions : $d(fg) = f dg + g df$.

Si l'on considère alors des produits et des sommes respectant ces conditions, on peut les développer de façon à arriver à "la forme la plus simple". Par exemple

$$\begin{aligned} & \left(\sum_i A_i dx_i \right) \left(\sum_{,jk} B_{jk} dx_j dx_k \right) + \left(\sum_{,ijk} C_{ijk} dx_i dx_j dx_k \right) \\ &= \sum_{,ijk} (A_i B_{jk} + C_{ijk}) dx_i dx_j dx_k . \end{aligned}$$

Il faut noter que Cartan suppose que les variables x_i dénotent des coordonnées de \mathbb{R}^n et ne considère pas d'autres possibilité de variétés (c'est important car \mathbb{R}^n étant simplement connexe, une forme fermée, i.e. localement exacte, y est globalement exacte). Il suppose également que les coefficients sont dérivables autant de fois qu'il est nécessaire dans les calculs. En fait, il introduit des hypothèses non seulement de différentiabilité mais même d'analyticité lorsque cela facilite les choses.

Les expressions de Pfaff dont traite le mémoire sont les expressions du premier degré (les 1-formes) $\omega = \sum_{i=1}^{i=n} A_i(x) dx_i$. Mais, ainsi que le notent Chern et Chevalley,

"the essential originality of Cartan consists in having introduced, besides the Pfaffian forms, the exterior differential forms of higher degree." ([114], p. 229)

7.3.3. *Le lien avec les intégrales multidimensionnelles.*

Dans ce mémoire fondateur, Cartan n'introduit pas directement l'*antisymétrie* au niveau des "expressions différentielles" symboliques elles-mêmes. Il considère ces expressions comme des symboles différentiels pouvant posséder des *valeurs*, et c'est au niveau de ces valeurs qu'il introduit l'antisymétrie.

Il s'agit là d'un point essentiel sur lequel Édouard Goursat a beaucoup insisté. Les formes différentielles $\omega = \sum A_{i_1 \dots i_h}(x) dx_{i_1} \dots dx_{i_h}$ de degré h à n variables sont faites pour être *intégrées* sur les sous-variétés M de dimension h de \mathbb{R}^n . Ce sont les quantités $\int_M \omega$ qui ont un sens et la spécificité du calcul sur les ω vient du fait que ces quantités doivent posséder un sens *intrinsèque* qui soit *invariant* par changement de variables. Nous l'avons déjà expliqué dans les sections 7.2.3 et 7.2.6, mais précisons-le maintenant à partir des textes originaux.

Dès son papier de 1915 [217], Goursat explique :

"L'étude des intégrales multiples étendues à des variétés d'un nombre quelconque de dimensions conduit tout naturellement à considérer des formes symboliques de différentielles qui figurent sous le signe d'intégration." ([217], p. 1)

et dans son traité de 1922 [218] il réaffirme

"Le calcul de ces quantités est, en somme, celui des expressions différentielles qui sont placées sous un signe d'intégrale multiple." ([218], p. 242)

Et il insiste sur le fait que le calcul symbolique des expressions différentielles est *sui generis* :

“Le produit $dx_1 \dots dx_p$ n’est plus un produit algébrique au sens propre du mot, mais un *produit symbolique* pour lequel les règles du calcul sont toutes différentes des règles ordinaires du calcul algébrique.” ([218], p. 84)

Comme nous l’avons vu, l’antisymétrie vient essentiellement du fait que dans une intégrale multiple $I = \int_S A(\alpha_1, \dots, \alpha_h) d\alpha_1 \dots d\alpha_h$ sur un domaine S de \mathbb{R}^h un changement de variables $\alpha : \mathbb{R}^h \rightarrow \mathbb{R}^h$, $\beta \mapsto \alpha(\beta)$, I devient

$$\int_{\varphi^{-1}(S)} A(\alpha_i(\beta_j)) J(\alpha, \beta) d\beta_1 \dots d\beta_h$$

où J , le Jacobien de I (aussi noté $\frac{D(\alpha)}{D(\beta)}$ ou simplement $D(\alpha)$) est le *déterminant* de la matrice jacobienne des dérivées premières $\frac{\partial \alpha_i}{\partial \beta_j}$ des α_i par rapport aux β_j . C’est l’intervention de ces déterminants garantissant l’invariance par changement de variables qui est à l’origine du calcul multilinéaire alterné anticommutatif des formes différentielles. Si par exemple l’on intervertit deux variables (x, y) , la matrice jacobienne du changement de variable est $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ et le jacobien est -1 et, comme le dit Goursat,

“Cette règle d’apparence paradoxale [$dydx = -dxdy$] est, comme on le voit, une conséquence de la signification même du produit $dxdy$.” ([218], p. 85)

Précisons encore un peu plus les choses. En tant que telle, une intégrale $\int_M \omega$ n’a pas de sens puisque ω dépend des n variables x_i alors que M est une sous-variété de dimension h . Pour lui donner un sens, il faut considérer que M est paramétrée (du moins localement)⁵² par des coordonnées locales $(\alpha_1, \dots, \alpha_h)$, paramétrage défini par une application $\psi : S \subset \mathbb{R}^h \rightarrow M \subset \mathbb{R}^n$ et interpréter chaque terme de I

$$\int_M A_{j_1 \dots j_h}(x) dx_{j_1} \dots dx_{j_h}$$

comme l’intégrale

$$\int_{\psi^{-1}(S)} A(x_j(\alpha_k)) J(x, \alpha) d\alpha_1 \dots d\alpha_h .$$

Par construction, cette définition est invariante par changements de coordonnées.

Présentons maintenant la “convention” de Cartan avant de la commenter. Si ω est de degré h ,

52. Cartan n’aborde pas dans cet article la dialectique du local et du global, bien qu’elle existât déjà à cette époque.

1. il introduit h variables $\alpha_1, \dots, \alpha_h$ (leur nombre est le degré) et suppose que les x_j sont des fonctions $x_j(\alpha)$ des α_i ; cela signifie qu'il considère des sous-variétés M de \mathbb{R}^n de dimension h qui sont *paramétrées* par les variables α ; ⁵³
2. il considère les $h!$ permutations possibles $\alpha_{i_1}, \dots, \alpha_{i_h}$ des $\alpha_1, \dots, \alpha_h$;
3. étant donné un monôme $Adx_{j_1} \dots dx_{j_h}$ de ω , il y remplace le dx_{j_l} occupant la $l^{\text{ème}}$ place (rang l) dans le produit par sa dérivée partielle $\frac{\partial x_{j_l}}{\partial \alpha_{i_l}}$ par rapport à la variable α_{i_l} occupant la $l^{\text{ème}}$ place dans la permutation $\alpha_{i_1}, \dots, \alpha_{i_h}$;
4. il affecte le résultat du signe $+1$ ou -1 suivant la signature s de la permutation ($s = (-1)^p$ où p est le nombre de transpositions nécessaires pour effectuer la permutation).

La valeur de ω est alors la *somme algébrique* des résultats. Pour un monôme $\omega_0 = Adx_{j_1} \dots dx_{j_h}$, $\omega_0(\alpha)$ est $A \frac{D(x_j)}{D(\alpha)}$. Cartan donne l'exemple simple de la 2-forme de 3 variables

$$\omega = A_1(x_1, x_2, x_3)dx_2dx_1 + A_2(x_1, x_2, x_3)dx_3dx_2.$$

Pour le premier terme, dx_2 est de rang 1 et dx_1 de rang 2 et, pour le deuxième terme, dx_3 de rang 1 et dx_2 de rang 2. Soient alors α_1, α_2 deux variables ($h = 2$). Pour la permutation identique (α_1, α_2) , on obtient donc $A_1 \frac{\partial x_2}{\partial \alpha_1} \frac{\partial x_1}{\partial \alpha_2} + A_2 \frac{\partial x_3}{\partial \alpha_1} \frac{\partial x_2}{\partial \alpha_2}$ et, pour la permutation (α_2, α_1) de signature -1 , on obtient $-A_1 \frac{\partial x_2}{\partial \alpha_2} \frac{\partial x_1}{\partial \alpha_1} - A_2 \frac{\partial x_3}{\partial \alpha_2} \frac{\partial x_2}{\partial \alpha_1}$, d'où la valeur de ω que nous noterons pour simplifier $\omega(\alpha)$,

$$\omega(\alpha) = A_1(\alpha) \left(\frac{\partial x_2}{\partial \alpha_1} \frac{\partial x_1}{\partial \alpha_2} - \frac{\partial x_2}{\partial \alpha_2} \frac{\partial x_1}{\partial \alpha_1} \right) + A_2(\alpha) \left(\frac{\partial x_3}{\partial \alpha_1} \frac{\partial x_2}{\partial \alpha_2} - \frac{\partial x_3}{\partial \alpha_2} \frac{\partial x_2}{\partial \alpha_1} \right)$$

où nous notons $A(\alpha)$ la valeur de $A(x_i)$ lorsque les variables x_i sont remplacées par leurs valeurs $x_i(\alpha)$.

Remarque. Dans le cas de la 1-forme $\omega = dy - p dx$ de 3 variables qui sera au cœur du chapitre 3, les variables deviennent des fonctions d'une variable t et parcourent une courbe gauche Γ dans \mathbb{R}^3 . La valeur $\omega(t)$ de ω est alors $\omega(t) = y'(t) - p(t)x'(t)$ et l'équation de Pfaff $\omega(t) = 0$ signifie que $p(t) = \frac{y'(t)}{x'(t)}$ est la tangente à la courbe $\gamma = \{x(t), y(t)\}$ du plan (x, y) . Nous verrons que cela signifie que Γ est la "relevée legendrienne" de γ . \square

Commentons un peu la convention de Cartan et montrons qu'elle correspond bien à la définition actuelle des formes que nous avons donnée plus haut. Prenons le cas le plus simple $\omega = dx_i$ dans \mathbb{R}^n et introduisons une variable α . Elle va paramétrer une courbe $\Gamma = \{x_i(\alpha)\}$ dans \mathbb{R}^n . En un point x de Γ , la paramétrisation définit

53. Comme nous l'avons dit, Cartan ne prend pas en considération la différence local/global. Le mieux est donc de dire que l'on travaille dans une carte locale de \mathbb{R}^n de coordonnées x_i .

un vecteur tangent $X = \sum_{i=1}^{i=n} \frac{\partial x_i}{\partial \alpha} \frac{\partial}{\partial x_i}$ de composantes $X_i = \frac{\partial x_i}{\partial \alpha}$. La valeur $\omega(\alpha)$ de ω sur cette courbe est par définition $\omega(\alpha) = \frac{\partial x_i}{\partial \alpha}$, c'est-à-dire effectivement $dx_i(X)$. Autrement dit, $\omega(\alpha)$ est la valeur de la 1-forme le long de la courbe Γ . On peut aussi dire

- (i) que α définit une application $\varphi : \mathbb{R} \rightarrow \mathbb{R}^n$,
- (ii) que l'on considère l'image réciproque $\varphi^*(\omega)$ de ω par φ , et
- (iii) que l'on écrit $\varphi^*(\omega) = \omega(\alpha) d\alpha$, $d\alpha$ étant la forme volume dV^1 de \mathbb{R} (dans ce cas l'élément de longueur puisqu'on est en dimension 1).⁵⁴

Considérons maintenant le cas le plus simple de dimension 2, $\omega = dx_i dx_j$ ($i \neq j$) dans $\mathbb{R}^{n \geq 2}$. Soient (α_1, α_2) paramétrant une surface $M = \{x_l(\alpha_1, \alpha_2)\}_{l=1, \dots, n}$ de \mathbb{R}^n . En un point x de M , la paramétrisation définit deux vecteurs tangents $X_1 = \sum_{i=1}^{i=n} \frac{\partial x_i}{\partial \alpha_1} \frac{\partial}{\partial x_i}$ et $X_2 = \sum_{i=1}^{i=n} \frac{\partial x_i}{\partial \alpha_2} \frac{\partial}{\partial x_i}$. On a par définition $\omega(\alpha) = \frac{\partial x_i}{\partial \alpha_1} \frac{\partial x_j}{\partial \alpha_2} - \frac{\partial x_i}{\partial \alpha_2} \frac{\partial x_j}{\partial \alpha_1}$. On vérifie immédiatement que $\omega(\alpha) = dx_i \wedge dx_j(X_1, X_2)$ au sens de la définition standard de la section 7.2.4. On peut aussi dire que

$$\begin{aligned} dx_i \wedge dx_j &= \left(\frac{\partial x_i}{\partial \alpha_1} d\alpha_1 + \frac{\partial x_i}{\partial \alpha_2} d\alpha_2 \right) \wedge \left(\frac{\partial x_j}{\partial \alpha_1} d\alpha_1 + \frac{\partial x_j}{\partial \alpha_2} d\alpha_2 \right) \\ &= \left(\frac{\partial x_i}{\partial \alpha_1} \frac{\partial x_j}{\partial \alpha_2} - \frac{\partial x_i}{\partial \alpha_2} \frac{\partial x_j}{\partial \alpha_1} \right) d\alpha_1 \wedge d\alpha_2 = \omega(\alpha) d\alpha_1 \wedge d\alpha_2 = \omega(\alpha) dV^2, \end{aligned}$$

que les (α_1, α_2) définissent une application $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^n$ et que l'on considère l'image réciproque $\varphi^*(\omega)$ de ω par φ , celle-ci s'écrivant $\varphi^*(\omega) = \omega(\alpha) dV^2$, dV^2 étant la forme volume de \mathbb{R}^2 .

Cela se généralise trivialement en dimension h et redonne la définition initiale de Cartan.

On voit donc que, d'emblée, Cartan pense les h -formes différentielles comme prenant des valeurs le long de sous-variétés M de dimension h , c'est-à-dire comme des entités devant être intégrées sur les M . L'intégration sera invariante par changement de variable car, comme y insiste Cartan, si l'on considère un monôme $\omega = A dx_{j_1} \dots dx_{j_h}$ et des paramètres $\alpha_1, \dots, \alpha_h$, la valeur $\omega(\alpha)$ est donc bien

“tout simplement le produit de A par le déterminant fonctionnel de x_{j_1}, \dots, x_{j_h} par rapport à $\alpha_1, \dots, \alpha_h$.” (p. 247)

Cartan formule alors certaines conséquences de l'antisymétrie :

- (i) $dx_{j_1} \dots dx_{j_h} \equiv 0$ si deux des différentielles sont égales ;
- (ii) un changement de signe si l'on transpose deux des dx_{j_i} ;
- (iii) $\omega \equiv 0$ si $h > n$;
- (iv) si ω est un monôme, toutes les puissances $\omega^p \equiv 0$ ($p \geq 2$) car les dx_i y apparaissent plusieurs fois ;

54. L'application linéaire tangente de φ est $D\varphi = \left[\frac{\partial x_i}{\partial \alpha} \right]$ et $D\varphi \left(\frac{\partial}{\partial \alpha} \right) = \sum_{i=1}^{i=n} \frac{\partial x_i}{\partial \alpha} \frac{\partial}{\partial x_i} = X$.

(v) si $\omega = \omega_1 + \dots + \omega_m$ (les ω_k étant des monômes) et si ω est de degré impair alors $\omega^p \equiv 0$ ($p \geq 2$) car chaque terme de ω^p comprenant un $\omega_k \omega_l$ sera associé au même terme avec $\omega_k \omega_l$ remplacé par $\omega_l \omega_k$ et $\omega_l \omega_k = -\omega_k \omega_l$ en degré impair ;

(vi) si ω est de degré pair il n'en va plus de même, par exemple

$$(\omega_1 + \dots + \omega_m)^2 = 2 \left(\sum_{k,l=1, k < l}^{k,l=m} \omega_k \omega_l \right).$$

Une “propriété capitale” sur laquelle insiste Cartan est le bon comportement des formes par changements de variables $x_i = x_i(y_k)$ et $y_k = y_k(x_i)$, les “déterminants fonctionnels” (les jacobiens) $\frac{D(x_i)}{D(y_k)}$ étant non nuls. En développant les dx_i en $dx_i = \sum_{k=1}^{k=n} \frac{\partial x_i}{\partial y_k} dy_k$, on transforme une h -forme ω en une h -forme $\bar{\omega}$. Mais la valeur $\omega(\alpha)$ reste *invariante* car les $\frac{\partial x_{j_l}}{\partial \alpha_{i_l}}$ se transforment en $\sum_{k=1}^{k=n} \frac{\partial x_{j_l}}{\partial y_k} \frac{\partial y_k}{\partial \alpha_{i_l}}$ qui leurs sont égales. Qui plus est, les produits $\omega = \omega_1 \dots \omega_m$ se comportent bien, i.e. $\bar{\omega} = \bar{\omega}_1 \dots \bar{\omega}_m$, d'après la propriété fondamentale des déterminants fonctionnels $\frac{D(x_i)}{D(\alpha_l)} = \frac{D(x_i)}{D(y_k)} \frac{D(y_k)}{D(\alpha_l)}$.

Remarque sur les notations. Les auteurs dont nous parlons dans cette section historique n'utilisent pas de notation spécifique pour le produit symbolique non commutatif $\omega \rho$ des formes. Il le notent comme un produit en insistant sur sa non-commutativité. Ensuite, on introduisit le symbole du produit extérieur $\omega \wedge \rho$. Cela peut introduire parfois quelques ambiguïtés sur les coefficients. Par exemple, pour une 2-forme $\sum A_{ij} dx_i dx_j$, on somme sur tous les couples (i, j) et l'on compte donc les termes $A_{ij} dx_i dx_j + A_{ji} dx_j dx_i = (A_{ij} - A_{ji}) dx_i dx_j$ pour $i < j$ alors que la somme $\sum A_{ij} dx_i \wedge dx_j$ signifie directement $\sum_{i < j} A_{ij} dx_i \wedge dx_j$. \square

7.3.4. La dérivée extérieure des formes de Pfaff.

Dans la seconde partie du mémoire, Élie Cartan introduit la *dérivée* extérieure $d\omega$ des 1-formes de Pfaff $\omega = \sum_{i=1}^{i=n} A_i dx_i$, concept fondamental introduit dans les années 1870 par Frobenius sous le nom de “*covariant bilinéaire*”. Il la note ω' et la définit comme la forme de degré 2, $\omega' = \sum_{i=1}^{i=n} dA_i dx_i$, les $dA_i = \sum_{k=1}^{k=n} \frac{\partial A_i}{\partial x_k} dx_k$ étant les différentielles totales des coefficients A_i . On a donc $\omega' = \sum_{i=1}^{i=n} \sum_{k=1}^{k=n} \frac{\partial A_i}{\partial x_k} dx_k dx_i$, les termes en $dx_i dx_i$ s'annulant et les termes $dx_i dx_k$ et $dx_k dx_i = -dx_i dx_k$ ($k \neq i$) se regroupant.

Si ω est une différentielle *exacte*, c'est-à-dire s'il existe une fonction $F(x_i)$ telle que $\omega = dF = \sum_{i=1}^{i=n} \frac{\partial F}{\partial x_i} dx_i$, alors $\omega' \equiv 0$. En effet, $\frac{\partial A_i}{\partial x_k} dx_k dx_i = \frac{\partial^2 F}{\partial x_k \partial x_i} dx_k dx_i$ et $\frac{\partial A_k}{\partial x_i} dx_i dx_k = \frac{\partial^2 F}{\partial x_i \partial x_k} dx_i dx_k$ et comme $\frac{\partial^2 F}{\partial x_k \partial x_i} = \frac{\partial^2 F}{\partial x_i \partial x_k}$ et $dx_k dx_i = -dx_i dx_k$, les deux termes s'annulent.

Cartan montre la réciproque (mais en oubliant une hypothèse), à savoir que si $\omega' \equiv 0$ alors ω est exacte au sens où il existe F telle que $\omega = dF$.⁵⁵ Cartan procède par récurrence. Le théorème est vrai en dimension 1 car si $\omega = A(x) dx$, ω' est automatiquement $\equiv 0$ et $\omega = dF$ avec $F(x) = \int_a^x A(\xi) d\xi$. Supposons que le théorème soit vrai en dimension $< n$ et soit $\omega = \sum_{i=1}^{i=n} A_i dx_i$. Si l'on fixe x_1 en une constante c , ω devient une 1-forme ω_1 en les $(n-1)$ variables $x_{i \geq 2}$ et $\omega'_1 = 0$ puisque $\omega' = 0$. D'après l'hypothèse de récurrence, il existe donc une fonction $u(c, x_{i \geq 2})$ telle que $\omega_1 = u(c, x_{i \geq 2})$. Si l'on fait alors varier la constante c (considérée comme un paramètre) afin de récupérer la variable x_1 , on obtient

$$\omega = du + \left(A_1 - \frac{\partial u}{\partial x_1} \right) dx_1 = \widetilde{A}_1 dx_1 + du.$$

Par hypothèse, $\omega' = 0$ et donc, puisque $(du)' = 0$,

$$d(\widetilde{A}_1) dx_1 = 0 = \sum_{i=1}^{i=n} \frac{\partial \widetilde{A}_1}{\partial x_i} dx_i dx_1 = \sum_{i=2}^{i=n} \frac{\partial \widetilde{A}_1}{\partial x_i} dx_i dx_1 \quad (\text{car } dx_1 dx_1 = 0)$$

ce qui implique $\frac{\partial \widetilde{A}_1}{\partial x_i} = 0$ pour $i \geq 2$. Par conséquent \widetilde{A}_1 est une fonction de la seule variable x_1 et l'on en déduit que $\omega = d\left(u + \int_a^{x_1} \widetilde{A}_1(\xi) d\xi\right)$.

Bref, la condition $\omega' \equiv 0$ est la *condition d'intégrabilité optimale* garantissant que l'équation de Pfaff $\omega = 0$ est, comme on dit, *complètement intégrable* et que $\omega = dF$, l'espace ambiant \mathbb{R}^n étant feuilleté par les hypersurfaces de niveau $F = \text{cste}$. Comme nous le verrons dans le chapitre 4 (en particulier aux sous-sections 2.7.4 et 3), tout le travail sur le problème de Pfaff a consisté à définir des conditions d'intégrabilité *générales* allant bien au-delà de ce cas optimal.

Cartan insiste sur le fait que la dérivée ω' est *covariante*, i.e. se comporte bien par rapport aux changements de variables, au sens où, si $x_i = x_i(y_k)$ et (α_1, α_2) paramétrisent une surface, $\bar{\omega}'(\alpha) = \omega'(\alpha)$. En effet

$$\bar{\omega} = \sum_{i=1}^{i=n} A_i \left(\sum_{k=1}^{k=n} \frac{\partial x_i}{\partial y_k} dy_k \right) = \sum_{k=1}^{k=n} \left(\sum_{i=1}^{i=n} A_i \frac{\partial x_i}{\partial y_k} \right) dy_k = \sum_{k=1}^{k=n} B_k dy_k$$

avec $B_k = \left(\sum_{i=1}^{i=n} A_i \frac{\partial x_i}{\partial y_k} \right)$. Soient $\alpha = (\alpha_1, \alpha_2)$ paramétrant une surface M . Sur M

55. En termes actuels, cela signifierait que si ω est fermée ($d\omega = 0$) alors ω est exacte ($\omega = dF$), ce qui est faux en général et n'est vrai ici, rappelons-le, que parce que \mathbb{R}^n est simplement connexe et a une cohomologie triviale.

$$\omega = \sum_{i=1}^{i=n} A_i dx_i = \sum_{i=1}^{i=n} \sum_{l=1}^{l=2} A_i(\alpha) \frac{\partial x_i}{\partial \alpha_l} d\alpha_l$$

$$\bar{\omega} = \sum_{k=1}^{k=n} B_k dy_k = \sum_{k=1}^{k=n} \sum_{l=1}^{l=2} B_k(\alpha) \frac{\partial y_k}{\partial \alpha_l} d\alpha_l$$

et comme les coefficients de $d\alpha_1$ et $d\alpha_2$ doivent être les mêmes, on a nécessairement

$$\sum_{i=1}^{i=n} A_i(\alpha) \frac{\partial x_i}{\partial \alpha_1} = \sum_{k=1}^{k=n} B_k(\alpha) \frac{\partial y_k}{\partial \alpha_1}$$

$$\sum_{i=1}^{i=n} A_i(\alpha) \frac{\partial x_i}{\partial \alpha_2} = \sum_{k=1}^{k=n} B_k(\alpha) \frac{\partial y_k}{\partial \alpha_2} .$$

Cartan dérive la première égalité par rapport à α_2 et la seconde par rapport à α_1 et fait la différence. D'où les égalités

$$\sum_{i=1}^{i=n} \left(A_i \frac{\partial^2 x_i}{\partial \alpha_2 \partial \alpha_1} + \frac{\partial A_i}{\partial \alpha_2} \frac{\partial x_i}{\partial \alpha_1} \right) = \sum_{k=1}^{k=n} \left(B_k \frac{\partial^2 y_k}{\partial \alpha_2 \partial \alpha_1} + \frac{\partial B_k}{\partial \alpha_2} \frac{\partial y_k}{\partial \alpha_1} \right)$$

$$\sum_{i=1}^{i=n} \left(A_i \frac{\partial^2 x_i}{\partial \alpha_1 \partial \alpha_2} + \frac{\partial A_i}{\partial \alpha_1} \frac{\partial x_i}{\partial \alpha_2} \right) = \sum_{k=1}^{k=n} \left(B_k \frac{\partial^2 y_k}{\partial \alpha_1 \partial \alpha_2} + \frac{\partial B_k}{\partial \alpha_1} \frac{\partial y_k}{\partial \alpha_2} \right)$$

$$\sum_{i=1}^{i=n} \left(\frac{\partial A_i}{\partial \alpha_2} \frac{\partial x_i}{\partial \alpha_1} - \frac{\partial A_i}{\partial \alpha_1} \frac{\partial x_i}{\partial \alpha_2} \right) = \sum_{k=1}^{k=n} \left(\frac{\partial B_k}{\partial \alpha_2} \frac{\partial y_k}{\partial \alpha_1} - \frac{\partial B_k}{\partial \alpha_1} \frac{\partial y_k}{\partial \alpha_2} \right)$$

$$\sum_{i=1}^{i=n} \left(\frac{\partial A_i}{\partial \alpha_2} \frac{\partial x_i}{\partial \alpha_1} - \frac{\partial A_i}{\partial \alpha_1} \frac{\partial x_i}{\partial \alpha_2} \right) = \sum_{k=1}^{k=n} \left(\frac{\partial B_k}{\partial \alpha_2} \frac{\partial y_k}{\partial \alpha_1} - \frac{\partial B_k}{\partial \alpha_1} \frac{\partial y_k}{\partial \alpha_2} \right) .$$

Mais

$$\sum_{i=1}^{i=n} \left(\frac{\partial A_i}{\partial \alpha_2} \frac{\partial x_i}{\partial \alpha_1} - \frac{\partial A_i}{\partial \alpha_1} \frac{\partial x_i}{\partial \alpha_2} \right) = \sum_{i=1}^{i=n} \left(\left(\sum_{k=1}^{k=n} \frac{\partial A_i}{\partial x_k} \frac{\partial x_k}{\partial \alpha_2} \right) \frac{\partial x_i}{\partial \alpha_1} - \left(\sum_{k=1}^{k=n} \frac{\partial A_i}{\partial x_k} \frac{\partial x_k}{\partial \alpha_1} \right) \frac{\partial x_i}{\partial \alpha_2} \right)$$

$$= \sum_{i=1}^{i=n} \sum_{k=1}^{k=n} \frac{\partial A_i}{\partial x_k} \left(\frac{\partial x_k}{\partial \alpha_2} \frac{\partial x_i}{\partial \alpha_1} - \frac{\partial x_k}{\partial \alpha_1} \frac{\partial x_i}{\partial \alpha_2} \right)$$

$$= \omega'(\alpha) .$$

De même, $\sum_{k=1}^{k=n} \left(\frac{\partial B_k}{\partial \alpha_2} \frac{\partial y_k}{\partial \alpha_1} - \frac{\partial B_k}{\partial \alpha_1} \frac{\partial y_k}{\partial \alpha_2} \right) = \bar{\omega}'(\alpha)$ et par conséquent $\omega'(\alpha) = \bar{\omega}'(\alpha)$ quelle que soit la paramétrisation α . Donc $\bar{\omega}' = \omega'$.

On peut évidemment généraliser la définition de ω' aux formes de degré quelconque. Ainsi que le note Goursat,

“Les formes dérivées jouent le même rôle, dans l'étude des intégrales multiples, que les différentielles totales exactes dans la théorie des intégrales curvilignes. [Ce sont des] différentielles totales symboliques.” ([217], p. 6, cf. aussi [218], p. 110)

Nous reviendrons de façon plus précise historiquement sur le “covariant bilinéaire” de Frobenius dans les sections 2.7.4 et 13 du chapitre 3.

7.3.5. Les “dérivées” d'ordre supérieur.

Une fois définie la dérivée ω' d'une 1-forme, Cartan n'itère pas simplement la procédure, ce qui conduirait à $d^2\omega = 0$,⁵⁶ et donne une tout autre définition des dérivées successives. Nous verrons que cela est lié aux *conditions d'intégrabilité* (surtout dues à Frobenius et Darboux dont Cartan loue les “belles recherches”) des formes de Pfaff qui sont au centre du mémoire, conditions que nous expliquerons au chapitre 4, en particulier à la section 2. Cartan introduit une chaîne d'invariants :

“Grâce à la notion des expressions différentielles symboliques, ce covariant $[\omega']$ est le premier terme d'une suite de covariants symboliques du troisième, quatrième, ..., degré, qui se déduisent intuitivement de l'expression de Pfaff et de sa dérivée par des multiplications.” (p. 242, souligné par Cartan)

Comme ω' est de degré 2, les puissances $(\omega')^p$ de ω' ne sont pas nécessairement nulles (du moins si $2p \leq n$ puisqu'une k -forme sur \mathbb{R}^n est nécessairement identiquement nulle si $k > n$). Cartan définit alors itérativement les dérivées $\omega^{(p)}$ d'ordre p de ω par les $(p+1)$ -formes $\omega^{(0)} = \omega$, $\omega^{(1)} = \omega'$, $\omega^{(2)} = \omega\omega'$, ..., $\omega^{(2m-1)} = \frac{1}{m!}(\omega')^m$, $\omega^{(2m)} = \omega\omega^{(2m-1)}$.⁵⁷ Comme $\omega^{(p)}$ est de degré $(p+1)$, on a $\omega^{(p)} \equiv 0$ si $p+1 > n$ et donc seules les dérivées $\omega^{(p)}$ pour $p \leq n-1$ sont à considérer. Le plus petit p tel que $\omega^{(p)} \equiv 0$ s'appelle *la classe* de la forme de Pfaff ω . Depuis Frobenius (chapitre 4, section 2), on sait que c'est le plus petit nombre de variables auquel on peut réduire ω par des changements de variables éliminant le maximum de variables initiales. Si ω ne peut être réduite, sa classe a la valeur maximale n .

À un extrême, $\omega' \equiv 0$ correspond donc au cas de la classe 1 où ω est une différentielle exacte dF . À l'autre extrême on aura deux cas suivant la parité de n . Si n est pair, $n = 2m$, on aura le cas *symplectique*⁵⁸ où la $2m = n$ -forme $(\omega')^m = m!\omega^{(n-1)}$ est non nulle et proportionnelle à la forme volume $d_n V = \bigwedge_{i=1}^{i=n} dx_i$. La

56. En revanche, dans [217], Goursat note que si l'on appelle dérivée d'ordre supérieur l'itération de d , alors $d^2 = 0$.

57. Insistons sur le fait que ces “dérivées” $\omega^{(p)}$ ne sont pas les $d^p\omega$ qui sont toujours $\equiv 0$ pour $p > 1$ puisque $d^2 = 0$.

58. Le terme “symplectique” a été introduit par Hermann Weyl, pour éviter des confusions entre le terme de “complexe” qui était alors utilisé et le terme de “nombre complexe”.

classe est alors n . Si au contraire n est impair, $n = 2m + 1$, on aura le cas *de contact* où la $2m + 1 = n$ -forme $\omega (\omega')^m = m! \omega^{(2m)}$ est non nulle et proportionnelle à la forme volume $d_n V = \bigwedge_{i=1}^{i=n} dx_i$. La classe est alors également n .

Cartan insiste sur le caractère unificateur de sa théorie des formes différentielles et de leurs dérivées :

“La considération de ces dérivées permet de trouver d’une manière pour ainsi dire intuitive tous les résultats déjà connus.” (p. 242)

Nous allons donc nous arrêter un instant sur les applications du calcul symbolique universel des formes différentielles.

8. Formes différentielles et physique classique

8.1. L’analyse vectorielle classique

Le formalisme de Cartan pour les formes différentielles a permis de profondément réinterpréter tout un ensemble de structures classiques peu à peu dégagées et systématisées par la physique au cours du XIX^e siècle pour exprimer des lois fondamentales de la *théorie des champs* comme le champ gravitationnel et le champ électromagnétique. Il éclaire en particulier les notions d’analyse vectorielle comme celles de gradient, de divergence et de rotationnel (dont nous avons rencontré d’importants exemples dans la section 4.6 du chapitre 4 du Vol I consacrée aux pinwheels de $V1$ comme “champs de phases”). C’est tout un édifice physico-mathématique qui se trouve ainsi couronné par un formalisme permettant de bien distinguer ce qui relève seulement du niveau de structure différentiable de ce qui relève en plus du niveau métrique.

8.1.1. Gradients, rotationnels, divergences.

Dans \mathbb{R}^3 la structure euclidienne permet d’identifier les vecteurs et les covecteurs tangents. Si ω est une 1-forme, on peut alors lui associer le champ de vecteurs tangents X_ω tel que pour tout autre champ de vecteurs Y on ait $\omega(Y) = \langle X_\omega, Y \rangle$ (produit scalaire définissant la structure euclidienne) et, dualement, on peut associer une ω_X à un $X = (X_x, X_y, X_z)$ (vecteur colonne dans la base $\{\partial_x, \partial_y, \partial_z\}$) par la même formule $\omega_X(Y) = \langle X, Y \rangle$. Évidemment $X_{\omega_X} = X$ et $\omega_{X_\omega} = \omega$. Si $f = f(x, y, z)$ est une fonction, $df = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$ (vecteur ligne dans la base $\{dx, dy, dz\}$) est une 1-forme et le champ X_{df} est le *gradient* ∇f de f . Comme df , il mesure la variation de f : il est nul sur les surfaces de niveau de f et maximal le long des lignes de plus grande pente qui leurs sont orthogonales. La formule scolaire

$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$ (vecteur colonne) dit en effet exactement que

$$\begin{aligned} \langle \nabla f, Y \rangle &= \frac{\partial f}{\partial x} Y_x + \frac{\partial f}{\partial y} Y_y + \frac{\partial f}{\partial z} Y_z = \frac{\partial f}{\partial x} dx(Y) + \frac{\partial f}{\partial y} dy(Y) + \frac{\partial f}{\partial z} dz(Y) \\ &= \left(\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz \right) (Y) = df(Y) . \end{aligned}$$

Nous allons vérifier que $\omega_X \wedge \omega_Y = \sigma_{X \wedge Y}$ avec $X \wedge Y$ le produit vectoriel défini ci-dessous et σ_A une 2-forme canoniquement associée à A dans \mathbb{R}^3 .

L'intégrale $\int_C \omega_X$ le long d'un chemin (une 1-chaîne) C s'appelait traditionnellement la "circulation" du champ X le long de C .

Par ailleurs, comme on a $\omega_X = X_x dx + X_y dy + X_z dz$, on peut aussi associer naturellement au champ de vecteurs X une 2-forme $\sigma = d\omega_X$. Comme on est en dimension $n = 3$, l'espace des 2-formes est de dimension $\frac{n(n-1)}{2} = 3$ et $\sigma = d\omega_X$ correspond également à une entité à 3 composantes dite "pseudo-vecteur" qui s'appelle son *rotationnel*.⁵⁹ Il mesure la "vorticité" du champ X au moyen de la formule

$$\nabla \wedge X = \text{rot}(X) = (\partial_y X_z - \partial_z X_y, \partial_z X_x - \partial_x X_z, \partial_x X_y - \partial_y X_x)$$

(vecteur colonne) où \wedge symbolise le *produit vectoriel*⁶⁰ de vecteurs défini par

$$A \wedge B = (A_y B_z - A_z B_y, A_z B_x - A_x B_z, A_x B_y - A_y B_x) .$$

En effet, du côté de ω_X on a bien (puisque $dx \wedge dx = dy \wedge dy = dz \wedge dz = 0$)

$$\begin{aligned} d\omega_X &= dX_x \wedge dx + dX_y \wedge dy + dX_z \wedge dz \\ &= (\partial_y X_x dy + \partial_z X_x dz) \wedge dx + (\partial_x X_y dx + \partial_z X_y dz) \wedge dy \\ &\quad + (\partial_x X_z dx + \partial_y X_z dy) \wedge dz \\ &= (\partial_y X_z - \partial_z X_y) dy \wedge dz + (\partial_z X_x - \partial_x X_z) dz \wedge dx + (\partial_x X_y - \partial_y X_x) dx \wedge dy . \end{aligned}$$

L'équivalence vient du fait que la base $\{\partial_x, \partial_y, \partial_z\}$ des vecteurs tangents est identifiable à la base $\{dy \wedge dz, dz \wedge dx, dx \wedge dy\}$ des 2-formes car on est en dimension 3. À travers cette identification, on peut associer à tout champ de vecteurs $A = (A_x, A_y, A_z)$ la 2-forme $\sigma_A = A_x dy \wedge dz + A_y dz \wedge dx + A_z dx \wedge dy$. La valeur $\sigma_A(X, Y)$ de σ_A sur le couple de vecteurs (X, Y) est ce qu'on appelle traditionnellement le "produit mixte" (A, X, Y) . Et l'intégrale $\int_S \sigma_A$ de σ_A sur une surface (une 2-chaîne) S s'appelait traditionnellement le "flux" du champ A à travers S . On a donc par définition $d\omega_X = \sigma_{\text{rot}(X)}$. À A on associe σ_A et à σ on associe A_σ et évidemment $\sigma_{A_\sigma} = \sigma$ et $A_{\sigma_A} = A$.

59. On parle de pseudo-vecteur parce que son sens dépend de l'orientation choisie pour l'espace ambiant.

60. Une autre notation conventionnelle est \times mais il est préférable dans notre contexte de garder le même symbole que pour le produit extérieur des formes différentielles.

Notons alors que l'on peut également considérer la 3-forme $d\sigma_A$ et, comme toutes les 3-formes sont des multiples de $dx \wedge dy \wedge dz$, on obtient

$$d\sigma_A = (\partial_x A_x + \partial_y A_y + \partial_z A_z) dx \wedge dy \wedge dz$$

le facteur de proportionnalité étant la *divergence* de A qui mesure les "sources" de A .

On notera que si X et A sont deux champs de vecteurs, la 3-forme $\sigma_A \wedge \omega_X = \omega_X \wedge \sigma_A$ est

$$\begin{aligned} \sigma_A \wedge \omega_X &= (A_x dy \wedge dz + A_y dz \wedge dx + A_z dx \wedge dy) \wedge (X_x dx + X_y dy + X_z dz) \\ &= (A_x X_x + A_y X_y + A_z X_z) dx \wedge dy \wedge dz = \langle A, X \rangle dx \wedge dy \wedge dz . \end{aligned}$$

Cette réinterprétation de l'analyse vectorielle classique a permis de retrouver toutes les formules de cette dernière comme des conséquences directes des formules universelles du calcul sur les formes différentielles. L'avantage était considérable non seulement sur le plan théorique conceptuel mais aussi sur le plan technique car les formes différentielles ne dépendent que de la structure différentielle et pas de la structure euclidienne de l'espace.

Par exemple la formule $d^2 = 0$ implique immédiatement deux identités constamment utilisées en physique depuis le XVIII^e siècle (et encore apprise aujourd'hui dans la plupart des manuels voulant faire l'économie du formalisme général des formes différentielles) :

1. Si l'on part de f , ∇f correspond à df et donc $\text{rot}(\nabla f)$ correspond à $d^2 f$ et $d^2 f \equiv 0$: le rotationnel d'un gradient est identiquement nul.
2. Si l'on part de X , son rotationnel $\text{rot}(X)$ correspond à $\sigma = \text{rot}(X) = d\omega_X$. La divergence de $\text{rot}(X)$ correspond à $d^2 \omega_X$ et $d^2 \omega_X \equiv 0$: la divergence d'un rotationnel est identiquement nulle.

D'autres formules classiques apparemment un peu compliquées apparaissent comme de simples conséquences du calcul sur les formes différentielles. Les propriétés de linéarité bien sûr, mais également des formules de comportement par rapport à des produits. Par exemple

1. Soit X un champ de vecteurs et f une fonction. Évidemment, $\omega_{fX} = f\omega_X$ et donc

$$d\omega_{fX} = d(f\omega_X) = df \wedge \omega_X + f d\omega_X .$$

On en déduit immédiatement que

$$\text{rot}(fX) = \nabla f \wedge X + f \text{rot}(X) .$$

2. Calculons de même $\text{div}(fA)$. On utilise $d\sigma_{fA} = \text{div}(fA) dx \wedge dy \wedge dz$. Mais $\sigma_{fA} = f\sigma_A$ et $d\sigma_{fA} = df \wedge \sigma_A + f d\sigma_A$. Or, pour un champ de vecteurs A et

une 1-forme α , la 3-forme $\alpha \wedge \sigma_A$ est

$$\begin{aligned}\alpha \wedge \sigma_A &= (\alpha_x dx + \alpha_y dy + \alpha_z dz) \wedge (A_x dy \wedge dz + A_y dz \wedge dx + A_z dx \wedge dy) \\ &= (\alpha_x A_x + \alpha_y A_y + \alpha_z A_z) dx \wedge dy \wedge dz \\ &= (X_\alpha, A) dx \wedge dy \wedge dz .\end{aligned}$$

Donc

$$\operatorname{div}(fA) = (\nabla f, A) + f \operatorname{div}(A) = A(f) + f \operatorname{div}(A) .$$

3. Considérons maintenant le produit extérieur $X \wedge Y$ de deux champs de vecteurs et calculons sa divergence. À chaque champ est associé une 1-forme ω_X et ω_Y et l'on obtient donc une 2-forme $\omega_X \wedge \omega_Y$. On vérifie que cette 2-forme est bien celle associée à $X \wedge Y$, i.e.

$$\begin{aligned}\omega_X \wedge \omega_Y &= \sigma_{X \wedge Y} \\ &= (X_y Y_z - X_z Y_y) dy \wedge dz + (X_z Y_x - X_x Y_z) dz \wedge dx \\ &\quad + (X_x Y_y - X_y Y_x) dx \wedge dy .\end{aligned}$$

En prenant $d\sigma_{X \wedge Y}$ on obtient la 3-forme $d\sigma_{X \wedge Y} = \operatorname{div}(X \wedge Y) dx \wedge dy \wedge dz$. Mais $d\sigma_{X \wedge Y} = d(\omega_X \wedge \omega_Y)$. Or, étant données deux 1-formes α et β , on a $d(\alpha \wedge \beta) = d\alpha \wedge \beta - \alpha \wedge d\beta$, et par ailleurs, pour deux champs de vecteurs A et X , la 3-forme $\sigma_A \wedge \omega_X$ est, nous l'avons vu, $(A, X) dx \wedge dy \wedge dz$. Par conséquent, comme $d\omega_X$ et $d\omega_Y$ correspondent aux rotationnels $\operatorname{rot}(X)$ et $\operatorname{rot}(Y)$, la formule de Leibniz

$$d(\omega_X \wedge \omega_Y) = d\omega_X \wedge \omega_Y - \omega_X \wedge d\omega_Y$$

implique la formule classique

$$\operatorname{div}(X \wedge Y) = \langle \operatorname{rot}(X), Y \rangle - \langle X, \operatorname{rot}(Y) \rangle .$$

8.1.2. Applications classiques du théorème de Stokes.

D'autres théorèmes clés partout utilisés depuis le XVIII^e siècle en physique deviennent des cas particuliers du théorème de Stokes dont les conséquences sont innombrables, en particulier en ce qui concerne des *lois de conservation*. Par exemple en dimension 3 :

1. Si γ est une courbe (différentiable) de M d'extrémités $x = a$ et $x = b$, son bord est la chaîne de dimension 0 $\partial\gamma = b - a$ et si f est une 0-forme, i.e. une fonction sur M , $\int_{b-a} f = \int_\gamma df$, autrement dit, $f(b) - f(a) = \int_\gamma \nabla f \cdot dx$: ce que l'on appelait autrefois la "circulation" du champ de gradient ∇f le long d'une courbe γ ne dépend que des extrémités de la courbe et pas du chemin qu'est la courbe. En particulier, si $a = b$, $\int_\gamma \nabla f \cdot dx = 0$: l'intégrale d'un champ de gradient sur une courbe fermée (i.e. un "cycle") est nulle. Ce n'est pas le cas pour un champ de vecteurs qui n'est pas un champ de gradient.

2. Si X est un champ de vecteurs sur M et si S est une surface de bord $\partial S = \gamma$, on a $\int_{\gamma} \omega_X = \int_S d\omega_X$, autrement dit, en notant dn l'élément de longueur sur les normales à S , $\int_{\gamma} X dx = \int_S \text{rot}(X) \cdot dn$: la circulation d'un champ de vecteurs le long d'un cycle est égale à ce qu'on appelait autrefois le "flux" de son rotationnel à travers toute surface bordée par ce cycle. En particulier si $X = \nabla f$ est de gradient, comme $\text{rot}(\nabla f) \equiv 0$, ce flux du rotationnel est nul. Par ailleurs, si S est fermée (sans bord), le flux du rotationnel est automatiquement nul quel que soit X .

3. Si A est un champ de vecteurs sur M et si V est un volume limité par une surface $S = \partial V$, on a $\int_S \sigma_A = \int_V d\sigma_A$, autrement dit $\int_S A \cdot dn = \int_V \text{div}(A) dv$, formule dite d'Ostrogradski-Gauss-Green. En particulier, si $\text{div}(A) = 0$ (entre autres si $A = \text{rot}(X)$ comme nous venons de le voir), le flux de A à travers S est nul.

8.1.3. L'émergence du non commutatif.

L'analyse vectorielle classique dont nous venons d'esquisser certains éléments a été une grande et difficile conquête du XIX^e siècle. Stimulée par la nécessité de comprendre des propriétés physiques très énigmatiques comme celles du "pseudo-vecteur" qu'est le champ magnétique, elles ont conduit à dégager progressivement un calcul vectoriel en quelque sorte "non commutatif". Nous avons déjà cité de nombreux mathématiciens et physiciens qui en ont été les inventeurs. On peut citer également William Hamilton (1805-1865) et ses quaternions, Hermann Grassmann (1809-1877), William Clifford (1845-1879), ou Josiah Willard Gibbs (1839-1903).

Un concept clé a été celui du produit vectoriel de deux vecteurs X, Y de \mathbb{R}^3 introduit par Gibbs en 1881, à la suite des travaux de Grassmann et de Hamilton. Nous avons vu qu'il consiste à porter dans la direction orthogonale au plan (X, Y) un vecteur $X \wedge Y$ ayant pour longueur l'aire du parallélogramme (X, Y) et la direction telle que le trièdre $(X, Y, X \wedge Y)$ soit d'orientation positive. Par définition, cette opération vectorielle dépend (i) de la structure euclidienne de \mathbb{R}^3 et (ii) de son orientation. Un des grands intérêts de sa traduction par la 2-forme $\sigma_{X \wedge Y}$ est de s'affranchir de la structure euclidienne.

Ces grands géomètres définirent également le produit mixte $[X, Y, Z]$ de trois vecteurs X, Y, Z comme le volume du parallélépipède (X, Y, Z) et comprirent qu'il s'agissait du *déterminant* de la matrice (X, Y, Z) calculé dans une base orthonormée

d'orientation positive. Ils prouvèrent la formule

$$\begin{aligned} [X, Y, Z] &= \langle X \wedge Y, Z \rangle \\ &= (X_y Y_z - X_z Y_y) Z_x + (X_z Y_x - X_x Y_z) Z_y + (X_x Y_y - X_y Y_x) Z_z \\ &= \det \begin{pmatrix} X_x & X_y & X_z \\ Y_x & Y_y & Y_z \\ Z_x & Z_y & Z_z \end{pmatrix} \end{aligned}$$

montrant la relation entre le produit vectoriel et le produit scalaire. Ainsi s'élabora tout le nouveau domaine de l'algèbre multilinéaire alternée que la théorie des formes différentielles détachera des structures euclidiennes sous-jacentes pour ne plus les faire dépendre que des structures vectorielles et différentiables.

La linéarité du produit vectoriel ($X \wedge (\lambda Y + Y') = \lambda(X \wedge Y) + X \wedge Y'$) ne soulevait pas de questions. En revanche sa non-commutativité et sa non-associativité en soulevaient de profondes. D'une part $Y \wedge X = -X \wedge Y$ montrait que le produit vectoriel était un produit très spécial et d'autre part le calcul montrait que

$$X \wedge (Y \wedge Z) = \langle X, Z \rangle Y - \langle X, Y \rangle Z ,$$

ce qui impliquait la non-associativité puisque

$$\begin{aligned} X \wedge (Y \wedge Z) &= \langle X, Z \rangle Y - \langle X, Y \rangle Z \\ (X \wedge Y) \wedge Z &= -Z \wedge (X \wedge Y) = -\langle Z, Y \rangle X + \langle Z, X \rangle Y . \end{aligned}$$

Mais, étant donnée la symétrie du produit scalaire, on obtient par permutation circulaire l'identité

$$\begin{aligned} X \wedge (Y \wedge Z) + Y \wedge (Z \wedge X) + Z \wedge (X \wedge Y) \\ &= \langle X, Z \rangle Y - \langle X, Y \rangle Z + \langle Y, X \rangle Z - \langle Y, Z \rangle X + \langle Z, Y \rangle X - \langle Z, X \rangle Y \\ &= 0 \end{aligned}$$

autrement dit l'identité de Jacobi, identité fondamentale des algèbres de Lie que nous retrouverons constamment par la suite à partir de la section 2 du chapitre 5 où nous en donnerons une présentation plus précise.

Muni du produit vectoriel \wedge , l'espace vectoriel \mathbb{R}^3 devient ainsi une algèbre de Lie. Le lien avec la structure euclidienne et l'orientation de \mathbb{R}^3 peut alors être considérablement éclairci si l'on remarque que cette algèbre de Lie est *isomorphe* à l'algèbre de Lie $\mathfrak{so}(3)$ du groupe des rotations $SO(3)$ de \mathbb{R}^3 , c'est-à-dire des isométries de \mathbb{R}^3 (les isomorphismes linéaires préservant la métrique euclidienne) préservant l'orientation. Les éléments de $\mathfrak{so}(3)$ sont les matrices antisymétriques

$m = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}$, le produit de deux matrices m et m' étant leur commutateur $[m, m'] = mm' - m'm$. L'isomorphisme entre (\mathbb{R}^3, \wedge) et $(\mathfrak{so}(3), [,])$ consiste à associer à $X = (X_x, X_y, X_z)$ la matrice antisymétrique $m_X = \begin{pmatrix} 0 & -X_z & X_y \\ X_z & 0 & -X_x \\ -X_y & X_x & 0 \end{pmatrix}$.

On vérifie immédiatement que l'on a bien $m_{X \wedge Y} = [m_X, m_Y]$.

On peut même aller plus loin dans l'interprétation géométrique en introduisant l'interprétation de $\mathfrak{so}(3)$ et de $SO(3)$ au moyen du corps non commutatif des *quaternions* \mathbb{H}_Q de Hamilton que nous rencontrerons dans la section 4.6.2 du chapitre 15.⁶¹ Cela explique le rôle exceptionnel de la dimension 3. \mathbb{H}_Q est de dimension 4 comme \mathbb{R} -vectoriel et les rotations de $SO(3)$ se retrouvent à partir de sa sphère unité, la structure de groupe et d'algèbre de Lie de $SO(3)$ et $\mathfrak{so}(3)$ étant intimement liée à la structure de corps de \mathbb{H}_Q . Or un théorème fondamental dû à Frobenius démontre qu'il n'existe de structure de corps sur \mathbb{R}^n que pour $n = 1$, $n = 2$ (corps des complexes $\mathbb{C} = \mathbb{R}^2$), $n = 4$ (corps non commutatif des quaternions \mathbb{H}_Q) et $n = 8$ (corps ni commutatif ni associatif des octonions de Graves et Cayley). Nous y reviendrons au chapitre 15, section 4.6.2.

8.2. Digression : l'exemple des équations de Maxwell

Permettons-nous ici une digression. Bien que cela soit en dehors de notre sujet, il est très intéressant de voir comment la théorie des formes différentielles a pu permettre d'unifier certaines grandes équations de la physique fondamentale, équations issues elles-mêmes d'une extraordinaire convergence de données expérimentales très diversifiées et de formalisations partielles. C'est en effet le même genre d'unification que nous voudrions essayer de rendre possible pour certaines des formalisations partielles de la géométrie de la vision en termes d'équations différentielles neuronales. Nous prendrons l'exemple des équations de Maxwell pour le champ électromagnétique.

8.2.1. L'unification des lois : Gauss, Thomson, Faraday, Ampère.

Les équations sont gravées sur une stèle de la tombe de leur auteur, James Clerk Maxwell (1831-1879) à Edimbourg (figure 17). Rappelons-en les principales étapes.

La deuxième équation de la stèle (dite de Maxwell-Gauss) régit le champ électrique E considéré pour l'instant comme un champ de vecteurs $E(a, t)$ sur \mathbb{R}^3 ($a = (x, y, z) \in \mathbb{R}^3$ et t est le temps).⁶² La notion de champ de vecteurs et les règles du

61. Traditionnellement \mathbb{H}_Q se note \mathbb{H} . Mais comme nous réservons la notation \mathbb{H} au groupe de Heisenberg nous indexons \mathbb{H} par Q .

62. Nous verrons plus bas que c'est en fait une 1-forme.

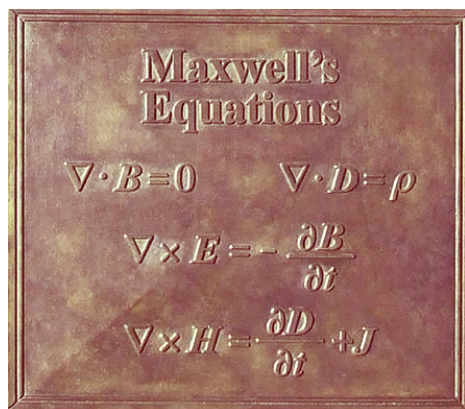


FIGURE 17. Les équations de Maxwell gravées sur sa tombe à Edimbourg.

calcul vectoriel existaient déjà à l'époque de Newton avec le champ de gravitation. Dans le cas des équations *dans un espace vide* en dehors des sources, cette équation, $\operatorname{div} E = \frac{\rho}{\varepsilon_0}$ relie la divergence de $E(a, t)$ à la densité de charges électriques $\rho(a, t)$, le facteur de proportionnalité étant $\frac{1}{\varepsilon_0}$ avec ε_0 la "permittivité diélectrique" du vide. Elle s'écrit déjà chez Maxwell comme le produit scalaire $\nabla \cdot E$ de l'opérateur vectoriel ∇ avec le vecteur E . La forme intégrale de cette équation est donnée par le théorème de Gauss disant que le flux de E à travers une surface fermée S délimitant un volume V est $\int_S E \cdot dn = \frac{1}{\varepsilon_0} \int_V \rho dv = \frac{Q_V}{\varepsilon_0}$ ($dv = dx dy dz$) où n est la normale sortante à S et Q_V est la charge électrique à l'intérieur de V . Comme le théorème de Stokes (Ostrogradski-Gauss-Green) dit que $\int_S E \cdot dn = \int_V \operatorname{div}(E) dv$, cela est bien équivalent à $\operatorname{div}(E) = \frac{1}{\varepsilon_0} \rho$.

Au-delà du cas du vide, c'est-à-dire lorsque l'on considère des milieux continus possédant des propriétés macroscopiques comme celles de polarisation ou d'aimantation, Maxwell a été conduit, et ce fut l'un de ses apports majeurs, à introduire à côté du champ E un champ de "déplacement" D , appelé aussi le champ *d'induction* électrique, mesurant la façon dont les charges du milieu considéré réagissent au champ électrique. L'équation des sources devient alors $\nabla \cdot D = \rho$ (la seconde équation de la stèle). Le lien entre E et D peut être compliqué si le milieu est un milieu continu hétérogène et peut contenir des termes de "polarisation", mais dans les milieux "linéaires" il est de la forme $D = \varepsilon(E)$ où ε est une matrice de permittivité et, si le milieu est homogène et isotrope, ε est simplement une constante qui agit comme un scalaire. Dans le cas du vide $D = \varepsilon_0 E$ et l'on retrouve l'équation $\nabla \cdot E = \frac{\rho}{\varepsilon_0}$.

La première équation de la stèle (dite de Maxwell-Thomson) est liée aux travaux de Thomson et concerne le champ magnétique B . Elle s'écrit $\nabla \cdot B = 0$ (i.e. $\operatorname{div} B = 0$) et exprime le fait fondamental qu'il n'y a pas de "sources ponctuelles" du champ

magnétique, autrement dit pas de “monopôles” magnétiques : aussi finement que l’on divise un matériau magnétique on n’obtient que des aimants microscopiques, i.e. des dipôles. Cette loi de *conservation* implique d’après Stokes que le flux de B à travers une surface fermée S est toujours nul : $\int_S B \cdot dn = \int_V \operatorname{div}(B) \, dv = 0$. Insistons sur le fait qu’une des grandes découvertes du XIX^e siècle a été que B n’est pas un vecteur mais un *pseudo-vecteur*.⁶³

Nous avons vu que la divergence d’un rotationnel est toujours nulle et que, réciproquement, tout champ dont la divergence est identiquement nulle est localement un rotationnel. Une conséquence essentielle de la loi de conservation est donc l’existence *locale* d’un *potentiel vecteur* A dont B est le rotationnel : $B = \nabla \wedge A$. Le potentiel vecteur A n’est défini qu’à un gradient ∇f près car le rotationnel d’un gradient est toujours nul. Le champ scalaire f s’appelle *une transformation de jauge* de la théorie.

Les *relations dynamiques* entre E et B ayant conduit au concept *unifié* de champ électro-magnétique sont exprimées par les découvertes majeures de Faraday et d’Ampère. D’abord, et c’est la troisième loi de la stèle (dite de Maxwell-Faraday), la variation d’un champ magnétique induit un champ électrique et donc un courant : $\nabla \wedge E = -\frac{\partial B}{\partial t}$. Ce champ est donné par son rotationnel $\nabla \wedge E = \operatorname{rot}(E)$. Sa forme intégrale est la loi de Faraday qui est l’un des grands exemples du théorème de Stokes. La circulation $\int_C E \cdot dl$ du champ E le long d’un circuit C , appelée “force électromotrice d’induction”, est égale au flux $\int_S \operatorname{rot}(E) \cdot dn$ de son rotationnel à travers toute surface S de bord C et à l’opposé de la variation du flux du champ magnétique.

Mais B est lui-même localement un rotationnel $B = \nabla \wedge A$, le potentiel vecteur A n’étant défini qu’à un gradient ∇f près. On a donc localement une relation entre E et $-\frac{\partial A}{\partial t}$ la différence devant être annulée par un rotationnel. C’est ainsi que s’introduit, puisque le rotationnel d’un gradient $\operatorname{rot}(\nabla V)$ est nul, en plus du potentiel vecteur qui est spécifique de B un *potentiel scalaire* V local qui est spécifique de E , la formule étant maintenant, en termes des deux potentiels, $E = -\nabla V - \frac{\partial A}{\partial t}$. A est défini à ∇f près et V à $-\frac{\partial f}{\partial t}$ près.

La quatrième équation fondamentale de la stèle (dite de Maxwell-Ampère) résume l’apport crucial d’Ampère et de Maxwell lui-même en disant qu’un champ magnétique peut être engendré par des courants électriques et par des variations du champ électrique. Dans le vide elle se formule par

$$\nabla \wedge B (= \operatorname{rot}(B)) = \mu_0 j + \mu_0 \varepsilon_0 \frac{\partial E}{\partial t} = \mu_0 \left(j + \varepsilon_0 \frac{\partial E}{\partial t} \right) = \mu_0 (j + j_D)$$

63. Nous verrons plus bas que c’est en fait une 2-forme.

où $j(a, t)$ est la densité de courant, $j_D = \varepsilon_0 \frac{\partial E}{\partial t}$ le “courant de déplacement” et μ_0 la “perméabilité magnétique” du vide. L’introduction de j_D était nécessaire pour unifier les équations en un tout cohérent.

La forme intégrale de cette dernière équation dit, toujours d’après Stokes, que la circulation $\int_C B dl$ du champ B le long d’un circuit C est égale au flux $\int_S \text{rot}(B) \cdot dn$ de son rotationnel à travers toute surface S de bord C et donc au flux du courant total $\mu_0 (j + \varepsilon_0 \frac{\partial E}{\partial t})$.

Dans les milieux continus, E est remplacé par le champ de “déplacement” D et B par le champ H d’“excitation magnétique”. $H = \frac{1}{\mu_0} B$ dans le vide mais il peut contenir en général des termes “d’aimantation”. L’équation devient alors (quatrième équation de la stèle) $\nabla \wedge H = j + \frac{\partial D}{\partial t}$. En fait, le lien entre d’une part E et D et d’autre part B et H est non seulement physiquement mais également *mathématiquement* subtil et il est tout à fait remarquable que l’on trouve les quatre champs chez Maxwell, E et B correspondant aux deux équations sans termes de contexte (sources et courant) et D et H aux deux équations faisant au contraire intervenir le contexte. Nous reviendrons sur ce point essentiel à la section 8.2.4 à propos de la dualité de Hodge.

Les deux dernières équations montrent que E et B sont dynamiquement inséparables et doivent être interprétés comme les composantes d’une entité unique, le *champ électro-magnétique* (EM). Il s’agit de la première grande *unification* de théories physiques.

Les relations de l’analyse vectorielle que nous avons explicitées permettent de déduire immédiatement des propriétés essentielles du champ EM. Par exemple, comme la divergence d’un rotationnel est identiquement nulle, l’équation de Maxwell-Ampère implique (cas du vide) $\nabla \cdot (\nabla \wedge B) = 0 = \mu_0 (\nabla \cdot j + \varepsilon_0 \nabla \cdot \frac{\partial E}{\partial t})$. Mais d’un côté on a $\nabla \cdot \frac{\partial E}{\partial t} = \frac{\partial (\nabla \cdot E)}{\partial t}$ car les dérivées spatiales et temporelles sont indépendantes et d’un autre côté, d’après l’équation de Maxwell-Gauss, on a $\nabla \cdot E = \frac{\rho}{\varepsilon_0}$ et donc $\nabla \cdot j + \frac{\partial \rho}{\partial t} = 0$. Cette équation

$$\text{div } j + \frac{\partial \rho}{\partial t} = 0$$

est l’équation de conservation de base du champ EM.

On peut aussi appliquer l’identité d’analyse vectorielle concernant le *double* rotationnel d’un champ de vecteurs⁶⁴ :

$$\nabla \wedge (\nabla \wedge E) = \nabla \langle \nabla, E \rangle - \langle \nabla, \nabla \rangle E = \nabla (\nabla \cdot E) - \Delta E .$$

64. On applique la formule $X \wedge (Y \wedge Z) = \langle X, Z \rangle Y - \langle X, Y \rangle Z$.

Mais d'après Maxwell-Faraday $\nabla \wedge E = -\frac{\partial B}{\partial t}$, d'après Maxwell-Gauss $\nabla \cdot E = \frac{\rho}{\varepsilon_0}$ et d'après Maxwell-Ampère $\nabla \wedge \frac{\partial B}{\partial t} = \mu_0 \frac{\partial j}{\partial t} + \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2}$. Donc

$$\begin{aligned} \nabla \wedge \left(-\frac{\partial B}{\partial t} \right) &= - \left(\mu_0 \frac{\partial j}{\partial t} + \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2} \right) = \nabla \left(\frac{\rho}{\varepsilon_0} \right) - \Delta E \\ \Delta E - \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2} &= \nabla \left(\frac{\rho}{\varepsilon_0} \right) + \mu_0 \frac{\partial j}{\partial t}. \end{aligned}$$

Mais cette équation linéaire aux dérivées partielles du second ordre est une *équation des ondes* associée à la vitesse de propagation c définie par $\frac{1}{c^2} = \mu_0 \varepsilon_0$.

De même,

$$\nabla \wedge (\nabla \wedge B) = \nabla (\nabla \cdot B) - \Delta B$$

et d'après Maxwell-Ampère $\nabla \wedge B = \mu_0 j + \mu_0 \varepsilon_0 \frac{\partial E}{\partial t}$, d'après Maxwell-Thomson $\nabla \cdot B = 0$ et d'après Maxwell-Faraday $\nabla \wedge E = -\frac{\partial B}{\partial t}$. Donc

$$\begin{aligned} \nabla \wedge \left(\mu_0 j + \mu_0 \varepsilon_0 \frac{\partial E}{\partial t} \right) &= \nabla \wedge \mu_0 j - \mu_0 \varepsilon_0 \frac{\partial^2 B}{\partial t^2} = -\Delta B \\ \Delta B - \mu_0 \varepsilon_0 \frac{\partial^2 B}{\partial t^2} &= -\nabla \wedge \mu_0 j \end{aligned}$$

qui est également une équation des ondes de même vitesse c , à savoir la vitesse de la lumière.

Les conséquences de l'unification de l'électro-magnétisme par Maxwell ont été considérables. D'abord sur le plan technique puisque c'est à partir de là qu'on a pu développer la physique des ondes électro-magnétiques et tout ce qui en découle. Ensuite sur le plan théorique lorsqu'on a découvert que le groupe de relativité naturel par rapport auquel les équations de Maxwell sont covariantes n'était pas le groupe de Galilée de la mécanique classique mais le groupe de Lorentz. Cette compétition entre deux groupes de relativité conduira Einstein à unifier la physique en appliquant le groupe de Lorentz à la mécanique, ce qui donnera la révolution de la relativité restreinte.

8.2.2. Le type des entités "champs".

Tout au long de cette épopée, les équations ont été présentées en termes d'entités de type vectoriel. Mais les spécialistes ont progressivement compris que ce n'était pas parce que ces entités possédaient trois composantes (E_x, E_y, E_z) et (B_x, B_y, B_z) associées aux trois directions de l'espace qu'elles étaient pour autant des "vrais" vecteurs. Pour être de vrais vecteurs encore eût-il fallu qu'elles se *transformassent* comme des vecteurs. Or de nombreuses expériences montrèrent que ce n'est pas le cas.

1. Par exemple lorsque l'on dilate l'espace par un changement d'échelle λ , E n'est pas dilaté de λ mais au contraire contracté de λ . On a ainsi compris

progressivement que son type n'est pas celui d'un vecteur mais d'un *covecteur*. La bonne description de E est non pas celle d'un champ de vecteurs mais d'un champ de covecteurs, autrement dit d'une 1-forme $\omega_E = E_x dx + E_y dy + E_z dz$. En effet, une 1-forme ω agit sur les vecteurs X par $\omega(X) = \langle \omega; X \rangle$ et cette valeur a un sens intrinsèque. Si un changement d'échelle transforme X en λX alors il doit également transformer ω en $\frac{1}{\lambda} \omega$ de façon à ce que $\langle \frac{1}{\lambda} \omega; \lambda X \rangle = \langle \omega; X \rangle$.

2. De même, on découvrit que B qui possédait le statut d'un pseudo-vecteur dépendant de l'orientation de l'espace ambiant⁶⁵ devait être représenté par la matrice antisymétrique

$$\sigma_B = \begin{pmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{pmatrix}$$

En fait B possède le type d'une 2-forme $\sigma_B = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy$.

Qui plus est, la formule fondamentale décrivant le mouvement d'une particule de charge électrique élémentaire q sous l'action d'un champ EM posait un problème fondamental car elle n'était pas compatible avec la relativité galiléenne. Elle s'écrivait en effet

$$F = q(E + v \wedge B)$$

F étant la force exercée sur la particule et v étant sa vitesse. Or F est invariante sous l'action du groupe de Galilée alors que v ne l'est pas et peut même être annulée par changement de repère inertiel. Dans un tel repère où $v = 0$, l'action du champ magnétique B sur la particule disparaît.

De nombreuses solutions ont été proposées pour résoudre cette aporie. La plus connue avant la solution d'Einstein et son instauration de la relativité restreinte par changement du groupe de relativité de la théorie était celle de l'*éther*. On postulait que l'espace était rempli d'un éther possédant des propriétés substantielles assez magiques (fluide subtil inobservable, ou solide élastique infiniment rigide mais sans viscosité, ou encore medium composé de tourbillons microscopiques) permettant aux ondes EM de se propager comme des vibrations. Cet éther faisait de l'*espace vide* un *milieu physique* et constituait un référentiel *absolu* (comme l'espace absolu de Newton qui était à la fois un éther mécanique et le "sensorium Dei" permettant la transmission instantanée des forces gravitationnelles). Du coup, la vitesse v relative à ce repère acquérait alors une valeur absolue (non galiléennement relative).

Mais cela ne suffisait pas. En effet, même dans l'hypothèse de l'existence d'un espace absolu, il restait la règle d'additivité des vitesses par changement de repère et la

65. Si un circuit électrique induit un B et si l'on inverse le sens du courant dans le circuit, le B induit change de sens.

vitesse de propagation des ondes EM (vitesse c de la lumière) devait donc dépendre du repère inertiel choisi. Or les expériences cruciales d'interférométrie d'Albert Abraham Michelson et Edward Morley (dans les années 1880) montrèrent que ce n'est pas le cas. On introduisit alors des corrections (contraction des longueurs dans la direction des vitesses, dilatation des durées, relativité de la simultanéité) permettant d'expliquer la constance de c . Cela conduisit aux nouvelles formules de composition des vitesses dues à Hendrik Lorentz (1853-1928) et Henri Poincaré jusqu'à ce qu'Einstein (1905) élimine définitivement l'éther qui "attribue des propriétés physiques à l'espace" et introduise le groupe de Lorentz comme nouveau groupe de relativité de la physique. Chez Einstein, les contractions des longueurs et les dilatations des durées ne sont plus des phénomènes réels induits par des mouvements dans l'éther mais des effets purement *cinématiques* de changements de repères inertiels.⁶⁶

8.2.3. La synthèse des formules (1) et (3) et l'espace-temps.

Une fois compris le type des entités mathématiques décrivant les champs, les équations de Maxwell (dans le vide pour simplifier) purent alors être exprimées de façon beaucoup plus synthétique. On a d'abord les lois n'impliquant ni les sources ni les courants.

- (i) Comme $d\sigma_B = \text{div}(B) dV$ ($dV = dx \wedge dy \wedge dz$), la première loi (Maxwell-Thomson) $\text{div} B = \nabla \cdot B = 0$ dit que $d\sigma_B = 0$, i.e. que σ_B est une 2-forme *fermée*. Et comme une forme fermée est localement exacte, σ_B est localement de la forme $\sigma_B = d\omega_A$, A étant un potentiel vecteur. A est défini à un gradient ∇f près car ω_A est définie à df près puisque $d(\omega_A + df) = d\omega_A + d^2f = d\omega_A$.
- (ii) La troisième loi (Maxwell-Faraday) $\nabla \wedge E = -\frac{\partial B}{\partial t}$ signifie quant à elle l'égalité des 2-formes $d\omega_E = -\frac{\partial \sigma_B}{\partial t}$. Elle montre comment les dérivées spatiales de E sont intrinsèquement liées à la dérivée temporelle de B .

Un nouveau pas essentiel dans la synthèse des formules a alors consisté à comprendre que ces deux lois établissant un rapport de *structure* entre E et B pouvaient être *unifiées* en une seule, à condition de passer de \mathbb{R}^3 à \mathbb{R}^4 c'est à dire d'un espace et d'un temps découplés à un *espace-temps* les unifiant en une entité unique. Le champ σ_B étant une 2-forme sur \mathbb{R}^3 , il peut être prolongé trivialement à \mathbb{R}^4 . Le champ ω_E étant une 1-forme sur \mathbb{R}^3 , on peut la transformer en une 2-forme sur \mathbb{R}^4 en considérant son produit extérieur $\omega_E \wedge dt$ par dt . C'est la remarque clé. Introduisons la 2-forme sur \mathbb{R}^4

$$\mathfrak{F} = \sigma_B + \omega_E \wedge dt$$

66. En relativité générale, la métrique locale de l'espace-temps est d'origine physique et donc l'espace-temps possède en tant que tel des propriétés physiques. Mais, comme y a insisté Einstein, cela n'est en rien un retour de quelque ancienne conception de l'éther. Cela signifie seulement qu'il existe en quelque sorte des "états métriques" de l'espace-temps qui sont en interaction avec les champs physiques.

correspondant à l'unification des champs exprimée en notation classique par la matrice antisymétrique

$$\begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix}$$

et calculons sa différentielle extérieure. On a (en signalant en indice si l'on est dans \mathbb{R}^3 ou \mathbb{R}^4)

$$d\mathfrak{F} = d_4\sigma_B + d_4\omega_E \wedge dt - \omega_E \wedge d^2t = d_4\sigma_B + d_4\omega_E \wedge dt .$$

Faisons le calcul explicitement. On a

$$\begin{aligned} \sigma_B &= B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy \\ \omega_E &= E_x dx + E_y dy + E_z dz . \end{aligned}$$

Quand on calcule $d_4\sigma_B$ on commence par

$$\left(\frac{\partial B_x}{\partial x} dx + \frac{\partial B_x}{\partial y} dy + \frac{\partial B_x}{\partial z} dz + \frac{\partial B_x}{\partial t} dt \right) \wedge dy \wedge dz , \text{ etc.}$$

On voit que l'on obtient ainsi $d_3\sigma_B = \text{div}(B) d_3V$ plus les termes

$$\frac{\partial B_x}{\partial t} dt \wedge dy \wedge dz + \frac{\partial B_y}{\partial t} dt \wedge dz \wedge dx + \frac{\partial B_z}{\partial t} dt \wedge dx \wedge dy .$$

Mais ces termes supplémentaires donnent $dt \wedge \frac{\partial \sigma_B}{\partial t}$. D'un autre côté, $d_4\omega_E \wedge dt$ donne $d_3\omega_E \wedge dt$ plus des termes $(\frac{\partial E_x}{\partial t} dt \wedge dx) \wedge dt$, etc. mais ces derniers termes contenant les dérivées temporelles de E s'annulent puisque $dt \wedge dt = 0$. On obtient donc en définitive (puisque changer dt de côté ne change pas de signe pour des 3-formes $dt \wedge dx \wedge dy$, etc. ⁶⁷)

$$\begin{aligned} d\mathfrak{F} &= d_4\sigma_B + d_4\omega_E \wedge dt = d_3\sigma_B + dt \wedge \frac{\partial \sigma_B}{\partial t} + d_3\omega_E \wedge dt \\ &= dt \wedge \frac{\partial \sigma_B}{\partial t} - \frac{\partial \sigma_B}{\partial t} \wedge dt \text{ (car } d_3\sigma_B = \text{div}(B) d_3V = 0 \text{ et } d_3\omega_E = -\frac{\partial \sigma_B}{\partial t}) \\ &= 0 . \end{aligned}$$

On voit ainsi comment, en unifiant les dérivées spatiales et temporelles en des dérivées spatio-temporelles, on peut absorber la relation dynamique entre E et B dans une propriété de fermeture de 2-forme \mathfrak{F} sur \mathbb{R}^4 .

67. C'est la formule d'(anti)commutativité $\rho \wedge \omega = (-1)^{kr} \omega \wedge \rho$.

Étant fermée, \mathfrak{F} est localement exacte et il existe donc localement des 1-formes α sur \mathbb{R}^4 telles que $\mathfrak{F} = d\alpha$. Notons $\alpha = \omega_A - Vdt$, ω_A étant une forme sur \mathbb{R}^3 . On obtient par conséquent

$$\mathfrak{F} = \sigma_B + \omega_E \wedge dt = d_3\omega_A + dt \wedge \frac{\partial\omega_A}{\partial t} - dV \wedge dt$$

soit $\sigma_B = d\omega_A$ et $\omega_E = -\frac{\partial\omega_A}{\partial t} - dV$. Cela correspond au potentiel vecteur magnétique A défini par $B = \nabla \wedge A$ et au potentiel scalaire électrique V défini par $E = -\frac{\partial\omega_A}{\partial t} - \nabla V$. Le potentiel EM (A, V) n'est défini qu'à une différentielle exacte df près, f étant une fonction sur \mathbb{R}^4 . Comme $d_4f = d_3f + \frac{\partial f}{\partial t}dt$, cette transformation de jauge change ω_A en $\omega_A + d_3f$, c'est-à-dire A en $A + \nabla f$ et $-V$ en $-V + \frac{\partial f}{\partial t}$. On retrouve ainsi les transformations de jauge introduites précédemment.

8.2.4. La synthèse des formules (2) et (4) et la théorie de Hodge.

La synthèse des lois (1) et (3) qui sont indépendantes des sources et des courants ne fait intervenir que la structure différentiable de \mathbb{R}^4 . Lorsqu'on essaye de la prolonger aux deux autres lois, on se rend compte qu'il faut introduire une structure *métrique* bien précise de l'espace temps qui n'est plus de nature purement géométrique mais imposée par la structure profonde, d'origine physique, des lois de l'EM. Cela nécessite l'utilisation de la *dualité de Hodge* développée dans les années 1930-1940 par William Hodge (1903-1975). Expliquons de quoi il s'agit.

Les lois (2) et (4) sont $\text{div}(E) = \frac{\rho}{\varepsilon_0}$ et $\text{rot}(B) = \mu_0 j + \mu_0 \varepsilon_0 \frac{\partial E}{\partial t}$ si l'on traite E et B comme des vecteurs. Mais nous avons vu que E est en fait une 1-forme ω_E et B une 2-forme σ_B . Or, en termes de formes, on sait définir la divergence d'un champ A au moyen de la 2-forme σ_A par $d\sigma_A = \text{div}(A) dV$ et le rotationnel d'un champ X au moyen de la 1-forme ω_X par $d\omega_X = \sigma_{\text{rot } X}$. Ce n'est donc pas ce qu'il nous faut ici. Pour formuler les deux équations en termes de formes différentielles, *il faut introduire la 1-forme ω_B et la 2-forme σ_E* . La loi (2) se formule alors par $d\sigma_E = \frac{1}{\varepsilon_0} \rho dV$ (3-forme) et la loi (4) par $d\omega_B = \mu_0 \sigma_j + \mu_0 \varepsilon_0 \frac{\partial \sigma_E}{\partial t}$ (2-forme).

Une façon de comprendre les liens entre d'un côté ω_E et σ_E et d'un autre côté σ_B et ω_B est d'introduire la dualité de Hodge qui, sur \mathbb{R}^n muni d'un produit scalaire \langle, \rangle , transforme canoniquement les k -formes en $(n - k)$ -formes. Commençons par \mathbb{R}^3 et prenons les coordonnées euclidiennes (x, y, z) par rapport à une base orthonormée (ce qui utilise le produit scalaire). Le produit scalaire \langle, \rangle se prolonge aux k -formes en décidant que Λ^0 a pour base orthonormée la fonction identique à 1, Λ^1 la base $\{dx, dy, dz\}$, Λ^2 la base $\{dy \wedge dz, dz \wedge dx, dx \wedge dy\}$ et Λ^3 la base $dV = dx \wedge dy \wedge dz$. Si f et f' sont deux 0-formes on a donc $\langle f, f' \rangle = ff'$; si $\omega = \omega_X$ et $\omega' = \omega_{X'}$ sont deux 1-formes on a $\langle \omega, \omega' \rangle = \langle X, X' \rangle$; si $\sigma = \sigma_A$ et $\sigma' = \sigma_{A'}$ sont deux 2-formes on a $\langle \sigma, \sigma' \rangle = \langle A, A' \rangle$; si $f dV$ et $f' dV$ sont deux 3-formes on a $\langle f, f' \rangle = ff'$. Soient alors $\alpha \in \Lambda^k$ et $\beta \in \Lambda^{3-k}$. Leur produit vectoriel $\alpha \wedge \beta$ est une 3-forme et s'écrit donc $\alpha \wedge \beta = \lambda dV$ où λ est une fonction. On définit alors la k -forme $*\beta \in \Lambda^k$ duale de

Hodge de β par la formule $\alpha \wedge \beta = \langle \alpha, * \beta \rangle dV$. Cela se généralise facilement à \mathbb{R}^n . On montre que $**\beta = (-1)^{k(n-k)}\beta$ pour la structure euclidienne.

Mais nous avons vu plus haut section 8.1 que si X et A sont deux champs de vecteurs, la 3-forme $\sigma_A \wedge \omega_X = \omega_X \wedge \sigma_A$ est donnée par $\langle A, X \rangle dV$. Cela montre que $*\omega_X = \sigma_X$ et $*\sigma_A = \omega_A$. La loi (2) se formule donc par $d(*\omega_E) = \frac{1}{\varepsilon_0} \rho dV$ (3-forme) et la loi (4) par $d(*\sigma_B) = \mu_0 (*\omega_j) + \mu_0 \varepsilon_0 \frac{\partial(*\omega_E)}{\partial t}$ (2-forme).

C'est de cette façon que l'on peut comprendre les champs de déplacement D et d'excitation magnétique H dans leur relation respective avec E et B (cf. section 8.2.1). Les champs D et H sont les *duaux de Hodge* des champs E et B . En termes de formes différentielles la relation $D = \varepsilon_0 E$ dans le vide ne signifie pas $\omega_D = \varepsilon_0 \omega_E$ (1-formes) mais $\sigma_D = \varepsilon_0 (*\omega_E)$ (2-formes) et, de même, la relation $H = \frac{1}{\mu_0} B$ ne signifie pas $\sigma_H = \frac{1}{\mu_0} \sigma_B$ (2-formes) mais $\omega_H = \frac{1}{\mu_0} (*\sigma_B)$ (1-formes).

On montre alors que l'on peut regrouper ces deux lois en utilisant la 2-forme sur \mathbb{R}^4

$$*\mathfrak{F} = *(\sigma_B + \omega_E \wedge dt) .$$

Cela est assez subtil et il ne serait pas pertinent d'entrer ici vraiment dans les détails. Disons simplement qu'il faut introduire un produit scalaire sur \mathbb{R}^4 et que celui-ci (i) n'est pas euclidien mais *minkowskien* et (ii) introduit un poids de la coordonnée t par rapport aux coordonnées spatiales qui dépend de la vitesse de la lumière c . Comme $c^2 = \frac{1}{\varepsilon_0 \mu_0}$, on peut simplifier l'exposition en choisissant des unités dans lesquelles $\varepsilon_0 = \mu_0 = c = 1$. La métrique de Minkowski est alors $ds^2 = dt^2 - dr^2$ où dr^2 est la métrique euclidienne de \mathbb{R}^3 . Pour pouvoir faire les calculs, on choisit une orientation de \mathbb{R}^4 , par exemple l'ordre (t, x, y, z) et on utilise le fait que si l'on prend un sous-espace de coordonnées (par exemple $\{t, y\}$) et si l'on choisit une orientation (par exemple l'ordre (t, y)) alors le complémentaire (en l'occurrence $\{x, z\}$) doit être orienté de façon à ce que la concaténation des deux sous-bases orientées donne une base directe (en l'occurrence (t, y, x, z) n'est pas directe et il faut orienter le plan $\{x, z\}$ dans l'ordre (z, x)). On obtient ainsi pour Λ^2 et Λ^3 les bases naturelles (celles de Λ^0 , Λ^1 et Λ^4 vont de soi) $dx \wedge dy$, $dt \wedge dz$, $dy \wedge dz$, $dt \wedge dx$, $dz \wedge dx$, $dt \wedge dy$ et $dx \wedge dy \wedge dz$, $-dt \wedge dx \wedge dy$, $-dt \wedge dy \wedge dz$, $-dt \wedge dz \wedge dx$.

Ensuite, il faut connaître les transformées de Hodge des k -formes de base pour k de 0 à 4. On utilise $**\alpha = -(-1)^{k(4-k)}\alpha$ pour $\alpha \in \Lambda^k$, le premier signe $-$ venant de la *signature négative* de la métrique minkowskienne. On trouve pour Λ^1 et Λ^3 $*dt = dx \wedge dy \wedge dz$, $*dx = dt \wedge dy \wedge dz$, $*dy = dt \wedge dz \wedge dx$, $*dz = dt \wedge dx \wedge dy$ et vice-versa car, pour $k = 1, 3$, $-(-1)^{k(4-k)} = +1$ ainsi que pour Λ^2 $*(dt \wedge dx) = -dy \wedge dz$, $*(dt \wedge dy) = -dz \wedge dx$, $*(dt \wedge dz) = -dx \wedge dy$, $*(dx \wedge dy) = dt \wedge dz$, $*(dy \wedge dz) = dt \wedge dx$, $*(dz \wedge dx) = dt \wedge dy$. Pour $k = 2$, $-(-1)^{k(4-k)} = -1$ et l'on vérifie que l'on a bien $**\alpha = -\alpha$. Par exemple $** (dt \wedge dx) = - * (dy \wedge dz) = -dt \wedge dx$.

On obtient alors ; puisque $\mathfrak{F} = \sigma_B + \omega_E \wedge dt$

$$\begin{aligned}\mathfrak{F} &= B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy + E_x dx \wedge dt + E_y dy \wedge dt + E_z dz \wedge dt \\ * \mathfrak{F} &= B_x dt \wedge dx + B_y dt \wedge dy + B_z dt \wedge dz + E_x dy \wedge dz + E_y dz \wedge dx + E_z dx \wedge dy \\ &= \sigma_E - \omega_B \wedge dt .\end{aligned}$$

Puis on calcule $d(*\mathfrak{F})$ comme on a calculé \mathfrak{F} et l'on trouve, puisque $dt \wedge \frac{\partial \sigma_E}{\partial t} = \frac{\partial \sigma_E}{\partial t} \wedge dt$,

$$\begin{aligned}d(*\mathfrak{F}) &= d_4 \sigma_E - d_4 \omega_B \wedge dt = d_3 \sigma_E + dt \wedge \frac{\partial \sigma_E}{\partial t} - d_3 \omega_B \wedge dt \\ &= \rho d_3 V + dt \wedge \frac{\partial \sigma_E}{\partial t} - \left(\sigma_j + \frac{\partial \sigma_E}{\partial t} \right) \wedge dt \\ &= \rho d_3 V - \sigma_j \wedge dt .\end{aligned}$$

En regroupant le courant j et la densité de charges ρ en un *quadri-courant spatio-temporel* J dont la 1-forme associée sur \mathbb{R}^4 est $\omega_J = \rho dt - \omega_j$, on peut décrire de façon unifiée la répartition et le mouvement des charges dans le \mathbb{R}^4 minkowskien. On a alors $*J = \rho d_3 V - \sigma_j \wedge dt$ et l'on obtient ainsi l'équation des sources-courants

$$d(*\mathfrak{F}) = *J .$$

On constate sur cet exemple la remarquable *réduction* de complexité des formalismes que les physiciens ont réussi à obtenir en approfondissant le *type* des entités mathématiques utilisées. Au départ Maxwell partit de 20 équations à 20 inconnues puis, à partir de 1865, réussit le tour de force de les ramener synthétiquement à ses quatre équations fondamentales. Puis l'usage des formes différentielles, le passage à l'espace-temps minkowskien et la dualité de Hodge permirent de ramener celles-ci à deux.

8.2.5. *Calculs symboliques et objectivité physique.*

On note ainsi la subtilité du jeu des opérations algébriques et de leurs signes dans ces calculs qui permettent de réduire considérablement la complexité des équations initiales. Saisissons cette occasion pour faire une remarque de nature philosophique. Quand on formalise des phénomènes physiques, on doit faire l'hypothèse que ces phénomènes "phénoménalisent" relativement à des moyens d'observation (quels que soient ces derniers) une ontologie "en soi" indépendante de son observabilité par des êtres cognitifs dotés de telles ou telles facultés cognitives. Toutefois la physique ne s'occupe pas de l'ontologie "en soi" de la réalité mais seulement de l'*objectivité* des phénomènes. Elle cherche seulement à comprendre les phénomènes objectifs (les "observables") en les formalisant mathématiquement afin de pouvoir en reconstruire toute la diversité de la façon la plus exacte possible au moyen de calculs. Or ces calculs ont une *structure* qui contraint leur générativité. Dans notre cas, nous avons utilisé les règles (d'origine leibnizienne) du calcul différentiel, les règles du calcul des formes multilinéaires alternées, les règles de la dérivation des formes différentielles

qui en découlent, le théorème de Stokes et la dualité de Hodge faisant intervenir la métrique minkowskienne de l'espace-temps.

Ces calculs symboliques très structurés permettent de “calculer” la diversité empirique des observables, mais on ne peut pas pour autant faire l'hypothèse qu'ils font partie de la réalité “en soi” : ce serait introduire une harmonie préétablie entre une ontologie platonicienne d'idéalités mathématiques et l'ontologie “en soi” (inobservable) de la nature. Cette difficulté théorique majeure ouvre la voie à une approche néo-transcendantale de l'objectivité physique. Pour des précisions, le lecteur pourra consulter nos études [419] et [436].

9. Faisceaux et cohomologie

Après cette incursion dans l'utilisation des formes différentielles en physique fondamentale et avant d'aborder un nouveau chapitre consacré à $V1$ comme structure de contact, nous allons revenir sur la notion de cohomologie des formes différentielles pour en présenter une approche alternative fondée sur cette notion de “nerf” d'un recouvrement ouvert que nous avons exposée à propos de la neurotopologie dans les sections 4.4 et 4.5.

La cohomologie telle que nous l'avons présentée correspond à ce que l'on appelle traditionnellement la cohomologie de de Rham⁶⁸ qui mesure l'obstruction empêchant une k -forme fermée d'être exacte. Mais il existe une autre approche consistant à se focaliser sur la façon dont on peut “recoller” des k -formes différentielles ω_i définies sur différents ouverts U_i d'une variété de base M . Il s'agit de la cohomologie dite de Čech.⁶⁹ Il est devenu habituel de la présenter en relation avec la notion de *faisceau*.

9.1. Le concept de faisceau

La notion de faisceau est une notion de topologie algébrique qui a été élaborée techniquement dans les années 1940 par Jean Leray pendant qu'il était prisonnier en Autriche à l'Oflag XVII. Il en fit un cours publié en 1945 et la développa au Collège de France de 1945 à 1953. Le concept a été très vite approfondi par Henri Cartan et Jean-Pierre Serre au début des années 1950, ainsi que par Roger Godement (cf. le classique [215]) et d'autres géomètres jusqu'à devenir omniprésent, d'abord en

68. Georges de Rham (1903-1990), l'un des plus éminents mathématiciens suisses, a été inspiré par les travaux de Poincaré en topologie et a été soutenu par Lebesgue. En 1931 il a soutenu sous la direction d'Élie Cartan une thèse où il démontre comment identifier les groupes de cohomologie des formes différentielles à des invariants topologiques.

69. Eduard Čech (1893-1960) était un mathématicien tchécoslovaque spécialiste de géométrie différentielle et de topologie, également connu pour la notion de “compactification” d'un espace topologique (1937) introduite par Andrey Tychonoff en 1930 et développée également par Marshall Stone (nous rencontrerons plusieurs fois les travaux de Stone à propos de la décomposition spectrale des opérateurs, cf. chapitre 16).

géométrie différentielle puis en géométrie algébrique avec Alexandre Grothendieck qui le généralisa puissamment dans le langage de la théorie des catégories et en fit la base de la théorie des topos.⁷⁰ Dans les livres universitaires il est introduit après de nombreuses autres notions, comme par exemple celle de fibré vectoriel. Il peut donc sembler étrange de l'introduire dans cet ouvrage dès à présent alors que nous ne développerons techniquement la notion de fibration qu'au chapitre suivant.⁷¹

Si nous le faisons c'est que cette notion de faisceau formalise une intuition très profonde qui est celle de recollement global de données locales, intuition déjà largement approfondie dans les sections précédentes de ce chapitre.⁷²

9.2. Phénoménologie du remplissement et du recollement

L'intuition est la suivante. On se donne un espace topologique de base M avec ses ouverts U et l'on considère des données locales d'un certain type définies sur ces ouverts. Elles forment des ensembles $\Gamma(U)$. On se donne donc un *type* d'entités fonctionnellement dépendant de domaines ouverts d'un espace de base. En mathématiques, ces entités auront des types que nous avons déjà rencontrés pour les variétés différentiables M : champs de vecteurs, formes différentielles, etc. C'est cette situation que les géomètres dont nous avons parlé ont voulu formaliser dans le cadre de la topologie algébrique développée depuis Poincaré par des spécialistes comme James Alexander.

Mais cette intuition dépasse de beaucoup les mathématiques. Dans le Vol I, chapitre 4, section 4.10, à propos de la structure en pinwheels de $V1$, nous avons parlé de "cartes" d'orientation, de direction, de dominance oculaire, de phase, de fréquence spatiale. Ce sont autant d'exemples de même type, mais neurologiquement implémentés !

Qui plus est, il existe beaucoup de cas relevant de la perception où des qualités sensibles (couleurs, textures, etc.) "remplissent" des domaines du champ visuel et les psychologues ont développé depuis longtemps une intuition des recollements de remplissements locaux en remplissements plus globaux. Ces cas représentent une sorte "d'archéologie phénoménologique" de la notion de faisceau.

Pour voir à quel point l'intuition de recollement est profonde dans ce domaine perceptif nous allons nous référer aux analyses phénoménologiques de Husserl par

70. Une très bonne introduction historique à la notion de faisceau est celle de Christian Houzel [258].

71. Ceci dit, nous avons déjà introduit la structure de fibration de l'aire $V1$ Vol I, chapitre 4, section 4.3.

72. Pour une étude plus détaillée de la dialectique local/global en mathématiques, le lecteur italoophone pourra se référer à notre item "Locale/Globale" de 1979 dans l'*Enciclopedia Einaudi* [411].

exemple dans les *Ideen* ([261] *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*), *Ding und Raum* [260], ou *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik* [262], des textes parmi les plus prodigieux sur la géométrie de la perception visuelle. Dans plusieurs études ([424], [429], [432], [433], [438]), nous avons commenté l'archéologie phénoménologique de la notion *géométrique* de faisceau et l'application de la théorie des topos à la théorie husserlienne des *jugements* perceptifs. Héritier du schématisme kantien, Husserl est l'un des grands précurseurs de cette corrélation entre intuition géométrique et logique des jugements qu'a formalisé la théorie des topos.

Dans le cas de la perception nous avons vu que le recollement de champs récepteurs locaux y intervient de façon constitutive. Contrairement à ce qui se passe en géométrie différentielle classique, la notion de recollement n'est pas imposée ici par la nécessité de prendre en compte des structures non globalement triviales mais par le câblage de l'architecture du hardware neuronal. Lorsque ces champs récepteurs sont des détecteurs de qualités (couleur, texture, grain, etc.) associés rétinitiquement (au moyen de connexions rétino-géniculo-corticales "verticales") aux positions dans le champ visuel, leurs recollements en domaines plus globaux (au moyen de connexions cortico-corticales "horizontales") conduisent au "remplissement" de ces domaines par ces qualités. Dans le Vol I, nous avons traité des blobs de couleur dans V1 et V2 (chapitre 4, section 4.10.4), des cellules détectant des contrastes spatiaux et chromatiques ("double-opponent cells", chapitre 4, section 4.12.2) et de la constance des couleurs dans l'aire V4 (chapitre 5, section 5.9).

C'est cette notion de remplissement ("Erfüllung") ou de recouvrement ("Überdeckung") de domaines spatiaux par des qualités sensibles que Husserl a thématiquement ("eidétiquement" comme il aimait le dire), évidemment sans référence aux neurosciences et sans chercher à l'axiomatiser. Ce qui l'intéressait était de *conceptualiser* le "rapport de fondation" intuitif des qualités sensibles "secondes" dans la qualité "première" qu'est l'extension spatiale, fondation qu'il pensait comme une donnée originaire de l'expérience perceptive.

Ses analyses phénoménologiques sont les héritières de la psychologie de Franz Brentano et sont étroitement liées à la psychologie de la Gestalt développée par Christian von Ehrenfels et Carl Stumpf (puis plus tard, entre autres, par Max Wertheimer, Wolfgang Köhler, Kurt Koffka). Mais par ailleurs il ne faut pas oublier que Husserl était un mathématicien élève de Leopold Kronecker et de Karl Weierstrass à Berlin, ayant soutenu une thèse *Contributions au calcul des variations* (1883) avant de devenir l'assistant de Weierstrass. Weierstrass était le premier à avoir théorisé le prolongement analytique comme recollement d'éléments analytiques locaux, point de vue approfondi ensuite par Hermann Weyl dans son célèbre ouvrage *Die Idee der Riemannschen Fläche* (1913). Husserl était l'ami de Cantor et de Hilbert.

Pour Husserl (qui connaissait fort bien Riemann, Helmholtz et leur débat que nous avons rappelé dans l'Introduction), le champ visuel est une variété riemannienne bidimensionnelle, d'extension finie et possédant un groupe d'automorphismes transitif (c'est-à-dire pouvant échanger deux points quelconques) :

“Le champ visuel est une multiplicité bidimensionnelle⁷³, congruente avec elle-même, continue, simplement homogène, finie, et s'entend, délimitée” ([260], p. 202).

Le principe de base à thématiser est que

“Tout (...) schème sensible (...), est une (...) figure spatiale, ‘sur laquelle’ ou ‘dans laquelle’ des qualités sensibles s'étendent” ([260], p. 347).

Et c'est la cohésion de l'extension, son “ordre” spatial (on dirait maintenant sa topologie), qui confère leur unité aux qualités : la spatialité est un format universel multisensoriel, une “forme fondamentale”, pour les qualités sensibles.

Husserl thématise alors plusieurs propriétés des remplissements.

1. La restriction des recouvrements à des sous-domaines et la transitivité des restrictions :

“La couleur recouvre l'étendue (...), s'ordonne dans l'étendue. À chaque fragment de l'étendue correspond un fragment de la coloration, et à chaque fragment du fragment à nouveau. Et toutes les colorations partielles sont ordonnées dans l'unité de la coloration globale, dont la forme d'ordre est précisément l'étendue globale” (pp. 400-401).

2. Le passage à la limite des restrictions. Les recouvrements de “petites parties de surface” doivent être poussés à la limite si l'on veut arriver à des remplissements par des qualités “simples”, c'est-à-dire de valeur constante. Par passage à la limite,

“la qualité simple devient ainsi le plein d'un point spatial” (p. 420).

Cette intuition de “plein d'un point spatial” est tout à fait remarquable.

3. Le recollement de différents remplissements locaux du champ visuel. Husserl le conceptualise à partir d'une phénoménologie de la *kinesthésie*. En effet, l'exemple le plus évident de domaine local du champ visuel est le spot attentionnel visuel S correspondant à la vision centrale. L'image remplissant le champ visuel est balayée par ce spot au moyen des contrôles kinesthésiques des mouvements des yeux, de la tête et du corps. Aujourd'hui les saccades oculaires et le contrôle oculo-moteur sont bien compris. À l'époque de Husserl ce n'était pas le cas mais la description phénoménologique reste très pertinente.

Pour expliciter le recollement, Husserl prend l'exemple d'un carré parcouru des yeux en suivant son bord $a \rightarrow b \rightarrow c \rightarrow d \rightarrow a$.

73. *Mannigfaltigkeit* est le terme allemand pour “variété” ou “multiplicité”.

“À chaque position de l’oeil [fixant l’un des sommets] correspond alors une figure pré-empirique [une image] déterminée” (p. 218)

et l’on obtient ainsi une suite d’images $I_a \rightarrow I_b \rightarrow I_c \rightarrow I_d \rightarrow I_a$ remplissant le spot visuel S et se transformant l’une en l’autre sous l’action de contrôles kinesthésiques $K_a \rightarrow K_b \rightarrow K_c \rightarrow K_d \rightarrow K_a$.⁷⁴ Le point essentiel est que les images locales I_α remplissent des domaines S_α différents qui se recoupent (“Überschiebung”), qu’elles “renvoient” les unes aux autres et que l’image globale s’obtient en les recollant au moyen d’un recollement des domaines spatiaux S_α qu’elles remplissent. Mais pour cela il faut pouvoir *identifier* des points correspondants dans les différents S_α et cette identification pose un problème fondamental d’identité. Dans le lexique husserlien la conscience d’identité (ou d’unité) relève de l’*intentionnalité* et Husserl exprime donc les renvois des I_α entre elles en termes intentionnels :

“Dans la série d’images qui s’écoule, l’une renvoie continûment à l’autre et toujours de nouveau à d’autres, (...) à travers elle passent l’intention d’unité et le remplissement d’unité.” (p. 219)

Il s’agit bien d’intersection des différents S_α puisque, dans un addendum, Husserl explique que

“l’identification court de telle façon qu’elle n’amène pas [les champs S_α] à coïncidence, mais au décalage (Überschiebung), une partie seulement [étant] en coïncidence.” (p. 413)

En termes formels, cela signifie qu’il existe une “identité” (en termes actuels un isomorphisme local) de recollement $\varphi_{\alpha\beta} : S_\beta \upharpoonright_{S_\alpha \cap S_\beta} \rightarrow S_\alpha \upharpoonright_{S_\alpha \cap S_\beta}$ qui identifie l’intersection $S_\alpha \cap S_\beta$ vue comme sous-domaine de S_α avec $S_\beta \cap S_\alpha$ vue comme sous-domaine de S_β .⁷⁵ Pour préciser les choses Husserl parle d’“intentionnalité transversible” et de “rayons intentionnels”.

“Les rayons intentionnels qui traversent les images (...) rattachent en une conscience d’unité des points correspondants des images qui se transforment continûment l’une en l’autre.” (p. 230)

Ces rayons intentionnels sont bien conçus par Husserl comme des isomorphismes de recollement puisqu’il parle à leur propos de “correspondance mono-univoque” (p. 238).

“Les points qui se trouvent sur le même rayon intentionnel exposent par leurs contenus un seul et même point d’objet.” (p. 238)

74. Aujourd’hui un aspect de ce problème est celui de l’“eye tracking” (oculométrie).

75. \upharpoonright est le symbole de restriction.

9.3. Formalisation du remplissage et du recollement

Nous allons voir que la description husserlienne des opérations de remplissage et de recollement constitue bien une “archéologie phénoménologique” de la notion formelle de faisceau.

De façon abstraite, un faisceau sur un espace de base M est d’abord caractérisé par les ensembles $\Gamma(U)$ de ses “sections” (les “remplissements” et “recouvrements” ci-dessus) sur les ouverts $U \subset M$.⁷⁶ On dit que les $\Gamma(U)$ forment un *préfaisceau* si les conditions suivantes sont satisfaites.

(PF) Si $s \in \Gamma(U)$ est une section sur U et si $V \subset U$, on peut considérer la restriction $s \upharpoonright_V$ de s à V . Cette restriction est une application $\Gamma(U) \rightarrow \Gamma(V)$. Si $V = U$ alors $s \upharpoonright_V = s$ et si $W \subset V \subset U$ et $s \in \Gamma(U)$ alors $(s \upharpoonright_V) \upharpoonright_W = s \upharpoonright_W$ (transitivité de la restriction). Cela signifie que Γ est un “foncteur contravariant” de la catégorie $\mathcal{O}(M)$ des ouverts de M dans la catégorie **Ens** des ensembles.

Définition. De façon générale une *catégorie* est une classe d’objets X et de transformations (dites morphismes) $f : X \rightarrow Y$ entre objets satisfaisant les conditions suivantes : tout objet admet un morphisme identité et les morphismes peuvent se composer $X \xrightarrow{f} Y \xrightarrow{g} Z$ donnant $X \xrightarrow{g \circ f} Z$. Les objets de $\mathcal{O}(M)$ sont les ouverts de M et les morphismes sont ici les inclusions d’ouverts. Un foncteur d’une catégorie dans une autre est une application des objets et des morphismes de l’une dans les objets et les morphismes de l’autre qui préserve la structure de catégorie, ce qu’expriment exactement les conditions (PF). _____□

Réciproquement, soit Γ un tel foncteur – donc un “préfaisceau” – sur M . Pour être un faisceau Γ doit satisfaire en plus les deux propriétés caractéristiques suivantes :

(F1) Deux sections localement égales sont globalement égales. Si $\mathcal{U} = (U_i)_{i \in I}$ est un recouvrement ouvert⁷⁷ de U et si $s, s' \in \Gamma(U)$ sont deux sections globales, alors, si $s \upharpoonright_{U_i} = s' \upharpoonright_{U_i}$ pour tout $i \in I$, on a l’égalité $s = s'$.

(F2) Une famille de sections s_i qui sont définies sur un recouvrement ouvert $\mathcal{U} = (U_i)_{i \in I}$ d’un ouvert U et égales sur les intersections $U_i \cap U_j$ peut être recollées en une section globale s sur U . Autrement dit si $s_i \in \Gamma(U_i)$ est une famille sur $\mathcal{U} = (U_i)_{i \in I}$ et si les s_i sont telles que i.e. si $s_i \upharpoonright_{U_i \cap U_j} = s_j \upharpoonright_{U_i \cap U_j}$ quand $U_i \cap U_j \neq \emptyset$, alors il existe une section globale $s \in \Gamma(M)$ (unique d’après (F1)) telle que $s \upharpoonright_{U_i} = s_i$ pour tout $i \in I$.

On peut montrer que si les propriétés (F1) et (F2) sont satisfaites alors on peut représenter le foncteur section Γ par une structure fibrée généralisée $\pi : \mathcal{E} \rightarrow M$ (appelée un espace “étalé”) de telle façon que $\Gamma(U)$ devienne l’ensemble des sections

76. Au départ, Jean Leray a défini les faisceaux sur les fermés, puis ensuite, pour des raisons techniques, Henri Cartan et Jean-Pierre Serre les ont définis sur les ouverts, tout en introduisant des fermés formant des “familles de supports”.

77. On suppose des recouvrements ouverts localement finis.

continues de π sur U .⁷⁸ La fibre \mathcal{E}_x de \mathcal{E} en $x \in M$ est la limite “inductive” (qui existe toujours pour les ensembles) :

$$\varinjlim_{V \subset U \in \mathfrak{V}_x} \{\Gamma(U), \Gamma(V \hookrightarrow U) : \Gamma(U) \rightarrow \Gamma(V)\}$$

(où \mathfrak{V}_x est le filtre des voisinages ouverts de x).⁷⁹ \mathcal{E}_x est l’ensemble des “germes” s_x des sections en x .⁸⁰ En tant qu’ensemble, \mathcal{E} est la somme disjointe des fibres \mathcal{E}_x . Si $s \in \Gamma(U)$, elle peut être interprétée comme l’application $x \in U \mapsto s_x \in \mathcal{E}_x$. La topologie de \mathcal{E} est alors définie comme la plus fine rendant toutes ces sections *continues*. On impose donc que les sections $s \in \Gamma(U)$ s’identifient aux sections continues s de π au-dessus de U et la topologie de \mathcal{E} est engendrée par les ouverts $s(U)$. La projection π devient alors par construction un homéomorphisme local. Pour une discussion approfondie de la topologie des espaces étalés le lecteur pourra consulter les notes [365] de Joël Merker.

Nous voyons à quel point le concept de faisceau formalise des propriétés d’essence – “synthétiques a priori” – de l’intuition pure spatiale telles que décrites par l’“eidétique descriptive” husserlienne.

De nombreux faisceaux sont très naturels. Par exemple si M est une variété différentiable, un faisceau privilégié est le faisceau \mathcal{O}_M des applications différentiables $f : U \subset M \rightarrow \mathbb{R}$. Il est dit “faisceau structural” de M et encode de nombreuses propriétés géométriques de M .

Les formes différentielles qui sont au centre de notre réflexion constituent des exemples typiques de faisceaux.

Les faisceaux étant faits pour recoller des données locales en données plus globales, on peut y encoder la liberté ou les contraintes régissant ces recollements. L’absence maximale de solidarité entre le local et le global correspond au cas où, pour tout $V \subset U$, l’application de restrictions $\mathcal{F}(U) \rightarrow \mathcal{F}(V)$ est surjective. Lorsqu’il y a une forte solidarité entre le local et le global ce n’est plus du tout le cas. Par exemple si M est une surface de Riemann compacte, le théorème de Liouville dit que toute fonction globalement holomorphe sur M est nécessairement constante. Les sections globales du faisceau \mathcal{F} des fonctions holomorphes sont donc constantes. Mais sur un petit ouvert V il y aura beaucoup de fonctions holomorphes non constantes $f \in \mathcal{F}(V)$. Par prolongement analytique elles se prolongeront à des ouverts U plus grands mais pas jusqu’à M .

78. Mais π n’est plus nécessairement localement triviale comme le seront les fibrations dont nous parlerons dans les prochains chapitres.

79. Intuitivement, le filtre \mathfrak{V}_x est l’ensemble des voisinages de plus en plus petits de x ordonnés par inclusion. La limite inductive consiste à prendre la limite des $\Gamma(U)$ lorsque U tend vers x dans \mathfrak{V}_x .

80. C’est le “plein d’un point spatial” de Husserl.

Une solidarité faible entre le local et le global est celle des fonctions continues et des fonctions différentiables. Leurs faisceaux ont la propriété que toute section au-dessus d'un fermé se prolonge à M tout entière.

Les faisceaux sur un espace de base M peuvent être munis de différentes structures en imposant que tous les ensembles de sections $\mathcal{F}(U)$ soient munis de structures du type considéré et que toutes les applications de restriction soient des morphismes de ce type de structure. On aura ainsi des faisceaux de groupes (abéliens ou non), des faisceaux d'anneaux, des faisceaux de modules sur un faisceau d'anneaux, des produits tensoriels de faisceaux, etc. Ces structures apparaissent naturellement dès que l'on considère des faisceaux de fonctions $f : U \rightarrow A$ où A est une structure algébrique d'un de ces types car alors $\mathcal{F}(U)$ en hérite par la règle $(f * g)(x) = f(x) * g(x)$ pour les opérations $*$ définissant la structure.

Remarque. Pour construire les espaces étalés des faisceaux munis de structures il faut toutefois vérifier que les limites inductives existent pour ces structures. Elles existent pour les structures algébriques de base : groupes, anneaux, modules. Cela permet de généraliser à de tels faisceaux les structures algébriques traditionnelles de sommes directes \oplus , de produits tensoriels \otimes , etc. \square

Les morphismes de faisceaux $\varphi : \mathcal{F} \rightarrow \mathcal{G}$ sur un même espace de base se définissent naturellement comme des systèmes de $\varphi(U) : \mathcal{F}(U) \rightarrow \mathcal{G}(U)$ compatibles aux restrictions et aux structures dont sont munis \mathcal{F} et \mathcal{G} .⁸¹ Si $f : M \rightarrow N$ est une application continue alors l'image inverse $f^{-1}(W)$ de tout ouvert de N est un ouvert de M et si donc \mathcal{F} est un faisceau sur M on peut naturellement définir son image directe $f_*\mathcal{F}$ comme le faisceau sur N donné par $f_*\mathcal{F}(W) = \mathcal{F}(f^{-1}(W))$. Définir l'image inverse $f^*\mathcal{G}$ d'un faisceau \mathcal{G} sur N est plus compliqué. On doit considérer les limites inductives (si elles existent) $f^*\mathcal{G}(U) = \varinjlim_{W \supset U} \mathcal{G}(W)$.

La notion de sous-faisceau ne pose pas de problème : si \mathcal{F} est un faisceau, \mathcal{G} est un sous-faisceau si les $\mathcal{G}(U)$ sont des sous-structures des $\mathcal{F}(U)$. En revanche la notion de quotient est plus subtile. En effet si \mathcal{G} est un sous-faisceau de \mathcal{F} , les quotients $\mathcal{F}(U)/\mathcal{G}(U)$ forment un préfaisceau qui n'est pas forcément un faisceau.⁸² Pour obtenir un faisceau \mathcal{F}/\mathcal{G} il faut "faisceautiser" les $\mathcal{F}(U)/\mathcal{G}(U)$. Lorsque les limites

81. En plus des grands textes classiques déjà cités, il existe sur le Web de nombreux textes pédagogiques d'introduction aux différentes propriétés des faisceaux. Nous avons cité ceux de Joël Merker (alias François de Marçay, [365]). Citons aussi ceux de Yann Ollivier [395]. Tous deux sont d'anciens normaliens professeurs à l'Université d'Orsay.

82. Un exemple pédagogique standard consiste à prendre le plan complexe épointé $M = \mathbb{C} - \{0\}$ qui n'est pas simplement connexe et à considérer le morphisme $f \mapsto e^{2\pi i f}$ qui va du faisceau des fonctions holomorphes sur M sur les fonctions holomorphes $\neq 0$ sur M . La fonction z est holomorphe et $\neq 0$ sur M mais n'est pourtant pas de la forme $e^{2\pi i f}$ car cela impliquerait que $\log z = f$ soit univoque, ce qui n'est pas le cas, $\log z$ n'étant déterminé qu'à $2\pi i$ près puisque $e^{2\pi i} = 1$. Et pourtant sur tout ouvert U de M ne faisant pas le tour de 0 , il existe une détermination univoque de $\log z$ et $z|_U = e^{2\pi i f_U}$ pour une certaine fonction f_U holomorphe sur U . Ces $z|_U$ ne se recollent donc pas correctement.

inductives existent pour le type de structure considéré, l'espace étalé $E_{\mathcal{P}}$ associé à un préfaisceau \mathcal{P} existe et la faisceautisation de \mathcal{P} consiste à prendre le faisceau $\mathcal{F}_{\mathcal{P}}$ des sections de $E_{\mathcal{P}}$. On montre que les fibres $(\mathcal{F}/\mathcal{G})_x$ sont bien les quotients $\mathcal{F}_x/\mathcal{G}_x$. Les sections $(\mathcal{F}/\mathcal{G})(U)$ de \mathcal{F}/\mathcal{G} sur U s'obtiennent en considérant des s de $\mathcal{F}(U)$ obtenues par recollements de $s_i \in \mathcal{F}(U_i)$ pour un recouvrement $\mathcal{U} = \{U_i\}_{i \in I}$ de U , deux telles sections s, s' étant considérées comme équivalentes si pour tout $x \in U_i \cap U_j$ il existe un voisinage V de x , $V \subset U_i \cap U_j$, tel que $s_i|_V = s'_j|_V \in \mathcal{G}(V)$.

Si $\varphi : \mathcal{F} \rightarrow \mathcal{G}$ est un morphisme de faisceaux sur M avec le minimum de structure algébrique permettant de parler de noyaux et de conoyaux,⁸³ on peut considérer le sous-faisceau $\ker \varphi$ de \mathcal{F} des noyaux $\varphi(U)$ des $\varphi(U) : \mathcal{F}(U) \rightarrow \mathcal{G}(U)$, le sous-faisceau de \mathcal{G} des images $\text{im } \varphi(U)$ et le faisceau quotient $\text{coker } \varphi$ des conoyaux qui faisceautise les $\mathcal{G}(U) / \text{im } \varphi(U)$. On obtient ainsi une *suite exacte* de faisceaux⁸⁴

$$0 \longrightarrow \ker \varphi \longrightarrow \mathcal{F} \xrightarrow{\varphi} \mathcal{G} \longrightarrow \text{coker } \varphi \longrightarrow 0 .$$

En revanche les suites

$$0 \longrightarrow \ker \varphi(U) \longrightarrow \mathcal{F}(U) \xrightarrow{\varphi} \mathcal{G}(U) \longrightarrow \text{coker } \varphi(U) \longrightarrow 0$$

sont exactes à gauche mais pas forcément à droite à cause de ce que nous avons vu à propos des quotients de faisceaux. Le morphisme $\mathcal{G}(U) \rightarrow \text{coker } \varphi(U)$ est localement surjectif mais pas forcément surjectif.

Si \mathcal{F} est un sous-faisceau de \mathcal{G} , on a une suite exacte

$$0 \longrightarrow \mathcal{F} \hookrightarrow \mathcal{G} \longrightarrow \mathcal{G}/\mathcal{F} \longrightarrow 0$$

car $\ker(\mathcal{F} \hookrightarrow \mathcal{G}) = 0$.

9.4. Cohomologies de Čech et de de Rham

Une cohomologie permet de mesurer, dans un certain contexte, l'obstruction au passage du local au global. Il existe de nombreuses théories cohomologiques adaptées à différents contextes et différents types de structure. Elles ont toutes un air de famille mais peuvent être techniquement assez différentes. D'où l'intérêt de pouvoir les comparer entre elles.⁸⁵

83. On supposera par la suite que c'est toujours le cas et que les faisceaux considérés ont au moins une structure de groupe abélien. On suppose alors que $\mathcal{F}(\emptyset) = 0$.

84. L'exactitude signifie (i) que $\ker \varphi \rightarrow \mathcal{F}$ est une injection (un sous-faisceau), ce qu'exprime le $0 \rightarrow \ker \varphi$ initial, et (ii) que $\mathcal{G} \rightarrow \text{coker } \varphi$ est une surjection (un faisceau quotient) ce qu'exprime le $\text{coker } \varphi \rightarrow 0$ final. De façon générale une suite de groupes abéliens $F \xrightarrow{f} G \xrightarrow{g} H$ est exacte en G non seulement si $g \circ f = 0$ (i.e. $\text{im}(f) \subseteq \ker(g)$) mais si en plus on a exactement $\text{im}(f) = \ker(g)$.

85. Le sujet devient vite très technique et fait intervenir le concept délicat de "suite spectrale" introduit également par Jean Leray.

Nous avons déjà dit quelques mots de la cohomologie de de Rham des formes différentielles. La cohomologie de Čech consiste quant à elle à partir des recouvrements ouverts $\mathcal{U} = \{U_i\}_{i \in I}$, de l'espace de base M , à prendre leurs nerfs (la structure combinatoire, le complexe simplicial) $\mathcal{N}(\mathcal{U})$ des intersections non vides et à construire des p -cochaînes à valeurs dans un faisceau de coefficients \mathcal{F} permettant d'y interpréter les opérations de restriction et de recollement. Par exemple une 0-cochaîne s consiste à associer à tout $i \in I$ un élément $s_i \in \mathcal{F}(U_i)$; une 1-cochaîne consiste à associer à tout couple $\{i, j\}$ tels que $U_i \cap U_j \neq 0$ un élément $s_{\{i, j\}} \in \mathcal{F}(U_i \cap U_j)$; une 2-cochaîne consiste à associer à tout triplet $\{i, j, k\}$ tels que $U_i \cap U_j \cap U_k \neq 0$ un élément $s_{\{i, j, k\}} \in \mathcal{F}(U_i \cap U_j \cap U_k)$; etc.⁸⁶ On peut généraliser ces cochaînes à des sommes formelles.⁸⁷

Pour pouvoir définir le cobord ∂ d'une p -cochaîne s^p comme une $(p+1)$ -cochaîne on utilise la structure de groupe additif des sommes formelles. Si s^0 est une 0-cochaîne $\{s_i^0\}$, son cobord est la 1-cochaîne

$$\partial s^0 = s^1 = \{s_{\{i, j\}}^1 = s_j^0 - s_i^0\} ;$$

si s^1 est une 1-cochaîne $\{s_{i, j}^1\}$, son cobord est la 2-cochaîne

$$\partial s^1 = s^2 = \{s_{\{i, j, k\}}^2 = s_{\{j, k\}}^1 - s_{\{i, k\}}^1 + s_{\{i, j\}}^1\} ;$$

et l'on itère la construction.⁸⁸ On voit que si s^1 est un cobord

$$s^1 = \partial s^0 = \{s_{\{i, j\}}^1 = s_j^0 - s_i^0\}$$

alors $\partial s^1 = \partial^2 s^0 = 0$ puisque

$$s_{\{j, k\}}^1 - s_{\{i, k\}}^1 + s_{\{i, j\}}^1 = s_k^0 - s_j^0 - s_k^0 + s_i^0 + s_j^0 - s_i^0 = 0 .$$

Les p -cocycles sont les p -cochaînes s^p de cobord $= 0$ et les p -cobords sont des p -cocycles. On peut alors définir le groupe de cohomologie $\check{H}^p(\mathcal{U}, \mathcal{F})$ comme le quotient du groupe des p -cocycles par le sous-groupe des p -cobords.

On prend ensuite la limite inductive $\check{H}^p(M, \mathcal{F}) = \varinjlim_{\mathcal{U}} \check{H}^p(\mathcal{U}, \mathcal{F})$ de cette cohomologie sur des recouvrements de plus en plus fins. Par exemple, $\check{H}^0(M, \mathcal{F})$ est simplement l'ensemble des 0-cocycles $\{s_i^0\}$ puisqu'il n'y a pas de 0-cobords. La condition $\partial s^0 = 0$ signifie que $s_j^0 = s_i^0$ sur $U_i \cap U_j$, autrement dit que les s_i^0 sont les restrictions aux U_i d'une section globale $s^0 \in \mathcal{F}(M)$. On a donc $\check{H}^0(M, \mathcal{F}) = \mathcal{F}(M)$.

86. Plus généralement, une p -cochaîne s^p à valeurs dans le faisceau de coefficients \mathcal{F} est définie par des $s_{\{i_0, \dots, i_p\}}^p \in \mathcal{F}\left(\bigcap_{i \in \{i_0, \dots, i_p\}} U_i\right)$.

87. Même si les intersections sont symétriques on considère les simplexes $\{i, j\}$, $\{i, j, k\}$, etc. comme ordonnés avec $s_{\{j, i\}} = -s_{\{i, j\}}$, etc.

88. Si s^p est définie par $s_{\{i_0, \dots, i_p\}}^p \in \mathcal{F}\left(\bigcap_{i \in \{i_0, \dots, i_p\}} U_i\right)$, $\partial s_{\{i_0, \dots, i_{p+1}\}}^p = \sum_{k=0}^{k=p+1} (-1)^k s_{\{i_0, \dots, \hat{i}_k, \dots, i_{p+1}\}}^p \in \mathcal{F}\left(\bigcap_{i \in \{i_0, \dots, \hat{i}_k, \dots, i_{p+1}\}} U_i\right)$ où \hat{i}_k signifie que l'indice i_k est absent.

Le calcul des groupes de cohomologie de Čech $\check{H}^p(M, \mathcal{F})$ n'est pas évident à cause des limites inductives. Heureusement dans certains cas ils sont les mêmes que les $\check{H}^p(\mathcal{U}, \mathcal{F})$ pour des recouvrements assez fins. C'est ce qu'affirme par exemple un théorème dû à Jean Leray :

Théorème de Leray. Si $\mathcal{U} = \{U_i\}_{i \in I}$ est un recouvrement ouvert de M tel que pour toute intersection finie non vide V des U_i , $\check{H}^p(V, \mathcal{F}) = 0$ pour $p \geq 1$, alors $\check{H}^p(M, \mathcal{F}) = \check{H}^p(\mathcal{U}, \mathcal{F})$. \diamond

Un des outils les plus essentiels de la cohomologie est que toute suite exacte "courte" de faisceaux $0 \rightarrow \mathcal{F} \rightarrow \mathcal{G} \rightarrow \mathcal{H} (= \mathcal{G}/\mathcal{F}) \rightarrow 0$ induit une suite exacte "longue" des groupes de cohomologie d'ordre croissant :

$$0 \rightarrow \check{H}^0(M, \mathcal{F}) \rightarrow \check{H}^0(M, \mathcal{G}) \rightarrow \check{H}^0(M, \mathcal{H}) \xrightarrow{\partial} \check{H}^1(M, \mathcal{F}) \rightarrow \check{H}^1(M, \mathcal{G}) \rightarrow \dots$$

Cela est lié au fait que si s est une p -cochaîne à valeurs dans \mathcal{H} elle est l'image d'une p -cochaîne \tilde{s} à valeurs dans \mathcal{G} et si son cobord ∂s est nul cela signifie que $\partial \tilde{s}$ est une $(p+1)$ -cochaîne à valeurs dans \mathcal{F} .

Un des résultats cohomologiques les plus importants de la géométrie différentielle est le théorème de de Rham disant que si M est une variété différentielle, sa cohomologie de Čech à coefficients dans le faisceau constant des réels est isomorphe à sa cohomologie de de Rham. Il existe toujours une application $H^p(M, \mathbb{R}) \rightarrow \check{H}^p(M, \mathbb{R})$ de la cohomologie des faisceaux vers la cohomologie de Čech, qui est un isomorphisme pour $n = 0, 1$. Pour les formes différentielles sur M , les k -formes ω^k sur les ouverts U forment un faisceau Ω^k , les k -formes fermées un sous faisceau \mathcal{Z}^k et l'opérateur d de dérivation extérieure définit un morphisme de faisceaux $d_* : \Omega^k \rightarrow \mathcal{Z}^{k+1}$ dont l'image est le sous-faisceau \mathcal{B}^{k+1} des $(k+1)$ -formes exactes. Les k -formes telles que nous les avons définies plus haut "à la Cartan" sont les sections *globales* de ces faisceaux. et la cohomologie de de Rham est donc

$$H_{dR}^k(M) = \check{H}^0(M, \mathcal{Z}^k) / d_*(\check{H}^0(M, \Omega^{k-1})) .$$

Mais comme les formes fermées sont localement exactes d'après le lemme de Poincaré, la suite "courte"

$$0 \rightarrow \mathcal{Z}^k \rightarrow \Omega^k \xrightarrow{d_*} \mathcal{Z}^{k+1} \rightarrow 0$$

est exacte.⁸⁹ On en déduit la suite exacte "longue"

$$\dots \rightarrow \check{H}^p(M, \Omega^k) \xrightarrow{d_*} \check{H}^p(M, \mathcal{Z}^{k+1}) \xrightarrow{\partial} \check{H}^{p+1}(M, \mathcal{Z}^k) \rightarrow \check{H}^{p+1}(M, \Omega^k) \rightarrow \dots$$

On voit ainsi apparaître un lien subtil entre les opérateurs d et ∂ , le premier concernant la cohomologie de de Rham et le second la cohomologie de Čech.

^{89.} \mathcal{Z}^0 est le faisceau des fonctions $f : M \rightarrow \mathbb{R}$ localement constantes (i.e. constantes sur les composantes connexes de M) et $\Omega^0 = C^\infty(M, \mathbb{R})$.

Le lien peut se préciser⁹⁰ en montrant que la “plasticité” de la structure différentiable (absence de solidarité local/global), implique l’existence de “partitions de l’unité” subordonnées aux recouvrements ouverts localement finis $\mathcal{U} = \{U_i\}_{i \in I}$. \mathcal{U} est localement fini si tout point x de M n’appartient qu’à un nombre fini de U_i ⁹¹ et une partition de l’unité est une famille de fonctions $C^\infty f_i$ de supports respectifs dans les U_i et telles que en tout point $x \in M$ la somme $\sum_{i \in I} f_i(x)$ (qui est bien définie puisque le recouvrement est localement fini) soit = 1. La fonction constante 1 qui est complètement délocalisée puisque partout la même est toutefois partout une superposition locale de f_i localisées dans les U_i .

Comme les formes différentielles forment un $\mathcal{C}^\infty(M)$ -module, tout p -cocycle est un p -cobord si $p \geq 1$. Pour le voir, prenons l’exemple $p = 1$. Si s est un 1-cocycle $s_{\{j,k\}}$, on peut considérer la 0-cochaîne $t_{\{m\}} = \sum_{i \in I} f_i s_{\{i,m\}}$ et l’on a

$$\partial t_{\{m,n\}} = t_{\{n\}} - t_{\{m\}} = \sum_{i \in I} f_i (s_{\{i,n\}} - s_{\{i,m\}})$$

et comme s est un cocycle, $\partial s = 0$, i.e. $s_{\{m,n\}} - s_{\{i,n\}} + s_{\{i,m\}} = 0$, soit $\partial t_{\{m,n\}} = (\sum_{i \in I} f_i) (s_{\{i,n\}} - s_{\{i,m\}}) = (\sum_{i \in I} f_i) s_{\{m,n\}} = s_{\{m,n\}}$. On a donc $s = \partial t$. Ce calcul simple se généralise trivialement à $p > 1$. On a donc $\check{H}^p(M, \Omega^k) = 0$ si $p \geq 1$. Pour $p = 0$, $\check{H}^0(M, \Omega^k)$ est tout simplement le groupe (en fait le $\mathcal{C}^\infty(M)$ -module) des k -formes. On voit ainsi qu’à cause du lemme de Poincaré et de l’existence de partitions de l’unité, la cohomologie de Čech est triviale pour le faisceau Ω^k des germes de k -formes différentielles.

Revenons alors pour $p \geq 1$ à la suite exacte “longue”

$$\dots \longrightarrow \check{H}^p(M, \Omega^k) \xrightarrow{d_*} \check{H}^p(M, \mathcal{Z}^{k+1}) \xrightarrow{\partial} \check{H}^{p+1}(M, \mathcal{Z}^k) \longrightarrow \check{H}^{p+1}(M, \Omega^k) \longrightarrow \dots$$

Comme $\check{H}^p(M, \Omega^k) = \check{H}^{p+1}(M, \Omega^k) = 0$, on a la suite exacte

$$0 \longrightarrow \check{H}^p(M, \mathcal{Z}^{k+1}) \xrightarrow{\partial} \check{H}^{p+1}(M, \mathcal{Z}^k) \longrightarrow 0$$

qui signifie que ∂ est un isomorphisme de $\check{H}^p(M, \mathcal{Z}^{k+1})$ avec $\check{H}^{p+1}(M, \mathcal{Z}^k)$. On en déduit en itérant que

$$\check{H}^1(M, \mathcal{Z}^{k-1}) \simeq \check{H}^2(M, \mathcal{Z}^{k-2}) \simeq \dots \simeq \check{H}^k(M, \mathcal{Z}^0).$$

Mais nous avons noté plus haut que les 0-cocycles sont les fonctions localement constantes, autrement dit les sections du faisceau constant \mathbb{R} , et que donc $\check{H}^k(M, \mathcal{Z}^0)$

90. Il s’agit de résultats classiques que l’on retrouve dans tous les cours d’introduction à la cohomologie. Nous suivons ici des notes de Yann Ollivier.

91. Cette notion est surtout utile lorsque l’espace M a la propriété que tout recouvrement ouvert admet un sous-recouvrement localement fini. On dit alors que M est *paracompact*, notion introduite en 1944 par Jean Dieudonné.

$= \check{H}^k(M, \mathbb{R})$. Si l'on revient alors au début de la suite "longue" de la cohomologie de Čech, on trouve

$$\check{H}^0(M, \Omega^{k-1}) \xrightarrow{d_*} \check{H}^0(M, \mathcal{Z}^k) \xrightarrow{\partial} \check{H}^1(M, \mathcal{Z}^{k-1}) \longrightarrow 0,$$

soit $\check{H}^1(M, \mathcal{Z}^{k-1}) = \check{H}^0(M, \mathcal{Z}^k) / d_*(\check{H}^0(M, \Omega^{k-1}))$. Mais par définition (cf. plus haut)

$$\check{H}^0(M, \mathcal{Z}^k) / d_*(\check{H}^0(M, \Omega^{k-1})) = H_{dR}^k(M)$$

et l'on obtient ainsi une version du classique théorème de de Rham :

Théorème de de Rham.

$$H_{dR}^k(M) \simeq \check{H}^1(M, \mathcal{Z}^{k-1}) \simeq \check{H}^k(M, \mathcal{Z}^0) \simeq \check{H}^k(M, \mathbb{R}).^{92} \quad \diamond$$

Ce résultat est remarquable car il montre que la cohomologie des formes différentielles, qui sont des entités exprimant l'essence du calcul différentiel symbolique sur M , "incarne" la cohomologie de Čech qui exprime l'essence des nerfs des recouvrements ouverts de M .

92. Ces deux cohomologies sont également isomorphes à la cohomologie "singulière" construite directement à partir des complexes simpliciaux des nerfs des recouvrements.

CHAPITRE 3

V1 comme structure de contact

Nous allons maintenant commencer la description de notre périple modélisateur piloté par des idées mises au point au tournant des années 1980-1990.¹ Il ressemble un peu à une circumnavigation dans un monde de terres et de mers neuromathématiques.

1. Structures de contact et vision

Comme nous l'avons vu plusieurs fois dans le Vol I, le point de départ de la neurogéométrie est que les neurones simples de $V1$ détectent (à une certaine échelle) ce que l'on appelle en géométrie différentielle depuis le milieu du XIX^e siècle des *éléments de contact*. Si le cortex visuel fait de la géométrie différentielle il le fait donc en implémentant grâce à ces détecteurs et à leur architecture fonctionnelle un calcul intégral-différentiel fondé sur une connectivité entre éléments de contact. Un exemple fondamental en est la *structure de contact*.

Jusqu'ici, pratiquement aucun spécialiste de la vision n'a vraiment analysé les relations frappantes qui existent entre les hypercolonnes d'orientation du cortex visuel primaire et les notions géométriques de fibration, d'espace de jets, ou de structure de contact. Toutefois, il y a heureusement des exceptions. Nous avons déjà cité plusieurs fois dans le Vol I les travaux de Jan Koenderink. Citons également ceux de William Hoffman, l'un des pionniers de l'application de la géométrie différentielle et de la théorie des groupes de Lie à la vision que nous avons évoqué dans la Préface du Vol I. En particulier dans son important article de 1989 "The visual cortex is a contact bundle" [254], il a formulé explicitement l'idée que les contours sont des relèvements de discontinuités du stimulus rétinien dans un fibré de contact rétinotopique implémenté par les hypercolonnes corticales. Selon lui, ce que les neurobiologistes appellent des "projections" rétine \rightarrow cortex sont en fait de tels "path liftings" :

1. En plus des travaux détaillés que nous citerons au fur et à mesure, une introduction contextualisée à nos modèles est le cours *Geometry of Neurosciences* donné en 2017 à Caltech par Matilde Marcolli et Doris Tsao [350], suivi en 2020 par le Programme *New Geometric Methods in Neuroscience* du Fields Institute de Toronto.

“A path on one manifold [the retina] is ‘lifted’ via a fibering to another manifold [the cortex] in a coherent fashion.” (p. 145)

Dans un autre travail Hoffman [253] a aussi évoqué le concept de fibration et conclu :

“Fibrations (...) are certainly present and operative in the posterior perceptual system if one takes account of the presence of ‘orientation’ micro-response fields and the columnar arrangement of cortex.” (p. 645)

Il a également introduit l’idée que les champs récepteurs rétiniens fournissent des cartes locales, les applications de changement de cartes entre cartes se recouvrant étant implémentées dans la connectivité fine de la rétine.

2. Relevées legendriennes, forme de contact et espace des 1-jets

Nous allons maintenant approfondir le formalisme et la philosophie des fibrations (Vol I, chapitre 4, section 4.3.) qui donnent un statut mathématique rigoureux à la notion d’“engrafted variables” de David Hubel (Vol 1, section 4.3.2.). Nous considérons des fibrations $\pi : V \rightarrow M$ qui paramétrisent des “fibres” de descripteurs au moyen des positions a dans un “espace de base” M et nous voulons *lifter le calcul différentiel* de M dans V .

2.1. L’espace des 1-jets

Nous travaillons, pour commencer, dans la fibration $\pi : V = M \times \mathbb{P}^1 \rightarrow M$ et en fait $\pi : V = \mathbb{R}^2 \times \mathbb{P}^1 \rightarrow \mathbb{R}^2$ si nous posons $M = \mathbb{R}^2$. M est ici l’idéalisée de la couche corticale et les descripteurs des fibres \mathbb{P}^1 sont les orientations. Nous ne tenons pas compte de l’application rétinotopique $\chi : R \rightarrow M$ menant du plan rétinien R à M .² Nous allons expliciter la *structure de contact* de V que nous avons déjà plusieurs fois évoquée dans le Vol I.

Dans le premier modèle que nous proposons de l’aire primaire $V1$, celle-ci est considérée comme une réalisation neuronale de l’espace des 1-jets des courbes du plan. Rappelons (Vol I, 4.3.5.) que si, dans un certain système de coordonnées locales (x, y) de M , une courbe régulière γ est le graphe $\{x, f(x)\}$ d’une fonction f à valeurs réelles sur \mathbb{R} , le 1-jet de f en x , noté $j^1 f(x)$, est le triplet (x, y, p) , avec $y = f(x)$ (la valeur de f en x) et $p = f'(x)$ (la valeur de la dérivée de f en x , c’est-à-dire la pente de la tangente au graphe de f au point $a = (x, f(x))$ de M). Un 1-jet est donc tout simplement un élément de contact $c = (a, p)$.

De façon générale, comme nous l’avons vu également dans cette section du Vol I, si M est une variété de dimension n , on peut considérer en chaque point a de

2. Nous avons vu Vol I, 4.2. (à propos de la structure hypercolumnaire de $V1$) que χ est une application conforme de type logarithme complexe.

M , l'ensemble des hyperplans (sous-espaces vectoriels de codimension 1) de l'espace tangent T_aM . Notons-le C_aM . C_aM est isomorphe à l'espace projectif \mathbb{P}^{n-1} . L'espace total recollant ces fibres s'appelle le *fibré de contact* de M . Notons-le CM . Dans notre cas, $M = \mathbb{R}^2$ est 2-dimensionnelle, les hyperplans sont les orientations (droites) et $CM = M \times \mathbb{P}^1$ est un fibré globalement trivial.

Pour $M = \mathbb{R}^2$, CM est presque l'espace des 1-jets $J^1(\mathbb{R}, \mathbb{R})$ associé au choix des coordonnées (x, y) et noté par commodité J^1M . En effet soient (ξ, η) les coordonnées naturelles des vecteurs du plan tangent T_aM associées à la base $\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$. Sur un ouvert ne contenant pas la droite "verticale" $\xi = 0$, une coordonnée locale de C_aM est $p = \frac{\eta}{\xi}$ et, au voisinage de $\xi = 0$, on peut prendre la coordonnée $p = \frac{\xi}{\eta}$. Un élément c de CM est donc repéré par les coordonnées $(x, y, p) = (a, p)$. On vérifie aisément que les changements de cartes associés à ces coordonnées naturelles sur CM sont des difféomorphismes. Ainsi, $CM = C\mathbb{R}^2$ est une variété différentiable de dimension 3 isomorphe à $V = M \times \mathbb{P}^1$.

Se donner une section s de ce fibré $\pi : E = CM \rightarrow M$ consiste alors à associer à chaque point a de la base M un élément $s(a)$ de la fibre C_aM au-dessus de a , c'est-à-dire une orientation. Les sections sont donc des champs de couples $(a, p) = (\text{position}, \text{orientation})$. Un cas particulier fondamental de section restreinte à une courbe différentiable γ tracée sur la variété de base M est obtenu en choisissant, au-dessus de chaque point de γ , l'orientation de la tangente à γ en ce point.

La différence entre CM et J^1M est que J^1M a pour fibre non pas \mathbb{P}^1 mais le \mathbb{R} des valeurs de $\tan(\theta)$, où l'angle θ est défini modulo π , est mesuré à partir du choix de l'axe des x et doit rester $\neq \frac{\pi}{2}$. Pour obtenir la fibre \mathbb{P}^1 de CM , il faut compactifier \mathbb{R} en ajoutant le point à l'infini. CM est le compactifié à l'infini (le compactifié projectif) de J^1M et sa fibre \mathbb{P}^1 correspond à la complétion $\mathbb{R} \cup \{\infty\}$ ³ de celle de J^1M . Dans le langage de la géométrie algébrique, J^1M est l'ouvert affine de CM complémentaire de la section à l'infini, le choix de cette section correspondant au choix d'un axe des x dans le plan M . Si l'on identifie \mathbb{P}^1 au quotient modulo π de \mathbb{S}^1 muni de la coordonnée angulaire θ , alors J^1M se déduit de $M \times \mathbb{S}^1$ par la projection stéréographique $\mathbb{S}^1 \rightarrow \mathbb{R}$, $\theta \rightarrow \tan(\theta)$.

Remarque. Dans la suite, nous confondrons éventuellement J^1M et CM en admettant que $\tan(\theta)$ peut prendre la valeur ∞ . Mais, après un très long chemin, nous reviendrons de façon approfondie sur leur différence dans la section 3.6 du chapitre 15 où nous esquisserons l'extraordinaire richesse des complétions projectives des espaces de jets d'ordre supérieur révélée par Richard Montgomery et Michail Zhitomirskii. □

Remarque. Pour une courbe régulière C du plan, le choix de coordonnées (x, y) est *arbitraire* et il existe en tout point des coordonnées locales telles que C soit

3. Le point ∞ identifie les limites $\pm\infty$ de \mathbb{R} . C'est une "coincidentia oppositorum".

un graphe de $y = f(x)$. Nous reviendrons également sur cette arbitrarité dans la section 3.6 du chapitre 15. \square

Revenons pour les préciser à nos remarques du Vol I, 4.3.5. Le jet d'ordre 1, noté $j^1 f(x)$, d'une fonction dérivable $f : \mathbb{R} \rightarrow \mathbb{R}$ au point x de \mathbb{R} est caractérisé par la donnée de l'abscisse du point x considéré, de la valeur de la fonction f en ce point, $y = f(x)$, et de la valeur de sa dérivée $p = f'(x)$. D'après le théorème des fonctions implicites, toute courbe γ différentiable est localement le graphe d'une telle fonction (sauf aux points singuliers à tangente verticale) et, en un point donné a , le 1-jet associé ne dépend pas du choix des coordonnées puisque la définition de la tangente est intrinsèque. On note ce 1-jet $j^1 \gamma(a)$.

Insistons encore une fois sur le fait que les espaces de jets sont fondamentaux car ils ramènent des calculs locaux (infinitésimaux) à des calculs *ponctuels*. La contrepartie de ce remarquable bénéfice est l'augmentation de la dimension (c'est-à-dire du nombre de variables indépendantes) : au lieu de considérer le plan de dimension 2 muni des coordonnées (x, y) et de calculer $y' = \frac{dy}{dx}$ (ce qui exige de connaître non seulement la valeur $y = f(x)$ de f en x mais aussi les valeurs de f dans un voisinage de x), on se place dans l'espace à trois dimensions (x, y, p) et l'on impose la *contrainte* $y' = p$. Autrement dit, on travaille sur la géométrie différentielle des courbes du plan mais en introduisant une dimension supplémentaire qui est en quelque sorte la "dimension *cachée*" des tangentes.

Cette idée très profonde remonte à Euler et à Lagrange, et surtout Hamilton qui, en introduisant les moments conjugués p_i des variables de position q_i d'un système mécanique comme des variables indépendantes, est passé de la formulation lagrangienne à la formulation hamiltonienne de la Mécanique. Nous aurons l'occasion d'y revenir plusieurs fois pour en souligner l'importance cardinale.

Si γ est une courbe (lisse) de M d'équations $x(s)$ et $y(s)$, le 1-jet $j^1 \gamma(a(s))$ de γ en $a(s) = (x(s), y(s))$ est l'élément de contact $(a(s), p(s))$ où $p(s) = \frac{y'(s)}{x'(s)}$ est la pente de la tangente à γ en $a(s)$. L'image de $j^1 \gamma$ s'appelle la *relevée legendrienne* de γ . Les relevées legendriennes des courbes γ de M représentent ces courbes non plus comme ensembles de points dans M mais, dualement, comme enveloppes de leur tangentes. C'est ce que l'on appelle la *dualité projective*. Il est remarquable que l'évolution biologique ait créé deux structures neurophysiologiques, la rétine et V1, afin d'implémenter la dualité projective pour les contours.

Remarque sur les notations. Dans ce qui suit, nous traiterons de modèles particuliers explicites et en même temps de situations plus générales. Pour éviter des inflations de notations nous adopterons les règles suivantes. Nous noterons $\pi : M \times P \rightarrow M$ les fibrations où l'on ne spécifie pas précisément la base M et la fibre P . Nous noterons \mathbb{V}_J à la fois $\pi_J : \mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R}^1 \rightarrow \mathbb{R}^2$ et $\pi_J : \mathbb{V}_J = \mathbb{R}^2 \times \mathbb{P}^1 \rightarrow \mathbb{R}^2$ obtenue en ajoutant le point à l'infini $p = \infty$. En toute rigueur il faudrait utiliser deux notations différentes. Mais ce qui est caractéristique de \mathbb{V}_J est que l'orientation

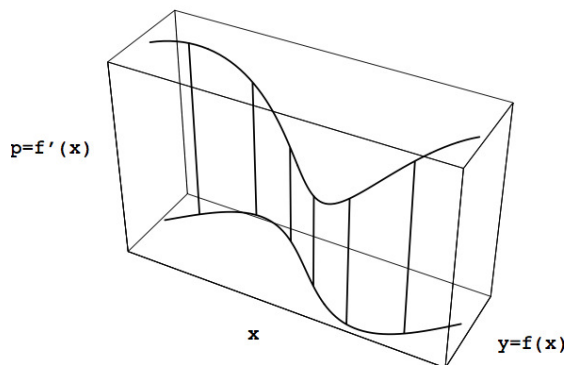


FIGURE 1. La relevée legendrienne d'une courbe différentiable γ , $y = f(x)$, de \mathbb{R}^2 dans l'espace de la fibration $\mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R}$. Au-dessus de chaque point $(x, y = f(x))$ de γ on sélectionne la direction tangente $p = f'(x)$.

d'angle θ y soit paramétrée par $\tan(\theta)$. Nous noterons $\mathbb{V}_S = \mathbb{R}^2 \times S^1 \rightarrow \mathbb{R}^2$ le modèle où les orientations sont paramétrées par l'angle θ ($V1$ correspondant au quotient modulo π).

2.2. Relevées legendriennes et condition d'intégrabilité

Dans ce qui suit, nous continuons à supposer $M = \mathbb{R}^2$ et nous travaillons dans l'espace des 1-jets \mathbb{V}_J .

Rappelons ce que nous avons esquissé Vol I, 4.3.6. et 4.3.7. Soit γ une courbe (différentiable⁴) dans \mathbb{R}^2 . Son 1-jet $j^1\gamma : \gamma \subset \mathbb{R}^2 \rightarrow J^1\mathbb{R}^2$ qui associe à tout point a de γ le 1-jet $j^1\gamma(a)$ de γ en ce point (c'est-à-dire l'élément de contact (a, p_a) où p_a est la tangente à γ en a), a pour image une courbe gauche Γ dans $\mathbb{V}_J = J^1\mathbb{R}^2$ qui est sa relevée legendrienne (l'enveloppe de ses tangentes). La figure 1 en montre un exemple.

Si $a(s)$ est une paramétrisation de γ , on a $p_{a(s)} = \frac{y'(s)}{x'(s)}$ et donc

$$\Gamma = (a(s), p(s)) = (a(s), p_{a(s)}(s)) .$$

Dans les coordonnées x et y et relativement à l'équation $y = f(x)$ de γ , l'équation de Γ est $(x, y, p) = (x, y = f, y' = f')$.

Supposons maintenant que nous n'ayons aucun accès à ce qui se passe dans l'espace de base \mathbb{R}^2 et que nous essayions de le reconstruire à partir de ce qui se passe dans l'espace total $\mathbb{V}_J \simeq J^1\mathbb{R}^2 \simeq \mathbb{R}^3$. À chaque courbe γ dans \mathbb{R}^2 est

4. Sauf mention du contraire, nous supposons que les courbes sont différentiables.

associée une courbe Γ dans \mathbb{V}_J . *Mais la réciproque est fausse.* Il est donc crucial de pouvoir caractériser, parmi les courbes gauches Γ dans \mathbb{V}_J , celles qui sont les relevées legendriennes des courbes γ dans \mathbb{R}^2 . Nous savons déjà comment le faire et c'est cette condition que nous voudrions maintenant approfondir. *Neurophysiologiquement*, γ modélise un contour dans le champ visuel projeté rétinotopiquement sur la couche 4 de V1, Γ modélise la chaîne des neurones simples de V1 activés par γ et la caractérisation visée est celle de l'architecture fonctionnelle de V1.

Soit $\Gamma = v(s) = (a(s), p(s))$ une courbe (paramétrée) dans \mathbb{V}_J . La projection $a(s)$ de Γ est une courbe γ dans \mathbb{R}^2 et Γ est le relèvement de γ si et seulement si $p(s) = p_{a(s)}$. De façon équivalente, si Γ est localement définie par des équations $y = f(x)$, $p = g(x)$, il existe une courbe γ dans \mathbb{R}^2 telle que $\Gamma = j^1\gamma$ si et seulement si $g(x) = f'(x)$, c'est-à-dire si et seulement si $p = y'$.

En géométrie différentielle, cette condition s'appelle une *condition d'intégrabilité*. Nous en avons explicité la signification dans Vol I, 5.4.1. Elle dit que pour être une relevée legendrienne dans \mathbb{V}_J , Γ doit être une *courbe intégrale de la structure de contact* de la fibration π au sens suivant. Soit $\mathbf{t} = (a, p; \alpha, \pi) = (x, y, p; \xi, \eta, \pi)$ un vecteur tangent à \mathbb{V}_J au point $v = (a, p) = (x, y, p)$. Si \mathbf{t} est tangent à la courbe d'équation $y = y(x)$ et $p = p(x)$, on a $\mathbf{t} = (x, y, p; 1, y', p')$. Si $p = y'$, c'est-à-dire si la condition d'intégrabilité est satisfaite, on a donc $\mathbf{t} = (x, y, p; 1, p, p')$. Étant donnée cette forme très spéciale, \mathbf{t} appartient au *noyau de la forme différentielle* $\omega_J = dy - pdx$ sur $\mathbb{V}_J = J^1\mathbb{R}^2$. En effet, soit $\varpi = \varpi_1 dx + \varpi_2 dy + \varpi_3 dp$ une 1-forme. Comme nous l'avons vu longuement au chapitre 2 à partir de la section 6.3.4, les 1-formes de base dx , dy , dp associées aux coordonnées locales (x, y, p) agissent très simplement sur les vecteurs tangents $\mathbf{t} = (\xi, \eta, \pi)$: dx sélectionne la première composante $\mathbf{t}_1 = \xi$, dy la seconde $\mathbf{t}_2 = \eta$ et dp la troisième $\mathbf{t}_3 = \pi$. On a donc $\varpi(\mathbf{t}) = \sum \varpi_i \mathbf{t}_i$. Dans notre cas $\omega_J = dy - pdx = -pdx + 1 \times dy + 0 \times dp$ appliquée à $\mathbf{t} = (1, p, p')$ donne $\omega_J(\mathbf{t}) = -p \times 1 + 1 \times p + 0 \times p' = -p + p = 0$.

C'est ici qu'intervient *une remarque simple mais cruciale*. Le noyau d'une forme différentielle (non nulle) sur un espace 3-dimensionnel est un *plan* et le vecteur tangent \mathbf{t} appartient par conséquent au plan $K_v = \ker(\omega_J)$ tangent à $\mathbb{V}_J = J^1\mathbb{R}^2$ en v (i.e. $K_v \subset T_v\mathbb{V}_J$).⁵

Ainsi, les tangentes aux courbes de $\mathbb{V}_J = J^1\mathbb{R}^2$ qui sont de la forme $j^1\gamma$ appartiennent au champ de plans $\mathcal{K} : v \rightarrow K_v$. Les plans K_v sont appelés *plans de contact*. Leur champ est appelé *structure de contact* sur $J^1\mathbb{R}^2$ ou $C\mathbb{R}^2$ et la 1-forme différentielle ω_J dont il est le noyau est appelée *forme de contact*. Les plans de contact K_v sont d'équation $-p\xi + \eta = 0$. Une base naturelle en est donnée par les deux vecteurs tangents de $T_v\mathbb{V}_J$

5. Dans ces modèles de fibrés de contact $V = CM$, il ne faut pas confondre les K_v avec les C_a . C_a est la fibre au-dessus de $a \in M$ et est un sous-espace de CM , alors que K_v est le plan de contact de CM en $v \in CM$ et est un sous-espace de $T_v(CM)$.

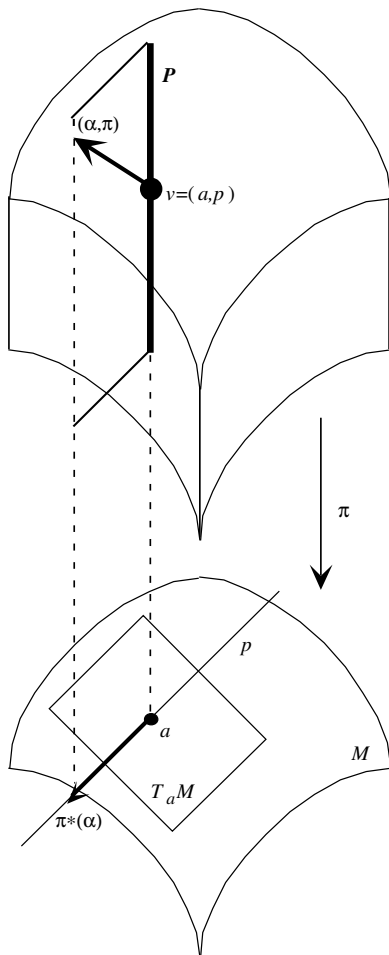


FIGURE 2. La structure de contact sur le fibré de contact CM d'une variété de base M . Soient $a \in M$ et p un hyperplan du plan tangent $T_a M$. p correspond à un point de la fibre $C_a M$ de CM au-dessus de a . Soit $\mathfrak{t} = (\alpha, \pi)$ un vecteur tangent à CM en $v = (a, p)$. $\mathfrak{t} = (\alpha, \pi)$ appartient au plan de contact $C_v M$ de CM en v si et seulement si la projection $\pi_*(\alpha)$ de la composante "horizontale" α de \mathfrak{t} est portée par p .

$$\mathfrak{t}_1 = \frac{\partial}{\partial x} + p \frac{\partial}{\partial y} = (\xi = 1, \eta = p, \pi = 0) \quad , \quad \mathfrak{t}_2 = \frac{\partial}{\partial p} = (\xi = 0, \eta = 0, \pi = 1) \quad .$$

Puisque les courbes de la forme $j^1 \gamma$ sont tangentes en tout point à ce champ de plans, on dit que ce sont des *courbes intégrales* de la structure de contact (cf. figure 2).

Réciproquement, considérons une courbe Γ de $J^1\mathbb{R}^2$ et paramétrons-la sous la forme

$$\Gamma = \{(x, y, p) | y = f(x), p = g(x)\}.$$

Le long de Γ , on a alors $\omega_J = (f' - g)dx$ (car $dy = f'dx$ et $pdx = gdx$). Supposons que cette courbe Γ soit une intégrale de la structure de contact, autrement dit que $\omega_J = 0$ sur Γ . Alors $f' = g$ et Γ est donc bien de la forme $j^1\gamma$.

La structure de contact \mathcal{K} est ainsi la structure géométrique qui permet de caractériser les relevées legendriennes de courbes planes parmi les autres courbes gauches de $\mathbb{V}_J = J^1\mathbb{R}^2$: pour être une relevée legendrienne dans \mathbb{V}_J , une courbe gauche Γ de \mathbb{V}_J doit être une *courbe intégrale de la structure de contact*.

Pour rendre cela intuitif, reprenons la figure 3 de Vol I, 5.2.6. La sous-figure (*haut-gauche*) montre une courbe dans $J^1\mathbb{R}^2$ qui est une relevée legendrienne. La sous-figure 3 (*haut-droit*) correspond au cas où on a ajouté à $p = f'(x)$ une constante p_0 (un angle de $\pi/2$, les orientations p devenant par conséquent orthogonales à la courbe γ décrite par les positions a). Les sous-figures 3 (*bas*) montrent deux autres exemples de courbes non intégrables dans \mathbb{V}_J , la première (à gauche) parce que p est constante alors que f' ne l'est pas, la seconde (à droite) parce que p varie plus rapidement que f' . Dans chacun des trois derniers cas, le pop-out perceptif de la courbe γ décrite par les positions a est impossible parce que la condition d'intégrabilité n'est pas satisfaite.

2.3. La forme ω et l'équation $\omega = 0$

Il faut comprendre toute la subtilité théorique (conceptuelle et pas seulement formelle) de la définition de la structure de contact \mathcal{K} comme noyau de la forme de contact ω . Elle est cruciale pour notre approche neurogéométrique. Il s'agit, revenons-y encore, de comprendre comment nous pouvons percevoir une courbe plane régulière γ comme *un objet global doté de cohésion*. Au niveau sensoriel rétinien γ est discrétisée en une chaîne de points comme dans un tableau "pointilliste" à la Georges Seurat ou "divisionniste" à la Paul Signac. Pour qu'il y ait cohésion, il faut interpoler entre les points par les petits segments qui les joignent et donc à la limite, comme nous venons de le voir, passer par dualité projective de γ comme ensemble de points à γ comme enveloppe de tangentes. Sur le plan neurophysiologique, nous avons vu dans le Vol 1 (chapitre 5, sections 5.1. et 5.2.) que la cohésion de γ comme entité globale est

- (i) assurée par les connexions cortico-corticales "horizontales" et observable au niveau psychophysique par le champ d'association de Field, Hayes et Hess,
- (ii) réalisée en tant que pop-out et saillance perceptive par la synchronisation, le "liage", le "binding", des activités des neurones connectés par des connexions horizontales.

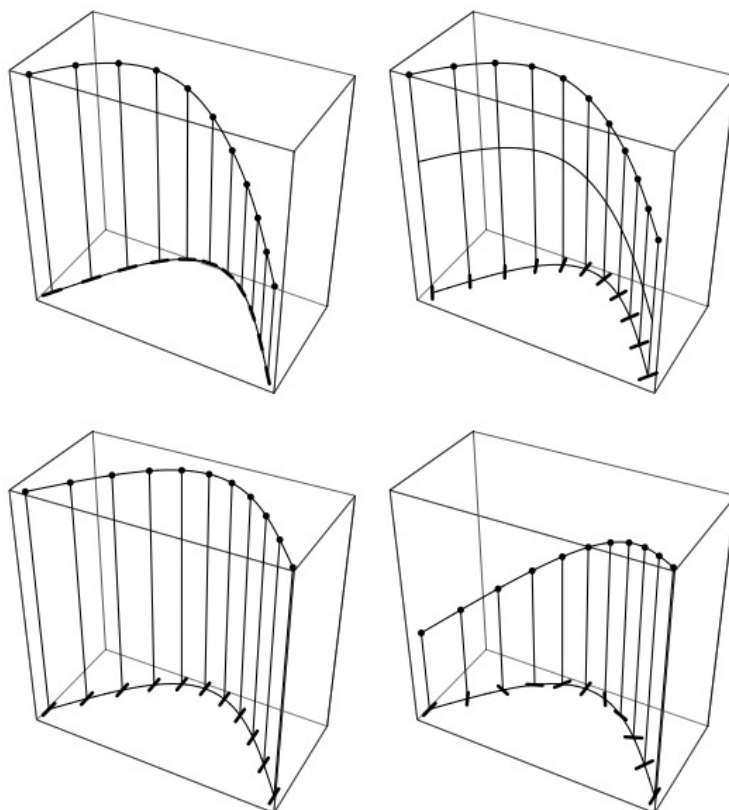


FIGURE 3. La condition d'intégrabilité. (*Haut-gauche*) la condition est satisfaite. (*Autres*) la condition n'est pas satisfaite. (*Haut-droit*) On ajoute un angle constant à la tangente ($p = f'(x) + p_0$). (*Bas-gauche*) p est constant alors que f' ne l'est pas. (*Bas-droit*) p tourne plus vite que f' .

Il y a donc une *différence fondamentale* entre γ comme “donnée” dans la sensation (input rétinien) et γ comme “construite” par la perception. On sait que cette différence est centrale depuis toujours dans toutes les philosophies de la perception. Épistémologiquement, elle renvoie à la différence entre l’“observation” de données et l’“explication” des observations. La neurogéométrie lui donne une réponse nouvelle et particulière qui renvoie en amont aux sources du calcul différentiel. L'idée de base est que la “construction” perceptive de γ peut être modélisée, comme en physique, en faisant de γ une solution d'une équation différentielle. Dans cette perspective, c'est le fait que le cortex visuel construit γ comme une *intégrale* ajustée à l'input qui explique la saillance perceptive.

Le problème est alors de trouver une telle équation différentielle. Mais cela semble être impossible puisque, par définition, elle doit être satisfaite par *toutes* les courbes planes γ assez régulières. C'est là qu'intervient l'idée clé d'augmenter la dimension

en introduisant une nouvelle variable indépendante p et de se situer dans le \mathbb{R}^3 des (x, y, p) . On peut alors chercher une équation différentielle (E) dans $\mathbb{R}^3_{(x,y,p)}$ telle que

- (i) toutes les γ assez régulières soient des *projections* sur $\mathbb{R}^2_{(x,y)}$ de courbes intégrales Γ de (E) dans $\mathbb{R}^3_{(x,y,p)}$,
- (ii) chaque γ ait essentiellement *une relevée* Γ (sa relevée legendrienne),
- (iii) (E) soit en quelque sorte *universelle* et ne fasse que permettre de faire des γ “données” les solutions “construites” d’un problème d’intégration.

La neurogéométrie renoue ici, pour la neurophysiologie, avec ce qui fut aux sources de la mécanique hamiltonienne en physique : faire des trajectoires observées les projections dans \mathbb{R}^3 de solutions d’équations de Hamilton dans l’espace cotangent $T^*\mathbb{R}^3$.

L’équation (E) s’obtient en écrivant sous la forme $\omega = 0$ la formule $y'(x) = \frac{dy}{dx}$ qui définit la fonction numérique “dérivée” comme quotient d’éléments différentiels symboliques dx et dy . La 1-forme $\omega = dy - pdx$ *préexiste* à l’équation $\omega = 0$ et elle est définie sur $\mathbb{R}^3_{(x,y,p)}$. Elle doit être conçue comme une 1-forme ω à *trois* éléments différentiels $\omega = Adx + Bdy + Cdp$ dont le coefficient C est nul. C’est cette propriété qui fait que l’équation $\omega = 0$ peut définir p comme relation entre dx et dy et donc concerner les courbes planes et leurs relevées legendriennes.

La structure de contact \mathcal{K} est un cadre *universel* pour les courbes planes $y = f(x)$ et c’est pourquoi elle peut modéliser l’architecture fonctionnelle d’une aire neuronale dédiée à la détection de contours encodés dans le signal rétinien. Pour mettre encore plus en relief cette universalité, on peut considérer que $\omega = 0$ est une 1-forme sur $\mathbb{R}^3_{(x,y,p)}$ dont les intégrales dépendent d’une fonction arbitraire $f(x)$. Le *stimulus* qu’est le contour $y = f(x)$ devient alors l’équivalent d’une “*constante d’intégration*” dans l’intégrale générale de $\omega = 0$ et l’on peut ainsi bien séparer l’*input* de la “machine à intégrer” qu’est l’architecture fonctionnelle.

Principe. “Intégrer $\omega = 0$ ” est l’équivalent mathématique, du côté du modèle, de “activer l’aire neuronale”, du côté de la neurophysiologie. _____◇
Ce principe permet de comprendre comment des architectures fonctionnelles peuvent simuler des intégrations en étant simplement activées.

Remarque. Cette géométrie de contact de \mathbb{V}_J correspond aux deux types fondamentaux de connexions intra- et inter-colonnes d’orientation dans $V1$ et admet une interprétation “physique”. Si l’orientation p est “active” en $a = (x, y)$, elle agit comme une “force” : la connexion horizontale excitatrice inter-hypercolonnes qui lui est associée fait qu’un incrément de la position x à p constant induit un incrément $dy = pdx$ de y . En revanche, un incrément de la tangente p à x constant ne change pas y et correspond à des connexions intra-hypercolonne inhibitrices. _____□

Remarque. Nous verrons à la section 2.6 que le tournant post-eulérien qu’a été la géométrisation révolutionnaire du calcul différentiel par Pfaff (avec Gauss et

Jacobi) au début du XIX^e siècle a été en partie inspiré par le problème, nettement plus compliqué que celui des courbes de $\mathbb{R}^2_{(x,y)}$, des *surfaces* de $\mathbb{R}^3_{(x,y,z)}$. Toute la difficulté a été de comprendre que si $z = f(x, y)$ est l'équation d'une surface de \mathbb{R}^3 , il faut traiter les dérivées partielles $p = \frac{\partial f}{\partial x}$ et $q = \frac{\partial f}{\partial y}$ comme de nouvelles variables indépendantes et considérer la 1-forme $\omega = dz - p dx - q dy$ comme une 1-forme à cinq éléments différentiels mais dont les coefficients de dp et dq sont nuls. L'équation différentielle $\omega = 0$ est alors automatiquement vérifiée pour toutes les surfaces régulières $z = f(x, y)$ et les relevées legendriennes de ces dernières (leurs 1-jets) sont des surfaces gauches dans \mathbb{R}^5 . □

2.4. SE(2)-invariance

Choisir des coordonnées (x, y, p) du fibré de contact $V = C\mathbb{R}^2$ pour l'identifier à $J^1(\mathbb{R}, \mathbb{R})$ pour les $p \neq \infty$ équivaut à choisir un point 0 du plan affine \mathbb{R}^2 comme origine et l'axe des x comme axe privilégié. Cela est justifié quand on étudie les contours illusoires reliant deux éléments de contact (a_1, p_1) et (a_2, p_2) , la droite (a_1, a_2) s'imposant comme un axe des x naturel. Certes, un tel choix brise la symétrie de \mathbb{R}^2 , mais rappelons (cf. Vol I, 4.3.8.) que cette brisure de symétrie est compensée par le fait que la structure de V admet une action naturelle du groupe euclidien $SE(2) = \mathbb{R}^2 \times SO(2)$ des déplacements du plan (préservant l'orientation). Ce groupe est le produit semi-direct⁶ du groupe des translations \mathbb{R}^2 et du groupe des rotations $SO(2)$.⁷

Soit (q, r_θ) un élément of $SE(2)$ où q est un point de M et r_θ la rotation d'angle θ . Le triplet (q, r_θ) agit sur les points a de M par

$$(q, r_\theta)(a) = q + r_\theta(a) .$$

Si (q, r_θ) et (s, r_φ) sont 2 éléments de $SE(2)$, leur produit (non commutatif) est donné par la formule de composition :

$$(s, r_\varphi) \circ (q, r_\theta) = (s + r_\varphi(q), r_{\varphi+\theta}) .$$

L'élément neutre est évidemment $(0, r_0 = Id)$ et l'inverse de (q, r_θ) est $(q, r_\theta)^{-1} = (-r_{-\theta}(q), r_{-\theta})$.⁸ Le produit est non commutatif car $(q, r_\theta) \circ (s, r_\varphi) = (q + r_\theta(s), r_{\theta+\varphi})$. Bien sûr $r_{\varphi+\theta} = r_{\theta+\varphi}$, mais $s + r_\varphi(q) \neq q + r_\theta(s)$.

6. Rappelons que, de façon générale, si G est un groupe et si H est un sous groupe de G qui opère sur un autre sous groupe, quant à lui normal, N de G , alors G est le produit semi direct $N \times H$ si sa loi de composition \circ est $(n', h') \circ (n, h) = (n'h'(n), h'h)$.

7. Nous reviendrons plus précisément sur cette notion au chapitre 5, section 3.2.

8. $(q, r_\theta)^{-1} \circ (q, r_\theta) = (-r_{-\theta}(q) + r_{-\theta}(q), r_{-\theta+\theta}) = (0, r_0)$ et $(q, r_\theta) \circ (q, r_\theta)^{-1} = (q + r_\theta(-r_{-\theta}(q)), r_{\theta-\theta}) = (0, r_0)$.

Une façon simple de se représenter cette action linéaire est de se placer dans \mathbb{R}^3 et de représenter (q, r_θ) par la matrice 3×3 $\begin{pmatrix} 1 & 0 \\ q & r_\theta \end{pmatrix}$ agissant sur le vecteur colonne $\begin{pmatrix} 1 \\ a \end{pmatrix}$ pour donner $\begin{pmatrix} 1 \\ q + r_\theta(a) \end{pmatrix}$.

La rotation r_θ agit (de façon non linéaire) sur les fibres de la fibration $V \rightarrow M$ par

$$(q, r_\theta)(p = \tan(\psi)) = \tan(\theta + \psi)$$

(où ψ est la coordonnée angulaire correspondant à p).⁹ Cette action supplémentaire sur la fibre garantit le fait que l'alignement des directions préférentielles est $SE(2)$ -invariant. Elle rend l'action intéressante au niveau de la théorie des représentations des groupes. L'action de $g = (q, r_\theta)$ sur $v = (a, p = \tan(\psi)) \in \mathbb{V}_J$ est donc en définitive $g(v) = (q + r_\theta(a), \tan(\theta + \psi))$. Tout cela est bien expliqué dans les figures 4.24 et 4.25 du Vol I.

C'est un petit exercice scolaire que de vérifier à la main que le champ \mathcal{K} des plans de contact K_v est bien $SE(2)$ -invariant. En effet, dans les coordonnées (x, y, p) , $g = (q, r_\theta)$ agit sur $v = (x, y, p = \tan(\psi)) \in \mathbb{V}_J$ par

$$g(v) = (q_x + \cos(\theta)x - \sin(\theta)y, q_y + \sin(\theta)x + \cos(\theta)y, \tan(\theta + \arctan(p)))$$

L'application linéaire tangente $D_v g$ de g en v a donc pour matrice

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & \frac{\partial}{\partial p} \tan(\theta + \arctan(p)) \end{pmatrix}$$

avec $\frac{\partial}{\partial p} \tan(\theta + \arctan(p)) = \frac{1}{\cos^2(\theta + \psi)} \frac{1}{1+p^2} = \frac{\cos^2(\psi)}{\cos^2(\theta + \psi)}$. On veut montrer que l'on a $D_v g(K_v) = K_{g(v)}$. Considérons alors les générateurs de K_v $\mathbf{t}_1 = \frac{\partial}{\partial x} + p \frac{\partial}{\partial y} = (1, p, 0)$, $\mathbf{t}_2 = \frac{\partial}{\partial p} = (0, 0, 1)$. On obtient

$$\begin{aligned} D_v g(\mathbf{t}_1) &= (\cos(\theta) - p \sin(\theta), \sin(\theta) + p \cos(\theta), 0) \\ &= \left(\cos(\theta) - \frac{\sin(\psi)}{\cos(\psi)} \sin(\theta), \sin(\theta) + \frac{\sin(\psi)}{\cos(\psi)} \cos(\theta), 0 \right) \\ &= \left(\frac{\cos(\theta + \psi)}{\cos(\psi)}, \frac{\sin(\theta + \psi)}{\cos(\psi)}, 0 \right) \\ &= \frac{\cos(\psi)}{\cos(\theta + \psi)} (1, \tan(\theta + \psi), 0) \\ D_v g(\mathbf{t}_2) &= \left(0, 0, \frac{\cos^2(\psi)}{\cos^2(\theta + \psi)} \right) = \frac{\cos^2(\psi)}{\cos^2(\theta + \psi)} (0, 0, 1) \end{aligned}$$

9. Dans V , p peut être ∞ et il n'y a pas de restriction sur l'angle.

qui sont proportionnels aux \mathfrak{t}_1 et \mathfrak{t}_2 en $g(v)$.

2.5. Structure de contact et courbure

Il est très important pour notre modélisation que la structure de contact soit définie par des *plans* et que donc les vecteurs tangents

$$\mathfrak{t} = (a, p; \alpha, \pi) = (x, y, p; \xi, \eta, \pi)$$

des relevées legendriennes Γ dans \mathbb{V}_J des courbes γ dans \mathbb{R}^2 puissent posséder une composante “*verticale*” $\pi = p'$. En effet, comme d’après la condition d’intégrabilité on a $p = y'$ (si x est la variable indépendante), cela signifie que $\pi = p' = y''$. Autrement dit, la composante “*verticale*” des vecteurs tangents des relevées legendriennes Γ dans \mathbb{V}_J est associée à la *courbure* des courbes γ dans \mathbb{R}^2 . Les courbes intégrales de la structure de contact encodent la courbure des courbes dans l’espace de base \mathbb{R}^2 et un vecteur tangent $\mathfrak{t} = (x, y, p; \xi, \eta, \pi)$ à \mathbb{V}_J en $v = (x, y, p)$ définit le *cercle osculateur* de la projection γ dans \mathbb{R}^2 des courbes Γ de \mathbb{V}_J tangentes à \mathfrak{t} .

Dans la mesure où la structure de contact modélise l’architecture fonctionnelle de $V1$ et donc le système des connexions “*horizontales*” cortico-corticales, on voit que le modèle implique que les connexions “*horizontales*” puissent induire des contours *courbes*. Or nous avons vu toute l’importance de ce fait dans la section Vol I, 5.2.2 consacrée aux expériences psychophysiques sur le “*champ d’association*”. Nous avons également vu à la section 6 de ce même chapitre que Steve Zucker et Ohad Ben-Shahar ont beaucoup insisté sur le rôle fonctionnel fondamental de la courbure. Si la co-axialité était strictement définie par des sous-espaces de dimension 1 des $T_v\mathbb{V}_J$, c’est-à-dire en quelque sorte par des “*droites de contact*” au lieu de plans de contact, alors $V1$ ne pourrait intégrer que des contours droits. C’est parce que la co-axialité n’est qu’approximative et définie par des sous-espaces de dimension 2 des $T_v\mathbb{V}_J$, c’est-à-dire par des *plans* de contact, que $V1$ peut intégrer des contours courbes.

Nous reviendrons sur ce problème à la section 4.

2.6. La non-intégrabilité de la structure de contact

Comme nous l’avons déjà noté dans la section Vol I, 5.4.3., il faut souligner un point essentiel à propos de la structure de contact \mathcal{K} . Elle est définie comme le champ de plans $v \in \mathbb{V}_J \mapsto K_v \subset T_v\mathbb{V}_J$ qui est le champ des noyaux d’une 1-forme $\omega_J = dy - p dx$ sur \mathbb{V}_J . Il est alors naturel de se demander si le champ $2D$ des plans K_v , qui possède beaucoup de *courbes* intégrales $1D$, est lui-même intégrable, autrement dit s’il existe des *surfaces* S de \mathbb{V}_J tangentes à K_v en chacun de leur point v , c’est-à-dire telles que $T_vS = K_v$. Cela n’est pas du tout le cas et ne peut pas l’être. Le champ $v \mapsto K_v$ est le prototype d’un champ *non* intégrable. On peut s’en convaincre intuitivement : le champ K_v se “*vriille*” trop pour que des surfaces

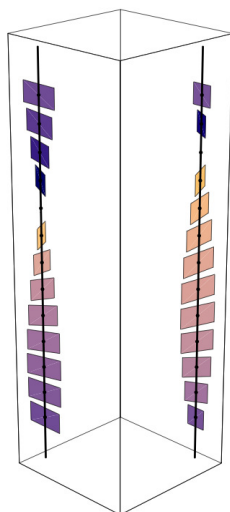


FIGURE 4. Le champ des plans de contact K_v “vrille” trop pour être intégrable.

intégrales puissent exister (cf. figure 4). La non-intégrabilité est, comme on dit, une condition de “twisting”.

De façon plus technique, on remarque que si $\mathbf{t} \in T_v \mathbb{V}_J$ est un vecteur tangent à \mathbb{V}_J de composantes $\mathbf{t} = (\xi, \eta, \pi)$, la condition d’annulation $\omega_J(\mathbf{t}) = 0$ définit, nous l’avons vu, le plan $-p\xi + \eta = 0$ i.e. $\eta = p\xi$ (ce qui n’est rien d’autre que $p = \frac{dy}{dx}$). Si l’on identifie \mathbb{V}_J à \mathbb{R}^3 , K_v devient le plan “vertical” au-dessus de la droite “horizontale” de pente p . Lorsque l’on se déplace dans la fibre \mathbb{R}_a au-dessus de a , ce plan tourne avec p (sa pente est égale à la “hauteur” dans la fibre) comme le montre la figure 4.

On démontre que cette non-intégrabilité du champ des plans de contact K_v résulte de ce que la *condition de Frobenius* $\omega_J \wedge d\omega_J = 0$ (ou $d\omega_J(\mathbf{t}, \mathbf{t}') = 0$ pour tous les \mathbf{t} et \mathbf{t}' tels que $\omega_J(\mathbf{t}) = \omega_J(\mathbf{t}') = 0$) *n’est pas* satisfaite. Cette condition provient du théorème disant qu’une condition nécessaire et suffisante d’intégrabilité est que, si $\{\mathbf{t}_1, \mathbf{t}_2\}$ est une base (locale) de K_v , le crochet de Lie $[\mathbf{t}_1, \mathbf{t}_2]$ appartienne à K_v , et

que donc K_v soit une *sous-algèbre de Lie* de $T_v\mathbb{V}_J$.¹⁰ Mais pour $\omega_J = dy - pdx$ on a

$$\begin{aligned} d\omega_J &= - \left(\frac{\partial p}{\partial x} dx \wedge dx + \frac{\partial p}{\partial y} dy \wedge dx + \frac{\partial p}{\partial p} dp \wedge dx \right) + d^2y - pd^2x \\ &= -dp \wedge dx = dx \wedge dp \end{aligned} \quad (1)$$

et donc

$$\omega_J \wedge d\omega_J = (-pdx + dy) \wedge dx \wedge dp = dy \wedge dx \wedge dp = -dx \wedge dy \wedge dp .$$

Mais cette 3-forme est une forme volume de \mathbb{V}_J et ne peut donc en aucun cas être nulle. Elle est même “maximalement” non nulle. On peut aussi remarquer que pour la base naturelle

$$\left\{ \mathbf{t}_1 = \frac{\partial}{\partial x} + p \frac{\partial}{\partial y} = (1, p, 0), \mathbf{t}_2 = \frac{\partial}{\partial p} = (0, 0, 1) \right\}$$

de K_v introduite plus haut, on a, pour $p \neq 0$, le crochet de Lie $[\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3 = -\frac{\partial}{\partial y} = (0, -1, 0)$ et que $\mathbf{t}_3 = (0, -1, 0) \notin K_v$ car, pour ce vecteur, $-p\xi + \eta = -0 + (-1) = -1 \neq 0$.

Le champ $-\mathbf{t}_3 = \chi = \frac{\partial}{\partial y}$ partout transverse au champ \mathcal{K} de plans de contact $\{\mathbf{t}_1, \mathbf{t}_2\}$ et satisfaisant $\omega_J(\chi) = 1$ s’appelle le *champ caractéristique* – ou *champ de Reeb* – du champ \mathcal{K} défini par ω_J . $\mathcal{K} \oplus \chi$ est une décomposition en somme directe de TV . Ce champ préserve ω_J car, d’après la “formule d’homotopie” de Cartan on a $\mathcal{L}_\chi \omega_J = i_\chi d\omega_J + d(i_\chi \omega_J)$ et, d’une part, puisque $i_\chi \omega_J = \omega_J(\chi) = 1$, on a $d(i_\chi \omega_J) = 0$, alors que, d’autre part, on a $i_\chi d\omega_J = d\omega_J(\chi, \bullet) = 0$.¹¹

La conséquence de la non-intégrabilité – appelée *non holonomie* – de la structure de contact \mathcal{K} est que ses intégrales sont nécessairement *1-dimensionnelles*. Autrement dit, \mathcal{K} est fonctionnellement dédiée à l’intégration des *courbes*. Qui plus est, la non-intégrabilité de \mathcal{K} implique que la géométrie *immanente* de \mathbb{V}_J , qui modélise l’architecture fonctionnelle de $V1$, est *complètement différente de la géométrie euclidienne même au niveau infinitésimal*. Nous verrons plus bas au chapitre 14 qu’il s’agit en fait d’une géométrie *sous-riemannienne*.

Faisons ici une remarque sur laquelle nous reviendrons à la section 2.7.6. À travers le simple changement de coordonnées $u \leftrightarrow -p, v \leftrightarrow x, w \leftrightarrow y$, la 1-forme $\omega_J = dy - pdx$ se transforme dans la structure de contact sur \mathbb{R}^3 (de coordonnées (u, v, w)) définie par la 1-forme $\omega_D = u dv + dw$. Cette dernière est la structure de contact standard et un célèbre théorème de Darboux dit que si ϖ est une 1-forme de contact

10. Il s’agit bien des mêmes conditions. En effet $d\omega_J(\mathbf{t}, \mathbf{t}') = \pi\xi' - \pi'\xi$ et comme $[\mathbf{t}, \mathbf{t}'] = [(\xi, \eta, \pi), (\xi', \eta', \pi')] = (0, \pi\xi' - \pi'\xi, 0)$, on a $\omega_J([\mathbf{t}, \mathbf{t}']) = \pi\xi' - \pi'\xi = d\omega_J(\mathbf{t}, \mathbf{t}')$. Donc dire que $d\omega_J(\mathbf{t}, \mathbf{t}') = 0$ pour tous $\mathbf{t}, \mathbf{t}' \in \mathcal{K}$ équivaudrait à dire que $\omega_J([\mathbf{t}, \mathbf{t}']) = 0$ sur la distribution \mathcal{K} , i.e. que $[\mathbf{t}, \mathbf{t}'] \in \mathcal{K}$ si $\mathbf{t}, \mathbf{t}' \in \mathcal{K}$. Lorsque c’est le cas on dit que la distribution est *involutive*. La distribution \mathcal{K} n’est pas involutive.

11. $d\omega_J(\chi, \mathbf{t}) = \chi_x \mathbf{t}_p - \chi_p \mathbf{t}_x = 0$ car $\chi_x = \chi_p = 0$.

sur une variété M de dimension 3, c'est-à-dire une 1-forme satisfaisant la condition de non-intégrabilité de Frobenius $\varpi \wedge d\varpi =$ forme volume, alors on peut toujours la mettre localement sous la forme standard par un changement différentiable de coordonnées approprié. Nous aurons par exemple à introduire plus bas la 1-forme de contact $\omega_S = \cos(\theta) dy - \sin(\theta) dx$ sur $\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1$. On a

$$\begin{aligned} d\omega_S &= -\sin(\theta) d\theta \wedge dy - \cos(\theta) d\theta \wedge dx \\ \omega_S \wedge d\omega_S &= (\cos(\theta) dy - \sin(\theta) dx) \wedge (-\sin(\theta) d\theta \wedge dy - \cos(\theta) d\theta \wedge dx) \\ &= -\cos^2(\theta) dy \wedge d\theta \wedge dx + \sin^2(\theta) dx \wedge d\theta \wedge dy \\ &= -dx \wedge dy \wedge d\theta \end{aligned}$$

et $\omega_S \wedge d\omega_S$ est bien une forme volume.¹² Soit (a_0, θ_0) un point de \mathbb{V}_S et centrons les coordonnées (x, y, θ) sur ce point. On développe ω_S en série de Taylor et on écrit

$$\omega_S = dy - \theta dx + \left(-\frac{\theta^2}{2} dy + \frac{\theta^3}{6} dx + \dots \right) = \omega_0 + \omega'$$

On considère alors l'interpolation $\omega_\lambda = \omega_0 + \lambda\omega'$, $\lambda \in [0, 1]$, entre ω_0 et ω_S et on construit une famille φ_λ de difféomorphismes locaux (de changements de variables) qui, pour chaque λ , ramène ω_λ sur ω_0 . Or $\omega_0 = \omega_J$ avec $p = \theta$, le développement au premier ordre de $p = \tan(\theta)$.

2.7. La “simplexité” de la notion de jet

Nous allons approfondir un peu les notions de jet et de structure de contact. Mais nous voudrions au préalable noter qu'elles fournissent un excellent exemple de ce qu'Alain Berthoz a appelé la *simplexité* à propos des structures neuronales produites par l'évolution biologique. Il s'agit d'une remarquable convergence entre un résultat de l'évolution biologique et un approfondissement fondationnel de structures formelles.¹³

Dans son ouvrage de 2009 *La simplexité* [54], Alain Berthoz a introduit le néologisme de “simplexité” pour qualifier tout un ensemble de solutions originales trouvées par l'évolution biologique pour traiter la complexité intraitable du réel. Ce qu'il a appelé le “détour” de la simplexité, consiste à

12. Nous avons utilisé le fait que $dx \wedge dx = dy \wedge dy = 0$, $dy \wedge d\theta \wedge dx = dx \wedge dy \wedge d\theta$ et $dx \wedge d\theta \wedge dy = -dx \wedge dy \wedge d\theta$.

13. Nous avons rencontré dans le Vol I une convergence remarquable du même ordre concernant cette fois les profils récepteurs des neurones corticaux visuels. Dans la section 3.6 du chapitre 3 nous évoquons des travaux comme ceux de Joseph Atick [21] ou Jean-Pierre Nadal [546] montrant comment des profils récepteurs biologiques concrets peuvent se retrouver à partir d'hypothèses générales abstraites sur l'optimisation de l'efficacité du codage informationnel des signaux.

“décomposer les problèmes compliqués en sous-problèmes plus simples, grâce à des modules spécialisés, quitte à devoir ensuite recomposer l’ensemble.” ([?], p.22)

Il donne l’exemple de l’usage de variables mécaniques composites dans la robotique bioinspirée :

“Au lieu de la variable simple qu’il [le roboticien] veut contrôler, [il peut utiliser] un mélange de variables combinant position, vitesse et accélération.” (p.31)

Et il insiste beaucoup sur le fait que “la géométrie est un outil de la simplicité” (p.165) car

“l’usage de la géométrie et, donc, de l’espace pour organiser l’activité neuronale conduit à des simplifications remarquables en matière de traitement, de flexibilité et d’adaptabilité cérébrale.” (p.166)

La notion de jet fournit un fort bel exemple de simplicité. Le problème complexe est de calculer des dérivées de certaines fonctions relativement aux variables de position rétinienne $a = (x, y)$. Comme nous y avons abondamment insisté il est trop complexe pour être effectué par des processeurs ponctuels comme les neurones. La micro-anatomie hypercolumnnaire et l’architecture fonctionnelle de l’aire visuelle primaire $V1$ constituent une solution simplexe trouvée par l’évolution à ce problème complexe qui permet de n’utiliser que des opérations simples de type “prendre la valeur d’une multivariable en un point” :

1. on ajoute une nouvelle variable indépendante p aux variables de position rétinienne $a = (x, y)$: il s’agit de la variable d’orientation à laquelle sont sélectifs les neurones “simples” de $V1$;

2. on organise en modules spécialisés (hypercolonnes d’orientation) les valeurs des triplets (a, p) (éléments de contact) ;

3. le traitement d’un input consiste alors simplement à mesurer les valeurs ponctuelles des (a, p) activés ;

4. mais il faut une “recomposition” complémentant la “décomposition” hypercolumnnaire et, pour recomposer, on introduit une architecture fonctionnelle qui garantit que prendre des valeurs ponctuelles relativement aux trois variables (a, p) équivaut bien à dériver relativement aux variables initiales (x, y) .

C’est bien cette simplicité que formalisent les concepts mathématiques de jet et de structure de contact (cf. notre étude [439]).

La solution “simplexe” au problème “complexe” du calcul *neuronal* d’une dérivée consiste ainsi à permettre au système neuronal d’accéder à une information géométrique locale non ponctuelle dans l’espace rétinien bidimensionnel R sous le format purement ponctuel du calcul d’une simple valeur numérique, mais cela à condition de calculer dans l’espace cortical tridimensionnel V des jets.

Calculer une dérivée directement dans R en utilisant la formule de la dérivée ou la calculer de façon “simplexe” dans V sont des opérations *formellement* équivalentes. Mais les modes de calcul, les “formats”, sont complètement différents.

Nous touchons ici du doigt l’apport de la neurophysiologie aux fondements de la géométrie. S’il est plausible, comme nous le croyons, que l’évolution biologique ait matérialisé dans la machine corticale d’organismes supérieurs une sorte de géométrie de contact, alors cette dernière apparaît comme une sorte de “décompilation” d’algorithmes neuronaux “simplexes”. La géométrie de contact reflèterait elle-même des “simplifications remarquables” inventées par la phylogénèse.

2.8. De Frobenius à Hörmander

Approfondissons maintenant un peu théoriquement les notions de jet et de structure de contact

Nous ne pouvons pas entrer ici dans la théorie générale, mais soulignons que ce que nous venons de voir pour \mathbb{V}_J et \mathbb{V}_S est général. Si M est une variété différentiable de dimension n et si \mathcal{K} est une distribution régulière d’hyperplans tangents de dimension $k < n$ (i.e., en tout point $x \in M$, K_x est un k -hyperplan de T_xM), il y a alors deux cas extrêmes :

1. \mathcal{K} est “involutive” ou intégrable au sens où les crochets de champs de vecteurs tangents appartenant à \mathcal{K} appartiennent aussi à \mathcal{K} et alors \mathcal{K} est intégrable et définit un feuilletage de M par des feuilles qui sont des sous-variétés de dimension k . C’est la condition de Frobenius et dans ce cas il est impossible de passer d’une feuille à l’autre en restant tangent à \mathcal{K} . En codimension 1 (i.e. les K_x sont des hyperplans) $\mathcal{K} = \ker(\omega)$ est le champ des noyaux d’une 1-forme ($\neq 0$) de Pfaff. L’intégrabilité, à savoir que par chaque point de M il passe une et une seule hypersurface intégrale de ω , équivaut au fait que ω est exacte à un facteur d’intégration près : il existe des fonctions telles que $\omega = f dh$. On a alors la condition de Frobenius $\omega \wedge d\omega = f dh \wedge (df \wedge dh) = 0$. Cela signifie que les chemins tangents à K_x en x sont tous tangents à l’hypersurface intégrale passant par x .

Sur l’impulsion de Max Born, Carathéodory démontra la réciproque, à savoir que s’il existe des points non connectables par des courbes intégrales de \mathcal{K} alors \mathcal{K} est intégrable (cf. Montgomery [375], p. xii). Le lien avec la *thermodynamique* est historique. La chaleur n’est pas une fonction d’état d’un système thermodynamique et la différentielle dQ n’est pas une 1-forme exacte, elle dépend des chemins suivis pour faire évoluer le système. Mais elle admet un facteur intégrant et s’écrit $dQ = T dS$ au moyen des deux fonctions d’état T (température) et S (entropie). Les transformations idéales “adiabatiques” (sans transfert thermique) sont des évolutions avec $dQ = 0$. C’est

à cause de cette analogie historique avec la thermodynamique que l'appellation de "Carnot-Carathéodory" intervient centralement en géométrie sous-riemannienne (cf. chapitre 5, section 7 et chapitre 6, section 8).

2. \mathcal{K} est complètement non holonome ("bracket generating") au sens où les crochets de champs de vecteurs tangents appartenant à \mathcal{K} engendrent *tout* le fibré tangent TM et alors \mathcal{K} est maximale non intégrable et, d'après le théorème de Chow-Rashevskii¹⁴, l'on peut joindre deux points quelconques de M par des courbes restant tangentes à \mathcal{K} . C'est la condition dite *d'Hörmander*. Le théorème de Chow n'est pas trivial et repose sur le fait que les commutateurs de champs tangents à \mathcal{K} permettent de se déplacer transversalement à \mathcal{K} . Il est en fait une conséquence du "ball-box theorem" des métriques de Carnot-Carathéodory que nous expliquerons à la section 8 du chapitre 6.

2.9. Champs et points caractéristiques des surfaces de V

2.9.1. Définition et exemple.

Si $F : V \rightarrow \mathbb{R}$ ($V = \mathbb{V}_J$ ou \mathbb{V}_S) est une fonction régulière sur V , génériquement $F = 0$ définit une surface S régulière de V (nous sommes en dimension 3). S ne peut pas être une intégrale de \mathcal{K} et en général le plan tangent $T_v S$ à S en un point v sera *transverse* au plan de contact K_v . Leur intersection sera une droite D_v et l'on obtient ainsi un champ $D \subset \mathcal{K}$ de directions tangentes à S dit *champ caractéristique*. Ses intégrales sont dites *courbes caractéristiques* de S . Mais il peut néanmoins exister des points exceptionnels v où $T_v S = K_v$. De tels points sont appelés *points caractéristiques* de S . Ils existent en général même si S est sans points singuliers au sens classique. Mais, dans la mesure où la structure de contact \mathcal{K} est complètement non intégrable, ils sont nécessairement rares génériquement et même isolés.

Pour illustrer ces champs caractéristiques nous allons considérer l'exemple très "simple" de la sphère unité S de $V = \mathbb{V}_J = \mathbb{R}^3$ d'équation

$$F(x, y, z) = x^2 + y^2 + z^2 - 1 = 0$$

l'axe z étant interprété comme celui de la variable p . Malgré la symétrie sphérique les calculs sont déjà assez compliqués et la géométrie est déjà non triviale.

Utilisons les coordonnées sphériques $x = \cos(\theta) \cos(\varphi)$, $y = \sin(\theta) \cos(\varphi)$, $z = p = \sin(\varphi)$. Elles font de S une image du cylindre $C = [0, 2\pi[\times [-\frac{\pi}{2}, \frac{\pi}{2}]$. Il s'agit d'un difféomorphisme sauf aux pôles où les cercles $[0, 2\pi[\times \{-\frac{\pi}{2}\}$ et $[0, 2\pi[\times \{\frac{\pi}{2}\}$ s'écrasent sur des points.

Le long de l'équateur E , $\varphi = 0$, le plan de contact K_v est le plan vertical (x, p) d'équation $y = 0$ et le plan tangent à S , $T_v S$, est le plan vertical sur la tangente à

14. Wei-Liang Chow 1939 et Petr Konstanovich Rashevskii 1938.

E . Il tourne donc autour de l'axe vertical. Si $\theta \neq \pm \frac{\pi}{2}$, la direction caractéristique D_v est la direction verticale et se projette sur un point sur le plan (x, y) . En particulier, si $\theta = 0$ (i.e. $x = \pm 1, y = 0, p = 0$) $T_v S$ est le plan vertical orthogonal à K_v . Si en revanche $\theta = \pm \frac{\pi}{2}$ (i.e. $x = 0, y = \pm 1, p = 0$), alors $T_v S = K_v$ et les points sont caractéristiques.

Au pôle Nord $p = +1$, $T_v S$ est horizontal et K_v est le plan vertical de pente 1 relativement à (x, y) . D_v est donc la direction horizontale (i.e. parallèle au plan (x, y)) de pente 1. Il en va de même au pôle Sud $p = -1$ avec la pente -1 .

On notera que l'équateur E est le *contour apparent* de la sphère S projetée sur le plan (x, y) parallèlement à la direction p . Il est la projection $\pi(\Sigma)$ du lieu Σ de S où la direction verticale de la projection π est tangente à S .

2.9.2. Le champ Δ sur C .

On peut calculer directement le champ de directions D sur S . Mais pour s'en faire d'abord une idée simple on peut construire le champ caractéristique Δ sur le cylindre C des (θ, φ) . On calcule les différentielles dx, dy, dp et on écrit la condition $dy = p dx$. Cela donne une relation entre $d\theta$ et $d\varphi$ que l'on écrit $\frac{d\theta}{\Delta_\theta} = \frac{d\varphi}{\Delta_\varphi}$, $(\Delta_\theta, \Delta_\varphi)$ étant les composantes de Δ (définies à un facteur d'échelle près). On obtient :

$$\begin{cases} dx = -\sin(\theta) \cos(\varphi) d\theta - \cos(\theta) \sin(\varphi) d\varphi \\ dy = \cos(\theta) \cos(\varphi) d\theta - \sin(\theta) \sin(\varphi) d\varphi \\ dp = \cos(\varphi) d\varphi \end{cases}$$

$$\begin{cases} dy = p dx \\ \cos(\theta) \cos(\varphi) d\theta - \sin(\theta) \sin(\varphi) d\varphi = -\sin(\theta) \cos(\varphi) \sin(\varphi) d\theta - \cos(\theta) \sin^2(\varphi) d\varphi \\ \frac{d\theta}{\sin(\varphi)(\sin(\theta) - \cos(\theta) \sin(\varphi))} = \frac{d\varphi}{\cos(\varphi)(\cos(\theta) + \sin(\theta) \sin(\varphi))} \\ \Delta_\theta = \sin(\varphi) (\sin(\theta) - \cos(\theta) \sin(\varphi)) \\ \Delta_\varphi = \cos(\varphi) (\cos(\theta) + \sin(\theta) \sin(\varphi)) \end{cases}$$

Étant donné un tel champ, une première information est donnée par les points fixes $\Delta_\theta = 0, \Delta_\varphi = 0$ et la linéarisation du champ Δ dans leur voisinage. Les points fixes sont solutions de $\sin(\varphi) = 0$ et $\cos(\theta) = 0$ ainsi que de $\cos(\varphi) = 0$ et $\sin(\theta) - \cos(\theta) \sin(\varphi) = 0$ (si $\sin(\varphi)$ et $\cos(\varphi)$ sont $\neq 0$, il n'y a pas de solutions). On obtient ainsi les six points fixes

$$\begin{cases} P_1 = \left(\varphi = \frac{\pi}{2}, \theta = \frac{\pi}{4} \right), \\ P_2 = \left(\varphi = \frac{\pi}{2}, \theta = \frac{5\pi}{4} \right), \\ P_3 = \left(\varphi = 0, \theta = \frac{\pi}{2} \right), \\ P_4 = \left(\varphi = 0, \theta = \frac{3\pi}{2} \right), \\ P_5 = \left(\varphi = -\frac{\pi}{2}, \theta = \frac{3\pi}{4} \right), \\ P_6 = \left(\varphi = -\frac{\pi}{2}, \theta = \frac{7\pi}{4} \right). \end{cases}$$

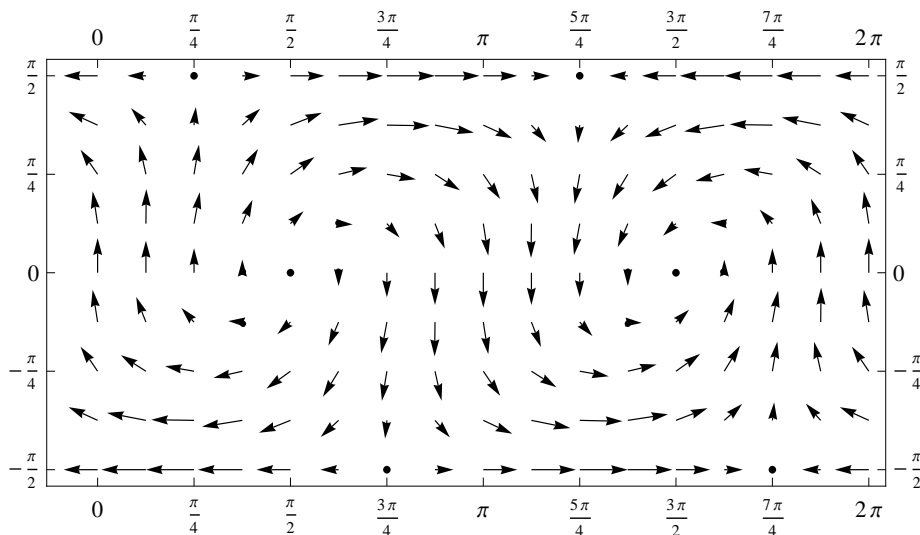


FIGURE 5. Le champ caractéristique Δ du cylindre $C = [0, 2\pi[\times [-\frac{\pi}{2}, \frac{\pi}{2}]$ relativement aux coordonnées sphériques (θ, φ) . Les points fixes y apparaissent comme des points épais.

La figure 5 montre le champ Δ . Les points fixes y sont visibles comme des points épais.

La linéarisation des points fixes s'obtient en développant Δ_θ et Δ_φ au premier ordre. On trouve :

1. Pour $P_1 = (\theta = \frac{\pi}{4}, \varphi = \frac{\pi}{2})$, $\Delta_\theta \sim \sqrt{2}(\theta - \frac{\pi}{4})$, $\Delta_\varphi \sim -\sqrt{2}(\varphi - \frac{\pi}{2})$: point selle (col) répulsif dans la direction θ et attractif dans la direction φ . La matrice A du champ, $A = \sqrt{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, a pour valeurs propres $\sqrt{2}$ et $-\sqrt{2}$ de signes opposés et les vecteurs propres sont les axes.
2. Pour $P_2 = (\theta = \frac{5\pi}{4}, \varphi = \frac{\pi}{2})$, $\Delta_\theta \sim -\sqrt{2}(\theta - \frac{5\pi}{4})$, $\Delta_\varphi \sim \sqrt{2}(\varphi - \frac{\pi}{2})$: point selle (col) attractif dans la direction θ et répulsif dans la direction φ .
3. Pour $P_3 = (\theta = \frac{\pi}{2}, \varphi = 0)$, $\Delta_\theta \sim \varphi$, $\Delta_\varphi \sim \varphi - (\theta - \frac{\pi}{2})$: point foyer dextrogyre répulsif. La matrice A du champ, $A = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$, a pour valeurs propres imaginaires conjuguées $\frac{1 \pm i\sqrt{3}}{2}$ de partie réelle > 0 (d'où la répulsivité).
4. Pour $P_4 = (\theta = \frac{3\pi}{2}, \varphi = 0)$, $\Delta_\theta \sim -\varphi$, $\Delta_\varphi \sim -\varphi + (\theta - \frac{3\pi}{2})$: point foyer lévogyre attractif. La matrice A du champ, $A = \sqrt{2} \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$, a pour valeurs propres imaginaires conjuguées $\frac{-1 \pm i\sqrt{3}}{2}$ de partie réelle < 0 (d'où l'attractivité).

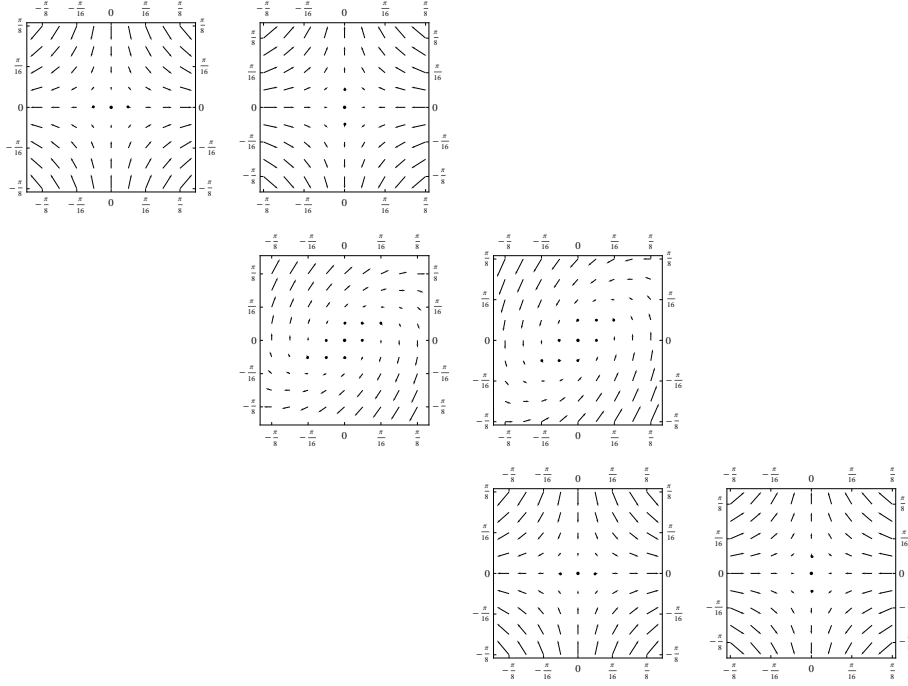


FIGURE 6. La linéarisation des points fixes (voir texte). Pour des raisons de largeur de l'image la diagonale des points 2-4-6 est beaucoup plus proche de la diagonale 1-3-5 qu'elle ne l'est en réalité.

5. Pour $P_5 = (\theta = \frac{3\pi}{4}, \varphi = -\frac{\pi}{2})$, $\Delta_\theta \sim \sqrt{2}(\theta - \frac{3\pi}{4})$, $\Delta_\varphi \sim -\sqrt{2}(\varphi + \frac{\pi}{2})$: point selle (col) répulsif dans la direction θ et attractif dans la direction φ .
6. Pour $P_6 = (\theta = \frac{7\pi}{4}, \varphi = -\frac{\pi}{2})$, $\Delta_\theta \sim -\sqrt{2}(\theta - \frac{7\pi}{4})$, $\Delta_\varphi \sim \sqrt{2}(\varphi + \frac{\pi}{2})$: point selle (col) attractif dans la direction θ et répulsif dans la direction φ .

La figure 6 montre la linéarisation des points fixes.

La figure 5 rend également évident le fait que sur l'équateur E , $\varphi = 0$, $\Delta_\theta = 0$ et le champ Δ est donc vertical de composante $\Delta_\varphi = \cos(\theta)$. En revanche pour $\varphi = \pm\frac{\pi}{2}$, $\Delta_\varphi = 0$ et le champ Δ est donc horizontal de composante $\Delta_\theta = \pm\sin(\theta) - \cos(\theta)$.

Les points col sont reliés aux points foyer par des trajectoires spéciales qui sont des séparatrices entre des régimes différents de trajectoires. La direction verticale répulsive issue du col $P_2 = (\theta = \frac{5\pi}{4}, \varphi = \frac{\pi}{2})$ va s'enrouler sur le foyer lévogyre attractif $P_4 = (\theta = \frac{3\pi}{2}, \varphi = 0)$ (trajectoire T_{24}). Il en va de même de celle issue du col $P_6 = (\theta = \frac{7\pi}{4}, \varphi = -\frac{\pi}{2})$ (trajectoire T_{64}). En revanche la trajectoire arrivant au col $P_1 = (\theta = \frac{\pi}{4}, \varphi = \frac{\pi}{2})$ dans sa direction verticale attractive provient du foyer dextrogyre répulsif $P_3 = (\theta = \frac{\pi}{2}, \varphi = 0)$ (trajectoire T_{31}). Il en va de même de la trajectoire arrivant au col $P_5 = (\theta = \frac{3\pi}{4}, \varphi = -\frac{\pi}{2})$ dans sa direction verticale attractive (trajectoire T_{35}).

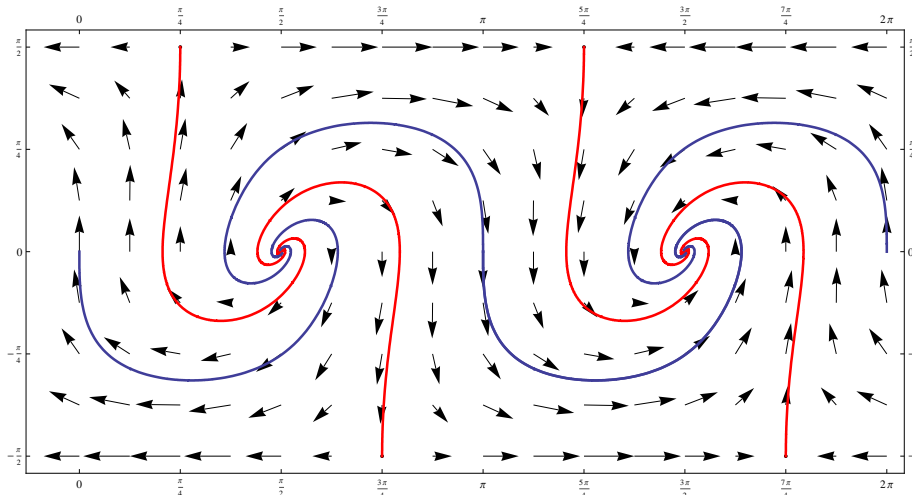


FIGURE 7. Le champ caractéristique Δ de la figure 5 avec, en rouge, ses trajectoires spéciales et, en noir, les trajectoires (symétriques par rapport à leur point d'inflexion) partant du foyer dextrogyre répulsif 3 ($\theta = \frac{\pi}{2}, \varphi = 0$) pour aboutir au foyer lévogyre attractif 4 ($\theta = \frac{3\pi}{2}, \varphi = 0$) en passant l'une par le point ($\theta = 0, \varphi = 0$) et l'autre par le point ($\theta = \pi, \varphi = 0$), points où le champ Δ est vertical et d'intensité maximale.

La figure 7 montre, en rouge, ces trajectoires spéciales T_{jk} du champ Δ ainsi que, en noir, les trajectoires (symétriques par rapport à leur point d'inflexion) issues du foyer dextrogyre répulsif P_3 ($\theta = \frac{\pi}{2}, \varphi = 0$) et aboutissant au foyer lévogyre attractif P_4 ($\theta = \frac{3\pi}{2}, \varphi = 0$) en passant l'une par le point ($\theta = 0, \varphi = 0$) et l'autre par le point ($\theta = \pi, \varphi = 0$), points où le champ Δ est vertical et d'intensité maximale. Grâce aux vecteurs de Δ représentés on reconstruit très bien la forme des trajectoires non symétriques par rapport à leur point d'inflexion.

Les symétries du champ Δ sont riches. On a en particulier les transformations suivantes.

1. La symétrie σ_0 par rapport au point central $P_0 = (\varphi = 0, \theta = \pi)$ change de signe (mod 2π) les coordonnées sphériques (θ, φ) et laisse les directions invariantes ainsi que leur vecteur représentatif. Elle laisse fixe P_0 et l'origine ($\theta = 0, \varphi = 0$) et échange les points fixes P_1 et P_6 , P_2 et P_5 , P_3 et P_4 .
2. La symétrie σ_3 par rapport au point fixe $P_3 = (\theta = \frac{\pi}{2}, \varphi = 0)$ change θ en $\pi - \theta$ et φ en $-\varphi$, ce qui change de signe le vecteur sur la direction Δ mais laisse la direction invariante. Elle laisse fixes P_3 (évidemment) et P_4 et échange P_1 et P_5 ainsi que P_2 et P_6 . Elle est identique à la symétrie σ_4 par rapport au point fixe $P_4 = (\theta = \frac{3\pi}{2}, \varphi = 0)$.

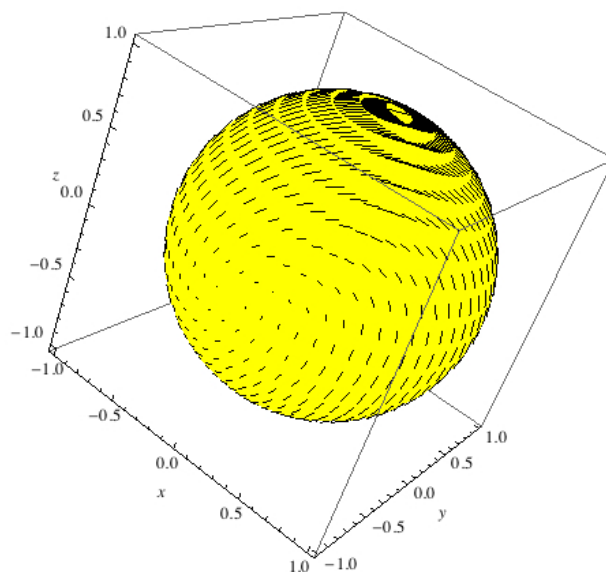


FIGURE 8. Le champ caractéristique D sur la sphère unité S . L'axe des z est celui de la variable p .

3. La composée de σ_0 et σ_3 est la translation τ de π de la coordonnée θ , φ restant constante. Elle laisse la direction Δ invariante mais inverse le vecteur représentatif.

Abstraitement, ces transformations forment un groupe multiplicatif d'ordre 4 composé de l'identité et de trois nilpotents $g^2 = Id$ (i.e. $g = g^{-1}$) chaque produit de deux d'entre eux donnant le troisième. Interprété additivement c'est le produit $\mathbb{F}_2 \times \mathbb{F}_2$ du corps à deux éléments \mathbb{F}_2 par lui-même, la nilpotence se traduisant par la caractéristique 2 du corps premier \mathbb{F}_2 .

2.9.3. *Le champ D sur S .*

Quand on projette le cylindre C des coordonnées sphériques (θ, φ) sur la sphère unité S , on obtient le champ caractéristique D représenté à la figure 8.

On le calcule directement en écrivant qu'il est à la fois dans les plans tangents de S (i.e. $dF = 0$ et donc $xdx + ydy + pdp = 0$) et dans les plans de contact $dy = p dx$. On obtient $(x + py) dx + pdp = 0$, ce qui donne

$$\frac{dx}{p} = -\frac{dp}{x + py} = \frac{dy}{p^2}.$$

Le champ caractéristique D de S est donc (à un facteur d'homothétie près) $D_x = p$, $D_y = p^2$, $D_p = -(x + py)$. Sur l'équateur E , $p = 0$ et donc $D_x = 0$, $D_y = 0$, $D_p = -x$. Le champ est vertical et il s'annule pour $x = 0$. Ces points $(x = 0, y = \pm 1, p = 0)$ sont les foyers attractifs et répulsifs de la figure 6. En leur voisinage la linéarisation de D est $D_x = p$, $D_y = 0$, $D_p = -(x \pm p)$ et les valeurs propres sont 0 et $\frac{\mp 1 \pm i\sqrt{3}}{2}$.

Comme les parallèles $[0, 2\pi[\times \{\frac{\pi}{2}\}$ et $[0, 2\pi[\times \{-\frac{\pi}{2}\}$ de C s'écrasent sur les pôles (respectivement Nord et Sud) de S , les points fixes P_1 et P_2 disparaissent en fusionnant et il ne reste plus que la direction caractéristique du pôle Nord horizontale de pente 1. De même les points fixes P_5 et P_6 disparaissent en fusionnant et il ne reste plus que la direction caractéristique du pôle Sud horizontale de pente -1 . On a donc un enchaînement des trajectoires spéciales $T_{31} \rightarrow T_{24}$ et $T_{35} \rightarrow T_{64}$.

La figure 9 représente sur la sphère transparente les trajectoires rouges et noires de la figure 7 ainsi que trois plans de contact : celui du pôle Nord, celui du foyer attractif $P_4 = (x = 0, y = -1, p = 0)$ où $D_v = T_v S$ et celui de l'intersection de l'équateur E avec un morceau de la trajectoire spéciale T_{64} (qui vient du pôle Sud après être partie du foyer répulsif $P_3 = (x = 0, y = 1, p = 0)$ en suivant T_{35}) et également un morceau de la trajectoire spéciale T_{24} (qui vient elle du pôle Nord après être partie du foyer répulsif P_3 en suivant T_{31}). La figure montre également la trajectoire noire passant par P_0 .¹⁵

La figure 10 montre sur la sphère opaque les trajectoires de la figure 9 en tant qu'intégrales du champ de directions D . La figure 11 les montrent de même sur la sphère transparente et sous un autre point de vue.

Enfin, la figure 12 montre trois points de vue sur les trajectoires des figures précédentes.

Les symétries σ_0 , σ_3 , τ du cylindre C des (θ, φ) ne sont plus vraiment des symétries car elles ne laissent pas le champ D invariant. Elles consistent pour chaque point $v \in S$ à prendre le cercle section de S par le plan passant par v et parallèle à un plan de coordonnées et à prendre le point diamétralement opposé à v sur ce cercle. Autrement dit ce sont les symétries par rapport aux axes.

1. σ_0 devient la symétrie A_x par rapport à l'axe des x : $x \rightarrow x$, $y \rightarrow -y$, $p \rightarrow -p$. Elle laisse fixes les points $(x = \pm 1, y = 0, p = 0)$ et échange les pôles. La transformation du champ est $D_x \rightarrow -D_x$, $D_y \rightarrow D_y$, $D_p \rightarrow D_p$ qui est la symétrie par rapport au plan (y, p) .
2. σ_3 devient la symétrie A_y par rapport à l'axe des y : $x \rightarrow -x$, $y \rightarrow y$, $p \rightarrow -p$. Elle laisse fixes les points $(x = 0, y = \pm 1, p = 0)$ et échange les pôles. La transformation du champ est $D_x \rightarrow -D_x$, $D_y \rightarrow D_y$, $D_p \rightarrow -D_p$ qui est la symétrie par rapport à l'axe des y .

15. Nous gardons les mêmes symboles P_j et T_{jk} que pour les trajectoires sur C .

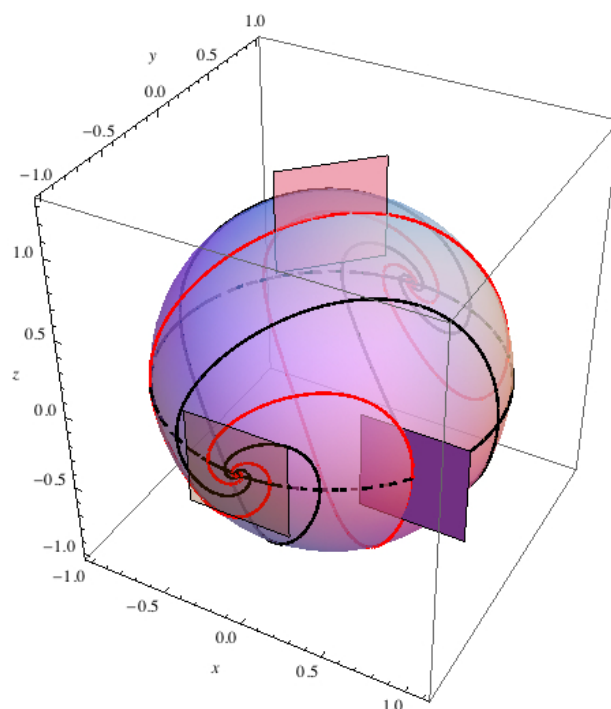


FIGURE 9. La sphère S et les trajectoires rouges et noires de la figure 7 ainsi que trois plans de contact : au pôle Nord, au foyer attractif $P_4 = (x = 0, y = -1, p = 0)$ où $D_v = T_v S$ et à l'intersection de l'équateur E avec la trajectoire spéciale (en rouge) T_{64} (qui vient du pôle Sud après être partie du foyer répulsif $P_3 = (x = 0, y = 1, p = 0)$ en suivant T_{35}). La figure représente également un morceau de la trajectoire spéciale (en rouge) T_{24} (qui vient, elle, du pôle Nord après être partie du foyer répulsif P_3 en suivant T_{31}) ainsi que (en noir) la trajectoire passant par P_0 . L'axe des z est celui de la variable p .

3. τ devient la symétrie A_p par rapport à l'axe des p : $x \rightarrow -x$, $y \rightarrow -y$, $p \rightarrow p$. Elle laisse fixes les pôles $(x = 0, y = 0, p = \pm 1)$ et échange les points $(x = \pm 1, y = 0, p = 0)$ ainsi que les points $(x = 0, y = \pm 1, p = 0)$. La transformation du champ est $D_x \rightarrow D_x$, $D_y \rightarrow D_y$, $D_p \rightarrow -D_p$ qui est la symétrie par rapport au plan (x, p) .

Ces transformations sont très visibles sur les figures 10 et 11. Leur loi de groupe est la même que celles des symétries de C dont elles proviennent.

2.9.4. Les cusps le long du contour apparent.

Nous avons noté que le cercle unité E du plan (x, y) est le contour apparent de la sphère S projetée verticalement sur ce plan. Le long de E les trajectoires de D sont verticales sauf aux foyers $(0, \pm 1, 0)$. Il est donc intéressant de regarder la projection des trajectoires de D . La figure 13 les représente et la figure 14 y adjoint la projection de D .

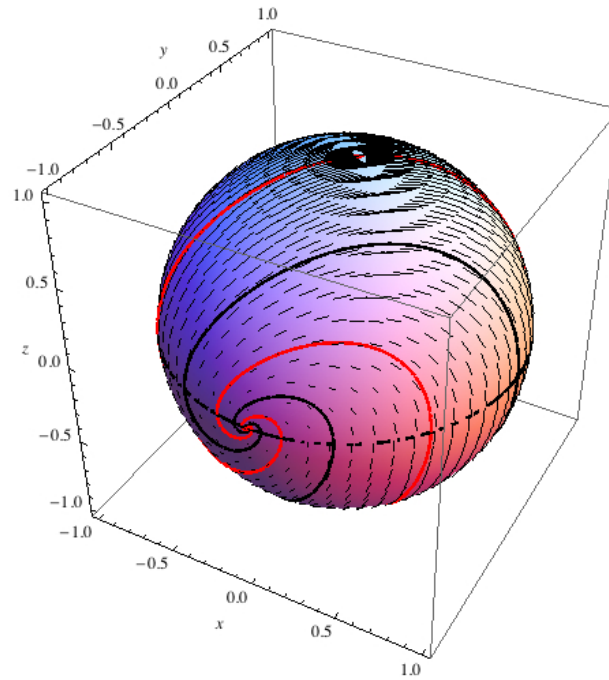


FIGURE 10. Les trajectoires de la figure 9 en tant qu'intégrales du champ de directions D . L'axe des z est celui de la variable p .

Sur la figure 14 on voit en chaque point (x, y) du plan les deux vecteurs D_v correspondant aux deux valeurs de $p = \pm\sqrt{1 - (x^2 + y^2)} = \pm \sin(\varphi)$. Leurs composantes sont $D_x = p = \pm \sin(\varphi)$ qui change de signe et $D_y = p^2 = \sin^2(\varphi)$ qui ne change pas de signe. Pour $p = 1$ (pôles) leur angle est de $\frac{\pi}{2}$. Pour $p = 0$ (i.e. sur le contour apparent E) leur angle s'aplatit en même temps que leur longueur s'annule (D_v se verticalise).

Ce qui est le plus frappant dans la figure 13 sont les *cusps* des trajectoires projetées sur le contour apparent. Ils ont pour tangente au point de rebroussement l'axe des x et s'inclinent donc de plus en plus sur E , jusqu'à y devenir tangents, au fur et à mesure que l'on passe des points $(x = \pm 1, y = 0)$ aux foyers $(x = 0, y = \pm 1)$. Au voisinage des foyers la structure est compliquée car on a une suite de cusps de plus en plus petits qui s'aplatissent sur leur tangente (parallèle à l'axe des x) en convergeant vers le foyer.

Pour bien comprendre ce qui se passe, considérons des $F(x, y, p)$ plus simples que la sphère, de type parabolôïde $p^2 = y$ ou $p^2 = x$.

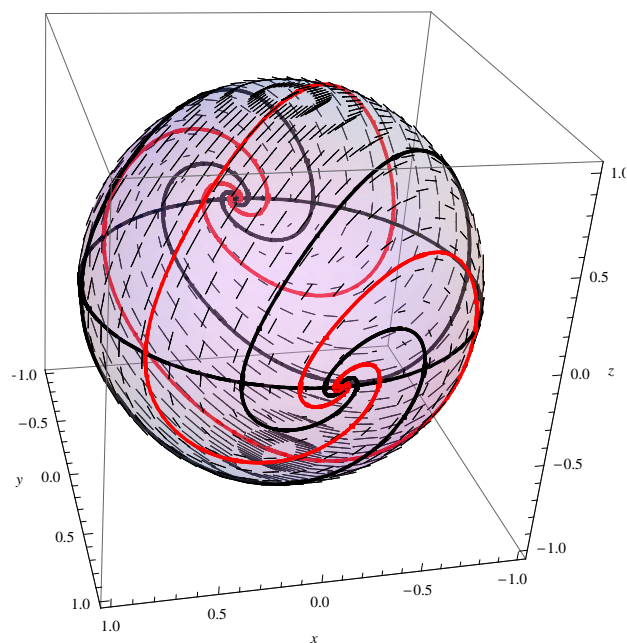


FIGURE 11. La figure 10 sur la sphère transparente et sous un autre point de vue. L'axe des z est celui de la variable p .

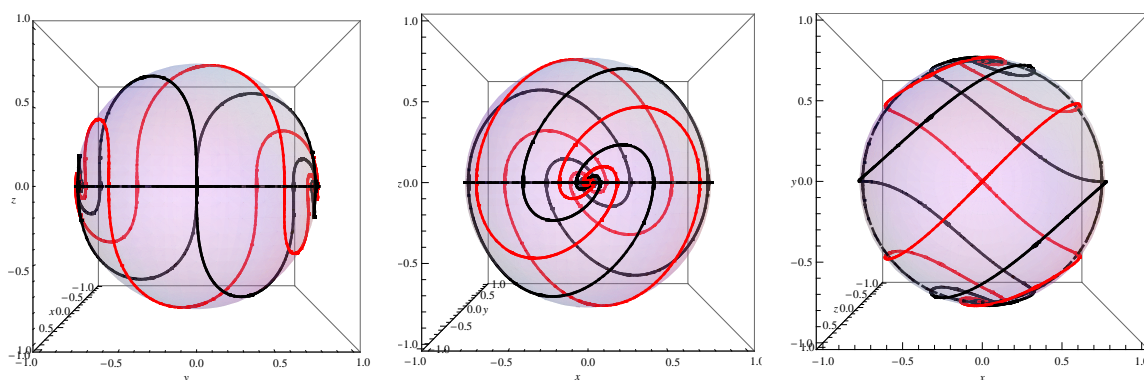
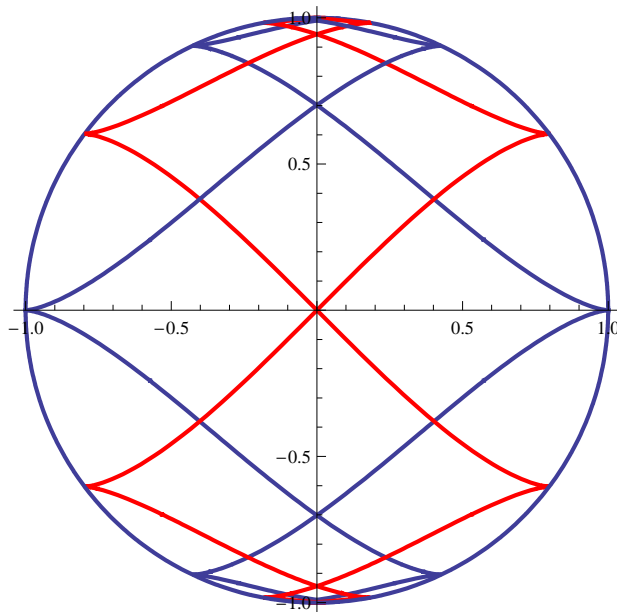


FIGURE 12. Trois points de vue différents sur les trajectoires des figures précédentes. L'axe des z est celui de la variable p .

- (i) Dans le premier cas $p^2 = y$ l'axe des x est une ligne pli, $2pdp = dy$ et donc $dy = pdx$ implique $dx = 2dp$, soit $x = 2p + x_0$. Les intégrales générales passant par l'axe des x en $(x = x_0, y = 0, p = 0)$ sont d'équations paramétriques $(x = 2p + x_0, y = p^2, p)$. Elles se projettent sur les paraboles

FIGURE 13. La projection des trajectoires de D sur le plan (x, y) .

$$(x = 2p + x_0, y = p^2)$$

tangentes à la ligne pli (l'axe des x) et cette dernière en est donc l'enveloppe. La figure 15 montre cette géométrie.

- (ii) Dans le deuxième cas $p^2 = x$ l'axe des y est une ligne pli, $2pdp = dx$ et donc $dy = pdx$ implique $dy = 2p^2dp$, soit $y = \frac{2}{3}p^3 + y_0$. Les intégrales générales passant par l'axe des y en $(x = 0, y = y_0, p = 0)$ sont d'équations paramétriques $(x = p^2, y = \frac{2}{3}p^3 + y_0, p)$. Elles se projettent sur les paraboles semi-cubiques $(x = p^2, y = \frac{2}{3}p^3 + y_0)$ qui présentent un cusp sur la ligne pli (l'axe des y). La figure 16 montre cette géométrie en changeant les signes ($p^2 = -x$) pour avoir une meilleure perspective dans l'affichage de l'image.

Notre exemple est une interpolation entre ces deux cas simples à travers le quart de cercle allant de points $(x = \pm 1, 0, 0)$ (cas (ii)) aux foyers $(x = 0, y = \pm 1, 0)$. Les foyers ne sont cependant pas des exemples du cas (i) car celui-ci n'est réalisé qu'en un point isolé et non pas sur un intervalle.

Les équations différentielles $\frac{dx}{dp} = -\frac{p}{x+py}$ et $\frac{dy}{dp} = -\frac{p^2}{x+py}$ donnant le champ caractéristique D étant beaucoup plus difficiles à intégrer, on peut essayer d'approximer la structure locale d'une trajectoire au voisinage du cusp d'un point pli

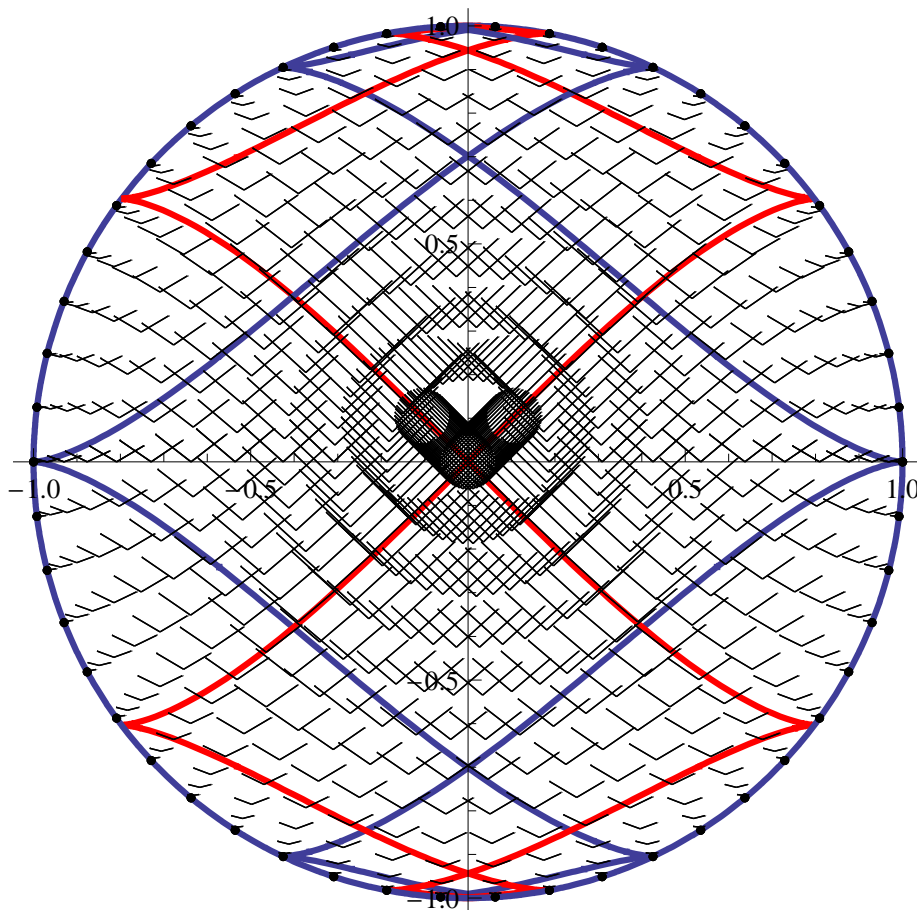


FIGURE 14. La projection de D et de ses trajectoires sur le plan (x, y) .

$(x_0, y_0, 0)$, $x_0^2 + y_0^2 = 1$. Soit $x = x_0 - \xi$, $y = y_0 - \eta$. On a

$$\begin{aligned} p^2 &= 1 - x^2 + y^2 = 1 - (x_0 - \xi)^2 + (y_0 - \eta)^2 \\ &= 1 - (x_0^2 + y_0^2) + 2x_0\xi + 2y_0\eta - \xi^2 - \eta^2 \\ &= 2x_0\xi + 2y_0\eta - \xi^2 - \eta^2 \\ &= f(\xi, \eta) . \end{aligned}$$

Si $x_0 \neq 0$, $\frac{\partial f}{\partial \xi}(x_0, y_0) = 2x_0 \neq 0$ et on peut prendre $\tilde{\xi} = \xi - \frac{1}{2x_0}\xi^2$ comme nouvelle variable ξ . On a $\tilde{\xi} = \frac{1}{2x_0}p^2 - \frac{y_0}{x_0}\eta - \frac{1}{2x_0}\eta^2$. Donc $d\tilde{\xi} = \frac{1}{x_0}pdp - \frac{y_0}{x_0}d\eta - \frac{1}{x_0}\eta d\eta$ et $d\eta = pd\xi$. On veut une approximation pour ξ, η petits du premier ordre. Alors, comme p est grand par rapport à ξ, η (comme $\sqrt{\varepsilon}$ est grand par rapport à ε si $\varepsilon \ll 1$), on a $\tilde{\xi} \sim \xi \sim \frac{1}{2x_0}p^2$ et $d\eta = pd\xi \sim \frac{1}{x_0}p^2$ et $\eta \sim \frac{1}{3x_0}p^3$.

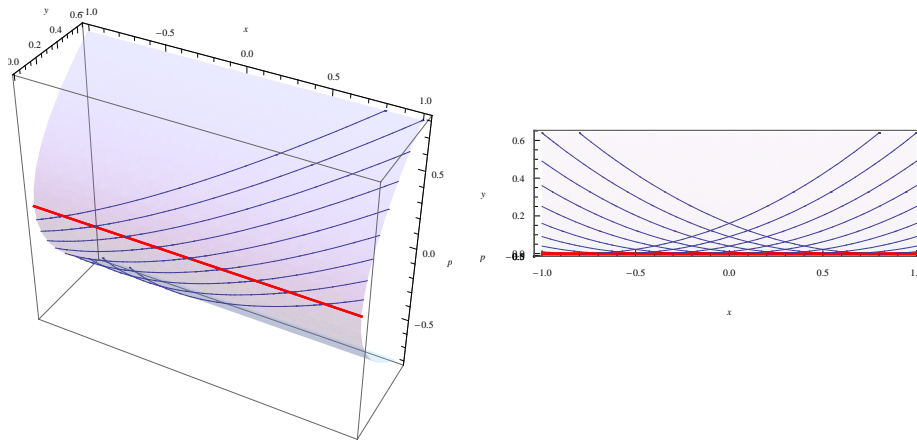


FIGURE 15. À gauche le parabolôide $y = p^2$ avec sa ligne pli en rouge et les trajectoires caractéristiques $(x = 2p + x_0, y = p^2, p)$. À droite leurs projections $(x = 2p + x_0, y = p^2)$. Ce sont des paraboles tangentes à la ligne pli qui est leur enveloppe. On a $p \in [-0.8, 0.8]$ et $x \in [-1, 1]$, x_0 variant de -0.8 à 0.8 par pas de 0.2 .

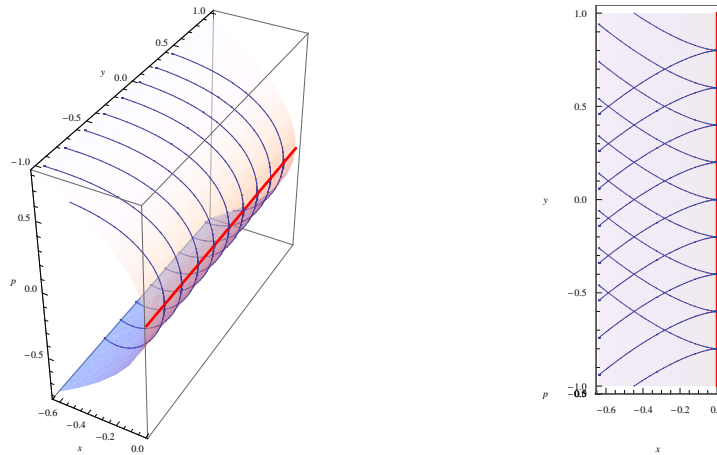


FIGURE 16. À gauche le parabolôide $-x = p^2$ avec sa ligne pli en rouge et les trajectoires caractéristiques $(x = -p^2, y = -\frac{2}{3}p^3 + y_0, p)$. À droite leurs projections $(x = -p^2, y = -\frac{2}{3}p^3 + y_0)$. Ce sont des paraboles semi-cubiques ayant un cusp sur la ligne pli. On a $p \in [-0.8, 0.8]$ et $y \in [-1, 1]$, y_0 variant de -0.8 à 0.8 par pas de 0.2 .

Bref, le cusp est approximé au premier ordre par le cusp $x = \frac{1}{2x_0}p^2$, $y = \frac{1}{3x_0}p^3$ d'équation semi-cubique $y^2 = \frac{8x_0}{9}x^3$. La figure 17 montre cette structure locale en $(x_0 = 1, y_0 = 0)$.

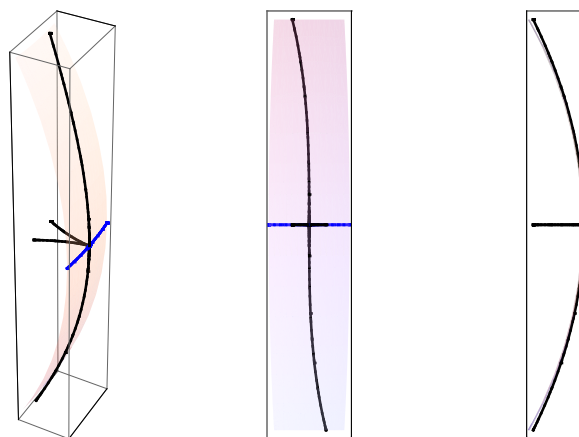


FIGURE 17. À gauche, la structure locale de la projection cuspidale d'une trajectoire de D au voisinage d'un point pli ($x_0 = 1, y_0 = 0, p = 0$) du contour apparent de S . Au milieu vue de face, à droite vue de profil.

2.10. Structure de contact et EDOs

La structure de contact permet d'approfondir considérablement la compréhension des équations différentielles ordinaires (EDOs) de type $F(x, y(x), y'(x)) = 0$ en les géométrisant. Classiquement, on cherchait des formules explicites pour les solutions $y(x)$ ou, si l'on n'en trouvait pas, au moins des méthodes numériques de calcul par approximations successives. Mais on peut aller beaucoup plus loin en passant dans l'espace des 1-jets. En effet, géométriquement, qu'est-ce qu'une solution $y(x)$ d'une telle EDO? Vue dans \mathbb{V}_J , c'est une courbe $(x, y(x), p(x))$ qui

- (i) est une relevée legendrienne ($p(x) = y'(x)$) et
- (ii) est tangente à la surface S d'équation $F = 0$.

Elle est donc essentiellement portée par les intersections des plans de contact et des plans tangents à S , c'est-à-dire par le champ caractéristique D . C'est par conséquent ce dernier qu'il faut comprendre en priorité avant que de chercher une expression explicite ou un algorithme de calcul pour $y(x)$.

Dans un article pionnier de 1972 [531] consacré aux intégrales singulières des équations différentielles *multiformes*, René Thom a expliqué comment il fallait revoir la notion classique d'*intégrale singulière* d'une EDO.¹⁶ Classiquement, les intégrales singulières étaient pensées comme les *enveloppes* des intégrales générales. Mais Thom insiste sur le fait qu'il faut raisonner géométriquement, se placer dans l'espace des 1-jets muni de sa forme de Pfaff $dy = p dx$ et considérer les relevées legendriennes des intégrales. C'est le contour apparent de la surface S d'équation $F(x, y, p) = 0$

16. Article republié dans le Volume III des *Œuvres Mathématiques* (2021) avec un commentaire détaillé de Marc Chaperon [107].

qui est déterminant. Il est obtenu en éliminant p des équations $F = 0$ et $\frac{\partial F}{\partial p} = 0$. Génériquement (théorème de Whitney), il ne comprend comme singularités que des projections de lignes de points plis, des cusps projections de points fronces isolés (où une ligne pli possède une tangente verticale) et des croisements transverses isolés de projections de lignes plis.

Thom explique en détail comment on calcule localement en utilisant le théorème des fonctions implicites et le théorème de préparation de Weierstrass-Malgrange permettant d'écrire $F = 0$ au voisinage d'un point pli sous la forme $p^2 + a_1(x, y)p + a_0(x, y) = 0$. Nous aurons l'occasion de revenir dans les prochains chapitres sur les formes normales des singularités. Le théorème de Whitney affirme que, au voisinage d'un point pli, l'équation $F = 0$ peut en fait se ramener à la forme normale $p^2 = f(x, y)$. Si $\frac{\partial f}{\partial x} \neq 0$ le long d'une ligne pli on est dans le cas que nous avons exemplifié plus haut et le contour apparent est une intégrale singulière constituée de cusps, c'est-à-dire un lieu de *points de rebroussement* des intégrales générales. Cela est dû au fait que la valeur de x déduite de $p^2 = f(x, y)$ par le théorème des fonctions implicites étant en p^2 et que, dy étant égal à $p dx$, y est en p^3 . Si en revanche $\frac{\partial f}{\partial x} \equiv 0$ le long d'une ligne pli (ce qui est une situation non générique), c'est y qui est en p^2 et donc x est en p . Les projections des courbes intégrales générales sont alors tangentes au contour apparent et celui-ci est une enveloppe. Thom étudie aussi la situation où $\frac{\partial f}{\partial x} \neq 0$ le long de la ligne pli mais s'annule en un point (x_0, y_0) . On a alors deux suites de cusps qui fusionnent en ce point. C'est le cas des foyers ($x_0 = 0, y_0 = \pm 1$) de l'exemple précédent.

Thom analyse aussi les points fronces du contour apparent où les intégrales générales se projettent sur des queues d'aronde dont les deux cusps parcourent les deux branches du cusp déployant le point fronce. Nous ne traiterons pas ce point ici. Nous allons plutôt approfondir la notion de structure de contact.

3. Champs de Reeb, structures symplectiques, métriques et complexes

En fait, la structure de contact de $J^1\mathbb{R}^2$ appartient à l'univers mathématique de ce que l'on appelle la *géométrie symplectique*. D'ailleurs, il existe toujours une "symplectisation" naturelle de la structure de contact.

3.1. Structures symplectiques

L'exemple prototypique de 2-forme symplectique est la forme symplectique canonique des *fibrés cotangents* T^*M d'une variété différentiable M de dimension n . Localement, T^*M est identifiable à $T^*\mathbb{R}^n$ avec des coordonnées x_i et p_i ($i = 1, \dots, n$) qui sont des variables conjuguées.

On considère la 1-forme $\omega = \sum_i p_i dx_i$. Sa dérivée est la 2-forme symplectique $\Omega = d\omega = \sum_i dp_i \wedge dx_i$ dont la puissance Ω^n est essentiellement la forme volume (au facteur $n!$ près). Cette 2-forme possède les deux propriétés définitionnelles d'une forme symplectique :

1. elle est *fermée* puisqu'elle est exacte $d\Omega = d^2\omega = 0$;
2. elle est *non dégénérée* au sens où si $\Omega(X, X') = 0$ pour $X = \sum_j \xi_j \partial_{x_j} + \sum_k \pi_k \partial_{p_k} \in T(T^*M)$ donné et $X' = \sum_i \xi'_i \partial_{x_i} + \sum_j \pi'_j \partial_{p_j} \in T(T^*M)$ quelconque, alors $X = 0$. En effet chaque terme $dp_i \wedge dx_i$ de Ω doit donner $\pi_i \xi'_i - \pi'_i \xi_i = 0$ pour tout ξ'_i, π'_i et donc $\xi_i = \pi_i = 0$.

Les structures symplectiques généralisent ce cas de base. Par définition, elles permettent de mesurer des *aires* dans les espaces tangents, ni des longueurs comme le permettent des normes, ni des angles comme le permettent des structures conformes.

3.2. Symplectisation de la structure de contact

En termes des coordonnées locales $(x, y, p; \xi, \eta, \pi)$, K_v est le plan d'équation $\eta = p\xi$. C'est, nous l'avons vu, le champ des noyaux de la 1-forme $\omega_J = -pdx + dy$. Mais la définition de la structure de contact au moyen de la 1-forme ω_J est *sous-déterminée*. En effet, ω_J contient plus d'information que son noyau (*i.e.* la distribution \mathcal{K} des plans de contact K_v) car les 1-formes ω_J et $f\omega_J$, où f est une fonction scalaire partout $\neq 0$, ont le même noyau et définissent donc la même distribution de plans. L'information supplémentaire encodée dans ω_J est la valeur numérique de ω_J le long du vecteur tangent "caractéristique" $\chi = -\mathbf{t}_3$ qui est transverse à K_v et de valeur $\omega_J(\chi) = 1$ (champ de Reeb).

On peut donc considérer le fibré $\pi_\Sigma : \Sigma \rightarrow J^1\mathbb{R}^2$ de base $J^1\mathbb{R}^2$, de fibre le groupe multiplicatif $\mathbb{R}^* = \mathbb{R} - \{0\}$ des valeurs des f et d'espace total Σ . On définit alors une 1-forme canonique Ω_J sur Σ de la façon suivante. Soient $s = (x, y, p, r) = (v, r)$ un point de Σ et $\sigma = (\xi, \eta, \pi, \rho) = (\mathbf{t}, \rho) \in T_s\Sigma$ un vecteur tangent en s à Σ . L'image directe $(\pi_\Sigma)_* \sigma$ de σ par la projection π_Σ est un vecteur tangent $\mathbf{t} \in T_v J^1\mathbb{R}^2$ à $J^1\mathbb{R}^2$ en $v = (x, y, p) = \pi_\Sigma(s)$. On pose alors $\Omega_J(\sigma) = r\omega_J(\mathbf{t})$. On a :

$$\begin{aligned} d\Omega_J &= rd\omega_J + dr \wedge \omega_J \\ &= rdx \wedge dp + dr \wedge (-pdx + dy) \\ &= rdx \wedge dp + pdx \wedge dr + dr \wedge dy . \end{aligned}$$

Cette 2-forme est fermée et non dégénérée. C'est par conséquent une forme symplectique sur Σ . Elle est fermée puisqu'elle est même exacte. Elle est non dégénérée car si $d\Omega_J(\sigma, \sigma') = 0$ pour tout σ' alors $\sigma = 0$. Soient en effet $\sigma = (\xi, \eta, \pi, \rho)$ et

$\sigma' = (\xi', \eta', \pi', \rho')$. On a :

$$\begin{aligned} d\Omega_J(\sigma, \sigma') &= r(\xi\pi' - \xi'\pi) + p(\xi\rho' - \xi'\rho) + \rho\eta' - \rho'\eta \\ &= -(r\pi + p\rho)\xi' + \rho\eta' + r\xi\pi' + (p\xi - \eta)\rho' . \end{aligned}$$

L'annulation des coefficients de ξ', η', π', ρ' implique alors, puisque l'on a $r \neq 0$ par hypothèse, la nullité de ξ, η, π, ρ . Donc $\sigma = 0$.

3.3. Formes $f\omega$, champs de Reeb et métriques

Lorsque l'on multiplie une forme de contact ω par une fonction différentiable $f \neq 0$, on garde son champ de plans de contact \mathcal{K} mais on change son champ de Reeb χ . Cela permet entre autres de changer la géométrie métrique. Pour notre 1-forme $\omega_J = dy - p dx$, le champ de Reeb χ transverse à \mathcal{K} et tel que $\omega_J(\chi) = 1$ est $-\mathbf{t}_3 = \partial_y$. En choisissant de rendre la base $\{\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3\}$ orthonormée, on choisit une *métrique* qui n'est pas la métrique euclidienne standard. Mais on peut se demander si on peut avec une $f \neq 0$ bien choisie s'arranger pour que $f\omega_J = \omega$ ait pour champ de Reeb un vecteur \mathbf{n} normal à \mathcal{K} pour la métrique euclidienne. Comme les vecteurs de \mathcal{K} sont de la forme $\mathbf{t} = (\xi, \eta = p\xi, \pi)$, \mathbf{n} devrait être de la forme $\mathbf{n} = (-p\alpha, \alpha, 0)$ et satisfaire les conditions $\omega(\mathbf{n}) = 1$ et $d\omega(\mathbf{n}, \mathbf{t}) = 0$ pour tout $\mathbf{t} \in \mathcal{K}$. On a $\omega(\mathbf{n}) = f\alpha(1 + p^2)$ pour la première condition. Pour la seconde on utilise

$$\begin{aligned} d\omega &= d(f\omega_J) = df \wedge \omega_J + f d\omega_J \\ &= (\partial_x f dx + \partial_y f dy + \partial_p f dp) \wedge (dy - p dx) + f dx \wedge dp \\ &= \partial_x f dx \wedge dy + \partial_p f dp \wedge dy - p \partial_y f dy \wedge dx - p \partial_p f dp \wedge dx + f dx \wedge dp \\ &= (\partial_x f + p \partial_y f) dx \wedge dy - \partial_p f dy \wedge dp - (p \partial_p f + f) dp \wedge dx . \end{aligned}$$

Donc

$$\begin{aligned} d\omega(\mathbf{n}, \mathbf{t}) &= (\partial_x f + p \partial_y f) (\mathbf{n}_x \mathbf{t}_y - \mathbf{n}_y \mathbf{t}_x) - \partial_p f (\mathbf{n}_y \mathbf{t}_p - \mathbf{n}_p \mathbf{t}_y) - (p \partial_p f + f) (\mathbf{n}_p \mathbf{t}_x - \mathbf{n}_x \mathbf{t}_p) \\ &= (\partial_x f + p \partial_y f) (-p\alpha p\xi - \alpha\xi) - \partial_p f (\alpha\pi - 0) - (p \partial_p f + f) (0 + p\alpha\pi) \\ &= -\alpha\xi (\partial_x f + p \partial_y f) (p^2 + 1) - \alpha\pi \partial_p f - \alpha\pi p (p \partial_p f + f) \\ &= 0 . \end{aligned}$$

La condition $d\omega(\mathbf{n}, \mathbf{t}) = 0$ pour tout $\mathbf{t} \in \mathcal{K}$ devient par conséquent la double condition sur f

$$\begin{aligned} \partial_x f + p \partial_y f &= 0 \\ \partial_p f + p (p \partial_p f + f) &= 0 . \end{aligned}$$

Le plus simple est de prendre une fonction $f(p)$ ne dépendant que de p . La première condition est alors automatiquement satisfaite et la seconde devient $(1 + p^2) f' +$

$pf = 0$ dont la solution est $f = \frac{c}{\sqrt{1+p^2}}$. On a alors

$$\omega(\mathbf{n}) = f\alpha(1+p^2) = \alpha c\sqrt{1+p^2} = 1.$$

Si l'on veut que \mathbf{n} soit unitaire pour la norme euclidienne, i.e. $\sqrt{p^2\alpha^2 + \alpha^2} = 1$ et $\alpha = \frac{1}{\sqrt{1+p^2}}$, alors il faut choisir $c = 1$ et donc $\omega = \frac{1}{\sqrt{1+p^2}}\omega_J$.

Remarque. On notera que le long d'une relevée legendrienne on a $p = \tan(\theta)$ et $\frac{1}{\sqrt{1+p^2}} = \cos(\theta)$ (pour $\cos(\theta) \geq 0$) et que par conséquent $f\omega_J = \cos(\theta)\omega_J$ est la 1-forme $\omega_S = \cos(\theta)dy - \sin(\theta)dx$ du second modèle \mathbb{V}_S évoqué plus haut. Le champ de Reeb est alors $\mathbf{n} = (-p\cos(\theta), \cos(\theta), 0) = (-\sin(\theta), \cos(\theta), 0)$. Nous y reviendrons plus bas, entre autres à la section 7 du chapitre 5. \square

3.4. Structures (presque) complexes

3.4.1. Conditions générales.

Il existe des liens très étroits entre les structures symplectiques, métriques et complexes sur une variété M de dimension $2n$. Soit Ω la 2-forme symplectique, c'est-à-dire, dans notre cas, $d\omega$. En prenant une base symplectique comme base orthonormée on définit une métrique riemannienne g . On a alors $d\omega(X, Y) = g(X, J(Y))$ pour un endomorphisme (unique) J des espaces tangents satisfaisant $J^2 = -I$. On a donc $d\omega(X, Y) = 0$, i.e. le plan $\text{Span}\{X, Y\}$ est "isotrope", si et seulement si $J(X)$ et Y sont g -orthogonaux. Mais les espaces tangents étant de dimension paire $2n$, ils sont isomorphes à \mathbb{C}^n et on peut choisir l'isomorphisme de façon à ce que J soit la multiplication par i (la rotation de $\frac{\pi}{2}$). On peut alors définir un produit scalaire hermitien par $g(X, Y) + id\omega(X, Y)$ dont la partie réelle (symétrique) donne la métrique et la partie imaginaire (antisymétrique) donne la 2-forme symplectique. On a ainsi trois groupes de structures géométriques qui interviennent : le groupe orthogonal $O(2n)$ des isométries de \mathbb{R}^{2n} , le groupe $GL(n, \mathbb{C})$ des automorphismes \mathbb{C} -linéaires de \mathbb{C}^n , le groupe $Sp(2n)$ des automorphismes symplectiques de \mathbb{R}^{2n} . Leur intersection est le groupe unitaire $U(n)$ de \mathbb{C}^n .

Nous reviendrons sur cette trinité géométrique à la section 2.2.2 du chapitre 15.

Revenons en attendant à ce que nous venons de voir pour \mathbb{V}_J et tentons de le généraliser. Une structure de contact \mathcal{K} définie par une 1-forme de contact ω sur une variété M de dimension 3 se trouve complétée par le champ de Reeb χ qui permet de lever la sous-détermination de ω . Nous avons vu comment symplectiser (\mathcal{K}, χ) . Étant donné le lien étroit entre structures symplectiques, métriques et structures complexes, on peut se demander dans quelle mesure on peut considérer M comme une sous-variété de dimension 3 d'une variété Σ de dimension 4 munie d'une structure complexe $\simeq \mathbb{C}^2$ sur ses espaces tangents $T_{x,r}$ pour laquelle les plans K_x seraient des droites complexes (de dimension réelle 2) de $T_{x,r}$. Pour ce faire, il

faut définir la multiplication par i (intuitivement la rotation de $\frac{\pi}{2}$) dans les plans de contact K_x , i.e. un endomorphisme linéaire J de K_x satisfaisant la relation caractéristique $J^2 = -1$. On peut en plus vouloir introduire une métrique rendant \mathcal{K} et χ orthogonaux comme l'explorent Shiing-Shen Chern et Richard Hamilton dans [115]. Cela est effectivement possible : il existe une métrique riemannienne g et un endomorphisme J de $T_x M$ satisfaisant les conditions (redondantes) suivantes (cf. par exemple le survey [117] de Jong Taek Cho).

1. $\omega(X) = g(X, \chi)$ pour tout $X \in T_x M$,¹⁷ ce qui assure que χ soit g -orthogonal à \mathcal{K} . $\omega(X)$ est alors la projection orthogonale de X sur $\mathbb{R}\chi$. Cette condition se ramène à $g(\chi, \chi) = 1$ car si $X = X_{\mathcal{K}} + X_{\chi}$ avec $X_{\chi} = x_{\chi}\chi$, alors $\omega(X) = \omega(x_{\chi}\chi) = x_{\chi}$ et donc $X = X_{\mathcal{K}} + \omega(X)\chi$. Donc

$$\begin{aligned} g(X, \chi) &= g(X_{\mathcal{K}} + \omega(X)\chi, \chi) = g(X_{\mathcal{K}}, \chi) + g(\omega(X)\chi, \chi) \\ &= \omega(X)g(\chi, \chi) = \omega(X) . \end{aligned}$$

2. $J(\chi) = 0$, autrement dit on prolonge J (défini initialement sur \mathcal{K}) par 0 sur $\chi \perp_g \mathcal{K}$.¹⁸ Cela signifie que pour obtenir $J(X)$ on projette d'abord X en $X_{\mathcal{K}}$ sur \mathcal{K} parallèlement à χ puis on applique J à $X_{\mathcal{K}}$.
3. $\omega \circ J = 0$. Cela est nécessaire car, si $X = X_{\mathcal{K}} + X_{\chi}$, $J(X) = J(X_{\mathcal{K}}) \in \mathcal{K}$ et donc $\omega(J(X_{\mathcal{K}})) = 0$.
4. $d\omega(X, J(Y)) = -d\omega(J(X), Y)$. La 2-forme $d\omega$ est ce que l'on appelle dans ce contexte la *forme de Levi*.
5. $J^2(X) = -X + \omega(X)\chi$, ce qui assure que $J^2(X) = -X$ si $X \in \mathcal{K}$. Si X est quelconque, $X = X_{\mathcal{K}} + X_{\chi}$ avec $X_{\chi} = x_{\chi}\chi$, alors $\omega(X) = \omega(x_{\chi}\chi) = x_{\chi}$ et donc $X = X_{\mathcal{K}} + \omega(X)\chi$. Par conséquent $J^2(X) = J^2(X_{\mathcal{K}}) = -X_{\mathcal{K}} = -X + \omega(X)\chi$.
6. $g(J(X), J(Y)) = g(X, Y) - \omega(X)\omega(Y)$, ce qui assure que g soit J -invariante sur \mathcal{K} puisque $\omega(X)\omega(Y) = 0$ sur \mathcal{K} . Autrement dit, sur \mathcal{K} , J est une isométrie. Cette dernière condition suffit car on a alors

$$g(J(X_{\mathcal{K}}), J(Y_{\mathcal{K}})) = g(X_{\mathcal{K}}, Y_{\mathcal{K}})$$

et comme $X_{\mathcal{K}} = X - \omega(X)\chi$ (cf. la condition (1))

$$\begin{aligned} g(J(X), J(Y)) &= g(X_{\mathcal{K}}, Y_{\mathcal{K}}) = g(X - \omega(X)\chi, Y - \omega(Y)\chi) \\ &= g(X, Y) - \omega(X)g(\chi, Y) - \omega(Y)g(\chi, X) + \omega(X)\omega(Y)g(\chi, \chi) \end{aligned}$$

et comme d'après (1) $g(\chi, Y) = \omega(Y)$ et $g(\chi, X) = \omega(X)$, on obtient $g(X, Y) - \omega(X)\omega(Y)$.

17. Lorsqu'un point x est sélectionné, g signifie la métrique g_x sur $T_x M$, χ signifie la valeur $\chi(x)$ du champ χ en x , etc.

18. \perp_g dénote l'orthogonalité relative à g . Par construction, χ est seulement transverse à \mathcal{K} mais la condition (1) transforme la transversalité en g -orthogonalité.

7. $d\omega(X, Y) = -d\omega(Y, X) = \alpha g(X, J(Y))$.¹⁹ Cette condition de base (que nous avons formulée d'emblée plus haut) assure la compatibilité entre $d\omega$ et g sur \mathcal{K} et garantit que $X \perp J(X)$ puisque $\alpha g(X, J(X)) = d\omega(X, X) = 0$.²⁰

Dans ce contexte, on peut introduire une structure presque complexe \tilde{J} sur $M \times \mathbb{R}$ avec \mathbb{R} de coordonnée r , structure associée à la symplectisation de \mathcal{K} . Pour cela on pose $\tilde{J}(X, f\partial_r) = (J(X) - f\chi, \omega(X)\partial_r)$ et on vérifie que $\tilde{J}^2 = -Id$. En effet

$$\begin{aligned} \tilde{J}^2(X, f\partial_r) &= J(J(X) - f\chi, \omega(X)\partial_r) \\ &= (J(J(X) - f\chi) - \omega(X)\chi, \omega(J(X) - f\chi)\partial_r) \\ &= (J^2(X) - 0 - \omega(X)\chi, -f\omega(\chi)\partial_r) \\ &= (-X + \omega(X)\chi - \omega(X)\chi, -f\partial_r) = (-X, -f\partial_r) \\ &= -(X, f\partial_r) . \end{aligned}$$

Remarque. Les spécialistes de géométrie complexe²¹ ont aussi étudié la construction suivante.²² Ils utilisent J pour introduire des vecteurs tangents à M qui sont des vecteurs complexes, c'est-à-dire des vecteurs appartenant au complexifié $TM \otimes_{\mathbb{R}} \mathbb{C}$ de TM , et ils définissent le fibré en droites complexes $\mathcal{K}_{\mathbb{C}}$ par

$$\mathcal{K}_{\mathbb{C}} = \{X - iJ(X) \mid X \in \mathcal{K}\} .²³$$

Autrement dit, ils traitent $J(X)$ non plus comme la rotation de $\frac{\pi}{2}$ de X dans \mathcal{K} mais comme une composante dans $i\mathcal{K}$. On a $\mathcal{K}_{\mathbb{C}} \cap \overline{\mathcal{K}_{\mathbb{C}}} = \{0\}$ ²⁴ et $\mathcal{K}_{\mathbb{C}} \oplus \overline{\mathcal{K}_{\mathbb{C}}} = \mathbb{C}\mathcal{K}$ puisque, pour tout $Z \in \mathcal{K}$, $(a + ib)Z$ s'écrit de façon unique sous la forme $(a + ib)Z = X - iJ(X) + Y + iJ(Y)$ avec $X + Y = aZ$ et $-J(X) + J(Y) = bZ$. Qui plus est, alors que \mathcal{K} est par définition non intégrable puisque c'est une structure de contact, $\mathcal{K}_{\mathbb{C}}$ est en revanche *intégrable* au sens où $[\mathcal{K}_{\mathbb{C}}, \mathcal{K}_{\mathbb{C}}] \subset \mathcal{K}_{\mathbb{C}}$. En effet, si l'on applique la formule (D) de la section 7.2.7 du chapitre 2

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]) ,$$

on voit que, pour $X, Y \in \mathcal{K}$, $\omega(X) \equiv \omega(Y) \equiv 0$ implique

$$d\omega(X, Y) = -\omega([X, Y]) \text{ et } d\omega(J(X), J(Y)) = -\omega([J(X), J(Y)]) .$$

19. Le facteur α est 1 ou 2 suivant les auteurs.

20. Ces sept conditions ne sont pas indépendantes. Il est facile de vérifier que (1), (5) et (7) entraînent les quatre autres.

21. i.e. dont le corps des scalaires est \mathbb{C} et non plus \mathbb{R} .

22. Cf. par exemple Cho [117] ainsi que le papier [392] de Pawel Nurowski et Jacek Tafel.

23. En tout point x , $(\mathcal{K}_{\mathbb{C}})_x$ est une droite complexe car si $X - iJ(X)$ est un \mathbb{C} -vecteur $\neq 0$ de référence dans $(\mathcal{K}_{\mathbb{C}})_x$ et si $Y - iJ(Y)$ est un autre \mathbb{C} -vecteur de $(\mathcal{K}_{\mathbb{C}})_x$, on a $Y - iJ(Y) = (a + ib)(X - iJ(X))$ avec $Y = aX + bJ(X)$ et $J(Y) = -bX + aJ(X)$. Autrement dit a et b sont les composantes de Y dans la base $\{X, J(X)\}$ de \mathcal{K}_x au point considéré.

24. Puisque $X - iJ(X) = Y + iJ(Y)$ implique à la fois $Y = X$ et $Y = -X$.

Mais d'après (4) $d\omega(X, Y) = d\omega(J(X), J(Y))$ et donc

$$\omega([X, Y]) = \omega([J(X), J(Y)]),$$

ce qui signifie que $[J(X), J(Y)] - [X, Y] \in \mathcal{K}$. De même, toujours d'après (4), $d\omega(X, J(Y)) = -d\omega(J(X), Y)$, donc $-\omega([X, J(Y)]) = \omega([J(X), Y])$, soit

$$[J(X), Y] + [X, J(Y)] \in \mathcal{K}.$$

Considérons alors

$$[X - iJ(X), Y - iJ(Y)] \in [\mathcal{K}_{\mathbb{C}}, \mathcal{K}_{\mathbb{C}}], \quad X, Y \in \mathcal{K}.$$

On montre que $[X, J(Y)] + [J(X), Y] = J([X, Y] - [J(X), J(Y)])$, ce qui implique

$$\begin{aligned} [X - iJ(X), Y - iJ(Y)] &= [X, Y] - i[X, J(Y)] - i[J(X), Y] - [J(X), J(Y)] \\ &= ([X, Y] - [J(X), J(Y)]) - i([X, J(Y)] + [J(X), Y]) \\ &= ([X, Y] - [J(X), J(Y)]) - iJ([X, Y] - [J(X), J(Y)]) \\ &\in \mathcal{K}_{\mathbb{C}}. \end{aligned}$$

Cette structure définit ce que l'on appelle une variété *Cauchy-Riemann* (CR). —□

3.4.2. L'exemple de \mathbb{V}_J .

Appliquons ce qui précède à notre modèle \mathbb{V}_J (de point courant $v = (x, y, p)$) défini par $\omega = dy - p dx$,²⁵ le crochet de Lie

$$[\mathbf{t}, \mathbf{t}'] = [(\xi, \eta, \pi), (\xi', \eta', \pi')] = (0, \pi\xi' - \pi'\xi, 0),$$

la base

$$\mathbf{t}_1 = \frac{\partial}{\partial x} + p \frac{\partial}{\partial y} = (1, p, 0), \quad \mathbf{t}_2 = \frac{\partial}{\partial p} = (0, 0, 1)$$

de $\mathcal{K} = \ker(\omega)$ augmentée de

$$\mathbf{t}_3 = -\frac{\partial}{\partial y} = (0, -1, 0) = [\mathbf{t}_1, \mathbf{t}_2],$$

et les commutateurs

$$[\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3, [\mathbf{t}_1, \mathbf{t}_3] = 0, [\mathbf{t}_2, \mathbf{t}_3] = 0.$$

Sa structure ne dépend que de p , i.e. est invariante par translations dans le plan base $\text{Span}\{x, y\}$. Dans la base $\{\partial_x, \partial_y, \partial_p\}$ de $T_v\mathbb{V}_J$, si $\mathbf{t} = \xi\partial_x + \eta\partial_y + \pi\partial_p \in T_v\mathbb{V}_J$, on a $\omega(\mathbf{t}) = -p\xi + \eta$. Si $\mathbf{t} = \tilde{\xi}\tilde{\mathbf{t}}_1 + \tilde{\eta}\tilde{\mathbf{t}}_2 + \tilde{\pi}\tilde{\mathbf{t}}_3$ dans la base $\{\tilde{\mathbf{t}}_1, \tilde{\mathbf{t}}_2, \tilde{\mathbf{t}}_3\}$, alors $\xi = \tilde{\xi}$, $\eta = p\tilde{\xi} - \tilde{\pi}$, $\pi = \tilde{\eta}$ et $\omega(\mathbf{t}) = -p\tilde{\xi} + p\tilde{\xi} - \tilde{\pi} = -\tilde{\pi}$. On a $d\omega = dx \wedge dp$, $\omega \wedge d\omega = -\text{Vol}$ et le champ de Reeb pour cette forme de contact ω de \mathcal{K} est $\chi = -\mathbf{t}_3 = (0, 1, 0)$. La décomposition de $\mathbf{t} = (\xi, \eta, \pi) \in T_v\mathbb{V}_J$ en somme $\mathbf{t} = \mathbf{t}_{\mathcal{K}} + \omega(\mathbf{t})\chi$ s'écrit

$$\mathbf{t} = (\xi, p\xi, \pi) + (0, -p\xi + \eta, 0).$$

25. Nous notons la forme de contact ω plutôt que ω_J pour éviter toute confusion de notation avec l'opérateur J .

Calculons dans la base $\{\partial_x, \partial_y, \partial_p\}$ la matrice symétrique de la métrique g et celle de la multiplication complexe J .

La condition (1) $\omega(\mathbf{t}) = g(\mathbf{t}, \chi)$ implique $-p\xi + \eta = g_{12}\xi + g_{22}\eta + g_{23}\pi$ pour tout \mathbf{t} et donc $g_{12} = -p$, $g_{22} = 1$, $g_{23} = 0$.

La condition (2) $J(\chi) = 0$ implique $J_{12} = J_{22} = J_{32} = 0$.

La condition (3) $\omega \circ J = 0$ implique

$$-p(J_{11}\xi + J_{12}\eta + J_{13}\pi) + (J_{21}\xi + J_{22}\eta + J_{23}\pi) = 0$$

pour tout \mathbf{t} et donc $-pJ_{11} + J_{21} = -pJ_{12} + J_{22} = -pJ_{13} + J_{23} = 0$, la seconde égalité étant déjà satisfaite suite à (2).

La condition (4) $d\omega(\mathbf{t}, J(\mathbf{t}')) = -d\omega(J(\mathbf{t}), \mathbf{t}')$ s'écrit

$$\xi(J_{3i}\mathbf{t}'_i) - \pi(J_{1i}\mathbf{t}'_i) = \xi'(J_{3i}\mathbf{t}_i) - \pi'(J_{1i}\mathbf{t}_i)$$

et implique $J_{33} = -J_{11}$, $J_{32} = J_{12} = 0$, les deux dernières relations étant déjà satisfaites.

La condition (5) $J^2(\mathbf{t}) = -\mathbf{t} + \omega(\mathbf{t})\chi$ implique quant à elle $A = J_{11}^2 + J_{13}J_{31} = -1$. En effet $J^2(\mathbf{t}) = (A\xi, pA\xi, A\pi)$ et $-\mathbf{t} + \omega(\mathbf{t})\chi = (-\xi, -p\xi, -\pi)$ et donc $A = -1$.

Les conditions (6) et (7) sont un peu plus sophistiquées. (7), à savoir $d\omega(\mathbf{t}, \mathbf{t}') = \xi\pi' - \pi\xi' = \alpha g(\mathbf{t}, J(\mathbf{t}'))$ donne le système linéaire

$$\begin{cases} J_{11}g_{11} - p^2J_{11} + J_{31}g_{13} = 0 \\ J_{11}g_{13} + J_{31}g_{33} = -\frac{1}{\alpha} \\ J_{13}g_{11} - p^2J_{13} - J_{11}g_{13} = \frac{1}{\alpha} \\ J_{13}g_{13} - J_{11}g_{33} = 0 \end{cases}$$

dont les solutions sont $g_{11} = \frac{J_{31}}{\alpha A} + p^2$, $g_{13} = -\frac{J_{11}}{\alpha A}$, $g_{33} = -\frac{J_{13}}{\alpha A}$ avec, nous l'avons vu, $A = J_{11}^2 + J_{13}J_{31} = -1$.

On vérifie alors que la condition (6) $g(J(X), J(Y)) = g(X, Y) - \omega(X)\omega(Y)$ est bien satisfaite par ces valeurs et donne des deux côtés les coefficients

$\eta\xi'$	$\eta\eta'$	$\eta\pi'$	$\pi\xi'$	$\pi\eta'$	$\pi\pi'$	$\xi\xi'$	$\xi\eta'$	$\xi\pi'$
0	0	0	$\frac{J_{11}}{\alpha}$	0	$\frac{J_{13}}{\alpha}$	$\frac{1+J_{11}^2}{\alpha J_{13}}$	0	$\frac{J_{11}}{\alpha}$

En définitive, pour \mathbb{V}_J et $\omega = \omega_J$, J possède deux degrés de liberté puisqu'il dépend de J_{11} , J_{13} et J_{31} reliés par la relation $J_{11}^2 + J_{13}J_{31} = -1$, les autres coefficients de J étant donnés par $J_{12} = J_{22} = J_{32} = 0$, $J_{21} = pJ_{11}$, $J_{23} = pJ_{13}$, $J_{33} = -J_{11}$. Quant à la métrique g , elle satisfait $g_{12} = -p$, $g_{22} = 1$, $g_{32} = 0$, ainsi que $g_{11} = -\frac{J_{31}}{\alpha} + p^2$, $g_{13} = \frac{J_{11}}{\alpha}$, $g_{33} = \frac{J_{13}}{\alpha}$. Bref, les matrices sont

$$J = \begin{pmatrix} J_{11} & 0 & J_{13} \\ pJ_{11} & 0 & pJ_{13} \\ J_{31} & 0 & -J_{11} \end{pmatrix}, \quad g = \begin{pmatrix} -\frac{J_{31}}{\alpha} + p^2 & -p & \frac{J_{11}}{\alpha} \\ -p & 1 & 0 \\ \frac{J_{11}}{\alpha} & 0 & \frac{J_{13}}{\alpha} \end{pmatrix};$$

Pour $p = 0$, $J_{33} = -J_{11}$ et $J_{11}^2 + J_{13}J_{31} = -1$ tous les autres coefficients de J étant nuls. Si $J_{11} = 0$, $-J_{13} = J_{31} = -1$, J est la projection sur le plan $\{x, p\}$ suivi de la rotation de $\frac{\pi}{2}$. Quant à la métrique, elle est dans ce cas $g = \begin{pmatrix} \frac{1}{\alpha} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\alpha} \end{pmatrix}$ qui est la métrique euclidienne pour $\alpha = 1$. Cela incite à garder ces valeurs pour $p \neq 0$. Dans ce cas

$$J = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & p \\ -1 & 0 & 0 \end{pmatrix}, \quad g = \begin{pmatrix} 1 + p^2 & -p & 0 \\ -p & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$J(\mathbf{t})$ consiste à projeter $\mathbf{t} = (\xi, \eta, \pi)$ en $\mathbf{t}_{\mathcal{K}} = (\xi, p\xi, \pi)$ sur le plan de contact \mathcal{K}_v et à prendre $J(\mathbf{t}_{\mathcal{K}})$. Quant à $J(\mathbf{t}_{\mathcal{K}})$, il consiste à projeter $\mathbf{t}_{\mathcal{K}}$ en $(\xi, 0, \pi)$ sur le plan $\{x, p\}$, à effectuer la rotation $(\xi, 0, \pi) \mapsto (\pi, 0, -\xi)$ et à reprojeter en sens inverse sur le plan de contact en $(\pi, p\pi, -\xi)$. En ce qui concerne la métrique, $\det(g) = 1$, le polynôme caractéristique est $\lambda^3 - (3 + p^2)\lambda^2 + (3 + p^2)\lambda - 1$ et la décomposition spectrale est la suivante. Le vecteur unitaire “vertical” $\partial_p = (0, 0, 1)$ est toujours vecteur propre de valeur propre 1 et dans le plan “horizontal” $\text{Span}\{\partial_x, \partial_y\}$, il y a deux vecteurs propres $\left(\frac{1}{2}(-p \pm \sqrt{4 + p^2}), 1, 0\right)$ de valeurs propres $\frac{1}{2}(2 + p^2 \mp p\sqrt{4 + p^2})$.

La sphère de rayon 1 a pour équation

$$(1 + p^2)\xi^2 - 2p\xi\eta + \eta^2 + \pi^2 = 1.$$

Elle intersecte l’axe des π en $\pi = \pm 1$ où le plan tangent est horizontal et son équateur est l’ellipse (E_p) d’équation

$$(1 + p^2)\xi^2 - 2p\xi\eta + \eta^2 = (\eta - p\xi)^2 + \xi^2 = \omega(\mathbf{t})^2 + \xi^2 = 1.$$

La figure 18 montre trois des ellipses (E_p) pour $p = 0$ (cercle), $p = 1.5$, $p = 3$. On voit qu’elle s’aplatit et s’allonge le long de son axe propre principal qui tourne vers l’axe des η .

3.4.3. Modification de l’exemple.

Nous avons vu plus haut que, en ce qui concerne les métriques, il est intéressant de regarder aussi la forme de contact $\omega = \frac{1}{\sqrt{1+p^2}}\omega_J$ dont le champ de Reeb est orthogonal aux plans de contact pour la métrique euclidienne. Le champ de Reeb est $\chi = \left\{-\frac{p}{\sqrt{1+p^2}}, \frac{1}{\sqrt{1+p^2}}, 0\right\}$ et $d\omega = \frac{1}{(1+p^2)^{\frac{3}{2}}}(pdy \wedge dp + dx \wedge dp)$.

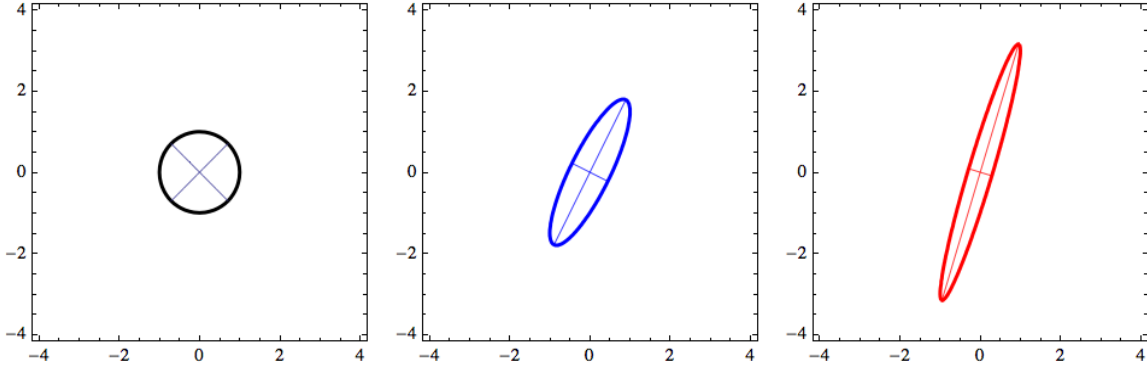


FIGURE 18. Trois exemples de l'ellipse (E_p) équateur de la sphère de rayon 1 de la métrique g de paramètre p . De gauche à droite, $p = 0$ (cercle), $p = 1.5$, $p = 3$.

La condition (1) $\omega(X) = g(X, \chi)$ impose $g_{12} = p(g_{11} - 1)$, $g_{22} = 1 + p^2(g_{11} - 1)$, $g_{23} = pg_{13}$.

La condition (2) $J(\chi) = 0$ impose $J_{12} = pJ_{11}$, $J_{22} = pJ_{21}$, $J_{32} = pJ_{31}$.

La condition (3) $\omega \circ J = 0$ impose quant à elle $J_{21} = pJ_{11} = J_{12}$, $J_{22} = pJ_{12} = p^2J_{11}$, $J_{23} = pJ_{13}$.

La condition (4) $d\omega(X, J(Y)) = -d\omega(J(X), Y)$ conduit aux relations $J_{32} = pJ_{31}$, $J_{33} = -(1 + p^2)J_{11}$.

La condition (5) $J^2(X) = -X + \omega(X)\chi$ impose en plus $J_{13}J_{31} = -(1 + p^2)J_{11}^2 - \frac{1}{1+p^2}$.

La condition (7) $d\omega(X, Y) = \alpha g(X, J(Y))$ conduit pour la métrique g aux contraintes

$$g_{11} = \frac{1 + (1 + p^2)^2 J_{11}^2 + ap^2 (1 + p^2)^{\frac{3}{2}} J_{13}}{\alpha (1 + p^2)^{\frac{5}{2}} J_{13}} = -\frac{J_{31}}{\alpha (1 + p^2)^{\frac{3}{2}}} + \frac{p^2}{1 + p^2},$$

$$g_{13} = \frac{J_{11}}{\alpha \sqrt{1 + p^2}},$$

$$g_{33} = \frac{J_{13}}{\alpha \sqrt{1 + p^2}}.$$

On vérifie alors la condition (6) $g(J(X), J(Y)) = g(X, Y) - \omega(X)\omega(Y)$ est bien satisfaite par ces valeurs et donne des deux côtés les coefficients

$\eta\xi'$	$\eta\eta'$	$\eta\pi'$
$\frac{pJ_{31}}{\alpha(1+p^2)^{\frac{3}{2}}}$	$\frac{p^2J_{31}}{\alpha(1+p^2)^{\frac{3}{2}}}$	$\frac{pJ_{11}}{\alpha\sqrt{(1+p^2)}}$
$\pi\xi'$	$\pi\eta'$	$\pi\pi'$
$\frac{J_{11}}{\alpha\sqrt{(1+p^2)}}$	$\frac{pJ_{11}}{\alpha\sqrt{(1+p^2)}}$	$\frac{J_{13}}{\alpha\sqrt{(1+p^2)}}$
$\xi\xi'$	$\xi\eta'$	$\xi\pi'$
$\frac{J_{31}}{\alpha(1+p^2)^{\frac{3}{2}}}$	$\frac{pJ_{31}}{\alpha(1+p^2)^{\frac{3}{2}}}$	$\frac{J_{11}}{\alpha\sqrt{(1+p^2)}}$

La condition (4) correspond quant à elle aux coefficients

$\eta\xi'$	$\eta\eta'$	$\eta\pi'$
$\frac{pJ_{31}}{(1+p^2)^{\frac{3}{2}}}$	$\frac{p^2J_{31}}{(1+p^2)^{\frac{3}{2}}}$	$-\frac{J_{11}}{\sqrt{(1+p^2)}}$
$\pi\xi'$	$\pi\eta'$	$\pi\pi'$
$-\frac{J_{11}}{\sqrt{(1+p^2)}}$	$-\frac{pJ_{11}}{\sqrt{(1+p^2)}}$	$-\frac{J_{13}}{\sqrt{(1+p^2)}}$
$\xi\xi'$	$\xi\eta'$	$\xi\pi'$
$\frac{J_{31}}{(1+p^2)^{\frac{3}{2}}}$	$\frac{pJ_{31}}{(1+p^2)^{\frac{3}{2}}}$	$-\frac{J_{11}}{\sqrt{(1+p^2)}}$

On voit que même sur un exemple aussi élémentaire (mais non complètement trivial) les possibilités de métriques et de structures presque complexes associées sont déjà assez riches. De même que ω , J et g ne dépendent que de p , mais elles varient le long des fibres de \mathbb{V}_J , le coefficient $\frac{1}{\sqrt{1+p^2}}$ de ω fonctionnant comme un facteur d'échelle. La multiplication complexe J possède deux degrés de liberté. Elle dépend des 3 paramètres J_{11} , J_{13} et J_{31} liés par la relation $J_{13}J_{31} = -(1+p^2)J_{11}^2 - \frac{1}{1+p^2}$ qui définit un hyperboloïde à deux nappes (H) dans le \mathbb{R}^3 de coordonnées $\{J_{11}, J_{13}, J_{31}\}$. Pour $J_{11} = 0$, on obtient dans le plan $\{J_{13}, J_{31}\}$ l'hyperbole équilatère (E) $J_{13}J_{31} = -\frac{1}{1+p^2}$ qui est située dans les quadrants $J_{13} > 0, J_{31} < 0$ et $J_{13} < 0, J_{31} > 0$ et intersecte la seconde diagonale aux points $\left\{ \pm \frac{1}{\sqrt{1+p^2}}, \mp \frac{1}{\sqrt{1+p^2}} \right\}$. Lorsque $p \rightarrow \infty$ l'hyperbole (E) s'écrase sur ses asymptotes et dégénère en l'union des axes, tandis que l'hyperboloïde (H) s'écrase sur le plan $\{J_{13}, J_{31}\}$ puisque $J_{11} \rightarrow 0$. La figure 19 représente les cas $p = 0$ et $p = 10$.

La métrique g possède également deux degrés de liberté si l'on fixe α , en prenant par exemple $\alpha = 1$, ce que nous allons faire. Pour $p = 0$, avec les paramètres $J_{11} = 0, J_{13} = 1$ et $J_{31} = -1$ on obtient la structure euclidienne. Si l'on regarde les

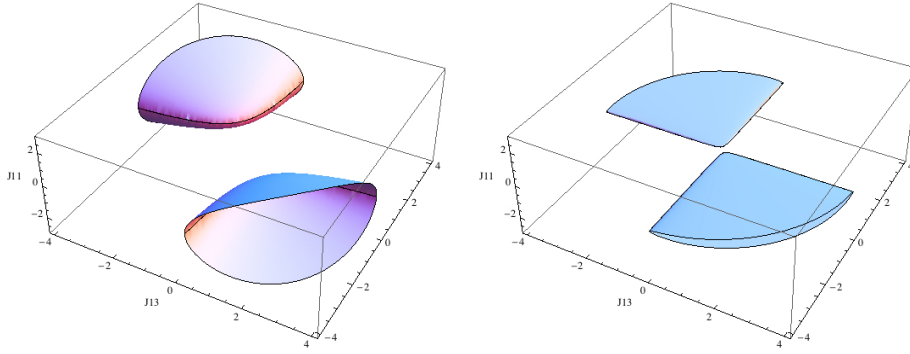


FIGURE 19. L'hyperboloïde à deux nappes (H) des valeurs des paramètres $\{J_{11}, J_{13}, J_{31}\}$ de la matrice J . À gauche le cas $p = 0$ et à droite le cas $p = 10$.

générateurs $\mathbf{t}_1 = \{1, p, 0\}$ et $\mathbf{t}_2 = \{0, 0, 1\}$ de K_v , on note que $g(\mathbf{t}_1, \mathbf{t}_2) = J_{11}$,

$$\|\mathbf{t}_1\|_g^2 = -J_{31}\sqrt{1+p^2}, \|\mathbf{t}_2\|_g^2 = \frac{J_{13}}{\sqrt{1+p^2}}$$

(on choisit $J_{13} \geq 0$ et $J_{31} \leq 0$). Si l'on souhaite que cette base de K_v soit orthogonale, il faut alors choisir $J_{11} = 0$, ce qui implique $J_{13}J_{31} = -\frac{1}{1+p^2}$. On a alors

$$\|\mathbf{t}_1\|_g^2 = \frac{1}{J_{13}(1+p^2)}\sqrt{1+p^2} = \frac{1}{J_{13}\sqrt{1+p^2}},$$

$$\|\mathbf{t}_2\|_g^2 = \frac{J_{13}}{\sqrt{1+p^2}}.$$

Si l'on veut des vecteurs g -unitaires, il faut alors scaler \mathbf{t}_1 et \mathbf{t}_2 . Pour J , on obtient

$$J = \begin{pmatrix} 0 & 0 & J_{13} \\ 0 & 0 & pJ_{13} \\ J_{31} & pJ_{31} & 0 \end{pmatrix} \text{ et par conséquent, pour les vecteurs } \mathbf{t} = \{\xi, p\xi, \pi\} \in K_v,$$

$J\mathbf{t} = \{J_{13}\pi, pJ_{13}\pi, (1+p^2)J_{31}\xi\} = \left\{J_{13}\pi, J_{13}p\pi, -\frac{1}{J_{13}}\xi\right\}$. Dans le plan de contact, J est la rotation de $\frac{\pi}{2}$ avec deux homothéties inverses de rapport J_{13} pour π et $\frac{1}{J_{13}}$ pour ξ et l'on vérifie que l'on y a donc bien $J^2 = -I$. J n'est pas une isométrie pour la structure euclidienne, mais elle l'est pour la métrique g puisque $\mathbf{t} = \{\xi, p\xi, \pi\} \in K_v$ et $J\mathbf{t}$ sont tous deux de g -norme au carré $\frac{J_{13}^2\pi^2 + \xi^2}{J_{13}\sqrt{1+p^2}}$. Par rapport à la métrique euclidienne, g est une métrique *conforme* : elle préserve les angles mais pas les normes.

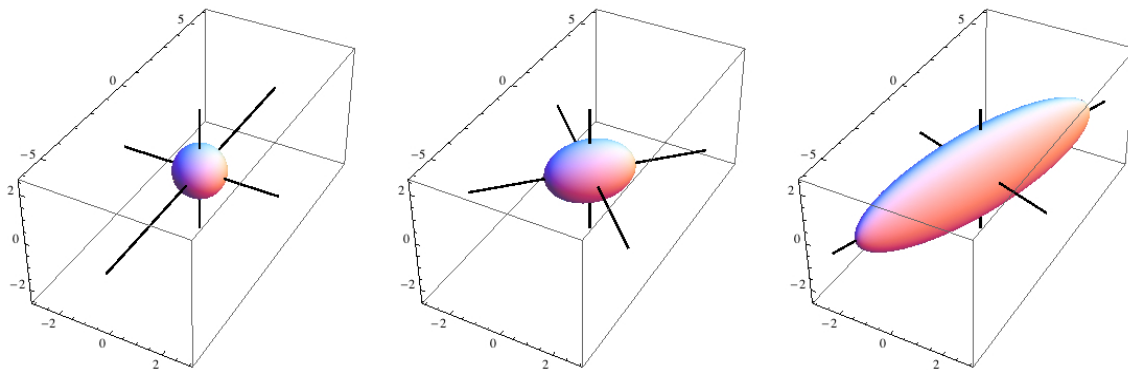


FIGURE 20. Les g -sphères unitaires de la métrique g dans le cas $J_{11} = 0$, $J_{13} = 1$, $J_{13} = -\frac{1}{1+p^2}$. De gauche à droite $p = 0, 1, 3$.

La matrice de g est assez compliquée mais sa diagonalisation est simple. En effet son système de vecteurs propres est $\left\{\frac{1}{p}, 1, 0\right\} = \frac{1}{p}\mathbf{t}_1$ de valeur propre $-\frac{J_{31}}{\sqrt{1+p^2}}$, $\{0, 0, 1\} = \mathbf{t}_2$ de valeur propre $\frac{J_{13}}{\sqrt{1+p^2}}$ et $\{-p, 1, 0\} = \chi\sqrt{1+p^2}$ de valeur propre 1. Comme χ est de norme 1 par définition, on voit comment la g -sphère unité évolue lorsque p varie. Si l'on veut obtenir la sphère unité euclidienne pour $p = 0$, il faut prendre $J_{13} = 1$, $J_{13} = -\frac{1}{1+p^2}$, ce qui implique d'ailleurs que J soit la rotation euclidienne de $\frac{\pi}{2}$ (sans homothétie) dans le plan de contact. On obtient alors pour sphère de g -rayon 1 l'ellipsoïde euclidien intersectant les axes orthogonaux \mathbf{t}_1 , \mathbf{t}_2 , χ en

$$\begin{aligned} & \pm (1+p^2)^{\frac{1}{4}} \{1, p, 0\}, \\ & \pm (1+p^2)^{\frac{1}{4}} \{0, 0, 1\}, \\ & \pm \frac{1}{\sqrt{1+p^2}} \{-p, 1, 0\}. \end{aligned}$$

La figure 20 montre les g -sphères unitaires pour $p = 0, 1, 3$.

Un autre choix naturel pour J_{13} et J_{31} est $J_{13} = \frac{1}{\sqrt{1+p^2}}$, $J_{31} = -\frac{1}{\sqrt{1+p^2}} = -J_{13}$, ce qui implique d'ailleurs que J soit la rotation de $\frac{\pi}{2}$ dans le plan de contact avec les homothéties de rapport $\frac{1}{\sqrt{1+p^2}}$ pour π et $\sqrt{1+p^2}$ pour ξ . Le système de vecteurs propres de g est $\left\{\frac{1}{p}, 1, 0\right\} = \frac{1}{p}\mathbf{t}_1$ de valeur propre $\frac{1}{\sqrt{1+p^2}}$, $\{0, 0, 1\} = \mathbf{t}_2$ de valeur propre $\frac{1}{1+p^2}$ et $\{-p, 1, 0\} = \chi\sqrt{1+p^2}$ de valeur propre 1. On obtient alors pour sphère de g -rayon 1 l'ellipsoïde euclidien intersectant les axes orthogonaux \mathbf{t}_1 , \mathbf{t}_2 , χ

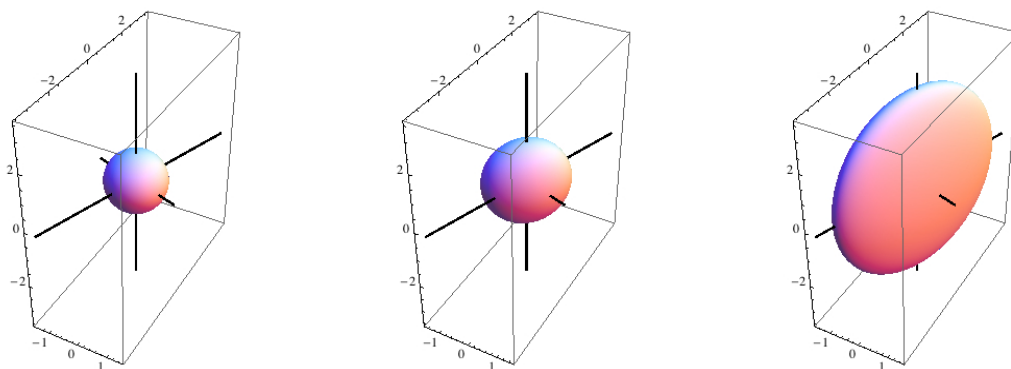


FIGURE 21. Les g -sphères unitaires de la métrique g dans le cas $J_{13} = \frac{1}{\sqrt{1+p^2}}$, $J_{13} = -\frac{1}{\sqrt{1+p^2}} = -J_{13}$. De gauche à droite $p = 0, 1, 3$.

en

$$\begin{aligned} & \pm \{1, p, 0\}, \\ & \pm \sqrt{1+p^2} \{0, 0, 1\}, \\ & \pm \frac{1}{\sqrt{1+p^2}} \{-p, 1, 0\}. \end{aligned}$$

La figure 21 montre les g -sphères unitaires pour $p = 0, 1, 3$.

Nous aurons à revenir plusieurs fois par la suite sur les métriques de nos modèles, en particulier dans les chapitres 10 et 14. Dans le premier cas, elles ne seront pas complètement ω -compatibles et, dans le second cas, elles le seront mais en tant que métriques *sous-riemanniennes* plutôt que riemanniennes.

3.4.4. Le cas de la métrique euclidienne.

On remarquera que $\omega = \frac{1}{\sqrt{1+p^2}}\omega_J$ est imposé par la condition (1) $\omega(X) = g(X, \chi)$ pour $\omega = f\omega_J$ ($f \neq 0$) si g est la métrique euclidienne E . En effet, comme

$$d\omega = -(\partial_p f) dy \wedge dp + (p\partial_p f + f) dx \wedge dp,$$

$d\omega(X, \chi) = 0$ pour tout X implique $\chi = \left(\frac{\partial_p f}{f^2}, \frac{p\partial_p f + f}{f^2}, 0\right)$ et donc (1) s'écrit

$$f(\eta - p\xi) = \xi \frac{\partial_p f}{f^2} + \eta \frac{p\partial_p f + f}{f^2}$$

pour tous ξ, η , ce qui implique $f^2 = \frac{1}{1+p^2}$. Nous avons vu plus haut que l'on a alors

$$\chi = \left(-\frac{p}{\sqrt{1+p^2}}, \frac{1}{\sqrt{1+p^2}}, 0 \right).$$

$$d\omega = \frac{1}{(1+p^2)^{\frac{3}{2}}} (pdy \wedge dp + dx \wedge dp).$$

La métrique euclidienne E ainsi que la rotation de $\frac{\pi}{2}$, $J_{\mathcal{K}}$, dans les plans de contact satisfont presque toutes les conditions.

1. $\omega(X) = g(X, \chi)$. Nous venons de le voir.
2. $J(\chi) = 0$. On écrit que l'on projette d'abord E -orthogonalement X en $X_{\mathcal{K}}$ sur \mathcal{K} au moyen de la projection $P_{\mathcal{K}}$ puis que l'on applique $J_{\mathcal{K}}$. Pour $P_{\mathcal{K}}$, on écrit que $P_{\mathcal{K}}(-p\xi, \xi, 0) = 0$ pour tout ξ ($\text{Span}\{\chi\}$ se projette sur $\{\chi\}$) et $P_{\mathcal{K}}(\xi, p\xi, \pi) = (\xi, p\xi, \pi)$ pour tous ξ, π ($P_{\mathcal{K}}$ est l'identité sur les plans de contact.) On suppose donc que $J_{\mathcal{K}}(\xi, p\xi, \pi) = \left(\frac{\pi}{\sqrt{1+p^2}}, \frac{p\pi}{\sqrt{1+p^2}}, -\xi\sqrt{1+p^2} \right)$ (on vérifie trivialement que $J_{\mathcal{K}}^2 = -I$ et que $J_{\mathcal{K}}$ est une E -isométrie) et l'on obtient

$$P_{\mathcal{K}} = \begin{pmatrix} \frac{1}{1+p^2} & \frac{p}{1+p^2} & 0 \\ \frac{p}{1+p^2} & \frac{p^2}{1+p^2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad P_{\mathcal{K}}(\xi, \eta, \pi) = \left(\frac{\xi + p\eta}{1+p^2}, \frac{p(\xi + p\eta)}{1+p^2}, \pi \right)$$

$$J(\xi, \eta, \pi) = J_{\mathcal{K}} \circ P_{\mathcal{K}}(\xi, \eta, \pi) = \left(\frac{\pi}{\sqrt{1+p^2}}, \frac{p\pi}{\sqrt{1+p^2}}, \frac{-(\xi + p\eta)}{1+p^2} \sqrt{1+p^2} \right)$$

$$= \left(\frac{\pi}{\sqrt{1+p^2}}, \frac{p\pi}{\sqrt{1+p^2}}, \frac{-(\xi + p\eta)}{\sqrt{1+p^2}} \right).$$

On a évidemment $J(\chi) = 0$ puisque, pour χ , $\pi = \xi + p\eta = 0$.

3. $\omega \circ J = 0$. C'est évident.
4. $d\omega(X, J(Y)) = -d\omega(J(X), Y) = d\omega(Y, J(X))$. En effet, si $Y = (\xi', \eta', \pi')$,

$$d\omega(X, J(Y)) = \frac{1}{(1+p^2)^{\frac{3}{2}}} \left[\begin{aligned} & p \left(\eta' \left(\frac{-(\xi' + p\eta')}{\sqrt{1+p^2}} \right) - \pi' \frac{p\pi'}{\sqrt{1+p^2}} \right) \\ & + \xi' \left(\frac{-(\xi' + p\eta')}{\sqrt{1+p^2}} \right) - \pi' \frac{\pi'}{\sqrt{1+p^2}} \end{aligned} \right]$$

$$= \frac{-1}{(1+p^2)^{\frac{3}{2}}} \left[(\xi + p\eta)(\xi' + p\eta') \frac{1}{\sqrt{1+p^2}} + \pi\pi' \sqrt{1+p^2} \right]$$

et cette formule est bien symétrique en X, Y .

5. $J^2(X) = -X + \omega(X)\chi$. La vérification est triviale. En effet,

$$J^2(X) = \left(\frac{-(\xi + p\eta)}{1 + p^2}, \frac{-p(\xi + p\eta)}{1 + p^2}, -\pi \right)$$

$$\omega(X)\chi = \frac{1}{\sqrt{1 + p^2}}(\eta - p\xi) \left(-\frac{p}{\sqrt{1 + p^2}}, \frac{1}{\sqrt{1 + p^2}}, 0 \right).$$

6. $g(J(X), J(Y)) = g(X, Y) - \omega(X)\omega(Y)$. Cela vient du fait que $J_{\mathcal{K}}$ est une E -isométrie.

7. La sous-condition de E -orthogonalité $X \perp J(X)$ est bien satisfaite car

$$\frac{\xi\pi}{\sqrt{1 + p^2}} + \frac{\eta p\pi}{\sqrt{1 + p^2}} - \frac{\pi(\xi + p\eta)}{\sqrt{1 + p^2}} = 0.$$

Mais la condition complète $d\omega(X, Y) = \alpha g(X, J(Y))$ ne peut pas être satisfaite avec un facteur α constant. En effet,

$$g(X, J(Y)) = \frac{\xi\pi'}{\sqrt{1 + p^2}} + \frac{\eta p\pi'}{\sqrt{1 + p^2}} - \frac{\pi(\xi' + p\eta')}{\sqrt{1 + p^2}}$$

alors que

$$d\omega(X, Y) = \frac{1}{(1 + p^2)^{\frac{3}{2}}} [p(\eta\pi' - \eta'\pi) + (\xi\pi' - \xi'\pi)].$$

La condition (7) ne peut donc être satisfaite qu'avec le facteur variable $\alpha = \frac{1}{1 + p^2}$.

4. Courbure, 2-jets et structure d'Engel

Nous avons vu au Vol I, 5.6. l'importance que des spécialistes comme Steve Zucker accordent aux détecteurs de courbure et nous avons posé la question de savoir s'il fallait traiter ces détecteurs de façon standard comme des "array processors" au sens de Koenderink, ou plutôt de façon "simplexe"²⁶ comme des processeurs ponctuels en introduisant les 2-jets de courbes dans \mathbb{R}^2 . Si l'on opte pour la seconde option, ce que semblent favoriser les données expérimentales, on doit introduire en plus des x, y, p ou des x, y, θ une 4ème variable indépendante κ pour la courbure et une 1-forme supplémentaire exprimant son interprétation comme courbure.

26. Pour le concept de "simplexité" introduit par Alain Berthoz, cf. Vol I, 5.4.6.

4.1. Structure d'Engel dans le cas $\{x, y, p\}$

Dans le cas $\{x, y, p\}$, on travaille dans $\mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R}$ avec la 1-forme de contact $\omega_J = dy - p dx$ et la base non holonome des plans de contact $\{\mathbf{t}_1 = \partial_x + p \partial_y, \mathbf{t}_2 = \partial_p\}$, le 3ème vecteur tangent de base \mathbf{t}_3 étant donné par le crochet de Lie $[\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3 = -\partial_y$.

On introduit une 4ème variable κ liée à la courbure, on travaille dans l'espace $\widetilde{\mathbb{V}}_J = \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}$, et on écrit que l'interprétation naturelle de κ est la dérivée seconde $f''(x)$ pour les courbes d'équation $y = f(x)$. L'espace des $\widetilde{\mathbb{V}}_S$ des $\{x, y, p, \kappa\}$ est l'espace des 2-jets $J^2(\mathbb{R}^2) = J^2(\mathbb{R}, \mathbb{R})$ et sa structure canonique, dite structure d'Engel,²⁷ est le système (dit de Pfaff) constitué des deux 1-formes ω_J (étendue trivialement de \mathbb{V}_J à $\widetilde{\mathbb{V}}_J$) et $\tau_J = dp - \kappa dx$.

Le noyau de τ_J dans $\widetilde{\mathbb{V}}_J$ est engendré par les 3 vecteurs tangents

$$\begin{cases} \mathbf{t}_1^\kappa = \partial_x + p \partial_y + \kappa \partial_p = \mathbf{t}_1 + \kappa \mathbf{t}_2 \\ \mathbf{t}_3 = -\partial_y \\ \mathbf{t}_4 = \partial_\kappa \end{cases}$$

alors que le noyau de ω_J étendu à $\widetilde{\mathbb{V}}_J$ est engendré par \mathbf{t}_1^κ , \mathbf{t}_2 et \mathbf{t}_4 . La distribution de plans est maintenant $\text{Span}\{\mathbf{t}_1^\kappa, \mathbf{t}_4\}$ ²⁸ et elle engendre toute l'algèbre de Lie puisque $[\mathbf{t}_1^\kappa, \mathbf{t}_4] = -\mathbf{t}_2 = -\partial_p$ et $[[\mathbf{t}_1^\kappa, \mathbf{t}_4], \mathbf{t}_1^\kappa] = \mathbf{t}_3 = -\partial_y$.

4.2. Structure d'Engel dans le cas $\{x, y, \theta\}$

Dans le cas de $SE(2)$, on travaille dans $\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1$ avec la 1-forme de contact $\omega_S = -\sin(\theta) dx + \cos(\theta) dy$ et la base non holonome des plans de contact $\{X_1 = \cos(\theta) \partial_x + \sin(\theta) \partial_y, X_2 = \partial_\theta\}$, le 3ème vecteur tangent de base X_3 étant donné par le crochet de Lie $[X_1, X_2] = X_3 = \sin(\theta) \partial_x - \cos(\theta) \partial_y$.

On introduit la courbure K et on travaille dans l'espace $\widetilde{\mathbb{V}}_S = \mathbb{R}^2 \times \mathbb{S}^1 \times \mathbb{R}$. Le système de Pfaff définissant la structure d'Engel est maintenant constitué des deux 1-formes ω_S et $\tau_S = d\theta - K ds$, s étant la paramétrisation des courbes du plan (x, y) par leur longueur d'arc. En effet, on sait que la courbure est donnée par $K = \frac{d\theta}{ds}$.

En coordonnées cartésiennes, pour une courbe d'équation $y = f(x)$, la courbure K est donnée par (formule de Newton)

$$K = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}}.$$

27. Friedrich Engel était l'un des principaux disciples de Sophus Lie. Nous l'avons déjà rencontré dans l'*Introduction* du Vol I à propos du débat Riemann-Helmholtz sur les fondements de la géométrie. Nous y reviendrons plus en détail à la section 2.2.

28. Si les X, Y, \dots sont des vecteurs, on note $\text{Span}\{X, Y, \dots\}$ le vectoriel engendré ("spanned") par les X, Y, \dots

Le lien entre les deux formules pour K est facile à écrire. Comme le vecteur tangent $(x'(s), y'(s))$ est de norme 1 par définition de s , on a :

$$\begin{aligned} x'(s) &= \cos(\theta) , \\ y'(s) &= \sin(\theta) = f'(x) x'(s) , \\ y''(s) &= f''(x) x'(s)^2 + f'(x) x''(s) , \\ f''(x) &= \frac{y''(s) - f'(x) x''(s)}{x'(s)^2} = \frac{\cos(\theta) \theta'(s) + \tan(\theta) \sin(\theta) \theta'(s)}{\cos(\theta)^2} \\ &= K \left(\frac{1}{\cos(\theta)} + \frac{\sin(\theta)^2}{\cos(\theta)^3} \right) = \frac{K}{\cos(\theta)^3} . \end{aligned}$$

Mais

$$1 + f'(x)^2 = 1 + \tan(\theta)^2 = \frac{1}{\cos(\theta)^2}$$

et donc

$$K = f''(x) \cos(\theta)^3 = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}} .$$

Pour exprimer la seconde 1-forme τ_S , on écrit

$$\begin{aligned} \tau_S &= d\theta - K ds \\ &= d\theta - K (\cos(\theta) dx + \sin(\theta) dy) . \end{aligned}$$

Le noyau de la 1-forme τ_S est engendré par les 3 vecteurs tangents

$$\begin{cases} X_1^K = \cos(\theta) \partial_x + \sin(\theta) \partial_y + K \partial_\theta = X_1 + K X_2 \\ X_3 = \sin(\theta) \partial_x - \cos(\theta) \partial_y \\ X_4^K = \partial_K \end{cases}$$

tandis que celui de ω_S étendue à \widetilde{V}_S est engendré par X_1, X_2 et X_4^K . La distribution des plans est maintenant $\text{Span} \{X_1^K, X_4^K\}$. Elle engendre toute l'algèbre de Lie puisque $[X_1^K, X_4^K] = -X_2 = -\partial_\theta$ et $[[X_1^K, X_4^K], X_1^K] = -X_3 = -\sin(\theta) \partial_x + \cos(\theta) \partial_y$.

Comme le montre la figure 22, dans l'espace \widetilde{V}_S , le vecteur tangent $X_1 + K X_2$ se déploie de façon hélicoïdale le long de l'axe des K .

5. La notion cruciale d'élément de contact

5.1. Les "écailles de poisson" sont de bonnes géomètres (Klein)

Les modèles neurogéométriques dont nous venons de parler reposent sur la notion d'"engrafted variable" de Hubel, c'est-à-dire, répétons-le encore et encore, de

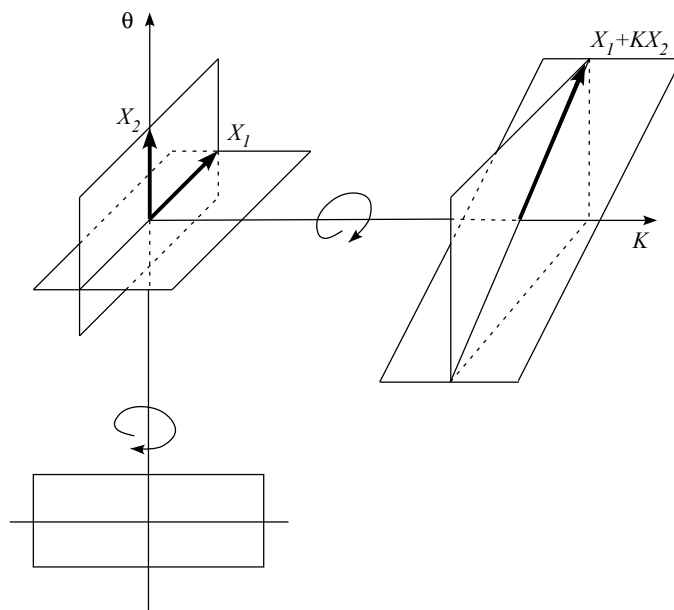


FIGURE 22. La structure d'Engel. La partie gauche correspond à $K = 0$ et à la structure de contact. Elle montre comment le plan de contact engendré par $\{X_1, X_2\}$ tourne le long de l'axe des θ . La partie droite montre comment le plan $\{X_1^K = X_1 + KX_2, X_4^K = \partial_K\}$ tourne le long du 4ème axe, celui des K . La figure est inexacte car elle représente en dimension 3 un espace de dimension 4.

fibration où, sur un espace de base M , on “greffe” des variables supplémentaires ayant vocation à être interprétées comme des dérivées partielles de fonctions sur M . La notion centrale est celle d'*élément de contact*, c'est-à-dire d'un point x de M et d'un hyperplan K_x de $T_x M$. Alors que les cellules rétinienne détectent des positions (i.e. idéalement des points x), les neurones corticaux simples de V1 détectent des

éléments de contact et leur connectivité cortico-corticale implémente une “géométrie d’intégrabilité” de ces éléments de contact. Pour comprendre au niveau *neuronal* comment le système visuel peut implémenter un calcul intégral-différentiel, il faut donc comprendre au niveau *mathématique* comment un calcul intégral-différentiel peut être équivalent à une géométrie de connectivité entre éléments de contact.

Or, comme nous l’avons déjà dit, il se trouve que cette équivalence a été découverte, démontrée, approfondie, systématisée, généralisée et amplifiée au cours du XIX^e siècle de Pfaff à Cartan et a constitué, avec celles de Galois et de Riemann, l’une des plus grandes révolutions mathématiques de ce siècle. Son application à la géométrie de la perception est analogue à ce qui s’est passé avec la géométrisation de la gravitation dans la relativité générale grâce à la reformulation révolutionnaire de la notion d’espace géométrique par Riemann dans son texte fondateur de 1854 [464].

Avant de nous plonger dans cette histoire fascinante, nous allons à titre de prélude évoquer la façon dont cette notion d’élément de contact fut pensée par Sophus Lie avec un sens aigu de *l’inversion de point de vue* qu’elle apportait. Lie considère des surfaces $z = f(x, y)$ (ou, plus généralement, $F(x, y, z) = 0$) dans l’espace \mathbb{R}^3 de coordonnées (x, y, z) , c’est-à-dire l’équivalent en dimension 3 de ce que nous avons traité en dimension 2 avec les courbes $y = f(x)$ du plan. De même que la structure de contact dont nous avons parlé consistait à introduire une variable supplémentaire p et la 1-forme $\omega = dy - pdx$ dont la structure de contact est le noyau, Lie introduit, en plus des (x, y, z) , les nouvelles variables p et q ayant vocation à être interprétées comme les dérivées partielles $\frac{\partial z}{\partial x}$ et $\frac{\partial z}{\partial y}$. En chaque point $v = (x, y, z, p, q)$ de ce \mathbb{R}^5 le noyau K_v de la 1-forme $\omega = dz - pdx - qdy$ est un hyperplan de dimension 4 de $T_v\mathbb{R}^5 \simeq \mathbb{R}^5$ qui contient la fibre $\mathbb{R}_{(p,q)}^2$ au-dessus du (x, y, z) considéré.

Dans un texte émouvant de 1894 [294], *L’œuvre géométrique de Sophus Lie*, l’ami de Lie Felix Klein (cf. plus bas la section historique 2 du chapitre 4) insiste sur l’importance de ce changement de perspective apporté avec cette “grande théorie” et cette “vue nouvelle, claire et pénétrante” (p. 19) par ce “grand géomètre” souvent incompris des mathématiciens “accoutumés plutôt au point de vue analytique” (p. 1). Klein explique que, inspiré par les conceptions géométriques de Monge et Plücker, Lie va introduire de “nouveaux éléments de l’espace” (p. 11) beaucoup plus généraux que des points et, au lieu d’appliquer les méthodes de l’analyse différentielle à la géométrie (calculs de courbure, de géodésiques, etc.) il va s’intéresser à “*la réciproque*” et développer au contraire

“l’application de l’intuition géométrique à l’Analyse.” (p. 3)

Les “éléments de surface” (“Flächen-element”) de Lie sont joliment comparés par Klein à “une écaille de poisson infiniment petite” (p. 13). Ils sont composés d’un point $a = (x, y, z)$ et d’un plan P en a considéré intuitivement comme un plan

infinitésimal abstrait (un possible plan tangent). Et comme l'équation d'un plan passant par $a = (x, y, z)$ est de la forme $z' - z = p(x' - x) + q(y' - y)$, on obtient un espace de dimension 5 de coordonnées $v = (x, y, z, p, q)$ satisfaisant l'équation $\omega = dz - p dx - q dy = 0$. Klein explique alors qu'en géométrie classique "ordinaire", la variété des P passant par un point a fixé est de dimension 2 (la fibre $\mathbb{R}_{(p,q)}^2$ au-dessus de a) et que ces variétés constituent une première classe (classe (1)) de variétés de contact. Mais, dans la géométrie de Lie, il y a deux autres classes de variétés de contact de dimension 2 qui interviennent. La classe (2) correspond au cas où, étant donnée une courbe γ dans \mathbb{R}^3 , la variété est celle des P tangents à γ aux points a de γ (il y en a une famille à 1-paramètre en chaque a , ceux tournant autour de la tangente), variété que l'on pourrait appeler, si l'on nous permet la métaphore, le "livre"²⁹ de reliure γ . La classe (3) correspond au cas où, étant donnée une surface S de \mathbb{R}^3 , la variété est celle des plans P tangents à S aux points a de S (il n'y en a qu'un pour chaque $a : P = T_a S$). Ces trois classes de variétés bidimensionnelles d'éléments de contact (fibres $\mathbb{R}_{(p,q)}^2$, "livres" ayant pour reliure une courbe γ , surfaces S) satisfont la même équation de Pfaff de base $\omega = dz - p dx - q dy = 0$. Lie introduit alors les *transformations de contact* (*Berührungstransformationen*) qui sont les changements de coordonnées dans \mathbb{R}^5 laissant $\omega = 0$ invariante. Ces transformations sont beaucoup plus générales que des changements de variables classiques ("ordinaires") dans \mathbb{R}^3 . En effet, elles transforment des éléments de contact en éléments de contact mais peuvent *changer leur classe*. Et bien sûr, si deux variétés de contact ont un contact (i.e. un élément de contact en commun), leurs transformées ont également un contact. Selon Klein,

"leur véritable et profonde portée restait ignorée"

alors que pourtant elles "sont d'une telle importance." (p. 15)

Et Klein continue :

"une des plus merveilleuses applications des transformations de contact se trouvent dans la théorie des équations aux dérivées partielles." (p. 17)

Il s'agit en particulier des équations aux dérivées partielles (EDP) du premier ordre. $f(x, y, z, p, q) = 0$. Paradoxalement, p et q peuvent même être absents de f (par exemple un point fixe $a = (x, y, z)$) car une telle équation porte sur des éléments de contact (a, P) qui ont toujours un contenu infinitésimal et différentiel (le plan P est un plan tangent). Une telle EDP définit une hypersurface Σ^4 de dimension 4 dans \mathbb{R}^5 et trouver une solution au sens de Lie c'est trouver une sous-variété de contact bidimensionnelle Γ^2 de Σ^4 , le fait de savoir quelle est la classe de Γ^2 étant non pertinent. Et Klein y insiste :

29. Une métaphore plus juste serait même celle de "fichier rotatif". Nous rencontrerons au chapitre 15, section 4.8 une métaphore analogue de "livre ouvert" dans les travaux d'Emmanuel Giroux sur la géométrie de contact.

“À ces nouveaux point de vue cette théorie prend un degré de profondeur inconnue auparavant et la véritable signification des mots *solution*, *solution générale*, *complète*, *singulière*, introduits par *Lagrange* et *Monge*, devient extrêmement claire.” (p. 18)

Une solution est “complète” si Σ^4 est feuilletée par une famille à deux paramètres de Γ^2 . Si dans cette famille à deux paramètres on prend une sous-famille à un paramètre, l’enveloppe de celle-ci est une solution “générale”.

Klein prend l’équation la plus simple $z = 0$. Chez Lie, cette équation ne définit pas le plan $z = 0$ dans \mathbb{R}^3 mais le $\Sigma^4 = \mathbb{R}^4$, $z = 0$, dans \mathbb{R}^5 . Les points du plan (x, y) considérés comme des Γ^2 , c’est-à-dire les fibres $\mathbb{R}_{(p,q)}^2$, forment une solution *complète* car elles feuilletent le $\Sigma^4 = \mathbb{R}^4$. Pour la solution *générale* on doit considérer les familles à un paramètre de points du plan, c’est-à-dire des courbes quelconques γ et considérer les “livres” Γ^2 de reliure γ . Enfin comme surface S il y a le plan $z = 0$ dans \mathbb{R}^3 qui est une solution *singulière*. Et Klein conclut en soulignant que

“l’immense importance et l’intérêt capital de ce simple exemple” (p. 19)

vient du fait que toute EDP du premier ordre peut s’y ramener par une transformation de contact.

Remarque. Le lecteur aura constaté à quel point l’épistémologie de la neurogéométrie reprend cette remarquable évaluation par Klein de la “révolution” accomplie par Lie. □

5.2. Des “point processors” peuvent être de bons géomètres

Les géomètres qui ont mis en place la “révolution copernicienne” de la géométrisation du calcul différentiel ont intuitionné puis progressivement formalisé cette idée de base que nous avons vu à l’œuvre dans la structure de contact. On a un espace de base M de positions et en chaque position a on peut considérer des données différentielles (a, D) d’un certain type. Cela constitue des fibrations. On va alors découpler la partie géométrique des déplacements dans M et la partie algébrique des opérations sur les D . Cela commence avec la réinterprétation des fonctions $f : M \rightarrow \mathbb{R}$ comme sections de la fibration $M \times \mathbb{R} \rightarrow M$, la structure algébrique de \mathbb{R} étant en fibre. Une fonction est ainsi une assignation d’une entité d’un certain type (un nombre réel) à chaque position. L’idée d’élément de contact si joliment comparée par Klein à de petites écailles de poissons consiste à considérer des fibrations dont les fibres sont constituées d’éléments différentiels exprimant des possibilités de *tangence* (c’est-à-dire justement de “contact”). Nous commenterons à la section 3.5 cette citation de Cartan où l’auteur souligne lui-même avec force ce caractère ponctuel de conditions de tangence, indépendamment de toute intégrale particulière :

“Le système (1) exprime que *les tangentes en un point quelconque de l'espace à une multiplicité M^p qui passe par ce point satisfont à certaines conditions qui ne dépendent que du point considéré.*” ([91] p. 248, souligné par Cartan)

Un problème d'intégration sera alors défini par un (ou des) champ(s) de telles conditions de tangence, par exemple par l'annulation d'une forme de Pfaff $\omega = 0$ et on pourra alors définir les intégrales comme des sous-variétés de M qui, lorsqu'on les relève dans les fibrations, satisfont toutes les conditions de tangence préconisées. Une intégrale est une sous-variété qui a la propriété que tout petit déplacement de position en elle maintient les contraintes de contact. Cette possibilité de définir ponctuellement des conditions de tangence par ce que Jan Koenderink appelait, à propos des espaces de jets, des “point processors” [299] (cf. la section 4 de l'Introduction) est la clef de la géométrie de l'intégrabilité.

Une fois conquis ce découplage entre déplacements dans la base et conditions ponctuelles de contact permettant de définir l'intégrabilité au moyen de champs de conditions de contact, tous ces grands géomètres des écoles allemandes et françaises, en parallèle avec ceux de l'école anglaise, ont alors découvert *l'algèbre linéaire* des éléments différentiels D constituant les fibres de ces fibrations. Il y a les espaces tangents, leurs sous-espaces, leurs changements de base, leurs grassmanniennes, leurs projectivisations, leurs duaux (espaces cotangents), leurs applications linéaires avec leurs noyaux, leurs images et leurs relations dimensionnelles, leurs endomorphismes, leurs automorphismes (groupes linéaires) avec leurs déterminants et leurs mineurs, leurs tenseurs, leurs formes multilinéaires antisymétriques, leurs espaces de jets d'ordre supérieur, etc. Aujourd'hui nous apprenons cela à l'école mais la découverte fut majeure.

Géométriser l'intégrabilité : une "révolution copernicienne"

1. Que signifie une "révolution copernicienne"

1.1. De Copernic à Galois

Dans la section 4 du chapitre 2 nous avons parlé du "reverse engineering" permettant de reconstruire à partir d'une activité neuronale dans un espace de très grande dimension un recouvrement topologique d'un espace de basse dimension par les champs récepteurs des neurones. Nous avons parlé à ce propos de "révolution copernicienne" en référence explicite à Kant.

Dans la section 7.3 de ce même chapitre nous avons parlé du "retournement structural" effectué par Cartan avec l'introduction des formes différentielles en tant qu'expressions abstraites et symboliques. Ce "retournement" concernait la relation entre (i) les méthodes d'intégration des équations de Pfaff $\omega = 0$ et (ii) leurs conditions de possibilité d'intégrabilité. Il consiste à *inverser* la relation de priorité entre (i) et (ii). On peut également parler à son propos de "révolution copernicienne". Nous donnons à cette expression un sens technique que nous voudrions préciser à travers quelques exemples.

La vraie révolution copernicienne a consisté à passer d'un repère géocentrique à un repère héliocentrique pour théoriser le mouvement des planètes. Elle remonte à l'antiquité sur le plan théorique (témoignage d'Archimède sur l'héliocentrisme d'Aristarque de Samos, 310-230 av. J.-C.) et triomphera techniquement, après le *De revolutionibus orbium coelestium* (1543) de Copernic (qui utilise encore des épicycles à la Ptolémée), avec les modèles géniaux de trajectoires planétaires elliptiques introduits par Kepler (*Astronomia nova*, 1609 et *Harmonices Mundi*, 1619) et ses spectaculaires calculs des "Tabulæ Rudolphinæ" (1627) où il recalcule les données de son maître Tycho Brahé dans un repère héliocentrique.

Cette inversion de système de référence a servi de métaphore pour d'autres inversions brutales de perspectives. Par exemple Kant l'utilisera pour qualifier sa philosophie transcendantale de l'objectivité. Dans la Préface à la seconde édition de la *Critique de la Raison pure* (1787) il explique :

"Il en est ici comme de l'idée première de Copernic : voyant qu'il ne pouvait venir à bout de l'explication des mouvements du ciel en admettant que toute l'armée des astres tournait autour du spectateur, il essaya de voir

s'il ne réussirait pas mieux en faisant tourner le spectateur, et en laissant en revanche les astres au repos. On peut faire un essai du même genre en métaphysique, au sujet de l'*intuition* des objets." ([281], III, 12, trad. A., J.-L. Delamarre, F. Marty)

Ce n'est plus simplement l'objet qui affecte un sujet réceptif mais le sujet qui constitue transcendantalement l'objet. Dans la Préface du Vol 1 (p. 2), nous avons d'emblée souligné que de nombreux neuroscientifiques de la vision (comme le prix Nobel John O'Keefe ou l'académicien Stanislas Dehaene) ont soutenu explicitement la conception kantienne pour les objets de perception. Kant passe ainsi de la fausse évidence que les objets de l'expérience sont déjà donnés en tant qu'objets dans la phénoménalité intuitive à une théorie des conditions de possibilité de l'expérience elle-même. D'où la célèbre affirmation concluant la section "Du principe suprême de tous les jugements synthétiques"¹ avec laquelle Kant résume ce que signifie la *constitution* des objets en tant qu'objets d'une connaissance possible au-delà de leur simple donnée phénoménale empirique :

"Les conditions de la *possibilité de l'expérience* en général sont en même temps conditions de la *possibilité des objets de l'expérience*, et ont de ce fait une validité objective dans un jugement synthétique *a priori*." ([281], III, 145, trad. A., J.-L. Delamarre, F. Marty)

Dans la section 7.3 du chapitre 2 nous avons évoqué le tournant analogue accompli par Galois devant l'impossibilité de généraliser aux degrés ≥ 5 les formules de "résolubilité par radicaux" des équations algébriques, formules connues depuis longtemps pour les degrés 2 (depuis l'Antiquité) et 3 (depuis Tartaglia et Cardan). Galois inverse la relation entre (i) les formules donnant les solutions d'un certain type d'équations et (ii) les conditions de possibilités d'une résolubilité d'un certain type. Il le fait en introduisant une structure (la structure de groupe de permutation des racines) permettant de formaliser ces conditions de possibilité. Le calcul des solutions (qui, de toute façon, pourrait se faire numériquement au moyen de méthodes d'approximations comme celle de Newton) devient alors subordonné à la théorie "structurale" des contraintes conditionnant les possibilités de résolubilité, contraintes devenant constitutives (au sens transcendantal) des solutions elles-mêmes.

Nous allons voir qu'il en va de même en théorie de l'intégrabilité des équations de Pfaff où la recherche de méthodes d'intégration explicite va devenir subordonnée à une analyse structurale géométrisant les conditions de possibilité de l'intégrabilité.

1. Rappelons la composition de la *Critique de la Raison pure* : I : Théorie des éléments / Deuxième partie : Logique transcendantale / Première division : Analytique transcendantale / Livre II : Analytique des principes / Chapitre II : Système de tous les principes de l'entendement pur / Troisième section : Du principe suprême de tous les jugements synthétiques.

1.2. La similitude des exemples

Mais avant que de nous lancer dans cette riche période de l’histoire de la géométrie différentielle, précisons ce qu’ont en commun ces différents exemples de “révolutions coperniciennes”.

1. Il y a des données et un problème à résoudre à propos de ces données.
2. Il y a des solutions “faibles” accessibles directement, c’est-à-dire sans restructuration des données.
3. On rencontre une *obstruction* : soit que les méthodes soient *ad hoc* et non généralisables, soit qu’elles ne soient pas satisfaisantes sur le plan de l’explication et de la compréhension théoriques.
4. On *inverse* alors le problème en se focalisant sur *les conditions de possibilité de la résolubilité* du problème. C’est la “révolution copernicienne”.
5. On découvre ensuite avec étonnement qu’il existe une structure propre, *sui generis* et universelle, de ces conditions de possibilité, structure *indépendante des données* et, en quelque sorte, *a priori* relativement à elles, le terme “a priori” n’ayant ici rien à voir avec une nécessité logique. En utilisant une expression venant de la physique, on peut dire qu’elle fonctionne comme une *background structure*, une “structure d’arrière fond”. C’est la découverte de cette structure *supplémentaire* — par définition “synthétique a priori” — qui fait de la “révolution copernicienne” une authentique révolution scientifique.
6. On revient alors aux données et on les restructure — on les “normalise” — à partir de ce supplément de structure et on relance à partir des données normalisées la procédure de résolution du problème considéré. Dans les cas les plus connus, on a découvert que la “background structure” synthétique a priori était *mathématisable* avec de nouvelles mathématiques *sui generis* et révolutionnaires et que la mathématisation des conditions de possibilité était d’une telle force qu’elle résolvait déjà presque à elle seule le problème.

1.3. L’exemple initial (Copernic, Kepler, Newton)

1. Les données sont les trajectoires très irrégulières des planètes comme “astres errants” et le problème à résoudre est celui de leur calcul et de leur explication théorique à partir de principes universels.
2. Les solutions “faibles” accessibles directement, c’est-à-dire sans restructuration des données, sont les modèles de Ptolémée (déjà remarquables) à base d’épicycles : ils sont fondés sur le principe universel que les mouvements premiers sont les mouvements circulaires uniformes et fournissent une méthode

d'approximations successives par décomposition du mouvement en "harmoniques", ce qui est quand même inouï puisque cela se retrouvera XVI siècles plus tard avec l'analyse de Fourier

3. L'obstruction est l'absence de lois générales explicatives permettant de réduire la contingence apparente de ces trajectoires empiriques "errantes" et "sans loi".
4. La "révolution copernicienne" est celle du changement de repère cosmologique global et le basculement vers un repère héliocentré *indépendant de l'observateur*.
5. Ce repère héliocentrique indépendant des données observationnelles fonctionne comme une background structure pour les mouvements planétaires.
6. Les données restructurées et "normalisées" issues des calculs de Kepler ne sont plus "errantes". Presque circulaires, elles conduisent à relancer la recherche d'une explication théorique à partir de mouvements circulaires uniformes. Kepler constate un écart et après des essais avec des cercles excentrés (où la position du centre est contingente et doit simplement s'adapter aux données) et des ovales il découvre (i) le modèle elliptique (où le centre n'est plus contingent puisqu'il est nécessairement l'un des foyers de l'ellipse mais où d'autres paramètres de l'ellipse comme l'excentricité et la taille sont contingents et doivent simplement s'adapter aux données) et (ii) les trois lois que Newton expliquera ensuite avec ses deux lois $f = m\gamma$ (loi du mouvement d'un corps de masse m soumis la force f , γ est l'accélération) et $F = G\frac{mm'}{r^2}$ (loi de la gravitation universelle : force attractive F exercée entre deux corps de masse m et m' séparés d'une distance de r , G est la constante gravitationnelle).

1.4. L'exemple de Kant et de la relativité galiléenne

1. Les données sont les *phénomènes* empiriques et le problème à résoudre à leur propos est celui de leur type d'objectivité et de leur explication comme *objets d'expérience possible* à partir de principes. La connaissance possible doit être réduite aux phénomènes (i.e. aux observables).
2. Les solutions "faibles", accessibles directement sans normalisation des phénomènes en objets d'expérience, sont triples : le réalisme ontologique dogmatique (les phénomènes manifestent une réalité en soi cognitivement accessible à travers eux), l'empirisme au sens du scepticisme anti-métaphysique de Hume (les phénomènes peuvent être décrits, classés, catégorisés et manifestent des régularités à partir desquelles on peut induire des lois empiriques probabilistes), le subjectivisme (les phénomènes ne sont que des apparences et toute connaissance est solipsiste). Toutes ces solutions ne tiennent pas compte du fait que le concept d'objet est *normatif* et qu'un phénomène n'est jamais

- donné comme objet et doit être *constitué* en tant que tel. Il n’est un objet d’expérience qu’une fois “déterminé” — normalisé — comme objet. En tant que phénomène, il reste “indéterminé”.
3. L’obstruction vient, d’une part, des contradictions internes du réalisme ontologique et du solipsisme subjectif et, d’autre part, des limites de l’empirisme qui ne peut pas satisfaire aux exigences de nécessité et d’universalité requises par une explication et une compréhension théoriques authentiques.
 4. On *inverse* alors le problème de la nature de l’objectivité en se focalisant sur les conditions de possibilité de l’expérience et sur la source de la normativité du concept d’objet. C’est la “révolution copernicienne” en métaphysique.
 5. On découvre alors qu’il existe une architecture de ces conditions de possibilité qui est universelle et indépendante des données. C’est la découverte fondamentale du *synthétique a priori* comme “background structure”.² Il existe un “format” de la manifestation des phénomènes sensibles, à savoir les “intuitions pures” que sont l’espace et le temps (l’Esthétique Transcendantale).³ Cette structure supplémentaire est mathématisable (géométrie de l’espace-temps) et doit s’articuler avec la structure complètement hétérogène qui est celle de la *logique* des jugements (la logique aristotélicienne à l’époque de Kant) de façon à ce que cette logique, dans son applicabilité à l’expérience, devienne *contrainte* par la background structure de l’expérience. En mécanique newtonienne, la relativité galiléenne est incompatible avec la logique car un énoncé comme “le corps K possède la vitesse v ” ne possède ni valeur de vérité logique ni contenu physique puisque par changement de repère inertial on peut assigner à v n’importe quelle valeur. Ces jugements violent le principe logique du tiers exclu (satisfait par les jugements normaux que Kant appelle pour cette raison “disjonctifs”, i.e. soit vrais, soit faux) et c’est pourquoi Kant a introduit pour en

2. Il pourra sembler étrange de se référer à Kant autrement que comme à un auteur classique dont l’œuvre est irrémédiablement dépassée puisque les critiques radicales, multiples et répétées, contre le synthétique a priori ont semblé avoir sonné le glas de l’actualité de toute approche transcendantale en philosophie des sciences. Pourtant, nous défendons depuis longtemps la thèse (cf. par exemple [419] et [436]) qu’il n’en est rien et que la philosophie transcendantale demeure bien au contraire la meilleure philosophie des sciences possibles dans la mesure où elle n’est rien d’autre que la métaphysique des groupes de relativité. Jules Vuillemin le disait déjà dans son ouvrage de 1955 *Physique et Métaphysique kantienne* [557] et, plus récemment, l’éminent kantien de l’Université de Stanford Michael Friedman l’a développé dans son vaste commentaire de 2013 *Kant’s Construction of Nature* [199] des *Metaphysische Anfangsgrunde der Naturwissenschaft*.

3. C’est ce que les neurosciences cognitives confirment, d’où le “kantisme” déclaré de spécialistes comme John O’Keefe ou Stanislas Dehaene évoqués plus haut.

tenir compte la notion de jugements "alternatifs".⁴ L'articulation se fait à travers la remarquable invention du *schématisme* transcendantal. Les catégories, qui sont un moyen terme entre types de jugements et déterminations d'objets, se schématisent grâce à l'esthétique transcendantale et se convertissent en "principes de l'entendement pur" comme les principes d'invariance ou le principe de causalité. Comme Kant l'explique dans les *Metaphysische Anfangsgründe der Naturwissenschaft*, dans la mécanique newtonienne l'esthétique transcendantale se traduit en particulier par la relativité galiléenne traitée dans la *Phoronomie* (Cinématique). Et effectivement les repères inertiels de l'espace-temps munis de leur groupe de Galilée constituent l'exemple princeps d'une "background structure". Il est essentiel de noter que la relativité galiléenne impose des limites indépassables à la connaissance possible. Un instant (une origine du temps), une position (origine de l'espace), une direction privilégiée ou une vitesse *absolus* sont inaccessibles. Mais cette inconnaissabilité intrinsèque possède une *structure*, précisément celle du groupe de Galilée, structure synthétique a priori supplémentaire qui convertit ce qu'il est impossible de connaître en un principe de détermination de ce qu'il est possible de connaître.⁵ Kant a été le premier à comprendre ce paradoxe et à articuler une logique des jugements avec cette structure d'objet qui lui est hétérogène. Il a ainsi ouvert une problématique cruciale restée longtemps complètement méconnue des philosophes des sciences. Un siècle et demi plus tard (1936),

4. Partie "Phénoménologie" des *Metaphysische Anfangsgründe der Naturwissenschaft*. Aujourd'hui, ces jugements dont la sémantique est indéfinie sont dits *contextuels*. Leurs dénotations et leurs valeurs de vérité sont de nature *pragmatique* et non pas sémantique car elles dépendent du contexte. L'énoncé contextualisé "le corps K possède la vitesse v relativement au repère inertiel R " possède, lui, une valeur de vérité bien définie. Il n'est plus "alternatif" mais "disjonctif". La relativité galiléenne concerne ces contextes qui représentent une composante "subjective" de l'expérience. Les repères inertiels dépendent de l'observateur mais les contenus physiques objectifs doivent en être indépendants. Ils doivent être *invariants* sous l'action du groupe de Galilée. La vitesse n'est pas un invariant et n'a pas de contenu physique mais l'accélération a un contenu physique, ce qu'exprime précisément la loi universelle $f = m\gamma$.

5. Comme nous l'avons souvent souligné (cf. par exemple [419]) cette conversion est mathématisée par le théorème fondamental qu'est le *théorème de Noether*. Dans la formulation lagrangienne de la mécanique où le comportement du système mécanique considéré est décrit par son lagrangien et les équations d'Euler-Lagrange (que nous rencontrerons nous aussi plusieurs fois en neurogéométrie), le lagrangien est invariant par le groupe de Galilée G et à chaque sous-groupe à 1-paramètre de G , est associée une grandeur invariante (l'énergie pour les translations temporelles, le moment pour les translations spatiales, le moment angulaire pour les rotations, le centre de masse pour les mouvements rectilignes uniformes). Nous y reviendrons dans les sections 3 et 6.2 du chapitre 12.

John von Neumann et Garrett Birkhoff retrouveront le problème avec la “logique quantique” qui articule la logique avec la géométrie des espaces de Hilbert qui lui est hétérogène.⁶ Les versions les plus contemporaines de cette relation profonde entre logique des prédicats et structures géométrique d’objets se rencontrent dans la *théorie des topos* inventée par Alexandre Grothendieck en géométrie algébrique au milieu des années 1960 et articulée peu après avec la logique par William Lawvere.

6. On revient aux données et on les restructure — on les “normalise” — à partir de cette architecture et on relance la procédure de résolution du problème considéré à partir de ces données normalisées. Kant en a donné lui-même l’exemple dans les *Metaphysische Anfangsgrunde der Naturwissenschaft* où il applique les principes transcendants de l’expérience à la mécanique de Newton. Une des grandes caractéristiques des théories physiques fondamentales exprimables par des lagrangiens, de la mécanique de Newton au modèle standard de la théorie quantique des champs (Glashow, Salam, Weinberg) en passant par la relativité générale, est d’avoir montré que la structure mathématique des groupes de relativité de plus en plus généraux (la mathématique du synthétique a priori) impliquait une part de plus en plus importante du contenu physique des théories (le synthétique a posteriori). Comme le disait Jean-Marie Souriau à propos de la quantification géométrique (cf. [509]) :

- “philosophiquement [la géométrisation] c’est ramener la physique à des symétries géométriques pour faire de la physique a priori”,
- “il n’y a rien de plus dans les théories physiques que les groupes de symétrie si ce n’est la construction mathématique qui permet précisément de montrer qu’il n’y a rien de plus”.

6. La mécanique quantique a réactivé de façon spectaculaire une problématique transcendantale (kantienne dans l’esprit mais évidemment non kantienne dans la lettre) car elle repose sur une réduction drastique des théories physiques aux observables (cf. Petitot [427]). Au niveau atomique, les trajectoires des électrons autour des noyaux ne sont pas des observables et l’on doit abandonner le modèle planétaire de Bohr (1913) où il existe des orbites électroniques stables (non radiatives) et quantifiées (ce qui est contradictoire avec l’électrodynamique classique selon laquelle ces trajectoires circulaires, puisqu’elles possèdent une accélération radiale, doivent être radiatives, dissiper de l’énergie, être instables et s’effondrer sur les noyaux). En revanche les raies spectrales atomiques sont observables et suivent les lois empiriques de Balmer-Rydberg. Dès son article fondateur de 1925, “Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen”, Werner Heisenberg “réinterprétera” la signification de la cinématique (phoronomie) en mécanique quantique en se restreignant aux raies spectrales observables et en montrant comment cela conduit à un calcul *non commutatif* très vite identifié par Max Born et Pascual Jordan (1925) comme un calcul matriciel. Cette *structure supplémentaire* qu’est l’algèbre matricielle est à l’origine de la “révolution copernicienne” quantique.

1.5. L'exemple de Galois

Revenons un peu plus en détail sur la théorie de Galois que nous avons déjà évoquée à la section 7.3.1 du chapitre 2. Elle est technique et subtile et apparemment hors sujet pour notre propos. Nous tenons néanmoins à en esquisser ici quelques éléments pour deux raisons. D'abord parce qu'ils sont extraordinairement instructifs⁷ et ensuite parce que cette théorie a servi d'inspiration pour la "théorie de Galois de l'intégrabilité des équations différentielles" qui se trouve au cœur de notre propos. Nous ne pourrions hélas présenter les choses que de façon assez imprécise et insuffisamment rigoureuse. La théorie de Galois peut se développer sur n'importe quel corps de base, mais nous ne parlerons ici que des corps de base K qui sont intermédiaires entre le corps des nombres rationnels \mathbb{Q} et le corps des nombres complexes \mathbb{C} et qui sont en plus de dimension finie comme espaces vectoriels sur \mathbb{Q} .⁸

1-a. Les données sont les équations algébriques

$$P(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0 = 0$$

où $P \in K[X]$ est un polynôme d'un certain degré n à coefficients dans un corps K ($K[X]$ est la notation standard pour l'anneau des polynômes à une indéterminée X sur K) et le problème le plus général est de trouver les solutions x de l'équation dans la clôture algébrique \overline{K} de K . On peut supposer que $P(x)$ commence par x^n car s'il commence par $a_n x^n$ on peut le diviser par a_n (on dit alors que $P(x)$ est "unitaire"). Le cas le plus classique était celui où K était le corps \mathbb{R} et \overline{K} le corps \mathbb{C} et l'on savait depuis longtemps (Raffaello Bombelli 1572, après Girolamo Cardano 1545) que certaines des solutions d'équations à coefficients réels étaient dans \mathbb{C} . On avait montré (théorème fondamental de l'algèbre de d'Alembert-Gauss) que P possède exactement n racines α_i (comptées avec leur multiplicité) sur \mathbb{C} et s'écrit donc $P(x) = \prod_{i=1}^{i=n} (x - \alpha_i)$ (on dit que $P(x)$ est "scindé" sur \mathbb{C}). Si l'on ne s'intéresse qu'aux valeurs des racines et pas à leur multiplicité, on peut supposer que toutes les racines sont simples (on dit alors que $P(x)$ est "séparable"), ce que nous supposerons dans la suite. Si l'on s'intéresse à la construction de racines irrationnelles, on peut supposer que $P(x)$ est "irréductible" sur \mathbb{Q} , autrement dit que toutes les racines α_i sont en dehors de \mathbb{Q} . Si l'on compare les deux formules donnant $P(x)$ on voit

7. Après avoir lu cette section, le lecteur intéressé pourra lire par exemple une réflexion comme celle de Jean-Jacques Szczeciniarz sur "The Mysterious Strength of the Galois Theory" [520] qui parcourt le chemin menant à la théorie de Galois de Grothendieck. Sous ses multiples formes la théorie de Galois est l'une des théories mathématiques les plus puissantes et les plus universelles.

8. C'est de la théorie de Galois en caractéristique 0. Il existe aussi des théories de Galois (plus compliquées) en caractéristique p sur les corps finis.

immédiatement que les coefficients sont les *fonctions symétriques élémentaires* des racines, $s_1 = \sum_{1 \leq i \leq n} \alpha_i$, $s_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} \alpha_{i_1} \cdots \alpha_{i_k}$, $s_n = \alpha_1 \cdots \alpha_n$ puisque $P(x) = \sum_{k=1}^{k=n} (-1)^k s_k x^{n-k}$. Lorsque $K = \mathbb{Q}$ (équations à coefficients rationnels), les α_i sont des nombres dits *algébriques* et, en adjoignant leur ensemble $\bar{\alpha}$ à \mathbb{Q} puis en prenant le corps engendré $\mathbb{Q}(\bar{\alpha}) = \mathbb{Q}(\alpha_1, \dots, \alpha_n)$, on obtient une extension “algébrique” $D_{\mathbb{Q}}P = L$ de \mathbb{Q} , appelée le “corps de décomposition” de $P(x)$. L est un espace vectoriel de dimension *finie* sur \mathbb{Q} et la dimension $\dim_{\mathbb{Q}} L$ (notée aussi $[\mathbb{Q} : K]$) est appelée le degré de l’extension. Chaque sous-extension $\mathbb{Q}(\alpha_i)$ de L , dite “simple” parce qu’engendrée *comme corps* par un seul élément α_i , est engendrée *comme espace vectoriel* sur \mathbb{Q} par les α_i^k , $k = 1, \dots, n-1$. On peut alors prendre de telles extensions comme nouveaux corps de base K et construire de la même façon d’autres extensions algébriques à partir de ces K . Si $P(x) \in K[X]$ a pour racines les α_i , $L = K(\bar{\alpha}) = D_K P$ s’obtient à partir de l’anneau de polynômes à n variables $K[X_1, \dots, X_n] = K[\bar{X}]$ en le quotientant par l’idéal \mathfrak{M} des polynômes $Q(\bar{X})$ s’annulant au point $\bar{\alpha}$. \mathfrak{M} est le noyau du morphisme d’anneaux $\varphi : K[\bar{X}] \rightarrow K(\bar{\alpha})$, $X_i \mapsto \alpha_i$ et comme $K(\bar{\alpha})$ est un corps, \mathfrak{M} est un idéal maximal. On obtient ainsi des extensions “galoisiennes”⁹. Si \mathfrak{S}_n est le groupe des permutations de n objets considéré comme agissant sur les (α_i) , le groupe de Galois G de $\bar{\alpha}$ sur K est le sous-groupe des $\sigma \in \mathfrak{S}_n$ stabilisant \mathfrak{M} (cf. plus bas).

- 1-b. En ce qui concerne le problème général de la résolution des équations, il existait un spectre de méthodes. D’abord à un extrême il existait, et pour tout degré, des méthodes numériques permettant de trouver par approximations successives des valeurs numériques aussi précises que l’on voulait des α_i . La plus connue était celle de Newton. À l’autre extrême ; il y avait pour les bas degrés, nous l’avons déjà dit plusieurs fois (chapitre 2, section 7.3.1), les solutions algébriques données par des formules symboliques connues depuis très longtemps, depuis les babyloniens pour l’équation du second degré, depuis, répétons-le, Niccolo Tartaglia (1500-1557) et Girolamo Cardano (1501-1576) pour le troisième degré, depuis Ludovico Ferrari (1522-1565), disciple de Cardano, pour le quatrième degré. Ces formules sont des formules de résolution “par radicaux” qui effectuent sur les coefficients des opérations algébriques standard et des extractions de racines $\sqrt[k]{}$. Et comme des formules symboliques exactes sont très supérieures en termes d’explication et de compréhension à

9. Une extension L/K est dite galoisienne si lorsqu’un $P(x)$ est irréductible sur K et a une racine dans L il a toutes ses racines dans L , i.e. $P(x)$ est scindé dans L . Par exemple $x^3 - 2$ est irréductible sur \mathbb{Q} et $L = \mathbb{Q}(\sqrt[3]{2})$ n’est pas galoisienne car les autres racines $j\sqrt[3]{2}$ et $j^2\sqrt[3]{2}$ (où j et j^2 sont les racines cubiques de l’unité) ne sont pas dans L . En revanche $M = \mathbb{Q}(\sqrt[3]{2}, j)$ est galoisienne.

de simples méthodes numériques, le problème de base devint par conséquent celui de la résolution par radicaux des équations algébriques. Il s'exprime en se demandant si l'extension algébrique $L = K(\bar{\alpha}) = D_K P$ de K est "radicale", i.e. peut s'obtenir au moyen d'une suite d'extensions emboîtées $K_0 = K \subset \dots \subset K_\ell \subset \dots \subset K_p = K(\bar{\alpha}) = L$ telle que, à chaque étape, on ait $K_\ell = K_{\ell-1}(\rho_\ell)$ où ρ_ℓ est racine d'ordre ν_ℓ dans $K_{\ell-1}$ (i.e. il existe un $r_{\ell-1} \in K_{\ell-1}$ tel que $\rho_\ell^{\nu_\ell} = r_{\ell-1}$). C'est la "résolubilité par radicaux" de L . Comme $\rho^\nu = r$ signifie $|\rho|^\nu e^{\nu \arg(\rho)} = |r| e^{\arg(r)}$ c'est-à-dire $\rho = \sqrt[\nu]{|r|} e^{\frac{\arg(r)}{\nu}}$, ρ est défini à $e^{i\frac{2\pi}{\nu}}$ près. Les racines ν -ème font donc intervenir les racines ν -ème de l'unité $\zeta = e^{i\frac{2\pi\mu}{\nu}}$ avec $\mu = 0, \dots, \nu - 1$ (ce sont dans le plan \mathbb{C} les sommets du polygone régulier de ν côtés de rayon 1 avec un premier sommet en 1) et les extensions $\mathbb{Q}(\zeta)$ sont les *corps cyclotomiques* déjà utilisés par Gauss et théorisés en tant que tels par Ernst Kummer (1810-1893) pour résoudre partiellement le théorème de Fermat (1847).¹⁰

- 2-a. Avant Abel et Galois il y avait des solutions "faibles" partielles et accessibles assez directement (bien que parfois techniques et ingénieuses) mais elles n'étaient pas subordonnées à la thématization, la théorisation et la mathématisation de conditions de possibilité de résolubilité. On disposait de nombreux outils pour les obtenir.
- On savait faire des changements de variables et transformer, réduire, décomposer les polynômes pour simplifier le problème. On connaissait les polynômes irréductibles sur K (i.e. indécomposables en produits de plusieurs polynômes à racines dans K).
 - On avait commencé à analyser la structure des fonctions symétriques de n variables, c'est-à-dire les $f(x_1, \dots, x_n)$ invariantes par l'action des permutations σ des x_i et à comprendre le rôle des permutations des racines. On savait (Gauss) que tout polynôme symétrique était un polynôme en les fonctions symétriques s_k .
 - On connaissait les "discriminants" $\Delta = \delta^2$ avec $\delta = \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j)$. Δ est un polynôme symétrique des racines, donc des s_k , et par conséquent est un élément du corps de base K . Ces discriminants deviennent vite très compliqués. Par exemple, pour le polynôme général du 4ème degré $ax^4 + bx^3 + cx^2 + dx + e$, on trouve

$$\begin{aligned} \Delta = & 256a^3e^3 - 128a^2e^2c^2 - 4b^3d^3 + 16ac^4e - 4ac^3d^2 - 192a^2bde^2 \\ & - 27b^4e^2 - 6ab^2d^2e + 144ab^2ce^2 + 144a^2cd^2e - 80abc^2de \\ & + 18b^3cde + 18abcd^3 + b^2c^2d^2 - 4b^2c^3e - 27a^2d^4 . \end{aligned}$$

10. Le lecteur intéressé par la démonstration du théorème de Fermat par Andrew Wiles pourra consulter ma présentation pédagogique [421].

– En 1770, Joseph-Louis Lagrange (1736-1802) et Alexandre-Théophile Vandermonde (1735-1796) avaient expliqué pourquoi le troisième degré pouvait se ramener au second degré et le quatrième au troisième par la considération d'équations auxiliaires appelées “résolvantes”. On avait donc cherché à généraliser cette possibilité à tout degré avec les “résolvantes de Lagrange”. Nous allons y revenir plus bas.

- 2-b. L'idée des résolvantes est simple et profonde. Dans une équation algébrique $P(x) = 0$ on connaît les coefficients a_i mais on ne connaît pas les racines α_j et l'on cherche des formules permettant de les calculer à partir des coefficients. On cherche donc à calculer ce que l'on ne connaît pas à partir de ce que l'on connaît. C'est déjà l'amorce d'une “révolution copernicienne” car si l'obstruction définissant ce qu'on ne connaît pas possède une structure *a priori*, alors l'exploration de cette structure peut permettre de mieux déterminer ce qu'on ne connaît pas et par conséquent ce qu'on peut connaître. L'outil de base est l'utilisation des *symétries*. Les fonctions symétriques $f(\bar{x})$ sont invariantes par les permutations $\sigma \in \mathfrak{S}_n$ des n variables x_i . Leurs valeurs $f(\bar{\alpha})$ sur les racines de $P(x)$ sont donc des fonctions des s_i et elles sont par conséquent connaissables directement à partir des coefficients puisque $(-1)^k s_k = a_{n-k}$. On peut alors considérer des fonctions $g(\bar{x})$ “presque” symétriques mais pas complètement. En appliquant les permutations $\sigma \in \mathfrak{S}_n$ on obtiendra m fonctions $g_k(\bar{x})$ et le produit $\Pi_g(x) = \prod_{k=1}^{k=m} (x - g_k(\bar{\alpha}))$ sera un polynôme de x symétrique en les $\bar{\alpha}$ dont les coefficients seront par conséquent connus à partir de ceux de $P(x)$. Si $m < n$ on aura ainsi réduit le degré des polynômes à résoudre et l'on peut espérer trouver une “tour descendante” de polynômes auxiliaires $\Pi_f(x)$ de degré strictement décroissant permettant de calculer les α_i par une tour de formules emboîtées. Mais pour comprendre ces procédures de “descente” il faut comprendre la *structure* de \mathfrak{S}_n . C'est en effet cette structure, qui se révélera être celle de groupe, qui constitue la “background structure” synthétique a priori fournissant une connaissance a priori sur la résolubilité du problème. Alors que les $f(\bar{x})$ sont \mathfrak{S}_n -invariantes, les $g(\bar{x})$ sont seulement H -invariantes pour des sous-groupes distingués¹¹ H de \mathfrak{S}_n et les $g_k(\bar{x})$ en sont les orbites isomorphes aux quotients \mathfrak{S}_n/H (d'ordre m).
3. On rencontra effectivement une *obstruction* qui conduisit à inverser la perspective : malgré de nombreux efforts, la méthode des résolvantes ne paraissait pas être généralisable. C'est Paolo Ruffini (1765-1822) qui donna le premier des arguments pour montrer qu'en fait on ne pouvait pas résoudre par radicaux l'équation générale de degré 5.

11. Rappelons qu'un sous-groupe H d'un groupe G est dit “normal” (ou “distingué”) s'il est stable par conjugaisons : pour tout $g \in G$, $gHg^{-1} = H$. Si G est commutatif, tout sous-groupe est normal.

- 4-a. On inversa alors le problème en se focalisant sur les conditions de possibilité de la résolubilité par radicaux. Ce fut la "révolution copernicienne" effectuée en 1824-1826 par Niels Henrik Abel (1802-1829) suivi en 1829-1830 par Évariste Galois (1811-1832) qui montra la solidarité fondamentale entre la résolubilité d'une équation par radicaux et la structure de son groupe de permutations des racines (groupe de Galois).¹² Ce changement de perspective est devenu légendaire et a été beaucoup commenté. Il consiste à passer d'une recherche de formules donnant les racines au groupe qui mesure le degré d'indiscernabilité des racines. Ce n'est plus l'équation donnée qui domine mais la théorie des groupes permettant de lire sur son groupe de Galois les conditions de possibilité de sa résolubilité par radicaux. La structure de groupe devient la connaissance synthétique a priori, indépendante des équations données, mais conditionnant leur résolubilité.¹³
- 4-b. Le groupe de Galois G de l'équation peut se définir de plusieurs façons. On part du groupe de permutation \mathfrak{S}_n de n objets et on le fait opérer comme groupe de permutations des variables et des racines. En termes de l'idéal maximal \mathfrak{M} de $K[\bar{X}]$ (cf. plus haut), \mathfrak{S}_n opère sur $K[\bar{X}]$ par permutations des X_i dans les polynômes $Q(\bar{X})$ et G est le sous-groupe de \mathfrak{S}_n laissant \mathfrak{M} globalement stable. En termes de l'extension $K \subset K(\bar{\alpha}) \subset \mathbb{C}$, G est le groupe des automorphismes de $K(\bar{\alpha})$ laissant K fixe. Si Σ est un tel automorphisme, il échange les α_i et si l'on pose $\Sigma(\alpha_i) = \alpha_{\sigma(i)}$, σ est une permutation et donc un élément de \mathfrak{S}_n . Le théorème de Galois dit que K est exactement le sous-corps $K(\bar{\alpha})^G$ de $K(\bar{\alpha})$ fixé par G . On voit ainsi s'introduire une idée fondamentale, celle d'une dialectique entre *généricité* et *spécialisation*. Quand on considère $K(\bar{\alpha}) = D_K P$, on considère les racines α_i comme des nombres. On considère un polynôme $P(x)$ d'une variable x qui prend des valeurs $P(\alpha)$ lorsque l'on spécialise x en α et l'on cherche les racines α annulant $P(\alpha)$. Mais l'on peut aussi considérer n variables indépendantes X_i et des polynômes génériques $R(\bar{X}) \in K[\bar{X}]$ que l'on spécialise en $R(\bar{\alpha})$. Nous allons voir tout le bénéfice que l'on peut en tirer.
- 4-c. Le problème fondamental que l'on cherche à résoudre, celui de la résolubilité par radicaux de l'extension $L = K(\bar{\alpha}) = D_K P$ de K , va alors être posé en termes de la structure du groupe de Galois. Cela est justifié car les résolvantes de Lagrange reposent en fait sur des propriétés de structure du groupe de Galois (cf. plus bas).

12. Les écrits de Galois présentés à l'Académie des sciences en 1829 sont perdus. Son article de 1830 a été redécouvert par Joseph Liouville qui l'a présenté à l'Académie en 1843.

13. Henri Poincaré a souvent insisté sur le caractère a priori du concept de groupe, entre autres, dans *La Science et l'Hypothèse* [450].

- 5-a. Galois invente ainsi la théorie abstraite des groupes comme une théorie universelle des conditions de possibilité de résolubilité par radicaux qui est indépendante de toute équation particulière et opère comme une “structure d’arrière fond” a priori relative aux équations. La découverte fondamentale fut que la structure abstraite supplémentaire de groupe permettait de caractériser la résolubilité par radicaux de l’équation par la propriété que son groupe soit “résoluble”. Elle a fait de la “révolution copernicienne” galoisienne une révolution mathématique.
- 5-b. Le résultat fondamental de la théorie de Galois est que si L/K est une extension galoisienne de dimension finie $d = \dim_K L = [L : K]$ sur K , alors
- (i) le groupe de Galois G est fini et $K = L^G$;
 - (ii) le cardinal $\#G$ de G est égal à d ;
 - (iii) les extensions intermédiaires $K \subset M \subset L$ sont en bijection (correspondance dite “de Galois”) avec les sous-groupes H de G : à M on associe $H = \text{Gal}(L/M)$ (L/M est évidemment galoisienne) et à H on associe $M = L^H$;
 - (iv) M est galoisienne si et seulement si H est un sous-groupe distingué de G et alors $\text{Gal}(M/K) = G/H$;
 - (v) si $P(x)$ est scindé sur L , G opère *transitivement* sur les racines si et seulement si $P(x)$ est irréductible sur K .
- 5-c. Le lien entre deux façons de concevoir la théorie, celle originale de Galois (qui repose sur les permutations de racines de polynômes) et celle devenue ensuite standard, surtout grâce à Emil Artin (1898-1962) (qui repose sur les automorphismes des extensions : $G = \text{Aut}_K L$), se fait au moyen du théorème d’Abel-Galois de l’*élément primitif* (1832), cette notion venant de Lagrange. Soit $L = K(\alpha_1, \dots, \alpha_n)$ une extension galoisienne (par exemple $L = K(\bar{\alpha}) = D_K P$ venant d’un polynôme irréductible $P(x)$). Alors elle est “simple”, autrement dit engendrée par un seul élément $\xi \in L : L = K(\xi)$ et le polynôme minimal $Q(x)$ définissant ξ est de degré d . La correspondance $\sigma \in G \mapsto \sigma(\xi)$ est une bijection entre G et son orbite $G(\xi)$ dans L , tous les éléments de l’orbite (les racines de $Q(x)$) étant conjugués par G . Qui plus est, l’orbite $G(\xi)$ fournit une base de L comme espace vectoriel sur K .
- 5-d. Nous avons vu que les extensions avec radicaux sont liées aux racines de l’unité. On peut montrer une sorte de réciproque. Soit L/K une extension galoisienne de groupe de Galois cyclique $G = \mathbb{Z}/p\mathbb{Z}$, p premier. Alors si K contient une (et donc toute) racine primitive p -ème de l’unité ζ , $L = K(\alpha)$, avec α une racine p -ème de K , i.e. $\alpha^p = a \in K$. Cela implique que toute L/K de groupe de Galois $G = \mathbb{Z}/p\mathbb{Z}$ soit résoluble. Soit σ un générateur de

G . Il suffit de construire un $\alpha \in L$, $\alpha \notin K$, tel que $\sigma(\alpha) = \zeta\alpha$ (cela exige évidemment $\alpha \notin K$ car si $\alpha \in K$, $\sigma(\alpha) = \alpha$ et $\alpha = \zeta\alpha$ implique $\alpha = 0$). En effet on aura alors $\sigma(\alpha^p) = \sigma(\alpha)^p = \zeta^p\alpha^p = \alpha^p$ et donc, α^p étant fixé par σ (et donc par G), il est dans K : $\alpha^p = a \in K$. Comme $\alpha \notin K$ et comme $K(\alpha)$ est de degré p , $K(\alpha) = L$ puisque L/K étant galoisienne on a $[L : K] = p$. Le plus simple est de prendre un α donné par une résolvante de Lagrange $R(x, \zeta) = \sum_{k=0}^{k=p-1} \zeta^{-k} \sigma^k(x)$ car on a automatiquement $\sigma(\alpha) = \zeta\alpha$. Il faut seulement trouver un x garantissant que $\alpha \notin K$. Pour cela, on part d'un générateur γ de L/K , on considère $\beta = \sum_{k=0}^{k=p-1} \sigma^k(\gamma)$. Comme $\sigma(\beta) = \beta$, $\beta \in K$. Si l'on prend alors $x = \gamma - \frac{\beta}{p}$, on a

$$R(x, 1) = \sum_{k=0}^{k=p-1} \sigma^k(x) = \left(\sum_{k=0}^{k=p-1} \sigma^k(\gamma) \right) - \beta = 0$$

et x est lui aussi un générateur de L/K . Il y a alors nécessairement un ζ telle que $R(x, \zeta) \neq 0$ car si les p sommes $R(x, \zeta)$ étaient $= 0$, on aurait $x = 0$ car le déterminant de ce système d'équations est non nul.

- 5-e. Le résultat le plus significatif est que $G \subset \mathfrak{S}_n$ est *résoluble* si et seulement si il existe une suite de sous-groupes distingués (i.e. normaux) $G_0 = \{\text{Id}\} \subset \dots \subset G_r \subset \dots \subset G_s = G$ telle que les quotients successifs soient *abéliens* (commutatifs).¹⁴ Une conséquence immédiate de cette définition est que si G est *simple* (i.e. sans sous-groupe normal non trivial) et *non commutatif* il ne peut pas être résoluble. La résolubilité a donc à voir avec la structure non commutative de G . On appelle groupe dérivé de G et on note $D(G)$ le quotient de G par le sous-groupe normal¹⁵ $[G, G]$ engendré par ses commutateurs (si $g, h \in G$, $[g, h] = ghg^{-1}h^{-1}$). Le fait que G soit résoluble est équivalent au fait que la suite des groupes dérivés de G aboutit au groupe trivial $\{\text{Id}\}$.
- 5-f. C'est alors la structure des groupes (dits "symétriques") de permutations \mathfrak{S}_n — autrement dit les propriétés du supplément de structure formalisant les conditions de possibilité de la résolubilité — qui devient la clé de la solution du problème. Leur théorie fournit un très riche ensemble de contraintes synthétiques a priori qui conditionnent les méthodes de résolubilité. Il est très instructif de s'y arrêter. Le groupe \mathfrak{S}_n est de cardinal $n!$ et est engendré par

14. Nous reviendrons en détail sur cette notion à la section 2 du chapitre 7.
 15. Répétons que si H est un sous-groupe d'un groupe G , le quotient G/H est l'ensemble des classes d'équivalences gH . Pour qu'il soit bien défini comme *groupe* il faut et il suffit que, si $gh \in gH$ et $g'h' \in g'H$, alors il existe $h'' \in H$ tel que $gh.g'h' = gg'h''$. Si G est commutatif $gh.g'h' = gg'.hh'$ et $h'' = hh'$ convient. Mais si G n'est pas commutatif, il faut pouvoir écrire $gh.g'h' (= g.hg'.h') = g.g'h'''.h'$ et donc $hg' = g'h'''$, autrement dit $h = g'h'''(g')^{-1}$, ce que garantit le fait que H soit normal. En effet nous avons vu que H est dit normal s'il est stable par conjugaison, autrement dit si tous les conjugués ghg^{-1} d'un élément h de H sont dans H .

les $n - 1$ transpositions $\tau_i = (i, i + 1)$, $i = 1, \dots, n - 1$ avec les $\frac{n(n-1)}{2}$ relations $\tau_i^2 = \text{Id}$, $\tau_i\tau_j = \tau_j\tau_i$ (commutation) si l'écart entre i et j est au moins 2, $(\tau_i\tau_{i+1})^3 = \text{Id}$ ou $\tau_i\tau_{i+1}\tau_i = \tau_{i+1}\tau_i\tau_{i+1}$ (relation de tresse). \mathfrak{S}_n est également engendré par la transposition τ_1 et le n -cycle $\gamma = (1, 2, \dots, n)$. Un sous-groupe distingué d'indice 2 dans \mathfrak{S}_n est le groupe, dit “alterné”, \mathfrak{A}_n des permutations paires, i.e. réalisables comme une composition d'un nombre pair de transpositions. Le quotient $\mathfrak{S}_n/\mathfrak{A}_n$ est le groupe d'ordre 2 et \mathfrak{A}_n est le groupe dérivé de \mathfrak{S}_n . Or il est *simple* si $n \neq 4$ et pour $n \geq 5$ il est *non abélien*. Le résultat clé qui résout le problème de la résolubilité par radicaux est alors le théorème :

Théorème. Pour $n \geq 5$, \mathfrak{A}_n est l'unique sous-groupe distingué de \mathfrak{S}_n et, comme \mathfrak{A}_n est simple et non abélien, \mathfrak{S}_n et \mathfrak{A}_n ne sont pas résolubles. _____◇

- La simplicité de \mathfrak{A}_5 (puis de \mathfrak{A}_n pour $n \geq 5$) se démontre de la façon suivante. Comme \mathfrak{S}_5 est d'ordre $5! = 120$ et \mathfrak{A}_5 d'indice 2, \mathfrak{A}_5 est d'ordre $\frac{120}{2} = 60 = 2^2 \cdot 3 \cdot 5$. Les p -cycles sont donc des 2-cycles (transpositions), des 3-cycles ou des 5-cycles puisque p doit diviser 60. Pour les 2-cycles (qui sont des produits de deux transpositions à supports disjoints car il faut un nombre pair de transpositions) on en trouve 15. Pour les 3-cycles on choisit 2 points qui restent fixes (10 possibilités) et un sens de permutation circulaire sur les 3 points qui restent, d'où 20 3-cycles. Pour les 5-cycles, on en trouve $4 \cdot 3 \cdot 2 = 24$. D'où les $1 + 15 + 20 + 24 = 60$ éléments. On démontre ensuite que tous les 2-cycles, 3-cycles et 5-cycles sont *conjugués* entre eux. Soit alors H un sous-groupe *normal*, $H \neq \{1\}$, de \mathfrak{A}_5 . S'il ne contient que des 2-cycles (resp. que des 3-cycles, resp. que des 5-cycles) il serait de cardinal $1 + 15 = 16$ (resp. $1 + 20 = 21$, resp. $1 + 24 = 25$) mais 16 (resp. 21, resp. 25) ne divise pas 60. Donc H doit contenir au moins deux types de cycles et son ordre est donc $\geq 1 + 15 + 20 = 36$ et par conséquent H est nécessairement d'ordre 60 puisque son ordre doit diviser 60. Ainsi, $H = \mathfrak{A}_5$ et \mathfrak{A}_5 est simple. On ne saurait trop méditer sur le fait que, une fois identifiée la structure de groupe de permutation comme structure supplémentaire d'arrière-fond, un théorème général d'irrésolubilité par radicaux qui était un véritable casse-tête pour les plus grands mathématiciens se ramène en définitive à des calculs énumératifs élémentaires.
- En revanche, \mathfrak{S}_2 est abélien, donc résoluble, et \mathfrak{S}_3 et \mathfrak{S}_4 sont résolubles, ce qui explique l'existence des formules trouvées pour les équations de degré 2, 3 et 4. \mathfrak{S}_3 est le groupe de symétrie du triangle équilatéral. Il comprend les rotations de $\frac{2\pi}{3}$ et $\frac{4\pi}{3}$ (les trois racines cubiques de l'unité ou encore les permutations circulaires) $\{1, j = \sigma_1 = (1, 2, 3), j^2 = \sigma_2 = (1, 3, 2)\}$ ainsi que les trois réflexions par rapport aux hauteurs-médiatrices (transpositions)

$$\{\tau_1 = (2, 3), \tau_2 = (3, 1), \tau_3 = (1, 2)\}$$

qui donnent trois sous-groupes d'ordre 2 $\{1, \tau_j\}$. La loi de groupe est dérivable des relations évidentes $\tau_k^2 = 1$, $\sigma_i^3 = 1$ ou $\sigma_1^2 = \sigma_2$ ou $\sigma_1\sigma_2 = 1$, $\tau_1\tau_2 = \sigma_2$, $\tau_2\tau_1 = \sigma_1$, $\tau_1\tau_3 = \sigma_1$, $\tau_3\tau_1 = \sigma_2$. Le sous-groupe alterné \mathfrak{A}_3 est le groupe cyclique d'ordre 3 $\{1, j, j^2\} = \{1, \sigma_1, \sigma_2\}$. \mathfrak{S}_3 est résoluble puisque $D(\mathfrak{S}_3) = \mathfrak{A}_3$ qui est abélien.

- 5-g. Comme nous l'avons dit plus haut, les extensions contenant des racines ν -ème de K sont liées aux extensions $L = K(\zeta)$ où ζ est une racine primitive ν -ème de l'unité (qui n'est pas dans K).¹⁶ $K(\zeta)$ est galoisienne et son groupe de Galois est un sous-groupe du groupe multiplicatif $(\mathbb{Z}/\nu\mathbb{Z})^*$. En effet, si $\sigma \in G$, alors $\sigma(\zeta) = \zeta^{\mu(\sigma)}$ avec $\mu(\sigma) \in (\mathbb{Z}/\nu\mathbb{Z})^*$ et

$$(\tau \circ \sigma)(\zeta) = \zeta^{\mu(\tau \circ \sigma)} = (\zeta^{\mu(\sigma)})^{\mu(\tau)} = \zeta^{\mu(\sigma)\mu(\tau)}.$$

Le morphisme $\mu : G \rightarrow (\mathbb{Z}/\nu\mathbb{Z})^*$ est injectif car ζ est primitive et donc si $\mu(\sigma) = 1$, $\sigma(\zeta) = \zeta$ donc $\sigma(\zeta^\mu) = \zeta^\mu$ pour tout μ et σ fixe tous les éléments de $L = K(\zeta)$ et par conséquent $\sigma = \text{Id}$. Si maintenant $L = K(\alpha)$ avec α une racine ν -ème de $a \in K$, alors, si K contient une racine primitive ν -ème de l'unité ζ , l'extension L/K est galoisienne et son groupe de Galois est un sous groupe du groupe additif $\mathbb{Z}/\nu\mathbb{Z}$. En effet les racines du polynôme $(X^\nu - a) \in K[X]$ définissant α sont les $\zeta\alpha$ et si $\zeta \in K$ elles sont alors toutes dans $L = K(\zeta)$ ce qui montre que L/K est bien galoisienne. Si $\sigma \in G$, $\sigma(\alpha)^\nu = \sigma(\alpha^\nu) = \sigma(a) = a$ (car les σ fixent les $a \in K$) et donc $\sigma(\alpha) = \zeta^{\mu(\sigma)}\alpha$. On a alors

$$(\tau \circ \sigma)(\alpha) = \zeta^{\mu(\tau \circ \sigma)}\alpha = \tau(\zeta^{\mu(\sigma)}\alpha) = \zeta^{\mu(\sigma)}\tau(\alpha)$$

car $\zeta^{\mu(\sigma)} \in K$ et donc

$$\zeta^{\mu(\tau \circ \sigma)}\alpha = \zeta^{\mu(\sigma)}\zeta^{\mu(\tau)}\alpha = \zeta^{\mu(\sigma)+\mu(\tau)}\alpha,$$

ce qui montre que $\mu(\tau \circ \sigma) = \mu(\sigma) + \mu(\tau)$ est bien un morphisme additif de G dans $\mathbb{Z}/\nu\mathbb{Z}$. Il est injectif car si $\mu(\sigma) = 0$, alors $\sigma(\alpha) = \zeta^{\mu(\sigma)}\alpha = \alpha$ et $\sigma = \text{Id}$.

6. On revient alors aux données et on relance la procédure de résolution par radicaux des équations à partir de leur groupe de Galois. Et l'on découvre que la structure même du groupe de Galois (le synthétique a priori) résout déjà une grande partie du problème de la résolution. Nous allons considérer le cas historique de l'équation du troisième degré (Tartaglia, Cardano) tel qu'il est exposé dans l'excellent texte pédagogique de Daniel Perrin [406]. \mathfrak{S}_3 est d'ordre $3! = 6$ et ses sous-groupes distingués sont le groupe alterné \mathfrak{A}_3 cyclique d'ordre 3 et d'indice 2 et les 3 groupes $\{1, \tau\}$ des transpositions (cf.

16. $\zeta \neq 1$ est primitive si ζ est d'ordre exactement ν . Par exemple pour $\nu = 4$, $\zeta = -1$ n'est pas primitive car elle est d'ordre 2 mais $\zeta = i$ est primitive. ζ n'est pas primitive si $\zeta^\mu = 1$ avec μ un diviseur de ν . Si ν est premier, toute $\zeta \neq 1$ est primitive.

plus haut). \mathfrak{S}_3 est résoluble. Si $P(y) = y^3 + ay^2 + by + c$, ($a, b, c \in K$), on peut éliminer le terme en y^2 par le changement de variable $x = y + \frac{a}{3}$ et l'on obtient $P(x) = x^3 + px + q$ avec $p = b - \frac{a^2}{3}$, $q = \frac{2a^3}{27} - \frac{ab}{3} + c$. Les fonctions symétriques élémentaires des racines sont $\alpha_1 + \alpha_2 + \alpha_3 = 0$, $\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1 = p$, $\alpha_1\alpha_2\alpha_3 = -q$. Supposons $P(x)$ irréductible sur K et soit $L = D_K P$ l'extension galoisienne associée. Le discriminant est $\Delta = -(4p^3 + 27q^2)$. Si Δ n'est pas un carré dans K , $\dim_K L = [L : K]$ est 6 et donc $G \simeq \mathfrak{S}_3$. Si en revanche Δ est un carré dans K , $[L : K] = 3$ et $G \simeq \mathfrak{A}_3$. Le corps fixé par \mathfrak{A}_3 est $K(\sqrt{\Delta}) = M$ et L est de degré 3 sur M . Si $\sigma \in G$ est la permutation circulaire $(1, 2, 3)$, on a dans les deux cas $\sigma \in G$. On suppose alors que K contient les racines cubiques de l'unité et on cherche une résolvante de Lagrange $\rho \in L$, $\rho \notin M$ telle que $\rho^3 = r \in M$ qui soit de la forme $\rho = \lambda\alpha_1 + \mu\alpha_2 + \nu\alpha_3$. On a $\sigma(\rho)^3 = \sigma(\rho^3) = \sigma(r) = r = \rho^3$ puisque r est fixé par $\sigma \in G$ et donc, comme $\sigma(\rho) \neq \rho$, on a $\sigma(\rho) = j\rho$ ou $\sigma(\rho) = j^2\rho$. On peut prendre par exemple $\rho = \alpha_1 + j\alpha_2 + j^2\alpha_3$ pour laquelle $\sigma(\rho) = j^2\rho$. On peut considérer également la seconde résolvante $\xi = \alpha_1 + j^2\alpha_2 + j\alpha_3$ qui est la transposée de ρ par la transposition $\tau = (2, 3)$ (la τ_1 ci-dessus). On a également $\xi^3 \in M$. Si donc on connaît ρ et ξ on peut immédiatement calculer les α_i avec le système d'équations $\alpha_1 + \alpha_2 + \alpha_3 = 0$, $\alpha_1 + j\alpha_2 + j^2\alpha_3 = \rho$, $\alpha_1 + j^2\alpha_2 + j\alpha_3 = \xi$ de déterminant $(j^2 - j) \neq 0$. On trouve $\alpha_1 = \frac{\rho + \xi}{3}$, $\alpha_2 = \frac{j^2\rho + j\xi}{3}$, $\alpha_3 = \frac{j\rho + j^2\xi}{3}$. Pour calculer ρ et ξ on utilise le fait que $\rho^3 + \xi^3$ et $\rho^3\xi^3$ sont des fonctions symétriques des α_i et donc solutions d'une équation du second degré dont les coefficients sont des polynômes en les coefficients p et q de $P(x)$. En fait $\rho\xi$ est déjà une fonction symétrique puisque

$$\begin{aligned} \rho\xi &= (\alpha_1 + j\alpha_2 + j^2\alpha_3)(\alpha_1 + j^2\alpha_2 + j\alpha_3) \\ &= \alpha_1^2 + \alpha_2^2 + \alpha_3^2 - (\alpha_2\alpha_3 + \alpha_3\alpha_1 + \alpha_1\alpha_2) \\ &= s_1^2 - 3s_2 = -3p . \end{aligned}$$

– Donc $\rho^3\xi^3 = -27p^3$. Pour calculer $\rho^3 + \xi^3$ on part de

$$(\rho + \xi)^3 = \rho^3 + \xi^3 + 3\rho\xi(\rho + \xi) .$$

Mais nous venons de voir que $\rho + \xi = 3\alpha_1$ et donc

$$\rho^3 + \xi^3 = 27\alpha_1^3 - 3(-3p)(3\alpha_1) = 27(\alpha_1^3 + p\alpha_1) .$$

Mais comme $\alpha_1^3 + p\alpha_1 + q = 0$ par hypothèse, $\rho^3 + \xi^3 = -27q$. Par conséquent, ρ^3 et ξ^3 sont solutions de l'équation $X^2 + 27qX - 27p^3$ dont le discriminant est

$$\tilde{\Delta} = (27q)^2 - 4(-27p^3) = 27(27q^2 + 4p^3) = -27\Delta .$$

D'où $\rho^3, \xi^3 = \frac{1}{2}(-27q \pm 3\sqrt{-3\Delta})$. On retrouve ainsi la célèbre formule de Cardano,

$$\begin{aligned}\alpha_1 &= \frac{\rho + \xi}{3} = \frac{1}{3} \left(\sqrt[3]{\frac{1}{2}(-27q + 3\sqrt{-3\Delta})} + \sqrt[3]{\frac{1}{2}(-27q - 3\sqrt{-3\Delta})} \right) \\ &= \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}}\end{aligned}$$

et de même pour les deux autres racines faisant intervenir j et j^2 , les racines devant être choisies de façon à satisfaire la contrainte $\rho\xi = -3p$.

7. L'équation quintique générale $P(x) = 0$ n'étant pas résoluble par radicaux, deux problèmes se posèrent immédiatement : (i) quelles sont les conditions sur les coefficients de $P(x)$ pour que l'équation soit résoluble par radicaux, (ii) avec quelles autres classes de fonctions que les radicaux peut-on trouver des formules de résolution. Pour (i) la réponse fut apportée entre autres par l'ami de Max Planck, Carl Runge, en 1885 (les formules sont compliquées). Pour (ii) Charles Hermite, Leopold Kronecker puis Felix Klein montrèrent que l'on pouvait utiliser la théorie des *fonctions elliptiques*. Ces questions difficiles sont encore ouvertes et conduisent jusqu'aux profondes recherches contemporaines sur la théorie de Galois "topologique" initiée par Vladimir Arnold en 1963 avec une démonstration topologique du théorème de Ruffini-Abel.

1.6. Notre exemple de neurotopologie

Après ces exemples hors sujet mais fort instructifs, nous allons maintenant traiter de problèmes qui sont au centre de notre réflexion. Le premier est celui de la neurotopologie à la section 4 du chapitre 2. Nous avons vu qu'il est très proche, à travers la tradition physiologique et psychophysique de Helmholtz, de ce que Kant disait sur la perception et les schèmes de concepts empiriques.

1. Les données sont les propriétés observables de la perception visuelle. Le problème est celui de la recherche de leurs causes externes.
2. Les solutions "faibles" sont les solutions "ascendantes" ("bottom-up") empiristes et associationnistes à la Hume.
3. L'obstruction est l'impossibilité de résoudre le problème inverse sans introduire des conditions de possibilité.
4. Le "reverse engineering" est la méthode "descendante" ("top-down") appropriée.
5. Elle fait apparaître des priors et même des hyperpriors (version de l'Esthétique Transcendantale kantienne) et des schèmes générateurs de percepts.

6. Le retour sur les propriétés observables de la perception visuelle est tout simplement la neurogéométrie à laquelle cet ouvrage est consacré...

1.7. Notre exemple de l'intégrabilité

Le second exemple est celui de l'intégrabilité des *équations aux différentielles totales* (EDT) de Pfaff.

1. Les données sont des EDT de Pfaff $\omega = 0$ où ω est une 1-forme et le problème à résoudre est celui de leur intégrabilité au moyen de fonctions les satisfaisant.
2. Il y a des solutions "faibles", directement accessibles, sans restructuration des données, pour les EDT exactes et les EDT admettant des facteurs intégrants.
3. Mais on rencontre une obstruction dans le cas général car il n'existe pas d'intégrales complètes.
4. On inverse alors le problème en se focalisant sur les conditions de possibilité de l'intégrabilité.
5. On découvre ensuite progressivement, de Pfaff à Cartan, qu'il existe une structure supplémentaire spécifique et universelle, indépendante des données et *a priori* relativement à elles, qui sert de "background structure". Il s'agit de la géométrie des champs d'hyperplans tangents \mathcal{K} définis par l'EDT et de leurs "dérivées" successives au sens de Cartan, formes différentielles de degré croissant dont la nullité permet d'éliminer des variables et de réduire la 1-forme ω à une forme normale qui est de contact ou symplectique. *Cette géométrie de l'intégrabilité fonctionne comme la structure synthétique a priori du calcul différentiel.*
6. On revient alors aux données et on les "normalise" à partir de ce supplément de structure en effectuant la réduction et en calculant les intégrales.

Nous allons maintenant explorer en détail comment s'est opérée cette nouvelle "révolution galoisienne".

2. Le problème de Pfaff de Monge à Cartan : la période héroïque

2.1. Le "moment fécond" de la géométrie de l'intégrabilité

Avant d'approfondir la géométrie de l'intégrabilité, fidèles à nos choix pédagogiques et méthodologiques, nous allons consacrer une section à des rappels historiques sur les conditions d'intégrabilité des 1-formes, c'est-à-dire des formes de Pfaff ω . Cette histoire va nous conduire de remarques pionnières de Gaspard Monge en 1784 aux grands textes d'Élie Cartan de 1899 [90] et 1901 [91]. Nous avons déjà

parlé du premier "Sur certaines expressions différentielles et le problème de Pfaff" à la section 7.3 du chapitre 2. Nous allons maintenant parler également du second "L'intégration des systèmes d'équations aux différentielles totales".

Cette plongée historique voudrait saisir un "moment fécond" particulièrement important pour notre entreprise modélisatrice de neurogéométrie. Par rapport à la tradition du calcul différentiel et de l'intégration fonctionnelle déjà richement développée au début du XIX^e siècle, la géométrisation des conditions d'intégrabilité est bien une "révolution". Initiée de façon encore intuitive et mal formalisée entre Pfaff et Grassmann, elle trouve son formalisme idoine avec Lie, Frobenius et Cartan.

Évidemment, aujourd'hui, ces travaux qui ont fait rupture en leur temps sont à leur tour devenus classiques et même désuets. Mais, selon nous, ils constituent *les bonnes structures susceptibles d'être neuralemement implémentables*. Notre insistance sur ce "moment fécond" n'est donc pas motivée par un intérêt simplement historique mais aussi et surtout par un intérêt méthodologique. De même que, bien qu'on ne puisse pas faire l'hypothèse que des planètes "calculent", on peut néanmoins faire l'hypothèse que leurs trajectoires satisfont un principe de moindre action et des contraintes symplectiques, de même, bien qu'on ne puisse pas faire l'hypothèse que des neurones "calculent", on peut néanmoins faire l'hypothèse que leurs activités satisfont un principe variationnel et des contraintes de contact.

2.2. Quelques repères biographiques

Donnons donc quelques brefs repères biographiques sur les auteurs que nous allons commenter. Nous le ferons en suivant en partie les fiches de *Wikipedia* et d'*Encyclopedia.com*, ainsi que les ouvrages de synthèse cités plus bas. Un excellent panorama des mathématiques du XIX^e siècle est la série de trois volumes *Matematika XIX veka* édités par Andreï Kolmogorov et Adolf Yushkevich, en particulier le volume 3 [304] pour ce qui nous occupe ici.

Gaspard Monge (1746-1818). Savant (mathématicien, physicien, chimiste, ingénieur) d'une importance considérable sous la Révolution et l'Empire, spécialiste de géométrie descriptive et de géométrie différentielle, franc-maçon. Jeune, il est inspiré par Euler et Lagrange, se lie à Vandermonde, d'Alembert et Condorcet, est élu en 1780 à l'Académie des Sciences. Professeur de grand renom dans des écoles d'ingénieurs (génie militaire), auteur de traités influents, il devient l'un des scientifiques majeurs de la période révolutionnaire. Ministre de la Marine en 1792-1793, il est l'un des fondateurs de l'École Polytechnique en 1794 où il enseigne la géométrie descriptive et dont il sera nommé directeur en 1797 par Bonaparte. Il joue un rôle essentiel dans le développement des sciences et des techniques de l'industrie de l'armement (artillerie). Suivant la trajectoire de Bonaparte, il participe à la campagne d'Égypte, est nommé Président du Sénat en 1806. Exclu et licencié de ses fonctions par la Restauration après Waterloo, il meurt en 1818.

Johann Friedrich Pfaff (1765-1825). Professeur de mathématiques aux universités d'Helmstedt et de Halle. Directeur de thèse en 1799 à l'université de Helmstedt de Carl Friedrich Gauss (qui resta proche de lui) et professeur d'August Ferdinand Möbius, il était un spécialiste du calcul différentiel symbolique à la Leibniz, des EDP et de la théorie des séries (Taylor, Lagrange). C'est son travail principal qui est au cœur de cette section, travail dont Gauss soulignait, dans son compte-rendu [205], qu'il était "une belle extension du calcul intégral". Le lecteur intéressé par son parcours pourra consulter l'étude de Jean Dhombres "La méthode fonctionnelle chez J.F. Pfaff : une filiation leibnizienne" de 1993 [143] où l'auteur explique en détail les travaux de début de carrière de Pfaff à Helmstedt sur le symbole d comme opérateur de construction de différentielles soumis à les lois de calcul symbolique ($da = 0$ et $d(ax) = adx$ si a est constante et x variable, $d(xy) = xdy + ydx$, etc.).

Julius Plücker (1801-1868). Mathématicien et physicien (spécialiste des rayons cathodiques et de la spectroscopie des gaz). Séjournant comme étudiant à Paris en 1823, il est influencé par les travaux de Monge et de l'école française de géométrie. Il retourne à Bonn où il fera sa carrière de professeur. Pour ce qui nous intéresse ici, il peut être considéré comme un précurseur de Grassmann et un inspirateur de Lie en tant qu'inventeur de la grassmannienne projective $G(1, 3)$ des droites projectives \mathbb{P}^1 de l'espace projectif \mathbb{P}^3 . \mathbb{P}^3 est l'espace des droites X de \mathbb{R}^4 passant par l'origine O , celles-ci étant identifiables aux points $\bar{x} = (x_0, x_1, x_2, x_3,) \neq O$ à homothétie $\lambda\bar{x}$ ($\lambda \neq 0$) près (\bar{x} définit la droite $O\bar{x}$). Les x_i sont les coordonnées dites "homogènes" de X . Un sous-espace affine \mathbb{A}^3 de \mathbb{P}^3 s'obtient en supposant (par exemple) $x_0 \neq 0$ et en normalisant par $x_0 = 1$, (x_1, x_2, x_3) devenant les coordonnées standard d'un point x_X et l'équation $x_0 = 0$ devenant celle du "plan à l'infini". Les droites Δ de \mathbb{P}^3 correspondent à des plans P_Δ de \mathbb{R}^4 passant par l'origine. Soient X et Y deux générateurs de P_Δ et considérons la matrice 2×4 (X, Y) à 2 colonnes et 4 lignes. On considère les (2×2) -mineurs $p_{ij} = x_i y_j - x_j y_i$ qui sont des quantités *antisymétriques* avec $p_{ii} = 0$ (si $x_i = y_i$) et $p_{ji} = -p_{ij}$. Les coordonnées de Plücker sont 6 des p_{ij} qui sont indépendants. On prend traditionnellement $p_{01}, p_{02}, p_{03}, p_{23}, p_{31}, p_{12}$. Si l'on change les générateurs (X, Y) en $(\lambda X, \mu Y)$ par homothétie (avec $\lambda, \mu \neq 0$) on obtient $\lambda\mu p_{ij}$; si l'on change les générateurs (X, Y) en prenant $(X' = \alpha X + \beta Y, Y' = \gamma X + \delta Y)$ (avec $\alpha\delta - \beta\gamma \neq 0$ car X' et Y' doivent être indépendants) on obtient $p'_{ij} = (\alpha\delta - \beta\gamma) p_{ij}$. Les 6 p_{ij} sont donc les coordonnées homogènes d'un point $\varphi(\Delta)$ de \mathbb{P}^5 : ce sont les coordonnées de Plücker. Dans le \mathbb{A}^3 défini par $x_0 = 1$, la droite Δ (non contenue dans le plan à l'infini) devient la droite passant par x_X et y_Y . Dans ce cas, les trois $p_{0j} = y_j - x_j$ sont les composantes de $y_Y - x_X$ définissant classiquement la direction d_Δ de Δ et, quant aux p_{23}, p_{31}, p_{12} , ils sont les composantes du produit vectoriel $x_X \wedge y_Y$ donnant le moment m_Δ de Δ par rapport à l'origine. La direction d_Δ et le moment m_Δ définissent univoquement Δ dans \mathbb{A}^3 . Mais l'on voit que le produit scalaire $d_\Delta \cdot m_\Delta = 0$ puisque d_Δ et m_Δ

sont orthogonaux. D'où la relation quadratique entre les coordonnées de Plücker $p_{01}p_{23} + p_{02}p_{31} + p_{03}p_{12} = 0$ (que l'on peut trivialement vérifier directement). Il s'agit de l'équation d'une quadrique de Klein \mathcal{Q} et l'on montre que $\varphi(\Delta)$ est une bijection entre l'espace des droites de \mathbb{P}^3 et cette quadrique \mathcal{Q} de \mathbb{P}^5 . Il est alors facile d'exprimer la "géométrie des droites" avec les coordonnées plückeriennes. Par exemple dire qu'un point Z appartient à la droite $\Delta = (X, Y)$, c'est dire que tous les mineurs 3×3 de la matrice 3×4 (Z, X, Y) s'annulent, ce qui donne un système linéaire de 4 équations à 4 inconnues. Ou encore, deux droites $\Delta = (X, Y)$ et $\Delta' = (X', Y')$ se coupent si et seulement si $d_{\Delta} \cdot m_{\Delta'} + d_{\Delta'} \cdot m_{\Delta} = 0$ (dans le cas affine \mathbb{A}^3), autrement dit, dans \mathbb{P}^3 , si et seulement si le déterminant de la matrice 4×4 $\det(X, Y, X', Y') = 0$. On peut *dualiser* toutes ces notions en considérant qu'une droite Δ est également l'intersection de deux plans de coordonnées homogènes (a^i) et (b^j) ce qui conduit aux coordonnées de Plücker duales p^{ij} dont l'utilisation simplifie des formules. Par exemple, la relation quadratique de Plücker s'écrit $\sum_{j=0}^{j=3} p_{0j} p^{0j} = 0$. De même, $Z \in \Delta$ s'exprime par les 4 équations $\sum_{i=0}^{i=3} p^{ij} z_j, j = 0, \dots, 3$.

Hermann Günther Grassmann (1809-1877). Très éclectique (mathématicien, physicien ayant travaillé sur la théorie des couleurs, mais aussi phonéticien, indianiste, spécialiste du sanskrit internationalement reconnu, membre de l'American Oriental Society, humaniste, éditeur), enseignant au lycée de Stettin (Prusse), il est l'un des principaux fondateurs de l'algèbre linéaire, du calcul tensoriel et, surtout, de l'algèbre extérieure. Son ouvrage pionnier sur ce qui deviendra l'algèbre linéaire est *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik* (1844) [219].¹⁷ Les vecteurs et leurs combinaisons linéaires étaient connus depuis longtemps (règle du parallélogramme) en dimension 2 et 3 mais dans des espaces affines, sans leur relation d'équipollence et sans définition axiomatique abstraite en toute dimension. Grassmann développe un calcul qui allie la "géométrie analytique" avec coordonnées et la "géométrie synthétique" sans coordonnées. On calcule abstraitement et symboliquement sur des symboles de sous-espaces (points, droites, plans, etc.) d'espaces linéaires et l'on effectue des opérations algébriques sur ces symboles (produits scalaires, produits extérieurs).¹⁸ Les principales découvertes de Grassmann sont les déterminants de n vecteurs comme volume du parallélépipède qu'ils engendrent et l'*algèbre extérieure* avec son produit extérieur radicalement non commutatif généralisant le produit vectoriel introduit par Gibbs et Clifford en dimension

17. "Ein neuer Zweig", une "nouvelle branche". On voit à quel point ces auteurs avaient conscience de découvrir et d'explorer de nouvelles structures mathématiques.

18. Si $e_i, i = 1, \dots, n$ sont des symboles, Grassmann considère l'ensemble E des combinaisons linéaires formelles $e = \sum_{i=1}^{i=n} \lambda_i e_i$ à coefficients dans \mathbb{R} et y définit la structure d'espace vectoriel, l'indépendance linéaire entre vecteurs, les bases, les sous-espaces V, W , les projections, les dimensions (avec la formule fondamentale $\dim(V \cap W) + \dim(V + W) = \dim V + \dim W$).

3.¹⁹ Parfois trop obscur, sa *Ausdehnungslehre* eut peu de succès mais fut utilisée par Hamilton pour sa théorie révolutionnaire des quaternions. Sa refonte – *Die Ausdehnungslehre. Vollständig und in strenger Form bearbeitet* de 1862 [220] – eut plus de succès. Formulées de façon difficile à comprendre, ces découvertes furent méconnues jusqu’à leur véritable reconnaissance par Hermann Hankel en 1866, Victor Schlegel, Sophus Lie, Felix Klein en 1869 et surtout par Clebsch qui, en 1871, le fit élire membre correspondant à l’Académie des Sciences de Göttingen. Klein et Gibbs publieront son œuvre et Maxwell et Clifford écriront sur lui. Ses *Œuvres Complètes* seront publiées par Friedrich Engel.

L’*Ausdehnungslehre* de Grassmann est encore un mixte d’intuitions géométriques et de calculs symboliques. Sa “théorie générale des formes” est un précurseur de la notion d’ensemble muni d’une structure algébrique. Par exemple il y aura un élément “générateur” (un vecteur lié générique pour un espace vectoriel) ayant la possibilité de se transformer de façon continue en engendrant une “extension” sur laquelle seront définies des opérations caractéristiques du type d’élément qu’est le générateur (pour les vecteurs libres, il y aura la structure vectorielle avec les sommes, les combinaisons linéaires, les produits extérieurs, et pour les vecteurs liés, il y aura la structure affine).

Le calcul algébrique de Grassmann sera intimement associé par Cartan au calcul différentiel dans la théorie des formes différentielles qui l’applique aux espaces tangents et cotangents des variétés.

Le lecteur intéressé par la philosophie des mathématiques de Grassmann pourra consulter les ouvrages de Dominique Flament [185], Gilles Chatelet [108] ou Paola Cantù [88], ainsi que leurs bibliographies.

Leopold Natani (1819-1905). Nous évoquerons plus bas ses contributions. Il est surtout connu pour son dictionnaire *Mathematisches Wörterbuch* en sept volumes publié avec Ludwig Hoffmann entre 1858 et 1867.

Alfred Clebsch (1833-1872). Étudiant à Königsberg puis (remarquable) professeur à Berlin, Karlsruhe, Giessen et surtout à Göttingen où il succéda à Riemann et finit recteur. Spécialiste d’abord d’hydrodynamique et d’élasticité, puis de géométrie algébrique, des intégrales et des fonctions abéliennes (dans une optique complémentaire, purement algébrique, à celle de Riemann), de la théorie des représentations de groupes et de la théorie des invariants (à la suite de Cayley, Salmon et Sylvester).²⁰ Les “coefficients de Clebsch-Gordan” pour les groupes de Lie compacts sont universellement utilisés. Cofondateur en 1868 avec son ami Carl Neumann des *Mathematische Annalen*. Bien qu’emporté jeune par une diphtérie, il eut des élèves aussi importants que Felix Klein ou Max Noether.

19. Grassmann introduit les multivecteurs $e_{i_1} \wedge \cdots \wedge e_{i_k}$ avec les règles $e_i \wedge e_i = 0$ et $e_i \wedge e_j = -e_j \wedge e_i$ si $i \neq j$.

20. Sur la théorie des invariants, le lecteur pourra consulter l’article de François Lê [318].

Gaston Darboux (1842-1917). Major de l'École Normale Supérieure et de l'École Polytechnique en 1861, suppléant de Joseph Liouville puis de Michel Chasles (qui fut son directeur de thèse et dont il devint le successeur à la Faculté des Sciences de Paris), il eut une carrière académique prestigieuse (Doyen de la Faculté des Sciences, Académie des sciences, dont il fut le secrétaire perpétuel, et plus d'une quinzaine d'Académies étrangères). Spécialiste des méthodes non seulement d'analyse mais également de géométrie en théorie des équations différentielles, il travailla aussi en cinématique et en dynamique. Célèbre comme spécialiste de la théorie des courbes et des surfaces particulières comme les cyclides de Dupin (les images par inversion de quadriques dans \mathbb{R}^3), il fut l'inventeur de la méthode du repère mobile que Cartan utilisera massivement et que nous rencontrerons plus bas dans la section 2 du chapitre 11.

Sophus Lie (1842-1899). Il est, avec Élie Cartan, l'un des génies mathématiques les plus centraux pour nous puisque plusieurs chapitres de cet ouvrage reposent entièrement sur la théorie des groupes et des algèbres de Lie. Étudiant à Kristiania (l'ancienne Oslo) il suit les cours de Sylow sur Abel et Galois. Il est influencé jeune (1868) par Poncelet et surtout Plücker, dont il se considérait être un disciple, et à qui il emprunta l'idée fondamentale que l'on pouvait créer de *nouvelles géométries* en prenant pour éléments de base non plus des points mais des droites, des plans et, plus généralement des sous-espaces. En 1869-1870 il visite Göttingen et Berlin (Kronecker, Kummer, Weierstrass) où il rencontre Felix Klein qui avait été l'étudiant de Plücker et avec qui il se lie d'amitié, amitié étroite qui durera jusque vers 1890. En 1870 les deux amis séjournent à Paris et rencontrent les géomètres Darboux, Chasles et Jordan et s'initient à l'école géométrique de Monge ainsi qu'à la théorie de Galois. À cause de la guerre avec la Prusse, ils doivent quitter Paris, Lie retournant en 1871 à Kristiania à travers l'Italie (après avoir été arrêté comme "espion" par les Prussiens et libéré grâce à l'intervention de Darboux) et Klein se retrouvant en 1872 à Erlangen où il conclura son célèbre "Programme d'Erlangen" définissant une géométrie comme l'ensemble des propriétés invariantes par l'action d'un groupe.

Professeur à Kristiania, où il se retrouve un peu trop isolé, Lie développe ses contributions à la théorie des EDP (cf. [333], [334]) en visant *une théorie de Galois de l'intégrabilité* puis se spécialise dans la théorie des groupes de transformations (cf. [335]). À partir de 1884, il rédige avec Friedrich Engel (1861-1941, doctorant à Leipzig d'Adolph Meyer et recommandé à Lie par Klein professeur à cette université depuis 1880) son monumental traité *Theorie der Transformationsgruppen* [169] qui sera publié en trois volumes de 1888 à 1893. En 1886 Lie succède à Klein à Leipzig où Engel se trouve aussi comme professeur depuis 1885. Il y fait école et forme de nombreux étudiants mais s'y adapte mal, lui le vigoureux norvégien habitué aux forêts du grand Nord.

Lie avait un caractère difficile. Il entretenait des relations conflictuelles avec ses rivaux comme Wilhelm Killing (qui s'était focalisé sur les *algèbres* de Lie, la structure infinitésimale des groupes de Lie au voisinage de leur élément neutre) ou Gaston Darboux. Il estima même qu'Engel l'avait trahi en faveur de Killing. Et quand, en 1889-1892, Klein revint sur son programme d'Erlangen et ses travaux avec lui des années 1869-1872 pour se battre contre l'école de Berlin où dominait l'“arithmétisation” de la géométrie (Kronecker, Kummer, Weierstrass, puis Frobenius qui en devint le leader après le retrait de Weierstrass en 1893), il résista à son enrôlement dans la géométrie “à la Klein” et revendiqua la spécificité de sa conception personnelle. En fait, Lie aura toujours des rapports conflictuels avec les mathématiciens allemands, et se détournera d'eux vers la fin de sa vie pour se tourner vers la France et des spécialistes de la théorie de Galois pour les équations différentielles comme Ernest Vessiot (1865-1952). Il sera reçu à l'Académie des sciences en 1892 et son grand continuateur sera Élie Cartan. À partir de 1889 Lie devient cliniquement déprimé, misanthrope, et il retourne à Kristiania en 1898 où il décèdera l'année suivante. Ses œuvres complètes *Gesammelte Abhandlungen* seront publiées en sept volumes entre 1922 et 1960 (et jusqu'en 1937 par Engel lui-même).

(Ferdinand) Georg Frobenius (1849-1917). Professeur à Berlin puis Zurich, membre de l'Académie des Sciences de Prusse (1893). Il fut à partir des années 1870 avec Schur et Stickelberger l'un des pionniers de la théorie des groupes *abstraites* établissant des liens structuraux entre la théorie de Galois des équations algébriques (groupes de permutations), la théorie des nombres et la géométrie (groupes de Lie). Il introduisit la notion de “caractère” d'une représentation²¹ que nous utiliserons plusieurs fois, en particulier au chapitre 7 et à la section 1.1.3 du chapitre 17. Il démontra ou redémontra des théorèmes célèbres comme le théorème de Sylow²², le théorème de Cayley-Hamilton²³ ou le théorème disant que les seuls espaces vectoriels \mathbb{R}^n que l'on peut munir d'une structure de corps (associative mais pas forcément commutative) sont \mathbb{R} , \mathbb{C} et \mathbb{H}_Q (les quaternions). En théorie des équations différentielles, il généralisa la théorie de Galois aux équations différentielles ordinaires (EDO) en résolvant un problème posé par Lazarus Fuchs (1833-1902) dans les années 1865-1868. Soit $L(f) = \sum_{k=0}^{k=n} a_k(z) f^{(k)}(z) = 0$ une EDO linéaire homogène d'ordre

21. Si $\rho : G \rightarrow GL(V)$ est une représentation linéaire des éléments d'un groupe G par des automorphismes d'un espace vectoriel V sur un corps de base K (GL signifie “groupe linéaire”), le caractère de ρ est le morphisme $\chi_\rho : G \rightarrow K$, $\chi_\rho(g) = \text{Trace}(\rho(g))$.

22. Si G est un groupe fini et p est un nombre premier, un p -sous-groupe H est un sous-groupe dont tous les éléments sont d'ordre une puissance p^k de p . Les p -sous-groupes de Sylow sont les p -sous-groupes maximaux. Si p^G est le facteur de p dans la décomposition de l'ordre de G en nombres premiers, les p -sous-groupes de Sylow existent, ils sont tous d'ordre p^G et tous isomorphes entre eux.

23. Si A est un endomorphisme d'un espace vectoriel V (i.e. une matrice si on a choisi une base de V), A satisfait son polynôme caractéristique $P(X) = \det(A - X.Id)$.

n pour une fonction $f(z)$ d'une variable complexe z , $f^{(k)}$ étant la dérivée d'ordre k et les $a_k(z)$ étant des fonctions méromorphes de z avec au plus un nombre fini de pôles. De même que pour les équations algébriques on cherchait des conditions pour la résolubilité par radicaux, on cherche ici des conditions pour une intégrabilité algébrique de $L(f) = 0$. Le résultat est que, au voisinage d'un point singulier $z_0 \in \mathbb{C}$, si toutes les solutions sont des fonctions algébriques de z , alors il existe un polynôme $P(f, z)$ en f , $P(f, z) = \sum_{\ell=0}^{\ell=m} A_\ell(z) f^\ell(z)$, de degré $m \geq n$ dont les coefficients $A_\ell(z)$ sont des fonctions rationnelles de z et tel que, localement en a , ses racines sont des solutions de $L(f) = 0$.

Ce sont évidemment les profonds travaux de Frobenius sur les EDP et le problème de Pfaff qui nous intéressent ici. Un excellent volume recouvrant l'ensemble de ses innovations est l'ouvrage de 2013 de Thomas Hawkins [243] *The Mathematics of Frobenius in Context. A Journey Through 18th to 20th Century Mathematics*.

Édouard Goursat (1858-1936). Étudiant entre autres de Gaston Darboux à l'École Normale Supérieure puis à la Faculté des Sciences, il se spécialisa en calcul différentiel et intégral puis en théorie des fonctions d'une variable complexe et des intégrales abéliennes. Il en devint l'un des leaders. Très proche d'Émile Picard, il enseigna à Toulouse, à l'ENS, puis à la Faculté des Sciences (où il succéda en 1897 à Picard) et, parallèlement, à l'École Polytechnique et à l'ENS de Saint Cloud. Récipiendaire de plusieurs prix, il fut élu à l'Académie des sciences en 1919. Les trois volumes de son classique *Cours d'Analyse mathématique* fut une référence essentielle au début du XX^e siècle. Mais l'ouvrage qui nous intéressera prioritairement ici sera ses *Leçons sur le problème de Pfaff* de 1922 [218].

Friedrich Engel (1861-1941). Nous avons parlé de lui à propos de Sophus Lie aux travaux desquels il fut associé de façon très étroite.

Élie Cartan (1869-1951). Avec Sophus Lie, il est le génie mathématique le plus important pour notre projet. Nous avons déjà expliqué certaines de ses découvertes (le calcul des formes différentielles) et nous en expliquerons de nombreuses autres dans les chapitres qui suivent. Élève extrêmement brillant, étudiant à l'École Normale Supérieure (promotion 1888) et à la Faculté des sciences (Sorbonne), il admirait Henri Poincaré et fut influencé, entre autres, par Édouard Goursat, Émile Picard, Paul Appell, Gaston Darboux, Charles Hermite, Jules Tannery. Son camarade de l'ENS Arthur Tresse, étudiant de Sophus Lie à Leipzig, le mis au courant des résultats de Killing et Engel sur les groupes de Lie simples. En 1892 le "génial" Sophus Lie était à Paris, invité par Darboux et Tannery, et rencontra Cartan. Ils eurent de nombreuses discussions. Cartan soutint sa thèse en 1894 sous la direction de Darboux et Lie sur la théorie des groupes de Lie et la classification des algèbres de Lie semi-simples sur \mathbb{C} . Il enseigna ensuite à Montpellier (1894), Lyon (1896), Nancy (1903) et Paris à partir de 1909 (la Sorbonne où il fut professeur de 1912 à

sa retraite en 1940, l'EPCI et l'ENS). Après sa retraite, il rejoindra le "Bureau des Longitudes" en 1945.

Ses innovations furent rapidement reconnues par des maîtres comme Henri Poincaré et Hermann Weyl, celui-ci affirmant plus tard en 1938 dans son compte-rendu [566] de *La théorie des groupes finis et continus et la géométrie différentielle traitées par la méthode du repère mobile* (Gauthier-Villars, Paris, 1937),

"Cartan is undoubtedly the greatest master in differential geometry."²⁴

Mais elles mirent un certain temps pour infuser dans la communauté. Comme le notent Chern et Chevalley [114] dans leur texte d'hommage nécrologique de 1952, avant Cartan les analystes s'intéressaient à l'intégrabilité plus du point de vue de la théorie des fonctions que du point de vue de la géométrie à la Lie-Darboux et après Cartan (André Weil et Bourbaki) les nouvelles générations s'intéresseront plus aux nouvelles orientations de l'algèbre et de la topologie. Mais, après un certain temps de latence, la reconnaissance de Cartan fut néanmoins remarquable : conférencier au "Congrès international des mathématiciens" en 1924 (Toronto), 1932 (Zürich) et 1936 (Oslo); docteur Honoris Causa des Universités de Liège (1934), Harvard (1936), Bruxelles-Louvain (1947), Bucarest (1948), Pise (1948); président de la Société Mathématique de France (1915), membre de l'Académie des sciences (1931) dont il sera élu président en 1946, membre des Académies de Belgique, de Pologne (1921), de Norvège (1926), des Lincei (1927), de Roumanie (1931), de la London Mathematical Society (1939), de la Royal Society of London (1947) et de la National Academy of Sciences (USA, 1949).

Comme nous l'avons déjà dit à la section 7 du chapitre 2, son œuvre est immense et couvre d'innombrables domaines

- du calcul intégral-différentiel (calcul extérieur des formes différentielles, théorie invariante des systèmes d'EDP, systèmes intégrables, systèmes en involution, "prolongements" permettant de transformer des solutions singulières en solutions générales par adjonction de nouvelles variables, théorie de Cartan-Kähler),
- de la géométrie et des représentations des groupes et algèbres de Lie (classification des algèbres semi-simples à la suite de Killing, formule de Maurer-Cartan),
- de la géométrie différentielle (méthode du repère mobile qui associe à un fibré sur lequel opère un groupe son fibré principal, connexions généralisant les métriques riemanniennes),
- de la géométrie des variétés riemanniennes (géométrie de Riemann-Cartan),
- des espaces symétriques et des espaces homogènes,
- des propriétés topologiques globales des groupes de Lie compacts,
- de la mécanique rationnelle (invariants intégraux),

24. Weyl trouvait cependant Cartan "hard reading" car il considérait acquise la tradition géométrique à la Darboux.

– de la physique théorique (approfondissement de la Relativité générale, introduction des spineurs).

Ses *Œuvres complètes* [99] ont été publiées en trois tomes (six volumes) de 1952 à 1955. Rappelons aussi (cf. chapitre 2, section 7), la "Notice sur les travaux scientifiques" [98] de 1939, l'hommage [114] de 1952 par Shiing-Shen Chern et Claude Chevalley, le texte de Paulette Libermann [332] et le Fonds Élie Cartan de l'Académie des Sciences. Un excellent ouvrage sur l'ensemble de ses travaux est le volume russe [12] de Maks Akivis et Boris Rosenfeld traduit par l'*American Mathematical Society*. Depuis 1980, il existe un prix Élie Cartan de l'Académie des Sciences.

2.3. EDO et EDP

Dans ce qui suit, nous allons maintenant évoquer l'histoire du problème de Pfaff, c'est-à-dire de la résolution des équations "aux différentielles totales" (EDT) $\omega = \sum_{i=1}^{i=n} A_i dx_i = 0$. Ce problème fait partie de l'immense domaine des équations différentielles (ED) dont il serait complètement hors de propos de reprendre ici l'histoire puisque celle-ci s'identifie à l'histoire de presque toutes les sciences mathématisées, en particulier la physique. Rappelons néanmoins, très brièvement, quelques notions basiques sans nous préoccuper de leur évolution historique.

2.3.1. EDO.

La classe la plus simple d'ED est celle des équations différentielles "ordinaires" (EDO). Cela signifie que l'on cherche une fonction d'une variable $y(x)$ ²⁵, fonction reliée à un certain nombre de ses dérivées par une fonction

$$F(x, y, y', \dots, y^{(n)}) = 0$$

n étant l'ordre de l'EDO.²⁶ Le cas le plus élémentaire est celui où y' est simplement une fonction $f(x)$ de x . En effet, si $f(x)$ est une fonction intégrable, alors la solution est immédiate et donnée par $y(x) = \int_a^x f(u) du + c$, a et c étant des constantes. La situation devient beaucoup plus compliquée lorsque y intervient dans F . Une foule de cas particuliers et de généralisations s'introduisent immédiatement. Par exemple, lorsque x n'intervient pas dans F on dit que l'équation est "autonome". Par ailleurs, on peut supposer que y est une variable multidimensionnelle $y \in \mathbb{R}^d$ de dimension d .

Une EDO à une variable d'ordre n peut toujours se ramener une EDO d'ordre 1 de dimension n . Il suffit en effet de poser $y = Y_0$, $y' = Y_1$, $y^{(n-1)} = Y_{n-1}$ et de considérer l'équation du premier ordre en $y = (Y_0, Y_1, \dots, Y_{n-1})$ donnée par

25. Dans les applications simples, x et y sont des variables réelles. La théorie est nettement plus compliquée lorsque ce sont des variables complexes et que l'on cherche des $y(x)$ holomorphes.

26. La notation $y(x)$ est traditionnelle en géométrie où il s'agit de trouver une courbe, et celle $x(t)$ en physique où il s'agit de trouver l'évolution temporelle d'une grandeur.

$F(x, Y_0, Y_1, \dots, Y_{n-1}, Y'_{n-1}) = 0$. Lorsqu'elle est autonome et qu'elle est résoluble en $y^{(n)}$ sous la forme $y^{(n)} = G(y, y', \dots, y^{(n-1)})$, elle s'écrit sous la forme $Y'_0 = Y_1, Y'_1 = Y_2, \dots, Y'_{n-1} = G(Y_0, Y_1, \dots, Y_{n-1})$. Cela est un exemple de champ de vecteurs sur l'espace \mathbb{R}^n de coordonnées Y_0, Y_1, \dots, Y_{n-1} , un tel champ étant défini par $Y'_0 = G_0(Y_0, Y_1, \dots, Y_{n-1}), Y'_1 = G_1(Y_0, Y_1, \dots, Y_{n-1}), \dots, Y'_{n-1} = G_{n-1}(Y_0, Y_1, \dots, Y_{n-1})$. La solution $y(x) = Y_0(x)$ est donc la première composante d'une trajectoire de ce champ.

Tout de suite se pose de nombreux problèmes. D'abord celui des conditions initiales et des conditions aux limites. On peut se fixer les valeurs de y et de ses dérivées en un point x_0 . On peut remplacer une condition initiale sur une dérivée par une valeur de la fonction en un autre point x_1 . Étant données de telles conditions, on s'interrogera sur l'existence et l'unicité, du moins locale, d'une solution (condition de Cauchy-Lipschitz), sur la régularité (continuité, ordre de différentiabilité) des solutions par rapport aux conditions initiales, sur la possibilité de prolonger globalement une solution locale, etc.

Si l'équation n'est pas résoluble par rapport $y^{(n)}$, si par exemple $F(x, y, y')$ est un polynôme $F_{x,y}(y')$ en y' , on décompose le plan (x, y) en domaines où $F_{x,y}(y')$ a un même nombre p de racines distinctes. Dans ces domaines il y aura localement p solutions différentes. Ces domaines seront séparés par des solutions singulières qui sont des enveloppes de trajectoires et correspondent aux points (x, y) où une paire de racines réelles de $F_{x,y}(y') = 0$ fusionnent et deviennent imaginaires conjuguées. Cela signifie que dans le $\mathbb{R}^3(x, y, p)$ les relevées legendriennes (x, y, y') sont sur la surface Σ d'équation $F(x, y, p) = 0$ qui admet un contour apparent lorsqu'on la projette sur le plan (x, y) . Σ admet des lignes "pli" le long desquelles la projection lui est tangente, ce qui correspond à la fusion de deux racines réelles et à leur disparition. Nous en avons déjà parlé à la section 2.10 du chapitre 3 à propos de l'article de Thom [531] "Sur les équations différentielles multiformes et leurs intégrales singulières" de 1972.

Pendant longtemps, les spécialistes se sont focalisés sur la recherche d'intégrales explicites des EDO. Les résultats ont été nombreux, mais, dans la mesure où les cas d'intégrabilité explicite au moyen de fonctions connues ou de fonctions spéciales sont très rares, il a vite fallu changer de perspective et essayer de dire quelque chose sur les intégrales même si on ne les connaît pas explicitement.

On a alors découvert que la complexité de la structure globale des champs de vecteurs, aussi dits systèmes dynamiques, est prodigieuse. Depuis Poincaré, elle fait l'objet de la *dynamique qualitative*. Un système dynamique $\frac{dx}{dt} = X$ (i.e. $\frac{dx_i(t)}{dt} = X_i(x_j)$) sur une variété M définit un flot Φ_t qui est un groupe à 1-paramètre de difféomorphismes (i.e. $\Phi_0 = Id_M, \Phi_{t'} \circ \Phi_t = \Phi_{t+t'}, \Phi_{-t} = (\Phi_t)^{-1}$). Φ_t consiste à suivre jusqu'au temps t toutes les trajectoires $x(t)$ issues des différents points x de M au temps $t = 0$. On étudie la distribution des trajectoires, leurs comportements

asymptotiques (la façon dont certaines trajectoires sont attirées par des points attractifs, des cycles limites, des attracteurs plus ou moins compliqués et "étranges", la façon dont d'autres trajectoires diffusent de façon chaotique), leur stabilité de Liapounov, leur stabilité structurelle, etc. Nous ferons quelques incursions dans cet univers foisonnant lorsque cela sera pertinent pour notre propos.

2.3.2. EDP.

Les EDP sont des équations différentielles concernant des fonctions de plusieurs variables. Elles sont beaucoup plus compliquées à résoudre et au lieu de dépendre de constantes d'intégration elles ont des conditions initiales et des conditions aux limites qui sont elles-mêmes des fonctions. Les difficultés d'intégration sont telles que l'on ne maîtrise bien la théorie que pour des classes très restreintes d'EDP.

Dans la section 8.2 du chapitre 2 nous avons présenté les équations de Maxwell qui sont prototypiques. En physique, certaines EDP du deuxième ordre linéaires sont omniprésentes comme l'équation de la chaleur (diffusion) dans un espace-temps $(t, x) : \frac{\partial u(t, x)}{\partial t} = \Delta_x u(t, x)$ (Δ_x étant le laplacien par rapport à la multi-variable spatiale x) avec une distribution initiale $u_0(x) = u(0, x)$, ou bien l'équation des ondes $\frac{\partial^2 u(t, x)}{\partial t^2} = \Delta_x u(t, x)$ avec un front d'onde initial $u_0(x) = u(0, x)$.²⁷ Nous avons déjà rencontré de telles EDP dans le Vol I à propos de la neurogéométrie et nous en rencontrerons plusieurs dans la suite. Il sera alors temps d'en expliquer certains aspects.

On peut remarquer qu'il existe un lien évident entre les systèmes d'EDO et certaines EDP. En effet si $\frac{dx_i(t)}{dt} = X_i(x_j)$ est un champ de vecteurs X sur une variété M , on peut lui associer une EDP linéaire du premier ordre, dite *équation de Liouville*, portant sur les fonctions $f : \mathbb{R} \times M \rightarrow \mathbb{R}$:

$$\frac{\partial f(t, x)}{\partial t} + \langle \nabla_x f, X \rangle = 0 .$$

Les solutions sont les fonctions laissées *invariantes* par le flot du champ X . En effet le long des trajectoires $x(t)$ du champ $f(t, x) = f(t, x(t))$ devient une fonction seulement de t et

$$\begin{aligned} \frac{df(t)}{dt} &= \frac{\partial f(t, x)}{\partial t} + \sum_i \frac{\partial f(t, x)}{\partial x_i} \frac{dx_i(t)}{dt} \\ &= \frac{\partial f(t, x)}{\partial t} + \langle \nabla_x f, X \rangle = 0 \end{aligned}$$

la somme \sum_i étant précisément le produit scalaire $\langle \nabla_x f, X \rangle$. Certaines méthodes classiques d'intégration d'EDP linéaires du premier ordre, dites *méthodes des caractéristiques*, s'inspirent de cette remarque. Nous les rencontrerons plusieurs fois.

27. Pour l'équation des ondes en optique géométrique et ondulatoire on pourra se référer à notre bibliographie dans [413]. Nous y reviendrons au chapitre 12, section 6.

Les EDP *linéaires* satisfont un principe de *superposition*, c'est-à-dire que les solutions forment un espace vectoriel. Cela permet de décomposer les solutions générales en sommes infinies de solutions simples (ce que l'on fait avec les séries de Fourier et les décompositions en harmoniques). Les EDP non linéaires ne possèdent pas cette propriété de superposition et c'est l'une des raisons des obstacles insurmontables que rencontre leur intégration. Un des exemples les plus connus est celui des équations de Navier-Stokes en mécanique des fluides (cf. la météorologie, l'océanographie, les tests aérodynamiques d'ailes d'avions, etc.), équations qui remontent à Euler, Navier ayant introduit la viscosité. Le théorème d'existence des solutions n'est même pas encore démontré (il fait partie des sept problèmes du Millénaire) et il suffit de penser à la turbulence avec sa fractalité multi-échelle de tourbillons pour comprendre sa difficulté. Mais de très nombreuses méthodes d'approximation numérique sont utilisables.

2.4. Les EDT : problème direct et problème inverse

Ce qu'on appela le problème de Pfaff au cours du XIX^e siècle était celui d'intégrer une équation dite "aux différentielles totales" (EDT)

$$\omega = \sum_{i=1}^{i=n} A_i(x_1, \dots, x_n) dx_i = 0$$

et, comme nous l'avons expliqué à la section 7.3 du chapitre 2, c'est pour unifier et systématiser les méthodes développées par Monge, Pfaff lui-même, puis, entre autres, Grassmann, Natani, Clebsch, Lie, Frobenius et Darboux que Cartan a créé la théorie moderne des formes différentielles.

L'appellation d'EDT est un peu ambiguë. En effet l'idéal serait que ω soit une *différentielle totale* $\omega = dF$, c'est-à-dire une 1-forme *exacte*, et que le problème de l'intégration se réduise à trouver F à partir des A_i . Mais, précisément, tout l'intérêt de la théorie est de comprendre ce que signifie l'intégration de $\omega = 0$ même quand ω n'est pas une différentielle totale. On continue toutefois de parler d'EDT.

Une vaste synthèse, excellente et complète, précédant juste les travaux de Cartan, se trouve dans le Volume 1 *Exact Equations and Pfaff's Problem* du traité classique de 1890 d'Andrew Russell Forsyth *Theory of Differential Equations* [188]. Une autre grande synthèse est le traité de 1922 d'Édouard Goursat [218] *Leçons sur le problème de Pfaff*.

Remarque. Avant d'entrer dans quelques détails techniques parfois un peu compliqués, soulignons qu'un exemple typique de problème de Pfaff est celui de la généralisation des relevées legendriennes de courbes planes aux relevées legendriennes de surfaces de $\mathbb{R}_{(x,y,z)}^3$. À l'époque on ne pouvait évidemment pas le formuler ainsi puisque tous les concepts étaient à inventer mais faisons le quand même

ici à des fins pédagogiques. Comme nous l'avons déjà esquissé dans une remarque de la section 2.3 du chapitre 3, toute la difficulté est de prendre conscience du fait que si $z = f(x, y)$ est l'équation d'une surface de \mathbb{R}^3 , les dérivées partielles $p = \frac{\partial f}{\partial x}$ et $q = \frac{\partial f}{\partial y}$ doivent être traitées comme de nouvelles variables indépendantes intervenant dans la 1-forme $\omega = dz - pdx - qdy$ qui est maintenant une 1-forme à cinq éléments différentiels dont les coefficients de dp et dq sont nuls. Cette 1-forme est non complètement intégrable (i.e. il n'existe pas d'hypersurfaces de dimension 4 de \mathbb{R}^5 sur laquelle ω s'annule) car $d\omega = -dp \wedge dx - dq \wedge dy$ et la 3-forme

$$\omega \wedge d\omega = dz \wedge dx \wedge dp + dz \wedge dy \wedge dq - pdx \wedge dy \wedge dq + qdx \wedge dy \wedge dp$$

est $\neq 0$ en dehors du \mathbb{R}^3 ($p = 0, q = 0$). Mais l'équation différentielle $\omega = 0$ est automatiquement vérifiée pour toutes les surfaces assez régulières $z = f(x, y)$ et admet donc une infinité d'intégrales de dimension 2, à savoir les relevées legendriennes de ces dernières (leurs 1-jets) qui sont des surfaces gauches dans \mathbb{R}^5 . Qui plus est, si $z = f(x, y)$ alors $dp = \frac{\partial^2 f}{\partial x^2} dx + \frac{\partial^2 f}{\partial y \partial x} dy$ et $dq = \frac{\partial^2 f}{\partial x \partial y} dx + \frac{\partial^2 f}{\partial y^2} dy$ et par conséquent

$$dp \wedge dx = \frac{\partial^2 f}{\partial y \partial x} dy \wedge dx \text{ et } dq \wedge dy = \frac{\partial^2 f}{\partial x \partial y} dx \wedge dy .$$

Mais comme $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$, la 2-forme $dp \wedge dx + dq \wedge dy = -d\omega$, qui est la 2-forme symplectique du $\mathbb{R}^4_{(x,y,p,q)}$ (identifiable au fibré cotangent $T^*\mathbb{R}^2_{(x,y)}$) trivialement étendue à $\mathbb{R}^5_{(x,y,z,p,q)}$, est nécessairement nulle sur toutes les relevées legendriennes des surfaces $z = f(x, y)$. Nous y reviendrons plus bas à la section 3.5.5 du chapitre 4. \square

Le problème de Pfaff est une réponse aux difficultés du problème de l'intégration au sens de Lagrange des EDP *du premier ordre* (en particulier non linéaires). Soit

$$f(x_i, z, p_i) = a \quad ((1))$$

une telle équation à $2m + 1$ variables imposée à une fonction $z(x_i)$ des m variables x_i , les p_i étant les m dérivées partielles $\frac{\partial z}{\partial x_i}$. La fonction z satisfait automatiquement l'EDT en $2m + 1$ variables

$$dz = \sum_i p_i dx_i. \quad ((2))$$

Intégrer (1) consiste à trouver des solutions $z = \varphi(x_i, c_i)$ (où les c_i sont les constantes d'intégration) en intégrant un système d'EDO. Comme le voulait Lagrange, il s'agit de réduire un problème de résolution d'EDP à la résolution d'un système d'EDO. Lagrange lui-même avait résolu le problème pour les équations du premier ordre *linéaires*. Car si $\sum_{k=1}^{k=m} C_k(x_i) p_k = 0$ est une telle équation, et si l'on utilise par exemple x_m comme variable indépendante, résoudre l'EDP équivaut à résoudre le

système d'EDO

$$\frac{dx_k}{dx_m} = \frac{C_k}{C_m}$$

pour $k = 1, \dots, m - 1$. Mais le problème général était trop complexe pour les EDP non linéaires et la situation était bloquée. Le passage effectué par Pfaff aux EDT débloqua la situation car toute EDP du premier ordre peut se ramener à une EDT. En effet, sous des conditions convenables, on peut utiliser l'EDP (1) pour exprimer par exemple p_m comme une fonction $p_m = \pi(z, x_i, p_k)$, $k = 1, \dots, m - 1$. Dans ce cas, l'EDT (2) devient une EDT (3) en seulement $n = 2m$ variables (z, x_i, p_k) ,

$$dz = \sum_{i=1}^{i=m-1} p_i dx_i + \pi(z, x_i, p_k) dx_m. \quad 28$$

Comme y insiste d'emblée Cartan dès son article de 1899 [90], il y a deux problèmes complémentaires, étroitement intriqués mais toutefois bien différents, dans le problème de Pfaff :

1. D'abord *réduire* par des changements de variables appropriés la 1-forme $\omega = \sum_{i=1}^{i=n} A_i dx_i$, conçue comme *expression symbolique*, à une forme la plus simple possible, ce qu'on appelait une "forme canonique" et, plus tard, une "forme normale". La méthode consiste à réduire pas à pas le nombre des différentielles dx_i de l'expression en baissant d'une unité à chaque pas en intégrant complètement un système d'EDO.
2. Ensuite *résoudre* par des méthodes uniformes et systématiques (non *ad hoc*) l'équation $\omega = 0$. Le résultat fondamental est qu'une telle EDT peut toujours se ramener à un système de $\frac{n}{2} = m$ équations intégrales si $n = 2m$ est pair et de $\frac{n-1}{2} = m$ équations intégrales si $n = 2m + 1$ est impair.

Alors que le second problème relevait de techniques d'intégration déjà éprouvées à l'époque et appartenant à l'analyse fonctionnelle, le premier relevait de cette géométrie de l'intégrabilité encore absente et qui constituera la principale découverte et innovation de la pléiade de mathématiciens inspirés évoqués plus haut, et au premier chef, Pfaff lui-même, puis Frobenius avec la découverte du "covariant bilinéaire" (la dérivée extérieure $d\omega$) de ω , et Lie avec les transformations de contact.

Ainsi posé, le problème de Pfaff de l'intégration des EDT est le *problème inverse* (compliqué) du problème direct (trivial) de la différentiation. Comme nous l'avons déjà vu à la section 7 du chapitre 2, si $F(x_i)$ est une fonction différentiable des

28. C'est la méthode de Pfaff lui-même. On peut aussi considérer l'EDT (2) comme une EDT à $2m + 1$ variables dont on connaît une intégrale première du système intégral équivalent. Ce sera la méthode de Natani.

variables x_i sur \mathbb{R}^n , l'EDT

$$\omega = dF = \sum_{i=1}^{i=n} \frac{\partial F}{\partial x_i} dx_i = 0 \quad (\text{DT})$$

a pour solutions triviales les hypersurfaces $F(x_i) = a = \text{cste}$. Dans ce cas, ω est une différentielle exacte. Le premier aspect du problème inverse est de trouver F connaissant ω . Pour cela on part de deux remarques simples.

1. Comme $\frac{\partial^2 F}{\partial x_k \partial x_i} = \frac{\partial^2 F}{\partial x_i \partial x_k}$, si $\omega = 0$ est une EDT exacte, alors $d\omega = \omega' \equiv 0$ ²⁹ et donc une condition nécessaire est que $\frac{\partial A_i}{\partial x_k} = \frac{\partial A_k}{\partial x_i}$, ce que l'on exprime en disant que la matrice antisymétrique \mathfrak{A} des $\mathbf{a}_{ik} = \frac{\partial A_i}{\partial x_k} - \frac{\partial A_k}{\partial x_i}$ est identiquement nulle.³⁰ Ce système (CI₁) de $\frac{1}{2}n(n-1)$ conditions est la plus forte des conditions d'intégrabilité.

$$\mathbf{a}_{ik} = \frac{\partial A_i}{\partial x_k} - \frac{\partial A_k}{\partial x_i} = 0, \quad i, k = 1, \dots, n. \quad ((\text{CI}_1))$$

2. L'EDT $\omega = 0$ est équivalente aux équations $\rho\omega = 0$ si $\rho(x_i)$ est une fonction partout non nulle. Si donc la 1-forme ω n'est pas exacte, on peut se demander s'il n'est pas néanmoins possible de trouver une fonction $\rho \neq 0$ telle que $\rho\omega = dF$ soit exacte. On appelle ρ un *facteur intégrant* et la recherche des facteurs intégrants a constitué une grande partie des recherches du XVIII^e siècle sur l'intégrabilité. C'était déjà un problème inverse non trivial.

Mais pour que $\omega = 0$ soit intégrable grâce à un facteur intégrant ρ , un système de conditions nécessaires doit être satisfait. En effet si $\rho A_i = \frac{\partial F}{\partial x_i}$ pour $i = 1, \dots, n$, alors $\frac{\partial^2 F}{\partial x_k \partial x_i} = \frac{\partial^2 F}{\partial x_i \partial x_k}$ implique $\frac{\partial(\rho A_i)}{\partial x_k} = \frac{\partial(\rho A_k)}{\partial x_i}$ et donc

$$\begin{aligned} \rho \frac{\partial A_i}{\partial x_k} + A_i \frac{\partial \rho}{\partial x_k} &= \rho \frac{\partial A_k}{\partial x_i} + A_k \frac{\partial \rho}{\partial x_i} \\ \rho \left(\frac{\partial A_i}{\partial x_k} - \frac{\partial A_k}{\partial x_i} \right) &= \rho \mathbf{a}_{ik} = A_k \frac{\partial \rho}{\partial x_i} - A_i \frac{\partial \rho}{\partial x_k}. \end{aligned}$$

On peut éliminer ρ en considérant trois indices i, k, l . On multiplie chaque équation par le A_m d'indice manquant et on somme les trois résultats :

29. Rappelons (chapitre 2, section 7.3.4) que ω' est la notation initiale de Cartan [90] pour la dérivée extérieure $d\omega$.

30. Tous les \mathbf{a}_{ik} sont nuls et pas seulement le déterminant de \mathfrak{A} .

$$\begin{aligned}
\rho A_l \mathbf{a}_{ik} &= A_l A_k \frac{\partial \rho}{\partial x_i} - A_l A_i \frac{\partial \rho}{\partial x_k} \\
\rho A_i \mathbf{a}_{kl} &= A_i A_l \frac{\partial \rho}{\partial x_k} - A_i A_k \frac{\partial \rho}{\partial x_l} \\
\rho A_k \mathbf{a}_{li} &= A_k A_i \frac{\partial \rho}{\partial x_l} - A_k A_l \frac{\partial \rho}{\partial x_i} \\
\rho (A_l \mathbf{a}_{ik} + A_i \mathbf{a}_{kl} + A_k \mathbf{a}_{li}) &= A_l A_k \frac{\partial \rho}{\partial x_i} - A_i A_k \frac{\partial \rho}{\partial x_l} \\
&\quad + A_i A_l \frac{\partial \rho}{\partial x_k} - A_l A_i \frac{\partial \rho}{\partial x_k} \\
&\quad + A_k A_i \frac{\partial \rho}{\partial x_l} - A_k A_l \frac{\partial \rho}{\partial x_i} \\
&= 0 .
\end{aligned}$$

D'où, puisque $\rho \neq 0$, les conditions nécessaires

$$\mathbf{a}_{ik} A_l + \mathbf{a}_{kl} A_i + \mathbf{a}_{li} A_k = 0, \quad i, k, l = 1, \dots, n, \quad ((\text{CI}_2))$$

qui ne dépendent que des coefficients de ω .³¹ Elles ne sont pas indépendantes et se réduisent à $\frac{1}{2}(n-1)(n-2)$ d'entre elles. Elles entraînent pour 4 indices i, j, k, l les relations

$$\mathbf{a}_{ij} \mathbf{a}_{kl} + \mathbf{a}_{ik} \mathbf{a}_{lj} + \mathbf{a}_{il} \mathbf{a}_{jk} = 0 .$$

Dans son ouvrage fondateur *Institutiones calculi integralis* de 1748 (publié en 1755, voir [176] E212, pour une édition ultérieure), Euler avait déjà introduit explicitement cette condition pour les EDT de trois variables $P(x, y, z) dx + Q(x, y, z) dy + R(x, y, z) dz$ sous la forme

$$P \left(\frac{\partial Q}{\partial z} - \frac{\partial R}{\partial y} \right) + Q \left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z} \right) + R \left(\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} \right) = 0 .$$

Si les conditions (CI₂) sont satisfaites, des équations pour ρ sont faciles à obtenir. Elles donnent la dérivée logarithmique $\frac{d\rho}{\rho}$ en fonction des coefficients de ω . On part de $\rho A_i = \frac{\partial F}{\partial x_i}$ et on prend les différentielles, d'où

$$A_i d\rho + \rho \sum_{k=1}^{k=n} \frac{\partial A_i}{\partial x_k} dx_k = \sum_{k=1}^{k=n} \frac{\partial^2 F}{\partial x_k \partial x_i} dx_k$$

31. Evidemment si les \mathbf{a}_{ik} sont nuls (condition (CI₁)) la condition (CI₂) est trivialement satisfaite.

et on y remplace $\frac{\partial^2 F}{\partial x_k \partial x_i}$ par $\frac{\partial^2 F}{\partial x_i \partial x_k} = \frac{\partial}{\partial x_i} (\rho A_k) = \rho \frac{\partial A_k}{\partial x_i} + A_k \frac{\partial \rho}{\partial x_i}$, puis on en déduit

$$A_i d\rho + \rho \sum_{k=1}^{k=n} \mathbf{a}_{ik} dx_k = \frac{\partial \rho}{\partial x_i} \sum_{k=1}^{k=n} A_k dx_k = \frac{\partial \rho}{\partial x_i} \frac{dF}{\rho},$$

soit, puisque $dF = 0$ si $\omega = 0$,

$$\frac{d\rho}{\rho} = -\frac{1}{A_i} \sum_{k=1}^{k=n} \mathbf{a}_{ik} dx_k.$$

On montre que les conditions (CI₂) sont également suffisantes et, depuis Euler, il existe d'efficaces méthodes inductives sur la dimension n qui permettent de calculer un facteur intégrant.

Forsyth ([188], p. 16 sq.) donne l'exemple simple suivant à 4 variables. Soit

$$\begin{aligned} \omega &= A_x dx + A_y dy + A_z dz + A_u du \\ &= z(y+z) dx + z(u-x) dy + y(x-u) dz + y(y+z) du \end{aligned}$$

On vérifie immédiatement que $\mathbf{a}_{xy} = z - (-z) = 2z$, $\mathbf{a}_{xz} = y + 2z - y = 2z$, $\mathbf{a}_{xu} = 0 - 0 = 0$, $\mathbf{a}_{yz} = (u-x) - (x-u) = 2(u-x)$, $\mathbf{a}_{yu} = z - (2y+z) = -2y$, $\mathbf{a}_{zu} = -y - y = -2y$ et que les conditions d'intégrabilité (CI₂) et leurs conséquences sont satisfaites. Par exemple

$$\mathbf{a}_{xy}\mathbf{a}_{zu} + \mathbf{a}_{xz}\mathbf{a}_{uy} + \mathbf{a}_{xu}\mathbf{a}_{yz} = 2z(-2y) + 2z(2y) + 0 = 0$$

Pour trouver un facteur intégrant on peut utiliser l'équation

$$\begin{aligned} -\frac{d\rho}{\rho} &= \frac{1}{A_1} \sum_{k=1}^{k=4} \mathbf{a}_{1k} dx_k \\ -\frac{d\rho}{\rho} &= \frac{1}{z(y+z)} (2z dy + 2z dz) = 2 \frac{d(y+z)}{y+z} \end{aligned}$$

ce qui donne le facteur intégrant $\rho = \frac{1}{(y+z)^2}$ et la différentielle exacte dF de $F = \frac{xz+uy}{y+z}$.

Remarque. Soit $\omega = dy - p dx = A_x dx + A_y dy + A_p dp$, $A_x = -p$, $A_y = 1$, $A_p = 0$, notre 1-forme de contact privilégiée. On a $\mathbf{a}_{xy} = 0$, $\mathbf{a}_{xp} = -1$, $\mathbf{a}_{yp} = 0$ et les conditions (CI₁) ne sont pas satisfaites. Par ailleurs $\mathbf{a}_{xy}A_p + \mathbf{a}_{yp}A_x + \mathbf{a}_{px}A_y = 1$ et la condition (CI₂) n'est pas non plus satisfaite (il n'y a pas de facteur intégrant). Nous retrouvons donc à la main, dans le style eulérien du XVIII^e siècle, que l'EDT $\omega = 0$ n'est pas intégrable au sens de l'époque. On voit le chemin immense qui restait à parcourir entre ce simple constat d'analyse différentielle et la très riche géométrie des structures de contact. Il s'agit bien de passer du constat d'obstructions à l'intégrabilité à

la géométrie de contraintes définissant les conditions de possibilité de l'intégrabilité. C'est le "retournement structural" galloisien — la "révolution copernicienne" — opéré entre Pfaff et Cartan. □

2.5. La "véritable" intégrabilité chez Monge

Pendant longtemps, le problème de l'intégrabilité des EDT $\omega = 0$ était considéré comme "impossible", "absurde", "sans signification" si les conditions (CI₂) garantissant l'existence d'un facteur intégrant n'étaient pas satisfaites. Gaspard Monge (1746-1818) fut l'un des premiers à lever l'obstruction dans son Mémoire [374] de 1784 de l'Académie Royale des Sciences dont le titre est tout à fait explicite : *"Supplément. Où l'on fait voir que les Équations aux différences ordinaires, pour lesquelles les conditions d'intégrabilité ne sont pas satisfaites, sont susceptibles d'une véritable intégration, et que c'est de cette intégration que dépend celle des équations aux différences partielles élevées"*.

Monge considère des EDO du premier ordre qui ne sont pas forcément linéaires. Lorsque le degré est > 1 , par exemple $Mdx^2 + Ndy^2$ il parle d'équations "élevées". Il considère aussi des équations d'ordre > 1 comme $Ld^2y + Md^2x + N = 0$. Il explique alors qu'en dimension 3 les conditions d'intégrabilité (qu'il formule explicitement) signifient que les solutions de l'EDT $\omega = 0$ sont des surfaces $F(x_i) = a$ de \mathbb{R}^3 mais que si les conditions ne sont pas satisfaites, il existe quand même des solutions mais que celles-ci sont des courbes gauches (courbes qu'il appelait "à double courbure") intersection de deux surfaces $G(x_i) = b$.³² Les conditions d'intégrabilité expriment qu'il existe des familles de telles solutions de dimension 1 formant des surfaces qui sont des solutions de dimension 2. Donnons un coup d'œil à son Mémoire.³³

Partant du cas de dimension $n = 2$, $\omega = M(x, y)dx + N(x, y)dy = 0$, Monge note que cette EDT définit "l'inclinaison de la tangente" à une courbe "étant pris un point à volonté" (p. 502) i.e. la pente de la tangente à la courbe-solution passant par un point donné. Puis, passant à la dimension $n > 2$, $\omega = Ldx + Mdy + Ndz + \dots = 0$ il note que les EDT "qui ne satisfont pas à certaines conditions, sont (\dots) considérées comme absurdes". Et il annonce alors son projet :

"Je me propose de faire voir qu'il n'y a aucune équation différentielle qui soit absurde, si toutefois l'on entend par ce mot (\dots) une propriété impossible, imaginaire. (\dots) Je montrerai qu'elles [les équations différentielles] sont toutes susceptibles d'une véritable intégration." (p. 504)

32. Fondateur de la géométrie descriptive, Monge était un remarquable spécialiste de la construction explicite d'intersections de surfaces dans \mathbb{R}^3 , expertise qui lui a peut-être permis de mieux comprendre où se trouvait le problème.

33. Nous modifions légèrement les notations de Monge pour qu'elles soient compatibles avec les nôtres et nous indiquons les numéros de page dans le texte.

Voilà une formulation parfaite d'une levée d'obstruction par élargissement du champ des possibles. Ce texte est tout à fait remarquable.

Monge part d'exemples simples d'obstructions à l'intégrabilité et explique qu'il ne s'agit pas d'"exceptions" mais au contraire d'exemples élémentaires d'un problème fondamental, celui qui prendra le nom trente ans plus tard (1815, trois ans avant la disparition de Monge) de "problème de Pfaff".

"L'on n'avait pas remarqué que c'était le commencement d'une chaîne immense, à laquelle tenaient les grandes difficultés du calcul différentiel." (p. 505)

Le retournement est spectaculaire : des cas d'obstructions jusque-là négligés deviennent au contraire le symptôme de problèmes fondationnels éminents. Et Monge explicite son idée directrice dans le cas $n = 3$:

"Parmi les équations aux différences ordinaires à trois variables, celles qui satisfont à la (...) *condition d'intégrabilité* appartiennent toutes à des surfaces courbes³⁴, & l'intégrale de chacune d'elles est une équation unique, complétée par une seule constante arbitraire³⁵; mais toutes les équations qui ne satisfont pas à cette condition, sont en nombre infini, & elles n'appartiennent pas à des surfaces; leurs lieux sont des courbes à double courbure³⁶, tracées dans l'espace³⁷, & l'intégrale de chacune d'elles est le système de deux équations simultanées³⁸." (p. 505)

Et Monge conclut :

"Ainsi l'objet des équations connues sous le nom de *conditions d'intégrabilité*, n'est pas, comme on l'a cru jusqu'ici, d'indiquer celles des équations différentielles dont les intégrales sont possibles, mais de faire connaître le nombre des équations finies simultanées dont doivent être composées les intégrales qui sont toujours possibles." (p. 506)

On voit apparaître ici une intuition profonde qui conduira progressivement aux structures de contact et à leurs nombreuses généralisations.

1. Les solutions de l'EDT sur \mathbb{R}^n ont une certaine *dimension* d et une *codimension* $c = n - d$. Lorsque les conditions d'intégrabilité sont satisfaites, la dimension est $d = n - 1$ et la codimension est $c = 1$, les solutions étant des hypersurfaces. Mais la codimension peut-être plus grande et les solutions sont alors l'intersection de p hypersurfaces avec $d = n - p$ et $c = p$. Le calcul de p à

34. Pour Monge, dire qu'une EDO "appartient" à des surfaces, c'est dire que leurs solutions sont des surfaces $F(x, y, z) = a$.

35. C'est l'équation $F(x, y, z) = a$.

36. Des courbes gauches dans \mathbb{R}^3 .

37. L'espace \mathbb{R}^3 .

38. La courbe gauche est l'intersection de deux surfaces $F(x, y, z) = a$ et $G(x, y, z) = b$.

partir des coefficients de l'EDT $\omega = 0$ et la classification des différents types d'EDT possibles au moyen de ce que l'on appellera leurs "formes normales" sera l'un des principaux problèmes résolus au XIX^e siècle.

2. Ensuite une autre grande étape consistera à *géométriser* le problème en termes de champs d'hyperplans tangents. C'est ce que nous avons vu avec la structure de contact. L'EDT linéaire $\omega = 0$ définit en tout point v de l'espace ambiant V l'hyperplan $K_v = \ker \omega$ de $T_v V$ et donc un champ d'*hyperplans tangents*. Toute la question est alors de savoir comment ce champ se "vrille" au voisinage de chaque point.

Et, après avoir explicité quelques exemples d'EDO du premier ordre non linéaires, Monge poursuit par ces lignes véritablement prophétiques :

"Ces considérations ouvrent un nouveau champ à l'analyse & à la géométrie, & elles donnent lieu à un calcul intégral qui mérite l'attention des Géomètres car on verra dans la suite que l'intégration des équations aux différences partielles élevées, ne dépend que de ce genre de calcul." (p. 515)

Donnons un petit exemple instructif du traitement par Monge d'une EDT du premier ordre linéaire dans \mathbb{R}^3 , $dz = xy(xdx + ydy)$ qui ne satisfait pas les conditions d'intégrabilité.³⁹ Si l'on pose $dz = pdx + qdy$, on doit résoudre $p = x^2y$ et $q = xy^2$. On résout la première équation à y constant ($dy = 0$) en posant $p = \frac{dz}{dx}$, $z(x, y)$ étant considéré comme une fonction $z_y(x)$ de x avec un paramètre y . On trouve $z = \frac{x^3y}{3} + \varphi(y)$. En dérivant par rapport à y reconsidéré comme une variable, on trouve $q = \frac{dz}{dy} = \frac{x^3}{3} + \varphi'(y)$ et comme $q = xy^2$, on doit satisfaire l'équation $\varphi'(y) = xy^2 - \frac{x^3}{3}$ entre x et y . Donc z est en fait une fonction d'une variable et définit une courbe gauche dans \mathbb{R}^3 . Vu la forme très particulière de l'équation initiale on peut préciser la nature géométrique du problème. L'équation initiale peut s'écrire $p = x^2y$ et $py - qx = x^2y^2 - x^2y^2 = 0$. Mais cette dernière équation implique que z soit de la forme $z = \psi(x^2 + y^2)$ et pour satisfaire $p = \frac{\partial z}{\partial x} = x^2y$, il faut $2x\psi'(x^2 + y^2) = x^2y$, c'est-à-dire $2\psi'(x^2 + y^2) = xy$. Toute solution est par conséquent une courbe gauche

"tracée sur une surface quelconque de révolution, l'axe coïncidant avec la ligne des z ." (p. 530)

Et Monge de conclure

"La Géométrie peut encore faire de très grands progrès, parce qu'on a le moyen de mettre en analyse des manières nouvelles d'engendrer les courbes,

39. $A_x = x^2y$, $A_y = xy^2$, $A_z = -1$, $a_{xy} = x^2 - y^2$, $a_{xz} = 0$, $a_{yz} = 0$ et les conditions (CI₁) ne sont pas satisfaites. Par ailleurs $a_{xy}A_z + a_{yz}A_x + a_{zx}A_y = -(x^2 - y^2)$ et la condition (CI₂) n'est pas non plus satisfaite.

& parce qu'on a la faculté d'entendre un grand nombre de propriétés de l'étendue, qui sont exprimées par les relations qu'ont entre elles des équations jusqu'ici regardées comme impossibles." (p. 573)

2.6. Autour du Mémoire de Pfaff (les années 1810-1820)

2.6.1. L'idée générale de l'intégration d'une équation de Pfaff.

Si l'on suit l'intuition de Monge, l'idée générale de l'intégration d'une équation de Pfaff (une EDT du premier ordre linéaire) $\omega = \sum_{i=1}^{i=n} A_i dx_i = 0$ est d'abord de comprendre comment les coefficients A_i déterminent la *dimension maximale* d_ω des solutions possibles, puis seulement ensuite de trouver des formules intégrales. Il y a tout un éventail de possibilités. À un extrême, quelle que soit ω , on a toujours les solutions triviales de dimension minimale $d = 0$ que sont les points $x_i = \text{cste}$ car dans ce cas tous les dx_i sont nuls. À l'autre extrême, si ω est exacte, $\omega = dF$, ou s'il existe un facteur intégrant ρ tel que $\rho\omega = dF$, les solutions sont des hypersurfaces $F = a$ de dimension maximale $d = n - 1$.

On peut aborder le problème de trouver les solutions de dimension maximale d_ω de deux façons, ascendante ou descendante. Dans la perspective ascendante, on part de la remarque qu'il existe toujours des solutions $x_i(t)$ de dimension 1. Certaines sont évidentes. En effet, on doit avoir $\sum_{i=1}^{i=n} A_i x'_i(t) dt = 0$ et donc $\sum_{i=1}^{i=n} A_i x'_i(t) = 0$. Soit par exemple \widehat{A}_i le produit des A_k sauf A_i . Si n est pair, $n = 2m$, prenons $x'_i(t) = \widehat{A}_i$ pour m des x_i et $x'_i(t) = -\widehat{A}_i$ pour les m autres. On peut ainsi annuler des paires (dx_1, dx_2) de dx_i ce qui consiste à considérer des courbes dans les hyperplans de dimension $n - 2$ définis par $x_1 = \text{cste}$ et $x_2 = \text{cste}$. On obtient des courbes solutions en résolvant ces systèmes d'EDO. Si n est impair, on peut prendre l'un des x_i constant et prendre une solution du système réduit à $n - 1$ variables.⁴⁰

Ensuite, on peut se demander s'il existe des familles de courbes intégrales formant des *surfaces* intégrales, puis des familles de surfaces intégrales formant des solutions de dimension 3, etc. jusqu'à la dimension $n - 1$. Si la condition d'intégrabilité n'est pas satisfaite on rencontre une obstruction à un certain niveau. Par exemple dans le cas de notre structure de contact familière $dy - p dx$ il n'existe pas de surfaces constituées de relevées legendriennes et qui satisfassent l'EDT.

Dans la perspective descendante, on essaye de réduire au maximum le nombre de variables en éliminant certaines des différentielles dx_i au moyen de changements

40. Exemple. Pour $dy - p dx$ on a les solutions triviales $dx = dy = 0$, p quelconque, c'est-à-dire les fibres au-dessus d'un point a du plan (x, y) et $dy = p = 0$, c'est-à-dire les droites $y = \text{cste}$ dans le plan (x, y) d'équation $p = 0$. Et il y a toutes les autres solutions non triviales que sont les relevées legendriennes.

de variables appropriés. On obtient des intégrales premières et les solutions de dimension d sont alors construites non plus comme des familles à $d - 1$ paramètres de courbes mais comme intersections de codimension $c = n - d$ de c hypersurfaces.

2.6.2. *Le résultat principal.*

C'est avec le célèbre *Mémoire* [447] de Johann Friedrich Pfaff (1765-1825), paru en 1815 dans les *Abhandlungen* de l'Académie Royale des Sciences de Berlin, que le problème inverse de l'intégrabilité des EDT (baptisées plus tard en son honneur équations "de Pfaff") se trouve formulé et systématisé. Le texte, rédigé en latin et intitulé "*Methodus generalis, aequationes differentiarum partialium, nec non aequationes differentiales vulgares, utrasque primi ordinis inter quocumque variables, complete integrandi*", étudie les équations $\omega = 0$ d'un nombre pair ($2m$) ou impair ($2m + 1$) de variables qui sont "inconditionnées" — on dirait maintenant "génériques" — au sens où il n'existe pas de relations particulières entre les coefficients.

Forsyth explique en détail dans son traité [188] de 1890 que Pfaff traita le cas impair, que Gauss [205] (1815) et Jacobi [268] (1827) démontrèrent les deux cas. La méthode est de réduire graduellement le nombre d'éléments différentiels de l'EDT en résolvant des systèmes "subsidiaries" d'EDO qui ont une solution précisément parce que les conditions d'intégrabilité *ne sont pas* satisfaites. Ce qui était considéré pendant longtemps comme une obstruction — et donc un obstacle paralysant — devient alors, comme Monge l'avait prophétisé, le moteur même de la résolution du problème — un fil directeur méthodologique ! La non-satisfaction des conditions d'intégrabilité signifie que les solutions existent mais sont de dimension *inférieure* (de codimension supérieure) à la dimension maximale possible $d_{\max} = n - 1$ (à la codimension minimale possible $c_{\min} = 1$) et le résultat fondamental est que, malgré cela, il existe toujours des solutions ayant au moins une certaine dimension (il existe évidemment toujours les solutions ponctuelles de dimension 0).

Soit donc $\omega = \sum_{i=1}^{i=n} A_i(\tilde{x}) dx_i = 0$, $\tilde{x} = (x_1, \dots, x_n)$,⁴¹ une équation de Pfaff (E) à n variables telle que les conditions d'intégrabilité (CI₁) et (CI₂) ne soient pas satisfaites. Même s'il n'existe pas par hypothèse de facteur intégrant ρ et de fonction $F(x_i)$ tels que $dF = \rho\omega$, (E) peut néanmoins avoir un *équivalent intégral* au sens où il existe plusieurs fonctions F_s , $s = 1, \dots, c$ telles que (E) soit équivalente au système d'équations $\{F_s = a_s\}$, $s = 1, \dots, c$; ce qui signifie que les solutions de (E) sont des sous-variétés de codimension c de \mathbb{R}^n . La question de l'intégration devient alors de trouver un système équivalent de ce type qui soit le plus simple possible. Le résultat fondamental de Pfaff est le théorème de réduction suivant :

41. Dans la suite nous notons parfois \tilde{x} plutôt que x une multivariable pour améliorer la lisibilité des changements de variables.

Théorème de réduction de Pfaff. On peut toujours réduire une EDT (E) à une EDT de Pfaff (E') équivalente à $m' \leq m = \frac{n}{2}$ éléments différentiels si $n = 2m$ est pair et $m' \leq m = \frac{n-1}{2}$ éléments différentiels si $n = 2m + 1$ est impair. —◇

Cela signifie que toute équation de Pfaff a des solutions de dimension au moins m si $n = 2m$ ou $n = 2m + 1$, m étant le minimum de la dimension maximale des solutions possibles.

2.6.3. Difféomorphismes et élimination d'éléments différentiels.

La méthode de Pfaff consiste à trouver de bons changements de variables permettant d'éliminer des symboles différentiels. Répétons que toute EDT

$$\omega = \sum_{i=1}^{i=n} A_i dx_i = 0$$

sur \mathbb{R}^n a évidemment les points $a = (x_i = a_i = \text{cste})$ comme solutions de codimension maximale n (de dimension minimale 0) puisqu'alors tous les dx_i sont nuls. Mais si par un changement de variables approprié $\omega = \sum_{i=1}^{i=n} A_i dx_i$ peut s'écrire $\omega = \sum_{r=1}^{r=m} U_r du_r$ avec $m < n$ alors les sous-variétés de codimension m (de dimension $n - m$) d'équations ($u_r = c_r = \text{cste}$) sont des solutions de dimension $n - m > 0$ puisque tous les du_r sont nuls. Si m est minimal, alors l'EDT $\omega = 0$ a des solutions de dimension au moins $n - m$. Le cas maximal est celui où ω est exacte, $\omega = dF$ admettant dans ce cas les solutions $F = c = \text{cste}$ de dimension maximale $n - 1$.

C'est donc l'élimination d'éléments différentiels qui se trouve au cœur de la méthode.

Les changements de variables sont des difféomorphismes φ de \mathbb{R}^n (plus ou moins locaux ou globaux suivant la situation considérée). Ils sont formulés fonctionnellement comme des $\{u_i(x_j)\}$, $i, j = 1, \dots, n$ dont le déterminant du jacobien $\left(\frac{\partial u_i}{\partial x_j}\right)$ est partout non nul. Au niveau de l'intuition géométrique ils signifient que l'on déforme l'espace \mathbb{R}^n tout en préservant la contrainte que la grille euclidienne des axes et des plans de coordonnées de dimensions $1, \dots, n - 1$ reste une bonne grille régulière sans singularités : les relations de *transversalité* doivent être préservées, même si elles ne sont plus des relations d'orthogonalité. Travailler à changement de coordonnées près c'est donc travailler à difféomorphisme près. Or les difféomorphismes φ forment un groupe G . Travailler sur une structure à difféomorphisme près c'est donc travailler sur les propriétés *G-invariantes* de cette structure et donc, dans bien des cas, réduire la situation à une forme normale canonique.⁴² Et comme les φ peuvent être extrêmement variés comme fonctions, on comprend pourquoi la question de

42. Si par exemple $X(x) = (X_i(x_j))$ est un champ de vecteurs différentiable sur \mathbb{R}^n et si $X(a) \neq 0$, alors, quelle que soit la complexité fonctionnelle des $X_i(x_j)$, on peut toujours trouver au voisinage de a un changement de variables tel que X devienne le champ $X_1(a) = 1$, $X_i(0) = 0$ pour $i > 1$. Cette trivialisatation est une forme normale universelle.

l'intégrabilité d'EDT de Pfaff $\omega = 0$ change de nature, n'est plus celui de trouver des intégrales *explicites* et se convertit en un problème essentiellement *géométrique* concernant des relations d'incidence (tangentialité, transversalité, etc.) entre divers éléments de contact.

On voit l'analogie frappante avec l'approche de Galois. On ne cherche plus des solutions explicites du problème considéré mais des conditions de résolubilité issues du groupe d'invariance du problème.

Suivant les cas, ces opérations pourront être globales (appliquées à \mathbb{R}^n tout entier), locales (appliquées à une carte locale d'une variété), ou même infinitésimales. Dans ce dernier cas, les changements de coordonnées deviennent *linéaires* et l'on travaille dans des espaces vectoriels, des sous-espaces vectoriels et des matrices. Les problèmes deviennent des problèmes d'algèbre linéaire et les invariants possédant un sens géométrique deviennent des invariants comme le fait qu'un déterminant soit non nul, qu'une trace soit nulle, qu'une matrice ait tel rang, etc. C'est pourquoi le "retournement galoisien" du calcul différentiel classique est inséparable de l'émergence et du développement de l'algèbre linéaire qui en sont contemporains.

2.6.4. La méthode de Pfaff.

La méthode de Pfaff consiste à trouver de bons changements de variables que l'on puisse obtenir en intégrant des systèmes d'EDO dérivables de la structure de ω . Pfaff considère $n - 1$ fonctions $u_r(x_i)$, $r = 1, \dots, n - 1$, fonctionnellement indépendantes entre elles de façon à ce que, avec (par exemple) x_n , les n variables $\tilde{u} = (u_r, x_n)$ définissent un changement de variables, ce qui signifie que le déterminant fonctionnel $\frac{D(x_1, \dots, x_{n-1})}{D(u_1, \dots, u_{n-1})}$ est $\neq 0$. On peut donc inverser les fonctions et les x_i , $i = 1, \dots, n - 1$ deviennent alors des fonctions $x_i(\tilde{u})$ des $\tilde{u} = (u_r, x_n)$ ⁴³ et ne sont plus indépendantes de x_n par rapport aux nouvelles coordonnées.⁴⁴ Prenons par exemple 3 variables (x, y, z) et un changement de variables (u, v, z) . Cela signifie que dans chaque plan horizontal $z = z_0$ de l'espace (u, v, z) les courbes $u(x, y_0, z_0)$ et $v(x_0, y, z_0)$ forment un système de coordonnées curvilignes et que ce système se déforme avec z . Réciproquement, dans chaque plan horizontal $z = z_0$ de l'espace (x, y, z) les courbes $x(u, v_0, z_0)$ et $y(u_0, v, z_0)$ forment un système de coordonnées curvilignes qui se déforme avec z .

43. On doit évidemment garder $x_n = x_n$.

44. Les $\frac{\partial x_i}{\partial x_n}$, $i \neq n$, ne sont donc plus forcément nulles.

Par un tel changement de variable, (E) se transforme en (E')

$$\begin{aligned}\tilde{\omega} &= \sum_{r=1}^{r=n-1} B_r(\tilde{u}) du_r + B_n(\tilde{u}) dx_n = 0, \text{ avec} \\ B_r(\tilde{u}) &= \sum_{i=1}^{i=n-1} A_i(\tilde{x}(\tilde{u})) \frac{\partial x_i(\tilde{u})}{\partial u_r} \\ B_n(\tilde{u}) &= \sum_{i=1}^{i=n-1} A_i(\tilde{x}(\tilde{u})) \frac{\partial x_i(\tilde{u})}{\partial x_n} + A_n(\tilde{x}(\tilde{u})).\end{aligned}$$

À condition que $\tilde{u} = \{u_r, x_n\}$ reste un vrai changement de variables, on peut lui demander de satisfaire certaines conditions. On cherche alors de voir si l'on peut en particulier imposer les conditions :

- (i) $B_n(\tilde{u}) = 0$ et
- (ii) trouver pour les $B_r(\tilde{u})$ un facteur intégrant "partiel" $M(x_n) \neq 0$ ne dépendant que de x_n .

Cela signifierait que $B_r(\tilde{u}) = M(x_n) C_r(u_r)$ où les C_r ne dépendent que des seules u_r et non plus de x_n et que les $B_r(\tilde{u})$ satisfont par conséquent des équations

$$\frac{1}{B_r(\tilde{u})} \frac{\partial B_r(\tilde{u})}{\partial x_n} = \frac{1}{M(x_n)} \frac{\partial M(x_n)}{\partial x_n} = \mu(x_n).^{45}$$

Avec une telle *séparation des variables* (u_r) et x_n entre les C_r et M , on aurait alors, puisque $B_n(\tilde{u}) = 0$, l'EDT

$$\tilde{\omega} = M(x_n) \sum_{r=1}^{r=n-1} C_r(u_r) du_r = 0$$

et donc l'EDT (E), déjà équivalente à (E'), deviendrait équivalente à (E'')

$$\tilde{\omega}_0 = \sum_{r=1}^{r=n-1} C_r(u_r) du_r = 0$$

puisque $M(x_n) \neq 0$, EDT qui ne dépend plus que des $n - 1$ variables u_r . Lorsqu'elle est applicable, cette méthode permet donc de réduire le nombre de variables de n à $n - 1$, ce qui permet d'amorcer une procédure itérative de réduction.

Reste à savoir comment satisfaire les conditions.(i) et (ii). Des calculs simples montrent que la possibilité de réduction dépend du déterminant Δ de la matrice \mathfrak{A}

45. $\frac{1}{B_r} \frac{\partial B_r}{\partial x_n} = \frac{1}{M C_r} C_r \frac{\partial M}{\partial x_n} = \frac{1}{M} \frac{\partial M}{\partial x_n}$ puisque C_r ne dépend pas de x_n .

antisymétrique $n \times n$ des $\mathbf{a}_{ik} = \frac{\partial A_i}{\partial x_k} - \frac{\partial A_k}{\partial x_{ik}}$. En dérivant les

$$B_r(\tilde{u}) = \sum_{i=1}^{i=n-1} A_i(\tilde{x}(\tilde{u})) \frac{\partial x_i(\tilde{u})}{\partial u_r}$$

par rapport à x_n et $B_n(\tilde{u}) = 0$ par rapport aux u_r , on déduit facilement (en utilisant le fait que le déterminant jacobien $\frac{D(x_1, \dots, x_{n-1})}{D(u_1, \dots, u_{n-1})}$ est $\neq 0$) le système

$$\begin{cases} \mu A_i = \mathbf{a}_{in} + \sum_k \mathbf{a}_{ik} \frac{\partial x_k(\tilde{u})}{\partial x_n}, & i, k = 1, \dots, n-1 \\ \mu A_n = \sum_k \mathbf{a}_{nk} \frac{\partial x_k(\tilde{u})}{\partial x_n} \end{cases}$$

c'est-à-dire $\mu(A_1, \dots, A_n)^T = \mathfrak{A} \left(\frac{\partial x_1(\tilde{u})}{\partial x_n}, \dots, \frac{\partial x_{n-1}(\tilde{u})}{\partial x_n}, 1 \right)^T$ (où $(\bullet)^T$ est le vecteur colonne transposé du vecteur ligne (\bullet)). Dans ce système n'interviennent plus que μ et les $\frac{\partial x_k(\tilde{u})}{\partial x_n}$. On voit que si l'on introduit les nouvelles variables $y_r = \frac{1}{\mu} \frac{\partial x_r(\tilde{u})}{\partial x_n}$ et $y_n = \frac{1}{\mu}$, on se ramène au système linéaire (S) $\mathfrak{A}y = A$ dont la solution dépend de la possibilité d'inverser \mathfrak{A} et donc de son déterminant Δ .

1. Si $\Delta = 0$, il faut approfondir les choses. C'est obligatoirement le cas si n est impair car \mathfrak{A} est antisymétrique. Nous allons y revenir.
2. Si en revanche n est pair et si les conditions d'intégrabilité sont *maximalement* non satisfaites, autrement dit si $\Delta \neq 0$, alors la méthode de réduction peut s'appliquer. On obtiendra dans ce cas $y = \mathfrak{A}^{-1}A$, soit $y_r = \frac{1}{\Delta} Y_r(A)$ où les $Y_r(A)$ se calculent par l'inversion de la matrice \mathfrak{A} et son application au vecteur A . Comme $\mathfrak{A}^{-1} = \frac{1}{\Delta} \left((-1)^{i+j} \mathfrak{M}_{ij} \right)^T$ où les \mathfrak{M}_{ij} sont les mineurs de \mathfrak{A} , les y_r ont bien $\frac{1}{\Delta}$ en facteur.

À l'époque, ces calculs matriciels avec des déterminants et des mineurs n'existaient pas et cela a été une grande conquête que de trouver les formules générales et universelles d'algèbre linéaire.

La méthode permet d'éliminer la différentielle dx_n et de calculer le facteur μ et les $n-1$ fonctions $u_r(x_i)$. On résout le système (S), ce qui donne μ et les y_r et le système de Pfaff *subsidaire* des $(n-1)$ EDO

$$\frac{\partial x_r(\tilde{u})}{\partial x_n} = \mu y_r(\tilde{u}) = \frac{y_r(\tilde{u})}{y_n(\tilde{u})}$$

donnant les x_r . On obtient ainsi l'EDT (E'')

$$\tilde{\omega}_0 = \sum_{r=1}^{r=n-1} C_r(u_r) du_r = 0$$

pour laquelle les $u_r = \text{cte}$ donnent $n - 1$ intégrales premières. D'où les formules pour les courbes intégrales. On notera que le système de Pfaff est un champ de vecteurs $\frac{dx_r(\tilde{u})}{dx_n} = \frac{y_r(\tilde{u})}{y_n(\tilde{u})}$ dans les variables u_r avec x_n comme variable "temporelle". Il peut s'écrire sous la forme des $n - 1$ équations $\frac{dx_r(\tilde{u})}{y_r(\tilde{u})} = \frac{dx_n}{y_n(\tilde{u})}$.

Dans le cas pair $n = 2m$, Δ est nécessairement un carré $\Delta = P^2$. La matrice P a pris le nom de *Pfaffien*, terme introduit par Arthur Cayley en 1852. Pour $n = 2$, $\Delta = \det \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} = a^2$ et $P = a$. Pour $n = 4$,

$$\Delta = \det \begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix} = (af - be + dc)^2$$

et $P = af - be + dc$.⁴⁶

Pour tester toute l'efficacité de la méthode, revenons alors au problème de l'intégration des EDP tel que nous l'avons présenté à la section 2.4. On suppose que l'EDP (1) $f(x_i, z, p_i) = a$ permet de définir la fonction $p_m = \varphi(z, x_i, p_k)$, $k = 1, \dots, m - 1$, ce qui conduit à l'EDT (3) en $n = 2m$ variables (z, x_i, p_k) , $dz = \sum_{i=1}^{i=m-1} p_i dx_i + \varphi(z, x_i, p_k) dx_m$. Renommons les variables en posant $x_{m+k} = p_k$ pour $k = 1, \dots, m - 1$ et $x_{2m} = z$. L'EDT (3) s'écrit alors $\omega = \sum_{i=1}^{i=n} A_i dx_i$ avec $A_k = p_k$ pour $k = 1, \dots, m - 1$, $A_m = \varphi$, $A_{m+k} = 0$, $k = 1, \dots, m - 1$, $A_{2m} = -1$. La matrice \mathfrak{A} des (a_{rs}) est donc donnée par

$$\left(\begin{array}{c|cccc} & x_k & x_m & x_{m+k} = p_k & x_{2m} = z \\ \hline A_k = p_k & 0 & -\frac{\partial \varphi}{\partial x_k} & I & 0 \\ A_m = \varphi & \frac{\partial \varphi}{\partial x_k} & 0 & \frac{\partial \varphi}{\partial p_k} & \frac{\partial \varphi}{\partial z} \\ A_{m+k} = 0 & -I & -\frac{\partial \varphi}{\partial p_k} & 0 & 0 \\ A_{2m} = -1 & 0 & -\frac{\partial \varphi}{\partial z} & 0 & 0 \end{array} \right)$$

et $\mathfrak{A}y = A$ est donc le système ($k = 1, \dots, m - 1$)

$$\begin{cases} A_k = p_k = -\frac{\partial \varphi}{\partial x_k} y_m + y_{m+k} \\ A_m = \varphi = \sum_k \frac{\partial \varphi}{\partial x_k} y_k + \sum_k \frac{\partial \varphi}{\partial p_k} y_{m+k} + \frac{\partial \varphi}{\partial z} y_{2m} \\ A_{m+k} = 0 = -y_k - \frac{\partial \varphi}{\partial p_k} y_m \\ A_{2m} = -1 = -\frac{\partial \varphi}{\partial z} y_m \end{cases} \quad ((S))$$

46. Notons que si $\theta = \sum_{i < j} a_{ij} dx_i \wedge dx_j$ est la 2-forme associée à la matrice antisymétrique $n \times n$ d'ordre pair ($n = 2m$) $A = (a_{ij})$, alors $\frac{1}{m!} \theta^m = P(A) dx_1 \wedge \dots \wedge dx_n$.

Mais, par définition, $\varphi(z, x_i, p_k) = p_m$ et, comme $f(x_i, z, p_i) = a$, si on dérive f par rapport aux (x_i, z, p_i) , on obtient ($i = 1, \dots, m$)

$$\begin{cases} \frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial p_m} \frac{\partial \varphi}{\partial x_i} = 0 \\ \frac{\partial f}{\partial p_i} + \frac{\partial f}{\partial p_m} \frac{\partial \varphi}{\partial p_i} = 0 \\ \frac{\partial f}{\partial z} + \frac{\partial f}{\partial p_m} \frac{\partial \varphi}{\partial z} = 0 \end{cases}$$

On en tire les $\frac{\partial \varphi}{\partial x_i}$, $\frac{\partial \varphi}{\partial p_i}$, $\frac{\partial \varphi}{\partial z}$ et, en les reportant dans le système (S) on obtient ($k = 1, \dots, m-1$; $i = 1, \dots, m$) :⁴⁷

$$\begin{cases} y_m = -\frac{\partial f}{\partial p_m} : \frac{\partial f}{\partial z} \\ y_k = -\left(\frac{\partial f}{\partial p_k} : \frac{\partial f}{\partial p_m}\right) \left(\frac{\partial f}{\partial p_m} : \frac{\partial f}{\partial z}\right) = -\frac{\partial f}{\partial p_k} : \frac{\partial f}{\partial z} \\ y_{m+k} = p_k + \left(\frac{\partial f}{\partial x_k} : \frac{\partial f}{\partial p_m}\right) \left(\frac{\partial f}{\partial p_m} : \frac{\partial f}{\partial z}\right) = p_k + \frac{\partial f}{\partial x_k} : \frac{\partial f}{\partial z} \\ \varphi = p_m = \sum_k \left(\frac{\partial f}{\partial x_k} : \frac{\partial f}{\partial p_m}\right) \left(\frac{\partial f}{\partial p_k} : \frac{\partial f}{\partial z}\right) \\ \quad - \sum_k \left(\frac{\partial f}{\partial p_k} : \frac{\partial f}{\partial p_m}\right) \left(p_k + \left(\frac{\partial f}{\partial x_k} : \frac{\partial f}{\partial z}\right)\right) - \left(\frac{\partial f}{\partial z} : \frac{\partial f}{\partial p_m}\right) y_{2m} \\ y_{2m} = -p_m \left(\frac{\partial f}{\partial p_m} : \frac{\partial f}{\partial z}\right) - \sum_k p_k \left(\frac{\partial f}{\partial p_k} : \frac{\partial f}{\partial z}\right) = -\sum_i p_i \left(\frac{\partial f}{\partial p_i} : \frac{\partial f}{\partial z}\right) \end{cases}$$

d'où, en utilisant la forme ci-dessus $\frac{dx_r(\tilde{u})}{y_r(\tilde{u})} = \frac{dx_n}{y_n(\tilde{u})}$ et en multipliant tous les dénominateurs par $\frac{\partial f}{\partial z}$, le système de Pfaff subsidiaire de $n = 2m$ EDO

$$\frac{dx_i}{\frac{\partial f}{\partial p_i}} = -\frac{dp_i}{\frac{\partial f}{\partial x_i} + p_i \frac{\partial f}{\partial z}} = \frac{dz}{\sum_i p_i \frac{\partial f}{\partial p_i}} = C. \quad ((P))$$

On vérifie que (P) possède évidemment $f = a$ comme intégrale première puisque, avec

$$\begin{aligned} dx_i &= C \frac{\partial f}{\partial p_i} \\ dp_i &= -C \frac{\partial f}{\partial x_i} + p_i \frac{\partial f}{\partial z}, dz = C \left(\sum_i p_i \frac{\partial f}{\partial p_i} \right), \end{aligned}$$

on a bien

$$df = \frac{\partial f}{\partial x_i} dx_i + \frac{\partial f}{\partial p_i} dp_i + \frac{\partial f}{\partial z} dz = 0.$$

On intègre ce système (P) et on trouve ses $2m-1 = n-1$ autres intégrales $u_\ell = a_\ell$ indépendantes qui, avec $f = a$, forment un système complet d'intégrales de (P). On utilise l'une d'elle pour transformer l'EDP initiale et abaisser son nombre de

47. Pour améliorer la lisibilité des formules nous écrivons certains quotients A/B sous la forme $A : B$.

variables d'une unité. En itérant ce processus et en intégrant les systèmes de Pfaff subsidiaires successifs, on obtient en définitive m intégrales premières de l'EDP initiale qui sont des fonctions de (z, x_i, p_i) et de m constantes. Avec $f = a$, elles permettent d'éliminer tous les p_i et de trouver une équation entre z et les x_i avec m constantes, équation qui fournit l'intégrale complète de l'EDP.

2.6.5. La méthode du Methodus.

Après avoir explicité son résultat principal en un langage un peu plus moderne que celui de l'époque, donnons quelques indications sur la structure du *Methodus* de Pfaff [447] qui, rappelons-le, est rédigé en latin. Pfaff part des travaux du "sagacissimus" Lagrange sur la méthode générale de résolution d'intégration des EDP à trois variables. Il note la difficulté du programme dont même le "peritissimus calculus artifex" Euler avouait "qu'il était encore loin de résoudre le problème général". Au-delà de quatre variables se font jour des "difficultés inextricables" que ni Euler, ni Lagrange, ni Legendre, ni Monge n'ont réussi à surmonter et Pfaff a donc une conscience aigüe de l'importance de son mémoire.

Pfaff suit la voie ouverte par Monge et son "egregia observatio" que, "contra opinionem antea vulgo receptam", le fait que les conditions d'intégrabilité ne soient pas satisfaites ne signifie pas l'impossibilité d'intégrer mais simplement qu'il faut *plusieurs* intégrales pour intégrer (i.e. les solutions peuvent être de codimension > 1).

Il explique que pour avancer, il a dû abandonner ("deserere") les méthodes existantes, changer de méthode et faire appel ("vocare") à un "autre principe" "qui ignore le critère d'intégrabilité". Ce principe "dérive des sources du calcul lui-même" et il a donc clairement conscience d'un retournement de méthode, d'une "révolution copernicienne" qui remonte aux sources pour emprunter une nouvelle route. Ce retour amont aux sources pour ouvrir un nouveau parcours est illustré par le schéma de la figure 1.

La remarque initiale est, nous l'avons dit, qu'une expression différentielle comme $dz = p dx + q dy$ doit être considérée comme une expression reliant non seulement 5 variables (x, y, z, p, q) mais 5 différentielles, les coefficients de dp et dq étant nuls. Il s'agit donc en fait d'une expression à $5 = 2m + 1$ différentielles, $m = 2$, que l'on a réduit à $3 = m + 1$ différentielles. C'est la continuation de ce que faisait déjà Euler, à savoir considérer les dérivées partielles p et q de $z(x, y)$ comme de nouvelles variables indépendantes, des "dérivées cachées" pour reprendre encore une fois la belle expression de Richard Montgomery, variables devant être interprétées seulement dans un second temps comme des dérivées partielles.⁴⁸

48. Nous reviendrons longuement sur cette notion fondamentale de "dérivée cachée" dans les sections 8 du chapitre 9 (sur le *Methodus* d'Euler de 1744, texte fondateur du Calcul des Variations) et 3.2 du chapitre 15 (sur l'histoire de cette notion jusqu'à aujourd'hui). C'est le leitmotiv de la neurogéométrie : les variables supplémentaires sont implémentées par des

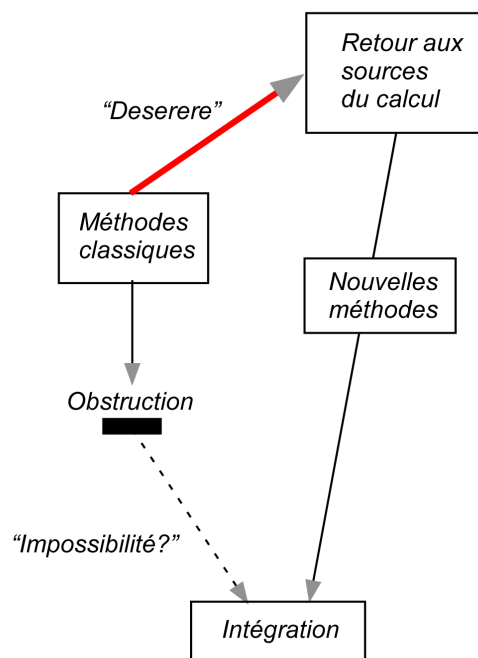


FIGURE 1. Le retour aux sources du calcul opéré par Pfaff pour ouvrir un nouveau parcours.

“Aequationes nimirum differentiarum partialium contemplari licet tanquam aequationes differentiales vulgaris generis truncatas inter plures variables, quam quae principaliter occurrunt, ipsis scilicet quotientibus differentialibus (p, q, \dots) variabilium loco habitis, quarum differentialia (dp, dq, \dots) ideo desunt, quoniam ea in zero ducta esse censentur.”

Autrement dit, si l’on a une fonction z de m variables x_i , on ajoute aux $m + 1$ “variables principales” (z, x_i) les m “variables accessoires” p_i que sont les dérivées (“quotients différentiales”⁴⁹) et l’EDT $dz = \sum_{i=1}^m p_i dx_i$ définit l’une des p_i (par exemple p_m) comme fonction des $2m$ autres variables “principales” et “accessoires”, p_m n’intervenant plus comme variable dans les calculs (“non in computum venit”). D’où le programme de recherche de bons changements de variables permettant d’éliminer des différentielles jusqu’à ce que l’on arrive à des formes normales.

Dans son *Methodus*, Pfaff va alors essentiellement développer toute une série de cas comprenant un nombre de plus en plus grand de variables. Dans le §3 il résout le “*Problema I. Aequationem differentiarum partialium inter tres variables complete*

détecteurs neuronaux post-rétiniens (corticaux) et leur opérativité comme “dérivées cachées” est garantie par l’architecture fonctionnelle de leur aire.

49. Les dérivées sont des “quotients” de différentielles.

integrare", i.e. l'intégration complète d'une EDT à trois variables $dz = pdx + qdy$, EDT qui doit être conçue en fait comme une équation différentielle à quatre variables (x, y, z, p) déterminant q en fonction de (x, y, z, p) , p n'étant pas forcément $\frac{\partial z}{\partial x}$. Il est très instructif d'observer la virtuosité avec laquelle il développe ses calculs. Il fait un changement de variables $(x, z, p) \rightarrow (a, b, c)$ de façon à ce que (x, z, p) deviennent des fonctions de (y, a, b, c) .⁵⁰ Il réécrit alors l'EDT et cherche les conditions sur (a, b, c) pour que y et dy disparaissent du calcul ("ut tam y , quam dy , ex calculo exeunt").

Il développe donc les trois différentielles dx, dz, dp sur dy, da, db, dc avec les coefficients (X, χ, χ', χ'') , etc. qui sont des fonctions explicites de (y, a, b, c) et implicites de (y, x, z, p) . Comme q est une fonction de (x, y, z, p) , dq se développe en $dq = q'dx + q''dy + q'''dz + q''''dp$ avec des coefficients qui sont des fonctions de (x, y, z, p) .⁵¹ L'EDT devient

$$\begin{aligned} dz &= pdx + qdy \\ Zdy + \zeta da + \zeta' db + \zeta'' dc &= p(Xdy + \chi da + \chi' db + \chi'' dc) + qdy \\ (Z - pX - q) dy &= (-\zeta + p\chi) da + (-\zeta' + p\chi') db + (-\zeta'' + p\chi'') dc . \end{aligned}$$

Pour faire disparaître la différentielle dy , il suffit donc que l'on ait $Z - pX - q = 0$ et l'EDT devient alors

$$(\zeta - p\chi) da + (\zeta' - p\chi') db + (\zeta'' - p\chi'') dc = 0 .$$

Pour éliminer non seulement dy mais également y , il suffit, quitte à multiplier l'expression par une fonction $F(y) \neq 0$, d'annuler les dérivées des coefficients C par rapport à y . Mais $\partial_y(FC) = F\partial_y C + C\partial_y F = 0$ équivaut à $\frac{\partial_y F}{F} = -\frac{\partial_y C}{C}$.⁵² D'où la condition

$$\frac{\partial_y(\zeta - p\chi)}{\zeta - p\chi} = \frac{\partial_y(\zeta' - p\chi')}{\zeta' - p\chi'} = \frac{\partial_y(\zeta'' - p\chi'')}{\zeta'' - p\chi''} .$$

Mais il est facile de calculer $\partial_y(\zeta - p\chi) = \partial_y\zeta - p\partial_y\chi - \chi\partial_y p$. En effet, par définition $X = \partial_y x$ et $\chi = \partial_a x$ et donc (commutativité des dérivées secondes) $\partial_y\chi = \partial_a X$. De même, $Z = \partial_y z$ et $\zeta = \partial_a z$ et $\partial_y\zeta = \partial_a Z$. Donc, puisque l'on a choisi $Z - pX = q$ pour éliminer y

$$\begin{aligned} \partial_y(\zeta - p\chi) &= \partial_y\zeta - p\partial_y\chi - \chi\partial_y p = \partial_a Z - p\partial_a X - \chi\partial_y p \\ &= \partial_a(Z - pX) + X\partial_a p - \chi\partial_y p = \partial_a q + X\partial_a p - \chi\partial_y p . \end{aligned}$$

50. Dans cette section historique consacrée au texte de son mémoire, nous gardons les notations de Pfaff pour que le lecteur se fasse une idée du style de l'époque.

51. Les ', '' , etc. sont chez Pfaff une numérotation et pas des ordres de dérivation.

52. Nous utilisons la notation standard $\frac{\partial}{\partial y} = \partial_y$ pour simplifier. Pfaff utilisait quant à lui la notation d^y .

Mais $\partial_a q$ peut se calculer facilement

$$\begin{aligned} dq &= q' dx + q'' dy + q''' dz + q'''' dp \\ \partial_a q &= q' \chi + q''' \zeta + q'''' \pi \end{aligned}$$

car $\partial_a y = 0$ puisque y est indépendante de a . Par ailleurs $\partial_y p = P$ et $\partial_a p = \pi$ et donc en définitive

$$\begin{aligned} \partial_y (\zeta - p\chi) &= \partial_a q + X \partial_a p - \chi \partial_y p \\ &= q' \chi + q''' \zeta + q'''' \pi + X \pi - \chi P \\ &= (q' - P) \chi + q''' \zeta + (q'''' + X) \pi . \end{aligned}$$

Si l'on revient au quotient $\frac{\partial_y(\zeta - p\chi)}{\zeta - p\chi}$, on voit qu'on peut trivialement le rendre égal à q'''' si $X = -q''''$ et $P = q' + pq''''$ ce qui implique $Z = pX + q = q - pq''''$. Ces valeurs satisfont les conditions et on peut donc les choisir. Avec ces valeurs, l'EDT ne contient plus ni y ni dy et devient une EDT des *seules* trois variables (a, b, c) (et non plus de quatre variables comme l'équation initiale) de la forme $da + Bdb + Cdc = 0$, B et C étant des fonctions de (a, b, c) , équation que l'on savait intégrer à l'époque (Lagrange). On fait $c = \text{cste}$ et on intègre $da + Bdb = 0$ avec un multiplicateur (facteur intégrant) M faisant de $da + Bdb$ la différentielle exacte dN d'une fonction $N(a, b; c)$ de (a, b) paramétrée par c .

Dans les §§ successifs de son *Methodus*, Pfaff résout en dimension croissante, avec de longs calculs explicites de réduction à une forme normale, les deux problèmes

- (i) d'intégrer une EDO du premier ordre de $n = 2m - 1$ et $n = 2m$ variables avec m équations,
- (ii) d'intégrer complètement une EDP à $(n + 1)$ variables (z, x_i) en rajoutant les variables p_i et l'EDT $dz = \sum_i p_i dx_i$ permettant de faire (par exemple) de p_n une fonction des $(z, x_{i \leq n}, p_{i \leq n-1})$.

2.6.6. L'apport de Jacobi.

Toujours, dans le cas *pair*, Jacobi⁵³ utilisa dans un article de 1827 du *Journal de Crelle* [268]⁵⁴ ce qui deviendra connu sous le nom de *méthode d'Hamilton-Jacobi* et montra qu'en introduisant des "valeurs initiales" des variables on pouvait, comme l'explique Forsyth, prendre les intégrales du premier système subsidiaire

"in a form which leads immediately to the transformation of the equation."
([188], p. 82)

53. Avec Lie, Frobenius et Cartan, Jacobi est l'un des grands géomètres classiques qui sera le plus mobilisé dans cet ouvrage.

54. Le *Journal de Crelle* est le nom donné pendant toute cette période au *Journal für die reine und angewandte Mathematik* (Journal de mathématiques pures et appliquées) fondé à Berlin en 1826 par August Leopold Crelle (1780-1855).

Jacobi montre que l'intégration du premier système subsidiaire (P) de la section précédente est suffisante pour intégrer l'EDP. L'idée est de prendre z et les $2m - 1 = n - 1$ intégrales $u_\ell = a_\ell$ de (P) comme $2m$ nouvelles variables et d'exprimer les x_i et les p_i en fonction d'elles, ce qui donne des $x_i(z, u_\ell)$ et des $p_i(z, u_\ell)$ satisfaisant identiquement (P). Si on fixe alors z , par exemple en prenant $z = 0$, les $x_i(z, u_\ell)$ et les $p_i(z, u_\ell)$ deviennent des fonctions $\xi_i(u_\ell)$ et $\pi_i(u_\ell)$ satisfaisant $f(\xi_i, 0, \pi_i) = a$ et $u_\ell(\xi_i, 0, \pi_i) = a_\ell$. On en tire des fonctions, $\xi_i(a, a_\ell)$ et $\pi_i(a, a_\ell)$, avec des constantes a et a_ℓ qui redonnent les $x_i(z, u_\ell)$ et les $p_i(z, u_\ell)$ lorsque a est remplacée par f et a_ℓ par u_ℓ . En faisant les substitutions nécessaires et en utilisant les EDO de (P), on montre alors que l'EDT initiale se transforme en $\sum_{i=1}^{i=m} \pi_i d\xi_i$ (attention au $i = m$) qui est la *forme normale* de l'EDT initiale et dont l'intégrale est $\xi_i = c_i = \text{cste}$. En éliminant les p_i des $m + 1$ équations $f(x_i, z, p_i) = a$ et $\xi_i(a, u_\ell) = a_\ell$ avec $u_\ell = u_\ell(x_i, z, p_i)$, on obtient une équation en les z et x_i et m constantes arbitraires a_ℓ qui est l'intégrale complète de l'EDP initiale.

Dans les cas où $n = 2m + 1$ est *impair* ou lorsque $\Delta = 0$, les calculs sont plus compliqués mais aboutissent au même type de réduction. On voit qu'une donnée clé est la matrice antisymétrique \mathfrak{A} . À un extrême, lorsque \mathfrak{A} est inversible, on peut appliquer la méthode de réduction que nous venons d'explicitier. À l'autre extrême, lorsque \mathfrak{A} est identiquement nulle l'EDT est exacte et les solutions sont de dimension $n - 1$ (hypersurfaces $F(x_i) = a$).

2.7. Les années 1860-1880

Pendant plus de 30 ans, le problème de Pfaff avancera peu. Puis, il subira un important regain d'intérêt à partir du début des années 1860 et plusieurs mathématiciens le reprendront et l'approfondiront en ajoutant au point de vue traditionnel sur l'intégration d'équations différentielles des nouvelles méthodes géométriques (champs d'éléments de contact, etc.) et algébriques (algèbre linéaire, matrices antisymétriques, déterminants, mineurs, etc.). C'est à partir de là que le "retournement galoisien" qu'est la "géométrisation" du calcul différentiel s'effectue véritablement.

2.7.1. Grassmann.

Si $\omega = \sum_{i=1}^{i=n} A_i dx_i = Adx$ est une forme de Pfaff de n variables⁵⁵ avec un système intégral équivalent optimal donné par p équations $u_r = \text{cste}$ (p minimal)

55. Grassmann utilise la notation compacte Adx (innovante à l'époque) en introduisant des variables "extensives" $x = (x_1, \dots, x_n)$, des produits scalaires et des produits extérieurs. C'est le début de l'algèbre linéaire moderne comme calcul symbolique.

alors l'EDT $\omega = 0$ peut s'écrire

$$\omega = Adx = \sum_{r=1}^{r=p} U_r du_r = U du = 0. \quad 56$$

Cela signifie que les solutions $u_r = a_r = \text{cste}$ sont de codimension $c = p$ minimale et de dimension $d = n - c = n - p$ maximale d_ω . Les $du_r = 0$ constitue le plus petit nombre d'équations *exactes* permettant de satisfaire l'EDT $\omega = 0$. Grassmann donne des conditions nécessaires et suffisantes pour l'existence des u_r et montre que nécessairement $2p \leq n$ ($2p - 1 < n$) et donc que $d_\omega \geq \frac{n}{2}$ ce qui est un premier résultat fondamental (déjà évoqué plus haut) : toute équation de Pfaff possède des solutions de dimension au moins m si $n = 2m$ ou $n = 2m + 1$, m étant le minimum de la dimension maximale des solutions.

Grassmann utilise des notations formelles symboliques comme $Adx = Udu$ ou $A = U \frac{du}{dx}$ ($A_i = \sum_r U_r \frac{\partial u_r}{\partial x_i}$) qui généralisent les notations classiques depuis Leibniz. Il dérive des A sous la forme de matrices $\frac{dA}{dx} = \left(\frac{\partial A_i}{\partial x_j} \right) = \frac{dU}{dx} \frac{du}{dx} + U \frac{d^2u}{dx^2}$ où $\frac{dU}{dx} \frac{du}{dx}$ est le produit scalaire $\sum_{r=1}^{r=p} \frac{dU_r}{dx} \frac{du_r}{dx}$ et $U \frac{d^2u}{dx^2}$ le produit scalaire $\sum_{r=1}^{r=p} U_r \frac{d^2u_r}{dx^2}$, chaque terme étant une matrice. Surtout, il introduit les *crochets* comme $\left[A \left(\frac{dA}{dx} \right)^{p-1} \right]$, $\left[A \left(\frac{dA}{dx} \right)^p \right]$, ou $\left[\left(\frac{dA}{dx} \right)^{p+1} \right]$ en notant que si le premier est en général $\neq 0$, les deux autres sont nécessairement nuls. Il y a là une première introduction de produits et de dérivées extérieures, à savoir que si $\omega = Adx$ ne dépend en fait vraiment que de p variables, alors $d^{p+1}\omega = 0$ et $\omega \wedge d^p\omega = 0$ car ce sont des formes de degré $p + 1$ mais $\omega \wedge d^{p-1}\omega$ est en général $\neq 0$ car c'est une forme de degré p .

Le cas *générique* (qu'il appelle "inconditionné") correspond à $2p = n$ ($p = m = n - p = d_\omega$) et, dans ce cas, Grassmann construit l'équation subsidiaire permettant de réduire l'EDT d'une variable. En supposant qu'il existe une intégrale, on trouve une EDT "inconditionnée" de $2p - 2 = n - 2$ variables et on itère. Dans le cas non générique $2p < n$ où l'on peut avoir des solutions de dimension supérieure au minimum des dimensions maximales, les n équations subsidiaires ne comprennent que $2p$ équations indépendantes. Grassmann suppose que $n - 2p$ variables sont constantes et il obtient une transformation de Adx en une EDT Udu de $n - 1$ variables. Si $2p$ est non seulement $< n$ mais $< n - 1$, le système intégral est composé de p équations et, en itérant, on arrive à un système *générique* de $2p$ variables auquel on peut appliquer la méthode déjà disponible.

Nous ne développons pas ici les calculs dans la notation de Grassmann car nous y reviendrons plus bas avec Cartan et Goursat dans des notations plus modernes.

56. Si ω est exacte, $\omega = dF$, alors $\omega = Udu$ avec $p = 1$, $u = F$, $U = 1$. S'il existe un facteur intégrant ρ tel que $\rho\omega = dF$, alors $U = \frac{1}{\rho}$. Les solutions sont les hypersurfaces $F = a$.

2.7.2. *Natani.*

Leopold Natani (cf. [384]) développa aussi des méthodes de réduction du nombre d'éléments différentiels d'une équation de Pfaff, méthodes exposées de façon plus simple et plus compréhensible que chez Grassmann. Si l'on est dans le cas générique d'une équation de $n = 2m$ variables avec m équations dans le système intégral équivalent à l'EDT, une intégrale réduit l'EDT à une EDT de $n - 1$ variables ayant un système intégral équivalent de $m - 1$ équations. En itérant la procédure, on en arrive à une EDT de $m+1$ variables complètement intégrable possédant une intégrale équivalente, i.e. des solutions de dimension m qui sont celles de l'EDT initiale de dimension maximale $d_\omega = m$.

2.7.3. *Clebsch.*

Dans ses longs deux articles [123] de 1862/63, Clebsch développa aussi des méthodes de réduction plus complètes que celles de Grassmann. Il s'intéressa en particulier aux différentes formes normales des EDT de Pfaff. Dans le cas pair $n = 2m$ avec la forme normale (appelée plus tard "symplectique") $\omega = \sum_{i=1}^{i=m} p_i dx_i$, on voit apparaître dans les calculs le crochet (appelé plus tard "crochet de Poisson") de deux fonctions f et g des n variables $(x_i, p_i) : \{f, g\} = \sum_{i=1}^{i=m} \left(\frac{\partial f}{\partial x_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial x_i} \right)$. Ce crochet implique la dérivée extérieure $d\omega = \sum_{i=1}^{i=m} dp_i \wedge dx_i$ puisque $\{f, g\} = -d\omega(\nabla f, \nabla g)$.

Nous allons revenir sur ces crochets à propos de Darboux.

2.7.4. *Frobenius.*

Une étape conceptuelle essentielle fut accomplie par Frobenius dans son mémoire de 1877 "Über das Pfaff'sche Problem" [200] où il discute les formes normales des équations de Pfaff plus que leurs processus d'intégration explicite (cf. aussi le mémoire [201] de 1879). Comme l'explique Forsyth

"The interest lies chiefly in the purely algebraical association of the number of terms in the reduced form with the critical conditions (...) satisfied by the coefficients of the original expression." ([188], p. 272)

Son point de vue relève donc plus de l'algèbre que de l'analyse différentielle dans la mesure où il cherche de bonnes transformations linéaires d'invariants différentiels. En particulier, Frobenius introduisit le fameux "covariant bilinéaire" qui sera interprété à partir de la 2-forme ω' (i.e. $d\omega$) par Cartan. Et comme l'explique Forsyth,

"He (...) changes the Pfaffian and the differential covariant into homogeneous algebraic forms subject to linear transformations." ([188], p. 87)

Nous avons déjà défini dans la section 7.3.4 du chapitre 2 la "dérivée" ω' de ω comme 2-forme $d\omega = \sum_i dA_i \wedge dx_i$. Elle s'introduit chez Frobenius de la façon suivante, d'abord en tant qu'entité bilinéaire puis ensuite en tant qu'entité covariante. Dans ce formalisme, ce qui deviendra au XX^e siècle la dualité entre vecteurs tangents

et vecteurs cotangents et par suite entre champs de vecteurs et 1-formes, n'existe pas encore. Dans une EDT de Pfaff $\omega = \sum_i A_i dx_i$ les différentielles dx_i ont encore un statut ambigu dû à leur interprétation comme infinitésimales leibniziennes : d'une part ce sont des symboles différentiels munis d'un produit extérieur non commutatif et d'autre part ce sont des variations infinitésimales des coordonnées x_i d'un point de \mathbb{R}^n , autrement dit des déplacements infinitésimaux dx d'un point x muni du produit normal des grandeurs (même si ces grandeurs sont infinitésimales). Nous avons déjà commenté cette ambiguïté dans les sections 6.2.1 et 6.3 du chapitre 2.

Pour lever ces ambiguïtés et bien comprendre la définition de ω' tout en restant au plus près des intuitions et des textes de l'époque, rappelons deux notions. D'abord celle, introduite à la section 7.3 du chapitre 2, de *valeur* (au sens de Cartan) d'une forme différentielle $\omega = \sum_i A_i dx_i$. Si $x = (x_i)$ est un point de \mathbb{R}^n et si δx est un accroissement infinitésimal de x , la valeur $\omega_x(\delta x)$ est l'infinitésimale du premier ordre $\omega_x(\delta x) = \sum_i A_i(x) \delta x_i$.⁵⁷ Ensuite, celle d'élément de contact introduite à la section 5. Il s'agit de ce que Felix Klein appelait joliment, nous l'avons vu, des "écailles de poisson infiniment petites", autrement dit des couples (x, E) d'un point $x \in \mathbb{R}^n$ et d'un sous-espace de dimension h de $T_x \mathbb{R}^n$. Si M est une sous-variété de dimension h passant par x , alors $T_x M$ est un élément de contact. L'élément de contact (x, E) est dit *intégral* si ω (en fait ω_x) s'annule sur E . Si E est de dimension $h = 1$, il n'y a qu'une seule façon de progresser à partir de x le long de E , mais si en revanche E est de dimension $h \geq 2$, il y a plusieurs façons et l'on peut alors *comparer entre elles* les différentes façons. C'est ainsi que s'introduit la dérivée ω' de ω . Soient en effet δx et $\mathfrak{d}x$ deux déplacements infinitésimaux dans E à partir de x et considérons leur parallélogramme. Il y a deux façon de calculer la variation des valeurs de ω . Si l'on suit d'abord $\mathfrak{d}x$ et ensuite δx on arrive à une variation de ω

$$\begin{aligned} \omega_{x+\delta x}(\mathfrak{d}x) - \omega_x(\mathfrak{d}x) &= \sum_i A_i(x + \delta x) \mathfrak{d}x_i - \sum_i A_i(x) \mathfrak{d}x_i \\ &= \sum_i \left(A_i(x) + \sum_k \frac{\partial A_i}{\partial x_k} \delta x_k \right) \mathfrak{d}x_i - \sum_i A_i(x) \mathfrak{d}x_i \\ &= \sum_{i,k} \frac{\partial A_i}{\partial x_k} \delta x_k \mathfrak{d}x_i . \end{aligned}$$

Si en revanche l'on suit d'abord δx et ensuite $\mathfrak{d}x$ on arrive à une variation de ω

$$\omega_{x+\mathfrak{d}x}(\delta x) - \omega_x(\delta x) = \sum_{i,k} \frac{\partial A_i}{\partial x_k} \mathfrak{d}x_k \delta x_i .$$

57. Si δx est considéré comme un vecteur tangent "infinitésimal" $\delta x = \varepsilon X$ avec $X \in T_x \mathbb{R}^n$ on retrouve bien la définition actuelle puisque $\omega_x(\delta x) = \varepsilon \omega_x(X)$.

La différence entre les deux variations est (avec le changement d'indice nécessaire pour pouvoir mettre en facteur $\delta x_k \mathfrak{D}x_i$) :

$$\begin{aligned} \sum_{i,k} \frac{\partial A_i}{\partial x_k} \delta x_k \mathfrak{D}x_i - \sum_{i,k} \frac{\partial A_i}{\partial x_k} \mathfrak{D}x_k \delta x_i &= \sum_{i,k} \frac{\partial A_i}{\partial x_k} \delta x_k \mathfrak{D}x_i - \sum_{i,k} \frac{\partial A_k}{\partial x_i} \mathfrak{D}x_i \delta x_k \\ &= \sum_{i,k} \left(\frac{\partial A_i}{\partial x_k} - \frac{\partial A_k}{\partial x_i} \right) \delta x_k \mathfrak{D}x_i \\ &= \sum_{i,k} \mathfrak{a}_{ik} \delta x_k \mathfrak{D}x_i \end{aligned}$$

ce qui est la valeur $d\omega(\delta x, \mathfrak{D}x)$ de la 2-forme $d\omega$ sur les éléments infinitésimaux $\delta x, \mathfrak{D}x$.

La *covariance* de ω' a été montrée à la section 7.3.4 du chapitre 2.

Dans [218], Goursat reformule d'une façon variationnelle simple et intuitive ce "covariant bilinéaire" ω' de Frobenius. Les formes étant faites pour être intégrées et ω étant une 1-forme, les objets de base sont les intégrales curvilignes $I = \int_{\gamma} \omega$ le long de courbes $\gamma : a \rightarrow b$. Si les $x_i(t)$ sont les coordonnées le long de γ , $t \in [t_0, t_1]$, $x(t_0) = a$, $x(t_1) = b$, on a $I = \int_{t_0}^{t_1} \sum_i A_i(x_i(t)) x'_i(t) dt$. Déformons alors un peu γ en une famille γ_{α} ($\gamma_0 = \gamma$) de façon à ce que les déformations à x constant soient transverses à γ . Les $x_i(t)$ deviennent des fonctions de deux variables $x_i(t, \alpha)$, les t_0, t_1 deviennent des fonctions $t_0(\alpha), t_1(\alpha)$ et l'intégrale I devient une fonctionnelle $I(\alpha)$ dont on peut calculer la dérivée en $\alpha = 0$:

$$\begin{aligned} I'(\alpha) &= \int_{t_0}^{t_1} \sum_i \left(\sum_j \frac{\partial A_i}{\partial x_j} \frac{\partial x_j}{\partial \alpha} \right) \frac{\partial x_i}{\partial t} dt + \int_{t_0}^{t_1} \sum_i A_i \frac{\partial^2 x_i}{\partial t \partial \alpha} dt \\ &\quad + \frac{dt_1}{d\alpha} \left[\sum_i A_i \frac{\partial x_i}{\partial t} \right]_{t=t_1} - \frac{dt_0}{d\alpha} \left[\sum_i A_i \frac{\partial x_i}{\partial t} \right]_{t=t_0} . \end{aligned}$$

Si on écrit $\delta I = I'(\alpha) \delta \alpha$ et si l'on impose que t_0, t_1 soient fixes, alors on obtient pour variation de la fonctionnelle I

$$\delta I = \int_{\gamma} \omega'$$

où l'intégrale de la 2-forme ω' le long de γ signifie que l'on intègre seulement le long de la variable t avec des $dx_i = \frac{\partial x_i}{\partial t} dt$ et pas le long de la variable α avec $\delta x_i = \frac{\partial x_i}{\partial \alpha} \delta \alpha$. En termes contemporains on dirait que l'on considère la *contraction* (le produit intérieur, cf. la section 7 du chapitre 2) $i_{\delta \alpha} \omega'$ de la 2-forme ω' par les vecteurs tangents $\delta \alpha$,

1-forme définie par convention, rappelons-le, par $i_{\delta\alpha}\omega'(\mathfrak{D}x) = \omega'(\mathfrak{D}x, \delta\alpha)$ et que l'on intègre le long de γ .⁵⁸

Le covariant bilinéaire ω' permet de définir la *condition d'intégrabilité*

$$\omega \wedge \omega' = 0$$

sur laquelle nous reviendrons plus longuement à la section suivante.

L'un des aspects essentiels de la démarche de Frobenius a été d'analyser, avec les outils, nouveaux à l'époque, de l'algèbre linéaire, la structure algébrique d'une part de la matrice antisymétrique $n \times n$ \mathfrak{A} des coefficients \mathfrak{a}_{ik} de ω' et d'autre part de la matrice antisymétrique $(n+1) \times (n+1)$ $\mathfrak{B} = \begin{pmatrix} \mathfrak{A} & A \\ -A^T & 0 \end{pmatrix}$ (A étant le vecteur colonne des A_i). Cela lui permet de définir précisément un invariant entier fondamental qui s'appellera la "classe" de l'équation de Pfaff et qui est le nombre *minimal* p de variables dont dépend véritablement l'EDT. Comme nous l'avons vu avec Cartan à la section 7.3.5 du chapitre 2, la classe est le p minimal tel que la p -ème dérivée $\omega^{(p)}$ soit $\equiv 0$. Soit m le rang de \mathfrak{A} et m' le rang de \mathfrak{B} , m et m' étant nécessairement pairs à cause de l'antisymétrie. Si $m = 2r$, alors $m' = 2r$ ou $m' = 2r + 2 = 2(r + 1)$. Alors $p = \frac{m+m'}{2}$, ce qui est un entier bien défini puisque m et m' sont pairs. Si $m = m'$, $p = m$ est pair et la forme normale de ω est $\sum_{i=1}^{\frac{p}{2}} p_i dq_i$ (cas symplectique). Si $m = m' + 2$, $p = m + 1$ est impair et la forme normale de ω est $dz + \sum_{i=1}^{\frac{p}{2}-1} p_i dq_i$ (cas de contact).

Cette caractérisation algébrique des deux cas, symplectique et de contact, est vraiment remarquable. Elle fait apparaître clairement les deux structures géométriques universelles qui vont dominer les recherches.

2.7.5. Lie.

Dans les années 1874-1878, Sophus Lie aborde le problème dans une perspective différente, encore plus géométrique et encore moins analytique. Il fait partie de ceux qui ouvrent la voie aux approches modernes. Comme le note Forsyth,

"(Lie) worked out a different stand point (... and his) results constitute a distinct addition to the theory." ([188], p. 86-87)

Dans son mémoire de 1877 "Theorie des Pfaff'schen Problems" [334], publié la même année que celui de Frobenius, il utilise les éléments de contact et introduit ses fameuses "transformations tangentielles" appelées aussi *transformations de contact* ("Berührungstransformationen", cf. la section 5). Cela lui permet de systématiser la théorie existante. Si l'on traite de l'EDT de Pfaff $\omega = dz - \sum_{i=1}^{i=m} p_i dx_i = 0$ à $n = 2m + 1$ variables (z, x_i, p_i) dans \mathbb{R}^n , une transformation de contact est un changement de variables $\tilde{z}(z, x_i, p_i)$, $\tilde{x}_i(z, x_i, p_i)$, $\tilde{p}_i(z, x_i, p_i)$ qui laisse l'EDT invariante,

58. Si l'on définit $i_{\delta\alpha}\omega'(\mathfrak{D}x)$ par $i_{\delta\alpha}\omega'(\mathfrak{D}x) = \omega'(\delta\alpha, \mathfrak{D}x)$, on change de signe.

autrement dit qui satisfait $\tilde{\omega} = d\tilde{z} - \sum_{i=1}^{i=m} \tilde{p}_i d\tilde{x}_i = \rho\omega = 0$. Nous reviendrons plus bas sur leur caractérisation en termes de crochets.

2.7.6. Darboux.

Comme Lie et Frobenius, Gaston Darboux approfondit également la théorie de Pfaff, en particulier dans son mémoire de 1882 "Sur le problème de Pfaff" [133]. Il étudie lui aussi plutôt les formes équivalentes de l'équation de Pfaff que son système intégral et il trouve lui aussi, ainsi que le rappelle Cartan

"les formules fondamentales de la théorie des transformations de contact."
([90], p. 241)

Il travaille également sur le covariant bilinéaire $\sum_{i,k} \mathbf{a}_{ik} \delta x_k \delta x_i$ et la réduction aux formes normales.

Pour donner un exemple de la littérature de l'époque (années 1870) nous allons résumer son mémoire. Darboux insiste d'emblée sur l'importance de la *méthode* de Pfaff :

"La méthode (\dots) a été longtemps négligée : les belles découvertes de Jacobi et de Cauchy ont seules attiré l'attention des géomètres qui s'occupent de cette théorie. (\dots) Cependant la méthode de Pfaff, qui est, d'ailleurs, la généralisation de celle que l'on doit à Lagrange pour le cas de deux variables indépendantes, offre de sérieux avantages. Elle substitue à des calculs souvent compliqués l'emploi de certaines idéalités différentielles qui donnent la clef et la solution intuitive des difficultés qui se présentent dans les autres méthodes." ([133], p. 14)

On ne saurait être plus clair. Encore une fois, c'est vraiment comme avec Galois : le fait de comprendre méthodiquement les conditions d'intégrabilité au moyen de nouvelles structures *sui generis* revient déjà à presque effectuer l'intégration.

Darboux est on ne peut plus clair sur le fait qu'il ne traite pas de l'intégration explicite. Il renvoie à Lie pour les méthodes permettant de

"réduire au plus petit nombre possible les intégrations que l'on a à effectuer successivement avant de parvenir à la solution complète d'une équation aux dérivées partielles" ([133], p. 15)

et il ajoute

"Je ne me suis nullement occupé des intégrations qui sont nécessaires pour ramener une expression différentielle à une forme réduite." ([133], p. 15)

Ce qu'il vise est la *structure géométrique intrinsèque* (invariante par changements de coordonnées) de l'intégrabilité :

"Dans le travail qu'on va lire, je me suis proposé d'expliquer la solution du problème de Pfaff sans rien emprunter à la théorie des équations aux dérivées

partielles et je me suis surtout attaché à mettre en évidence des propriétés d'invariance qui jouent un rôle fondamental dans cette solution." ([133], p. 15)

Le retournement méthodologique faisant passer d'un *calcul d'intégrales* à des *invariants de structure* est spectaculaire. L'analogie avec Galois continue à s'imposer. De même que la structure des groupes de permutations et de leurs sous-groupes est indépendante de la résolution de telle ou telle équation algébrique au moyen de telle ou telle formule et qu'elle en mathématise les conditions de possibilité au moyen d'une théorie algébrique universelle qui, relativement au problème de la résolution des équations, fonctionne, répétons-le encore une fois, comme une "background structure" synthétique a priori, de même la géométrie de l'intégrabilité est indépendante de la résolution de telle ou telle EDP au moyen de telle ou telle formule et en fournit la "background structure" synthétique a priori qui en mathématise les conditions intrinsèques d'intégrabilité.

Darboux note les analogies de son mémoire avec les travaux contemporains d'une part de Frobenius [200] (invariants et covariant bilinéaire), mais en insistant encore une fois sur le fait que

"[son] exposition [est] plus affranchie du calcul" ([133], p. 15)

et d'autre part de Lie [334], mais en insistant sur le fait que sa méthode est tout à fait différente.

Dans la première section (écrite en 1876), Darboux étudie les formes réduites. Il définit à son tour le covariant bilinéaire $\omega' = d\omega$ d'une expression de Pfaff $\omega = A dx = \sum A_i dx_i$ de n variables qu'il note θ_d , le symbole d désignant la notation choisie pour les différentielles. Comme on le faisait à l'époque, nous l'avons vu, ω' est donnée par $\delta\theta_\delta - \mathfrak{d}\theta_\delta = \sum_{i,k} \mathfrak{a}_{ik} \mathfrak{d}x_i \delta x_k$ et Darboux insiste à son tour sur son invariance.

Dans la seconde section, Darboux étudie les relations entre les formes réduites et retrouve les groupes de Lie. Il introduit le système de Pfaff invariant donné par l'égalité de 2-formes $d\omega = \lambda\omega \wedge dt$ (en notations actuelles), système qu'il écrit de façon compacte sous la forme

$$\mathfrak{A}dx = \lambda A dt \tag{(1)}$$

(où \mathfrak{A} est la matrice des \mathfrak{a}_{ik}).⁵⁹ Ces systèmes associés permettent de réduire progressivement ω à une forme normale.

Dans la troisième section, Darboux étudie alors le déterminant et les mineurs de la matrice \mathfrak{A} . Si $\Delta = \det(\mathfrak{A}) \neq 0$ (ce qui exige n pair), le système (1) est déterminé

59. $d\omega = \lambda\omega \wedge dt$ s'écrit $\sum_{i,k} \mathfrak{a}_{ik} dx_i \wedge dx_k = \lambda \sum A_i dx_i \wedge dt$. Le coefficient de dx_i est $-\lambda A_i dt$ dans le second membre et $-\sum_k \mathfrak{a}_{ik} dx_k$ dans le premier membre, d'où (1).

et on peut le résoudre en faisant $dx = \lambda \mathfrak{A}^{-1}(A) dt$. Si $\mathfrak{A}^{-1} = (\alpha_{ik})$, alors

$$\frac{dx_i}{\sum_{k=1}^{k=n} \alpha_{ik} A_k} = \lambda dt, \quad i = 1, \dots, n$$

$$\frac{dx_i}{dt} = \lambda \left(\sum_{k=1}^{k=n} \alpha_{ik} A_k \right) = X_i(x_k)$$

et l'on se retrouve à intégrer un champ de vecteurs X_i (un système d'EDO). Si $\lambda = 0$, la solution est $dx_i = 0$, $i = 1, \dots, n$, les x_i doivent être constantes et les trajectoires sont des points fixes. Si $\lambda \neq 0$, les trajectoires sont de dimension 1 et représentables comme intersections de $(n - 1)$ hypersurfaces $F_\ell = c_\ell$ (c_ℓ constantes) qui fournissent $(n - 1)$ intégrales premières de (1). On peut prendre ces dernières comme $(n - 1)$ nouvelles variables indépendantes y_ℓ , leur adjoindre une autre variable indépendante y_n et faire le changement de variables. Le nouveau système de Pfaff (1) associé à $\tilde{\omega} = \sum_i B_i dy_i$ devient $\mathfrak{b}_{n\ell} dy_n = \lambda B_\ell dt$ et $\lambda B_n dt = 0$. Donc $B_n = 0$, $\frac{\partial B_n}{\partial y_i} = 0$ et

$$\mathfrak{b}_{n\ell} dy_n = \lambda B_\ell dt = -\frac{\partial B_\ell}{\partial y_n} dy_n,$$

c'est-à-dire

$$(\partial B_\ell / \partial y_n) / B_\ell = \frac{\partial \text{Log}(B_\ell)}{\partial y_n} = -\lambda \frac{dt}{dy_n}.$$

Il y a donc *séparation des variables* : $y_\ell = C y_\ell^0$ où la constante C dépend de y_n mais pas de y_ℓ^0 et où complémentirement y_ℓ^0 ne dépend pas de y_n . En prenant C comme nouvelle variable y_n , on obtient $\tilde{\omega} = y_n \left(\sum_{\ell=1}^{\ell=n-1} y_\ell^0 dy_\ell \right)$. Cela revient à dire qu'on a trouvé un "facteur intégrant" non pas pour faire de ω une différentielle totale exacte mais pour faire de l'EDT $\omega = 0$ une EDT de $(n - 1)$ variables $\sum_{\ell=1}^{\ell=n-1} y_\ell^0 dy_\ell = 0$.

Dans le cas où n est impair, nécessairement $\Delta = 0$ et l'on considère le système de Pfaff (1) avec $\lambda = 0$ qui admet une solution dx non triviale. On peut appliquer des techniques analogues à celles du cas précédent si tous les mineurs d'ordre $(n - 1)$ de \mathfrak{A} sont non nuls car alors le système (1) comprend $n - 1$ équations indépendantes. Suivant les cas, on aboutit à une EDT de la forme $dy_n + \sum_{\ell=1}^{\ell=n-1} y_\ell^0 dy_\ell = 0$ ou $\sum_{\ell=1}^{\ell=n-1} y_\ell^0 dy_\ell = 0$.

Dans les quatrième et cinquième sections, Darboux traite les cas où le système de Pfaff (1) est *indéterminé* parce que tous les mineurs d'ordre $(n - 1)$ de \mathfrak{A} sont nuls (ce qui est automatique, à cause de l'antisymétrie, lorsque n est pair et que $\Delta = 0$). En faisant $\lambda = 0$, en réduisant (1) à p équations indépendantes et en ajoutant $(n - p - 1)$ équations supplémentaires $df_r = 0$ de façon à obtenir un nouveau système (1) déterminé, puis en lui appliquant les méthodes précédentes, on se ramène aux formes normales symplectiques $\omega = \sum_{k=1}^{k=m} p_k dy_k$ ($2m \leq n$) ou de contact $\omega = dz - \sum_{k=1}^{k=m} p_k dy_k$ ($2m + 1 \leq n$). Génériquement (i.e. si les coefficients

de ω ne satisfont pas de relations spéciales) $2m = n$ dans le premier cas et $2m+1 = n$ dans le second.

Dans la sixième section, Darboux précise un peu les choses. Dans le cas de contact, le système de Pfaff (1) devient $dy_k = 0$, $dp_k = -\lambda p_k dt$, $0 = \lambda dt$ et donc $dy_k = 0$, $dp_k = 0$ donnent $2m$ intégrales premières $y_k = \text{cste}$, $p_k = \text{cste}$ pour $2m+1$ variables et donc des courbes intégrales. Dans le cas symplectique, (1) devient $dy_k = 0$, $dp_k = \lambda p_k dt$ et λ peut être $\neq 0$. On obtient $2m-1$ intégrales premières dont m (intégrant les $dy_k = 0$) sont de la forme $y_k = \text{cste}$ et $(m-1)$ (intégrant les $\frac{dp_k}{p_k} = \lambda dt$) sont de la forme $p_k = C p_1$, $k = 2, \dots, m$.

Enfin dans la septième section Darboux aborde la question technique de trouver

“les intégrations qui sont nécessaires pour ramener une expression différentielle donnée à sa forme canonique.” ([133], p. 30)

Il cite Lie et Mayer mais reste assez général, déclarant

“dans ce travail, je ne m’occuperai que des propriétés d’invariance relatives à une forme différentielle.” ([133], p. 30)

Dans la seconde partie de son mémoire, Darboux introduit tout un ensemble de crochets. Pour le cas symplectique $n = 2m$ pair, $\omega = \sum_{k=1}^{k=m} p_k dx_k$, on a

$$\mathfrak{A} = \begin{array}{c|cc} & x_1 \cdots x_m & p_1 \cdots p_m \\ \hline x_1 & & \\ \vdots & 0 & -\delta_{ij} \\ x_m & & \\ \hline p_1 & \delta_{ij} & 0 \\ \vdots & & \\ p_m & & \end{array}$$

et il définit l’invariant qu’il note

$$\left\{ \begin{array}{l} \omega \\ df \end{array} \right\} = \det \left(\begin{array}{cc} \mathfrak{A} & A \\ \frac{\partial f}{\partial x} & 0 \end{array} \right) = \sum_{k=1}^{k=m} p_k \frac{\partial f}{\partial p_k}$$

le crochet (f) introduit par Clebsch étant $\frac{1}{\Delta} \left\{ \begin{array}{l} \omega \\ df \end{array} \right\} = \left\{ \begin{array}{l} \omega \\ df \end{array} \right\}$ car dans ce cas $\Delta = 1$. Cet invariant $\left\{ \begin{array}{l} \omega \\ df \end{array} \right\}$ n’est rien d’autre que $\omega(\nabla f)$.

Il introduit aussi un second invariant

$$\left\{ \begin{array}{l} df \\ dg \end{array} \right\} = \det \left(\begin{array}{cc} \mathfrak{A} & \frac{\partial f}{\partial x} \\ \frac{\partial g}{\partial x} & 0 \end{array} \right) = \sum_{k=1}^{k=m} \left(\frac{\partial f}{\partial x_k} \frac{\partial g}{\partial p_k} - \frac{\partial f}{\partial p_k} \frac{\partial g}{\partial x_k} \right)$$

qui n’est rien d’autre que ce qu’on appelle maintenant le crochet de Poisson $\{f, g\}$. Comme $d\omega = \sum_{k=1}^{k=m} dp_k \wedge dx_k$, on a

$$\{f, g\} = -d\omega(\nabla f, \nabla g) .$$

Le crochet de Clebsch était, comme nous l'avons vu plus haut,

$$(f, g) = -\frac{1}{\Delta} \det \begin{pmatrix} \mathfrak{A} & \frac{\partial f}{\partial x} \\ \frac{\partial g}{\partial x} & 0 \end{pmatrix} = -\left\{ \frac{df}{dg} \right\} = -\{f, g\} = d\omega(\nabla f, \nabla g) .$$

Dans le cas de contact $n = 2m + 1$ impair, Darboux considère le crochet qu'il note

$$[f, g] = \sum_{k=1}^{k=m} \left(\frac{\partial f}{\partial p_k} \left(\frac{\partial g}{\partial x_k} + p_k \frac{\partial g}{\partial z} \right) - \frac{\partial g}{\partial p_k} \left(\frac{\partial f}{\partial x_k} + p_k \frac{\partial f}{\partial z} \right) \right) .$$

Évidemment, si f et g sont indépendantes de z alors $[f, g] = (f, g)$ et par ailleurs $[f, z] = \sum_{k=1}^{k=m} p_k \frac{\partial f}{\partial p_k} = (f) = -\{f, z\} = \{z, f\}$.

On vérifie (c'est fastidieux, mais c'est un bon exercice scolaire) que ces crochets satisfont les relations de Mayer et Clebsch

$$[f, [g, h]] + [g, [h, f]] + [h, [f, g]] = \frac{\partial f}{\partial z} [g, h] + \frac{\partial g}{\partial z} [h, f] + \frac{\partial h}{\partial z} [f, g] .$$

Si les fonctions f, g, h ne dépendent pas de z , il s'agit de la relation de Jacobi caractéristique des algèbres de Lie. Si seulement f et g sont indépendantes de z et si $h = z$, alors on obtient une relation due à Clebsch

$$\begin{aligned} [f, [g, z]] + [g, [z, f]] + [z, [f, g]] &= \frac{\partial f}{\partial z} [g, z] + \frac{\partial g}{\partial z} [z, f] + \frac{\partial h}{\partial z} [f, g] \\ (f, (g)) - (g, (f)) - ((f, g)) &= (f, g) , \text{ i.e.} \\ (f, (g)) - (g, (f)) &= (f, g) + ((f, g)) . \end{aligned}$$

On a trivialement $(x_k) = 0$, $(p_k) = p_k$, $(p_j, p_k) = 0$, $(x_j, x_k) = 0$, $(p_j, x_k) = 0$ si $j \neq k$ et $(p_k, x_k) = 1$.

À partir de là, Darboux expose les transformations de contact de Lie et démontre le théorème de Lie à leur sujet. Nous y reviendrons plus bas à la section 3.3.4 dans la synthèse de Cartan.

On voit ainsi comment s'est mis en place ce mixte de géométrie des variétés et d'algèbre linéaire dans les espaces tangents et cotangents qui a permis la géométrisation du calcul différentiel et de l'intégrabilité des EDT et des EDP. La performance est remarquable, car cette géométrie différentielle n'existait pas et, contrairement au calcul différentiel d'un côté et à la géométrie de l'autre côté, *n'avait pas de précurseurs*. Comme le soulignent Chern et Chevalley [114],

"The difficulty of conceiving the proper concepts at the early stage of development can hardly be overestimated." ([114], p. 221)

3. La systématisation de Cartan et Goursat

3.1. Classe et réduction

Après ce survol de la période héroïque des premiers grands travaux sur le problème de Pfaff, nous allons résumer la synthèse qu'en a fait Élie Cartan dans ses grands articles de 1899 [90] et 1901 [91] avec les compléments ajoutés par Édouard Goursat dans son traité de 1922 [218]. La géométrisation abstraite y est omniprésente. Comme le notent Chern et Chevalley [114],

“the idea of studying the abstract structure of mathematical objects which hides itself beneath the analytical clothing under which they appear at first was also the mainspring of Cartan's theory of differential equations. [He] was able to see the geometrical content of very complicated calculations.” ([114], p. 220)

Le premier problème général est, nous l'avons vu, celui de la réduction à une forme normale de l'EDT $\omega = \sum_{i=1}^{i=n} A_i dx_i$. Il dépend de la *classe* p de l'EDT qui est le nombre minimal de variables dont dépend vraiment ω , ce qui est aussi le p minimal tel que la $(p+1)$ -forme $\omega^{(p)}$ soit $\equiv 0$. Dire que ω est de classe 0, c'est dire que $\omega \equiv 0$. Dire que ω est de classe 1, c'est dire que $\omega' = d\omega \equiv 0$, i.e. que ω est fermée et donc exacte ($\omega = df$) sur \mathbb{R}^n car \mathbb{R}^n est simplement connexe. Si $\omega = df \neq 0$, une des dérivées partielles de f (par exemple $\frac{\partial f}{\partial x_1}$) est $\neq 0$ et en prenant $y_1 = f$, $y_i = x_i$, $i > 1$, comme nouvelles variables⁶⁰, on réduit ω à la forme normale $\tilde{\omega} = dy_1 = df$. Dire que ω est de classe 2, c'est dire que la 3-forme $\omega^{(2)} = \omega\omega' \equiv 0$ avec

$$\omega^{(2)} = \left(\sum_i A_i dx_i \right) \left(\sum_{j < k=1}^{j,k=n} \mathbf{a}_{kj} dx_j dx_k \right) .$$

Pour 3 variables, être de classe 2 signifie donc que

$$A_1 \mathbf{a}_{32} + A_2 \mathbf{a}_{13} + A_3 \mathbf{a}_{21} = 0 .$$

À l'autre extrême, dire que ω est de classe maximale n (i.e. $\omega^{(n)} \equiv 0$), c'est dire que l'on ne peut réduire le nombre n de variables par aucun changement de variables. Comme nous l'avons vu à la section 7 du chapitre 2, si n est pair, $n = 2m$, on a le cas symplectique où la $2m = n$ -forme $(\omega')^m = m! \omega^{(n-1)}$ est non nulle et proportionnelle à la forme volume. Si au contraire n est impair, $n = 2m + 1$, on a le cas de contact où la $2m + 1 = n$ -forme $\omega (\omega')^m = m! \omega^{(2m)}$ est non nulle et proportionnelle à la forme volume.

Supposons que la classe p soit $< n$ et qu'il existe donc des changements de variables (y_i) , $i = 1, \dots, n$, permettant d'abaisser le nombre de variables. Le problème

60. Il s'agit d'un bon changement de variable car $\frac{D(y_1, x)}{D(x_i)} = \frac{\partial f}{\partial x_i} \neq 0$.

est de passer du "il existe un changement de variables" à "tel changement de variables est opératoire". Une technique de base pour trouver de bons changements de variables est de chercher des fonctions f satisfaisant $\omega^{(p-1)}df = 0$ (il s'agit d'une $(p+1)$ -forme⁶¹, $p+1 \leq n$) ou $\omega^{(p-2)}df = 0$ (il s'agit d'une p -forme). Cette équation induit un système complet d'EDP linéaires du premier ordre Σ_1 que doit satisfaire f et qui s'appelle le *système adjoint*. En prenant certaines de ses intégrales pour nouveaux y_i on peut abaisser le nombre de variables.

Il s'agit là d'une méthode basique, pédagogiquement explicitée dans Goursat [217] et [218]. Supposons que, τ étant une t -forme sur \mathbb{R}^n , il existe des variétés intégrales M^{n-1} de dimension $(n-1)$ (de codimension 1) de $\tau = 0$ ($t \leq n-1$), ce qui signifie que l'intégrale de τ sur toutes les sous-variétés N^t de dimension t de M sont nulles. Soit $f(x_i) = c$ l'équation d'une de ces intégrales maximales, supposons $\frac{\partial f}{\partial x_1} \neq 0$ et prenons $(y_1 = f, x_{i \geq 2})$ comme nouvelles variables (c'est possible car le jacobien est $\frac{\partial f}{\partial x_1} \neq 0$). Dans ces nouvelles coordonnées, τ devient $\tilde{\tau}$. Mais, par définition, $\int_{\tilde{M}} \tilde{\tau} = 0$ sur les \tilde{M} d'équation $y_1 = c$ satisfaisant donc $dy_1 = 0$. Mais cette dernière condition signifie que la différentielle dy_1 doit être contenue dans tous les termes de $\tilde{\tau}$, condition elle-même équivalente à $\tilde{\tau}dy_1 \equiv 0$.⁶² En revenant aux variables (x_i) , on trouve donc la condition (invariante) $\tau df \equiv 0$ sur la $(t+1)$ -forme τdf .

Dans le cas où $t = n-1$ (i.e. τ s'annule sur M), τdf est d'ordre n et est donc égale à la forme volume multipliée par un coefficient dont l'annulation donne une EDP en f . Écrivons $\tau = \sum_i T_i(x_i) d\hat{x}_i$ où $d\hat{x}_i$ dénote le produit $dx_1 \dots dx_n$ auquel on a ôté dx_i . On a $\tau df = \sum_i T_i \frac{\partial f}{\partial x_i} dx_i d\hat{x}_i$ et $dx_i d\hat{x}_i = \pm dx_1 \dots dx_n$ suivant les cas. Si n est impair, l'EDP est $\sum_i T_i \frac{\partial f}{\partial x_i} = 0$, et, si n est pair, l'EDP est $\sum_i (-1)^i T_i \frac{\partial f}{\partial x_i} = 0$.

Si l'on prend pour nouvelles coordonnées x_1, \dots, x_{n-1} ($n-1$) intégrales distinctes de $\tau df \equiv 0$ et une x_n arbitraire, alors

$$\begin{aligned} \tau &= T dx_1 \dots dx_{n-1} \\ \tau' &= (-1)^{n-1} \frac{\partial T}{\partial x_n} dx_1 \dots dx_n . \end{aligned}$$

61. Car $((p-1)+1)+1 = p+1$.

62. Il s'agit d'un lemme facile (cf. Goursat [218], p. 97). Si τ est une k -forme, $k \geq 2$, et si ω est une 1-forme, $\tau \equiv \tau_1 \omega$ (τ_1 une $(k-1)$ -forme) équivaut à $\tau \omega \equiv 0$. Evidemment si $\tau \equiv \tau_1 \omega$, $\tau \omega \equiv \tau_1 \omega^2 \equiv 0$. Réciproquement, soit $\tau \omega \equiv 0$ et supposons que ω contienne x_n et que l'on puisse résoudre dx_n relativement à ω en écrivant $dx_n \equiv \sum_{i=1}^{i=n-1} \alpha_i dx_i + \alpha \omega$ (ce sera toujours le cas par rapport à un x_i). On peut alors écrire $\tau \equiv \tau_2 + \tau_1 \omega$, τ_1 et τ_2 ne contenant pas x_n . Donc $\tau \omega \equiv 0 \equiv \tau_2 \omega + \tau_1 \omega^2 \equiv \tau_2 \omega$. Mais comme τ_2 ne contient pas x_n , $\tau_2 \omega \equiv 0$ implique $\tau_2 \equiv 0$ et donc $\tau \equiv \tau_1 \omega$.

Si $\frac{\partial T}{\partial x_n} \neq 0$, on prend T comme variable x_n et alors τ s'écrit sous la forme réduite $\tau = x_n dx_1 \dots dx_{n-1}$. Si $\frac{\partial T}{\partial x_n} = 0$ (i.e. T est indépendant de x_n), alors $\tau' = 0$ et $\tau = T(x_1 \dots x_{n-1}) dx_1 \dots dx_{n-1}$ est en fait une $(n-1)$ -forme de $(n-1)$ variables.

Si τ est un produit extérieur $\bigwedge_{\ell=1}^{\ell=t} \tau_\ell$ de ℓ EDT de Pfaff $\tau_\ell = \sum_i T_{\ell i}(x_j) dx_i$, alors la condition $\tau df \equiv 0$ s'écrit $df = \sum_\ell \lambda_\ell(x_i) \tau_\ell$.

Si l'on cherche des intégrales M^{n-s} de codimension s (de dimension $n-s$), de $\tau = 0$ ($t \leq n-s$) et si $f_k = c_k$, $k = 1, \dots, s$, sont des équations de M , on peut choisir les f_k comme s nouvelles coordonnées y_k et le même raisonnement conduit à la condition $\tau df_1 \dots df_s \equiv 0$ sur la $(t+s)$ -forme $\tau df_1 \dots df_s$ ($t+s \leq n$). On obtient ainsi un système d'EDP.

On peut en particulier considérer le cas des formes de Pfaff $\omega = \sum_i A_i(x) dx_i$, i.e. le cas $t = 1$ ($s+1 \leq n$), où, à tous les points x satisfaisant $f_k = c_k$, les δx tangents satisfont $df_k(\delta x) = 0$. Cela signifie que $\omega = \sum_k \mu_k(x) df_k$, ce qui implique le système de s équations $A_i(x) = \sum_k \mu_k(x) \frac{\partial f_k(x)}{\partial x_i}$, système qui pourra être résolu si les déterminants de tous les mineurs d'ordre $(s+1)$ de la matrice $n \times (s+1)$ $\begin{pmatrix} A_i \\ \frac{\partial f_k}{\partial x_i} \end{pmatrix}$ sont nuls. Si l'on suppose que l'on peut résoudre (au moyen du théorème des fonctions implicites) les équations $f_k = c_k$ sous la forme $x_k = \xi_k(x_\ell)$, $k = 1, \dots, s$ et $\ell = s+1, \dots, n$, alors, dans les variables x_ℓ , l'équation $\omega = 0$ équivaut aux $(n-s)$ EDP du premier ordre $A_\ell + \sum_k A_k \frac{\partial \xi_k}{\partial x_\ell}$. Goursat [218] note que "ce système est d'une forme très particulière" parce que dans chaque EDP on ne dérive que par rapport à une seule variable x_ℓ . Si $s = n-1$, autrement dit si l'on considère les courbes intégrales, le système se réduit à la seule équation $A_n + \sum_{k=1}^{k=n-1} A_k \frac{\partial \xi_k}{\partial x_n}$ et l'on peut choisir arbitrairement $(n-2)$ des fonctions $\xi_k(x_n)$.

En fait, on étend à des τ d'ordre t quelconque le concept de facteur intégrant en cherchant s'il existe des fonctions ρ telles que $\rho\tau = d\sigma$ soit une forme exacte. Une condition nécessaire est que $d(\rho\tau) = 0$ et comme $d(\rho\tau) = \rho d\tau + d\rho \wedge \tau$ on obtient un système d'EDP que doit satisfaire la fonction ρ .

C'est cette méthodologie que Cartan applique judicieusement à des dérivées $\omega^{(k)}$ bien choisies des formes de Pfaff.

Dans le cas symplectique $n = 2m$, $\omega = \sum_{i=1}^{i=m} p_i dx_i$, $\omega^{(n-1)}$ est la n -forme volume $dV = \bigwedge_{i=1}^{i=m} dp_i \wedge dx_i$, et par conséquent $\omega^{(n-1)} df = 0$ est toujours vérifié puisqu'il s'agit d'une $(n+1)$ -forme. Le premier système adjoint Σ_1 est donc vide. Pour le second système adjoint Σ_2 , comme

$$\omega^{(n-2)} = \sum_i (dp_1 \wedge dx_1) \wedge \dots \wedge p_i dx_i \wedge \dots \wedge (dp_n \wedge dx_n) ,$$

on obtient en revanche

$$\omega^{(n-2)}df = - \left(\sum_i p_i \frac{\partial f}{\partial p_i} \right) dV$$

et donc Σ_2 se réduit à l'EDP $\sum_i p_i \frac{\partial f}{\partial p_i} = 0$ qui possède les $(2m-1)$ intégrales x_i et $\frac{p_j}{p_1}$ ($j \neq 1$).⁶³ Pour $\omega^{(n-3)}dfdg$ on trouve $(f, g) dV$, (f, g) étant le crochet de Clebsch (l'opposé du crochet de Poisson).

Dans le cas de contact $n = 2m + 1$,

$$\omega = dz - \sum_{i=1}^{i=m} p_i dx_i ,$$

$$\omega' = \sum_i dx_i \wedge dp_i ,$$

$$\omega^{(n-2)} = \omega^{(2m-1)} = \bigwedge_{i=1}^{i=m} dx_i \wedge dp_i ,$$

$$\omega^{(n-1)} = \omega^{(2m)} = dz \bigwedge_{i=1}^{i=m} dx_i \wedge dp_i = dV .$$

Comme $\omega^{(2m)}$ est d'ordre maximal $2m + 1$, $\omega^{(n-1)}df = 0$ est toujours satisfait et le système adjoint Σ_1 est vide. En revanche $\omega^{(n-2)}df = \omega^{(2m-1)}df$ donne pour second système adjoint Σ_2 la simple EDP $\frac{\partial f}{\partial z} = 0$. Et pour $\omega^{(n-3)}dfdg = \omega^{(2m-2)}dfdg$ on trouve $-\{f, g\} dV$, $\{f, g\}$ étant le crochet (cf. plus haut Darboux 2.7.6)

$$\{f, g\} = \sum_{i=1}^{i=m} \left(\frac{\partial f}{\partial p_i} \left(\frac{\partial g}{\partial x_i} + p_i \frac{\partial g}{\partial z} \right) - \frac{\partial g}{\partial p_i} \left(\frac{\partial f}{\partial x_i} + p_i \frac{\partial f}{\partial z} \right) \right) .$$

3.2. Un exemple

Un bon exemple valant mieux que de longs discours, commençons par un des beaux exemples donnés par Cartan. Soit

$$\omega = x_1 x_3 dx_2 + x_1 x_2 dx_3 + (x_1 + x_3 x_5) dx_4 + x_3 x_4 dx_5 = 0$$

l'EDT à analyser. Elle est à 5 variables et 4 éléments différentiels. Calculons en détail (à titre d'exercice pédagogique, même si cela est un peu fastidieux) ses dérivées successives $\omega' = \omega^{(1)}$, etc.⁶⁴ Le calcul est un peu long mais immédiat, il suffit de

63. En effet, $\sum_i p_i \frac{\partial \left(\frac{p_j}{p_1} \right)}{\partial p_i} = p_i \left(-\frac{p_j}{(p_1)^2} \right) + p_i \frac{1}{p_1} = 0$.

64. Nous gardons les notations de Cartan où les produits extérieurs de différentielles sont notés comme des produits normaux sans le symbole \wedge (nous n'utilisons \wedge que lorsque c'est nécessaire pour la lisibilité des formules).

l'écrire. Pour $\omega^{(1)}$ on trouve

$$\begin{aligned}\omega^{(1)} &= d(x_1x_3) dx_2 + d(x_1x_2) dx_3 + d(x_1 + x_3x_5) dx_4 + d(x_3x_4) dx_5 \\ &= (x_1dx_3 + x_3dx_1) dx_2 + (x_1dx_2 + x_2dx_1) dx_3 \\ &\quad + (dx_1 + x_3dx_5 + x_5dx_3) dx_4 + (x_3dx_4 + x_4dx_3) dx_5 \\ &= x_3dx_1dx_2 + x_2dx_1dx_3 + dx_1dx_4 + x_5dx_3dx_4 + x_4dx_3dx_5\end{aligned}$$

car

$$\begin{cases} x_1dx_3dx_2 + x_1dx_2dx_3 = 0 \\ x_3dx_5dx_4 + x_3dx_4dx_5 = 0 \end{cases}$$

Pour $\omega^{(2)} = \omega\omega^{(1)}$ on trouve

$$\begin{aligned}\omega^{(2)} &= (x_1x_3dx_2 + x_1x_2dx_3 + (x_1 + x_3x_5) dx_4 + x_3x_4dx_5) \wedge \\ &\quad (x_3dx_1dx_2 + x_2dx_1dx_3 + dx_1dx_4 + x_5dx_3dx_4 + x_4dx_3dx_5) \\ &= x_1x_3x_2dx_2dx_1dx_3 + x_1x_3dx_2dx_1dx_4 + x_1x_3x_5dx_2dx_3dx_4 \\ &\quad + x_1x_3x_4dx_2dx_3dx_5 + x_1x_2x_3dx_3dx_1dx_2 + x_1x_2dx_3dx_1dx_4 \\ &\quad + (x_1 + x_3x_5) x_3dx_4dx_1dx_2 + (x_1 + x_3x_5) x_2dx_4dx_1dx_3 \\ &\quad + (x_1 + x_3x_5) x_4dx_4dx_3dx_5 + x_3x_4x_3dx_5dx_1dx_2 \\ &\quad + x_3x_4x_2dx_5dx_1dx_3 + x_3x_4dx_5dx_1dx_4 + x_3x_4x_5dx_5dx_3dx_4 \\ &= x_1x_3x_5dx_2dx_3dx_4 + x_1x_3x_4dx_2dx_3dx_5 \\ &\quad + x_3^2x_5dx_4dx_1dx_2 + x_3x_5x_2dx_4dx_1dx_3 + x_1x_4dx_4dx_3dx_5 \\ &\quad + x_3^2x_4dx_5dx_1dx_2 + x_3x_4x_2dx_5dx_1dx_3 + x_3x_4dx_5dx_1dx_4.\end{aligned}$$

car

$$\begin{cases} x_1x_3x_2dx_2dx_1dx_3 + x_1x_2x_3dx_3dx_1dx_2 = 0 \\ x_1x_3dx_2dx_1dx_4 + (x_1 + x_3x_5) x_3dx_4dx_1dx_2 = x_3x_5x_3dx_4dx_1dx_2 \\ x_1x_2dx_3dx_1dx_4 + (x_1 + x_3x_5) x_2dx_4dx_1dx_3 = x_3x_5x_2dx_4dx_1dx_3 \\ (x_1 + x_3x_5) x_4dx_4dx_3dx_5 + x_3x_4x_5dx_5dx_3dx_4 = x_1x_4dx_4dx_3dx_5 \end{cases}$$

Pour $\omega^{(3)} = \frac{1}{2} (\omega^{(1)})^2$ on trouve

$$\begin{aligned}\omega^{(3)} &= \frac{1}{2} (x_3dx_1dx_2 + x_2dx_1dx_3 + dx_1dx_4 + x_5dx_3dx_4 + x_4dx_3dx_5) \wedge \\ &\quad (x_3dx_1dx_2 + x_2dx_1dx_3 + dx_1dx_4 + x_5dx_3dx_4 + x_4dx_3dx_5) \\ &= \frac{1}{2} \left(\begin{aligned} &x_3x_5dx_1dx_2dx_3dx_4 + x_3x_4dx_1dx_2dx_3dx_5 + x_4dx_1dx_4dx_3dx_5 \\ &+ x_5x_3dx_3dx_4dx_1dx_2 + x_4x_3dx_3dx_5dx_1dx_2 + x_4dx_3dx_5dx_1dx_4 \end{aligned} \right) \\ &= x_3x_5dx_1dx_2dx_3dx_4 + x_3x_4dx_1dx_2dx_3dx_5 - x_4dx_1dx_3dx_4dx_5\end{aligned}$$

car $dx_3dx_4dx_1dx_2 = dx_1dx_2dx_3dx_4$, etc. puisqu'il faut un nombre pair (4) de permutations pour les identifier. Et enfin, pour $\omega^{(4)} = \omega\omega^{(3)}$ on trouve

$$\begin{aligned}\omega^{(4)} &= (x_1x_3dx_2 + x_1x_2dx_3 + (x_1 + x_3x_5)dx_4 + x_3x_4dx_5) \wedge \\ &\quad x_3x_5dx_1dx_2dx_3dx_4 + x_3x_4dx_1dx_2dx_3dx_5 - x_4dx_1dx_4dx_3dx_5 \\ &= -x_1x_3x_4dx_2dx_1dx_4dx_3dx_5 + (x_1 + x_3x_5)x_3x_4dx_4dx_1dx_2dx_3dx_5 \\ &\quad + x_3x_4x_3x_5dx_5dx_1dx_2dx_3dx_4 \\ &\equiv 0\end{aligned}$$

L'EDT $\omega = 0$ est donc de classe $p = 4$.

On peut donc réduire $\omega = 0$ à une EDT à 4 variables de forme normale symplectique $\tilde{\omega} = p_1dq_1 + p_2dq_2$. L'analyse géométrique est donc *théoriquement* effectuée. Mais pour résoudre l'EDT *concrètement*, encore faut il trouver de bons changements de variables. Pour ce faire, on résout des "systèmes complets adjoints" Σ_1 et Σ_2 . Cela signifie qu'on cherche des fonctions f satisfaisant $\omega^{(p-1)}df = 0$ ou $\omega^{(p-2)}df = 0$ et, là, le calcul n'est plus automatique et immédiat car il faut résoudre des EDP. Comme $\omega^{(p)}$ est de degré $p + 1$, $\omega^{(p-1)}$ est de degré p et pour une fonction f quelconque $\omega^{(p-1)}df$ est de degré $p + 1$. Ici, $p + 1 = 5$ est le nombre de variables initial et $\omega^{(p-1)}df = \omega^{(3)}df$ est une 5-forme, $\omega^{(3)}df = Cdx_1dx_2dx_3dx_4dx_5$. Pour annuler $\omega^{(3)}df$ il faut donc annuler l'unique coefficient C . On trouve donc pour Σ_1 l'unique EDP

$$x_4 \frac{\partial f}{\partial x_2} - x_3x_4 \frac{\partial f}{\partial x_4} + x_3x_5 \frac{\partial f}{\partial x_5} = \boxed{1} = 0$$

(où \boxed{n} symbolise l'expression correspondante). Elle possède quatre ($4 = p$) intégrales indépendantes $x_1, x_3, x_3x_5, x_4 + x_2x_3$. Si l'on prend comme nouvelles variables $x_1, x_2, x_3, y_4 = x_4 + x_2x_3$ et $y_5 = x_4x_5$, on peut alors éliminer la variable x_2 et réduire l'EDT à la forme normale symplectique $\tilde{\omega} = x_1dy_4 + x_3dy_5$ (cela se vérifie immédiatement).⁶⁵

Nous allons calculer une autre forme normale.à partir de $\omega^{(p-2)}$. Comme $\omega^{(p-2)}$ est de degré $p - 1$, $\omega^{(p-2)}df$ est de degré p . On a $\omega^{(p-2)} = \omega^{(2)} = \omega\omega^{(1)}$ et les fonctions f satisfaisant $\omega^{(p-2)}df = 0$ doivent annuler les coefficients des monômes d'ordre 4 de $\omega^{(p-2)}df$. On trouve le système Σ_2 d'EDP suivant⁶⁶ :

65. Il s'agit d'un bon changement de variable si $x_4 \neq 0$ car $\frac{D(x_1, x_2, x_3, y_4, y_5)}{D(x_1, x_2, x_3, x_4, x_5)} = x_4$.

66. Dans les formules du type " $dx_1dx_2 \dots dx_k$: *coefficient*" exprimant les termes d'une k -forme différentielle ϖ , on écrit à gauche les différentielles et à droite les coefficients. Les EDP associées à l'équation $\varpi = 0$ consistent à annuler les coefficients.

$$\begin{aligned}
dx_1 dx_2 dx_3 dx_4 &: x_1 x_3 x_5 \frac{\partial f}{\partial x_1} - x_2 x_3 x_5 \frac{\partial f}{\partial x_2} + x_3^2 x_5 \frac{\partial f}{\partial x_3} \\
&= x_3 x_5 \left(x_1 \frac{\partial f}{\partial x_1} - x_2 \frac{\partial f}{\partial x_2} + x_3 \frac{\partial f}{\partial x_3} \right) = x_3 x_5 \boxed{2} = 0 \\
dx_1 dx_2 dx_3 dx_5 &: x_1 x_3 x_4 \frac{\partial f}{\partial x_1} - x_2 x_3 x_4 \frac{\partial f}{\partial x_2} + x_3^2 x_4 \frac{\partial f}{\partial x_3} \\
&= x_3 x_4 \left(x_1 \frac{\partial f}{\partial x_1} - x_2 \frac{\partial f}{\partial x_2} + x_3 \frac{\partial f}{\partial x_3} \right) = x_3 x_4 \boxed{2} = 0 \\
dx_2 dx_1 dx_4 dx_5 &: x_3 x_4 \frac{\partial f}{\partial x_2} - x_3^2 x_4 \frac{\partial f}{\partial x_4} + x_3^2 x_5 \frac{\partial f}{\partial x_5} \\
&= x_3 \left(x_4 \frac{\partial f}{\partial x_2} - x_3 x_4 \frac{\partial f}{\partial x_4} + x_3 x_5 \frac{\partial f}{\partial x_5} \right) = x_3 \boxed{1} = 0 \\
dx_1 dx_3 dx_4 dx_5 &: -x_1 x_4 \frac{\partial f}{\partial x_1} - x_3 x_4 \frac{\partial f}{\partial x_3} + x_2 x_3 x_4 \frac{\partial f}{\partial x_4} - x_2 x_3 x_5 \frac{\partial f}{\partial x_5} \\
&= -x_4 \boxed{2} - x_2 \boxed{1} = 0 \\
dx_2 dx_3 dx_4 dx_5 &: -x_1 x_4 \frac{\partial f}{\partial x_2} + x_1 x_3 x_4 \frac{\partial f}{\partial x_4} - x_1 x_3 x_5 \frac{\partial f}{\partial x_5} \\
&= -x_1 \boxed{1} .
\end{aligned}$$

En dehors de l'origine, le système à résoudre se ramène au système complet adjoint Σ_2 de deux EDP

$$\begin{cases} \boxed{1} = x_1 \frac{\partial f}{\partial x_1} - x_2 \frac{\partial f}{\partial x_2} + x_3 \frac{\partial f}{\partial x_3} = 0 \\ \boxed{2} = x_4 \frac{\partial f}{\partial x_2} - x_3 x_4 \frac{\partial f}{\partial x_4} + x_3 x_5 \frac{\partial f}{\partial x_5} = 0 \end{cases}$$

possédant trois ($3 = p - 1$) intégrales indépendantes $\frac{x_1}{x_3}$, $x_2 x_3 + x_4$, $x_4 x_5$. Nous avons utilisé plus haut les intégrales $x_2 x_3 + x_4$, $x_4 x_5$. Utilisons maintenant l'intégrale $f = \frac{x_1}{x_3}$. Faisons $f = a$, c'est-à-dire $x_1 = a x_3$ et $dx_1 = a dx_3$. L'EDT devient

$$\tilde{\omega}_1 = a x_3^2 dx_2 + a x_3 x_2 dx_3 + x_3 (a + x_5) dx_4 + x_3 x_4 dx_5$$

qui dépend de 4 variables. Recalculons les dérivées de $\tilde{\omega}_1$

$$\begin{aligned}
\tilde{\omega}_1^{(1)} &= 2ax_3dx_3dx_2 + ax_3dx_2dx_3 + (a + x_5)dx_3dx_4 + x_3dx_5dx_4 \\
&\quad + x_3dx_4dx_5 + x_4dx_3dx_5 \\
&= ax_3dx_3dx_2 + (a + x_5)dx_3dx_4 + x_4dx_3dx_5 \\
\tilde{\omega}_1^{(2)} &= (ax_3^2dx_2 + ax_3x_2dx_3 + x_3(a + x_5)dx_4 + x_3x_4dx_5) \wedge \\
&\quad (ax_3dx_3dx_2 + (a + x_5)dx_3dx_4 + x_4dx_3dx_5) \\
&= ax_3^2(a + x_5)dx_2dx_3dx_4 + ax_3^2x_4dx_2dx_3dx_5 \\
&\quad + ax_3^2(a + x_5)dx_4dx_3dx_2 + x_3x_4(a + x_5)dx_4dx_3dx_5 \\
&\quad + ax_3^2x_4dx_5dx_3dx_2 + x_3x_4(a + x_5)dx_5dx_3dx_4 \\
&\equiv 0.
\end{aligned}$$

$\tilde{\omega}_1$ est donc de classe $p = 2$. On peut alors chercher les f_1 telles que la 2-forme $\tilde{\omega}_1^{(p-2)}df_1 = \tilde{\omega}_1df_1 = 0$. En annulant les coefficients des $dx_i dx_j$, on obtient le système de six EDP

$$\begin{aligned}
dx_2dx_3 &: ax_3^2 \frac{\partial f_1}{\partial x_3} - ax_2x_3 \frac{\partial f_1}{\partial x_2} = 0 \\
dx_2dx_4 &: ax_3^2 \frac{\partial f_1}{\partial x_4} - x_3(a + x_5) \frac{\partial f_1}{\partial x_2} = 0 \\
dx_2dx_5 &: ax_3^2 \frac{\partial f_1}{\partial x_5} - x_3x_4 \frac{\partial f_1}{\partial x_2} = 0 \\
dx_3dx_4 &: ax_2x_3 \frac{\partial f_1}{\partial x_4} - x_3(a + x_5) \frac{\partial f_1}{\partial x_3} = 0 \\
dx_3dx_5 &: ax_2x_3 \frac{\partial f_1}{\partial x_5} - x_3x_4 \frac{\partial f_1}{\partial x_3} = 0 \\
dx_4dx_5 &: -x_3x_4 \frac{\partial f_1}{\partial x_4} + x_3(a + x_5) \frac{\partial f_1}{\partial x_5} = 0.
\end{aligned}$$

Ce système peut s'écrire :

$$\begin{aligned}
\frac{\frac{\partial f_1}{\partial x_2}}{ax_3^2} &= \frac{\frac{\partial f_1}{\partial x_3}}{ax_2x_3} = \frac{\frac{\partial f_1}{\partial x_4}}{x_3(a + x_5)} = \frac{\frac{\partial f_1}{\partial x_5}}{x_3x_4} \\
\frac{\frac{\partial f_1}{\partial x_2}}{ax_3} &= \frac{\frac{\partial f_1}{\partial x_3}}{ax_2} = \frac{\frac{\partial f_1}{\partial x_4}}{a + x_5} = \frac{\frac{\partial f_1}{\partial x_5}}{x_4} = 1, \text{ si } x_3 \neq 0
\end{aligned}$$

et a pour intégrale (car $a = \frac{x_1}{x_3}$)

$$f_1 = ax_2x_3 + x_4(a + x_5) = b = x_1x_2 + x_4\left(\frac{x_1}{x_3} + x_5\right).$$

Cette équation $f_1 = b$ donne $x_5 = -a + \frac{b - ax_2x_3}{x_4}$. Avec cette valeur et $x_1 = ax_3$, l'EDT devient

$$\begin{aligned} \tilde{\omega}_2 &= ax_3^2 dx_2 + ax_3x_2 dx_3 + x_3(a + x_5) dx_4 + x_3x_4 dx_5 \\ &= ax_3^2 dx_2 + ax_3x_2 dx_3 + x_3 \frac{b - ax_2x_3}{x_4} dx_4 \\ &\quad + x_3x_4 \left(-\frac{a}{x_4} (x_3 dx_2 + x_2 dx_3) - \frac{1}{x_4^2} (b - ax_2x_3) dx_4 \right) \\ &= \left(ax_3^2 - x_3x_4 \frac{a}{x_4} x_3 \right) dx_2 + \left(ax_3x_2 - x_3x_4 \frac{a}{x_4} x_2 \right) dx_3 \\ &\quad + \left(x_3 \frac{b - ax_2x_3}{x_4} - x_3x_4 \frac{1}{x_4^2} (b - ax_2x_3) \right) dx_4 \\ &= 0. \end{aligned}$$

Cette EDT de 3 variables est identiquement nulle (i.e. de classe $p - 4 = 0$). Cela signifie que ω s'annule sur les sous-variétés de dimension 3 de \mathbb{R}^5 d'équations $f = a$ et $f_1 = b$ et qu'elle est donc de la forme $\omega = Ada + Bdb$. Mais comme $a = \frac{x_1}{x_3}$, $da = \frac{1}{x_3} dx_1 - \frac{x_1}{x_3^2} dx_3$ et comme $b = x_1x_2 + x_4\left(\frac{x_1}{x_3} + x_5\right)$,

$$db = x_1 dx_2 + x_2 dx_1 + \left(\frac{x_1}{x_3} + x_5\right) dx_4 + x_4 dx_5 + x_4 \left(\frac{1}{x_3} dx_1 - \frac{x_1}{x_3^2} dx_3\right)$$

et donc

$$\begin{aligned} \omega &= Ada + Bdb = \left(A \frac{1}{x_3} + Bx_2 + B \frac{x_4}{x_3} \right) dx_1 + Bx_1 dx_2 \\ &\quad + \left(-A \frac{x_1}{x_3^2} - Bx_4 \frac{x_1}{x_3^2} \right) dx_3 + B \left(\frac{x_1}{x_3} + x_5 \right) dx_4 + Bx_4 dx_5 \\ &= x_1x_3 dx_2 + x_1x_2 dx_3 + (x_1 + x_3x_5) dx_4 + x_3x_4 dx_5. \end{aligned}$$

On en tire d'abord $Bx_1 = x_1x_3$, et donc $B = x_3$ si $x_1 \neq 0$, puis $\frac{A}{x_3} + x_2x_3 + x_4 = 0$, et donc $A = -x_3(x_2x_3 + x_4)$ et l'on vérifie que l'on a bien

$$\begin{aligned} -A \frac{x_1}{x_3^2} - Bx_4 \frac{x_1}{x_3^2} &= x_3(x_2x_3 + x_4) \frac{x_1}{x_3^2} - x_4 \frac{x_1}{x_3} = x_1x_2, \\ B \left(\frac{x_1}{x_3} + x_5 \right) &= x_3 \left(\frac{x_1}{x_3} + x_5 \right) = x_1 + x_3x_5 \end{aligned}$$

et $Bx_4 = x_3x_4$.

On obtient ainsi en définitive une autre forme normale de ω :

$$\omega = -x_3(x_2x_3 + x_4)d\left(\frac{x_1}{x_3}\right) + x_3d\left(x_1x_2 + x_4\left(\frac{x_1}{x_3} + x_5\right)\right)$$

à comparer à celle que nous avons obtenue précédemment en utilisant d'autres intégrales

$$\omega = x_1d(x_4 + x_2x_3) + x_3d(x_4x_5) .$$

Cet exemple explicite la façon dont s'effectue la réduction à une forme normale dans le cas général. On voit très bien comment s'y articulent

- (i) l'analyse géométrique du type de l'EDT (dérivées successives, classe),
- (ii) la résolution fonctionnelle des systèmes adjoints permettant de trouver des changements de variables effectifs.

Remarque. Nous avons supposé que les coefficients x_3x_4 , x_3x_5 , des expressions [1] et [2] étaient différents de 0. Cela est vrai *génériquement*. Lorsque ces conditions ne sont pas satisfaites, on obtient des solutions *singulières*. Nous allons y revenir plus bas. □

3.3. L'intégration des structures de contact

3.3.1. Solutions génériques.

Une fois une EDT de Pfaff $\omega = 0$ mise sous forme normale (i.e. la classe p est égale au nombre n de variables), il s'agit d'analyser son intégrabilité. On peut supposer que l'on se trouve dans le cas où n est impair car si $n = 2m$ est pair, $\omega = \sum_{i=1}^{i=m} p_i dx_i$ avec $p_1 \neq 0$, et $\omega = 0$ équivaut à $\frac{1}{p_1}\omega = 0$ qui est de classe impaire $n - 1$. Nous considérons donc les formes normales de contact

$$\omega = dz - \sum_{i=1}^{i=m} p_i dx_i = 0$$

de classe $n = 2m + 1$ sur le $\mathbb{R}^n = \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^m$ de coordonnées (x_i, z, p_i) fibré sur l'espace de base (x_i, z) de dimension $m + 1$. Il s'agit de généraliser ce que nous avons vu au chapitre 3 avec la forme de contact $\omega = dy - p dx$ sur \mathbb{R}^3 .

L'EDT $\omega = 0$ définit des éléments de contact (v, K_v) de dimension $2m = n - 1$ et l'on cherche des solutions qui sont des sous-variétés de dimension d de \mathbb{R}^n partout tangentes aux champ des K_v . Si $z = \varphi(x_i)$ est une hypersurface S dans l'espace de base (S est de dimension m), elle induit une solution Σ de dimension m (et donc de codimension $m + 1$) qui généralise les relevées legendriennes de la section 2 du chapitre 3 et qui est définie en prenant au-dessus de tout point $s =$

$(x_0 = x_{i,0}, z_0 = \varphi(x_{i,0}))$ de S le point $v = \left(s, p_i = \frac{\partial \varphi}{\partial x_i}(x_{i,0})\right)$. Ce point est bien dans K_v puisque

$$\omega = dz - \sum_{i=1}^{i=m} p_i dx_i = dz - \sum_{i=1}^{i=m} \frac{\partial \varphi}{\partial x_i} dx_i = 0$$

si $z = \varphi(x_i)$. D'autres solutions de dimension m (et de codimension $m+1$) sont les fibres au-dessus des points de la base $s = (x_{i,0}, z_0)$.

Comme nous l'avons vu, pour Cartan une intégrale de dimension d est une sous-variété M de \mathbb{R}^n , paramétrée par $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^n$, telle que $\omega \upharpoonright_M \equiv 0$ (i.e. $\psi^*(\omega) \equiv 0$). Il montre alors que nécessairement $d \leq m$. Pour ce faire, il commence par supposer qu'il existe $h+1$ relations entre z et les x_i , ce qui définit des sous-variétés S de dimension $m-h$ de l'espace de base des (x_i, z) . Pour simplifier, il suppose que les x_{h+k} , $k = 1, \dots, m-h$ sont $m-h$ variables indépendantes et que $x_j = \xi_j(x_{h+k})$, $j = 1, \dots, h$, et $z = \zeta(x_{h+k})$ sont des fonctions des x_{h+k} . L'annulation de ω conduit au système de $m-h$ EDP

$$dx_{h+k} : \frac{\partial \zeta}{\partial x_{h+k}} - \sum_{j=1}^{j=h} p_j \frac{\partial \xi_j}{\partial x_{h+k}} - p_{h+k} = 0, \quad k = 1, \dots, m-h$$

qui conditionnent les p_i . Cela fait $m-h$ relations supplémentaires qui, génériquement (i.e. si les ξ_j et ζ n'ont pas de relations spéciales entre elles et avec leurs dérivées), sont indépendantes entre elles et indépendantes des $h+1$ autres relations. D'où un système de $h+1+m-h = m+1$ relations. La codimension c de M est $\geq m+1$ et la dimension d de M est $\leq 2m+1 - (m+1) = m$.

Le h ci-dessus peut varier de 0 à m . Si $h = 0$, $z = \zeta(x_k)$, $k = 1, \dots, m$, est fonction de tous les x_i et définit une hypersurface S de dimension m dans l'espace de base des (x_i, z) . Tous les p_i sont déterminés et égaux aux $\frac{\partial \zeta}{\partial x_i}$. Ce sont les *relevées legendriennes*. À l'autre extrême, si $h = m$, tous les x_i et z sont fixés en un $x_0 = x_{i,0}$, $z_0 = \varphi(x_{i,0})$ (i.e. les ξ_j et ζ sont constantes), les p_i sont quelconques et l'on obtient la *fibre* au-dessus de (x_0, z_0) . Pour $0 < h < m$ on obtient des solutions mixtes entre les relevées legendriennes et les fibres. Par exemple si $h = m-1$, autrement dit si l'on considère une *courbe* γ dans l'espace de base paramétrée par x_m , le système se réduit à l'unique EDP

$$\frac{\partial \zeta}{\partial x_m} - \sum_{j=1}^{j=m-1} p_j \frac{\partial \xi_j}{\partial x_m} - p_m = 0.$$

Si l'on écrit $x_m = \xi_m(x_m)$ et donc $\frac{\partial \xi_m}{\partial x_m} = 1$, on voit que cette EDP s'écrit comme le produit scalaire $\left\langle \left(\frac{\partial \xi_j}{\partial x_m}, 1, \frac{\partial \zeta}{\partial x_m} \right), (p_j, p_m, -1) \right\rangle = 0$ qui signifie que le vecteur $(p_j, p_m, -1)$ de \mathbb{R}^{m+1} est orthogonal au vecteur tangent à $\gamma \left(\frac{\partial \xi_j}{\partial x_m}, 1, \frac{\partial \zeta}{\partial x_m} \right)$ de \mathbb{R}^{m+1} .

Ensuite, pour obtenir

"tous les systèmes de relations satisfaisant à cette équation" (7.3, p. 268),

Cartan généralise le traitement au cas où certains p_i peuvent être pris comme variables indépendantes. Il suppose que λ variables dans les (x_i, z, p_i) peuvent être prises comme variables indépendantes et que les $2m + 1 - \lambda = n - \lambda$ autres en sont des fonctions.⁶⁷ Il considère donc des solutions de dimension $d = \lambda$. Il montre que, quitte à faire des transformations élémentaires, on peut supposer que z n'est pas une variable indépendante et de même pour des x_i et p_i de même i . Cela implique $d \leq m$. Il note alors $x_1, \dots, x_\alpha, p_{\alpha+1}, \dots, p_\lambda$ les variables indépendantes et démontre la formule générale pour les solutions génériques

$$z = w - \sum_{\mu=\alpha+1}^{\mu=\lambda} p_\mu \frac{\partial w}{\partial p_\mu} + \sum_{\nu=\lambda+1}^{\nu=m} v_\nu \left(\sum_{\mu=\alpha+1}^{\mu=\lambda} p_\mu \frac{\partial u_\nu}{\partial p_\mu} \right)$$

où $z = w - \sum_{\mu=\alpha+1}^{\mu=\lambda} x_\mu p_\mu$, $x_\nu = u_\nu$ ($\nu = \lambda + 1, \dots, m$), $p_\nu = v_\nu$ ($\nu = \lambda + 1, \dots, m$), les w, u_ν, v_ν étant des fonctions des variables indépendantes. Si $\lambda = m$, i.e. si l'on considère des solutions de dimension maximale, les $x_1, \dots, x_\alpha, p_{\alpha+1}, \dots, p_m$ sont les variables indépendantes et l'on a

$$z = w - \sum_{\mu=\alpha+1}^{\mu=m} p_\mu \frac{\partial w}{\partial p_\mu}$$

$$p_k = \frac{\partial w}{\partial x_k}, \quad k = 1, \dots, \alpha$$

$$x_\mu = -\frac{\partial w}{\partial p_\mu}, \quad \mu = \alpha + 1, \dots, m .$$

Si de plus $\alpha = m$, alors $z = w(x_i)$, $i = 1, \dots, m$, $p_i = \frac{\partial w}{\partial x_i}$ et l'on retrouve les relevées legendriennes.

3.3.2. Solutions singulières.

En plus des solutions génériques, une EDT peut avoir des solutions singulières comme par exemple les $p_i = 0$ pour $\omega = \sum_i p_i dx_i = 0$ où l'on annule tous les coefficients des dx_i . Elles sont caractérisées par le théorème disant que si ω est de classe $2m = n$ elles annulent les coefficients de $\omega^{(n-2)}$ et, si ω est de classe $2m+1 = n$, ceux de $\omega^{(n-1)}$.⁶⁸ Pour expliciter cette condition revenons à l'exemple de l'EDT à 5 variables de classe 4 :

$$\omega = x_1 x_3 dx_2 + x_1 x_2 dx_3 + (x_1 + x_3 x_5) dx_4 + x_3 x_4 dx_5 .$$

67. Pour des raisons techniques, Cartan suppose que les fonctions considérées sont non seulement différentiables mais analytiques.

68. Annuler les coefficients est une condition plus forte qu'annuler ω .

Il s'agit d'annuler les coefficients de la 3-forme $\omega^{(n-2)} = \omega^{(2)}$

$$\begin{aligned}\omega^{(2)} &= x_1 x_3 x_5 dx_2 dx_3 dx_4 + x_1 x_3 x_4 dx_2 dx_3 dx_5 \\ &+ x_3^2 x_5 dx_4 dx_1 dx_2 + x_3 x_5 x_2 dx_4 dx_1 dx_3 + x_1 x_4 dx_4 dx_3 dx_5 \\ &+ x_3^2 x_4 dx_5 dx_1 dx_2 + x_3 x_4 x_2 dx_5 dx_1 dx_3 + x_3 x_4 dx_5 dx_1 dx_4 .\end{aligned}$$

Les solutions les moins contraignantes sont $(x_1 = 0, x_3 = 0)$, $(x_3 = 0, x_4 = 0)$, $(x_4 = 0, x_5 = 0)$. La solution $(x_1 = 0, x_3 = 0)$ annule tous les coefficients de ω , donc ω , et est donc une solution singulière. La solution $(x_3 = 0, x_4 = 0)$ n'annule pas tous les coefficients de ω ($x_1 x_2$ et x_1 ne le sont pas), mais annule ω puisqu'elle annule à la fois les différentielles dx_3 et dx_4 et les coefficients de dx_2 et de dx_5 , et elle est donc également une solution singulière. En revanche, $(x_4 = 0, x_5 = 0)$ n'annule pas ω et n'est donc pas une solution du tout.

Nous reviendrons plus bas (section 3.5.7) sur la notion de solution singulière dans le cas d'un *système* d'EDT.

3.3.3. EDP du premier ordre avec contraintes.

Dans la quatrième partie de [90], Cartan traite des solutions d'une EDT de Pfaff lorsque les variables sont *contraintes* de satisfaire certaines équations supplémentaires. Cela est essentiel pour comprendre la résolution des EDP du premier ordre.

Nous avons vu plus haut le lien entre d'un côté une EDP du premier ordre (1) $f(x_i, z, p_i) = 0$ portant sur une fonction $z = \varphi(x_i)$ de m variables x_i et de dérivées partielles $\frac{\partial \varphi}{\partial x_i} = p_i$ et d'un autre côté l'EDT (2) de $2m + 1$ variables $\omega = dz - \sum p_i dx_i = 0$. Si l'on a un *système* d'EDP, (1) devient un système de h équations $f_k(x_i, z, p_i) = 0$, $k = 1, \dots, h$. Génériquement, (1) détermine h des x_i par rapport aux $m - h$ autres et ω devient une $\tilde{\omega}$ ne dépendant plus que de ces $m - h$ variables. La classe de $\tilde{\omega}$ est alors redéfinie comme le plus petit p tel que $\omega^{(p)} df_1 \dots df_h \equiv 0$ et la théorie se développe comme précédemment à partir de là.

3.3.4. Transformations de contact.

Revenons sur la notion de transformation de contact ("Berührungstransformationen", cf. section 5) introduite par Lie. Rappelons qu'il s'agit d'un changement de variables $\tilde{z}(x_i, z, p_i)$, $\tilde{x}_i(x_i, z, p_i)$, $\tilde{p}_i(x_i, z, p_i)$ qui laisse l'EDT $\omega = dz - \sum_{i=1}^{i=m} p_i dx_i = 0$ invariante et donc transforme ω en $\tilde{\omega} = d\tilde{z} - \sum_{i=1}^{i=m} \tilde{p}_i d\tilde{x}_i = \rho\omega$. En développant $\tilde{\omega}$ par rapport aux (x_i, z, p_i) et en l'égalant à $\rho\omega$, on trouve les équations

$$\begin{aligned}\frac{\partial \tilde{z}}{\partial z} - \sum_{i=1}^{i=m} \tilde{p}_i \frac{\partial \tilde{x}_i}{\partial z} &= \rho \\ \frac{\partial \tilde{z}}{\partial x_i} - \sum_{k=1}^{k=m} \tilde{p}_k \frac{\partial \tilde{x}_k}{\partial x_i} &= -\rho p_i \\ \frac{\partial \tilde{z}}{\partial p_i} - \sum_{k=1}^{k=m} \tilde{p}_k \frac{\partial \tilde{x}_k}{\partial p_i} &= 0 .\end{aligned}$$

En calculant les dérivées secondes de \tilde{z} et en écrivant leur commutativité ($\frac{\partial^2 \tilde{z}}{\partial x_i \partial x_j} = \frac{\partial^2 \tilde{z}}{\partial x_j \partial x_i}$, etc.) on en tire les relations

$$\begin{aligned} \sum_k \left(\frac{\partial \tilde{p}_k}{\partial p_j} \frac{\partial \tilde{x}_k}{\partial p_i} - \frac{\partial \tilde{p}_k}{\partial p_i} \frac{\partial \tilde{x}_k}{\partial p_j} \right) &= 0 \text{ pour tout } i, j \\ \sum_k \left(\frac{\partial \tilde{p}_k}{\partial x_j} \frac{\partial \tilde{x}_k}{\partial p_i} - \frac{\partial \tilde{p}_k}{\partial p_i} \frac{\partial \tilde{x}_k}{\partial x_j} \right) &= -\rho \delta_{ij} \text{ pour tout } i, j \\ \sum_k \left(\frac{\partial \tilde{p}_k}{\partial x_j} \frac{\partial \tilde{x}_k}{\partial x_i} - \frac{\partial \tilde{p}_k}{\partial x_i} \frac{\partial \tilde{x}_k}{\partial x_j} \right) &= 0 \text{ pour tout } i, j. \end{aligned}$$

À partir de là, on peut montrer (cf. Forsyth [188], p. 234-236) le théorème de Lie caractérisant les transformations de contact en termes de *crochets* pour la structure de contact (cf. plus haut section 2.7.6)

$$\{f, g\} = \sum_k \left(\frac{\partial f}{\partial p_k} \left(\frac{\partial g}{\partial x_k} + p_k \frac{\partial g}{\partial z} \right) - \frac{\partial g}{\partial p_k} \left(\frac{\partial f}{\partial x_k} + p_k \frac{\partial f}{\partial z} \right) \right)$$

les fonctions de crochet nul étant dites *en involution*.⁶⁹

Théorème (Lie). Une condition nécessaire et suffisante pour qu'un changement de variables $\tilde{z}(x_i, z, p_i)$, $\tilde{x}_i(x_i, z, p_i)$, $\tilde{p}_i(x_i, z, p_i)$ soit une transformation de contact est que leurs crochets satisfassent les relations

$$\begin{aligned} \{\tilde{z}, \tilde{x}_i\} &= 0, \\ \{\tilde{z}, \tilde{p}_i\} &= -\rho p_i, \\ \{\tilde{x}_i, \tilde{x}_j\} &= 0, \\ \{\tilde{p}_i, \tilde{p}_j\} &= 0 \text{ pour } i, j = 1, \dots, m, \\ \{\tilde{x}_i, \tilde{p}_j\} &= 0 \text{ si } i \neq j, \\ \{\tilde{x}_i, \tilde{p}_i\} &= -\rho, \\ \rho &= \frac{\partial \tilde{z}}{\partial z} - \sum_{i=1}^{i=m} \tilde{p}_i \frac{\partial \tilde{x}_i}{\partial z}. \end{aligned}$$

En particulier, les $m + 1$ nouvelles variables \tilde{x}_i et \tilde{z} sont en involution. \diamond

Cela peut se montrer facilement une fois que l'on a démontré (cf. Cartan [98], p. 321 sq.) que $\{f, g\}$ est donné par la formule

$$\omega^{(2m-2)} df dg = -\{f, g\} dz dx_1 dp_1 \cdots dx_m dp_m.$$

En effet il s'ensuit immédiatement que, si \tilde{f} et \tilde{g} sont les transformées de f et g alors leur crochet est donné par $\{\tilde{f}, \tilde{g}\} = \rho \{f, g\}$. Or, puisque les (x_i, z, p_i) sont

69. Si f et g sont indépendants de z , $\{f, g\} = \sum_k \left(\frac{\partial f}{\partial p_k} \frac{\partial g}{\partial x_k} - \frac{\partial g}{\partial p_k} \frac{\partial f}{\partial x_k} \right)$ est l'opposé du crochet de Poisson traditionnel. C'est le (f, g) de Clebsch.

des variables indépendantes, il est trivial de vérifier que $\{z, x_i\} = 0$, $\{z, p_i\} = -p_i$, $\{x_i, x_j\} = 0$, $\{p_i, p_j\} = 0$ pour $i, j = 1, \dots, m$, $\{x_i, p_j\} = 0$ si $i \neq j$, $\{x_i, p_i\} = -1$.

3.4. Les invariants intégraux de Poincaré-Cartan

Il est intéressant d'étudier les propriétés d'invariance de formes différentielles le long des trajectoires $\frac{dx_i(t)}{dt} = X_i(x_j)$ d'un champ de vecteurs X . Cela conduit à ce que l'on appelle maintenant les *invariants intégraux* de Poincaré-Cartan. Développés en particulier dans le chapitre XXII du tome III du grand traité d'Henri Poincaré *Les méthodes nouvelles de la Mécanique céleste* ([449], 1899), ils sont fondamentaux entre autres en géométrie symplectique et en mécanique hamiltonienne. Si ω est une r -forme, on considère ses intégrales $I_M = \int_M \omega$ sur les sous-variétés M de dimension r de \mathbb{R}^n et on fait des I_M des fonctions $I_M(t)$ en considérant le déplacement $M(t)$ de $M(0) = M$ le long des trajectoires du champ. La forme ω est un invariant intégral "absolu" si les $I_M(t)$ sont constantes pour toute M . Elle est dite un invariant intégral "relatif" si les $I_M(t)$ sont constantes pour toute les M fermées ($M = \partial N$, N de dimension $r+1$), ce qui permet d'appliquer le théorème de Stokes $\int_{M=\partial N} \omega = \int_N d\omega$. Le cas des f correspond aux 1-formes exactes df de fonctions $f(x)$: on écrit $f(x) = \int_{\gamma:x_0 \rightarrow x} df$ (où $\gamma : x_0 \rightarrow x$ est une courbe joignant x_0 à x , l'intégrale ne dépendant que de x_0 et x et pas de γ puisque \mathbb{R}^n est simplement connexe) et $f(t, x) = \int_{\gamma(t):x_0(t) \rightarrow x(t)} df$.

Les invariants intégraux forment un ensemble stable par combinaisons linéaires et produits extérieurs. Ils permettent d'exprimer des *lois de conservation*.

En particulier, si ω est une n -forme $\omega = F(x) dV$ (où dV est la forme volume $\bigwedge_{i=1}^n dx_i$), $I_M = \int_M F(x) dV$ et, pour calculer la variation de I_M quand on déplace le domaine M le long du champ pendant $\delta t \neq 0$, on fait le changement de variable $y(x) = x + X\delta t$ et on utilise le jacobien $\frac{D(y)}{D(x)}$ au premier ordre en δt . Or, comme la matrice jacobienne $\frac{\partial y}{\partial x}$ est $\frac{\partial y}{\partial x} = \left(\delta_{ij} + \frac{\partial X_i}{\partial x_j} \delta t \right)$, le jacobien au premier ordre est $1 + \sum_i \frac{\partial X_i}{\partial x_i} \delta t$. On en tire au premier ordre

$$\begin{aligned} \delta I_M &= \int_M F(x + X\delta t) \left(1 + \sum_i \frac{\partial X_i}{\partial x_i} \delta t \right) dV - \int_M F(x) dV \\ &= \int_M \left(F(x) + \sum_j \frac{\partial F}{\partial x_j} X_j \delta t \right) \left(1 + \sum_i \frac{\partial X_i}{\partial x_i} \delta t \right) dV - \int_M F(x) dV \\ &\sim \int_M \left(\sum_i F(x) \frac{\partial X_i}{\partial x_i} + \sum_j \frac{\partial F}{\partial x_j} X_j \right) \delta t dV . \end{aligned}$$

Pour que $\omega = F(x) dV$ soit un invariant intégral, il faut et il suffit que $\delta I_M = 0$ pour tout domaine M et, comme $\delta t \neq 0$, cela équivaut à (cf. l'analyse vectorielle classique, section 8.1 du chapitre 2)

$$\begin{aligned} \sum_i F(x_i) \frac{\partial X_i}{\partial x_i} + \sum_j \frac{\partial F}{\partial x_j} X_j &= 0 \\ F \operatorname{div}(X) + X(F) &= 0 \\ \operatorname{div}(FX) &= 0. \end{aligned}$$

Un exemple prototypique est celui des équations de Hamilton en mécanique symplectique (que nous utiliserons de façon cruciale dans les prochains chapitres), où l'on se donne des variables conjuguées x_i et p_i ($i = 1, \dots, m$) coordonnées de $T^*\mathbb{R}^m$, la 1-forme $\omega = \sum_i p_i dx_i$, un hamiltonien $H : T^*\mathbb{R}^m \rightarrow \mathbb{R}$ et les équations de Hamilton

$$\begin{cases} \frac{dx_i(t)}{dt} = \frac{\partial H}{\partial p_i} \\ \frac{dp_i(t)}{dt} = -\frac{\partial H}{\partial x_i} \end{cases}$$

associées au *gradient symplectique* de H . Par construction, la 2-forme symplectique $\tilde{\omega} = d\omega = \sum_i dp_i \wedge dx_i$ est un invariant intégral. Donc $(d\omega)^m$, qui est essentiellement la forme volume (au facteur $m!$ près), est un invariant intégral et par conséquent le volume de l'espace des phases est conservé le long des trajectoires. C'est le célèbre *théorème de Liouville*.

À l'autre extrême, si $\omega = \sum_i A_i dx_i$ est une forme de Pfaff, elle sera un invariant intégral du champ X si et seulement si, pour tout $i = 1, \dots, n$, le coefficient A_i satisfait l'EDP linéaire du premier ordre

$$A_i(x) \frac{\partial X_i}{\partial x_i} + \sum_j \frac{\partial A_i}{\partial x_j} X_j = A_i(x) \frac{\partial X_i}{\partial x_i} + X(A_i) = 0.$$

En effet le long d'une courbe γ , $I_\gamma = \int_\gamma \sum_i A_i(x) dx_i$ et

$$\begin{aligned} \delta I_\gamma &= \int_\gamma \sum_i A_i(x + X\delta t) \left(1 + \frac{\partial X_i}{\partial x_i} \delta t\right) dx_i - \int_\gamma \sum_i A_i(x) dx_i \\ &= \int_\gamma \sum_i \left(A_i(x) + \sum_j \frac{\partial A_i}{\partial x_j} X_j \delta t \right) \left(1 + \frac{\partial X_i}{\partial x_i} \delta t\right) dx_i - \int_\gamma \sum_i A_i(x) dx_i \\ &\sim \int_\gamma \sum_i \left(A_i(x) \frac{\partial X_i}{\partial x_i} + \sum_j \frac{\partial A_i}{\partial x_j} X_j \right) \delta t dx_i. \end{aligned}$$

Si ω est un invariant intégral du champ X , $\omega(X)$ est évidemment une intégrale première de X .

3.5. La pleine géométrisation de l'intégrabilité

3.5.1. *Éléments de contact intégraux et associés.*

Dans son second grand mémoire de 1901 [91], Cartan approfondit considérablement la *géométrisation* des conditions d'intégrabilité d'un système de s EDT de Pfaff de n variables

“lorsque ce système n'est pas complètement intégrable.” ([91], p. 241)

Il veut

“arriver à des résultats précis et généraux” ([91] p. 241)

et, pour ce faire, il va utiliser les *éléments de contact* et les *covariants bilinéaires* de Frobenius et Darboux que nous avons introduits plus haut dans les sections 5 et 2.7.4. Il vise une systématization des travaux antérieurs (il cite entre autres ceux de Biermann, 1885⁷⁰) montrant que, génériquement, la dimension maximale r des sous-variétés⁷¹ intégrales permettant de feuilleter \mathbb{R}^n est $\lfloor \frac{n}{s+1} \rfloor$ (la partie entière de $\frac{n}{s+1}$), le reste $n - (s+1) \lfloor \frac{n}{s+1} \rfloor$ donnant le nombre de fonctions arbitraires que l'on peut choisir. Dans notre exemple de base $\omega = dy - p dx$, $n = 3$, $s = 1$, $\lfloor \frac{n}{s+1} \rfloor = \lfloor \frac{3}{2} \rfloor = 1$, $n - (s+1) \lfloor \frac{n}{s+1} \rfloor = 3 - 2 = 1$ et les solutions génériques sont les relevées legendriennes, donc des courbes gauches de dimension 1 dans \mathbb{R}^n , qui dépendent d'une fonction arbitraire $y = f(x)$ dans le plan.

Remarque. On ne saurait trop insister sur cette conséquence de l'inversion de point de vue (la “révolution copernicienne-galoisienne”) apporté par la géométrisation de l'intégrabilité à la Lie-Frobenius-Cartan. La classe des *phénomènes empiriques* considérés (ici les courbes régulières du plan $y = f(x)$ que sont les contours perçus) apparaissent comme les “constantes d'intégration” de l'intégration d'une *équation de structure universelle “synthétique a priori”* (ici l'équation de Pfaff $\omega = dy - p dx = 0$) dans un espace idoine (ici l'espace des 1-jets). On peut approfondir ainsi la correspondance des vocabulaires.

70. Goursat [218] cite aussi de son côté les travaux d'Engel de 1889 et de Weber entre 1898 et 1902.

71. Cartan utilise le terme de “multiplicité” qui a été progressivement remplacé par celui de “variété”.

	Neurogéométrie	Géométrie de l'intégrabilité
Données empiriques purement phénoménales	Contours comme simples ensembles de neurones de l'aire corticale V1 activés par la détection de contrastes locaux par les cellules ganglionnaires de la rétine. Ils ne sont pas encore des contours <i>perçus</i> globalement en tant que tels.	Courbes planes arbitraires $y = f(x)$ comme simples ensembles d'éléments de contact.
Structure synthétique a priori	Connectivité de l'architecture fonctionnelle corticale.	Equation de Pfaff $\omega = 0$ de l'espace des 1-jets \mathbb{V}_J .
Phénomènes objectifs, globalisés et intégrés	Contours perçus en tant que tels.	Courbes planes arbitraires $y = f(x)$ comme "constantes (fonctionnelles) d'intégration de l'équation $\omega = 0$ dont leurs relevées legendriennes sont les intégrales.

□

Soit donc (S) $\omega_\alpha = \sum_{i=1}^{i=n} A_{\alpha,i} dx_i = 0$, $\alpha = 1, \dots, s$, un système de $s \leq n - 1$ EDT de Pfaff *indépendantes*.⁷² De tels systèmes sont très généraux puisque

"tout système d'équations aux dérivées partielles peut se ramener à un système d'équations aux différentielles totales, en regardant au besoin certaines des dérivées partielles des fonctions inconnues comme de nouvelles variables indépendantes." ([91] p. 247)

Suivant la méthodologie générale exposée plus haut, on cherche des variétés intégrales M^p du système (S) qui sont de dimension p (et donc de codimension $n - p$) paramétrées par p variables u_ℓ , $\ell = 1, \dots, p$, les EDT (S) exprimant

72. On suppose $s \leq n - 1$ car sinon le système (S) serait surdéterminé et n'aurait que les points comme solutions.

“des conditions auxquelles doivent satisfaire les coordonnées x_1, \dots, x_n d’un point de la multiplicité dans un déplacement quelconque sur cette multiplicité.” ([91] p. 247)

Ces M^p sont définies par $(n-p)$ équations $f_k = 0$, $k = 1, \dots, n-p$, (S) étant une conséquence des $2(n-p)$ équations $f_k = 0$ et $df_k = 0$ (non nécessairement indépendantes). Si M^p est paramétrée par des u_ℓ au moyen des fonctions $x_i = \xi_i(u)$, les $\xi_i(u)$ satisfont le système des sp EDP $\sum_{i=1}^{i=n} A_{\alpha,i}(u) \frac{\partial \xi_i(u)}{\partial u_\ell} = 0$. Si $s = n$, et si le déterminant D des coefficients $A_{\alpha,i}$ n’est pas identiquement nul, alors les solutions de (S) sont les points a . Mais sur l’hypersurface d’équation $D = 0$, il peut y avoir des intégrales de (S) de dimension ≥ 1 . Si $s = n-1$, le système (S) est un système d’EDO à n variables possédant des intégrales M^1 qui sont des courbes $x(t)$ dépendant de $(n-1)$ paramètres arbitraires.⁷³

La géométrie de l’intégrabilité va s’exprimer au moyen d’éléments de contacts (EC) de forme (a, E^p) où a est un point de \mathbb{R}^n et E^p un sous-espace vectoriel de dimension p de $T_a\mathbb{R}^n$.⁷⁴ Ces (a, E^p) sont des espaces vectoriels constitués de (a, E^1) de dimension 1 de $T_a\mathbb{R}^n$ qui sont des éléments “linéaires” (EL) identifiés à des droites passant par a au moyen de l’identification $T_a\mathbb{R}^n \simeq \mathbb{R}^n$.⁷⁵ Les noyaux $E_\alpha = \ker(\omega_\alpha)$ des ω_α sont des EC E_α^{n-1} de dimension $(n-1)$ qui sont donc des hyperplans et le système (S) définit en chaque point a un $E_S = \bigcap_\alpha E_\alpha^{n-1}$ qui, génériquement, sera un E_S^{n-s} car les E_α^{n-1} seront transverses les uns aux autres. En certains points singuliers, les E_S pourront avoir une dimension supérieure à $(n-s)$ si certains des s hyperplans E_α^{n-1} coïncident. Comme l’exprime Cartan dans cette citation que nous déjà commentée plus haut section 5,

“Le système (S) exprime que les tangentes en un point quelconque de l’espace à une multiplicité M^p qui passe par ce point satisfont à certaines conditions qui ne dépendent que du point considéré.” ([91] p. 248, souligné par Cartan)

Ainsi en chaque point a on se donne un EC $E_{a,S} \subset T_a\mathbb{R}^n$ constitué d’EL *intégraux*, ce qui définit ce que l’on nomme aujourd’hui une *distribution* de dimension $n-s$.

73. Si les $(n-1)$ équations du système (S) sont bien indépendantes, la matrice $n \times (n-1)$ des coefficients $(A_{\alpha,i})$ est de rang $(n-1)$ (par exemple relativement aux $(n-1)$ premières variables x_i) et permet de définir le système de $(n-1)$ équations $\sum_{i=1}^{i=n-1} A_{\alpha,i} x'_i(t) = -A_{\alpha,n} x'_n(t)$, système linéaire de la forme $\Omega \tilde{x}' = W x'_n$ où Ω est la $(n-1) \times (n-1)$ matrice des $A_{\alpha,i \leq n-1}$, \tilde{x}' le $(n-1)$ -vecteur des $x'_{i \leq n-1}$ et W le $(n-1)$ -vecteur des $-A_{\alpha,n}$. Les solutions sont les EDO données par $\tilde{x}' = \Omega^{-1} W x'_n$ qui donne par intégration les $x'_{i \leq n-1}(t)$ une fois choisie arbitrairement $x_n(t)$.

74. Dans notre exemple de base, les plans de contact étaient noté K pour “contact”. Ici nous gardons la notation de Cartan, E pour “élément”. Suivant que le contexte détermine a et p , on notera un EC (a, E^p) , (a, E) , E_a^p ou E .

75. Les $T_a\mathbb{R}^n$ déplacent l’origine 0 de \mathbb{R}^n en a et sont une façon de traiter \mathbb{R}^n comme un espace affine.

Intégrer (S) signifie alors trouver des sous-variétés M^q de \mathbb{R}^n de dimension $q \leq n - s$ qui sont tangentes à $E_{a,S}$ en chacun de leurs points a , i.e. telles que

$$T_a M. \subset E_{a,S} \subset T_a \mathbb{R}^n .$$

Si un $E_a^p \subset T_a M$ ($p \leq q$) , Cartan dit que E_a^p "appartient" à M . Les EC E^p appartenant à une M^p intégrale sous-jacente sont dits *intégraux*. Ils satisfont (S) mais également, comme nous allons le voir, des conditions *supplémentaires*.

Remarque. Il existe trois classes naturelles d'EC de dimension p : (i) les sous-espaces E^p de dimension p des espaces tangents $T_a \mathbb{R}^n$, (ii) les E^p satisfaisant (S), (iii) les E^p intégraux tangents à une M^p intégrale sous-jacente. Cartan ne donne pas de nom aux E^p de classe (ii). On pourrait les appeler "presqu'intégraux", comme on parle de structure "presque complexe". Les E^1 presqu'intégraux sont automatiquement intégraux _____ □

Remarque. Il faut noter qu'ici les EDT sont quelconques et qu'il n'y a pas de contraintes de type "forme normale". _____ □

Remarque sur les dimensions. Comme nous l'avons vu plusieurs fois (sections 6.2.1 et 6.3 du chapitre 2 et aussi 2.7.4), la dualité entre vecteurs tangents et vecteurs cotangents (entre champs de vecteurs et 1-formes), n'étant pas encore à l'époque totalement désambiguïsée, les différentielles dx_i sont à la fois des symboles différentiels et des variations infinitésimales des x_i . Dans les EC intégraux, Élie Cartan considère donc les différentielles dx_i comme des coefficients directeurs de droites, c'est à dire comme des δx_i satisfaisant $\omega(\delta x) = 0$. En conséquence, quand il calcule des dimensions, ce sont des dimensions *projectives*, un sous-espace vectoriel E de dimension p étant un espace projectif de dimension $p - 1$. _____ □

Cartan reformule alors, dans ce contexte systématisé des EC intégraux, la notion de "covariant bilinéaire" de Frobenius. Évidemment si M est une variété intégrale et si E^p est un EC intégral qui appartient à M , alors E^p est constitué de EL intégraux. Mais le fait que tous ces EL intégraux soient tangents à *une même* M sous-jacente impose de fortes conditions entre eux :

"il (E^p) satisfait aussi à d'autres conditions qui peuvent être établies indépendamment de toute multiplicité intégrale particulière." ([91] p. 249, souligné par Cartan)

Dire qu'un EC E_a^p est un EC *intégral* en tant qu'EC de dimension $p > 1$ et pas seulement en tant qu'ensemble d'EL intégraux linéairement indépendants les uns des autres, c'est dire en effet que, lorsque que l'on considère deux déplacements infinitésimaux ($\delta x, \mathfrak{d}x$) à partir de a qui lui appartiennent (i.e. deux EL intégraux), on peut trouver une sous-variété W de dimension 2 à laquelle ils sont tous les deux tangents. Quelle que soit W , elle sera définie par des fonctions et la relation fondamentale de *commutativité* des dérivées secondes se traduira par des conditions entre les deux EL intégraux .Cartan expose alors, en se référant à Frobenius [200] et

à Darboux [133], la formule que nous avons démontrée en détail à la section 2.7.4 :

$$d\omega(\delta x, \mathfrak{d}x) = 0, \text{ i.e. } \omega' = 0.$$

Pour intégrer le système (S), il faut par conséquent lui ajouter le système (F) des s équations supplémentaires $\omega'_\alpha = 0$. Deux EL intégraux satisfaisant (F) sont qualifiés par Cartan d’“associés” (on dit aussi “en involution”) et les EC intégraux sont donc constitués d’EL intégraux associés entre eux.

Dans [218] (p. 88) Goursat note que si deux EL intégraux (i.e. $\omega(\delta x) = 0$ et $\omega(\mathfrak{d}x) = 0$) ne sont pas forcément en involution, d’un autre côté deux EL en involution (i.e. $d\omega(\delta x, \mathfrak{d}x) = 0$) ne sont pas forcément intégraux. Si E^p est un EC intégral, les EL en involution avec E^p forment l’EC *polaire* $\mathfrak{P}(E^p)$ de E^p , qui n’est pas forcément intégral. Évidemment $E^p \subset \mathfrak{P}(E^p)$ puisque, E^p étant intégral, tous ses EL sont associés entre eux.

La double contrainte pour deux EL d’être intégraux et en involution correspond à la condition d’intégrabilité de Frobenius $\omega \wedge d\omega = 0$.

Remarque. En dimension 3, la condition d’association possède une interprétation géométrique très simple et très intuitive en termes d’analyse vectorielle classique (cf. la section 8.1 du chapitre 2). Si $\omega = P(x, y, z) dx + Q(x, y, z) dy + R(x, y, z) dz$, on peut considérer X_ω qui est le vecteur (P, Q, R) associé à ω par $\omega(Y) = \langle X_\omega, Y \rangle$, ainsi que le pseudovecteur $\text{rot}(X_\omega)$ associé à $d\omega$. Dans ce cas, $d\omega(\delta x, \mathfrak{d}x) = 0$ signifie que le rotationnel $\text{rot}(X_\omega)$ est dans le plan des EL $(\delta x, \mathfrak{d}x)$. L’intégrabilité de ω non seulement sur des courbes mais également sur des surfaces S signifie qu’en tout point x de S le plan tangent $T_x S$ contient $\text{rot}_x(X_\omega)$. Mais $\omega = 0$ signifie que, pour tout EL δx de $T_x S$, $\langle X_\omega, \delta x \rangle = 0$ et donc l’orthogonalité $X_\omega \perp T_x S$. Par conséquent, la double condition d’intégrabilité $\omega = 0$ et $\omega' = 0$ implique $X_\omega \perp \text{rot}(X_\omega)$, c’est-à-dire que X_ω soit orthogonal à son rotationnel. Comme $\omega \wedge d\omega = \langle X_\omega, \text{rot}(X_\omega) \rangle$, la condition $X_\omega \perp \text{rot}(X_\omega)$, i.e. $\langle X_\omega, \text{rot}(X_\omega) \rangle = 0$, est bien la condition d’intégrabilité de Frobenius $\omega \wedge d\omega = 0$. \square

C’est la structure géométrique des EC et de leurs relations qu’il s’agit alors de comprendre. Pour ce faire, Cartan va avoir recours à des éléments de géométrie projective. Les EL sont des droites des espaces tangents $T_a M \subset T_a \mathbb{R}^n$ et les projectivisés $\mathbb{P}_a M$ de ces $T_a M$ sont leurs sphères⁷⁶ avec l’identification antipodale des points diamétralement opposés. Traités comme accroissement infinitésimaux en a , les coefficients directeurs δx_i de la droite $\mathbb{R}\delta x$ de $T_a \mathbb{R}^n$ deviennent les coordonnées homogènes d’un point $\delta x^{\mathbb{P}}$ de $\mathbb{P}_a \mathbb{R}^n$ (qui est de dimension $n - 1$). Le système d’EDT (S) définit un sous-espace projectif $\mathbb{P}_a E_S$ qui est génériquement de dimension $n - s - 1$ et la remarque fondamentale est que $\delta x_i \mathfrak{d}x_k - \delta x_k \mathfrak{d}x_i = \delta x \wedge \mathfrak{d}x$ sont les coordonnées plückeriennes (cf. la notice sur Julius Plücker dans 2.2) de la droite projective joignant $\delta x^{\mathbb{P}}$ et $\mathfrak{d}x^{\mathbb{P}}$ dans $\mathbb{P}_a E_S$. L’équation supplémentaire d’association $\omega' = 0$, i.e.

76. On utilise pour simplifier la structure euclidienne de \mathbb{R}^n .

$d\omega(\delta x, \mathfrak{d}x) = 0$, exprime alors que cette droite $(\delta x^{\mathbb{P}}, \mathfrak{d}x^{\mathbb{P}})$ appartient à ce qui s'appelle un *complexe linéaire* en géométrie projective. Or il existe une géométrie très spéciale et très contrainte de ces complexes linéaires.

3.5.2. Cas d'une seule équation de Pfaff ($s = 1$).

Dans le cas où le système (S) se réduit à une seule équation de Pfaff ($s = 1$), Goursat systématisé dans [218] ce que nous avons déjà vu en ordre un peu dispersé dans la section historique 2 et introduit pour ce faire 4 systèmes covariants naturels.

Le premier système $(S)_1$ est défini par $\mathfrak{A}\delta x = 0$ où \mathfrak{A} est la matrice des (\mathfrak{a}_{ik}) et il caractérise les EL δx *singuliers* qui sont associés à (en involution avec) *tous* les autres EL $\mathfrak{d}x$ au même point. On peut aussi voir $(S)_1$ comme l'EDT $\mathfrak{A}dx = 0$ ou le produit intérieur $i_{\delta x}\omega' = 0$. Si n est pair et si le déterminant Δ de \mathfrak{A} est non nul, $\Delta \neq 0$, il n'existe pas d'EL singulier, $(S)_1$ équivaut à $dx = 0$ et a pour solutions les points a ($(S)_1$ est donc complètement intégrable). Si en revanche $\Delta = 0$ (ce qui est toujours le cas si n est impair), il y a toujours des EL singuliers mais le système $(S)_1$ reste complètement intégrable. À cause de son antisymétrie, \mathfrak{A} est de rang pair $2p$ et $(S)_1$ comprend $2p$ équations indépendantes. Par changement de variables on se ramène soit au cas de contact ($\tilde{\omega} = dz - \sum_{k=1}^{k=2p} p_k dy_k$) soit au cas symplectique ($\tilde{\omega} = \sum_{k=1}^{k=2p} p_k dy_k$) et le système complet des intégrales de $(S)_1$ s'obtient en écrivant que $df = 0$ est une combinaison linéaire des $2p$ équations indépendantes de $(S)_1$. Comme nous l'avons vu avec Darboux (section 2.7.6), cela revient à annuler tous les mineurs d'ordre $(2p + 1)$ de la matrice $n \times (n + 1)$ $\begin{pmatrix} \mathfrak{A} \\ \nabla f \end{pmatrix}$, ce qui donne un système d'EDO.

Le second système $(S)_2$ définit les EL *singuliers intégraux* et permet de calculer la *classe* de ω . On adjoint à $(S)_1$ l'EDT $\omega = 0$ elle-même (i.e. le système (S)). Si $\omega = 0$ n'est pas une conséquence de $(S)_1$, $(S)_2$ comprend $2p + 1$ équations indépendantes et l'on est dans le cas de contact $\tilde{\omega} = dz - \sum_{k=1}^{k=2p} p_k dy_k$ de classe $2p + 1$. $(S)_1$ correspond à $dy_k = 0, k = 1, \dots, 2p$ et les EL portés par les droites parallèles à l'axe z sont singuliers (associés à tous les EL car $d\tilde{\omega} = \sum_{k=1}^{k=2p} dy_k \wedge p_k$). Quant au système $(S)_2$ il correspond à $dz = 0, dy_k = 0, k = 1, \dots, 2p$, et a pour solution les points. Il est complètement intégrable. Si en revanche $\omega = 0$ est une conséquence de $(S)_1$, $(S)_2 = (S)_1$ et l'on est dans le cas symplectique $\tilde{\omega} = \sum_{k=1}^{k=2p} p_k dy_k$ de classe $2p$.

Le troisième système $(S)_3$ est le *système de Pfaff* qui définit les EL δx qui, sans être forcément intégraux, sont associés (en involution) avec tous les EL *intégraux* en ce point. Pour cela, il faut et il suffit que la condition d'association $\omega' = 0$, ou plus précisément $\omega'(\delta x, \mathfrak{d}x) = 0$, soit une conséquence de $\omega(\mathfrak{d}x) = 0$. Les équations de $(S)_3$ sont donc

$$\sum_k \mathfrak{a}_{ik} dx_k = A_i dt .$$

Enfin le quatrième système $(S)_4$ définit les EL *caractéristiques*, c'est-à-dire les EL *intégraux* qui sont associés à tous les autres EL *intégraux* au même point. On adjoint donc à $(S)_3$ le système (S) $\omega = 0$. $(S)_4$ est l'équivalent pour l'EDT $\omega = 0$ de ce qu'est le système de Pfaff $(S)_3$ pour la forme ω . Nous allons y revenir plus bas.

En fait les quatre systèmes ne sont pas indépendants et, suivant les cas, se ramènent à deux d'entre eux. En effet si $\omega = 0$ n'est pas une conséquence de $(S)_1$ alors $(S)_3$ ne peut être satisfait que si $dt = 0$. Car supposons que $(S)_3$ puisse être satisfait avec $dt \neq 0$. Il existe alors une solution à $\sum_k \mathbf{a}_{ik} \lambda_k = A_i$, i.e. il existe un vecteur λ tel que $A = \mathfrak{A}\lambda$ (et réciproquement). Mais alors $\omega = 0$ est une conséquence de $(S)_1$. Mais si $(S)_3$ ne peut être satisfait que par $dt = 0$, alors $(S)_3 \Leftrightarrow (S)_1$ et $(S)_4 \Leftrightarrow (S)_2$ (classe impaire). Si par ailleurs $(S)_3$ peut être satisfait avec $dt \neq 0$ alors $\omega = 0$ est une conséquence de $(S)_1$, $(S)_2 \Leftrightarrow (S)_1$ et $(S)_4 \Leftrightarrow (S)_3$ (classe paire).

3.5.3. Prolongements de variétés intégrales.

Pour comprendre géométriquement les conditions d'intégrabilité du système (S) , Cartan adopte la méthode "ascendante". Comme $s \leq n - 1$, il existe toujours des *courbes* intégrales M^1 de (S) . Supposons que M^p soit une variété intégrale et demandons-nous à quelle condition M^p peut être *prolongée* en une variété intégrale M^{p+1} . Une condition nécessaire est évidemment que chaque EC intégral E^p appartenant à M^p soit un sous-EC d'un EC intégral E^{p+1} , ce qui présuppose que $p + 1 \leq n - s$, la dimension maximale possible des variétés intégrales. Supposons que cette condition nécessaire soit satisfaite non seulement pour les points de M^p mais pour *tous* les points de \mathbb{R}^n . Dans le cas *régulier*, Cartan démontre alors le théorème non trivial ([91] pp. 256-261) que toute variété intégrale régulière M^p peut être prolongée en une variété intégrale M^{p+1} et même que, s'il existe en tout point une famille à μ paramètres de E^{p+1} intégraux contenant les E^p intégraux, alors il existe une famille à μ paramètres de M^{p+1} intégrales contenant M^p .

3.5.4. Inégalités dimensionnelles.

Cartan va alors étudier la géométrie des distributions \mathcal{E} de dimension générique $n - s$ en introduisant un certain nombre de dimensions et en précisant leurs possibles relations mutuelles. Les points $a = E^0$ de \mathbb{R}^n ont $n = n_0$ degrés de liberté. Les EL (a, E^1) contenant $a = E^0$ en ont $n - 1$, les EL (a, E^1) intégraux en ont $n_1 \leq n - 1$ avec *génériquement* $n_1 = n - s - 1$. De même les E^2 intégraux contenant un E^1 intégral en auront génériquement n_2 , etc. jusqu'aux n_r degrés de liberté d'un E^r intégral contenant un E^{r-1} intégral, où les E^r sont les EC intégraux génériquement de dimension maximale r . Cartan appelle *genre* cette dimension maximale d'intégrabilité r . Cette situation générique n'est pas forcément valable partout. Certains EC intégraux peuvent être singuliers et par exemple être des E^{n_r-1} intégraux possédant une famille de E^{n_r} intégraux les contenant dépendant de plus de n_r paramètres. Mais, comme le remarque Cartan, la propriété d'être singulier est définie par des conditions d'*égalité*

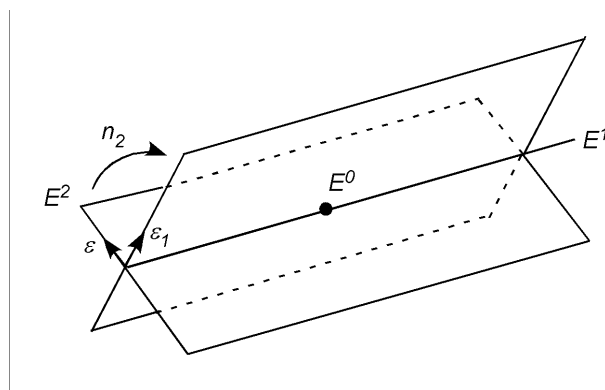


FIGURE 2. Une famille de dimension $n_2 = 1$ d'EC intégraux E^2 contenant un E^1 intégral donné passant par un E^0 donné. (E^1, ε) et (E^1, ε_1) sont deux tels E^2 .

et, dans l'espace d'EC considéré, est donc satisfaite seulement sur des sous-variétés. Or comme une sous-variété est de dimension inférieure, elle est de mesure nulle, et par conséquent ces conditions sont non génériques.

Remarque. Si un EC intégral E^r est singulier, il existe un E^{r+1} qui le contient. Mais cela n'implique pas pour autant que, si E^r appartient à une variété intégrale M^r , alors il existe un prolongement intégral M^{r+1} de M^r auquel E^{r+1} appartiendrait. En effet, le nombre r est la dimension maximale d'une sous-variété intégrale. \square

Dans une perspective ascendante, Cartan considère donc des chaînes d'inclusion $E^0 \subset E^1 \subset \dots \subset E^r$ d'EC intégraux et à chaque maillon de la chaîne s'introduisent n_i degrés de liberté (qui sont des dimensions projectives).

“Les nombres n, n_1, n_2, \dots, n_r jouent un grand rôle dans l'étude de l'indétermination de la multiplicité intégrale à r dimension la plus générale.” ([91] p. 263)

Il va alors préciser les contraintes dimensionnelles sur les n_i . Il montre d'abord ([91] p. 264) que

$$n_p \geq n_{p+1} + 1 \text{ pour } p = 0, \dots, r - 1 .$$

Cela est évident pour $p = 0$ puisque, comme $n_1 \leq n - 1$, on a bien $n_0 = n \geq n_1 + 1$. Cela se montre facilement pour $p \geq 1$. Soit E^p un EC intégral régulier passant par un E^{p-1} intégral donné, $p \leq r - 1$. Il est inclus dans une n_p -variété de E^p intégraux. Il y a donc $(n_p + 1)$ EL intégraux $\varepsilon, \varepsilon_1, \dots, \varepsilon_{n_p}$ indépendants de E^{p-1} et indépendants entre eux. L'EL intégral ε engendre un de ces $E^p = (E^{p-1}, \varepsilon)$ intégraux et ensuite il y a n_p degrés de liberté pour faire tourner ce E^p autour de E^{p-1} en préservant son caractère intégral, ce qui engendre $(E^{p-1}, \varepsilon_1), \dots, (E^{p-1}, \varepsilon_{n_p})$. La figure 2 montre un exemple pour $p = 2$ et $n_2 = 1$. Un tel cas se rencontre avec les E^1 que sont les fibres de notre structure de contact standard, les E^2 étant les plans de contact verticaux tournant autour de la fibre.

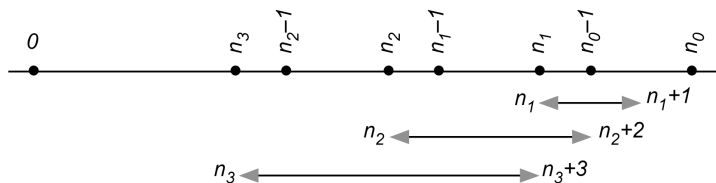


FIGURE 3. $n_p \geq n_{p+1} + 1$ pour $p = 0, \dots, r - 1 = 3$.

Mais $(E^{p-1}, \varepsilon) = E^p$ est un EC intégral qui appartient lui-même à une n_{p+1} -variété de E^{p+1} intégraux dont chacun peut être défini par un EL intégral indépendant de $(E^{p-1}, \varepsilon) = E^p$ mais dépendant de $(E^{p-1}, \varepsilon, \varepsilon_1, \dots, \varepsilon_{n_p})$. On doit donc pouvoir trouver $n_{p+1} + 1$ EL de ce type qui sont indépendants et donc $n_p \geq n_{p+1} + 1$.

La conséquence de ces inégalités dimensionnelles est que

$$n \geq n_1 + 1 \geq n_2 + 2 \geq \dots \geq n_r + r \geq r .$$

La figure 3 montre une chaîne pour $r = 4$.

Cartan montre de même que si l'on considère maintenant les EC intégraux E^{p+1} (et non plus E^p) passant par un EC intégral régulier E^{p-1} donné ($p \leq r - 1$), ils dépendent de $n_p + n_{p+1} - 1$ paramètres. Par exemple, si $E^0 = a$ et si $n_1 = 1$, il y a un plan de E^1 intégraux passant par $E^0 = a$. Comme $n_1 - 1 \geq n_2$, $n_2 = 0$ et il n'y a donc qu'un seul E^2 intégral passant par $E^0 = a$, le plan de ces E^1 . La dimension de la variété de ces E^2 est donc 0, ce qui correspond bien à $n_1 + n_2 - 1 = 0$. On peut itérer cette relation et montrer que si $p \leq r - i$ et si l'on se donne un E^{p-1} intégral régulier, les E^{p+i} intégraux passant par E^{p-1} dépendent d'un nombre de paramètres égal à

$$\begin{aligned} & (n_p - i) + (n_{p+1} - (i + 1)) + \dots + (n_{p+i-1} - 1) + n_{p+i} \\ &= n_p + n_{p+1} + \dots + n_{p+i} - \frac{i(i + 1)}{2} . \end{aligned}$$

En particulier si $p = 1$, $E^{p-1} = E^0 = a$ et $i = r - 1$, les E^r intégraux de dimension maximale r dépendent de $n_1 + \dots + n_r - \frac{(r-1)r}{2}$ paramètres.

Cartan démontre ensuite deux autres théorèmes le conduisant à introduire une autre suite de nombres caractéristiques $s_p = n_p - n_{p+1} - 1$ dont il dit :

“les nombres de cette suite jouent un très grand rôle.” ([91] p. 269)

Théorème.

$$\begin{aligned} n - n_1 - 1 &\geq n_1 - n_2 - 1 \geq \dots \geq n_{r-1} - n_r - 1 \geq n_r \geq 0 \\ s &\geq s_1 \geq \dots \geq s_{r-1} \geq s_r \geq 0 . \end{aligned}$$



Évidemment, si l'un des s_ν pour $\nu < r$ est nul alors $s_\nu = \dots = s_r = 0$. Ces s_p mesurent

"l'indétermination de l'intégrale générale M^r d'un système (S) de genre r ."
 ([91] p. 286)

Cartan appelle les s_p les *caractères* du système (S). La somme $s + s_1$ est le nombre d'équations indépendantes exprimant qu'un EL intégral $d\bar{x}$ est associé à un EL intégral $\delta\bar{x}$ arbitraire. et se calcule à partir des mineurs de la matrice des $A_{\alpha,i}$ et des $\mathbf{a}_{\alpha,ik} = \frac{\partial A_{\alpha,i}}{\partial x_k} - \frac{\partial A_{\alpha,k}}{\partial x_i}$.

Théorème. Si s_ν est le plus grand $s_\nu = 0$ (i.e. ν est minimal pour cette propriété) alors le genre r est $r = \nu + n_\nu$. Dans ce cas, il passe par tout EC intégral non singulier $E^{\nu-1}$ de dimension $(\nu - 1)$ un unique EC intégral E^r de dimension génériquement maximale r . On a $s_\nu = \dots = s_r = 0$. ◇

En effet

$$n_\nu = n_{\nu+1} + 1 = n_{\nu+2} + 2 = \dots = n_{\nu+r-\nu} + r - \nu = n_r + r - \nu = r - \nu .$$

On peut être un peu plus précis dans le cas *générique* où les coefficients des EDT du système (S) sont sans "particularisation". En effet à chaque étape E^{p-1} d'une chaîne ascendante $E^0 \subset E^1 \subset \dots \subset E^r$ d'EC intégraux, on introduit un nouvel EL intégral qui doit être associé aux EL d'une base de E^{p-1} ce qui ajoute les s conditions $\omega'_\alpha = 0$ du système (F) et si $n_{p-1} \geq s$ on obtient $n_{p+1} = n_p - 1 - s$ (si $n_{p-1} < s$, il n'existe pas de E^{p+1} intégral). Bref, on passe de n_p à n_{p+1} en enlevant $(s + 1)$ à chaque pas et l'on a $s = s_1 = \dots = s_{r-1}$ et $s_r = n_r = n - r(s + 1)$. D'où la formule de Biermann disant que, génériquement, $r = \lfloor \frac{n}{s+1} \rfloor$, avec $n_r = n - r(s + 1)$ donnant le nombre de fonctions arbitraires que l'on peut choisir. Le cas traité ci-dessus de $s_\nu = \dots = s_r = 0$ correspond au cas où n est multiple de $(s + 1)$.

Remarque. Si $s = 1$, autrement dit si le système (S) se réduit à une seule EDT, $r = \lfloor \frac{n}{2} \rfloor$ et c'est donc la parité de n qui est déterminante. Si $n = 2m$ (cas symplectique), $r = m$, $s = s_1 = \dots = s_{m-1} = 1$, $s_m = n_m = n - 2m = 0$ et une M^{m-1} intégrale n'est incluse que dans une seule M^m intégrale de dimension maximale. Si $n = 2m + 1$ (cas de contact), $r = m$, $s = s_1 = \dots = s_{m-1} = 1$, $s_m = n_m = n - 2m = 1$ et on peut choisir arbitrairement une fonction arbitraire de m variables (relevées legendriennes). □

3.5.5. *La solution générale dans le cas $n = 5$.*

La question se pose alors de construire une *solution générale* du système (S), du moins localement au voisinage d'un point régulier $E^0 = a$. Cela revient à *feuilleter* un voisinage V de a dans \mathbb{R}^n au moyen de variétés intégrales M^r de dimension maximale, chaque point de V appartenant à une de ces M^r . Cartan montre par de longs calculs que, génériquement, ces variétés intégrales M^r de dimension maximale feuilletant V dépendent de n_r fonctions arbitraires de r arguments, de s fonctions

arbitraires de $r - 1$ arguments, etc., et de s constantes arbitraires. Rappelons qu'une des caractéristiques des EDP et EDT est que leurs solutions ne dépendent pas seulement de constantes arbitraires mais aussi de *fonctions* arbitraires et qu'il est essentiel de comprendre comment.

Le cas $n = 3$ $\omega = dy - pdx = 0$ est notre cas des relevées legendriennes de courbes. Revenons sur le cas $n = 5$ avec $\omega = dz - pdx - qdy = 0$, x et y étant les variables indépendantes. Nous l'avons évoqué d'emblée dans la remarque de la section 2.4 comme généralisation directe des relevées legendriennes. On a $m = r = 2$, $s = s_1 = 1$, $s_2 = n_2 = 5 - 4 = 1$. Les EC presque-intégraux, i.e. les noyaux de ω , sont de dimension 4 et les intégrales de dimension maximale $r = m = 2$ sont des surfaces M^2 dans \mathbb{R}^5 . Si S est une surface $z = z(x, y)$ de $\mathbb{R}^3_{(x,y,z)}$, sa relevée legendrienne dans \mathbb{R}^5 est $(x, y, z = z(x, y), p = \frac{\partial z}{\partial x}, q = \frac{\partial z}{\partial y})$. Soit M^1 une courbe intégrale de (S) donnée par 5 fonctions

$$\left(x(\tau), y(\tau), z(\tau) = z(x(\tau), y(\tau)), p(\tau) = \frac{\partial z}{\partial x}(\tau), q(\tau) = \frac{\partial z}{\partial y}(\tau) \right)$$

d'une variable τ . Elle satisfait

$$z'(\tau) = p(\tau)x'(\tau) + q(\tau)y'(\tau) = \frac{\partial z}{\partial x}(\tau)x'(\tau) + \frac{\partial z}{\partial y}(\tau)y'(\tau) .$$

Le genre est

$$r = m = 2 = \left[\frac{n}{s+1} \right] = \left[\frac{5}{2} \right] ,$$

et

$$\begin{aligned} n_0 &= n = 5 , \\ n_1 &= n - (s+1) = 5 - 2 = 3 \end{aligned}$$

(n_1 est la dimension projective des E^4 presque-intégraux). Pour n_2 , les E^2 intégraux doivent satisfaire deux équations $\omega = 0$ et $\omega' = 0$ (condition d'association) et

$$n_2 = n - 2(s+1) = 5 - 4 = 1 .$$

On vérifie que l'on a bien $n_p \geq n_{p+1} + 1$ puisque

$$\begin{aligned} n_0 &= 5 > n_1 + 1 = 3 + 1 = 4 , \\ n_1 &= 3 > n_2 + 1 = 1 + 1 = 2 , \end{aligned}$$

ainsi que $n_p + p \geq r = 2$ puisque

$$\begin{aligned}
n_0 &= 5 > 2, \\
n_1 + 1 &= 3 + 1 = 4 > 2, \\
n_2 + 2 &= 1 + 2 = 3 > 2 ; s_0 = s = n_0 - n_1 - 1 = 5 - 3 - 1 = 1
\end{aligned}$$

et aussi

$$\begin{aligned}
s_0 &= s = n_0 - n_1 - 1 = 5 - 3 - 1 = 1 \\
&\geq s_1 = n_1 - n_2 - 1 = 3 - 1 - 1 = 1 \\
&\geq s_2 = n_2 = 1 \geq 0
\end{aligned}$$

La solution générale dépendra d'une ($n_2 = s_2 = 1$) fonction arbitraire de $r = m = 2$ arguments, d'une ($s = 1$) fonction arbitraire de un ($r - 1$) argument et d'une ($s = 1$) constante. Voyons comment.

L'origine 0 est un point E^0 régulier de (S). Le E^2 défini par les 3 équations indépendantes $dz = 0$, $dp = 0$, $dq = 0$ est évidemment un E^2 presque-intégral en 0 puisque les coefficients $p = 0$ et $q = 0$ sont nuls en 0. Mais il est également intégral parce que $\omega' = dx \wedge dp + dy \wedge dq = 0$ puisque les différentielles $dp = 0$, $dq = 0$ sont nulles. Le E^1 intégral régulier d'équations $dz = 0$, $dp = 0$, $dq = 0$, $dy = 0$ est inclus dans ce E^2 et $E^0 = 0$ satisfait en plus $dz = 0$, $dp = 0$, $dq = 0$, $dy = 0$, $dx = 0$. Le E^4 presque-intégral $E^4 = \ker(\omega)$ en 0 est d'équation $dz = 0$ puisque $p = q = 0$. Les EL intégraux associés à E^1 constituent un E^3 d'équations $dz = 0$, $dp = 0$. En effet la relation d'association s'écrit $\mathfrak{d}x\delta p - \mathfrak{d}p\delta x + \mathfrak{d}y\delta q - \mathfrak{d}q\delta y$ et si $\delta z = \delta p = \delta q = \delta y = 0$ (appartenance à E^1), il reste à satisfaire $-\mathfrak{d}p\delta x = 0$ pour δx quelconque et il faut par conséquent $\mathfrak{d}p = 0$. À partir de cette chaîne

$$E^0 = 0 \subset E^1 \subset E^2 \subset E^3 \subset E^4 \subset E^5 = T_0\mathbb{R}^5,$$

on peut, pour chaque point a d'un voisinage V de 0 construire localement une variété intégrale M^2 passant par a et définie par trois fonctions $z(x, y)$, $p(x, y)$, $q(x, y)$ calculables à partir d'une fonction arbitraire de (x, y) , $q = \chi(x, y)$, d'une fonction arbitraire de x , $p = \pi(x)$ pour $y = 0$ et d'une constante arbitraire $z = c$ pour $x = y = 0$, les données $\chi(0, 0)$, $\pi(0)$ et c étant assez petites pour que l'on reste dans le voisinage V du point régulier 0. On construit d'abord à partir de $\pi(x)$ et $\chi(x, 0)$ une fonction $\zeta(x)$ telle que $\zeta(0) = c$. Comme $\zeta'(x) = \pi(x)$, on a donc $\zeta(x) = c + \int_0^x \pi(u) du$. Ensuite, on construit les fonctions $z(x, y)$ et $p(x, y)$ à partir de $z(x, 0) = \zeta(x)$ et $p(x, 0) = \pi(x)$. Comme $\frac{\partial z}{\partial y} = \chi(x, y)$ et $\frac{\partial p}{\partial y} = \frac{\partial \chi}{\partial x}$, on trouve

$$\begin{cases}
z(x, y) = c + \int_0^x \pi(u) du + \int_0^y \chi(x, v) dv \\
p(x, y) = \pi(x) + \int_0^y \frac{\partial \chi}{\partial x}(x, v) dv \\
q(x, y) = \chi(x, y)
\end{cases}$$

Remarque. Pour notre cas de base $\omega = dy - p dx = 0$ ($n = 3$), on construit la solution générale dans un voisinage V de 0 à partir d'une fonction arbitraire $p = \pi(x)$ et d'une constante arbitraire c , les données $\pi(0)$ et c étant assez petites pour que l'on reste dans le voisinage V . La solution est $y(x) = c + \int_0^x \pi(u) du$ et repose sur la dualité projective (cf. section 2.1) calculant y à partir de p .

3.5.6. *Éléments caractéristiques.*

Parmi les EC intégraux E^p issus d'un point a régulier, certains sont dits *caractéristiques* (cf. plus haut le système $(S)_4$ de Goursat), au sens où ils sont associés à *tous* les EL intégraux issus de a (i.e. à $E_{a,s}$). Ils doivent donc satisfaire le système (S) (ils sont intégraux) et les s équations d'association $d\omega_\alpha(\delta x, \mathfrak{d}x) = 0$ pour tout $\delta x \in E^p$ et tout $\mathfrak{d}x$ satisfaisant (S).

Si E^p est caractéristique, tout E^r intégral régulier de dimension maximale (r est le genre de (S)) contient E^p . En effet, supposons qu'un E^r intégral ne contienne pas E^p . Alors $E^p + E^r = E^q$ serait de dimension $q \geq r + 1$. Mais comme E^p est caractéristique, E^q serait *intégral* et donc E^r serait inclus dans un E^q intégral de dimension $> r$ et serait donc *singulier*.

Si par tout point régulier il passe un EC caractéristique alors *toutes* les variétés intégrales M^r régulières de dimension maximale passant par un point régulier contiennent l'EC de ce point. C'est en particulier le cas si par tout point passe un EL intégral E^1 caractéristique. Ce champ de droites admet des courbes intégrales que l'on appelle des *courbes caractéristiques* et toute variété intégrale régulière M^r est engendrée par une famille à $(r - 1)$ paramètres de courbes caractéristiques. Donc si M^{r-1} est une variété intégrale qui n'est pas engendrée par des courbes caractéristiques, l'unique M^r prolongeant M^{r-1} s'obtient en faisant passer par chaque point de M^{r-1} la courbe caractéristique.

3.5.7. *Éléments singuliers.*

Nous avons défini plus haut à la section 3.3.2 la notion de solution singulière d'une EDT. Précisons cette notion dans le cas d'un système (S) d'EDT. Il s'agit essentiellement d'un problème dimensionnel (cf. Goursat [218]). Soit E^1 (support d'un δx) un EL *non* caractéristique. Les EL intégraux E'^1 associés à E^1 (support d'un $\mathfrak{d}x$) sont caractérisés par $2s$ équations en les $\mathfrak{d}x_i$ (les s équations de (S) $\omega_\alpha(\mathfrak{d}x) = 0$ et les s équations d'association $i_{\delta x} d\omega_\alpha = 0$). Soit $2s - q$ le nombre d'équations linéairement indépendantes (ce nombre dépend de la forme particulière de (S)). Génériquement, les $\mathfrak{d}x$ formeront un espace de dimension $(n - 2s + q)$. On dit que δx est *singulier* lorsque le nombre d'équations linéairement indépendantes est $< 2s - q$ sans que pour autant les s relations d'association soient des conséquences de (S) (dans ce cas $q = s$).

Goursat [218] donne l'exemple suivant à 8 variables. Le système (S) est

$$(G) \begin{cases} \omega_1 = dz - p dx - q dy \\ \omega_2 = dp - r dx - s dy \\ \omega_3 = dq - s dx - t dy \end{cases}$$

correspondant aux 2-jets de fonctions $z(x, y)$ (dans ce cas, p et q sont les dérivées premières de z et r, s et t les dérivées secondes). Ce système ne possède pas de caractéristiques. Montrons-le à la main (exercice scolaire). En effet

$$\begin{aligned} d\omega_2(\delta\bar{x}, \mathfrak{d}\bar{x}) &= \mathfrak{d}r\delta x - \mathfrak{d}x\delta r + \mathfrak{d}s\delta y - \mathfrak{d}y\delta s \\ d\omega_3(\delta\bar{x}, \mathfrak{d}\bar{x}) &= \mathfrak{d}s\delta x - \mathfrak{d}x\delta s + \mathfrak{d}t\delta y - \mathfrak{d}y\delta t \end{aligned}$$

et l'on veut en particulier $d\omega_2(\delta\bar{x}, \mathfrak{d}\bar{x}) = 0$ pour tout $\mathfrak{d}\bar{x}$ intégral. Mais

$$\mathfrak{d}\bar{x} = (\mathfrak{d}x = 0, \mathfrak{d}y = 0, \mathfrak{d}z = 0, \mathfrak{d}p = 0, \mathfrak{d}q = 0, \mathfrak{d}r, \mathfrak{d}s, \mathfrak{d}t)$$

est toujours intégral et il faut donc $\mathfrak{d}r\delta x + \mathfrak{d}s\delta y = 0$ et $\mathfrak{d}s\delta x + \mathfrak{d}t\delta y = 0$ pour tous $\mathfrak{d}r, \mathfrak{d}s, \mathfrak{d}t$, ce qui implique $\delta x = \delta y = 0$. Comme $\delta\bar{x}$ est intégral, $\delta x = \delta y = 0$ implique $\delta z = \delta p = \delta q = 0$ et par conséquent $d\omega_1(\delta\bar{x}, \mathfrak{d}\bar{x}) = 0$. Il reste donc à satisfaire $\mathfrak{d}x\delta r + \mathfrak{d}y\delta s = 0$ et $\mathfrak{d}x\delta s + \mathfrak{d}y\delta t = 0$ pour tout $\mathfrak{d}\bar{x}$ intégral. Mais comme il existe pour des raisons dimensionnelles des EL avec des $\mathfrak{d}x, \mathfrak{d}y$ linéairement indépendants, il faut en plus $\delta r = \delta s = \delta t = 0$. Donc la seule solution est $\delta\bar{x} = 0$ et il n'y a pas d'EL caractéristique.

En revanche le système (S) admet des EL singuliers. En effet,

$$d\omega_1(\delta\bar{x}, \mathfrak{d}\bar{x}) = \mathfrak{d}p\delta x - \mathfrak{d}x\delta p + \mathfrak{d}q\delta y - \mathfrak{d}y\delta q$$

et si $\delta\bar{x}$ et $\mathfrak{d}\bar{x}$ sont intégraux, $\delta p = r\delta x + s\delta y$, $\mathfrak{d}p = r\mathfrak{d}x + s\mathfrak{d}y$, $\delta q = s\delta x + t\delta y$ et $\mathfrak{d}q = s\mathfrak{d}x + t\mathfrak{d}y$ et par conséquent

$$\begin{aligned} d\omega_1(\delta\bar{x}, \mathfrak{d}\bar{x}) &= (r\mathfrak{d}x + s\mathfrak{d}y)\delta x - \mathfrak{d}x(r\delta x + s\delta y) + (s\mathfrak{d}x + t\mathfrak{d}y)\delta y - \mathfrak{d}y(s\delta x + t\delta y) \\ &= r\mathfrak{d}x\delta x + s\mathfrak{d}y\delta x - r\mathfrak{d}x\delta x - s\mathfrak{d}x\delta y + s\mathfrak{d}x\delta y + t\mathfrak{d}y\delta y - s\mathfrak{d}y\delta x - t\mathfrak{d}y\delta y \\ &= 0 \end{aligned}$$

ce qui montre que l'association $d\omega_1 = 0$ découle de (G) (autrement dit, $2s - q = 6 - 1 = 5$). Pour continuer, on suppose que, par exemple, la variable r est donnée en fonction des 7 autres et on calcule $d\omega_2(\delta\bar{x}, \mathfrak{d}\bar{x})$ à partir de cette hypothèse en développant δr et $\mathfrak{d}r$. On obtient un développement en $\mathfrak{d}x, \mathfrak{d}y, \mathfrak{d}s, \mathfrak{d}t$. Pour trouver les éléments singuliers, on écrit que les deux équations en $\mathfrak{d}\bar{x}$ $d\omega_2(\delta\bar{x}, \mathfrak{d}\bar{x}) = 0$ et $d\omega_3(\delta\bar{x}, \mathfrak{d}\bar{x}) = 0$ ne sont pas indépendantes, ce qui revient à dire que leurs coefficients en $\mathfrak{d}x, \mathfrak{d}y, \mathfrak{d}s, \mathfrak{d}t$ sont proportionnels. Les calculs montrent d'abord que l'on doit avoir

$$(\delta y)^2 + \frac{\partial r}{\partial s} \delta x \delta y - \frac{\partial r}{\partial t} (\delta x)^2 = 0$$

et, qu'en reportant une des solutions $dy = Kdx$ dans les autres équations, on obtient une nouvelle EDT reliant δx , δs , δt . En rajoutant ces deux EDT supplémentaires au système (G), on obtient un système de Pfaff de 5 équations en 7 variables définissant les EL intégraux singuliers de (G).

*
* *

Nous arrêterons là ce retour amont sur l'histoire de la géométrisation de l'intégrabilité des équations de Pfaff. Nous espérons qu'il aura permis au lecteur de prendre la mesure du profond bouleversement du calcul différentiel que cette géométrisation a engendré. Elle permet de comprendre comment, ainsi que la perception l'atteste partout et toujours de façon éclatante, le cortex visuel peut, grâce au hardware de ses architectures fonctionnelles, faire du calcul intégro-différentiel.

V1 comme groupe de Lie

Nous allons maintenant introduire une nouvelle donnée géométrique fondamentale, à savoir que la structure de contact de $\mathbb{V}_J = J^1\mathbb{R}^2$ peut être reconstruite comme une structure *invariante par translation* dans \mathbb{V}_J considéré comme un *groupe de Lie*.

1. Rendre \mathcal{K} invariante

Nous avons vu à la section 2.4 du chapitre 3 que le groupe de Lie $SE(2)$ agit naturellement sur la distribution \mathcal{K} des plans de contact de \mathbb{V}_J . Nous allons renforcer cette invariance en montrant que \mathcal{K} peut être en fait reconstruit comme le champ de plans obtenu à partir du plan de contact K_0 à l'origine au moyen des *translations* (à gauche) d'une structure de groupe de Lie sur \mathbb{V}_J . Cette structure peut paraître un peu curieuse au prime abord, mais nous verrons qu'elle est en fait l'une des plus universelles en physique, celle du *groupe de Heisenberg*. Nous allons d'abord l'introduire puis nous l'analyserons ensuite en détail dans le cadre de la théorie des groupes et des algèbres de Lie.

Reprenons ce que nous avons vu Vol I, section 5.4.4. et définissons un *produit* dans \mathbb{V}_J par la formule :

$$(x, y, p) \cdot (x', y', p') = (x + x', y + y' + px', p + p') .$$

Il est immédiat de vérifier que cette loi de composition est associative, que l'origine $(0, 0, 0)$ de \mathbb{V}_J est un élément neutre et que tout $v = (x, y, p)$ admet pour inverse $v^{-1} = (-x, -y + px, -p)$. Il s'agit donc bien d'une structure de groupe et, les opérations qui la définissent étant trivialement différentiables, c'est une structure de groupe de Lie (cf. la section suivante 2 pour une définition générale). À cause de l'asymétrie du terme px' , ce produit est *non commutatif*.

Considérons alors la *translation à gauche* L_v de \mathbb{V}_J par $v \in \mathbb{V}_J$ définie par $L_v(v') = v \cdot v'$. C'est un difféomorphisme (non linéaire) de \mathbb{V}_J dont l'application linéaire tangente en 0, T_0L_v , est

$$\begin{aligned} T_0L_v : \quad T_0\mathbb{V}_J &\rightarrow T_v\mathbb{V}_J \\ \mathbf{t} = (\xi, \eta, \pi) &\mapsto T_0L_v(\mathbf{t}) = (\xi, \eta + p\xi, \pi) \end{aligned}$$

On vérifie alors aussitôt qu'en translatant la base $\left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p} \right\}_0$ de $T_0\mathbb{V}_J$ on obtient la base $\left\{ \frac{\partial}{\partial x} + p\frac{\partial}{\partial y}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p} \right\}$ de $T_v\mathbb{V}_J$, c'est-à-dire, avec les notations de la section 2.2 du chapitre 3, $\{\mathbf{t}_1, -\mathbf{t}_3, \mathbf{t}_2\}$. Donc la base $\left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial p} \right\}$ de K_0 est translatée dans la base $\{\mathbf{t}_1, \mathbf{t}_2\}$ de K_v et la structure de contact $\mathcal{K} = \{K_v\}$ n'est rien d'autre que le champ invariant à gauche des plans tangents translatés à gauche de K_0 .

Nous allons maintenant entrer un peu plus dans les détails en introduisant d'abord de façon plus théorique les notions de groupes de Lie et d'algèbres de Lie dont nous avons décrit l'origine historique dans la section 2 du chapitre 4.

2. Groupes et algèbres de Lie (prélude)

2.1. Définition

Un groupe de Lie G est par définition une variété différentiable munie d'une structure de groupe dont les opérations sont des applications différentiables. Dès que l'on quitte les exemples triviaux comme les groupes additifs \mathbb{R}^n , il peut y apparaître des interactions subtiles entre les deux structures respectivement algébrique et différentiable.¹ Par exemple un sous-groupe de Lie H de G sera par définition un sous-groupe algébrique qui sera aussi une sous-variété différentiable. Mais dans des groupes *compacts* certains sous-groupes algébriques seront tellement "repliés" dans G "faute de place" qu'ils ne pourront pas être des sous-variétés. L'exemple standard le plus simple est celui des droites sur le tore. Considérons le plan \mathbb{R}^2 et une droite D_p de pente p . D_p est évidemment un sous-groupe algébrique et une sous-variété et donc un sous-groupe de Lie. Quotientons maintenant \mathbb{R}^2 par le sous-groupe discret \mathbb{Z}^2 des points à coordonnées entières. On obtient le tore $\mathbb{T} = \mathbb{R}^2/\mathbb{Z}^2$ qui est un groupe de Lie compact. Dans \mathbb{T} les projections Δ_p des droites D_p s'enroulent et donnent des sous-groupes algébriques. Si la pente p est rationnelle Δ_p se boucle au bout d'un nombre fini de tours et Δ_p est un sous-groupe de Lie de \mathbb{T} . En revanche, si la pente p est *irrationnelle*, Δ_p ne se boucle jamais, spirale indéfiniment et induit

1. Pour des précisions sur ce qui suit, le lecteur intéressé pourra consulter tout ouvrage comprenant une introduction à la géométrie différentielle, à la théorie des formes différentielles sur les variétés et à la théorie des groupes de Lie. Personnellement, nous apprécions particulièrement les "bibles" de Vladimir Arnold [20] et Ralph Abraham, Jerrold Marsden [2], l'incontournable Spivak [511] ainsi que l'ouvrage d'introduction à la conception d'Élie Cartan de Richard Sharpe [496]. Parmi des textes plus récents, citons par exemple la présentation de Martin Andler [17], le cours de Bernhard Keller [284], ou celui de Jean-Bernard Zuber [574]. Mais il y en a beaucoup d'autres. Il y a aussi d'innombrables "graduate texts". Parmi tant d'autres, citons par exemple les cours de Master de Nicolas Perrin [407] et Jean-François Dat [134]. Nous fournirons nous-mêmes plusieurs éléments de la théorie au fur et à mesure de la progression de notre réflexion.

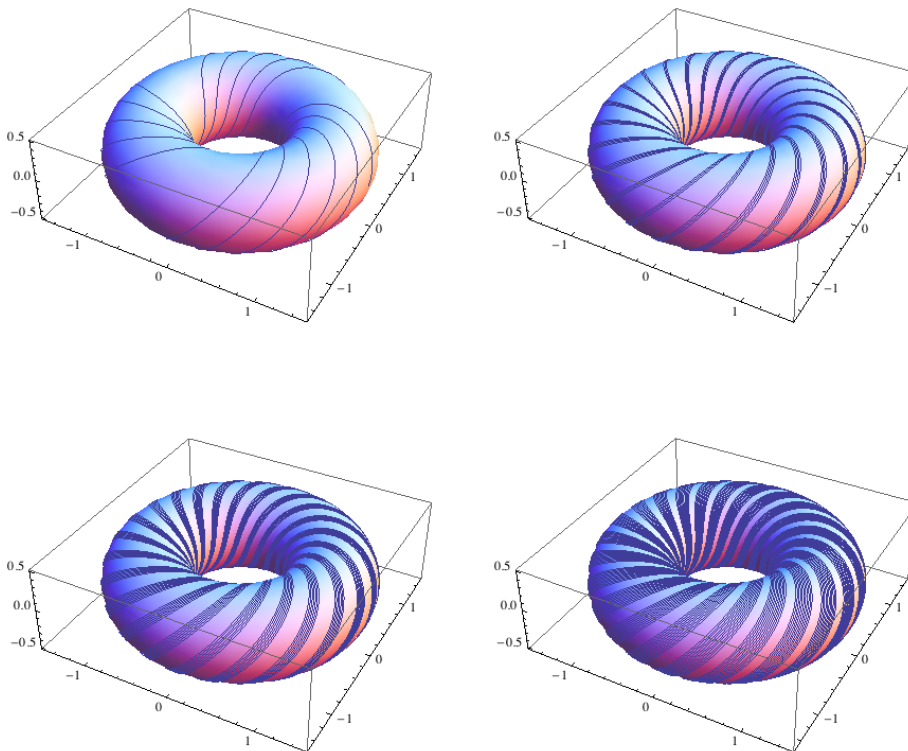


FIGURE 1. Droite de pente irrationnelle Δ_p s'enroulant densément sur un tore \mathbb{T} . La figure montre l'enroulement de segments de Δ_p de plus en plus longs. \mathbb{T} est le tore d'équations paramétriques $(x = (1 + \frac{1}{2} \cos(\psi)) \cos(\varphi), y = (1 + \frac{1}{2} \cos(\psi)) \sin(\varphi), z = \frac{1}{2} \sin(\psi))$ image du carré $\varphi, \psi \in [0, 2\pi)$ et Δ_p est l'image de la droite à pente irrationnelle $\psi = \pi\varphi$. Les effets de moiré sont dus aux problèmes de précision numérique des graphiques.

un sous-groupe algébrique *dense* qui ne peut pas être une sous-variété. Dans ce cas, Δ_p n'est pas un sous-groupe de Lie de \mathbb{T} . Cela est dû au fait que, si $t \in \Delta_p$, tout petit voisinage ouvert V de t dans \mathbb{T} (disons un petit disque $B(t, \varepsilon)$) intersecte Δ_p en une *infinité* de petits ouverts disjoints dont la famille est dense dans V et donc $V \cap \Delta_p$ ne peut jamais se réduire à un simple petit voisinage de t dans Δ_p . Comme le disent les géomètres, Δ_p est "immergée" dans \mathbb{T} mais n'est pas "plongée" dans \mathbb{T} . La figure 1 montre comment, en s'enroulant, une Δ_p se met progressivement à recouvrir \mathbb{T} .

Dans cet exemple, on voit que c'est la densité de Δ_p qui fait problème, autrement dit que la fermeture topologique $\overline{\Delta_p} = \mathbb{T} \neq \Delta_p$ et que donc le sous-groupe algébrique Δ_p n'est pas *topologiquement fermé* dans \mathbb{T} . En fait, c'est la seule obstruction.

On montre en effet qu'un sous-groupe algébrique H d'un groupe de Lie G est un sous-groupe de Lie si et seulement si il est fermé. Qu'un sous-groupe de Lie soit nécessairement fermé se voit facilement. En effet comme H est une sous-variété, en tout point $h \in H$ il existe un voisinage ouvert V de h dans G tel que $V \cap H$ est fermé dans V (c'est ce qui faisait obstruction dans le contre-exemple précédent). Soit alors $g \in \overline{H}$ un point adhérent à H . Pour montrer que H est fermé, il faut montrer que $g \in H$. Soit V_e un voisinage ouvert de e_G (l'élément neutre de G) comme ci-dessus. Le translaté gV_e de V_e est un voisinage ouvert de g et donc, puisque $g \in \overline{H}$, il intersecte H . Soit $h \in gV_e \cap H$. Comme $h \in gV_e$, on a évidemment $g \in hV_e^{-1}$ et donc $g \in hV_e^{-1} \cap \overline{H}$. Mais comme on peut choisir V_e symétrique (i.e. tel que $V_e = V_e^{-1}$), on a $g \in hV_e \cap \overline{H}$. Mais $hV_e \cap \overline{H}$ est l'adhérence de $hV_e \cap H$ dans hV_e et justement $hV_e \cap H$ est fermé dans hV_e par hypothèse. Donc $g \in H$ et H est bien fermé.

La réciproque est beaucoup moins évidente et a fait l'objet d'un théorème d'Élie Cartan approfondi par John von Neumann. Nous y reviendrons à la section 9.

2.2. L'algèbre de Lie sur l'espace tangent $\mathcal{G} = T_e G$

Une conséquence essentielle de la structure de groupe de Lie est que la structure d'algèbre de Lie des champs de vecteurs induit une structure d'algèbre de Lie sur l'espace tangent $\mathcal{G} = T_e G$ de G en l'origine e . Celle-ci exprime la structure de groupe au niveau infinitésimal. En effet, comme G est une variété différentiable, il possède des champs de vecteurs différentiables et nous avons vu à la section 6.3 du chapitre 2 que ces champs peuvent être interprétés comme des opérateurs de dérivation sur les fonctions différentiables (à valeurs réelles) f sur G . Dans la section 6.3.3 de ce même chapitre, nous avons défini le crochet de Lie $[X, Y]$ de deux champs X et Y par l'expression $X.Y - Y.X$, calculé la formule $[X, Y]_i = \sum_j \left(X_j \frac{\partial Y_i}{\partial x_j} - Y_j \frac{\partial X_i}{\partial x_j} \right)$ et montré l'identité de Jacobi

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0 .$$

Les champs de vecteurs X sur G forment donc une algèbre de Lie.

Mais G est en plus un groupe et tout groupe de Lie G est "homogène", c'est-à-dire géométriquement identique en chaque point, car il opère sur lui-même par translations à gauche et à droite. Soit $g \in G$, on lui associe la translation à gauche L_g sur G définie par $L_g : h \mapsto gh$ (nous avons utilisé plus haut les translations L_v de \mathbb{V}_J). Bien qu'inversibles, les translations à gauche ne sont pas pour autant des automorphismes de G car elles ne préservent pas l'élément neutre e . Mais elles sont néanmoins compatibles avec la loi de produit puisque $L_f \circ L_g = L_{fg}$.

La structure de groupe fait donc que, parmi tous les champs X sur G , il y a les champs *invariants* à gauche qui s'obtiennent en translatant un vecteur tangent \mathfrak{t} de \mathcal{G} . Or le commutateur de deux champs invariants est lui-même invariant. En plus de sa structure vectorielle d'espace tangent, \mathcal{G} possède donc la structure d'algèbre

de Lie induite par celle des champs invariants et cette algèbre est de dimension *finie* n égale à la dimension de G comme variété différentiable, alors que l'algèbre des champs de vecteurs sur une variété est de dimension infinie.

Il est essentiel de noter que si G n'est pas commutatif, cette structure d'algèbre de Lie n'est pas triviale et que donc, même si les $\frac{\partial}{\partial x_i}$ engendrent \mathcal{G} comme espace vectoriel, ils ne l'engendrent pas *comme algèbre de Lie*. En effet les champs $\frac{\partial}{\partial x_i}$ ne seront pas invariants en général. Or, insistons-y, la structure d'algèbre de Lie de \mathcal{G} est induite par celle des champs invariants.

2.3. La parallélisabilité

Les translations d'une base de $\mathcal{G} = T_e G$ fournissent un repère *global* (G -invariant) de G et donc une *trivialisat*ion globale canonique du fibré tangent TG . Beaucoup des propriétés géométriques des groupes de Lie découlent de cette propriété. Elle s'exprime en disant que G est une variété "parallélisable". De façon générale une variété différentiable M de dimension n est dite parallélisable s'il existe n champs de vecteurs X_i qui sont *partout* linéairement indépendants.

Il s'agit d'une contrainte forte qui implique, remarquons-le, qu'une variété non parallélisable n'admet aucune structure de groupe compatible avec sa structure différentiable. On voit donc que la compatibilité des structures algébrique et différentiable est vraiment très contraignante.

Par exemple les sphères de dimension paire ne sont pas parallélisables car tout champ de vecteurs s'y annule nécessairement quelque part. Elles n'admettent aucune structure de groupe de Lie. En dimension 2 le théorème est dû à Brouwer (1912) et explique le fameux phénomène de "l'épi" ou de "l'œil de cyclone" : si un champ sur une sphère \mathbb{S}^2 est imagé comme une "chevelure" sur la sphère, alors, quelle que soit la façon dont on "peigne" la sphère, il y aura toujours un épi quelque part ; et de même, s'il y a partout du vent à la surface du globe terrestre, il y a toujours un œil de cyclone quelque part. La généralisation à toutes les sphères de dimension paire est due à Poincaré et Hopf et utilise une formule remarquable de géométrie différentielle. Soit X un champ de vecteurs sur \mathbb{S}^n ne possédant que des zéros isolés x_i (en nombre fini car \mathbb{S}^n est compacte). Si x_0 est un zéro, on peut considérer un petit voisinage U de x_0 ne contenant pas d'autre zéro et regarder l'application $\varphi : U \rightarrow \mathbb{S}^n$, $\varphi_x = \frac{X(x)}{\|X(x)\|}$. Le degré de cette application s'appelle l'index (ou l'indice) de x_0 . La formule dit alors que la somme des index $\sum_i \text{index}(x_i)$ est égal à la caractéristique d'Euler-Poincaré $\chi(\mathbb{S}^n)$ de \mathbb{S}^n (cf. chapitre 15, section 4.7). Et comme celle-ci est 2 en dimension paire, X possède nécessairement des 0.

En revanche, en dimension impaire, il existe des champs partout non nuls. C'est évident par exemple le long d'un cercle \mathbb{S}^1 . Cela est dû au fait que si l'on considère la sphère unité \mathbb{S}^{2m+1} de \mathbb{R}^{2m+2} , on peut utiliser le fait que $\mathbb{R}^{2m+2} = \mathbb{C}^{m+1} = \prod_{k=1}^{m+1} \mathbb{C}_k$

et regrouper les coordonnées en $(m + 1)$ paires (x_k, y_k) correspondant à des variables complexes z_k dans les \mathbb{C}_k . Si l'on considère le vecteur unitaire $z = (z_k)$ de \mathbb{S}^{2m+1} , alors le vecteur unitaire (iz_k) obtenu en tournant les z_k de $\frac{\pi}{2}$ (dans le sens direct) dans le plan \mathbb{C}_k est orthogonal à z et donc, une fois translaté par z , devient identifiable à un vecteur tangent à \mathbb{S}^{2m+1} . D'où un champ de vecteurs tangents qui sont partout non nuls puisque unitaires. Mais cela ne signifie pas pour autant que \mathbb{S}^{2m+1} soit parallélisable.

Un célèbre théorème dû à John Frank Adams en 1962 [3] (cf. le séminaire Bourbaki [378] de Bernard Morin) dit que les seules sphères parallélisables (nécessairement de dimension impaire) sont \mathbb{S}^1 , \mathbb{S}^3 et \mathbb{S}^7 . Cela est dû au fait que, comme nous l'avons déjà vu à la section 8.1.3 du chapitre 2, ce sont les sphères unités de \mathbb{R}^2 , \mathbb{R}^4 et \mathbb{R}^8 qui, avec \mathbb{R} , sont les seuls \mathbb{R}^n admettant une structure de corps, le corps commutatif des nombres complexes \mathbb{C} pour \mathbb{R}^2 , le corps non commutatif des quaternions de Hamilton \mathbb{H}_Q pour \mathbb{R}^4 , le corps non commutatif et non associatif des octonions de Cayley \mathbb{O} pour \mathbb{R}^8 . La structure de groupe de Lie de \mathbb{S}^1 est claire (groupe multiplicatif des complexes de module 1 ou groupe des rotations du plan). Celle de \mathbb{S}^3 vient de son isomorphisme avec le groupe spécial unitaire $SU(2)$ des automorphismes unitaires (donc de déterminant 1) de \mathbb{C}^2 . En effet \mathbb{S}^3 vue comme sphère de \mathbb{C}^2 est d'équation $|u|^2 + |v|^2 = 1$, $u, v \in \mathbb{C}$. On peut ainsi associer à $(u, v) \in \mathbb{S}^3$ la matrice de $SU(2)$ $\begin{pmatrix} u & v \\ -\bar{v} & \bar{u} \end{pmatrix}$ (de déterminant $u\bar{u} + v\bar{v} = |u|^2 + |v|^2 = 1$) et c'est un difféomorphisme permettant de transférer à \mathbb{S}^3 la structure naturelle de groupe de Lie de $SU(2)$. On obtient ainsi la multiplication des quaternions unitaires. La structure des octonions est plus subtile car la non-associativité crée des problèmes.

3. \mathbb{V}_J comme groupe de Heisenberg polarisé \mathbb{H}_{pol}

3.1. La loi de groupe

Revenons maintenant au produit dans \mathbb{V}_J donné par la formule :

$$(x, y, p) \cdot (x', y', p') = (x + x', y + y' + px', p + p')$$

l'origine $(0, 0, 0)$ de \mathbb{V}_J étant l'élément neutre et $v^{-1} = (-x, -y + px, -p)$ l'inverse de $v = (x, y, p)$. À cause de l'asymétrie du terme px' , ce produit – qui est l'addition avec un terme de “couplage” – est *non* commutatif. En fait \mathbb{V}_J est un produit *semi-direct* $\mathbb{V}_J = \mathbb{R}^2 \rtimes \mathbb{R}$, le plan base \mathbb{R}^2 de la fibration $\pi : \mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ constituant le sous-groupe commutatif des translations. Nous allons préciser cette notion plus bas. Il faut noter que cette multiplication non commutative n'est pas distributive par rapport à l'addition standard et ne fait donc pas de \mathbb{V}_J une algèbre au sens

standard. On vérifie en effet que

$$\begin{aligned}
& (x + x', y + y', p + p').(x'', y'', p'') \\
&= (x + x' + x'', y + y' + y'' + (p + p')x'', p + p' + p'') \\
&\neq \\
& (x, y, p).(x'', y'', p'') + (x', y', p').(x'', y'', p'') \\
&= (x + x'', y + y'' + px'', p + p'') + (x' + x'', y' + y'' + p'x'', p' + p'') \\
&= (x + x' + 2x'', y + y' + 2y'' + (p + p')x'', p + p' + 2p'') .
\end{aligned}$$

Il est important de noter que le *centre* Z de \mathbb{V}_J (le sous-groupe des éléments commutant avec tous les autres) est constitué de l'axe des y . En effet (x', y', p') commute avec tous les éléments de \mathbb{V}_J si et seulement si pour tout (x, y, p) on a $px' = p'x$, ce qui implique $x' = p' = 0$. La variable y se trouve ainsi singularisée par une propriété algébrique du groupe. En fait \mathbb{V}_J est une *extension centrale* du groupe additif des (x, p) par le groupe Z des y .

Rappelons que, en termes abstraits, on dit qu'un groupe G est une *extension* de H et K , ou encore que G , H et K sont reliés par une *suite exacte courte*²

$$\{e\} \xrightarrow{\varphi_1} H \xrightarrow{\varphi_2} G \xrightarrow{\varphi_3} K \xrightarrow{\varphi_4} \{e\}$$

($\{e\}$ est le groupe trivial réduit à son élément neutre e), si les φ_i sont des morphismes de groupes et que, à chaque étape, l'image $\text{im}(\varphi_i)$ de φ_i est le noyau $\ker(\varphi_{i+1})$ de φ_{i+1} . Comme $\varphi_1(e)$ est l'élément neutre e_H de H puisque φ_1 est un morphisme, $\ker(\varphi_2) = e_H$ et φ_2 est donc injective. Comme $\ker(\varphi_4) = K$, $\text{im}(\varphi_3) = K$ et φ_3 est donc surjective. Et en plus $\text{im}(\varphi_2) = \ker(\varphi_3)$, ce qui exige que $\text{im}(\varphi_2)$ soit un sous-groupe normal et donc $K \simeq G/H$. Dans notre cas, on a

$$0 \xrightarrow{\varphi_1} Z \xrightarrow{\varphi_2} \mathbb{V}_J \xrightarrow{\varphi_3} \mathbb{V}_J/Z = W \xrightarrow{\varphi_4} 0 .$$

W peut s'identifier à l'ensemble des $(x, 0, p)$, mais ce dernier *n'est pas* un sous-groupe car $(x, 0, p).(x', 0, p') = (x + x', px', p + p')$ n'est pas de la forme $(x'', 0, p'')$. Ce n'est que modulo Z , i.e. modulo les $(0, y, 0)$, que l'on a bien

$$(x, 0, p).(x', 0, p') = (x + x', px', p + p') \equiv (x + x', 0, p + p') \text{ mod } Z .$$

En fait les $(x, 0, p)$ peuvent s'interpréter comme une section "plate" $\sigma_W : W \longrightarrow \mathbb{V}_J$ de la projection $\mathbb{V}_J \xrightarrow{\varphi_3} \mathbb{V}_J/Z = W$ qui sélectionne dans chaque fibre (x, \mathbb{R}, p) de \mathbb{V}_J au-dessus de (x, p) l'élément $(x, 0, p)$. Mais l'image $\sigma_W(W)$ de cette section n'est pas un sous-groupe. Comme on dit, σ_W ne permet pas de "scinder" la suite exacte. Nous reviendrons longuement sur ce type de structure au chapitre 17 quand nous traiterons des représentations de \mathbb{V}_J .

2. Nous avons déjà rencontré la notion de suite exacte dans la section 9 du chapitre 2 à propos de la cohomologie des faisceaux.

L'algèbre de Lie de \mathbb{V}_J est l'espace vectoriel $\mathcal{V}_J = T_0\mathbb{V}_J$ des vecteurs tangents $\mathbf{t} = (\xi, \eta, \pi)$ avec le crochet de Lie

$$[\mathbf{t}, \mathbf{t}'] = [(\xi, \eta, \pi), (\xi', \eta', \pi')] = (0, \pi\xi' - \pi'\xi, 0) .$$

Comme

$$\begin{cases} \mathbf{t}_1 = \frac{\partial}{\partial x} + p\frac{\partial}{\partial y} = (1, p, 0) \\ \mathbf{t}_2 = \frac{\partial}{\partial p} = (0, 0, 1) \\ \mathbf{t}_3 = -\frac{\partial}{\partial y} = (0, -1, 0) \\ [\mathbf{t}_1, \mathbf{t}_2] = (0, -1, 0) = \mathbf{t}_3 \end{cases}$$

la table des commutateurs est

$$\begin{cases} [\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3 \\ [\mathbf{t}_1, \mathbf{t}_3] = 0 \\ [\mathbf{t}_2, \mathbf{t}_3] = 0 \end{cases}$$

Si $\mathbf{t} = (0, \eta, 0)$ alors $[\mathbf{t}, \mathbf{t}'] = 0$ pour tout \mathbf{t}' et \mathbf{t} est dans le *centre* \mathcal{Z} de \mathcal{V}_J .

Remarque. On notera que $\frac{\partial}{\partial x} = \mathbf{t}_1 + p\mathbf{t}_3$ et donc que, dans l'algèbre de Lie \mathcal{V}_J , si l'on a bien $[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}] = [\mathbf{t}_1 + p\mathbf{t}_3, -\mathbf{t}_3] = 0$ et $[\frac{\partial}{\partial y}, \frac{\partial}{\partial p}] = [-\mathbf{t}_3, \mathbf{t}_2] = 0$, on a en revanche $[\frac{\partial}{\partial x}, \frac{\partial}{\partial p}] = [\mathbf{t}_1 + p\mathbf{t}_3, \mathbf{t}_2] = \mathbf{t}_3 \neq 0$, ce qui exprime que le champ $\frac{\partial}{\partial x}$ n'est pas invariant par translations du groupe \mathbb{V}_J . \square

On voit que, en tant que groupe de Lie, \mathbb{V}_J est un groupe *nilpotent* de "niveau" ("step") 2, ce qui signifie que tous les crochets de Lie de la forme $[[\mathbf{t}, \mathbf{t}'], \mathbf{t}'']$ dans \mathcal{V}_J s'annulent. En effet,

$$[[\mathbf{t}, \mathbf{t}'], \mathbf{t}'] = [(0, \pi\xi' - \pi'\xi, 0), (\xi'', \eta'', \pi'')] = (0, 0.\xi'' - \pi''.0, 0) = 0 .$$

Remarque. On peut aussi reconstruire \mathbb{V}_J à partir de la remarque suivante. Considérons les champs de vecteurs ∂_p et $p\partial_y$ sur $\mathbb{R}_{(y,p)}^2$. Il n'y a aucune loi de groupe par rapport à laquelle ces champs sont invariants car le rang décroît en 0. Mais si l'on rajoute la variable x et si on lifte ∂_p en ∂_p et $p\partial_y$ en $\partial_x + p\partial_y$ alors ces champs deviennent invariants par la structure de groupe de \mathbb{V}_J . On parle alors d'un relèvement de \mathbb{R}^2 dans \mathbb{V}_J . C'est une autre façon de voir les choses, susceptible d'une vaste généralisation développée par des auteurs comme Elias Stein, Linda Rothschild ou Gerald Folland. \square

3.2. \mathbb{V}_J comme produit semi-direct $\mathbb{R}^2 \rtimes \mathbb{R}$

Revenons sur le fait que \mathbb{V}_J est un produit *semi-direct* $\mathbb{V}_J = \mathbb{R}^2 \rtimes \mathbb{R}$, le plan base \mathbb{R}^2 de la fibration $\pi : \mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ s'identifiant au sous-groupe additif des translations.³ Ce $\mathbb{R}_{(a)}^2$ des $a = (x, y, 0)$ est le sous-groupe *normal* de \mathbb{V}_J qui est le noyau de la seconde projection $\pi_p : \mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}, (x, y, p) \mapsto p$. C'est un

3. Cf. la section 2.4 du chapitre 3 pour la définition de base d'un produit semi-direct.

sous-groupe puisque $(x, y, 0).(x', y', 0) = (x + x', y + y', 0) \in \mathbb{R}_{(a)}^2$ et $(x, y, 0)^{-1} = (-x, -y, 0) \in \mathbb{R}_{(a)}^2$. Dire que c'est un sous-groupe normal, c'est dire qu'il est stable par conjugaison par n'importe quel élément v de \mathbb{V}_J , autrement dit que pour tout $v = (x', y', p')$, $v.a.v^{-1}$ est un $a'' = (x'', y'', 0) \in \mathbb{R}_{(a)}^2$.⁴ Cela découle du fait qu'il est l'image réciproque du sous-groupe normal $\{0\}$ de \mathbb{R} par π_p qui est un morphisme de groupes, et cela se vérifie immédiatement. En effet,

$$(x', y', p').(x, y, 0).(x', y', p')^{-1} = (x' + x, y' + y, p'x).(-x', -y' + p'x', -p') = (x, y + p'x, 0) .$$

Précisons à ce propos les sous-groupes de \mathbb{V}_J , outre évidemment $\{0\}$ et \mathbb{V}_J lui-même.

1. Les sous-groupes à 1 paramètre (noté λ) de la forme

$$h_\lambda = \left\{ x = \lambda\xi_0, y = \lambda\eta_0 + \frac{1}{2}\lambda^2\xi_0\pi_0, p = \lambda\pi_0 \right\}_\lambda$$

avec $(\xi_0, \eta_0, \pi_0) \in \mathcal{V}_J$. On vérifie en effet immédiatement que $h_\lambda h_{\lambda'} = h_{(\lambda+\lambda')}$ puisque

$$\begin{aligned} & \left(\lambda\xi_0, \lambda\eta_0 + \frac{1}{2}\lambda^2\xi_0\pi_0, \lambda\pi_0 \right) \left(\lambda'\xi_0, \lambda'\eta_0 + \frac{1}{2}\lambda'^2\xi_0\pi_0, \lambda'\pi_0 \right) \\ &= \left(\lambda\xi_0 + \lambda'\xi_0, \lambda\eta_0 + \frac{1}{2}\lambda^2\xi_0\pi_0 + \lambda'\eta_0 + \frac{1}{2}\lambda'^2\xi_0\pi_0 + \lambda\pi_0\lambda'\xi_0, \lambda\pi_0 + \lambda'\pi_0 \right) \\ &= \left((\lambda + \lambda')\xi_0, (\lambda + \lambda')\eta_0 + \left(\frac{1}{2}\lambda^2 + \frac{1}{2}\lambda'^2 + \lambda\lambda' \right) \xi_0\pi_0, (\lambda + \lambda')\pi_0 \right) \\ &= \left((\lambda + \lambda')\xi_0, (\lambda + \lambda')\eta_0 + \frac{1}{2}(\lambda + \lambda')^2\xi_0\pi_0, (\lambda + \lambda')\pi_0 \right) ; \end{aligned}$$

Ce sont les exponentiations $\exp(\lambda((\xi_0, \eta_0, \pi_0)))$ qui intègrent les champs de vecteurs invariants, exponentielle que nous définirons rigoureusement plus bas à la section 9. Le sous-groupe est normal s'il exponentie le centre \mathcal{Z} de \mathcal{V}_J .

2. Les sous-groupes $\exp(\mathcal{W})$ où \mathcal{W} est un plan qui est une sous-algèbre de Lie de \mathcal{V}_J qui doit contenir \mathcal{Z} . Soient $W_1 = (\xi_1, \eta_1, \pi_1)$ et $W_2 = (\xi_2, \eta_2, \pi_2)$ des générateurs du plan \mathcal{W} . Ils ne sont pas colinéaires. Pour que

$$[W_1, W_2] = (0, \pi_1\xi_2 - \pi_2\xi_1, 0)$$

4. Rappelons que pour un groupe G abélien (commutatif) tous les sous-groupes sont normaux car $v.a.v^{-1} = v.v^{-1}.a = a$.

appartienne à \mathcal{W} , il faut soit que $\pi_1\xi_2 - \pi_2\xi_1 = 0$ et alors $[W_1, W_2] = 0$, soit qu'il existe $\alpha_1, \alpha_2 \in \mathbb{R}$ tels que $[W_1, W_2] = \alpha_1 W_1 + \alpha_2 W_2$, i.e.

$$\begin{cases} \alpha_1\xi_1 + \alpha_2\xi_2 = 0 \\ \alpha_1\pi_1 + \alpha_2\pi_2 = 0 \\ \alpha_1\eta_1 + \alpha_2\eta_2 = \pi_1\xi_2 - \pi_2\xi_1 \end{cases}$$

Si l'on veut que le système linéaire des deux premières équations

$$\begin{pmatrix} \xi_1 & \xi_2 \\ \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = 0$$

ait une solution non triviale, il faut que $\det \begin{pmatrix} \xi_1 & \xi_2 \\ \pi_1 & \pi_2 \end{pmatrix} = \xi_1\pi_2 - \pi_1\xi_2 = 0$ et donc $[W_1, W_2] = 0$. Considérons alors le vecteur de \mathcal{W}

$$\begin{aligned} \pi_2 W_1 - \pi_1 W_2 &= (\pi_2\xi_1, \pi_2\eta_1, \pi_2\pi_1) - (\pi_1\xi_2, \pi_1\eta_2, \pi_1\pi_2) \\ &= (\pi_2\xi_1 - \pi_1\xi_2, \pi_2\eta_1 - \pi_1\eta_2, \pi_2\pi_1 - \pi_1\pi_2) \\ &= (0, \pi_2\eta_1 - \pi_1\eta_2, 0) . \end{aligned}$$

Il est de la forme $(0, \eta, 0)$ et donc si $\eta \neq 0$, $\mathcal{Z} \subset \mathcal{W}$. Si $\eta = 0$, on considère le vecteur

$$\xi_2 W_1 - \xi_1 W_2 = (0, \xi_2\eta_1 - \xi_1\eta_2, 0)$$

qui est de la forme $(0, \eta', 0)$ et donc si $\eta' \neq 0$, $\mathcal{Z} \subset \mathcal{W}$. Mais on ne peut pas avoir à la fois $\eta = 0$ et $\eta' = 0$, car sinon le produit extérieur $W_1 \wedge W_2$ de W_1 et W_2 qui est de composantes

$$W_1 \wedge W_2 = (\pi_2\eta_1 - \pi_1\eta_2, \pi_1\xi_2 - \xi_1\pi_2, \xi_1\eta_2 - \xi_2\eta_1)$$

serait nul et W_1 et W_2 seraient colinéaires ce qui est faux par hypothèse.

Notons maintenant $\mathbb{R}_{(p)}$ le \mathbb{R} des $(0, 0, p)$. C'est un sous-groupe mais ce n'est pas un sous-groupe normal. On constate que, d'un côté, l'on a $(x, y, 0) \cdot (0, 0, p) = (x, y, p)$ mais que, d'un autre côté, on peut faire opérer $\mathbb{R}_{(p)}$ sur $\mathbb{R}_{(a)}^2$ par l'automorphisme de conjugaison

$$p(a) = (0, 0, p) \cdot (x, y, 0) \cdot (0, 0, p)^{-1} = (x, y + px, 0) .$$

Si $a = (x, y, 0)$ est identifié à $a = (x, y)$, par $p(a) = (x, y + px)$, on voit que le produit de \mathbb{V}_J est donné par

$$\begin{aligned} (x, y, p) \cdot (x', y', p') &= (x + x', y + y' + px', p + p') \\ &= ((x, y) + p(x', y'), p + p') . \end{aligned}$$

L'action $p(\bullet)$ est un morphisme de groupes du sous-groupe $\mathbb{R}_{(p)}$ dans le groupe $\text{Aut}(\mathbb{R}_{(a)}^2)$ des automorphismes de $\mathbb{R}_{(a)}^2$.⁵ C'est elle qui introduit une asymétrie entre les variables x et y .

La définition générale d'un produit semi-direct $G = H \rtimes K$ est que

- (i) H est un sous-groupe normal,
- (ii) K est un autre sous-groupe (en général non normal) tel que $K \cap H = \{e\}$ (l'élément neutre de G) et identifiable au groupe quotient G/H (qui est bien défini parce que H est normal⁶),
- (iii) $G = H.K$, i.e. tout élément g de G s'écrit de façon unique comme le produit $g = h.k$ d'un élément h de H et d'un élément k de K , ce qui fait que G est le produit direct $G = H \times K$ en tant qu'ensemble (pas en tant que groupe),
- (iv) si on fait agir K par les automorphismes de conjugaison sur H au moyen de $k(h') = kh'k^{-1}$, alors le produit de G est donné par $(h, k) \cdot (h', k') = (h.k(h'), k.k')$.

C'est exactement ce qui se passe ici avec $H = \mathbb{R}_{(a)}^2$ et $K = \mathbb{R}_{(p)}$. \mathbb{V}_J est donc une extension, autrement dit une *suite exacte courte*

$$0 \xrightarrow{\varphi_1} H = \mathbb{R}_{(a)}^2 \xrightarrow{\varphi_2} G = \mathbb{V}_J \xrightarrow{\varphi_3} K = \mathbb{R}_{(p)} \xrightarrow{\varphi_4} 0$$

le quotient $K = \mathbb{R}_{(p)}$ étant lui même identifiable à un sous-groupe (non normal), celui des $(0, 0, p)$.

Remarque. Comme nous avons vu plus haut que \mathbb{V}_J est aussi une extension centrale de son centre $Z : 0 \longrightarrow Z \longrightarrow \mathbb{V}_J \longrightarrow \mathbb{V}_J/Z \longrightarrow 0$, nous constatons que \mathbb{V}_J est doublement une extension, en accord avec sa structure de produit. \square

Remarque. La structure de produit semi-direct éclaire ce qui se passe dans le passage entre le plan de base $\mathbb{R}_{(a)}^2$ et le fibré \mathbb{V}_J que l'on peut identifier à la fibration $\pi : \mathbb{V}_J = \mathbb{R}_{(a)}^2 \times \mathbb{R}_{(p)} \rightarrow \mathbb{R}_{(a)}^2$. \mathbb{V}_J est une extension du plan de base $\mathbb{R}_{(a)}^2$, et l'on peut dire en termes de neurogéométrie que l'évolution a "inventé" une extension *corticale* du plan *rétinien*. \square

3.3. La condition d'Hörmander

L'espace des champs invariants (par translation à gauche) est engendré comme algèbre de Lie de dimension 3 par la base naturelle

$$\left\{ \mathbf{t}_1 = \frac{\partial}{\partial x} + p \frac{\partial}{\partial y} = (1, p, 0), \mathbf{t}_2 = \frac{\partial}{\partial p} = (0, 0, 1) \right\}$$

5. En effet, $p(0) = 0$, $p(-a) = -p(a)$, $p(a + a') = p(a) + p(a')$, $(-p)(a) = (x, y - px, 0)$ et $(-p) \circ p(a) = (x, y + px - px, 0) = a$.

6. Cf. la définition de la normalité dans la section 1.5 du chapitre 4 consacrée à Galois.

de l'espace vectoriel de dimension 2 qu'est le plan de contact K_v . En effet, comme nous l'avons vu, le crochet $[\mathbf{t}_1, \mathbf{t}_2]$ est

$$[\mathbf{t}_1, \mathbf{t}_2] = (0, -1, 0) = \mathbf{t}_3 = -\frac{\partial}{\partial y}$$

et tous les autres crochets s'annulent.⁷ Il s'agit là d'un fait fondamental. La base $\{\mathbf{t}_1, \mathbf{t}_2\}$ de la distribution \mathcal{K} est, comme on dit, "bracket generating", c'est-à-dire engendre au sens des algèbres de Lie *tout* le fibré tangent $T\mathbb{V}_J$. Cette condition s'appelle la *condition d'Hörmander*.⁸ Elle est la propriété cruciale conduisant à généraliser en dimensions supérieures et à des variétés différentiables générales notre structure de contact élémentaire \mathbb{V}_J .

3.4. Représentations matricielles

Une remarque très utile pour les calculs est que l'algèbre de Lie \mathcal{V}_J peut être représentée de façon fort commode par celle des matrices nilpotentes

$$\mathbf{t} \Leftrightarrow m(\mathbf{t}) = m(\xi, \eta, \pi) = \begin{pmatrix} 0 & \pi & \eta \\ 0 & 0 & \xi \\ 0 & 0 & 0 \end{pmatrix}.$$

Il y a bien nilpotence puisque

$$m^2(\xi, \eta, \pi) = \begin{pmatrix} 0 & \pi & \eta \\ 0 & 0 & \xi \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & \pi & \eta \\ 0 & 0 & \xi \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \pi\xi \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

et $m^3 = 0$. Ces matrices $m(\mathbf{t})$ triangulaires supérieures de diagonale nulle forment un sous-espace vectoriel, une sous-algèbre et une sous-algèbre de Lie de l'algèbre $M_3(\mathbb{R})$ des matrices carrées 3×3 à coefficients dans \mathbb{R} . L'algèbre $M_3(\mathbb{R})$ est un espace vectoriel de dimension 9 sur \mathbb{R} muni en plus d'une multiplication compatible (la multiplication des matrices) et d'une structure l'algèbre de Lie dont le crochet $[A, B]$ est le commutateur $AB - BA$. L'application m est une *représentation* de l'algèbre de Lie \mathcal{V}_J dans $M_3(\mathbb{R})$, un isomorphisme d'algèbres de Lie sur son image $m(\mathcal{V}_J)$ qui est une sous-algèbre de Lie de dimension 3 de $M_3(\mathbb{R})$.

On notera que la nilpotence des $m(\xi, \eta, \pi)$ fait intervenir la structure *multiplicative* de $M_3(\mathbb{R})$ qui n'a pas de sens pour les \mathbf{t} ainsi représentés, seul le crochet de Lie $[\mathbf{t}, \mathbf{t}']$ ayant un sens pour ces vecteurs \mathbf{t} .

7. Soulignons encore que, à l'origine $v = 0$, on a $t_1 = \frac{\partial}{\partial x}$, $t_2 = \frac{\partial}{\partial p}$ mais cela ne veut pas dire pour autant que l'on a $[t_1, t_2] = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial p} \right] = 0$ car le champ $\frac{\partial}{\partial x}$ n'est pas invariant et, rappelons-le, les commutateurs sont ceux des champs invariants.

8. Cf. le monumental traité de Lars Hörmander (1931-2012, médaille Fields 1962) *The analysis of linear partial differential operators* (1983-1985).

On vérifie que l'on a bien

$$\begin{aligned} [m(\mathbf{t}), m(\mathbf{t}')] &= \begin{pmatrix} 0 & \pi & \eta \\ 0 & 0 & \xi \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & \pi' & \eta' \\ 0 & 0 & \xi' \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & \pi' & \eta' \\ 0 & 0 & \xi' \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & \pi & \eta \\ 0 & 0 & \xi \\ 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & \pi\xi' \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & \pi'\xi \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \pi\xi' - \pi'\xi \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &= m([\mathbf{t}, \mathbf{t}']) . \end{aligned}$$

On vérifie aussi que les doubles commutateurs s'annulent tous :

$$m[[\mathbf{t}, \mathbf{t}'], \mathbf{t}'] = m([\mathbf{t}, \mathbf{t}']) \cdot m(\mathbf{t}'') - m(\mathbf{t}'') \cdot m([\mathbf{t}, \mathbf{t}']) = 0 .$$

Les éléments $v = (x, y, p)$ du groupe \mathbb{V}_J sont de la même façon représentés par les matrices

$$v \Leftrightarrow M(v) = M(x, y, p) = \begin{pmatrix} 1 & p & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} = I + m(x, y, p).$$

(I étant la matrice unité) le produit $v.v'$ devenant le produit de matrices $M(v.v') = M(v) \cdot M(v')$. Les matrices triangulaires supérieures $M(v)$ de diagonale $(1, 1, 1)$ forment un sous-groupe multiplicatif du groupe linéaire $GL_3(\mathbb{R})$ (et même du groupe spécial linéaire $SL_3(\mathbb{R})$ puisqu'elles sont de déterminant 1) mais pas un sous-groupe additif (car dans $M(v) + M(v')$ la diagonale devient $(2, 2, 2)$, en fait $M(v) + M(v') = M(v + v') + I$).

M est un isomorphisme de groupes multiplicatifs de \mathbb{V}_J sur son image $M(\mathbb{V}_J) = \mathfrak{M}$ qui est un sous-groupe de dimension 3 (en tant que sous-variété) de $GL_3(\mathbb{R})$. On remarquera que $m(\xi, \eta, \pi)$ consiste à dériver une trajectoire $M(x(s), y(s), p(s))$ par rapport à une variable s et à écrire $\xi = x'(s)$, $\eta = y'(s)$ et $\pi = p'(s)$. Les $m(\xi, \eta, \pi)$ sont donc les vecteurs tangents des M . Le fait que $M(x, y, p) = I + m(x, y, p)$ signifie que, en tant qu'espace affine isomorphe à \mathbb{R}^3 , l'espace $\mathfrak{M} \simeq \mathbb{V}_J$ des M satisfait $T_v \mathfrak{M} \simeq \mathfrak{M}$. Étant donné cet isomorphisme, on pourra parfois, même si les variables (ξ, η, π) concernent les vecteurs tangents \mathbf{t} et les variables (x, y, p) les points v , écrire $m(x, y, p)$.

Remarque. Nous ne ferons plus vraiment de différence entre \mathbf{t} et $m(\mathbf{t})$ lorsque le contexte sera clair. □

L'avantage de ces représentations de \mathbb{V}_J et \mathcal{V}_J est qu'elles se font dans la même algèbre de Lie $M_3(\mathbb{R})$ qui contient à la fois le groupe de Lie multiplicatif $GL_3(\mathbb{R})$ et les sous-algèbres de Lie dont on a besoin. Cela simplifie beaucoup les choses.

Remarque. À la section 4.1 du chapitre 2, nous avons étendu la forme de contact $\omega_J = dy - p dx$ sur \mathbb{R}^3 à la structure d'Engel sur \mathbb{R}^4 en adjoignant une variable supplémentaire (la courbure) κ et la 1-forme $\tau_J = dp - \kappa dx$. On vérifie que cette structure est *invariante* par la loi de groupe :

$$(x, y, p, \kappa) (x', y', p', \kappa') = \left(x + x', y + y' + px', p + p', \kappa + \kappa' + yx' + \frac{1}{2}px'^2 \right) .$$

La loi est bien associative, l'élément neutre est le 0 et l'inverse est

$$\left(-x, -y + px, -p, -\kappa + yx - \frac{1}{2}px^2 \right)$$

La structure d'Engel est donc également un groupe de Lie, le groupe d'Engel $J^2(\mathbb{R}, \mathbb{R})$

des matrices réelles 4×4 $\begin{pmatrix} 1 & p & y & \kappa \\ 0 & 1 & x & \frac{x^2}{2} \\ 0 & 0 & 1 & x \\ 0 & 0 & 0 & 0 \end{pmatrix}$ qui généralise $\mathbb{H}_{pol} \simeq \mathbb{V}_J = J^1(\mathbb{R}, \mathbb{R})$. \square

4. Algèbres de Lie nilpotentes VS semi-simples (introduction)

Nous venons de voir des exemples concrets très simples de groupes et algèbres de Lie. Il est bon de donner tout de suite quelques précisions générales supplémentaires. Si \mathcal{G} est une algèbre de Lie, ce sont les commutateurs $[X, Y]$ qui contiennent toute la structure et ce que l'on appelle la "représentation adjointe" $\text{ad}_X(Y) = [X, Y]$ de \mathcal{G} dans $\text{End}(\mathcal{G})$ (représentation sur laquelle nous allons revenir en détail à la section 6) code une information essentielle.⁹ Il faut souligner que ad_X est une *dérivation* de \mathcal{G} car, pour le produit qu'est le crochet de Lie $[\cdot, \cdot]$, il satisfait la règle de Leibniz

$$\text{ad}_X([Y, Z]) = [\text{ad}_X(Y), Z] + [Y, \text{ad}_X(Z)]$$

qui n'est rien d'autre que l'identité de Jacobi

$$[X, [Y, Z]] = [[X, Y], Z] + [Y, [X, Z]] .$$

Ces dérivations sont dites *internes*. Quant à $\text{ad} : X \mapsto \text{ad}_X$, c'est un morphisme d'algèbres de Lie car on vérifie immédiatement que $\text{ad}_{[X, Z]} = [\text{ad}_X, \text{ad}_Z]$. Ces propriétés reformulent pour les champs G -invariants ce que nous avons vu à la section 6.3.3 du chapitre 2 à propos des champs de vecteurs quelconques comme dérivées de Lie.

9. $\text{End}(\mathcal{G})$ est l'algèbre des endomorphismes de \mathcal{G} .

Comme pour les anneaux, il est alors utile d'introduire la notion d'*idéal*¹⁰ : un sous-espace \mathcal{I} de \mathcal{G} est un idéal si pour tout $Y \in \mathcal{I}$ et tout $X \in \mathcal{G}$, $[X, Y] \in \mathcal{I}$. $\{0\}$ et \mathcal{G} lui-même sont évidemment toujours des idéaux et l'identité de Jacobi implique que $\mathcal{D}(\mathcal{G}) = [\mathcal{G}, \mathcal{G}]$ est un idéal puisque si $[Y, Z] \in [\mathcal{G}, \mathcal{G}]$ et $X \in \mathcal{G}$,

$$[X, [Y, Z]] = -[Y, [Z, X]] - [Z, [X, Y]] \in [\mathcal{G}, \mathcal{G}] .$$

On l'appelle l'idéal *dérivé* de \mathcal{G} . Si \mathcal{I} est un idéal on peut alors quotienter \mathcal{G} par \mathcal{I} .¹¹ Par exemple $\mathcal{D}(\mathcal{G})$ est le plus petit idéal tel que l'algèbre quotient $\mathcal{G}/\mathcal{D}(\mathcal{G})$ soit abélienne. Le centre \mathcal{Z} de \mathcal{G} composé des Z commutant avec tous les X de \mathcal{G} est trivialement un idéal (puisque $0 \in \mathcal{Z}$). Le noyau de tout morphisme d'algèbre de Lie $\varphi : \mathcal{G} \rightarrow \mathcal{K}$ est également trivialement un idéal.¹²

Si \mathcal{I} est un idéal, $[\mathcal{G}, \mathcal{I}]$ est aussi un idéal car si $U = [X, Y] \in [\mathcal{G}, \mathcal{I}]$ ($X \in \mathcal{G}$ et $Y \in \mathcal{I}$) et $V \in \mathcal{G}$, alors, par Jacobi,

$$[V, U] = [V, [X, Y]] = [V, [X, Y]] - [X, [V, Y]]$$

mais $[X, Y]$ et $[V, Y]$ sont dans \mathcal{I} car \mathcal{I} est un idéal et donc $[V, U] \in [\mathcal{G}, \mathcal{I}]$. On peut alors construire par récurrence une suite d'idéaux $\mathcal{G}_{(j)}$ définis par

$$\mathcal{G}_{(0)} = \mathcal{G}, \quad \mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}_{(0)}] = [\mathcal{G}, \mathcal{G}], \quad \mathcal{G}_{(2)} = [\mathcal{G}, \mathcal{G}_{(1)}], \dots, \quad \mathcal{G}_{(j)} = [\mathcal{G}, \mathcal{G}_{(j-1)}], \text{ etc.}$$

$\mathcal{G}_{(j)}$ est l'idéal de j commutateurs successifs emboîtés. Dire que $\mathcal{G}_{(1)} = \{0\}$, c'est dire que \mathcal{G} est commutative (i.e. abélienne).

On dit que \mathcal{G} est *simple* si les seuls idéaux sont $\{0\}$ et \mathcal{G} . C'est l'opposé de \mathcal{G} abélienne où $[\mathcal{G}, \mathcal{G}] = \{0\}$ et où *tout* sous-espace est un idéal. Une algèbre simple est maximale non commutative puisque $[\mathcal{G}, \mathcal{G}] = \mathcal{G}$. Pour une algèbre abélienne $ad_X \equiv 0$ pour tout X et donc $ad : \mathcal{G} \rightarrow \text{End}(\mathcal{G})$ est $\equiv 0$. Au contraire, pour une algèbre simple, ad est injective car son noyau est l'idéal égal au centre \mathcal{Z} de \mathcal{G} et celui-ci est réduit à $\{0\}$. On dit plus généralement que \mathcal{G} est *semi-simple* si elle est une somme directe d'idéaux simples, cela étant équivalent au fait que le seul idéal *abélien* soit $\{0\}$.

Le prototype des algèbres simples sont les algèbres de Lie $\mathfrak{so}(n)$ des groupes de rotation $SO(n)$ pour $n \neq 4$ ($\mathfrak{so}(4) = \mathfrak{so}(3) \oplus \mathfrak{so}(3)$ n'est que semi-simple) et leur structure est d'une richesse fascinante.

Les algèbres de Lie les plus proches des algèbres abéliennes sont les algèbres *nilpotentes* où lorsque l'on itère la formation des commutateurs on arrive toujours

10. Rappelons que dans un anneau A , un sous-groupe additif I est un idéal (à gauche) si $b \in I$ et $a \in A$ impliquent $a.b \in I$.

11. On prend le quotient \mathcal{G}/\mathcal{I} des espaces vectoriels et on vérifie que les commutateurs passent au quotient : si $X, Y \in \mathcal{G}$ et si $X', Y' \in \mathcal{I}$ (i.e. si $X + X' \equiv X$ modulo \mathcal{I} et $Y + Y' \equiv Y$ modulo \mathcal{I}), alors $[X + X', Y + Y'] = [X, Y] + ([X, Y'] + [X', Y] + [X', Y'])$, les trois crochets entre parenthèses sont dans \mathcal{I} car \mathcal{I} est un idéal, et donc $[X + X', Y + Y'] \equiv [X, Y]$ modulo \mathcal{I} .

12. Si $Y \in \ker(\varphi)$ (i.e. $\varphi(Y) = 0$) et si $X \in \mathcal{G}$, alors $[X, Y] \in \ker(\varphi)$ puisque $\varphi([X, Y]) = [\varphi(X), \varphi(Y)] = [\varphi(X), 0] = 0$.

à 0 au bout d'un nombre fini de pas. Cela signifie que pour tout X , ad_X n'est pas forcément nulle mais a une puissance $k(X)$ nulle : $(ad_X)^{k(X)} = 0$. Cela est aussi équivalent au fait que la suite des idéaux $\mathcal{G}_{(j)}$ stationne en $\{0\}$ à partir d'un certain rang fini, ce rang étant appelé le "niveau" ("step") de \mathcal{G} . Tous les sous-espaces ne sont pas des idéaux comme dans le cas abélien mais il y a néanmoins "beaucoup" d'idéaux. Un théorème de structure fondamental, sur lequel nous reviendrons à la section 1 du chapitre 7, dit :

Théorème de structure. Il existe une suite croissante d'idéaux \mathcal{G}_j ¹³ de dimension exactement $j = 0, 1, \dots, n = \dim \mathcal{G}$ tels que $[\mathcal{G}, \mathcal{G}_{j+1}] \subset \mathcal{G}_j$. \diamond

En particulier $[\mathcal{G}, \mathcal{G}_1] = \mathcal{G}_0 = \{0\}$ (car $\dim \mathcal{G}_0 = 0$) et donc \mathcal{G} et \mathcal{G}_1 commutent et \mathcal{G}_1 est inclus dans le centre Z de \mathcal{G} . Celui-ci n'est donc pas nul car il est de dimension $\geq \dim \mathcal{G}_1 = 1$.

Les algèbres nilpotentes non commutatives les plus simples sont celle de niveau 2 satisfaisant $\mathcal{G}_{(0)} = \mathcal{G}$, $\mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}_{(0)}] \neq 0$ et $\mathcal{G}_{(2)} = [\mathcal{G}, \mathcal{G}_{(1)}] = 0$. C'est précisément le cas de l'algèbre de Heisenberg.

5. Le groupe de Heisenberg non polarisé \mathbb{H}

5.1. L'exponentielle du groupe

Nous avons vu comment l'on passe d'un groupe de Lie G à son algèbre de Lie \mathcal{G} qui en est sa version infinitésimale. Mais il existe dans l'autre sens la possibilité de passer d'une algèbre de Lie à des groupes de Lie G qui en sont des versions intégrales obtenues par exponentiation. Le cas le plus élémentaire est celui, commutatif, où \mathcal{G} est le groupe additif des $x \in \mathbb{R}$ et G le groupe multiplicatif des e^x . Les opérations se correspondent bien puisque $e^{x+y} = e^x e^y$, le passage de G à \mathcal{G} s'effectuant par la fonction logarithme inverse de la fonction exponentielle puisque $\log(e^x) = x$. Étant donnée \mathcal{G} , il y a a priori plusieurs possibilités pour G car \mathcal{G} ne peut déterminer G que localement au voisinage de e . Mais un théorème fondamental montre que, à isomorphisme près, il n'existe qu'un seul groupe de Lie *simplement connexe* $G^{s.c.}$ d'algèbre de Lie \mathcal{G} . $G^{s.c.}$ est le revêtement universel de tous les G intégrant \mathcal{G} .

Nous reviendrons plus en détail sur l'exponentielle à la section 9. Nous ne considérons ici que notre exemple simple où les $\mathfrak{t} = (\xi, \eta, \pi) \in \mathcal{V}_J$ sont représentés par des matrices $m(\mathfrak{t}) = m(\xi, \eta, \pi)$. Comme $m(\mathfrak{t})$ est une matrice, son exponentielle est donnée par le développement en série standard

$$\exp(x) = 1 + x + \frac{x^2}{2} + \dots + \frac{x^n}{n!} + \dots$$

13. Les \mathcal{G}_j ne sont pas les $\mathcal{G}_{(j)}$.

Mais comme les matrices $m(\mathbf{t})$ sont nilpotentes avec $m^3 = 0$, le développement de l'exponentielle s'arrête en fait à l'ordre 2 et l'on obtient la formule :

$$\begin{aligned} \exp(m(\xi, \eta, \pi)) &= I + m(\xi, \eta, \pi) + \frac{1}{2}m(0, \pi\xi, 0) \\ &= M\left(\xi, \eta + \frac{1}{2}\pi\xi, \pi\right). \end{aligned}$$

Alors que les $M(x, y, p) = I + m(x, y, p)$ donnent $\mathbb{V}_J \simeq \mathbb{H}_{pol}$, les $\exp(m(\mathbf{t}))$ donnent un autre groupe \mathbb{H} qui est une *variante* de \mathbb{H}_{pol} . Par abus de notation, notons simplement $\mathbf{t} = (\xi, \eta, \pi)$ la matrice $m(\mathbf{t}) = m(\xi, \eta, \pi)$ et h l'exponentielle $\exp(\mathbf{t})$. La loi de multiplication $h.h' = \exp(\mathbf{t}) \cdot \exp(\mathbf{t}')$ est, à cause de la non-commutativité et de la forme des commutateurs, donnée par la *formule de Baker-Campbell-Hausdorff*

$$\exp(\mathbf{t}) \cdot \exp(\mathbf{t}') = \exp\left(\mathbf{t} + \mathbf{t}' + \frac{1}{2}[\mathbf{t}, \mathbf{t}']\right).$$

On vérifie en effet que l'on a bien

$$\exp(\mathbf{t}) \cdot \exp(\mathbf{t}') = M\left(\xi + \xi', \eta' + \frac{1}{2}\pi'\xi' + \pi\xi' + \eta + \frac{1}{2}\pi\xi, \pi + \pi'\right)$$

(le terme asymétrique $\pi\xi'$ marque la non commutativité) et

$$\begin{aligned} \exp\left(\mathbf{t} + \mathbf{t}' + \frac{1}{2}[\mathbf{t}, \mathbf{t}']\right) &= \exp\left(\left(\xi + \xi', \eta + \eta' + \frac{1}{2}(\xi'\pi - \xi\pi'), \pi + \pi'\right)\right) \\ &= M\left(\xi + \xi', \eta + \eta' + \frac{1}{2}(\xi'\pi - \xi\pi') + \frac{1}{2}(\xi + \xi')(\pi + \pi'), \pi + \pi'\right). \end{aligned}$$

Or

$$\eta + \eta' + \frac{1}{2}(\xi'\pi - \xi\pi') + \frac{1}{2}(\xi + \xi')(\pi + \pi') = \eta' + \frac{1}{2}\pi'\xi' + \pi\xi' + \eta + \frac{1}{2}\pi\xi.$$

La loi de groupe de \mathbb{H} est donc, réécrite en termes des variables (x, y, p) ,

$$(x, y, p) \cdot (x', y', p') = \left(x + x', y + y' + \frac{1}{2}(px' - xp'), p + p'\right)$$

qui est une version *symétrisée* de la loi de groupe de $\mathbb{V}_J \simeq \mathbb{H}_{pol}$. Il s'agit du *groupe de Heisenberg*, sur lequel nous reviendrons longuement au chapitre 14. C'est à cause de cette relation avec \mathbb{H} , que le groupe \mathbb{H}_{pol} qui asymétrise la relation entre x et p s'appelle le groupe de Heisenberg *polarisé*. Le centre Z des deux groupes $\mathbb{V}_J \simeq \mathbb{H}_{pol}$ et \mathbb{H} (les éléments qui commutent avec tous les autres) est le même, l'axe des y (cf. plus haut). La symétrisation rend les opérations plus simples car $0 = (0, 0, 0)$ devient l'élément neutre et l'inverse de (x, y, p) devient tout simplement l'inverse pour l'addition $(-x, -y, -p)$.

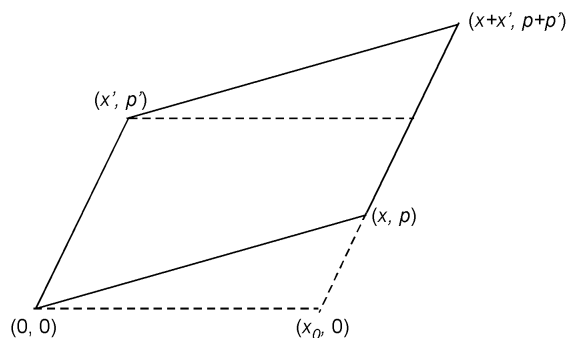


FIGURE 2. Illustration du terme d'aire dans le produit du groupe de Heisenberg.

$\mathbb{V}_J \simeq \mathbb{H}_{pol}$ et \mathbb{H} ayant la même algèbre de Lie et étant simplement connexes, ils doivent être isomorphes comme nous l'avons expliqué plus haut, et c'est bien le cas. L'isomorphisme $x = \xi$, $y = \eta + \frac{1}{2}\pi\xi$, $p = \pi$ transforme \mathbb{H} en \mathbb{H}_{pol} comme nous allons le préciser à la section 5.3.

Remarque. Dans la littérature, on trouve souvent le groupe de Heisenberg écrit un peu différemment, à un facteur 4 près, avec la loi :

$$(x, y, p) \cdot (x', y', p') = (x + x', y + y' + 2(px' - xp'), p + p') \cdot \square$$

5.2. \mathbb{H} et le problème isopérimétrique

On remarquera que le terme $\frac{1}{2}(-px' + xp')$ est l'aire du triangle

$$\Delta = \{(0, 0), (x, p), (x + x', p + p')\}$$

dans le plan (x, p) . En effet, la droite passant par les points (x, p) et $(x + x', p + p')$ est d'équation $P = \frac{p'}{x'}X + \frac{px' - xp'}{x'}$ et coupe donc l'axe des x en $(x_0 = \frac{-px' + xp'}{p'}, 0)$. Comme le montre la figure 2, l'aire du parallélogramme

$$\Pi = \{(0, 0), (x, p), (x + x', p + p'), (x', p')\}$$

qui est le double de celle de Δ est aussi celle du parallélogramme de côtés

$$\{(0, 0), (0, x_0)\} \text{ et } \{(0, 0), (x', p')\}.$$

Or celui-ci est de hauteur p' et son aire est donc

$$x_0 p' = \frac{-px' + xp'}{p'} \cdot p' = -px' + xp'.$$

C'est pourquoi le groupe de Heisenberg est étroitement lié au problème *isopérimétrique*¹⁴ sur lequel nous reviendrons en détail dans le chapitre 9 consacré au calcul des variations. Il s'agit, étant donnée une aire A bordée par une courbe fermée C de longueur L , d'un côté de maximiser l'aire A à périmètre L fixé et, d'un autre côté et dualement, de minimiser le périmètre L à aire A fixée.

Pour \mathbb{V}_J , la 1-forme de contact est $\omega_J = dy - p dx$ et sa nullité exprime que la variation de la troisième variable y le long d'une courbe $(x(t), p(t))$ dans $\mathbb{R}_{(p,x)}^2$ est donnée par l'intégrale $y = \int p dx$, autrement dit que $y(t)$ satisfait

$$\dot{y}(t) = p(t) \dot{x}(t) .$$

Dans \mathbb{H} , la 1-forme de contact ω_J est remplacée par la 1-forme de contact symétrisée $\omega = dz - \frac{1}{2}(p dx - x dp)$ dont la nullité exprime que la variation de la troisième variable¹⁵ z le long d'une courbe $(x(t), p(t))$ est donnée par l'intégrale

$$z = \int \frac{1}{2} (p dx - x dp) ,$$

autrement dit que $z(t)$ satisfait

$$\dot{z}(t) = \frac{1}{2} (p(t) \dot{x}(t) - x(t) \dot{p}(t)) .$$

Mais $z = A$ car $A = \int_C \frac{1}{2} (p dx - x dp)$, cette égalité étant un cas élémentaire du théorème de Stokes.¹⁶ Par ailleurs, $L = \int_C ds$, le ds étant celui dans le plan $\mathbb{R}_{(p,x)}^2$. Minimiser le périmètre L avec une aire A fixée revient à trouver dans \mathbb{H} un lacet $(x(t), p(t), z(t))$ qui est une courbe intégrale du champ des noyaux de ω allant de (x_0, p_0, z_0) à $(x_0, p_0, z_0 + A)$ (où (x_0, p_0) est un point de C pris comme origine) en étant de longueur minimale. On se trouve ainsi conduit à un problème de géodésiques sous-riemanniennes dans \mathbb{H} muni de sa structure de contact naturelle. Nous développerons très longuement cette problématique dans les chapitres suivants.

5.3. $\mathbb{H}_{pol} \simeq \mathbb{V}_J$ et \mathbb{H} : les coordonnées canoniques

Explicitons un peu les relations entre les deux groupes de Lie $\mathbb{H}_{pol} \simeq \mathbb{V}_J$ et \mathbb{H} qui ont le même espace sous-jacent $\mathbb{R}^2 \times \mathbb{R}$ et les mêmes algèbres de Lie tout en ayant des lois de composition un peu différentes. Nous savons que $\mathbb{H}_{pol} \simeq \mathbb{V}_J$ et \mathbb{H} sont isomorphes.

Par définition, \mathbb{H} exponentie l'algèbre de Lie \mathcal{V}_J . Lorsqu'on écrit un élément $h \in \mathbb{H}$ sous la forme

-
14. Qui remonte à l'antiquité sous le nom du célèbre problème de Didon à propos de la fondation de Carthage.
 15. Nous la notons ici z pour éviter la confusion avec y .
 16. L'intégrale d'une 2-forme $d\alpha$ sur S est l'intégrale de la 1-forme α le long du bord $C = \partial S$ de S : $\int_S d\alpha = \int_{\partial S} \alpha$. Ici, $d\omega = dp \wedge dx$ est l'élément d'aire, ce qui prouve l'affirmation.

$$h = \exp(\mathbf{t}) = \exp(m(\xi, \eta, \pi)) = M\left(\xi, \eta + \frac{1}{2}\pi\xi, \pi\right),$$

on dit que les coordonnées (ξ, η, π) sont les “coordonnées exponentielles canoniques de première espèce” de h . Mais comme $\mathbb{H}_{pol} \simeq \mathbb{V}_J$ et \mathbb{H} ont le même espace sous-jacent, on peut traiter $h = \exp(\mathbf{t})$ comme un élément $h = v = M(x, y, p)$ de \mathbb{V}_J , le changement de coordonnées étant $x = \xi$, $y = \eta + \frac{1}{2}\pi\xi$, $p = \pi$. Cela correspond à d’autres coordonnées exponentielles canoniques de \mathbb{H} , dites “de seconde espèce”. Elles consistent à considérer non plus les $h = \exp(m(\xi, \eta, \pi))$ mais les $h' = \exp((\xi, 0, 0)) \cdot \exp((0, \eta, 0)) \cdot \exp((0, 0, \pi))$ et à poser que les coordonnées de h' sont (ξ, η, π) . Les éléments h et h' ne sont pas identifiables à cause de non-commutativité. D’après la formule de Baker-Campbell-Hausdorff,

$$\begin{aligned} \exp((\xi, 0, 0)) \cdot \exp((0, \eta, 0)) \cdot \exp((0, 0, \pi)) &= \\ \exp\left((\xi, \eta, 0) + \frac{1}{2}[(\xi, 0, 0), (0, \eta, 0)]\right) \cdot \exp((0, 0, \pi)) &= \\ \exp((\xi, \eta, 0)) \cdot \exp((0, 0, \pi)) &= \exp((\xi, \eta, 0)) \cdot \exp((0, 0, \pi)) \\ \exp\left((\xi, \eta, \pi) + \frac{1}{2}[(\xi, \eta, 0), (0, 0, \pi)]\right) &= \exp\left(\left(\xi, \eta - \frac{1}{2}\pi\xi, \pi\right)\right) \\ &= M\left(\xi, \eta - \frac{1}{2}\pi\xi + \frac{1}{2}\pi\xi, \pi\right) \\ &= M(\xi, \eta, \pi). \end{aligned}$$

Autrement dit, si l’on revient aux notations (x, y, p) , $M(x, y, p)$ est le point de \mathbb{V}_J de coordonnées canoniques de seconde espèce (x, y, p) où (x, y, p) est considéré comme un élément de l’algèbre de Lie \mathcal{V}_J (nous avons expliqué plus haut que les coordonnées de \mathcal{V}_J peuvent s’écrire comme des coordonnées de \mathbb{V}_J parce que l’espace $\mathfrak{M} \simeq \mathbb{V}_J$ des M satisfait $T_v\mathfrak{M} \simeq \mathfrak{M}$).

5.4. Le groupe de Heisenberg \mathbb{H} et la mécanique quantique

Nous rencontrons ainsi au cœur de la neurogéométrie le groupe de Heisenberg qui se trouve par ailleurs à la base de la mécanique quantique. Essayons d’en comprendre la raison formelle.¹⁷

Au tout début de la mécanique quantique, Schrödinger, pour adapter aux nouvelles données expérimentales les formalismes de la mécanique classique où tout

17. Purement “formelle” puisque les phénomènes naturels concernés n’ont rien à voir entre eux.

repose sur l'espace des phases avec ses coordonnées de position q et leurs moments conjugués p , représenta ces quantités de base comme des opérateurs linéaires \hat{q} et \hat{p} agissant sur la fonction d'onde ψ décrivant l'état du système quantique considéré. L'état ψ est un vecteur d'un espace de Hilbert \mathcal{H} approprié¹⁸ et, dans la représentation positionnelle, est une fonction $\psi(q) \in \mathcal{H} = L^2(\mathbb{R}, \mathbb{C})$ où $L^2(\mathbb{R}, \mathbb{C})$ est l'espace de Hilbert des fonctions $\psi(q)$ à valeurs complexes de carré intégrable, i.e. telle que l'intégrale $\int_{\mathbb{R}} |\psi(q)|^2 dq$ soit finie. Schrödinger a montré que \hat{q} et \hat{p} agissent respectivement comme multiplication par q et $-i\hbar$ fois la dérivation par rapport à q : $\hat{q}(\psi(q)) = q\psi(q)$ et $\hat{p}(\psi(q)) = -i\hbar \frac{d\psi(q)}{dq}$.^{19 20} Cette traduction des coordonnées classiques de l'espace des phases en opérateurs "externes" agissant sur un vecteur d'état ψ décrivant l'état "interne" du système quantique est à la base du passage de la mécanique classique à la mécanique quantique. Quand \hbar tend vers 0 les deux formalismes confluent l'un vers l'autre, ce que l'on appelle "l'approximation semi-classique".

Les opérateurs \hat{q} et \hat{p} sont *auto-adjoints*.²¹ En effet le produit scalaire de \mathcal{H} est donné par $\langle \varphi | \psi \rangle = \int_{\mathbb{R}} \varphi(q) \psi(q)^* dq$ (où $\psi(q)^*$ est le complexe conjugué de $\psi(q)$) et si A est un opérateur, son adjoint A^* est défini par $\langle A\varphi | \psi \rangle = \langle \varphi | A^*\psi \rangle$ pour tous $\varphi, \psi \in \mathcal{H}$. Comme q est une variable réelle, $q = q^*$ et on a trivialement

$$\int_{\mathbb{R}} q\varphi(q) \psi(q)^* dq = \int_{\mathbb{R}} \varphi(q) q^* \psi(q)^* dq = \int_{\mathbb{R}} \varphi(q) (q\psi(q))^* dq .$$

Pour \hat{p} on utilise le fait que $\frac{d(\varphi\psi)}{dq} = \frac{d\varphi}{dq}\psi + \varphi\frac{d\psi}{dq}$ et que les ψ doivent s'annuler à l'infini pour être de carré intégrable. On a donc bien, avec

$$\int_{\mathbb{R}} \frac{d(\varphi(q) \psi(q)^*)}{dq} dq = [\varphi(q) \psi(q)^*]_{-\infty}^{+\infty}$$

(la différence des valeurs de $\varphi(q) \psi(q)^*$ à $\pm\infty$), l'égalité :

$$\begin{aligned} \langle \hat{p}\varphi | \psi \rangle &= -i\hbar \int_{\mathbb{R}} \frac{d\varphi(q)}{dq} \psi(q)^* dq = i\hbar \int_{\mathbb{R}} \varphi(q) \frac{d\psi(q)^*}{dq} dq - i\hbar [\varphi(q) \psi(q)^*]_{-\infty}^{+\infty} \\ &= i\hbar \int_{\mathbb{R}} \varphi(q) \frac{d\psi(q)^*}{dq} dq = \int_{\mathbb{R}} \varphi(q) (-i\hbar)^* \frac{d\psi(q)^*}{dq} dq = \langle \varphi | \hat{p}^* \psi \rangle . \end{aligned}$$

18. Nous développerons cette notion dans la section 4 du chapitre 16.

19. La constante de Dirac \hbar est $\hbar = \frac{h}{2\pi}$, où h est la constante de Planck qui mesure le quantum d'action et dont la valeur est $h \sim 6,626 \times 10^{-34}$ Joule seconde.

20. Il est nécessaire que les $\psi(q)$ soient à valeurs dans \mathbb{C} puisque le facteur i intervient. Travailler sur \mathbb{C} est essentiel pour comprendre les phénomènes empiriques d'*interférence* considérés comme si spectaculaires et si contre-intuitifs au début de la mécanique quantique.

21. La théorie des opérateurs sur un espace de Hilbert sera précisée dans le chapitre 16 et le chapitre 17. Nous n'avons besoin ici que de rudiments particulièrement élémentaires.

La relation de non-commutation de \widehat{q}, \widehat{p} est caractéristique. On a

$$\begin{aligned} [\widehat{q}, \widehat{p}](\psi(q)) &= \widehat{q}\widehat{p}(\psi(q)) - \widehat{p}\widehat{q}(\psi(q)) = q \left(-i\hbar \frac{d\psi(q)}{dq} \right) - \left(-i\hbar \frac{d(q\psi(q))}{dq} \right) \\ &= -i\hbar q \frac{d\psi(q)}{dq} + i\hbar \left(q \frac{d\psi(q)}{dq} + \psi(q) \right) = i\hbar \psi(q) . \end{aligned}$$

C'est la célèbre formule de Heisenberg $[\widehat{q}, \widehat{p}] = i\hbar I$ (I est l'opérateur identité de \mathcal{H}).

La situation est un peu plus subtile qu'elle n'en a l'air. D'abord les opérateurs \widehat{q} et \widehat{p} sont non bornés et non partout définis, \widehat{p} présuppose par exemple que $\psi(q)$ soit dérivable ce qui n'est pas le cas des $\psi \in \mathcal{H}$ en général. Par ailleurs, cette relation ne peut pas exister entre des opérateurs linéaires définis sur un Hilbert \mathcal{H} de dimension finie n . En effet, la trace des matrices $n \times n$ des opérateurs (i.e. la somme de leurs éléments diagonaux) satisfait $\text{trace}(AB) = \text{trace}(BA)$ et donc $\text{trace}([A, B]) = 0$. La trace de la formule deviendrait donc $0 = i\hbar n$, ce qui est impossible. Les représentations linéaires de \widehat{q} et \widehat{p} comme opérateurs sur \mathcal{H} exigent donc que \mathcal{H} soit de dimension infinie (en général un espace fonctionnel L^2).

Heisenberg était également arrivé à la conclusion que les opérateurs auto-adjoints (les "observables") de position \widehat{q} et de moment \widehat{p} satisfaisaient la relation de commutation $[\widehat{q}, \widehat{p}] = i\hbar I$. Cette relation se mit donc à définir l'algèbre de Lie dite de Heisenberg.

Notons alors que, puisque $[\widehat{q}, I] = [\widehat{p}, I] = 0$, l'algèbre de Heisenberg initiale $[\widehat{q}, \widehat{p}] = i\hbar I$ est formellement notre algèbre de Lie \mathcal{V}_J à 3 générateurs \mathfrak{t}_i satisfaisant $[\mathfrak{t}_1, \mathfrak{t}_2] = \mathfrak{t}_3$ et $[\mathfrak{t}_1, \mathfrak{t}_3] = [\mathfrak{t}_2, \mathfrak{t}_3] = 0$ avec $\mathfrak{t}_1 = \widehat{q}$, $\mathfrak{t}_2 = \widehat{p}$, $\mathfrak{t}_3 = i\hbar I$. Mais les deux représentations sont très différentes.

Géométrie différentielle réelle	Analyse dans des Hilbert sur \mathbb{C}
Dimension finie	Dimension infinie
Champs invariants par translation	Opérateurs auto-adjoints
$\mathfrak{t}_1 = \partial_x + p\partial_y$	$\mathfrak{t}_1 = \widehat{q}$
$\mathfrak{t}_2 = \partial_p$	$\mathfrak{t}_2 = \widehat{p}$
$\mathfrak{t}_3 = -\partial_y$	$\mathfrak{t}_3 = i\hbar I$
$[\mathfrak{t}_1, \mathfrak{t}_2] = \mathfrak{t}_3, [\mathfrak{t}_1, \mathfrak{t}_3] = [\mathfrak{t}_2, \mathfrak{t}_3] = 0$	$[\mathfrak{t}_1, \mathfrak{t}_2] = \mathfrak{t}_3, [\mathfrak{t}_1, \mathfrak{t}_3] = [\mathfrak{t}_2, \mathfrak{t}_3] = 0$

6. Représentations adjointes et co-adjointes

6.1. Invariance par translations à gauche et non-holonomie

La structure de groupe de Lie induit tout un ensemble de structures dérivées permettant d'encoder de façon très "lisible" certaines de ses propriétés. Cela est en

particulier très utile pour *classer* les structures possibles. Un immense travail a été accompli dans ce domaine à partir de la seconde moitié du XIX^e siècle.

Commençons par l'homogénéité due à l'action du groupe sur lui-même. Comme nous l'avons déjà vu à la section 1, la translation à gauche L_v de \mathbb{V}_J définie par $L_v(v') = v.v'$ est un difféomorphisme non linéaire de \mathbb{V}_J dont l'application tangente en 0 est l'application linéaire

$$\begin{aligned} T_0L_v : \mathcal{V}_J = T_0\mathbb{V}_J &\rightarrow T_v\mathbb{V}_J \\ \mathbf{t} = (\xi, \eta, \pi) &\mapsto T_0L_v(\mathbf{t}) = (\xi, \eta + p\xi, \pi) \end{aligned}$$

de matrice

$$T_0L_v = \begin{pmatrix} 1 & 0 & 0 \\ p & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Cela montre que la base naturelle $\left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p} \right\}$ du fibré tangent $T\mathbb{V}_J = TJ^1\mathbb{R}^2$ associée au système de coordonnées $\{x, y, p\}$ n'est pas invariante à gauche et n'est pas une base de $T\mathbb{V}_J$ comme algèbre de Lie. C'est l'origine de la non-holonomie. Pour obtenir une base invariante à gauche il faut translater au moyen des L_v la base $\left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p} \right\}_0$ en 0 et cela donne la base $\left\{ \frac{\partial}{\partial x} + p\frac{\partial}{\partial y}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p} \right\}$, c'est-à-dire $\{\mathbf{t}_1, -\mathbf{t}_3, \mathbf{t}_2\}$. On a vu que $[\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3$, les autres crochets étant nuls.

Considérons maintenant un vecteur \mathbf{t} de K_0 . Comme $\eta = p\xi$ et $p = 0$, on a $\eta = 0$. Son translaté $T_0L_v(\mathbf{t})$ est par conséquent donné par $(\xi, p\xi, \pi)$. Comme $\eta = p\xi$, $T_0L_v(\mathbf{t})$ est un élément du plan de contact K_v et, répétons-le (cf. section 1), la structure de contact $\mathcal{K} = \{K_v\}$ n'est rien d'autre que le champ invariant à gauche des plans tangents translatsés à gauche de K_0 . De manière équivalente, on peut dire que \mathcal{K} est le champ des noyaux de la 1-forme ω_J qui est invariante à gauche. En effet à l'origine $\omega_J = dy - p dx$ est simplement $\omega_0 = dy$. Si l'on translate ω_0 au point v on obtient $\omega_v = T_0L_v^*(\omega_0)$ défini par la formule $\omega_v(\mathbf{t}) = \omega_0(T_0L_v^{-1}(\mathbf{t}))$ pour $\mathbf{t} = (\xi, \eta, \pi) \in T_v\mathbb{V}_J$. Mais

$$T_0L_v^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -p & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

et

$$T_0L_v^{-1}(\mathbf{t}) = \begin{pmatrix} 1 & 0 & 0 \\ -p & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \pi \end{pmatrix} = \begin{pmatrix} \xi \\ -p\xi + \eta \\ \pi \end{pmatrix}.$$

Donc

$$\omega_v(\mathbf{t}) = dy(\xi, -p\xi + \eta, \pi) = -p\xi + \eta = dy(\mathbf{t}) - p dx(\mathbf{t}) = \omega_{J,v}(\mathbf{t}).$$

Comme ω_J est L -invariante, $d\omega_J$ l'est aussi ainsi que $\omega_J \wedge d\omega_J = -dx \wedge dy \wedge dp$, ce qui montre que la mesure de Lebesgue $dxdydp$ sur \mathbb{V}_J est L -invariante. On dit que c'est une *mesure de Haar* (à gauche) de \mathbb{V}_J .²²

Remarque. On obtient une théorie équivalente en considérant les translations à droite $R_v(v') = v'.v$ au lieu des translations à gauche. On obtient $T_0R_v = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix}$ et les formules sont les mêmes en échangeant x et p . La translatée de $\omega_0 = dy$ est maintenant $\tilde{\omega}_J = dy - xdp$ et on a encore que $\tilde{\omega}_J \wedge d\tilde{\omega}_J$ est une forme volume. Le mesure de Lebesgue de \mathbb{V}_J est donc à la fois L -invariante et R -invariante. C'est une mesure de Haar à la fois à gauche et à droite. On dit que le groupe de Lie \mathbb{V}_J est *unimodulaire*. \square

6.2. Les représentations Ad , ad , Ad^* , ad^* et leurs orbites

Exemplifions sur notre exemple très simple quelques traits généraux des groupes de Lie sur lesquels nous reviendrons à la section 4.1 du chapitre 14.

6.2.1. Représentation adjointe.

La translation à gauche L_v translate la situation en 0 dans la situation équivalente en v . On peut revenir en 0 en utilisant la translation à droite $R_{v^{-1}}$. On obtient ainsi ce que l'on appelle un *automorphisme intérieur* du groupe de Lie $\mathbb{V}_J = J^1\mathbb{R}^2$:

$$A_v : \begin{array}{ccc} v' & \mapsto & v.v'.v^{-1} \\ (x', y', p') & \mapsto & (x', y' + px' - p'x, p') \end{array} .$$

Il est trivial de vérifier à la main qu'il s'agit bien d'un automorphisme de groupes.²³

Comme 0 est un point fixe de A_v , l'application tangente $Ad_v = T_0A_v$ de A_v en 0 est un automorphisme de l'algèbre de Lie $\mathcal{V}_J = T_0\mathbb{V}_J$. Sa matrice (le jacobien de

22. Pour des raisons de technicité, nous ne donnerons la définition générale d'une mesure de Haar qu'à la section 3.2 du chapitre 7.

23. $A_v(0) = 0$ et $A_v(v'.v'') = A_v(v').A_v(v'')$. En effet

$$\begin{aligned} A_v(v'.v'') &= A_v(x' + x'', y' + y'' + p'x'', p' + p'') \\ &= (x' + x'', y' + y'' + p'x'' + p(x' + x'') - (p' + p'')x, p' + p'') \\ A_v(v').A_v(v'') &= (x', y' + px' - p'x, p').(x'', y'' + px'' - p''x, p'') \\ &= (x' + x'', y' + px' - p'x + y'' + px'' - p''x + p'x'', p' + p'') \end{aligned}$$

et

$$y' + y'' + p'x'' + p(x' + x'') - (p' + p'')x = y' + px' - p'x + y'' + px'' - p''x + p'x'' .$$

Par ailleurs,

$$A_v(A_{v^{-1}}(v')) = v(v^{-1}v)v^{-1} = (vv^{-1})v'(vv^{-1}) = v'$$

et idem pour $A_{v^{-1}}(A_v(v')) = v'$ ce qui montre que $A_{v^{-1}}$ est l'inverse de A_v .

A_v en 0) est donnée par :

$$Ad_v = \begin{pmatrix} 1 & 0 & 0 \\ p & 1 & -x \\ 0 & 0 & 1 \end{pmatrix} .$$

On notera que cette matrice jacobienne ne dépend pas de y , c'est-à-dire du centre Z de \mathbb{V}_J .

Il est trivial de vérifier que cette application $v \mapsto Ad_v$ de \mathbb{V}_J dans le groupe $\text{Aut}(\mathcal{V}_J)$ des automorphismes de \mathcal{V}_J (application non injective puisqu'indépendante de Z) est une *représentation* (c'est-à-dire un morphisme de groupes). En effet elle ne dépend que des x et p et pour ces variables $v.v'$ est l'addition et pour les éléments p et $-x$ des matrices il en va de même. On l'appelle la *représentation adjointe* de \mathbb{V}_J . Son application tangente est un morphisme d'algèbres de Lie, noté $ad_{\mathfrak{t}}$, de l'algèbre de Lie \mathcal{V}_J dans l'algèbre de Lie $\text{End}(\mathcal{V}_J)$ (endomorphismes de \mathcal{V}_J) de $\text{Aut}(\mathcal{V}_J)$. Si $\mathfrak{t} = (\xi, \eta, \pi) \in \mathcal{V}_J$, la matrice de $ad_{\mathfrak{t}}$ est

$$ad_{\mathfrak{t}} = \begin{pmatrix} 0 & 0 & 0 \\ \pi & 0 & -\xi \\ 0 & 0 & 0 \end{pmatrix} .$$

On a donc

$$ad_{\mathfrak{t}}(\mathfrak{t}') = (0, \xi'\pi - \xi\pi', 0) = [\mathfrak{t}, \mathfrak{t}']$$

et le crochet de Lie peut ainsi être *reconstruit* à partir de la représentation adjointe.

Remarque. Cette forme des $ad_{\mathfrak{t}}$ permet de confirmer que \mathbb{V}_J est *unimodulaire* (i.e. que les mesures de Haar à gauche et à droite sont les mêmes, à savoir la mesure de Lebesgue). En effet, un groupe de Lie G est unimodulaire si et seulement si les ad_X pour $X \in \mathcal{G}$ sont de trace nulle, ce qui est bien le cas ici.²⁴ \square

Les *orbites* de la représentation adjointe sont faciles à calculer. Si $v = (x, y, p)$ varie dans $\mathbb{V}_J = J^1\mathbb{R}^2$, et si $\mathfrak{t} = (\xi, \eta, \pi) \in \mathcal{V}_J = T_0\mathbb{V}_J$ est au contraire fixé, alors $Ad_v(\mathfrak{t}) = (\xi, p\xi + \eta - x\pi, \pi)$ engendre la droite $\tilde{\mathfrak{t}} = (\xi, \mathbb{R}, \pi)$ quand $\xi \neq 0$ ou $\pi \neq 0$. Quand $\xi = \pi = 0$, $Ad_v(\mathfrak{t}) = \mathfrak{t}$ et tous les éléments $\mathfrak{t} = (0, \eta, 0)$ sont des points fixes.

6.2.2. Représentation co-adjointe.

Il est facile de *dualiser* ces constructions. Soit $\{dx, dy, dp\}$ la base de l'espace vectoriel des 1-formes sur $\mathbb{V}_J = J^1\mathbb{R}^2$ associée au système de coordonnées $\{x, y, p\}$. Au point 0 on obtient une base du dual \mathcal{V}_J^* de l'algèbre de Lie \mathcal{V}_J , et donc si ϖ est

24. Rappelons que nous donnerons la définition générale d'une mesure de Haar à la section 3.2 du chapitre 7.

une 1-forme sur \mathcal{V}_J elle peut s'écrire²⁵

$$\varpi = \xi^* dx + \eta^* dy + \pi^* dp = (\xi^*, \eta^*, \pi^*) .$$

Si $\mathbf{t} \in \mathcal{V}_J$, il est conventionnel de noter $\langle \varpi, \mathbf{t} \rangle$ la valeur $\varpi(\mathbf{t})$ afin de souligner la dualité entre vecteurs tangents et 1-formes (vecteurs cotangents). On définit alors la *représentation co-adjointe* par

$$\langle Ad_v^*(\varpi), \mathbf{t} \rangle = \langle \varpi, Ad_{-v}(\mathbf{t}) \rangle .$$

Rappelons que $v^{-1} = (-x, -y + px, -p)$ et que $Ad_v(\mathbf{t}) = (\xi, p\xi + \eta - x\pi, \pi)$ et donc $Ad_{v^{-1}}(\mathbf{t}) = (\xi, -p\xi + \eta + x\pi, \pi)$

Il est facile de voir qu'il s'agit bien d'une représentation du groupe \mathbb{V}_J sur \mathcal{V}_J^* . Comme

$$\begin{aligned} \langle \varpi, Ad_{-v}(\mathbf{t}) \rangle &= \left\langle \xi^* dx + \eta^* dy + \pi^* dp, \xi \frac{\partial}{\partial x} + (-p\xi + \eta + x\pi) \frac{\partial}{\partial y} + \pi \frac{\partial}{\partial p} \right\rangle \\ &= \xi^* \xi + \eta^* (-p\xi + \eta + x\pi) + \pi^* \pi \\ &= (\xi^* - \eta^* p) \xi + \eta^* \eta + (\pi^* + \eta^* x) \pi \end{aligned}$$

on obtient $Ad_v^*(\varpi) = (\xi^* - \eta^* p, \eta^*, \pi^* + \eta^* x)$, la matrice de Ad_v^* étant

$$Ad_v^* = \begin{pmatrix} 1 & -p & 0 \\ 0 & 1 & 0 \\ 0 & x & 1 \end{pmatrix} .$$

Si on représente les ϖ comme des vecteurs colonnes, c'est la matrice transposée $(Ad_{v^{-1}})^t$.

Les *orbites* de la représentation co-adjointe sont les plans $(\mathbb{R}, \eta^*, \mathbb{R})$ si $\eta^* \neq 0$ (c'est-à-dire les plans parallèles au plan (ξ^*, π^*) avec une ordonnée η^*). Si $\eta^* = 0$, tous les points $(\xi^*, 0, \pi^*)$ du plan (ξ^*, π^*) sont des points fixes.

En prenant l'application tangente de la représentation co-adjointe on obtient l'adjointe ad^* de l'application ad . Il s'agit d'un morphisme d'algèbres de Lie de \mathcal{V}_J dans $\text{End}(\mathcal{V}_J^*)$ défini par

$$\begin{aligned} ad_{\mathbf{t}}^*(\varpi)(\mathbf{t}') &= \langle ad_{\mathbf{t}}^*(\varpi), \mathbf{t}' \rangle = \langle \varpi, ad_{-\mathbf{t}}(\mathbf{t}') \rangle = \langle \varpi, -[\mathbf{t}, \mathbf{t}'] \rangle \\ &= \langle \xi^* dx + \eta^* dy + \pi^* dp, (0, -\xi' \pi + \xi \pi', 0) \rangle = \eta^* (-\xi' \pi + \xi \pi') . \end{aligned}$$

Mais comme $\xi' = dx(\mathbf{t}')$ et $\pi' = dp(\mathbf{t}')$, on obtient

$$ad_{\mathbf{t}}^*(\varpi) = -\eta^* \pi dx(\mathbf{t}') + \eta^* \xi dp(\mathbf{t}') = (-\eta^* \pi, 0, \eta^* \xi) ,$$

la matrice de $ad_{\mathbf{t}}^*$ étant

25. Nous choisissons la notation ξ^* , etc. pour maintenir la relation entre les lettres romaines et grecques tout en indiquant la nature duale des vecteurs cotangents par rapport aux vecteurs tangents.

$$ad_{\mathfrak{t}}^* = \begin{pmatrix} 0 & -\pi & 0 \\ 0 & 0 & 0 \\ 0 & \xi & 0 \end{pmatrix} .$$

6.3. La forme symplectique de Kirillov

Expliquons sur cet exemple élémentaire de groupe nilpotent la remarque de base d'Alexandre Kirillov, sur laquelle nous reviendrons plusieurs fois, à savoir que les orbites de la représentation co-adjointe sont canoniquement munies d'une *structure symplectique*.²⁶ La représentation co-adjointe $ad_{\mathfrak{t}}^*(\varpi)$ satisfait par définition

$$ad_{\mathfrak{t}}^*(\varpi)(\mathbf{u}) = \langle ad_{\mathfrak{t}}^*(\varpi), \mathbf{u} \rangle = \langle \varpi, -ad_{\mathfrak{t}}(\mathbf{u}) \rangle = -\varpi([\mathfrak{t}, \mathbf{u}]) = \varpi([\mathbf{u}, \mathfrak{t}])$$

pour $\mathfrak{t}, \mathbf{u} \in \mathcal{V}_J$ et $\varpi \in \mathcal{V}_J^*$. Si \mathcal{O} est l'orbite co-adjointe $Ad^*(\varpi)$ de ϖ dans \mathcal{V}_J^* , \mathcal{O} est l'espace homogène \mathbb{V}_J/S_{ϖ} où S_{ϖ} est le stabilisateur S_{ϖ} de ϖ dans \mathbb{V}_J et la projection canonique $\mathbb{V}_J \rightarrow \mathcal{O} = \mathbb{V}_J/S_{\varpi}$ est $v \mapsto Ad_v^*(\varpi)$, la fibre au-dessus de ϖ étant S_{ϖ} . Au niveau infinitésimal, l'espace tangent $T_{\varpi}\mathcal{O}$ s'identifie à \mathcal{V}_J/S_{ϖ} où S_{ϖ} est l'algèbre de Lie de S_{ϖ} . Cela revient à noter que tout $\mathfrak{t} \in \mathcal{V}_J$ induit un vecteur tangent \mathfrak{t}_{ϖ}^* à \mathcal{V}_J^* en tout ϖ en prenant $\mathfrak{t}_{\varpi}^* = \left. \frac{d}{ds} \right|_{s=0} \exp(-s\mathfrak{t})\varpi$ puis à quotienter ces \mathfrak{t}_{ϖ}^* par l'action co-adjointe.

Supposons que \mathcal{O} ne soit pas réduite à un point (sinon il n'y a rien à montrer), autrement dit que $\eta_0^* \neq 0$ et que \mathcal{O} est donc le plan des (ξ^*, η_0^*, π^*) .

Sur \mathcal{V}_J , ϖ induit une forme bilinéaire anti-symétrique B_{ϖ} par

$$B_{\varpi}(\mathfrak{t}, \mathbf{u}) = -\varpi([\mathfrak{t}, \mathbf{u}]) = ad_{\mathfrak{t}}^*(\varpi)(\mathbf{u}) .$$

Son noyau est S_{ϖ} car si \mathfrak{t} est tel que $B_{\varpi}(\mathfrak{t}, \mathbf{u}) = 0$, et donc $\varpi([\mathfrak{t}, \mathbf{u}]) = 0$, pour tout $\mathbf{u} \in \mathcal{V}_J$, alors $\exp(\mathfrak{t})$ est dans le stabilisateur S_{ϖ} de ϖ et donc $\mathfrak{t} \in S_{\varpi}$ (et réciproquement). Qui plus est, B_{ϖ} est S_{ϖ} -invariante car si $w \in S_{\varpi}$ alors

$$\begin{aligned} \langle \varpi, [Ad_w(\mathfrak{t}), Ad_w(\mathbf{u})] \rangle &= \langle \varpi, Ad_w([\mathfrak{t}, \mathbf{u}]) \rangle = \langle Ad_w^* \varpi, [\mathfrak{t}, \mathbf{u}] \rangle \\ &= \langle \varpi, [\mathfrak{t}, \mathbf{u}] \rangle \text{ car } w \in S_{\varpi} . \end{aligned}$$

On peut donc considérer sur \mathcal{O} la 2-forme

$$\sigma_{\mathcal{O}}(\varpi)(\mathfrak{t}_{\varpi}^*, \mathbf{u}_{\varpi}^*) = -\varpi([\mathfrak{t}, \mathbf{u}]) = ad_{\mathfrak{t}}^*(\varpi)(\mathbf{u})$$

obtenue par passage au quotient. C'est une forme bien définie qui est *symplectique* car elle est bilinéaire, antisymétrique (c'est trivial), non dégénérée et fermée. Elle est bien définie et non dégénérée puisque, comme nous venons juste de le voir, si \mathfrak{t}_{ϖ}^*

26. Né en 1938, Alexandre Kirillov est l'un des principaux spécialistes russes d'analyse fonctionnelle. Doctorant du maître Israel Gelfand et directeur de thèse de la médaille Fields 2006 Andrei Yuryevich Okounkov (bel exemple de transmission), sa "méthode des orbites" a révolutionné la théorie des représentations irréductibles des groupes de Lie nilpotents. Son traité de 1972 *Elements of the theory of representations* [288] est devenu un classique.

est tel que $\sigma_{\mathcal{O}}(\varpi)(\mathbf{t}_{\varpi}^*, \mathbf{u}_{\varpi}^*) = 0$ pour tout \mathbf{u}_{ϖ}^* , alors $\varpi([\mathbf{t}, \mathbf{u}]) = 0$ pour tout $\mathbf{u} \in \mathcal{V}_J$ et $\exp(\mathbf{t})$ est dans S_{ϖ} tandis que, justement, \mathcal{O} est le quotient \mathbb{V}_J/S_{ϖ} .

Par ailleurs, et c'est sa propriété cruciale, $\sigma_{\mathcal{O}}$ est *fermée* à cause de l'identité de Jacobi. Montrons-le de façon *abstraite* par un calcul général sans tenir compte du fait qu'ici $G = \mathbb{V}_J$ est nilpotent de dimension 3. La différentielle extérieure $d\sigma_{\mathcal{O}}$ de $\sigma_{\mathcal{O}}$ est une 3-forme G -invariante sur \mathcal{O} et agit donc sur des triplets $(\mathbf{r}^*, \mathbf{t}^*, \mathbf{u}^*)$ de champs de vecteurs tangents à \mathcal{O} G -invariants, ces \mathbf{t}^* venant de $\mathbf{t} \in \mathcal{G}$. Ces champs \mathbf{t}^* peuvent être interprétés, nous le savons, comme des dérivations sur les fonctions différentiables sur \mathcal{O} , leur valeur \mathbf{t}_{ϖ}^* en $\varpi \in \mathcal{O}$ étant $\mathbf{t}_{\varpi}^* = ad_{\mathbf{t}}^*(\varpi)$. Notons-les dans ce cas $\partial_{\mathbf{t}^*}$. Une formule générale dit que, si θ est une k -forme différentielle sur une variété différentiable M et si $X_j, j = 0, \dots, k$, sont $k + 1$ champs de vecteurs sur M , la dérivée extérieure $d\theta$ de θ est la $(k + 1)$ -forme

$$d\theta(X_0, \dots, X_k) = \sum_{j=0}^{j=k} (-1)^j X_j \theta \left(X_0, \dots, \overset{\emptyset}{X_j}, \dots, X_k \right) + \sum_{0 \leq i < j \leq k} (-1)^{i+j} \theta \left([X_i, X_j], X_0, \dots, \overset{\emptyset}{X_i}, \dots, \overset{\emptyset}{X_j}, \dots, X_k \right)$$

où $\overset{\emptyset}{X_j}$ signifie que X_k est absent. En appliquant ici cette formule générale, on obtient :

$$d\sigma_{\mathcal{O}}(\mathbf{r}^*, \mathbf{t}^*, \mathbf{u}^*) = \circlearrowleft \partial_{\mathbf{r}^*}(\sigma_{\mathcal{O}}(\mathbf{t}^*, \mathbf{u}^*)) - \circlearrowleft \sigma_{\mathcal{O}}([\mathbf{r}^*, \mathbf{t}^*], \mathbf{u}^*)$$

où \circlearrowleft symbolise la somme des permutations circulaires. Comme

$$\sigma_{\mathcal{O}}(\mathbf{t}^*, \mathbf{u}^*) = -\varpi([\mathbf{t}, \mathbf{u}]) \text{ ,}$$

on a

$$\partial_{\mathbf{r}^*}(\sigma_{\mathcal{O}}(\mathbf{t}^*, \mathbf{u}^*)) = \langle -ad_{\mathbf{r}}^*(\varpi), [\mathbf{t}, \mathbf{u}] \rangle = \langle \varpi, [\mathbf{r}, [\mathbf{t}, \mathbf{u}]] \rangle \text{ .}$$

Par ailleurs,

$$\sigma_{\mathcal{O}}([\mathbf{r}^*, \mathbf{t}^*], \mathbf{u}^*) = \langle \langle \varpi, [\mathbf{r}, [\mathbf{t}, \mathbf{u}]] \rangle \rangle$$

et donc

$$d\sigma_{\mathcal{O}}(\mathbf{r}^*, \mathbf{t}^*, \mathbf{u}^*) = \circlearrowleft 2 \langle \varpi, [\mathbf{r}, [\mathbf{t}, \mathbf{u}]] \rangle = 2 \langle \varpi, \circlearrowleft [\mathbf{r}, [\mathbf{t}, \mathbf{u}]] \rangle = 0$$

car l'identité de Jacobi s'écrit précisément $\circlearrowleft [\mathbf{r}, [\mathbf{t}, \mathbf{u}]] = 0$.

Le fait que $\sigma_{\mathcal{O}}$ soit une forme symplectique sur \mathcal{O} implique que \mathcal{O} est de dimension *paire* car seules les variétés de dimension paire peuvent admettre une forme symplectique.

Pour calculer $\sigma_{\mathcal{O}}$, notons que, pour les générateurs de $T_{\varpi}\mathcal{O}$ associés aux coordonnées ξ^* et π^* , comme $ad_{(\xi, 0, 0)}^*(\varpi) = (0, 0, \eta_0^* \xi)$, le vecteur tangent $(1, 0, 0)_{\varpi}^* \in T_{\varpi}\mathcal{O}$ peut s'interpréter comme $\eta_0^* \partial_{\xi^*}$ et, comme $ad_{(0, 0, \pi)}^*(\varpi) = (-\eta_0^* \pi, 0, 0)$, le vecteur tangent $(0, 0, 1)_{\varpi}^* \in T_{\varpi}\mathcal{O}$ peut s'interpréter comme $-\eta_0^* \partial_{\pi^*}$. On a donc

$$\sigma_{\mathcal{O}}(\varpi)(\eta_0^* \partial_{\xi^*}, -\eta_0^* \partial_{\pi^*}) = -\varpi(- (0, 1, 0)) = \eta_0^* \neq 0$$

ce qui implique $\sigma_{\mathcal{O}}(\varpi) = \frac{1}{\eta_0^*} d\xi^* \wedge d\pi^*$ qui est une forme symplectique \mathbb{V}_J -invariante sur l'orbite co-adjointe \mathcal{O} .

Remarque. Les travaux fondamentaux d'Alexandre Kirillov (cf. [290]), puis de Bertram Kostant, puis de Louis Auslander, Lajos Pukánszky et bien d'autres, ont généralisé ces formalismes valables sans restrictions pour les groupes nilpotents, d'abord aux groupes résolubles, puis à beaucoup d'autres groupes de Lie par exemple compacts. Le domaine d'investigation est immense, et concerne entre autres ce que l'on appelle la "*quantification géométrique*". Nous y reviendrons plusieurs fois, en particulier à la section 8.6 du chapitre 16 et à la section 6.4 du chapitre 17. —□

7. \mathbb{V}_S comme groupe des déplacements du plan

Nous avons considéré jusqu'ici le cas $\mathbb{V} = J^1M = M \times \mathbb{R}$ (avec pour simplifier $M = \mathbb{R}^2$) mais pas $\mathbb{V} = M \times \mathbb{S}^1$. Or les rares neurophysiologistes qui se sont intéressés à la modélisation mathématique de V1 l'ont identifié non pas à un $M \times \mathbb{R}$ mais précisément à un $M \times \mathbb{S}^1$. Le passage de la tangente p à une variable angulaire θ (avec $p = \tan(\theta)$) conduit à une autre façon, plus naturelle, de penser la structure de contact en utilisant $\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1$, la 1-forme

$$\omega_S = -\sin(\theta) dx + \cos(\theta) dy$$

et la distribution \mathcal{K}_S de plans tangents K_v de $T_v\mathbb{V}_S$ ($v \in \mathbb{V}_S$) définie par $K_v = \text{Span}\{X_1, X_2\}$ ²⁷, X_1 et X_2 étant les deux vecteurs tangents de base

$$\begin{cases} X_1 = \cos(\theta) \partial_x + \sin(\theta) \partial_y \\ X_2 = \partial_\theta \end{cases}$$

dont le crochet de Lie est

$$[X_1, X_2] = \sin(\theta) \partial_x - \cos(\theta) \partial_y = X_3 .$$

Remarque sur les notations. Quand cela sera préférable, nous utiliserons la notation X_i plutôt que \mathfrak{t}_i pour les vecteurs tangents de base afin de ne pas confondre les deux structures de contact \mathcal{K}_J et \mathcal{K}_S . —□

La distribution \mathcal{K}_S des plans de contact est encore maximale non intégrable car, comme

$$d\omega_S = \cos(\theta) dx \wedge d\theta + \sin(\theta) dy \wedge d\theta ,$$

la 3-forme

$$\omega_S \wedge d\omega_S = -dx \wedge dy \wedge d\theta$$

ne peut pas être nulle puisque c'est une forme volume. La condition de Frobenius $\omega_S \wedge d\omega_S = 0$ n'étant pas remplie, il n'existe pas de surface intégrale de \mathcal{K}_S dans \mathbb{V}_S

27. Rappelons que $\text{Span}\{X_i\}$ est le vectoriel engendré par les X_i .

(mais il existe beaucoup de courbes intégrales de \mathcal{K}_S : les relevées legendriennes Γ dans \mathbb{V}_S des courbes γ dans \mathbb{R}^2). Quant au champ de vecteurs $-X_3 = \chi$, il satisfait

$$\omega_S(\chi) = -(-\sin(\theta)\sin(\theta) + \cos(\theta)(-\cos(\theta))) = 1 .$$

C'est le champ caractéristique (champ de Reeb) du champ des plans de contact K_v .

Les deux structures de contact $\pi_J : \mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ et $\pi_S : \mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1 \rightarrow \mathbb{R}^2$ se ressemblent mais sont néanmoins très différentes. Pour s'en convaincre, il suffit de regarder les algèbres de Lie. Pour \mathbb{V}_J on a l'algèbre de Lie \mathcal{V}_J de base $\{\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3\}$ ($\mathbf{t}_1 = \frac{\partial}{\partial x} + p\frac{\partial}{\partial y}$, $\mathbf{t}_2 = \frac{\partial}{\partial p}$, $\mathbf{t}_3 = -\frac{\partial}{\partial y}$) avec $[\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3$ et $[\mathbf{t}_1, \mathbf{t}_3] = [\mathbf{t}_2, \mathbf{t}_3] = 0$. C'est, nous l'avons vu, une algèbre nilpotente. En revanche pour \mathbb{V}_S on a l'algèbre \mathcal{V}_S de base $\{X_1, X_2, X_3\}$

$$\begin{cases} X_1 = \cos(\theta)\partial_x + \sin(\theta)\partial_y \\ X_2 = \partial_\theta \\ X_3 = \sin(\theta)\partial_x - \cos(\theta)\partial_y \end{cases}$$

avec

$$\begin{cases} [X_1, X_2] = X_3 \\ [X_1, X_3] = 0 \\ [X_2, X_3] = X_1 \end{cases}$$

Elle n'est pas nilpotente. Nous reviendrons en détail sur ce point à la section 2 du chapitre 11 où nous utiliserons une représentation matricielle commode pour faire les calculs. Mais on peut remarquer que pour θ petit, on a au premier ordre $p \sim \theta$, $\sin(\theta) \sim \theta$ et $\cos(\theta) \sim 1$, ce qui fait que $\omega_S = -\sin(\theta)dx + \cos(\theta)dy$ s'approxime par $\omega \sim -\theta dx + dy$ qui est la 1-forme $\omega_J = dy - p dx$. La première structure est donc en quelque sorte "*tangente*" à la seconde. Nous verrons plus bas à la section 8.3 du chapitre 14 qu'on l'appelle le "cône tangent" de cette dernière.

L'espace $\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1$ étant isomorphe au groupe de Lie $SE(2) = \mathbb{R}^2 \rtimes SO(2)$ des déplacements du plan euclidien, il est alors intéressant de traiter à son tour, comme nous l'avons fait pour le premier modèle \mathbb{V}_J à la section précédente, ce second modèle \mathbb{V}_S comme un groupe de Lie.

Nous avons vu plus haut à la section 2.4 que si (q, r_θ) est un élément de $SE(2)$ où q est un point de \mathbb{R}^2 et r_θ la rotation d'angle θ , (q, r_θ) agit sur les points a de \mathbb{R}^2 par

$$(q, r_\theta)(a) = q + r_\theta(a) .$$

Si (q, r_θ) et (s, r_φ) sont deux éléments de $SE(2)$, leur produit (non commutatif) est donné par la formule de composition :

$$(s, r_\varphi) \circ (q, r_\theta) = (s + r_\varphi(q), r_{\varphi+\theta}) .$$

Si l'on transfère cette multiplication à \mathbb{V}_S on obtient le produit :

$$\begin{pmatrix} x_1 \\ y_1 \\ \theta_1 \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ y_2 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \cos(\theta_1) - y_2 \sin(\theta_1) \\ y_1 + x_2 \sin(\theta_1) + y_2 \cos(\theta_1) \\ \theta_1 + \theta_2 \end{pmatrix} .$$

L'origine $(0, 0, 0)$ est l'élément neutre et l'inverse de (x, y, θ) est

$$(-x \cos(\theta) - y \sin(\theta), x \sin(\theta) - y \cos(\theta), -\theta) .$$

L'algèbre de Lie de \mathbb{V}_S est le vectoriel $\mathcal{V}_S = T_0\mathbb{V}_S \simeq \mathbb{R}^3$ muni du crochet de Lie (où les $X = (\xi, \eta, \tau)$ dénotent les vecteurs tangents de $T_0\mathbb{V}_S$) que l'on pourra comparer à celui du premier modèle qui était $[\mathfrak{t}, \mathfrak{t}'] = (0, \xi'\pi - \xi\pi', 0)$:

$$[X, X'] = [(\xi, \eta, \tau), (\xi', \eta', \tau')] = (-\tau\eta' + \eta\tau', \tau\xi' - \xi\tau', 0) .$$

Cela montre que si $\tau = \tau' = 0$ alors $[X, X'] = 0$ et par conséquent

$$[[\mathcal{V}_S, \mathcal{V}_S], [\mathcal{V}_S, \mathcal{V}_S]] \equiv 0 .$$

Comme nous le verrons à la section 2 du chapitre 7, cela signifie que \mathcal{V}_S , même si elle n'est pas nilpotente, est du moins *résoluble*, ce qui n'est pas trop éloigné de la nilpotence.

Comme dans la section précédente, la translation à gauche L_v définie par $L_v(v') = v.v'$ est un difféomorphisme de \mathbb{V}_S dont l'application tangente en 0 est l'application linéaire :

$$\begin{array}{ccc} T_0L_v : & T_0\mathbb{V}_S & \rightarrow & T_v\mathbb{V}_S \\ & X = (\xi, \eta, \tau) & \mapsto & T_0L_v(X) = (\xi \cos(\theta) - \eta \sin(\theta), \xi \sin(\theta) + \eta \cos(\theta), \tau) . \end{array}$$

Dans la base (x, y, θ) la matrice de T_0L_v est donc

$$T_0L_v = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} r_\theta & 0 \\ 0 & 1 \end{pmatrix} .$$

On vérifie immédiatement que la base $\{\partial_x, \partial_y, \partial_\theta\}_0$ de $T_0\mathbb{V}_S$ est translatée en v par T_0L_v sur la base de $T_v\mathbb{V}_S$

$$\{\cos(\theta) \partial_x + \sin(\theta) \partial_y = X_1, -\sin(\theta) \partial_x + \cos(\theta) \partial_y = -X_3, \partial_\theta = X_2\}_v$$

et que donc la base $\{X_1, X_2, -X_3\}$ est invariante à gauche.

De même, si l'on translate à gauche la valeur en 0, $\omega_0 = dy$, de la 1-forme $\omega_S = -\sin(\theta) dx + \cos(\theta) dy$, on obtient, puisque

$$(i) \ \omega_v = T_0L_v^*(\omega_0) \text{ est définie par } \omega_v(X') = \omega_0(T_0L_v^{-1}(X')) \text{ pour } X' = (\xi', \eta', \tau') \in T_v\mathbb{V}_S ,$$

(ii)

$$T_0L_v^{-1} = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

(iii)

$$T_0L_v^{-1}(X') = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi' \\ \eta' \\ \tau' \end{pmatrix} = \begin{pmatrix} \xi' \cos(\theta) + \eta' \sin(\theta) \\ -\xi' \sin(\theta) + \eta' \cos(\theta) \\ \tau' \end{pmatrix},$$

la 1-forme invariante à gauche

$$\omega_v(X') = dy(T_0L_v^{-1}(X')) = -\xi' \sin(\theta) + \eta' \cos(\theta) = -\sin(\theta) dx + \cos(\theta) dy$$

qui n'est rien d'autre que ω_S . Cette dernière est donc bien invariante à gauche. Et comme la 3-forme $\omega_S \wedge d\omega_S$ L -invariante est la forme volume $-dx \wedge dy \wedge d\theta$, la mesure de Lebesgue $dx dy d\theta$ (ou $dx dy \frac{d\theta}{2\pi}$ si on normalise $d\theta$) est L -invariante et est donc une mesure de Haar à gauche de \mathbb{V}_S . On vérifie que c'est aussi une mesure de Haar à droite et que \mathbb{V}_S est donc un groupe de Lie unimodulaire.

En revenant de v en 0 par la translation à droite $R_{v^{-1}}$, on obtient l'automorphisme intérieur :

$$A_v : v' \mapsto v.v'.v^{-1}$$

$$v.v'.v^{-1} = \begin{pmatrix} x + (x' \cos(\theta) - x \cos(\theta')) - (y' \sin(\theta) - y \sin(\theta')) \\ y + (x' \sin(\theta) - x \sin(\theta')) + (y' \cos(\theta) - y \cos(\theta')) \\ \theta' \end{pmatrix}.$$

Évidemment 0 est un point fixe de A_v et l'application tangente $Ad_v = T_0A_v$ de A_v en 0 est un automorphisme de l'algèbre de Lie $\mathcal{V}_S = T_0\mathbb{V}_S$ qui définit la représentation adjointe. On a

$$Ad_v = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & y \\ \sin(\theta) & \cos(\theta) & -x \\ 0 & 0 & 1 \end{pmatrix}.$$

En reprenant l'application tangente de la représentation adjointe on obtient un morphisme d'algèbres de Lie, ad_X , de l'algèbre de Lie \mathcal{V}_S dans l'algèbre de Lie $\text{End}(\mathcal{V}_S)$ de $\text{Aut}(\mathcal{V}_S)$. Si $X = (\xi, \eta, \tau) \in \mathcal{V}_S = T_0\mathbb{V}_S$, la matrice de ad_X est

$$ad_X = \begin{pmatrix} 0 & -\tau & \eta \\ \tau & 0 & -\xi \\ 0 & 0 & 0 \end{pmatrix}.$$

On retrouve ainsi le crochet de Lie par

$$ad_X(X') = (-\tau\eta' + \eta\tau', \tau\xi' - \xi\tau', 0) = [X, X'] .$$

Remarque. Nous reviendrons plus bas en détail, à la section 3 du chapitre 11, sur cette structure et nous expliciterons les orbites des représentations adjointe et co-adjointe. \square

Remarque. Sur le plan métrique, notons que \mathbb{V}_S correspond au fait que l'on prend comme vecteur tangent à une courbe γ du plan de base, le vecteur de norme 1 de composantes $(\cos(\theta), \sin(\theta))$, ce qui revient à paramétrer γ par sa longueur d'arc. Nous y reviendrons également au chapitre 11. \square

8. Extension en dimension 3

Il est intéressant de regarder ce que devient ce formalisme lorsque l'on passe de la dimension 2 à la dimension 3. On considère donc non plus des courbes dans le plan \mathbb{R}^2 de coordonnées (x, y) mais des courbes $\gamma(s)$ (s est la longueur d'arc) dans l'espace \mathbb{R}^3 de coordonnées (x, y, z) . Le vecteur tangent unitaire $\dot{\gamma}(s) = \mathbf{t}(s)$ n'est plus de composantes $(\cos(\theta), \sin(\theta))$, $\theta \in \mathbb{S}^1$, mais est défini par deux angles $(\theta, \varphi) \in \mathbb{S}^2$ et il est de composantes $(\cos(\theta) \sin(\varphi), \sin(\theta) \sin(\varphi), \cos(\varphi))$. On a donc

$$\frac{dx}{\cos(\theta) \sin(\varphi)} = \frac{dy}{\sin(\theta) \sin(\varphi)} = \frac{dz}{\cos(\varphi)} = ds.$$

Le trièdre mobile de Frenet le long de γ est constitué du vecteur tangent unitaire $\mathbf{t}(s)$, du vecteur normal unitaire $\mathbf{n}(s) = \frac{1}{K} \dot{\mathbf{t}}(s) = \frac{1}{K} \ddot{\gamma}(s)$ où K est la courbure et le vecteur binormal $\mathbf{b}(s) = \mathbf{t}(s) \wedge \mathbf{n}(s)$ (produit vectoriel). On montre que $\dot{\mathbf{b}}(s)$ est proportionnel à $\mathbf{n}(s)$ et on définit la torsion $T(s)$ par $\dot{\mathbf{b}}(s) = -T(s) \mathbf{n}(s)$. On a alors $\dot{\mathbf{n}}(s) = -K \mathbf{t}(s) + T \mathbf{b}(s)$. $K = \|\mathbf{t}(s) \wedge \mathbf{t}'(s)\|$ et $T = \frac{\det(\mathbf{t}(s), \mathbf{t}'(s), \mathbf{t}''(s))}{K^2}$. Plus précisément

$$\dot{\mathbf{t}}(s) = \begin{pmatrix} -\sin(\theta) \sin(\varphi) \dot{\theta} + \cos(\theta) \cos(\varphi) \dot{\varphi}, \\ \cos(\theta) \sin(\varphi) \dot{\theta} + \sin(\theta) \cos(\varphi) \dot{\varphi}, \\ -\sin(\varphi) \dot{\varphi} \end{pmatrix}$$

et donc $K^2 = \sin(\varphi) \left(\dot{\theta} \right)^2 + (\dot{\varphi})^2$. Par ailleurs,

$$\mathbf{b}(s) = \frac{1}{K} \begin{pmatrix} -\cos(\varphi) \sin(\varphi) \cos(\theta) \dot{\theta} - \sin(\theta) \dot{\varphi}, \\ -\cos(\varphi) \sin(\varphi) \sin(\theta) \dot{\theta} + \cos(\theta) \dot{\varphi}, \\ \sin^2(\varphi) \dot{\theta} \end{pmatrix}$$

On en tire la formule pour $\dot{\mathbf{b}}(s)$

$$\left(\begin{array}{c} -\frac{(\cos(\varphi)\cos(\theta)\dot{\varphi}-\sin(\varphi)\sin(\theta)\dot{\theta})\left(\dot{\theta}\left(4\cos(\varphi)\dot{\varphi}^2+\sin(\varphi)\left(\sin(2\varphi)\dot{\theta}^2-2\dot{\varphi}\right)\right)+2\sin(\varphi)\dot{\varphi}\ddot{\theta}\right)}{2\left(\dot{\varphi}^2+\sin^2(\varphi)\dot{\theta}^2\right)^{3/2}}, \\ -\frac{(\cos(\varphi)\sin(\theta)\dot{\varphi}+\cos(\theta)\sin(\varphi)\dot{\theta})\left(\dot{\theta}\left(4\cos(\varphi)\dot{\varphi}^2+\sin(\varphi)\left(\sin(2\varphi)\dot{\theta}^2-2\dot{\varphi}\right)\right)+2\sin(\varphi)\dot{\varphi}\ddot{\theta}\right)}{2\left(\dot{\varphi}^2+\sin^2(\varphi)\dot{\theta}^2\right)^{3/2}}, \\ \frac{\sin(\varphi)\dot{\varphi}\left(\dot{\theta}\left(4\cos(\varphi)\dot{\varphi}^2+\sin(\varphi)\left(\sin(2\varphi)\dot{\theta}^2-2\dot{\varphi}\right)\right)+2\sin(\varphi)\dot{\varphi}\ddot{\theta}\right)}{2\left(\dot{\varphi}^2+\sin^2(\varphi)\dot{\theta}^2\right)^{3/2}} \end{array} \right)$$

et la valeur de la torsion

$$T = \frac{1}{K^2} \left(-\cos(\varphi)\sin^2(\varphi)\dot{\theta}^3 - 2\cos(\varphi)\dot{\varphi}^2\dot{\theta} + \sin(\varphi)\ddot{\varphi}\dot{\theta} - \sin(\varphi)\dot{\varphi}\ddot{\theta} \right).$$

Si maintenant on regarde les relevées legendriennes des courbes $\gamma(s)$ dans le fibré de dimension cinq $\mathbb{R}^3 \times \mathbb{S}^2$, elles satisfont le système de Pfaff des trois 1-formes

$$\begin{cases} \omega = -\sin(\theta)\sin(\varphi)dx + \cos(\theta)\sin(\varphi)dy \\ \tau = -\cos(\varphi)dx + \cos(\theta)\sin(\varphi)dz \\ \rho = -\cos(\varphi)dy + \sin(\theta)\sin(\varphi)dz \end{cases}$$

dont chacune est combinaison des deux autres.²⁸ Elles sont les courbes intégrales de la distribution \mathcal{K} de codimension 2 composée de sous-espaces tangents de dimension 3. Trois générateurs naturels de \mathcal{K} sont les trois champs de vecteurs

$$\begin{cases} X_{\mathbb{R}^3} = \cos(\theta)\sin(\varphi)\partial_x + \sin(\theta)\sin(\varphi)\partial_y + \cos(\varphi)\partial_z \\ X_{\theta} = -\frac{1}{\sin(\varphi)}\partial_{\theta} \\ X_{\varphi} = \partial_{\varphi} \end{cases}$$

Ils sont unitaires pour les métriques euclidiennes de \mathbb{R}^3 et \mathbb{S}^2 .²⁹

On en tire facilement la structure de leur algèbre de Lie en considérant les commutateurs

$$\begin{cases} X_4 = [X_{\mathbb{R}^3}, X_{\theta}] = -\sin(\theta)\partial_x + \cos(\theta)\partial_y \\ X_5 = [X_{\mathbb{R}^3}, X_{\varphi}] = -\cos(\theta)\cos(\varphi)\partial_x - \sin(\theta)\cos(\varphi)\partial_y + \sin(\varphi)\partial_z \\ X_6 = [X_{\theta}, X_{\varphi}] = -\frac{\cos(\varphi)}{\sin^2(\varphi)}\partial_{\theta} \end{cases}$$

28. Cette section étant simplement indicative, nous ne nous attardons pas sur les valeurs exceptionnelles $\sin(\varphi) = 0$, etc. pour lesquelles ces formes peuvent être automatiquement nulles.

29. Nous verrons à la section 6.3 du chapitre 6 que le tenseur métrique de \mathbb{S}^2 est

$$g_{11} = \sin^2(\varphi), g_{22} = 1, g_{12} = g_{21} = 0,$$

d'où le coefficient $\frac{1}{\sin(\varphi)}$ de X_{θ} .

Les 5 premiers sont indépendants car leur déterminant est $-\frac{1}{\sin(\varphi)} \neq 0$ et la condition d'Hörmander est vérifiée. Il est amusant à titre d'exercice de calculer les crochets d'ordre supérieur.

Le groupe des déplacements de \mathbb{R}^3 , $SE(3) = \mathbb{R}^3 \rtimes SO(3)$ agit naturellement sur $\mathbb{R}^3 \times \mathbb{S}^2$. Si (q, r) est un élément de $SE(3)$ où q est un point de \mathbb{R}^3 et r une rotation, (q, r) agit sur les points (a, s) de $\mathbb{R}^3 \times \mathbb{S}^2$ par

$$(q, r)(a, s) = (q + r(a), r(s)) .$$

9. Exponentielle générale

Revenons dans ce contexte élargi à la notion d'exponentielle. De façon générale, l'exponentielle est l'application $\exp : \mathcal{G} \longrightarrow G$, $X \mapsto g = \exp(X)$, qui *intègre* les champs de vecteurs tangents G -invariants (à gauche). Répétons qu'elle généralise à tous les groupes de Lie le lien qui existe grâce à l'exponentielle et au logarithme entre $\mathcal{G} = (\mathbb{R}, +)$, le groupe additif des réels et $G = (\mathbb{R}^*, \times)$, le groupe multiplicatif des réels > 0 . Si $X \in \mathbb{R}$, $e^X = 1 + X + \frac{X^2}{2} + \dots + \frac{X^n}{n!} + \dots$ et si $x \in \mathbb{R}^*$, $\log(1+x) = x - \frac{x^2}{2} + \dots + (-1)^{n+1} \frac{x^n}{n} + \dots$. Des identités formelles montrent alors que $\log(e^X) = X$ et $e^{\log(x)} = x$. Tout le problème de la généralisation aux groupes de Lie généraux est d'arriver à tenir compte de la non-commutativité.

Si X est un élément de l'algèbre de Lie \mathcal{G} , il engendre un groupe à un paramètre $\exp(tX)$ où $t \in \mathbb{R}$ avec

$$\begin{cases} \exp(tX) \cdot \exp(sX) = \exp((t+s)X) \\ \exp(-tX) = (\exp(tX))^{-1} . \end{cases}$$

Autrement dit, $\exp(tX)$ est une représentation du groupe additif $(\mathbb{R}, +)$ dans G . C'est une trajectoire qui part de l'élément neutre $e = e_G$ de G avec la "vitesse" X et dont la vitesse reste la translatée de X . Si l'on considère le champ X^L obtenu en translatant X partout dans G , $\exp(tX)$ est la trajectoire passant par e , la trajectoire passant par g étant $g \cdot \exp(tX)$.³⁰

L'application \exp est un difféomorphisme local entre \mathcal{G} et G qui envoie 0 sur e . Cela peut se démontrer à partir de théorèmes généraux sur l'existence et l'unicité de solutions de champs de vecteurs tangents C^∞ sur les variétés différentiables. Soit \mathfrak{P}_G l'algèbre de Lie des groupes à 1-paramètres de G . Si $\mathfrak{p}(t) \in \mathfrak{P}_G$, comme $\mathfrak{p}(t)$ est un morphisme de groupes, sa dérivée est la translatée le long de $\mathfrak{p}(t)$ de sa dérivée

30. Nous retrouverons des représentations du groupe additif $(\mathbb{R}, +)$ à la section 1.3 du chapitre 17, mais cette fois dans des espaces de Hilbert.

en 0 :

$$\begin{aligned} \mathfrak{p}'(t) &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\mathfrak{p}(t + \varepsilon) - \mathfrak{p}(t)) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\mathfrak{p}(t) \mathfrak{p}(\varepsilon) - \mathfrak{p}(t)) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathfrak{p}(t) (\mathfrak{p}(\varepsilon) - e) = \mathfrak{p}(t) \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\mathfrak{p}(\varepsilon) - \mathfrak{p}(0)) \\ &= \mathfrak{p}(t) \mathfrak{p}'(0) . \end{aligned}$$

On montre alors que l'application $\mathfrak{P}_G \rightarrow \mathcal{G}$ qui associe à $\mathfrak{p}(t) \in \mathfrak{P}_G$ l'élément $X = \mathfrak{p}'(0)$ est un isomorphisme. On peut donc utiliser son inverse et associer à X le $\mathfrak{p}_X(t) \in \mathfrak{P}_G$ tel que $\mathfrak{p}'(0) = X$. On *définit* alors $\exp(X)$ par $\mathfrak{p}_X(1)$. Ainsi définie, \exp est une application C^∞ et son application tangente $T_0 \exp$ en 0 est l'identité.³¹ En effet, l'algèbre de Lie \mathcal{G} étant un espace vectoriel, elle s'identifie à son espace tangent et si $Y \in T_0 \mathcal{G} \simeq \mathcal{G}$,

$$\begin{aligned} (T_0 \exp)(Y) &= \left. \frac{d}{dt} (t \rightarrow \exp(tY)) \right|_{t=0} = \left. \frac{d}{dt} (t \rightarrow \mathfrak{p}_{tY}(1)) \right|_{t=0} \\ &= \left. \frac{d}{dt} (t \rightarrow \mathfrak{p}_Y(t)) \right|_{t=0} = \mathfrak{p}'_Y(0) = Y . \end{aligned}$$

Par conséquent, \exp est un difféomorphisme local.

L'image $\exp(\mathcal{G})$ est incluse dans la composante connexe G_e de e dans G et, si G est compact (ce qui n'est pas le cas pour \mathbb{V}_J), est obligatoirement la composante G_e tout entière. Mais si G est nilpotent, connexe et simplement connexe (ce qui est en revanche le cas pour \mathbb{V}_J), alors \exp est un difféomorphisme *global*.

On vérifie que l'application \exp se comporte bien vis-à-vis des morphismes au sens où, si $\varphi : G \rightarrow H$ est un morphisme de groupes de Lie et si $T_e \varphi : \mathcal{G} \rightarrow \mathcal{H}$ est le morphisme d'algèbres de Lie qui est son application linéaire tangente, alors $\varphi \circ \exp_G = \exp_H \circ T_e \varphi$, autrement dit le diagramme suivant est commutatif

$$\begin{array}{ccc} \mathcal{G} & \xrightarrow{\exp} & G \\ T_e \varphi \downarrow & & \downarrow \varphi \\ \mathcal{H} & \xrightarrow{\exp} & H \end{array} .$$

Si l'on applique cela à la conjugaison $\varphi = A_g : G \rightarrow G$, $g' \rightarrow g.g'.g^{-1}$, on obtient, puisque $T_e \varphi$ est la représentation adjointe Ad_g , la relation

$$\begin{aligned} A_g \circ \exp &= \exp \circ Ad_g , \text{ i.e.} \\ g . \exp(X) . g^{-1} &= \exp(Ad_g(X)) . \end{aligned}$$

Si on l'applique à la représentation adjointe $Ad : G \rightarrow \text{Aut}(\mathcal{G})$, on obtient, puisque l'application linéaire tangente de Ad est $ad : \mathcal{G} \rightarrow \text{Aut}(\mathcal{G})$, la relation (dite de

31. Nous reviendrons à la section 3.5 du chapitre 7 sur l'application tangente $T_X \exp$ en $X \neq 0$ qui est très délicate à calculer par des formules générales.

Baker-Hausdorff)

$$\begin{aligned} Ad \circ \exp_G &= \exp_{\text{Aut}(G)} \circ ad, \text{ i.e.} \\ Ad_{\exp(X)} &= \exp(ad_X). \end{aligned}$$

Il est amusant de se convaincre de cette formule directement par un calcul formel. On veut montrer que la conjugaison

$$e^X Y e^{-X} = \sum_{n \geq 0} \frac{1}{n!} (ad_X)^n(Y).$$

On remplace X par sX (i.e. on utilise le groupe à 1-paramètre engendré par X) et l'on dérive la conjugaison $C(s) = e^{sX} Y e^{-sX}$ par rapport à s . En utilisant le fait que X commute avec ad_X et e^{sX} , il est trivial de montrer par induction que

$$\frac{d^n}{ds^n} (e^{sX} Y e^{-sX}) = e^{sX} (ad_X)^n(Y) e^{-sX}.$$

On écrit alors le développement de Taylor de $C(s) = \sum_{n \geq 0} \frac{s^n}{n!} \frac{d^n C}{ds^n}(0)$. Mais $\frac{d^n C}{ds^n}(0) = (ad_X)^n(Y)$. D'où la formule en faisant $s = 1$.

Les propriétés de l'exponentielle permettent de démontrer le *théorème de Cartan-von Neumann* évoqué à la section 2, théorème disant que tout sous-groupe algébrique topologiquement fermé H d'un groupe de Lie G est un sous-groupe de Lie (i.e. une sous-variété différentiable). Pour en donner une intuition, supposons que $G = GL_n(\mathbb{R})$, et donc $\mathcal{G} = \mathfrak{M}_n(\mathbb{R})$ l'algèbre de Lie des matrices $n \times n$. Soit H un sous-groupe fermé et \mathcal{H} le sous-ensemble des $X \in \mathcal{G}$ tels que le sous-groupe à 1-paramètre $\exp(\tau X)$ soit dans H . Il faut d'abord montrer que \mathcal{H} est un sous-espace vectoriel (et même une sous-algèbre de Lie) de $\mathcal{G} = \mathfrak{M}_n(\mathbb{R})$ puisque ce serait le cas si H était un sous-groupe de Lie. Le fait que $X \in \mathcal{H}$ implique $\lambda X \in \mathcal{H}$ est trivial. Pour la somme $(X + Y)$, si $X, Y \in \mathcal{H}$ et $\tau \in \mathbb{R}$, on peut ramener τX et τY dans un voisinage V_0 aussi petit que l'on veut de $0 \in \mathcal{G}$ en les divisant par un entier N assez grand. On choisit V_0 assez petit pour que \exp soit un difféomorphisme local entre $(\mathcal{G}, 0)$ et $(G, 1)$ et on effectue des approximations au premier ordre de $\left(e^{\frac{\tau X}{N}} e^{\frac{\tau Y}{N}}\right)^N$ qui montrent que la limite de ces éléments de \mathcal{H} est $e^{\tau(X+Y)}$. Mais comme \mathcal{H} est fermé, cette limite est dans \mathcal{H} et donc $(X + Y) \in \mathcal{H}$ et \mathcal{H} est un sous-espace vectoriel de $\mathcal{G} = \mathfrak{M}_n(\mathbb{R})$.

Comme \mathcal{H} est un sous-espace vectoriel de $\mathcal{G} = \mathfrak{M}_n(\mathbb{R})$, on peut considérer un supplémentaire \mathcal{S} et écrire $\mathfrak{M}_n(\mathbb{R}) = \mathcal{H} \oplus \mathcal{S}$. Soit alors

$$\Phi : \mathcal{G} = \mathfrak{M}_n(\mathbb{R}) \rightarrow G = GL_n(\mathbb{R})$$

l'application définie pour $Z = X + S$ par $\Phi(Z) = e^X e^S$. Comme son application tangente en 0 est l'identité, elle est un difféomorphisme local sur un V_0 et l'on montre que si V_0 est assez petit, ce difféomorphisme local envoie $V_0 \cap \mathcal{H}$ difféomorphiquement

sur $\Phi(V_0) \cap H$ et comme $V_0 \cap \mathcal{H}$ est une sous-variété, il en va de même de $\Phi(V_0) \cap H$. Le fait que $\Phi(V_0 \cap \mathcal{H}) \subseteq \Phi(V_0) \cap H$ est évident par définition de \mathcal{H} . Dans l'autre sens, pour des V_0 bien choisis, $\Phi(V_0) \cap H \subseteq \Phi(V_0 \cap \mathcal{H})$ car, pour N assez grand, les boules $B(0, \frac{1}{N})$ ont la propriété que si $Z \in B(0, \frac{1}{N})$ a son image $\Phi(Z)$ dans H alors $Z \in \mathcal{H}$. Car sinon, il existerait pour N aussi grand qu'on veut un $Z_N = X_N + S_N \in B(0, \frac{1}{N})$ avec $Z_N \in \mathcal{H}$ (et donc $S_N \neq 0$) dont l'image $\Phi(Z_N) = e^{X_N} e^{S_N}$ serait quand même dans H . Les suites X_N et S_N tendent vers 0 car les $B(0, \frac{1}{N})$ se contractent sur 0 et $\mathcal{H} \cap \mathcal{S} = \{0\}$, mais comme les $S_N \neq 0$ on peut les normaliser en posant $U_N = \frac{S_N}{\|S_N\|} \in \mathcal{S}$ et extraire de la suite des U_N (qui est une suite dans la boule unité qui est compacte) une sous suite convergeant vers un vecteur unité U . Mais alors, c'est le point clé, $e^{\tau U}$ est un élément de H car c'est une limite d'éléments de H et H est fermé. Donc $U \in \mathcal{H} \cap \mathcal{S}$ et, comme $\mathcal{H} \cap \mathcal{S} = \{0\}$, $U = 0$. Mais cela est impossible puisque U est de norme 1.

10. La forme de Maurer-Cartan

Disons maintenant un mot de ce qu'on appelle la *forme de Maurer-Cartan* Λ associée à un groupe de Lie V d'algèbre de Lie \mathcal{V} . Il s'agit d'une équation fondamentale et universelle.

Pour simplifier supposons que V soit de dimension 3 comme \mathbb{V}_J et gardons les notations utilisées pour \mathbb{V}_J . Soit $dv = (dx, dy, dp)$ la différentielle des coordonnées de V . Elle peut être considérée comme une 1-forme $dv \in T^*V \otimes \mathcal{V}$ (où \otimes est le produit tensoriel) sur V à valeurs non plus dans \mathbb{R} mais dans l'algèbre de Lie $\mathcal{V} = T_0V$ au sens où, si $\mathfrak{t}_v = (\xi, \eta, \pi) \in T_vV$ est un vecteur tangent de V en v , alors $dv(v)(\mathfrak{t}_v)$, qui est la valeur en \mathfrak{t}_v de la 1-forme dv prise au point v de V , est alors un vecteur tangent à V en 0. Mais, précisément parce que les coordonnées $v = (x, y, p)$ ne sont pas holonomes, dv n'est pas invariante à gauche. La forme de Maurer-Cartan Λ consiste à partir de la 1-forme $dv(0)$ sur T_0V et à la translater de façon à obtenir une 1-forme qui est invariante à gauche par construction. Par définition, $\Lambda(v) = (T_vL_{v^{-1}})^* dv(0)$ où

$$T_vL_{v^{-1}} = (T_0L_v)^{-1} : T_vV \rightarrow T_0V .$$

On a donc par définition

$$\Lambda(v) : T_vV \xrightarrow{T_vL_{v^{-1}}} T_0V \xrightarrow{dv(0)} \mathbb{R} .$$

Mais pour tout vecteur tangent $\mathfrak{t} = (\xi, \eta, \pi) \in T_0V$, la valeur $dv(0)(\mathfrak{t})$ (qui est par construction un vecteur de T_0V) n'est rien d'autre que \mathfrak{t} *lui-même* puisque dx, dy et dp extraient les composantes de \mathfrak{t} .

Par exemple, dans le cas de \mathbb{V}_J et \mathcal{V}_J , on obtient

$$\Lambda(v)(\mathfrak{t}) = T_0L_v^{-1}(\mathfrak{t}) = (\xi, -p\xi + \eta, \pi) .$$

Cela montre que Λ est la 1-forme à valeurs dans \mathcal{V}_J de composantes :

$$\Lambda = (dx, \omega_J, dp) .$$

En utilisant cette expression explicite de Λ , il est facile de vérifier que Λ satisfait l'équation fondamentale dite de *Maurer-Cartan*

$$d\Lambda + \frac{1}{2} [\Lambda \wedge \Lambda] = 0$$

où $[\Lambda \wedge \Lambda]$ combine le produit extérieur \wedge des formes avec le crochet $[\cdot, \cdot]$ de l'algèbre de Lie où elles prennent leurs valeurs.³² En effet, Λ s'écrit

$$\Lambda = dx \otimes \partial_x + \omega_J \otimes \partial_y + dp \otimes \partial_p \in T^*\mathbb{V}_J \otimes \mathcal{V}_J$$

où la présence du terme $\omega_J \otimes \partial_y$ au lieu de $dy \otimes \partial_y$ exprime la non-holonomie.³³ Comme $d^2x = d^2y = 0$ et $d\omega_J = dx \wedge dp$ d'après l'équation (1) de la section 4 du chapitre 3, on obtient $d\Lambda = (dx \wedge dp) \otimes \partial_y$. Par ailleurs, par définition du produit extérieur des 1-formes à valeurs dans une algèbre de Lie, on a :

$$\begin{aligned} [\Lambda \wedge \Lambda] &= (dx \wedge dx) \otimes [\partial_x, \partial_x] + (dx \wedge \omega_J) \otimes [\partial_x, \partial_y] + (dx \wedge dp) \otimes [\partial_x, \partial_p] \\ &\quad + (\omega_J \wedge dx) \otimes [\partial_y, \partial_x] + (\omega_J \wedge \omega_J) \otimes [\partial_y, \partial_y] + (\omega_J \wedge dp) \otimes [\partial_y, \partial_p] \\ &\quad + (dp \wedge dx) \otimes [\partial_p, \partial_x] + (dp \wedge \omega_J) \otimes [\partial_p, \partial_y] + (dp \wedge dp) \otimes [\partial_p, \partial_p] . \end{aligned}$$

Mais $dx \wedge dx = \omega_J \wedge \omega_J = dp \wedge dp = 0$ et $[\partial_x, \partial_x] = [\partial_y, \partial_y] = [\partial_p, \partial_p] = 0$ pour des raisons générales d'antisymétrie ; $dx \wedge \omega_J = dx \wedge dy$ et $dp \wedge \omega_J = dp \wedge dy - pdp \wedge dx$ par définition de ω_J ; $[\partial_x, \partial_y] = [\partial_y, \partial_p] = 0$ et $[\partial_x, \partial_p] = -\partial_y$ à cause de la structure d'algèbre de Lie. Toutes ces égalités impliquent immédiatement $[\Lambda \wedge \Lambda] = -2d\Lambda$.

Remarque. Si l'on définit le crochet $[\Lambda, \Lambda]$ comme la 2-forme à valeurs dans \mathcal{V}_J par $[\Lambda, \Lambda](\mathbf{t}, \mathbf{u}) = [\Lambda(\mathbf{t}), \Lambda(\mathbf{u})]$, alors $[\Lambda \wedge \Lambda] = 2[\Lambda, \Lambda]$ et la forme de Maurer-Cartan s'écrit

$$d\Lambda + [\Lambda, \Lambda] = 0 .$$

Par exemple pour les termes correspondant à x et y , on obtiendra dans $[\Lambda \wedge \Lambda]$

$$\begin{aligned} &(dx \wedge \omega_J)(\mathbf{t}, \mathbf{u}) [\partial_x, \partial_y] + (\omega_J \wedge dx)(\mathbf{t}, \mathbf{u}) [\partial_y, \partial_x] \\ &= 2(dx \wedge \omega_J)(\mathbf{t}, \mathbf{u}) [\partial_x, \partial_y] = 2(dx(\mathbf{t})\omega_J(\mathbf{u}) - dx(\mathbf{u})\omega_J(\mathbf{t})) [\partial_x, \partial_y] \end{aligned}$$

alors que dans $[\Lambda, \Lambda]$ on obtiendra

$$dx(\mathbf{t})\omega_J(\mathbf{u}) [\partial_x, \partial_y] + \omega_J(\mathbf{t})dx(\mathbf{u}) [\partial_y, \partial_x] = (dx(\mathbf{t})\omega_J(\mathbf{u}) - dx(\mathbf{u})\omega_J(\mathbf{t})) [\partial_x, \partial_y] .$$

32. $\Lambda = (dx, \omega_J, dp)$ a pour composante relative à y la forme de contact ω_J L -translatée de $dy(0)$, mais elle est plus compliquée que ω_J et à valeurs dans \mathcal{V} . La formule de Maurer-Cartan est fautive pour ω_J puisque $d\omega_J = dx \wedge dp \neq 0$ tandis que $\omega_J \wedge \omega_J = 0$.

33. Cette expression signifie bien sûr que, dans la base $(\partial_x, \partial_y, \partial_p)$ de \mathcal{V}_J , si $\mathbf{t} = (\xi, \eta, \pi) \in T_vV$ est un vecteur tangent à V en v , alors

$$\Lambda_v(\mathbf{t}) = dx(t)\partial_x + \omega_J(t)\partial_y + dp(t)\partial_p \in \mathcal{V}_J .$$

Cette formulation alternative est courante en physique. _____□

Ces formalismes se généralisent aux groupes de Lie G quelconques. L'équation de Maurer-Cartan est une équation *universelle*. Comme nous allons l'expliquer, elle exprime qu'une certaine connexion est de courbure nulle. Pour le comprendre, il faut dire un mot des connexions de Cartan générales, y compris celles qui ne sont pas de courbure nulle. Nous allons le faire dans un contexte général. Depuis Riemann, la notion de métrique sur une variété différentiable M est définie comme la donnée supplémentaire de métriques sur les plans tangents T_xM , métriques variant différentiablement avec le point $x \in M$. Élie Cartan a considérablement généralisé cette révolution riemannienne avec le concept général de connexion.

Géométrie des connexions

1. Au “cœur” de l’espace : géométrie différentielle et groupes de Lie

La forme de Maurer-Cartan relève d’une conception fondamentale de la géométrie qu’Élie Cartan considérait comme l’un des trois grands points de vue possibles sur la géométrie. Elle conduit selon lui au “cœur” de l’intuition spatiale où la géométrie différentielle de l’intégrabilité s’intrique avec les groupes de Lie.

Et, revenons-y, c’est bien la découverte que ce “cœur” géométrique est, pour la basse dimension de l’espace visuel, neuralemement implémenté qui fonde, comme nous l’avons développé dans ces deux derniers chapitres, la neurogéométrie.

Dans sa conférence au Congrès International des Mathématiciens à Oslo en 1936, “Le rôle de la théorie des groupes de Lie dans l’évolution de la géométrie moderne” [96], Cartan commente les

“trois points de vue principaux qui ont dominé l’évolution de la géométrie depuis le début de ce siècle.” ([96], p. 93)

1. Selon Felix Klein et son “programme d’Erlangen” (“Erlanger Program”) héritier de la géométrie euclidienne et des géométries non euclidiennes, une géométrie étudie les propriétés de configurations géométriques d’un espace M qui sont *invariantes* par rapport à un certain groupe. Tout groupe de Lie agissant sur lui-même par translations définit une géométrie de Klein. Ces géométries sont homogènes et ont par définition un groupe de Lie de dimension finie d’automorphismes globaux.

2. Selon Riemann (puis Christoffel, Lipschitz, Levi-Civita, et plusieurs autres) une géométrie est la donnée en chaque point de propriétés métriques infinitésimales cohérentes (c’est le tenseur métrique $ds^2 = g_{ij}dx_i dx_j$ définissant sur chaque espace tangent $T_x M$ d’une variété différentiable M une forme quadratique variant différentiablement avec x). En chaque point, le groupe euclidien des isométries garde un sens infinitésimal et local, mais, globalement, une variété riemannienne de courbure non constante n’a pas d’isométries globales. Son groupe d’automorphismes est de dimension infinie. La géométrie riemannienne a permis à Einstein d’élaborer la relativité générale.¹

1. Même si notre point de vue en géométrie s’inspire essentiellement de Cartan, nous reviendrons de façon récurrente sur la géométrie riemannienne. Riemann est un “prodigieux génie”

3. Selon Cartan, le troisième point de vue est celui du “parallélisme” et du “transport parallèle”.² Il donne un sens aux “translations infinitésimales” rendant possible de “raccorder entre eux” les espaces tangents, munis de leur structure spécifique (euclidienne ou autre), aux points infiniment voisins. En géométrie riemannienne (connexion de Levi-Civita), il permet

“d’attribuer aux courbes de l’espace riemannien toutes les propriétés de courbure et de torsion des courbes plongées dans l’espace euclidien.” ([96], p. 97)

Cartan donne l’exemple de la structure euclidienne de $M = \mathbb{R}^3$. On part de M comme variété différentiable, c’est-à-dire sans structure affine ni structure métrique. Autrement dit, on part d’une structure différentiable sous-jacente où seule la structure différentielle (infinitésimale, locale) a un sens. Puis, pour ajouter des structures supplémentaires (comme ici la structure métrique affine invariante par le groupe euclidien $SE(3)$), on ajoute des structures infinitésimales sur les $T_x M$ que l’on coordonne de façon à pouvoir les globaliser par intégration. La théorie des formes différentielles devient alors l’outil de base. En chaque point x de M , $T_x M$ doit être muni d’une structure euclidienne et pour comparer $T_x M$ et $T_{x'} M$ en $x' = x + \delta x$, il faut connaître les 6 composantes (rotation et translation) de la transformation affine euclidienne δS de $T_x M$ transformant $T_x M$ en $T_{x'} M$. Cela revient à se donner 6 formes de Pfaff : τ_1, τ_2, τ_3 pour les translations et ρ_1, ρ_2, ρ_3 pour les rotations. L’ensemble des (τ, ρ) peut être interprété comme une forme de Pfaff ω sur M à valeurs dans l’algèbre de Lie \mathcal{G} de $G = SE(3)$.

Et Cartan souligne (au sens propre, le texte est en italiques)

“La seule connaissance de ces six formes, dont les trois premières [translations] sont linéairement indépendantes, permet de reconstruire, à un déplacement près, l’espace euclidien, sans qu’on ait besoin de connaître la nature des coordonnées $[x_i]$.” ([96], p. 98)

“L’espace euclidien est complètement défini comme un support de déplacements infinitésimaux $[\delta S]$ attachés aux différents couples de points infiniment voisins.” ([96], p. 98)

Si $G = SE(3)$ est le groupe qui définit la géométrie de Klein globale de \mathbb{R}^3 , la variété différentiable de base M (\mathbb{R}^3 comme pur continuum tridimensionnel) apparaît ainsi comme le “support” de transformations infinitésimales de G . Si R_x et $R_{x'}$ sont des rotations de $T_x M$ et $T_{x'} M$, le déplacement infinitésimal δS est remplacé par son conjugué $R_x^{-1}(\delta S)R_{x'}$.

(Jean Dieudonné) doté d’une profondeur philosophique incomparable. Pour des “Réflexions sur Riemann, philosophe mathématicien” le lecteur pourra consulter l’article [521] de Jean-Jacques Szczeciniarz.

2. C’est évidemment pourquoi, nous ne saurions trop y insister, Cartan est si crucial pour la neurogéométrie puisque $V1$ implémente primitivement un transport parallèle et non pas une métrique riemannienne.

Les équations de Maurer-Cartan (remontant à Darboux pour le cas euclidien)

$$d\omega + \frac{1}{2}[\omega \wedge \omega] = 0$$

sont les équations de compatibilité que doit satisfaire ω . Elles expriment le fait que si l’on intègre ω le long de courbes fermées infiniment petites on obtient la transformation identique. Cartan souligne à propos de ces équations :

“Nous sommes [...] à la fois au cœur de la géométrie différentielle générale et au cœur de la théorie des groupes finis et continus.” ([96], p. 99)

Dans le cas de l’espace euclidien qui est “plat” et simplement connexe, il en va de même le long de courbes fermées quelconques. Mais en général, ce ne sera pas le cas et on aura une holonomie non triviale (i.e. une propriété de non-holonomie).

La philosophie de l’espace de Cartan est implicitement fondée sur une philosophie des fibrations, même si la théorie générale des fibrations lui est postérieure. Si $\pi : V \rightarrow M$ est une fibration ayant pour base une variété M et pour fibres V_x au-dessus des points $x \in M$ un espace d’entités d’un certain type muni d’un certain groupe d’automorphismes G (le “groupe structural”), le problème est de relier entre elles les fibres voisines lorsqu’on se déplace dans la base M . Cela se fait en définissant les éléments de G qui permettent de suivre les transformations de V_x lors de déplacements infinitésimaux δx . On veut donc lifter le calcul différentiel de M dans V en utilisant le groupe structural G . D’où le concept de “connexion”, une connexion \mathcal{C} sur un fibré $\pi : V \rightarrow M$ consistant à se donner, en chaque point $v \in V$, un relèvement \mathcal{C}_v – appelé “plan horizontal” – de $T_{\pi(v)}M$ dans T_vV . Comme le notent Chern et Chevalley, le problème est

“to tie up the fiber with the differentiable structure of the base space.” ([114], p. 243)

En termes de G , on doit associer à toute courbe différentiable $x(t)$ dans M telle que $x(0) = x_0$, une courbe $g(t)$ dans G qui permettent de suivre les relevées $v(t)$ de $x(t)$ dans V à partir d’un point v_0 quelconque de la fibre V_{x_0} . Le vecteur tangent $\dot{v}(t) \in T_{v(t)}V$ se projette sur sa composante “horizontale” $\dot{x}(t) \in T_{x(t)}M$ et il faut donc connaître sa composante “verticale” tangente à la fibre $V_{x(t)}$. Elle se calcule à partir d’un système différentiel donnant la courbe $g(t)$. Il s’agit d’une stratégie générale. Comme le notent encore Chern et Chevalley,

“When one takes the notion of a connection as the guiding principle in differential geometry, the fundamental problem is to define the fiber bundle and the connection in every geometric problem.” ([114], p. 243)

2. Les structures de contact comme connexions

Un des grands intérêts de l'interaction entre la structure de contact \mathcal{K} de \mathbb{V}_J et sa structure de groupe de Lie est que l'on peut considérer que \mathcal{K} , ainsi que la 1-forme invariante à gauche ω_J dont elle est le noyau, définissent une *connexion* au sens de Cartan sur \mathbb{V}_J . Celle-ci est associée *non plus* à la projection π de \mathbb{V}_J sur le plan (x, y) ³ mais à la projection $\pi' : \mathbb{V}_J \rightarrow M$ de \mathbb{V}_J sur le plan de base (x, p) . Nous l'avons anticipé dans le chapitre 5 du Vol I, section 5.4.2. La fibre de π' est maintenant l'axe des y . Comme toute connexion \mathcal{C} sur un fibré $V \rightarrow M$ ⁴, elle doit, répétons-le, permettre de comparer les plans tangents voisins au moyen d'un transport parallèle des vecteurs tangents qui est obtenu en choisissant en chaque point v au-dessus de chaque point b de la base M un sous-espace "horizontal" de $T_v V$ se projetant isomorphiquement sur $T_b M$. Ici, $b = (x, p)$ et les plans de contact $\mathcal{K} = \ker(\omega)$, qui sont "verticaux" pour la projection sur le plan (x, y) , sont maintenant les plans "horizontaux" de $\mathcal{C} = \mathcal{K}$ au-dessus du plan (x, p) .

On peut montrer que la connexion est compatible avec la projection $\pi' : V \rightarrow V/W = M$, où W est le sous-groupe d'isotropie de ω_0 (dans notre cas $\omega_0 = dy$) pour la représentation co-adjointe.⁵ Par définition, W est l'ensemble

$$W = \{v \in V \mid Ad_v^*(\omega_0) = \omega_0\} .$$

C'est un groupe de Lie de dimension 1 et son algèbre de Lie \mathcal{W} est l'ensemble

$$\mathcal{W} = \{\mathfrak{t} \in \mathcal{W} \mid ad_{\mathfrak{t}}^*(\omega_0) = 0\} .$$

Dans notre cas, comme

$$\omega_0 = dy = (\xi^* = 0, \eta^* = 1, \pi^* = 0) ,$$

on a, pour $v = (x, y, p)$,

$$Ad_v^*(\omega_0) = (\xi^* - \eta^* p, \eta^*, \pi^* + \eta^* x) = (-p, 1, x) .$$

Pour satisfaire l'identité $Ad_v^*(\omega_0) = \omega_0 = (0, 1, 0)$ on doit avoir $x = 0$ et $p = 0$. Donc W est l'axe des y , la loi de groupe restreinte étant tout simplement l'addition $y + y'$. On vérifie que, comme $ad_{\mathfrak{t}}^*(\omega_0) = (-\pi, 0, \xi)$ en vertu de $ad_{\mathfrak{t}}^*(\theta) = (-\eta^* \pi, 0, \eta^* \xi)$ puisque $\eta^* = 1$, l'identité $ad_{\mathfrak{t}}^*(\omega_0) = (-\pi, 0, \xi) = 0$ conduit à $\xi = 0$ et $\pi = 0$ qui est aussi l'axe des y , mais considéré cette fois comme l'algèbre de Lie \mathcal{W} de W . Cela est général. C'est pourquoi, quand on adopte ce point de vue, celui de la projection $\pi' : V \rightarrow V/W = M$ de V sur la base M , on appelle naturellement les plans de contact K_v les plans "horizontaux" car ils sont précisément les plans

3. Ici x est une coordonnée et non pas un point de M .

4. Nous changeons la notation de \mathcal{K} en \mathcal{C} pour souligner dans notre cas le changement du plan de base de (x, y) en (x, p) .

5. $\omega_0 = dy = (\xi^* = 0, \eta^* = 1, \pi^* = 0)$ est la valeur en $0 = (x = 0, y = 0, p = 0)$ de la 1-forme $\omega = dy - p dx = (\xi^* = -p, \eta^* = 1, \pi^* = 0)$.

horizontaux de la connexion. Cette représentation comme connexion approfondit la *dualité projective* déjà évoquée à la section 2.1 du chapitre 3. Répétons que dans la perspective classique, le plan de base est, dans notre cas, le plan (x, y) et la fibre est l'axe des tangentes p . Les courbes γ sont données comme des fonctions $y = f(x)$ et la tangente p est calculée par dérivation. Dans la perspective alternative, le plan de base est le plan (x, p) et la fibre est l'axe des y qui est le centre du groupe. Les courbes γ sont données comme des fonctions $p = g(x)$, c'est-à-dire comme enveloppes de leurs tangentes, et y est calculé par intégration au moyen de l'intégrale $y = \int y'dx = \int pdx$.

On peut montrer que $d\omega$ (dont, comme nous allons le voir dans la prochaine section, la restriction aux plans de contact désormais traités comme "horizontaux" donne la *courbure* de la connexion ω) est une *forme symplectique* sur la nouvelle base $V/W = M$. C'est évident dans notre cas puisque $d\omega_J = dx \wedge dp$ est la forme symplectique standard sur le plan de base $\{x, p\} = \mathbb{V}_J/W$. Et comme $dx \wedge dp$ est l'élément d'*aire*, on peut dire qu'ici la courbure de la structure de contact, considérée comme connexion, est l'aire.

3. Connexions et dérivation covariante : mini *vademecum*

3.1. Cadre général et variantes

3.1.1. Un calcul différentiel fibré.

Ce lien entre structures de contact et connexions nous fournit une occasion pertinente de préciser certains aspects de la théorie des connexions en géométrie différentielle, théorie fondamentale qui, certes, dépasse de très loin le propos de cet ouvrage mais qui, en même temps, se trouve au cœur des formalismes de fibration que nous utilisons et dont il est par conséquent justifié de parler avec un peu plus de généralité que ne l'exigent nos modèles élémentaires. Nous allons évoquer quelques notions de base qui sont "common knowledge" et que l'on trouve dans tous les manuels élémentaires et dans les articles d'introduction comme ceux de Wikipedia.

En un mot, on pourrait dire que la théorie des connexions est la généralisation du calcul différentiel aux fibrations. C'est, si l'on peut dire, un "calcul différentiel fibré". Son formalisme est omniprésent en physique où il sert de cadre à la fois à la relativité générale et à la théorie quantique des champs. Cela nous permet de plonger la neurogéométrie dans un cadre universel de type "physique" et d'approfondir l'analogie frappante, décelée depuis l'époque de la Gestalttheorie, entre la géométrie de la perception visuelle et la théorie des champs en physique. La neurogéométrie est à une "physique neuronale" de potentiels d'action dont les propagations sont régies par une connectivité implémentant des structures différentielles ce que les théories des champs sont à une physique de particules dont les trajectoires sont régies par des champs implémentant des structures différentielles.

La littérature sur les connexions est immense. Outre les références déjà citées à la section 7.2 du chapitre 2 à propos des formes différentielles (Arnold [20], Abraham-Marsden [2], Spivak [511], Guillemin [236], Bachman [23]), nous pouvons citer, parmi les grands classiques, le Charles Misner, Kip Thorne & John Wheeler [369], pour la relativité générale, Claude Itzykson & Jean-Bernard Zuber [266] et Chris Quigg [457] pour la théorie quantique des champs et les théories de jauge, évidemment Alain Connes [125], [126] pour le lien avec la géométrie non commutative. Nous recommandons aussi l'introduction [127] du physicien Robert Coquereaux disponible sur internet. Nous nous permettons également de référer à nos mémoires [419], [427], [436], [441], ainsi qu'à leurs bibliographies.

Il y a plusieurs façons de définir les connexions sur des fibrés; elles se sont précisées progressivement à partir de leur introduction par Cartan puis Weyl. Nous ne pouvons pas trop entrer dans les détails et nous resterons donc assez intuitifs et pas vraiment rigoureux. Un certain nombre de cas ont été développés avec de nombreuses variantes de définitions et de notations qui sont progressivement devenues traditionnelles dans différents domaines de référence. Nous allons en évoquer certains avec des notations qui sont plus conventionnelles que celles, spécifiquement adaptées à notre propos, que nous avons adoptées jusqu'ici.⁶

3.1.2. *Horizontal/vertical, externe/interne.*

Le contexte général est celui de fibrations $\pi : V \rightarrow M$ et il faut donc considérer la base M avec ses points x , les fibres V_x , et l'espace total V . Si l'on intuitionne M comme "horizontale", les fibres V_x sont intuitionnées comme "verticales" se projetant sur le point x . Souvent également, et il s'agit d'une autre analogie fondamentale, on considère que les éléments des fibres sont des entités "internes" alors que les points de la base sont des positions "externes". Un fibré décrit alors la façon dont des entités internes peuvent être paramétrées par des positions externes, ou, de façon équivalente, comment un domaine de l'espace externe M peut-être "rempli" par des qualités internes. Pour la perception visuelle, un exemple typique en est la fibration de l'espace des couleurs sur le champ visuel. Encore une fois, c'est exactement l'intuition des "engrafted variables" de David Hubel (Vol I, chapitre 4, section 4.3.2.) qui a joué un rôle fondateur dans les modèles neurogéométriques. Un exemple technologique analogue est la fibration d'un écran d'ordinateur dont les positions externes sont les pixels et les variables internes les codes RGB des couleurs. En physique (théorie quantique des champs) ces fibrations se rencontrent partout, l'espace externe étant l'espace-temps et les fibres les variables internes quantiques.

3.1.3. *La base M et ses fibrés de repères.*

6. Nous traiterons ici essentiellement des définitions classiques des connexions "à la Cartan". Des généralisations existent. Un exemple est celle proposée par Dmitri Alekseevsky et Peter Michor dans [14]. D'après les auteurs "the starting idea is to consider a Cartan connection as a deformation of a local Lie group structure on the manifold".

Le cas le plus simple est celui où l'on ne considère que la base M , sans variables internes. En chaque point $x \in M$, les cartes locales définissent des coordonnées traditionnellement notées x^μ avec des indices $\mu, \nu, \dots \in \{1, \dots, n = \dim M\}$ en position haute. Ces coordonnées locales définissent un repère $\left\{\frac{\partial}{\partial x^\mu}\right\}$ holonome du plan tangent $T_x M$, repère souvent noté $\{\partial_\mu\}$ pour simplifier. Lorsque M est munie d'un niveau de structure supplémentaire (par exemple une structure de variété riemannienne), on considère des repères associés à la structure supplémentaire des $T_x M$ (par exemple des repères orthonormés pour la structure euclidienne des $T_x M$ définissant la métrique) qui sont en général "non holonomes"⁷ Nous les noterons \mathbf{e}_μ (en accord avec une des conventions courantes). Un vecteur tangent $X \in T_x M$ s'écrira donc, suivant les cas, $X^\mu \partial_\mu$ ou $X^\mu \mathbf{e}_\mu$, ces expressions étant des sommes conformément à la règle de sommation d'Einstein disant que lorsque le même indice apparaît en position haute et basse on somme sur ses valeurs.⁸

Remarque sur les notations. Dans ce mini vademecum nous parlons de structures générales. Pour ne pas créer d'ambiguïtés éventuelles avec les structures très particulières dont nous traitons la plupart du temps, nous utiliserons pour certains éléments des fontes gothiques. _____ □

La caractéristique des repères holonomes $\{\partial_\mu\}$ est la commutativité $[\partial_\mu, \partial_\nu] = 0$ venant de l'égalité des dérivées secondes des fonctions $\frac{\partial^2 f}{\partial x^\mu \partial x^\nu} = \frac{\partial^2 f}{\partial x^\nu \partial x^\mu}$. Nous les appellerons "*h*-repères" (*h* pour "holonomes", on pourrait aussi dire "commutatifs"). Pour des repères non holonomes $\{\mathbf{e}_\mu\}$ compatibles avec une certaine structure, les crochets ne sont pas nuls et l'on a, puisque le crochet de deux vecteurs tangents est encore un vecteur tangent,

$$[\mathbf{e}_\mu, \mathbf{e}_\nu] = c_{\mu\nu}{}^\rho \mathbf{e}_\rho,$$

formule où les scalaires $c_{\mu\nu}{}^\rho$ sont appelés les "constantes de structure".⁹ Nous les appellerons "*nh*-repères" (*nh* pour "non holonomes", on pourrait aussi dire "non commutatifs"). La non-commutativité empêche qu'il existe d'autres coordonnées

7. Cartan appelait certains de ces repères des repères "mobiles", en particulier lorsqu'il s'agissait de repères orthonormés adaptés à des sous-espaces, comme les trièdres de Frenet le long d'une courbe, ou des repères constitués du plan tangent à une surface et de la normale, etc.

8. Dans cette notation conventionnelle, les vecteurs de base sont à droite et les composantes à gauche. Certains physiciens préfèrent l'ordre inverse $\mathbf{e}_\mu X^\mu$. Notons aussi que, jusqu'ici, nous n'avions pas tenu compte de la nature des indices : nous notions X_i les composantes de X et donc X^k une famille de X . Si l'on adopte la convention de cette section, il faut être plus rigoureux.

9. En fait, ces "constantes" varient en général avec x . On notera la position haute décalée à droite de l'indice ρ . Cela est dû au fait que, lorsqu'on calcule en termes de composantes (ce que font habituellement les physiciens, contrairement aux géomètres, comme, par exemple, Abraham-Marsden [2], qui préfèrent des calculs plus intrinsèques), on se retrouve vite confronté à un déluge d'indices et il est donc très important qu'ils soient bien positionnés afin de garantir une bonne lisibilité des formules et éviter les ambiguïtés.

locales y^μ telles que $\mathbf{e}_\mu = \frac{\partial}{\partial y^\mu}$, autrement dit on ne peut pas obtenir localement les \mathbf{e}_μ par changement de coordonnées locales.¹⁰ Avec la dérivée de Lie, la formule $[\mathbf{e}_\mu, \mathbf{e}_\nu] = c_{\mu\nu}{}^\rho \mathbf{e}_\rho$ s'écrit $\mathcal{L}_{\mathbf{e}_\mu} \mathbf{e}_\nu = c_{\mu\nu}{}^\rho \mathbf{e}_\rho$.

Quand on dualise, la base duale de $\{\partial_\mu\}$ devient la base duale $\{dx^\mu\}$ de T_x^*M et une base $\{\mathbf{e}_\mu\}$ devient une base duale $\{\mathbf{e}^\mu\}$. Une 1-forme A s'écrira donc $A_\mu dx^\mu$ ou $A_\mu \mathbf{e}^\mu$. Supposons que le nh -repère $\{\mathbf{e}_\mu\}$ satisfasse les équations de structure $[\mathbf{e}_\mu, \mathbf{e}_\nu] = c_{\mu\nu}{}^\rho \mathbf{e}_\rho$ et considérons les différentielles $d\mathbf{e}^\rho$ des 1-formes de base \mathbf{e}^ρ . On veut montrer qu'elles satisfont une équation duale. Sur les vecteurs tangents de base \mathbf{e}_μ (qui peuvent être interprétés comme des opérateurs de dérivation) on a, d'après la formule

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y])$$

de la section 7.2.7 du chapitre 2,

$$d\mathbf{e}^\rho(\mathbf{e}_\mu, \mathbf{e}_\nu) = \mathbf{e}_\mu \mathbf{e}^\rho(\mathbf{e}_\nu) - \mathbf{e}_\nu \mathbf{e}^\rho(\mathbf{e}_\mu) - \mathbf{e}^\rho(c_{\mu\nu}{}^\sigma \mathbf{e}_\sigma) .$$

Mais $\mathbf{e}^\rho(\mathbf{e}_\nu) = \delta_\nu^\rho$ est constant et donc sa dérivée $\mathbf{e}_\mu \mathbf{e}^\rho(\mathbf{e}_\nu)$ par \mathbf{e}_μ est nulle, et il en va de même pour $\mathbf{e}_\nu \mathbf{e}^\rho(\mathbf{e}_\mu)$. Il reste donc

$$d\mathbf{e}^\rho(\mathbf{e}_\mu, \mathbf{e}_\nu) = -c_{\mu\nu}{}^\rho .$$

D'un autre côté, calculons $c_{\sigma\tau}{}^\rho \mathbf{e}^\sigma \wedge \mathbf{e}^\tau$. On trouve

$$c_{\sigma\tau}{}^\rho \mathbf{e}^\sigma \wedge \mathbf{e}^\tau(\mathbf{e}_\mu, \mathbf{e}_\nu) = 2c_{\mu\nu}{}^\rho .$$

D'où l'expression de la formule de Maurer-Cartan de la section 10 du chapitre 5

$$d\mathbf{e}^\rho + \frac{1}{2} c_{\sigma\tau}{}^\rho \mathbf{e}^\sigma \wedge \mathbf{e}^\tau = 0$$

cette fois appliquée aux nh -repères et développée en termes de constantes de structure.

Pour les dérivées de Lie, on obtient, d'après les formules de la section 7.2.12 du chapitre 2,

$$\begin{aligned} \mathcal{L}_{\mathbf{e}_\mu} \mathbf{e}^\rho(\mathbf{e}_\nu) &= \mathbf{e}_\mu(\mathbf{e}^\rho(\mathbf{e}_\nu)) - \mathbf{e}^\rho([\mathbf{e}_\mu, \mathbf{e}_\nu]) = -\mathbf{e}^\rho(c_{\mu\nu}{}^\sigma \mathbf{e}_\sigma) = -c_{\mu\nu}{}^\rho \\ \mathcal{L}_{\mathbf{e}_\mu} \mathbf{e}^\rho &= c_{\nu\mu}{}^\rho \mathbf{e}^\nu . \end{aligned}$$

Remarque. La convention de sommation d'Einstein peut induire des ambiguïtés pour les k -formes différentielles $k \geq 2$. En effet, par exemple, une 2-forme s'écrit habituellement $\sum_{\mu < \nu} A_{\mu\nu} dx^\mu \wedge dx^\nu$. Si on l'écrit $A_{\mu\nu} dx^\mu \wedge dx^\nu$, on obtient les termes $A_{\mu < \nu, \nu} dx^\mu \wedge dx^\nu$ et $A_{\nu > \mu, \mu} dx^\nu \wedge dx^\mu = -A_{\nu > \mu, \mu} dx^\mu \wedge dx^\nu = A_{\mu < \nu, \nu} dx^\mu \wedge dx^\nu$ (à cause de l'antisymétrie des $A_{\mu\nu}$). On double donc la somme. Pour une k -forme, il

10. Évidemment, en un point particulier $x \in M$, il y a toujours des coordonnées locales y^μ telles que $\mathbf{e}_\mu = \frac{\partial}{\partial y^\mu}$ dans $T_x M$. Mais cette égalité ponctuelle ne peut pas être prolongée infinitésimalement et localement à cause de la non commutativité des \mathbf{e}_μ .

faut diviser par $k!$ la formule avec sommation d'Einstein pour obtenir la formule classique où les indices μ sont ordonnés par ordre croissant. \square

3.1.4. Fibrés principaux.

Au-delà des fibrés de repères, le cas de fibré le plus important est celui des fibrés *principaux* $\pi : \mathfrak{P} \rightarrow M$. Nous noterons en général \mathfrak{P} l'espace total et π la projection, éventuellement $\pi_{\mathfrak{P}}$ si nécessaire.¹¹ Dans notre modèle de référence \mathbb{V}_J , le groupe euclidien $SE(2)$ agit sur le plan de base et le sous-groupe des rotations $SO(2)$ agit sur les fibres et, comme nous l'expliquerons à la section 7 et plus en détail dans l'ensemble du chapitre 11, le fibré principal "associé" est le groupe euclidien $SE(2) \rightarrow \mathbb{R}^2$ lui-même, ce qui donnera notre second modèle noté \mathbb{V}_S . Dire que \mathfrak{P} est principal de groupe structural G (en général G est un groupe de Lie) c'est dire que G agit *transitivement* sur ses fibres \mathfrak{P}_x , autrement dit que \mathfrak{P}_x constitue une unique orbite de l'action de G . Quitte à quotienter G par le sous-groupe normal H des éléments agissant comme l'identité¹², on peut considérer que les fibres sont difféomorphes à G , comme par exemple lorsque les rotations de $SO(2)$ agissent sur les cercles. Pour des raisons que nous expliciterons plus bas, il est conventionnel de faire agir G à *droite* sur les fibres. Si donc $\mathfrak{p} \in \mathfrak{P}$ est un point de \mathfrak{P} et $g \in G$ un élément de G , l'action de g sur \mathfrak{p} sera notée $\mathfrak{p}g$. L'action de G est "verticale" et on peut faire agir le *même* $g \in G$ "en parallèle" sur toutes les fibres. Il s'agit de ce que les physiciens des théories de jauge appellent une "transformation de jauge globale", G étant appelé le "groupe de jauge global". Nous allons y revenir.

Mais on peut aussi faire agir différents $g(x)$ dans les différentes fibres \mathfrak{P}_x , ce que les physiciens appellent des "transformations de jauge locales". Cela se fait en deux étapes. Dans une première étape on considère des automorphismes verticaux Φ de \mathfrak{P} , c'est-à-dire des difféomorphismes verticaux qui préservent les fibres et qui sont compatibles avec l'action de G au sens où $\Phi(\mathfrak{p}g) = \Phi(\mathfrak{p})g$. Un tel Φ définit une application $\varphi : \mathfrak{P} \rightarrow G$ par la formule $\Phi(\mathfrak{p}) = \mathfrak{p}\varphi(\mathfrak{p})$ et comme $\Phi(\mathfrak{p}g) = \Phi(\mathfrak{p})g$ implique $\mathfrak{p}g\varphi(\mathfrak{p}g) = \mathfrak{p}\varphi(\mathfrak{p})g$, on doit avoir, puisque $\mathfrak{p}g_1 = \mathfrak{p}g_2$ implique $g_1 = g_2$, $g\varphi(\mathfrak{p}g) = \varphi(\mathfrak{p})g$, soit

$$\varphi(\mathfrak{p}g) = g^{-1}\varphi(\mathfrak{p})g.$$

Ce comportement par rapport à l'action de G s'appelle la propriété d'*equivariance*.

Comme par définition \mathfrak{P} est localement trivial, localement, une trivialisatoin de \mathfrak{P} au-dessus d'un voisinage U de $x \in M$ identifie $\pi^{-1}(U)$ au produit direct $U \times G$ où

11. Comme les possibilités de notations sont limitées mais qu'il faut néanmoins les diversifier, nous utiliserons dans cette section des caractères gothiques pour les fibrés.

12. Si $h \in H$, on a $h(\mathfrak{p}) = \mathfrak{p}$ pour tout $\mathfrak{p} \in \mathfrak{P}_x$. Si $g \in G$, alors

$$ghg^{-1}(\mathfrak{p}) = gh(g^{-1}(\mathfrak{p})) = gg^{-1}(\mathfrak{p}) = \mathfrak{p}$$

et donc $ghg^{-1} \in H$. Par conséquent H est un sous-groupe normal.

G agit sur lui-même par translations à droite. Mais dans de nombreux cas \mathfrak{P} n'est pas globalement trivial et n'est donc pas identifiable à $M \times G$.

Il faut bien insister sur le fait que, même si les fibres \mathfrak{P}_x sont difféomorphes à G , en l'absence du choix d'un *point distingué* \star_x de \mathfrak{P}_x pouvant être identifié à l'élément neutre e_G , \mathfrak{P}_x n'est pas *intrinsèquement* identifiable à G . Par exemple un cercle n'est identifiable au groupe des rotations $SO(2)$ qu'une fois choisi un point comme origine des angles. À cause de la trivialité locale, le choix de points distingués \star_x peut se faire au moyen de sections locales $\sigma : U \rightarrow \mathfrak{P}$ sélectionnant des $\sigma(x) = \star_x$. De telles sections, appelées "jauge locale" par les physiciens, définissent des trivialisations locales $U \times G$ de \mathfrak{P} . En effet, si $x \in U$ et $\mathfrak{p} \in \mathfrak{P}_x$, il existe un seul $g_{\mathfrak{p}} \in G$ tel que $\star_x g_{\mathfrak{p}} = \mathfrak{p}$ et l'on peut donc coder \mathfrak{p} par $(x, g_{\mathfrak{p}})$. Les transformations de jauge locale consistent alors simplement, dans une deuxième étape, à changer les sections $\sigma(x) = \star_x$ en faisant opérer à droite un $g(x)$ approprié. Elles sont définies par une application différentiable $\varphi_M : M \rightarrow G$ qui "descend" φ de \mathfrak{P} sur M grâce à la propriété d'équivariance en posant $\varphi_M(x) = \varphi(\star_x)$. On prend alors $g(x) = \varphi_M(x)$.

On notera que les sections trivialisantes σ permettent de définir une action à gauche de G (qui dépend de σ). L'action à droite reste $(x, g_{\mathfrak{p}})g = (x, g_{\mathfrak{p}}g)$ et l'action à gauche est $g(x, g_{\mathfrak{p}}) = (x, gg_{\mathfrak{p}})$. Bien sûr, ces deux actions commutent.

Remarque. Il n'existe pas forcément de trivialisation globale de \mathfrak{P} et il faudra donc *recoller* des choix de jauge locale. Cela conduit à des problèmes de *cohomologie* des groupes au sens explicité à la section 9.4 du chapitre 2. Il s'agit de la cohomologie de Čech du groupe structural G . En particulier les trivialisations locales de \mathfrak{P} sur un recouvrement ouvert adapté $\mathcal{U} = \{U_i\}$ définissent un cocycle à coefficients dans G et \mathfrak{P} est globalement trivial si et seulement si ce cocycle est un cobord. \square

En plus des transformations de jauge locale définies par des automorphismes verticaux de \mathfrak{P} on peut introduire des transformations plus générales qui sont des difféomorphismes *échangeant* les fibres au moyen d'un difféomorphisme ψ de la base M . Cela signifie que si $\mathfrak{p} \in \mathfrak{P}_x$, $\Phi(\mathfrak{p}) \in \mathfrak{P}_{\psi(x)}$, la compatibilité avec l'action de G étant toujours $\Phi(\mathfrak{p}g) = \Phi(\mathfrak{p})g$.

Le groupe de Lie G agit infinitésimalement et verticalement sur les fibres de \mathfrak{P} à travers l'algèbre de Lie \mathcal{G} . Notons \mathfrak{g}_α les éléments d'une base de \mathcal{G} . Un élément \mathfrak{g} de \mathcal{G} s'écrira donc $\mathfrak{g} = \xi^\alpha \mathfrak{g}_\alpha$. En tout point $\mathfrak{p} \in \mathfrak{P}$, chaque $\mathfrak{g} \in \mathcal{G}$ induit un vecteur tangent "vertical" $\mathfrak{t}_{\mathfrak{g}} \in T_{\mathfrak{p}}\mathfrak{P}$ que l'on notera aussi $\overrightarrow{\mathfrak{g}}(\mathfrak{p})$. On définit ainsi sur \mathfrak{P} des champs $\overrightarrow{\mathfrak{g}}$ (aussi dits "fondamentaux") qui sont G -invariants à droite. Les $\overrightarrow{\mathfrak{g}_\alpha}$ forment une base des espaces tangents $T_{\mathfrak{p}}\mathfrak{P}$. Si l'on choisit une trivialisation locale $\sigma : U \rightarrow \mathfrak{P}$ alors il y a aussi une action à gauche de G et on obtient, du moins localement, des champs $\overleftarrow{\mathfrak{g}}(\mathfrak{p})$ G -invariants à gauche qu'il faut ensuite recoller.

Les vecteurs tangents généraux $\mathfrak{t} \in T_{\mathfrak{p}}\mathfrak{P}$ à \mathfrak{P} en \mathfrak{p} ont une composante "verticale" tangente à la fibre \mathfrak{P}_x dans $T_{\mathfrak{p}}(\mathfrak{P}_x)$. Ils ont par ailleurs une *projection* "horizontale"

$d\pi(\mathfrak{t})$ dans l'espace tangent de la base $T_{\pi(\mathfrak{p})}M$. Mais cela ne signifie pas pour autant qu'ils ont une composante "horizontale" dans $T_{\mathfrak{p}}\mathfrak{P}$. C'est même précisément le rôle des connexions que d'en fournir une.

Les connexions dites "principales" sur un fibré principal $\pi : \mathfrak{P} \rightarrow M$ seront définies par des formes ayant des propriétés particulières. Mais nous pouvons déjà remarquer ici que toute forme définie sur la base M peut être relevée dans \mathfrak{P} . En effet l'application linéaire tangente de π est une application linéaire π_* de $T\mathfrak{P}$ dans TM qui consiste à projeter $T_{\mathfrak{p}}\mathfrak{P}$ sur $T_{\pi(\mathfrak{p})}M$. Soit alors ω une k -forme sur M . On en déduit immédiatement une k -forme $\omega^{\mathfrak{P}}$ sur \mathfrak{P} en posant $\omega^{\mathfrak{P}}(\mathfrak{t}^1, \dots, \mathfrak{t}^k) = \omega(\pi_*(\mathfrak{t}^1), \dots, \pi_*(\mathfrak{t}^k))$. Ces formes $\omega^{\mathfrak{P}}$ relevant les formes de la base sont très spéciales. D'abord elles sont *horizontales* au sens où elles s'annulent dès que l'un de leurs arguments \mathfrak{t}^i est vertical, car alors $\pi_*(\mathfrak{t}^i) = 0$. Il est équivalent de dire que leur produit intérieur $i_{\vec{\mathfrak{g}}}\omega^{\mathfrak{P}}$ avec n'importe quel $\vec{\mathfrak{g}}$ est nul. Ensuite elles sont *invariantes* au sens où leur dérivée de Lie $\mathcal{L}_{\vec{\mathfrak{g}}}\omega^{\mathfrak{P}}$ est toujours nulle. Cela découle immédiatement de la formule de Cartan (chapitre 2, section 7)

$$\mathcal{L}_{\vec{\mathfrak{g}}}\omega^{\mathfrak{P}} = i_{\vec{\mathfrak{g}}}d\omega^{\mathfrak{P}} + d(i_{\vec{\mathfrak{g}}}\omega^{\mathfrak{P}})$$

car $i_{\vec{\mathfrak{g}}}\omega^{\mathfrak{P}} = 0$ et, $d\omega^{\mathfrak{P}}$ étant également horizontale, $i_{\vec{\mathfrak{g}}}d\omega^{\mathfrak{P}} = 0$.

3.1.5. Fibrés vectoriels.

Les autres fibrés les plus courants sont les fibrés vectoriels $\pi_{\mathfrak{V}} : \mathfrak{V} \rightarrow M$ dont les fibres sont des espaces vectoriels \mathfrak{V}_x tous isomorphes à une fibre type \mathfrak{W} . Souvent, ils sont construits à partir d'un fibré principal $\pi : \mathfrak{P} \rightarrow M$ de groupe structural G agissant à droite sur les fibres \mathfrak{P}_x et d'une action à gauche de G sur \mathfrak{W} , c'est-à-dire d'une *représentation* (un morphisme de groupes ¹³) $\rho : G \rightarrow GL(\mathfrak{W})$. On considère alors les couples $(\mathfrak{p}, \mathfrak{w})$ à l'équivalence près par l'action de G . Autrement dit, si $g \in G$, on identifie $(\mathfrak{p}, \mathfrak{w})$ et $(\mathfrak{p}g, \rho(g^{-1})\mathfrak{w})$ (on peut écrire $\rho(g^{-1})\mathfrak{w} = g^{-1}\mathfrak{w}$ pour simplifier). Comme g agit à droite par $\mathfrak{p}g$ sur \mathfrak{p} et à gauche par $g^{-1}\mathfrak{w}$ sur \mathfrak{w} , ces deux actions s'annulent sur $(\mathfrak{p}, \mathfrak{w})$. On notera que G n'agit plus naturellement à droite sur \mathfrak{V} .

Si une trivialisations locale $\sigma : U \rightarrow \mathfrak{P}$ sélectionne des $\sigma(x) = \star_x$ dans les fibres \mathfrak{P}_x de \mathfrak{P} , alors on peut faire agir $g_{\mathfrak{p}}^{-1}$ sur $(\mathfrak{p}, \mathfrak{w})$ pour obtenir

$$(\mathfrak{p}g_{\mathfrak{p}}^{-1}, \rho(g_{\mathfrak{p}})\mathfrak{w}) = (\star_x, \rho(g_{\mathfrak{p}})\mathfrak{w}) ,$$

ce qui élimine la référence à \mathfrak{P} et se ramène à l'action $\rho(G)$ sur les fibres $\mathfrak{V}_x \simeq \mathfrak{W}$. Mais il faut tenir compte du fait que cette simplification dépend des trivialisations locales, c'est-à-dire des choix de jauges locales, et que, souvent, il n'existe pas de trivialisations globale. Comme le note Robert Coquereaux ([127], p. 115), ce n'est pas G qui agit naturellement sur \mathfrak{V} mais le groupe de jauge.

13. Nous reviendrons longuement sur cette notion centrale dans le chapitre 17, section 1.

Le fibré tangent TM en est un exemple typique ainsi que tous les fibrés tensoriels qui s'en déduisent, en particulier les fibrés de formes différentielles. Le fibré tangent TM a pour groupe naturel d'automorphismes des fibres T_xM le groupe linéaire $GL(n)$ (si M est de dimension n). On peut le munir, répétons-le, de niveaux de structure plus ou moins forts, par exemple une structure euclidienne si M est riemannienne et alors T_xM aura pour groupe d'automorphismes le groupe orthogonal $O(n)$, réduit à $SO(n)$ si M est orientable.

Ces fibrés vectoriels construits à partir de TM ont toutefois une propriété extrêmement spéciale : les éléments “verticaux” dans les fibres sont en fait des déplacements “horizontaux” dans la base et l’“interne” se dialectise avec l’“externe” ! Nous allons y revenir car il s'agit d'un point clé (en particulier en relativité générale).

3.1.6. Symétries internes.

Les fibrés vectoriels $\pi : \mathfrak{V} \rightarrow M$, comme ceux de la théorie quantique des champs, sont des fibrés dont les fibres sont de “vrais” espaces internes associés à un fibré principal dont le groupe structural G exprime des “symétries internes”. Ils sont caractérisés par une *représentation* (c'est-à-dire un morphisme de groupes) $\rho : G \rightarrow GL(\mathfrak{W})$ de G dans le groupe d'automorphismes de leur fibre type \mathfrak{W} .

3.1.7. Fibrés quelconques.

On rencontre aussi parfois des fibrés quelconques dont les fibres n'ont pas forcément de structure linéaire, et sont sans groupe structural.

3.2. Transport parallèle infinitésimal

Comme nous l'avons dit et redit, la notion de connexion en géométrie différentielle consiste à définir des *transports parallèles infinitésimaux* reliant entre elles les fibres d'un fibré. Nous allons un peu l'explicitier en insistant sur son élaboration chez Cartan en termes de géométrie différentielle “pure”. Un autre des pères fondateurs de cette notion est Hermann Weyl mais celui-ci se focalisait surtout le problème de l'espace physique (“das Raumproblem”, cf. plus bas la section 3.2.2) comme en relativité générale.¹⁴

14. Hermann Weyl (1885-1955) fut l'un des plus prodigieux scientifiques de la première moitié du XX^e siècle, à la fois très grand géomètre, physicien inspiré et profond métaphysicien (fortement inspiré par Kant et Fichte, puis surtout Husserl, et aussi Cassirer). Comme s'amuserent à le souligner de façon très “bourbachique” Claude Chevalley et André Weil dans leur hommage de 1957 [116], il était “bon romantique”, aimait les “analogies” et la “métaphysique allemande”. Formé de 1904 à 1913 à Göttingen par David Hilbert et Hermann Minkowski (qui furent les directeurs de sa thèse de 1908) et aussi Felix Klein, il commença à travailler sur la théorie des opérateurs (en particulier auto-adjoints). De 1911 à 1912 il rendit complètement rigoureux le concept de surface de Riemann en se fondant sur le concept de prolongement analytique permettant de passer du local au global (*Die Idee der Riemannschen Fläche* [561], nous en avons déjà parlé dans une note de la section 4.2 du chapitre 2). En 1913, il fut nommé

3.2.1. *L'intuition géométrique initiale.* Il a été un peu délicat de bien stabiliser les définitions. L'intuition initiale de Cartan que nous avons évoquée plus haut était celle des surfaces S plongées dans l'espace euclidien \mathbb{R}^3 et munies de leur métrique induite. Dans ce cas, les plans tangents $T_s S$ sont des plans *affines* dans \mathbb{R}^3 devenant des plans vectoriels quand le point de contact s est pris pour origine. Pour connecter deux plans tangents voisins on fait alors "rouler sans glissement" le premier $T_s S$ sur le second $T_{s+\delta s} S$. On a bien un isomorphisme entre les deux, mais comme l'origine de $T_{s+\delta s} S$ est $s+\delta s$ et non pas l'origine s de $T_s S$, cet isomorphisme est *affine* et non pas linéaire, d'où d'ailleurs le nom de connexion "affine". Quand on essaye d'abstraire cette situation en éliminant l'espace ambiant \mathbb{R}^3 , en traitant S comme une variété bidimensionnelle *non plongée* et en traitant les espaces tangents $T_s S$ comme des espaces vectoriels *intrinsèques*, il faut recoder dans ce nouveau contexte le caractère affine de la connexion. Nous allons voir comment.

Remarque. Nous constaterons que les formalismes sont très analogues aux structures de contact : noyaux de 1-formes, relèvements de courbes, etc. Mais, comme nous l'avons remarqué plus haut, section 2, les fibrations qui interviennent sont très différentes. Dans notre cas de référence, la structure de contact est liée à la projection sur le plan (x, y) et la connexion à la projection duale sur le plan (x, p) . En tant que K_v , les plans de la distribution $\mathcal{K} = \mathcal{C}$ se projettent sur des *droites* du plan de base $\mathbb{R}^2_{(x,y)}$, ne sont donc pas du tout des isomorphismes et contiennent les fibres. Au contraire, en tant que C_v , ils se projettent isomorphiquement sur le plan de base $\mathbb{R}^2_{(x,p)}$ et sont *transverses* aux fibres □

3.2.2. *Das Raumproblem.* La notion de connexion telle qu'elle a été précisée de façon puissante par Cartan et Weyl approfondit considérablement le *Raumproblem*

à l'ETH de Zürich (l'École Polytechnique) et y rencontra Einstein. Il s'ensuivit ses travaux remarquables sur la relativité générale (*Raum, Zeit, Materie*, 1918, cf. [563]). Il y travailla aussi sur les fondements des mathématiques et publia, également en 1918, son célèbre *Das Kontinuum : Kritische Untersuchungen über die Grundlagen der Analysis* [562], conception intuitionniste et constructiviste, anti-cantorienne, du continu. En 1921 il rencontra Schrödinger et se passionna pour les débuts de la mécanique quantique et les possibilités d'y appliquer la théorie des groupes et la théorie des représentations de groupes par des opérateurs dans des espaces de Hilbert. Son grand mémoire de 1926 (*Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare Transformationen, I, II, III*, [564]) sur les représentations irréductibles des groupes de Lie compacts semi-simples, qu'il considérait comme son œuvre la plus importante, est effectivement un classique. Et en 1928, il publia son fameux *Gruppentheorie und Quantenmechanik* [565]. Après avoir refusé en 1923 la succession de Klein à Göttingen, il accepta en 1930, à la demande expresse de ce dernier, la succession de Hilbert, mais la dérive dramatique de l'Allemagne le conduisit à s'installer en 1933 à Princeton. Pendant une vingtaine d'années il continua à approfondir ses réflexions mathématiques-physiques-métaphysiques dans de grands textes comme *Mind and Nature* (1934), *The classical groups, their invariants and representations* (1939), *Philosophy of mathematics and natural sciences* (1949), *Symmetry* (1951). Les articles de la *Stanford Encyclopedia* et de *Wikipedia* sur Weyl sont excellents.

de Riemann-Helmholtz-Lie dont nous avons parlé d'emblée dans le Vol I (chapitre 2, section 2.1.) en nous référant au beau texte de Joël Merker [362]. Il s'agit de penser l'origine de la métrique de l'espace physique après que Riemann ait découplé la structure différentiable d'une variété M de ses possibles propriétés métriques. Sur le plan philosophique, c'est Hermann Weyl qui a sans doute le mieux formulé le problème comme dépassement de la doctrine kantienne de l'espace et du temps. Avec Riemann, l'élément infinitésimal de longueur ds doit être, dans des coordonnées locales (x_i) appropriées, de la forme $ds = \left(\sum_i (dx_i)^k\right)^{\frac{1}{k}}$ la solution "pythagoricienne" $k = 2$ étant la plus simple. Helmholtz déduisit cette dernière des propriétés des mouvements des corps rigides. Plus tard (1921¹⁵) Hermann Weyl démontra que les seules algèbres de Lie satisfaisant ces contraintes sont les algèbres de Lie des transformations orthogonales par rapport à une forme quadratique non dégénérée.

La relativité générale raviva avec une force inouïe l'actualité de ce problème et, aussi bien chez Cartan que chez Weyl, conduisit à la repenser en termes de connexions, en l'occurrence en termes de transports parallèles infinitésimaux reliant entre elles les fibres $T_x M$ du fibré tangent. Pour Weyl, une métrique riemannienne découle

(i) de l'existence en chaque point de coordonnées locales "géodésiques" relativement auxquelles les composantes d'un vecteur tangent restent constantes par transport parallèle infinitésimal, et en plus

(ii) du fait que la longueur d'un tel vecteur reste également constante, i.e. que la formule du ds reste la même dans ces coordonnées, ce qui relie la métrique au transport parallèle comme nous allons le voir. Ainsi qu'y insiste Thomas Hawkins [242]¹⁶, en géométrie riemannienne on peut comparer de façon *absolue* les longueurs des vecteurs tangents en des points différents de l'espace et cette géométrie n'est donc pas "vraiment" *infinitésimale*. C'est d'ailleurs pourquoi, selon Cartan et Weyl, on peut aller au-delà dans la recherche fondationnelle d'une géométrie "vraiment" infinitésimale.

Il est essentiel de comprendre qu'il existe *deux* façons complémentaires d'enrichir la structure différentiable d'un espace :

1. par des déplacements infinitésimaux *parallèles* des directions,
2. par des déplacements infinitésimaux *congruents* des longueurs.

L'intuition euclidienne couple les deux dans un espace vectoriel au moyen d'un produit scalaire qui définit *à la fois* des angles (directions) et des normes (longueurs). L'intuition riemannienne localise l'intuition euclidienne et du coup permet de continuer à comparer des longueurs mais *ne permet plus* de comparer des orientations

15. Dans la quatrième édition de *Raum, Zeit, Materie*.

16. Cet article de Thomas Hawkins est une excellente synthèse historique des débats entre Einstein, Weyl et Cartan sur le *Raumproblem*.

en tant que telles, sauf à les subordonner à la métrique au moyen de la connexion métrique de Levi-Civita. C'est elle qui se trouve implémentée *physiquement* en relativité générale.

Que la géométrie de l'espace-temps soit riemannienne n'est donc pas une nécessité. Weyl ; qui manifestait un

“deep seated, philosophically inclined interest in the mathematical foundations of theoretical physics.” (Hawkins, [242], p. 76)

est connu pour avoir envisagé, au moyen du transport parallèle, la possibilité d'une géométrie *conforme* de l'espace-temps où les mesures des longueurs pourraient changer d'échelle en chaque point mais pas la mesure des angles. C'est l'origine des théories de jauge dont Weyl espérait pouvoir dériver une théorie unifiée de la relativité générale et de l'électro-magnétisme. Il abandonnera cet espoir et élaborera ensuite la première théorie de jauge acceptable non plus en introduisant en chaque point un changement d'échelle dans la mesure des longueurs mais un changement de la phase de la fonction d'onde de l'électron dans la théorie quantique de Dirac du champ électromagnétique. Nous allons y revenir dans la section 4.2.

Nous verrons également à la section 6.5 qu'Einstein tenta également une unification en introduisant un “parallélisme à distance” ou “parallélisme absolu” (“Fern-Parallelismus”) permettant de comparer les directions en des points différents de l'espace-temps. Mais Weyl estimait que le caractère “à distance” du “Fernparallelismus” était incompatible avec le principe imposant la construction de la géométrie physique exclusivement à partir de transformations infinitésimales.

Weyl est l'un des plus grands penseurs de la *géométrisation* de la physique (cf. des textes comme celui de Christophe Eckes [162]). En ce qui concerne la métaphysique de l'espace physique, il est particulièrement intéressant parce qu'il connaissait très bien Leibniz, Kant, Husserl et Cassirer et n'était donc pas métaphysiquement analphabète.¹⁷ Pour lui, la structure C^∞ de l'espace reste *a priori*, ainsi que le *type* de la métrique (il maintient donc une forme de synthétique a priori kantien mais purement *local*), tandis que le *champ* métrique induit par la distribution de la matière et de

17. Pour la contribution d'Hermann Weyl à la compréhension des fondements de la géométrie, le lecteur pourra se référer également aux textes d'Armand Borel [65], Claude Chevalley et André Weil [116], Erhard Scheibe [488], d'Erhard Scholz [490], [491].

l'énergie (le "Strukturfeld") est, quant à lui, *a posteriori*.¹⁸ Après une vaste unification des révolutions successives de la compréhension de la géométrie non seulement chez Riemann, mais aussi chez Frobenius, Schur, Hurwitz, Lie, Cartan, il est passé, sous l'impulsion de la géométrisation de la mécanique quantique, à la théorie des groupes et aux théories de jauge.

Dans son ouvrage fondamental *The Reign of Relativity* [474] ainsi que dans sa belle contribution "Hermann Weyl and 'First Philosophy' : Constituting Gauge Invariance" dans *Constituting Objectivity* [475], Tom Ryckman insiste sur le primat du transport parallèle chez Weyl :

"To Weyl, parallel transport was the paradigm comparison relation of infinitesimal geometry for it satisfied the epistemological demand that all integral (and so, not immediately surveyable) relations between finitely separated points cannot be posited but must be constructed from a specified infinitesimal displacement along a given curve connecting them." ([474], p. 289)

Il insiste sur l'influence de la philosophie transcendantale de Kant et Husserl sur la façon dont Weyl pensait les fondements de la géométrie :

"transcendental idealism informed Weyl's construction of a 'purely infinitesimal geometry'." ([475], p. 279)

Il rappelle (p. 292-293) que

"writing to Husserl on 26 March 1921, Weyl could report that he had finally captured the '*a priori* essence of space' (*apriorische Wesen des Raumes*), through a notable deepening of its mathematical foundations (*eine merkliche Tieferlegung der Fundamente*)'."

3.3. Connexions générales et espaces horizontaux

Précisons maintenant toutes ces idées un peu plus techniquement. La définition géométriquement la plus intuitive d'une connexion \mathcal{C} est celle, due à Charles Ehresmann, que nous avons utilisée plus haut. Étant donné un fibré général $\pi : \mathfrak{P} \rightarrow M$, elle consiste à définir \mathcal{C} comme la donnée, en chaque point $\mathfrak{p} \in \mathfrak{P}$, d'un relèvement $C_{\mathfrak{p}}$ dans $T_{\mathfrak{p}}\mathfrak{P}$ de l'espace tangent de la base $T_{\pi(\mathfrak{p})}M$, relèvement qui doit être *transverse aux fibres* et est alors appelé l'espace "horizontal". Les espaces horizontaux $C_{\mathfrak{p}}$ et

18. Nous défendons depuis longtemps une thèse analogue (cf. par exemple [419]). Par rapport à la physique newtonienne, le synthétique a priori chez Kant est le nom philosophique du fait que l'espace-temps = "espace euclidien global" \times "temps universel" fonctionne pour la réalité physique empirique comme ce que l'on appelle aujourd'hui une "background structure" munie de son groupe de relativité (relativité galiléenne). Il suffit de lire la *Phoronomie des Metaphysische Anfangsgründe der Naturwissenschaft* (1786, [280]) pour s'en convaincre. Toute la physique moderne repose sur du synthétique a priori. Celui-ci est simplement plus sophistiqué mathématiquement que le groupe de Galilée.

les espaces verticaux $T_{\mathfrak{p}}(\mathfrak{F}_{x=\pi(\mathfrak{p})})$ tangents aux fibres forment une décomposition en somme directe $T_{\mathfrak{p}}\mathfrak{F} = T_{\mathfrak{p}}(\mathfrak{F}_{x=\pi(\mathfrak{p})}) \oplus C_{\mathfrak{p}}$, $\mathfrak{t} = \mathfrak{t}_{ver} + \mathfrak{t}_{hor}$, du fibré tangent $T\mathfrak{F}$ du fibré \mathfrak{F} . On obtient ainsi une distribution “horizontale”. Elle définit une *projection* $\omega_{\mathcal{C}}$ du fibré tangent $T_{\mathfrak{p}}\mathfrak{F}$ sur la distribution “verticale” des $T_{\mathfrak{p}}(\mathfrak{F}_{x\pi(\mathfrak{p})})$ parallèlement aux sous-espaces horizontaux $C_{\mathfrak{p}}$, projection qui est évidemment idempotente et nulle sur les vecteurs tangents horizontaux. On peut considérer $\omega_{\mathcal{C}}$ comme la 1-forme sur \mathfrak{F} à valeurs dans les espaces verticaux $T_{\mathfrak{p}}(\mathfrak{F}_{x=\pi(\mathfrak{p})})$ qui associe à $\mathfrak{t} \in T_{\mathfrak{p}}\mathfrak{F}$ sa composante verticale $\omega_{\mathcal{C}}(\mathfrak{t}) = \mathfrak{t}_{ver}$. On a bien évidemment $\omega_{\mathcal{C}}(\mathfrak{t}) = 0$ si \mathfrak{t} est horizontal et $\omega_{\mathcal{C}}(\mathfrak{t}) = \mathfrak{t}$ si \mathfrak{t} est vertical.

La distribution \mathcal{C} étant très peu contrainte, il n’y a aucune raison pour que le crochet de Lie de deux champs de vecteurs horizontaux $\mathfrak{t}, \mathfrak{u} \in \Gamma(\mathcal{C})$ soit lui-même horizontal.¹⁹ En général, $\Gamma(\mathcal{C})$ n’est pas une sous-algèbre de Lie de $\mathfrak{X}(\mathfrak{F}) = \Gamma(T\mathfrak{F})$. Ce phénomène se manifeste par la *courbure* de \mathcal{C} . La courbure $\Omega_{\mathcal{C}}$ est la 2-forme sur \mathfrak{F} à valeurs dans les espaces verticaux $T_{\mathfrak{p}}(\mathfrak{F}_x)$ qui associe à un couple $\mathfrak{t}, \mathfrak{u} \in \mathfrak{X}(\mathfrak{F})$ de champs de vecteurs sur \mathfrak{F} , le champ vertical $\Omega_{\mathcal{C}}(\mathfrak{t}, \mathfrak{u}) = [\mathfrak{t}_{hor}, \mathfrak{u}_{hor}]_{ver}$ qui est la composante verticale du crochet des composantes horizontales des champs. Nous allons y revenir plus en détail. On voit que la courbure est identiquement nulle si et seulement si on a toujours $[\mathfrak{t}_{hor}, \mathfrak{u}_{hor}]_{ver} = 0$, autrement dit si et seulement si $[\mathfrak{t}_{hor}, \mathfrak{u}_{hor}]$ est horizontal et donc si $\Gamma(\mathcal{C})$ est une sous-algèbre de Lie de $\mathfrak{X}(\mathfrak{F}) = \Gamma(T\mathfrak{F})$.

La donnée de \mathcal{C} permet de définir automatiquement un *transport parallèle* le long des chemins $\gamma(t) : x_0 \rightarrow y$ dans la base M , c’est-à-dire des isomorphismes $\mathfrak{h}_{\gamma(t)}$ entre les fibres \mathfrak{F}_{x_0} et $\mathfrak{F}_{\gamma(t)}$ qui se composent en accord avec la composition des chemins. En effet, soit $\dot{\gamma}(0)$ le vecteur tangent à $\gamma(t)$ en x_0 et soit \mathfrak{p}_0 un point de \mathfrak{F} au-dessus de x_0 . Parmi tous les vecteurs tangents de $T_{\mathfrak{p}_0}\mathfrak{F}$ se projetant sur $\dot{\gamma}(0)$ il n’y en a qu’un, $\dot{\Gamma}(0)$, qui soit horizontal. On peut ainsi relever $\gamma(t)$ en une courbe horizontale dans \mathfrak{F} à partir de tout point initial $\mathfrak{p}_0 \in \mathfrak{F}_{x_0}$. Si $y = x_0$, i.e. si γ est un lacet, \mathfrak{h}_{γ} est donc un automorphisme de la fibre \mathfrak{F}_{x_0} munie de son niveau de structure spécifique. On obtient ainsi un sous-groupe (de Lie) \mathfrak{H} du groupe (de Lie) des automorphismes G de la fibre (sous-groupe défini à conjugaison près quand on change de point base x_0) appelé le groupe d’*holonomie*. Si l’on se restreint aux lacets *contractiles*, on obtient un sous-groupe \mathfrak{H}_0 de \mathfrak{H} . Bien sûr $\mathfrak{H}_0 = \mathfrak{H}$ si M est simplement connexe. On obtient alors un morphisme de groupes *surjectif* entre le groupe fondamental (ou groupe de Poincaré) $\pi_1(M)$ de M (les lacets à homotopie près) et le quotient $\mathfrak{H}/\mathfrak{H}_0$.²⁰

19. Rappelons que $\Gamma(T\mathfrak{F})$ (aussi noté $\mathfrak{X}(\mathfrak{F})$) est l’espace fonctionnel des champs de vecteurs tangents sur \mathfrak{F} , c’est-à-dire des sections du fibré tangent $T\mathfrak{F}$. $\Gamma(\mathcal{C})$ est l’espace des sections de la distribution \mathcal{C} i.e. des champs de vecteurs de $\Gamma(T\mathfrak{F})$ qui sont partout tangents à \mathcal{C} , autrement dit qui sont horizontaux.

20. Introduit par Poincaré en 1892, la notion de groupe fondamental est une notion princeps de la *topologie algébrique* permettant de définir des invariants topologiques par des structures

La connexion est “plate” ou de courbure nulle si \mathfrak{H}_0 est trivial. C’est le cas pour la connexion Λ de Maurer-Cartan des groupes de Lie. Cette propriété est locale car \mathfrak{H}_0 étant défini par des lacets contractiles, ceux-ci peuvent être aussi petits que l’on veut.

3.4. Dérivation covariante (prélude)

Une connexion \mathcal{C} permet de redéfinir le calcul différentiel en utilisant le transport parallèle par les plans horizontaux de \mathcal{C} . Par exemple, pour la dérivée extérieure des formes différentielles, l’idée clé est de définir la nouvelle opération de dérivation, dite *dérivation covariante*, comme la dérivée extérieure standard restreinte aux \mathcal{C} -composantes des vecteurs tangents. Plus précisément, si $\theta(\mathbf{t}_1, \dots, \mathbf{t}_k)$ est une k -forme sur \mathfrak{P} , sa dérivée covariante $D\theta$ est donnée par

$$D\theta(\mathbf{t}_1, \dots, \mathbf{t}_{k+1}) = d\theta(\mathbf{t}_{1,hor}, \dots, \mathbf{t}_{k+1,hor})$$

où \mathbf{t}_{hor} est la projection de $\mathbf{t} \in T_p\mathfrak{P}$ sur le plan horizontal C_p . Il en va de même pour d’autres entités géométriques de l’analyse vectorielle classique que nous avons évoquées à la section 8.1 du chapitre 2. Par exemple, si $f : \mathfrak{P} \rightarrow \mathbb{R}$ est une fonction à valeurs réelles sur \mathfrak{P} , son gradient relatif à \mathcal{C} est donné par

$$\nabla_{\mathcal{C}}(f) = \sum \mathbf{t}_{hor,i}(f)\mathbf{t}_{hor,i}$$

algébriques sur lesquelles on peut calculer. Si M est un espace topologique connexe par arcs, son groupe fondamental $\pi_1(M)$ consiste à traiter les lacets γ_x basés en un point base $x \in M$ (c’est-à-dire les applications continues $\gamma_x : [0, 1] \rightarrow M$ telles que $\gamma_x(0) = \gamma_x(1) = x$) à équivalence d’homotopie près, γ_x^2 étant homotope à γ_x^1 si l’on peut déformer continûment γ_x^2 en γ_x^1 . Les lacets peuvent se composer par concaténation, $\gamma_x^2 \circ \gamma_x^1$ étant défini par $\gamma_x^2 \circ \gamma_x^1(t) = \gamma_x^1(2t)$ si $t \in [0, \frac{1}{2}]$ et $\gamma_x^2 \circ \gamma_x^1(t) = \gamma_x^1(2t - 1)$ si $t \in [\frac{1}{2}, 1]$. La composition est compatible à l’équivalence et définit une structure de groupe (i) dont l’élément neutre est la classe du lacet constant e_x , c’est-à-dire la classe des lacets pouvant se contracter de façon continue sur e_x , et (ii) où l’inverse de γ_x est γ_x parcouru en sens inverse. Si l’on change de point base, on obtient un $\pi_1(M)$ isomorphe : il suffit de choisir un chemin α de x à y (qui existe car M est connexe par arcs) et d’associer au lacet γ_y basé en y le lacet $\gamma_x = \alpha^{-1} \circ \gamma_y \circ \alpha$ basé en x . $\pi_1(M)$ compte en quelque sorte le nombre de tours que font les lacets autour de “trous” de M . $\pi_1(M)$ est trivial lorsque M est “sans trous”, c’est-à-dire simplement connexe : tout lacet peut se contracter sur le lacet constant. Pour le cercle, $\pi_1(\mathbb{S}^1) = \mathbb{Z}$ compte le nombre de tours qu’accomplit le lacet autour du cercle dans le sens direct ou indirect. Pour le tore, $\pi_1(\mathbb{T}) = \mathbb{Z} \times \mathbb{Z}$ compte le nombre de tours qu’accomplit le lacet déformé en une concaténation de méridiens et de parallèles. Une propriété fondamentale du π_1 est d’être un *foncteur* de la catégorie des espaces topologiques connexes par arcs dans la catégorie des groupes. Si $f : M \rightarrow N$ est une application continue et si $\gamma_x : [0, 1] \rightarrow M$ est un lacet de M basé en x , $f \circ \gamma_x = f_*(\gamma_x)$ est un lacet de N basé en $f(x)$. L’application f_* est compatible avec l’équivalence d’homotopie et définit un morphisme de groupes $f_* : \pi_1(M) \rightarrow \pi_1(N)$ avec $Id_{M*} = Id_{\pi_1(M)}$ et $(h \circ f)_* = h_* \circ f_*$ si $M \xrightarrow{f} N \xrightarrow{h} P$.

où $\{\mathbf{t}_{hor,i}\}$ est une base de \mathcal{C} .²¹ Par construction, $\nabla_{\mathcal{C}}$ est tangent à \mathcal{C} (i.e. horizontal) et définit par conséquent un champ de vecteurs dont les trajectoires sont toutes des courbes intégrales de \mathcal{C} . De même, si $\mathbf{t}_{hor} = \sum \varphi_i \mathbf{t}_{hor,i}$ est un champ de vecteurs tangent à \mathcal{C} , sa *divergence* relative à \mathcal{C} est donnée par

$$\operatorname{div}_{\mathcal{C}} \mathbf{t}_{hor} = \sum \mathbf{t}_{hor,i}(\varphi_i).$$

Quant au laplacien, il est donné par

$$\Delta_{\mathcal{C}}(f) = \operatorname{div}_{\mathcal{C}}(\nabla_{\mathcal{C}}(f)) = \sum \mathbf{t}_{hor,i}^2(f).$$

On voit que ces structures consistent à partir des structures classiques de gradient, de divergence et de laplacien dans la base M de la connexion et à les relever dans les plans horizontaux.

3.5. Connexions principales

La notion de connexion présentée ci-dessus est valable pour tout fibré. Lorsque l'on considère un fibré *principal* \mathfrak{P} de groupe structural G , l'action de G signifie que les structures et propriétés de \mathfrak{P} sont "homogènes" dans chaque fibre, autrement dit que ce qui se passe en \mathfrak{p} peut être translaté en $\mathfrak{p}g$ par l'action à droite R_g . On utilise alors cette "équivalence" pour définir les connexions *principales*. L'idée est de définir les espaces horizontaux comme les noyaux d'une forme de Pfaff ω sur \mathfrak{P} à valeurs dans l'algèbre de Lie \mathcal{G} et G -*équivariante*. Comme nous l'avons vu, cette "forme de connexion" est simplement la projection des vecteurs tangents $\mathbf{t} \in T_{\mathfrak{p}}\mathfrak{P}$ sur l'espace vertical $T_{\mathfrak{p}}(\mathfrak{P}_{x=\pi(\mathfrak{p})})$ parallèlement à l'espace horizontal $C_{\mathfrak{p}}$ conformément à la décomposition en somme directe $T_{\mathfrak{p}}\mathfrak{P} = T_{\mathfrak{p}}(\mathfrak{P}_{x=\pi(\mathfrak{p})}) \oplus C_{\mathfrak{p}}$. Par définition, ω s'annule donc sur les espaces horizontaux. Elle est "verticale" au sens où si $\mathbf{t} \in T_{\mathfrak{p}}\mathfrak{P}$, $\omega(\mathbf{t}) = \omega(\mathbf{t}_{ver})$. Mais cette propriété ne définit que les noyaux de ω . Reste à définir l'espace dans lequel ω prend ses valeurs $\omega(\mathbf{t}_{ver})$. C'est là qu'intervient la structure de fibré principal. Nous avons vu que si $\mathfrak{g} \in \mathcal{G}$, \mathfrak{g} induit un automorphisme infinitésimal, c'est-à-dire un champ de vecteurs $\vec{\mathfrak{g}}(\mathfrak{p})$ G -invariant sur \mathfrak{P} . On peut donc choisir $\omega(\vec{\mathfrak{g}}) = \mathfrak{g}$, cette propriété d'équivariance signifiant que le noyau de ω en $\mathfrak{p}g$ est le translaté par g du noyau en \mathfrak{p} , autrement dit que $C_{\mathfrak{p}g} = (C_{\mathfrak{p}})g$ (où l'on note $(T_{\mathfrak{p}}\mathfrak{P})g$ l'action de l'application linéaire tangente à la translation $\mathfrak{p} \rightarrow \mathfrak{p}g$).²² On peut aussi exprimer cela en disant que ω est invariante par la représentation adjointe (à droite) de G , $Ad(g)(R_g^*\omega) = \omega$.

21. Ici $\nabla_{\mathcal{C}}$ généralise le gradient et nous utilisons donc le symbole nabla ∇ . Nous verrons plus bas que ce symbole est utilisé traditionnellement pour les dérivations covariantes en général.

22. On vérifie la cohérence dimensionnelle. Si M est de dimension n et G de dimension N , $\omega_{\mathfrak{p}}$ est définie sur un espace de dimension $n + N$ et est à valeurs dans un espace de dimension N . Comme elle est surjective, son noyau est donc de dimension n .

L'équivariance d'une connexion principale permet de relever horizontalement de façon *cohérente* les vecteurs tangents de la base. En effet, si $X \in T_x M$ et si $\mathfrak{p} \in \mathfrak{P}_x$ est un point au-dessus de x (donc $\pi(\mathfrak{p}) = x$), il existe un seul vecteur horizontal $\mathfrak{t}_{\mathfrak{p}}(X) \in C_{\mathfrak{p}}$ au-dessus de X et l'équivariance de la distribution horizontale \mathcal{C} fait que $\mathfrak{t}_{pg}(X) = \mathfrak{t}_{\mathfrak{p}}(X)g$.

La *courbure* Ω de ω est alors la différentielle $d\omega$ calculée sur les espaces *horizontaux*. Il s'agit d'une 2-forme à valeurs dans \mathcal{G} définie de la façon suivante. Soient $\mathfrak{t}, \mathfrak{u}$ deux vecteurs tangents en $\mathfrak{p} \in \mathfrak{P}$. On considère leurs projections horizontales \mathfrak{t}_{hor} et \mathfrak{u}_{hor} parallèlement à l'espace vertical (l'espace tangent à la fibre) en \mathfrak{p} et on pose $\Omega(\mathfrak{t}, \mathfrak{u}) = d\omega(\mathfrak{t}_{hor}, \mathfrak{u}_{hor})$, ce qui précise la formule $\Omega_{\mathcal{C}}(\mathfrak{t}, \mathfrak{u}) = [\mathfrak{t}_{hor}, \mathfrak{u}_{hor}]_{ver}$ vue plus haut. Par construction, c'est une forme *horizontale équivariante*.²³

On montre (théorème d'Ambrose-Singer) que les $\Omega(\mathfrak{t}, \mathfrak{u}) \in \mathcal{G}$ engendrent dans \mathcal{G} la sous-algèbre de Lie qu'est l'algèbre de Lie \mathfrak{h} du groupe d'holonomie $\mathfrak{H} \subset G$. On montre également que l'on a la *formule* de Maurer-Cartan

$$\begin{aligned}\Omega &= d\omega + \frac{1}{2}[\omega \wedge \omega] \\ \Omega(\mathfrak{t}, \mathfrak{u}) &= d\omega(\mathfrak{t}, \mathfrak{u}) + [\omega(\mathfrak{t}), \omega(\mathfrak{u})]\end{aligned}$$

et que l'équation de Maurer-Cartan $d\omega + \frac{1}{2}[\omega \wedge \omega] = 0$ signifie donc exactement que la courbure est nulle.

3.6. Les potentiels vecteurs et le formalisme physique

Nous allons maintenant introduire, pour le lecteur susceptible d'être intéressé, les formulations de la théorie des connexions habituelles en physique. Elles sont moins géométriques et utilisent des indices dans des bases holonomes (*h*-repères) ou non holonomes (*nh*-repères).

À cause de son équivariance, on peut écrire une connexion principale ω sur un fibré principal $\mathfrak{P} \rightarrow M$ de groupe structural G comme une 1-forme $A \in \Gamma(T^*M) \otimes \mathcal{G}$ définie sur l'espace de base M et à valeurs dans l'algèbre de Lie \mathcal{G} de G , A étant appelée en physique un *potentiel vecteur*. Pour ce faire, il faut effectuer un choix de jauge locale au moyen d'une section $\sigma : U \rightarrow \mathfrak{P}$ du fibré $\mathfrak{P} \rightarrow M$. Grâce à son application tangente $\sigma_* : TU \rightarrow T\mathfrak{P}$, σ permet de relever de façon univoque les vecteurs tangents à M , $X \in T_x M$, en vecteurs tangents $\sigma_*(X) \in T_{\sigma(x)}\mathfrak{P}$ et de définir A par $A(X) = \omega_{\sigma(x)}(D\sigma(X)) \in \mathcal{G}$. Insistons sur le fait que A dépend du choix de jauge σ .

En termes de composantes, si $\{x^\mu\}$ sont des coordonnées locales en x et si $\{\mathfrak{g}_\alpha\}$ est une base de \mathcal{G} , A s'écrira

$$A = A_\mu^\alpha dx^\mu \otimes \mathfrak{g}_\alpha$$

23. Ω est *horizontale équivariante* alors que ω est *verticale équivariante*.

(ou pour simplifier $A = A_\mu^\alpha dx^\mu \mathfrak{g}_\alpha$). Donc $A \in \Gamma(T^*M) \otimes \mathcal{G}$ est une 1-forme sur M à valeurs dans \mathcal{G} qui peut s'écrire soit sous la forme $A^\alpha \otimes \mathfrak{g}_\alpha$ où les $A^\alpha = A_\mu^\alpha dx^\mu \in \Gamma(T^*M)$ sont des 1-formes scalaires sur M , soit sous la forme $A = A_\mu dx^\mu$ où les A_μ sont des éléments $A_\mu^\alpha \mathfrak{g}_\alpha$ de \mathcal{G} , soit encore sous la forme $A = A_\mu^\alpha dx^\mu \otimes \mathfrak{g}_\alpha$ où les A_μ^α sont des coefficients (des fonctions sur M).

3.7. Connexions linéaires ou affines (avec “déluge d'indices”)

Lorsque l'on passe des fibrés principaux à des fibrés vectoriels associés $\pi_{\mathfrak{V}} : \mathfrak{V} \rightarrow M$, on développe le même formalisme mais en tenant compte du fait que G agit sur les fibres à travers une représentation $\rho : G \rightarrow GL(\mathfrak{V})$ sur la fibre type \mathfrak{V} . Donc, relativement à des bases équivariantes des fibres, $\rho(g)$ (souvent notée simplement g) devient une matrice $(g)_i^j$ inversible et $\rho(\mathfrak{g})$ (encore notée \mathfrak{g}) devient une matrice $(\mathfrak{g})_i^j$ quelconque. La connexion A devient ainsi une *matrice* A_i^j de connexions *scalaires* sur M dont les composantes dépendent de quadruplets d'indices (c'est le début du “déluge” d'indices) :

$$A_i^j = A_\mu^\alpha dx^\mu (\mathfrak{g}_\alpha)_i^j = (A_\mu^\alpha dx^\mu \otimes \mathfrak{g}_\alpha)_i^j .$$

Si l'on se focalise sur cette interprétation, il est parfois préférable de changer l'ordre des produits et d'écrire

$$A_i^j = A_{i,\mu}^j dx^\mu = A_\mu^\alpha (\mathfrak{g}_\alpha)_i^j dx^\mu .$$

Les A_i^j expriment que, les $\mathfrak{g} \in \mathcal{G}$ étant représentés à travers ρ comme des matrices, A devient représentable comme une matrice et que donc A est une 1-forme sur M à valeurs dans l'algèbre $\text{End}(\mathfrak{V})$ des endomorphismes de \mathfrak{V} .²⁴

Reprenons alors, dans ce contexte et de façon plus élaborée, la notion, vue plus haut, de *dérivation covariante* ∇ . On veut généraliser le fait qu'une fonction $f : M \rightarrow \mathbb{R}$ sur M à valeurs réelles a pour différentielle la 1-forme df à valeurs réelles. Pour ce faire, on considère f comme une section du fibré vectoriel trivial $M \times \mathbb{R} \rightarrow \mathbb{R}$ et df comme une 1-forme sur M à valeurs dans $M \times \mathbb{R}$. La généralisation s'impose alors d'elle-même. On considère le $C^\infty(M)$ -module $\Gamma(\mathfrak{V})$ des sections de \mathfrak{V} et l'on impose que l'opérateur ∇ transforme les sections \mathfrak{v} de \mathfrak{V} en 1-formes sur M à valeurs dans \mathfrak{V} en vérifiant la propriété de linéarité et la formule de Leibniz²⁵

$$\nabla(f\mathfrak{v}) = f\nabla\mathfrak{v} + df \otimes \mathfrak{v} .$$

24. Rappelons que pour tout espace vectoriel V , $\text{End}(V) \simeq V \otimes V^*$, ce que reflète la position des indices A_i^j de la matrice exprimant un élément $A \in \text{End}(V)$ dans une base de V .

25. Par construction, l'opérateur ∇ n'est donc *pas* $C^\infty(M)$ -linéaire pour la structure de $C^\infty(M)$ -module de $\Gamma(\mathfrak{V})$, mais seulement \mathbb{R} -linéaire. Il comprend le terme $df \otimes \mathfrak{v}$ qui ne dépend pas de la connexion ∇ .

Cette formule signifie que si $\mathbf{v}(x) \in \mathfrak{V}_x$ est une section et si $X \in T_x M$, alors $(\nabla \mathbf{v})_x(X) \in \mathfrak{V}_x$ et

$$(\nabla(f\mathbf{v}))_x(X) = f(x)(\nabla \mathbf{v})_x(X) + df_x(X)\mathbf{v}(x) \in \mathfrak{V}_x .$$

∇ est donc un opérateur $\nabla : \Gamma(\mathfrak{V}) \rightarrow \Lambda^1(M) \otimes \Gamma(\mathfrak{V})$.

Comme nous venons de le noter, ∇ n'est pas $\mathcal{C}^\infty(M)$ -linéaire puisque

$$\nabla(f\mathbf{v}) = f\nabla\mathbf{v} + df \otimes \mathbf{v} \neq f\nabla\mathbf{v} .$$

Mais si ∇_1 et ∇_2 sont deux connexions, la différence $\nabla_1 - \nabla_2$ est, elle, un opérateur $\mathcal{C}^\infty(M)$ -linéaire (mais pas une connexion puisque précisément le terme $df \otimes \mathbf{v}$ a disparu). En effet,

$$(\nabla_1 - \nabla_2)(f\mathbf{v}) = f\nabla_1\mathbf{v} + df \otimes \mathbf{v} - f\nabla_2\mathbf{v} - df \otimes \mathbf{v} = f(\nabla_1 - \nabla_2)(\mathbf{v}) .$$

L'espace fonctionnel des connexions est donc comme un "espace affine" qui, une fois choisie la différentielle de base d comme origine, devient un "espace vectoriel" ; tout cela étant compris par rapport à l'algèbre des "scalaires" $\mathcal{C}^\infty(M)$.

Grâce à la partie $\Lambda^1(M)$, on peut, si l'on fixe un champ de vecteurs $X \in \mathfrak{X}(M) = \Gamma(TM)$ sur M , considérer $\nabla \mathbf{v}(X)$, qui est une section de \mathfrak{V} . On la note traditionnellement $\nabla_X \mathbf{v}$ et l'on voit que l'on a ainsi défini une dérivée de \mathbf{v} le long d'un champ X de la base M . Cela n'est possible que parce que les $T_x M$ de la base peuvent être relevés de façon équivariante en champs horizontaux dans le fibré : ∇_X est le champ de vecteurs horizontaux qui relève $X \in \mathfrak{X}(M)$ dans $\mathfrak{X}(\mathfrak{P})$.

Cela permet de définir la courbure de la connexion ∇ à partir des champs de vecteurs X sur la base M . Comme nous l'avons vu, les champs horizontaux n'ont pas de commutateurs horizontaux en général²⁶ et la courbure est précisément donnée par la 2-forme sur M à valeurs dans $\mathfrak{X}(\mathfrak{P})$, $\Omega(X, Y) = [\nabla_X, \nabla_Y]_{ver}$. Et comme $[\nabla_X, \nabla_Y]_{hor} = \nabla_{[X, Y]}$, on obtient immédiatement la formule

$$\begin{aligned} \Omega(X, Y) &= [\nabla_X, \nabla_Y]_{ver} = [\nabla_X, \nabla_Y] - [\nabla_X, \nabla_Y]_{hor} \\ &= [\nabla_X, \nabla_Y] - \nabla_{[X, Y]} \end{aligned}$$

sur laquelle nous allons revenir.

Relativement à des coordonnées locales $\{x^\mu\}$ dans M définissant les bases holonomes ∂_μ de vecteurs tangents à M , on peut ainsi définir les $\nabla_\mu \mathbf{v} = \nabla_{\partial_\mu} \mathbf{v}$.

Reste à définir ∇ à partir de la connexion A . Pour ce faire, on explicite la définition compacte

$$\nabla \mathbf{v} = A\mathbf{v} + d\mathbf{v}$$

26. En revanche, les champs verticaux \mathfrak{t} ont des commutateurs verticaux car ils sont définis par $\pi_*(\mathfrak{t}) = 0$.

formulant la différence entre la dérivation standard $d\mathbf{v}$ et la dérivation covariante $\nabla\mathbf{v}$. On considère d'abord des sections $\mathbf{f}_i(x)$ (\mathbf{f} pour "fibre") fournissant des bases des fibres \mathfrak{F}_x et on définit $\nabla\mathbf{f}_i$ par

$$\nabla\mathbf{f}_i = A_i^j \mathbf{f}_j$$

ce qui signifie que $(\nabla\mathbf{f}_i)_x(X) \in \mathfrak{F}_x$ est donné par $(A_i^j)_x(X) \mathbf{f}_j \in \mathfrak{F}_x$. Cela est bien cohérent car les $A_i^j = A^\alpha(\mathbf{g}_\alpha)_i^j$ étant des matrices agissant sur les fibres, on peut les faire agir sur les vecteurs \mathbf{f}_j . Il s'ensuit que pour $\mathbf{v} = v^i \mathbf{f}_i$ (i.e. $\mathbf{v}(x) = v^i(x) \mathbf{f}_i(x)$), on doit poser, conformément à la formule de Leibniz,

$$\nabla\mathbf{v} = \nabla(v^i \mathbf{f}_i) = v^i \nabla\mathbf{f}_i + dv^i \otimes \mathbf{f}_i = v^i A_i^j \otimes \mathbf{f}_j + dv^i \otimes \mathbf{f}_i = (v^j A_j^i + dv^i) \otimes \mathbf{f}_i$$

ce qui est bien cohérent puisqu'il s'agit d'une somme de produits tensoriels de 1-formes sur M avec des vecteurs venant des fibres.²⁷ Si on développe les 1-formes A_j^i sur M en $A_{j,\mu}^i dx^\mu$, on obtient

$$\nabla\mathbf{v} = (v^j A_{j,\mu}^i + \partial_\mu v^i) dx^\mu \otimes \mathbf{f}_i .$$

En termes de composantes, si $X = X^\mu \partial_\mu$,

$$\begin{aligned} \nabla_X \mathbf{v} &= (v^j A_j^i(X) + dv^i(X)) \otimes \mathbf{f}_i = (v^j A_{j,\mu}^i + \partial_\mu v^i) X^\mu \otimes \mathbf{f}_i \\ \nabla_\mu \mathbf{v} &= (v^j A_{j,\mu}^i + \partial_\mu v^i) \otimes \mathbf{f}_i . \end{aligned}$$

On notera que l'opérateur $\nabla_X \mathbf{v}$ est linéaire relativement à la structure de $\mathcal{C}^\infty(M)$ -module de $\Gamma(TM)$ puisque $dx^\mu(fX) = f dx^\mu(X)$, dx^μ sélectionnant simplement la μ -ième composante, X^μ et fX^μ , de X et fX . Par ailleurs, il satisfait par construction la formule de Leibniz. On a donc

$$\begin{cases} \nabla_{fX} \mathbf{v} = f \nabla_X \mathbf{v} \\ \nabla_X (f\mathbf{v}) = f \nabla_X \mathbf{v} + df(X) \otimes \mathbf{v} \end{cases}$$

On peut alors dire qu'une section \mathbf{v} est "parallèle" pour la connexion A si $\nabla\mathbf{v} = 0$, autrement dit si toutes les 1-formes $v^j A_j^i + dv^i$ sont nulles. Comme les A_j^i sont données, on obtient des équations différentielles pour les v^j . Comme $\nabla\mathbf{v} = 0$ généralise $df = 0$ pour les fonctions et que $df = 0$ signifie que f est constante, on peut dire que les sections \mathbf{v} "parallèles" sont en quelque sorte "constantes" pour les déplacements horizontaux.

On peut facilement généraliser la dérivation covariante ∇ , et d'abord au dual \mathfrak{F}^* de \mathfrak{F} . Si $\{\mathbf{f}^i\}$ est la base duale de $\{\mathbf{f}_i\}$, on pose

$$\nabla\mathbf{f}^i = -A_j^i \otimes \mathbf{f}^j .$$

27. On a dû intervertir i et j dans le premier terme pour pouvoir mettre \mathbf{f}_i en facteur.

On généralise ensuite aux tenseurs et, en particulier, aux formes différentielles. On obtient ainsi une dérivation graduée

$$\nabla : \Lambda^k(M) \otimes \Gamma(\mathfrak{V}) \rightarrow \Lambda^{k+1}(M) \otimes \Gamma(\mathfrak{V})$$

en faisant toutefois attention que, tant qu'un produit n'est pas défini sur les vecteurs de \mathfrak{V} , $\nabla(\rho \wedge \tau)$ avec $\rho \in \Lambda^k(M) \otimes \Gamma(\mathfrak{V})$ n'est défini que pour des $\tau \in \Lambda^l(M)$ scalaires (i.e. à valeurs dans \mathbb{R}). Et $\nabla(\rho \wedge \tau)$ satisfait dans ce cas

$$\nabla(\rho \wedge \tau) = (\nabla\rho) \wedge \tau + (-1)^k \rho \wedge d\tau .$$

Considérons par exemple une 1-forme $\rho \in \Lambda^1(M) \otimes \Gamma(\mathfrak{V})$, $\rho = \rho_\mu^i dx^\mu \otimes \mathfrak{f}_i$. On a (puisque $d^2x^\mu = 0$)

$$\begin{aligned} \nabla\rho &= \nabla(\rho_\mu^i dx^\mu \otimes \mathfrak{f}_i) = \nabla(\rho_\mu^i \mathfrak{f}_i) \wedge dx^\mu + \rho_\mu^i d^2x^\mu \otimes \mathfrak{f}_i \\ &= \nabla(\rho_\mu^i \mathfrak{f}_i) \wedge dx^\mu = \rho_\mu^i \nabla(\mathfrak{f}_i) \wedge dx^\mu + (d\rho_\mu^i \wedge dx^\mu) \otimes \mathfrak{f}_i \\ &= \rho_\mu^i A_i^j \mathfrak{f}_j \wedge dx^\mu + (d\rho_\mu^i \wedge dx^\mu) \otimes \mathfrak{f}_i \end{aligned}$$

qui est bien une 2-forme à valeurs dans \mathfrak{V} . Dans cette formule on a utilisé $d^2x^\mu = 0$, mais si l'on considère dans la base M un nh -repère \mathbf{e}_μ au lieu d'un h -repère ∂_μ , alors on doit faire intervenir les constantes de structures car $d\mathbf{e}^\mu = -\frac{1}{2}c_{\sigma\tau}^{\mu} \mathbf{e}^\sigma \wedge \mathbf{e}^\tau$.

On montre que la formule

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y])$$

de la section 7.2.7 du chapitre 2, se généralise bien à ∇ :

$$\nabla\rho(X, Y) = \nabla_X(\rho(Y)) - \nabla_Y(\rho(X)) - \rho([X, Y]) .$$

La *courbure* de la connexion ∇ est alors définie comme le carré

$$\nabla^2 : \Gamma(\mathfrak{V}) \rightarrow \Lambda^2(M) \otimes \Gamma(\mathfrak{V}) .$$

L'opérateur ∇^2 a la propriété remarquable d'être $\mathcal{C}^\infty(M)$ -linéaire pour la structure de $\mathcal{C}^\infty(M)$ -module de $\Gamma(\mathfrak{V})$ alors que, comme nous l'avons noté, l'opérateur ∇ est seulement \mathbb{R} -linéaire. En effet, d'après la formule donnant $\nabla(\rho \wedge \tau)$ pour $\rho = \nabla\mathbf{v} \in \Lambda^1(M) \otimes \Gamma(\mathfrak{V})$ et $\tau = f \in \Lambda^0(M)$, on a

$$\begin{aligned} \nabla^2(f\mathbf{v}) &= \nabla(\nabla(f\mathbf{v})) = \nabla(f\nabla\mathbf{v} + df \otimes \mathbf{v}) \\ &= f\nabla^2\mathbf{v} - \nabla\mathbf{v} \wedge df + \nabla\mathbf{v} \wedge df + d^2f \otimes \mathbf{v} = f\nabla^2\mathbf{v} . \end{aligned}$$

Comme le note Coquereaux ([127], p. 160), c'est pour cette raison que la courbure est un *tenseur*. Ce caractère "tensoriel relativement à $\mathcal{C}^\infty(M)$ " signifie que le \otimes sur \mathbb{R} qu'est $\Lambda^2(M) \otimes \Gamma(\mathfrak{V})$ est en fait un \otimes sur $\mathcal{C}^\infty(M)$, $\Lambda^2(M) \otimes_{\mathcal{C}^\infty(M)} \Gamma(\mathfrak{V})$, la caractéristique d'un produit tensoriel étant que les scalaires peuvent être transférés entre les termes : le "scalaire" f appliqué à \mathbf{v} peut être transféré à ∇^2 .

En termes du potentiel de jauge A de la connexion ∇ , les composantes de la courbure s'expriment en calculant les $\nabla^2 \mathbf{f}_i$. On trouve (en renommant les indices quand nécessaire)

$$\begin{aligned}\nabla^2 \mathbf{f}_i &= \nabla (\nabla \mathbf{f}_i) = \nabla (A_i^j \mathbf{f}_j) = \nabla (A_{i,\mu}^j dx^\mu \otimes \mathbf{f}_j) \\ &= \nabla \mathbf{f}_j \wedge A_i^j + dA_i^j \otimes \mathbf{f}_j = A_j^k \wedge A_i^j \otimes \mathbf{f}_k + dA_i^j \otimes \mathbf{f}_j \\ &= A_k^j \wedge A_i^k \otimes \mathbf{f}_j + dA_i^j \otimes \mathbf{f}_j = (A_k^j \wedge A_i^k + dA_i^j) \otimes \mathbf{f}_j .\end{aligned}$$

Cela est bien cohérent car, d'un côté, les $\nabla \mathbf{f}_j$ étant des 1-formes vectorielles sur M et les A_i^j des 1-formes scalaires, les $\nabla \mathbf{f}_j \wedge A_i^j$ sont des 2-formes vectorielles. alors que, d'un autre côté, les dA_i^j étant des 2-formes scalaires, les $dA_i^j \otimes \mathbf{f}_j$ sont également des 2-formes vectorielles. On obtient ainsi une matrice $F = (F_i^j)$ de 2-formes scalaires

$$F_i^j = dA_i^j + A_k^j \wedge A_i^k$$

qui est, comme plus haut, la *courbure*²⁸ et que l'on peut noter de façon compacte

$$F = dA + A \hat{\wedge} A ,$$

le symbole $\hat{\wedge}$ exprimant que l'on utilise *à la fois* le produit vectoriel de 1-formes sur M et le produit de matrices dans $\text{End}(\mathfrak{V})$ car A est une 1-forme sur M à valeurs dans $\text{End}(\mathfrak{V})$.²⁹ F est donc une 2-forme sur M à valeurs dans $\text{End}(\mathfrak{V})$ et nous retrouvons les définitions précédentes mais dans un contexte élargi et avec des notations indicielles à la manière des physiciens.

Remarque. Le lien avec la formule $\Omega = d\omega + \frac{1}{2}[\omega \wedge \omega]$ de la courbure pour les connexions principales vient du fait que $A \hat{\wedge} A = \frac{1}{2}[A \wedge A]$. □

Comme F est une 2-forme sur M , on peut considérer ses valeurs sur les nh -repères $\{\mathbf{e}_\mu\}$ et l'écrire sous la forme $F = \frac{1}{2}F_{\mu\nu} \mathbf{e}^\mu \wedge \mathbf{e}^\nu$ (le facteur $\frac{1}{2}$ corrige la sommation d'Einstein et $F = \frac{1}{2}F_{\mu\nu} dx^\mu \wedge dx^\nu$ pour les h -repères $\mathbf{e}_\mu = \partial_\mu$) les $F_{\mu\nu}$ étant des éléments de $\text{End}(\mathfrak{V}_x)$. On trouve les formules classiques

$$\begin{aligned}F_{\mu\nu} &= [\nabla_\mu, \nabla_\nu] - \nabla_{[\mu, \nu]} , \text{ avec } \nabla_\mu = \nabla_{\mathbf{e}_\mu} \text{ (} nh\text{-repère)} \\ &= [\nabla_\mu, \nabla_\nu] \text{ si } \mathbf{e}_\mu = \partial_\mu \text{ (} h\text{-repère)} .\end{aligned}$$

On montre que si l'on change de jauge au moyen d'une application $g : M \rightarrow G$ qui transforme $\sigma(x)$ en $\sigma(x)g(x)$, la connexion A se transforme en $A' = g^{-1}Ag + g^{-1}dg$ et la courbure F en $F' = g^{-1}Fg$. Ces deux formules

$$\begin{aligned}A' &= g^{-1}Ag + g^{-1}dg \\ F' &= g^{-1}Fg\end{aligned}$$

sont omniprésentes en théorie quantique des champs.

28. Le symbole traditionnel F vient de "field" : en physique, les champs sont les courbures des connexions définies par les potentiels vecteurs.

29. Les physiciens utilisent la simple notation $F = dA + A \wedge A$.

Une formule importante, et même essentielle en relativité générale, est l'*identité de Bianchi*.³⁰ Pour son interprétation philosophique dans le classique *Gravitation* de Misner, Thorne & Wheeler [369], on pourra consulter notre étude [419]. Elle concerne la différentielle normale dF de F qui est une 3-forme sur M à valeurs dans $\text{End}(\mathfrak{V})$ et qui s'écrit

$$dF = F \wedge A - A \wedge F = [F \wedge A]$$

le crochet étant le crochet de Lie des endomorphismes et de leurs matrices. Elle dit que la dérivée de la courbure F s'obtient algébriquement de façon simple à partir de A et de F . Pour la montrer, on écrit dF et on utilise l'identité $d^2A = 0$. Plus précisément

$$F_i^j = dA_i^j + A_k^j \wedge A_i^k$$

est une 1-forme scalaire et on a donc

$$dF_i^j = d^2A_i^j + dA_k^j \wedge A_i^k - A_k^j \wedge dA_i^k.$$

Pour éliminer les différentielles dans le second membre on utilise $d^2A_i^j = 0$ et $dA_k^j = F_k^j - A_j^i \wedge A_k^i$, d'où

$$dF_i^j = (F_k^j - A_j^i \wedge A_k^i) \wedge A_i^k - A_k^j \wedge (F_i^k - A_j^l \wedge A_i^l).$$

$F_k^j \wedge A_i^k$ donne $F \wedge A$, $-A_j^i \wedge F_i^k$ donne $-A \wedge F$ et les deux autres termes s'éliminent à cause de l'associativité des produits.

4. Théories de jauge

4.1. Théories de jauge abéliennes (équations de Maxwell)

Le formalisme des connexions est omniprésent en physique dans les théories de jauge qui se trouvent au cœur de la théorie quantique des champs. Depuis les travaux pionniers de Chen Ning Yang et Robert Mills (1954) sur l'invariance de jauge concernant l'isospin, il existe dans les théories de jauge deux classes de champs :

1. Les champs fermioniques de matière qui sont interprétés comme des sections de fibrés \mathfrak{F} sur l'espace-temps M . Les coordonnées des fibres sont les degrés internes de liberté. Le groupe structural (i.e. le groupe de symétrie des fibres) exprime les symétries internes des particules. Les groupes de Lie caractéristiques des interactions classiques sont $U(1)$ pour l'électromagnétisme, $SU(2)$ pour

30. Luigi Bianchi (1856-1928) fut l'élève de Enrico Betti et Ulisse Dini à la Scuola Normale Superiore de Pise où il devint plus tard professeur. Inspiré par Riemann, Lie et Klein, il est crucial pour la relativité générale. Les identités de Bianchi de 1902 approfondissent des identités découvertes par son ami Gregorio Ricci en 1880. Nous retrouverons Bianchi à la section 3.3 du chapitre 7 à propos de la classification des algèbres de Lie réelles de dimension 3.

les interactions nucléaires faibles, $SU(3)$ pour les interactions nucléaires fortes (chromodynamique).

2. Les champs bosoniques de jauge qui sont des champs d'interactions véhiculées par des particules virtuelles d'échange (des bosons) et sont interprétés comme des connexions sur ces fibrés. Les bosons sont les photons γ pour l'électromagnétisme, les bosons W^\pm , Z pour les interactions faibles, les gluons pour la chromodynamique.³¹ Les particules véhiculant les interactions sont par conséquent les quanta des champs de connexions sur les fibrés de matière.

L'exemple le plus simple, que nous avons traité en détail à la section 17 du chapitre 2, est celui des équations de Maxwell. On est dans le cas où

$$G = SO(2) = U(1)$$

est abélien. Rappelons que le champ électrique E est une 1-forme

$$\omega_E = E_x dx + E_y dy + E_z dz$$

sur \mathbb{R}^3 et le champ magnétique B une 2-forme

$$\sigma_B = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy .$$

Résumons les lois.

- La première loi (Maxwell-Thomson, concernant l'absence de sources magnétiques), $\text{div } B = 0$, signifie que $d\sigma_B = 0$ (σ_B est une 2-forme fermée) et donc que, localement, σ_B est de la forme $\sigma_B = d\omega_A$, où ω_A est la 1-forme associée à un potentiel vecteur magnétique A .
- La deuxième loi (Maxwell-Gauss, concernant les sources électriques), $\text{div } E = \frac{\rho}{\varepsilon_0}$, signifie que $d\sigma_E = \frac{1}{\varepsilon_0} \rho dV$ (où σ_E est la 2-forme $*\omega_E$ associée à la 1-forme ω_E par dualité de Hodge et dV la forme volume de \mathbb{R}^3).
- La troisième loi (Maxwell-Faraday, reliant géométriquement E à la dynamique de B , sans les sources), $\nabla \wedge E = -\frac{\partial B}{\partial t}$, signifie l'égalité des 2-formes $d\omega_E = -\frac{\partial \sigma_B}{\partial t}$ et donc l'égalité $\omega_E = -\frac{\partial \omega_A}{\partial t} - dV$ avec V un potentiel scalaire électrique.
- La quatrième loi (Maxwell-Ampère, reliant physiquement B à la dynamique de E à travers les sources), $\text{rot } (B) = \mu_0 (j + \varepsilon_0 \frac{\partial E}{\partial t})$, signifie que $d\omega_B = \mu_0 \sigma_j + \mu_0 \varepsilon_0 \frac{\partial \sigma_E}{\partial t}$ (où ω_B est la 1-forme $*\sigma_B$ associée à σ_B par dualité de Hodge et σ_j la 2-forme du vecteur courant j).

31. Les photons sont de masse nulle (ils peuvent donc agir à distance infinie), de spin 1 et n'ont ni charge électrique, ni charge de "couleur". Les bosons faibles W^\pm , Z ont une masse élevée, un spin 1, une charge électrique nulle pour Z et ± 1 pour W^\pm et pas de charge de couleur. Les huit gluons de la chromodynamique quantique sont sans masse, de spin 1, n'ont pas de charge électrique mais évidemment une (anti)charge de couleur. Ils confinent les quarks dans les protons, neutrons et autres hadrons. Il y a aussi le boson de Higgs de masse élevée, de spin 0, sans charge électrique et conférant, à travers le mécanisme de Higgs, de la masse aux bosons électrofaibles et aux fermions.

On regroupe ω_E et σ_B en une 2-forme \mathfrak{F} sur \mathbb{R}^4 (l'espace-temps comme cadre unitaire),

$$\mathfrak{F} = \sigma_B + \omega_E \wedge dt$$

(où σ_B est maintenant le prolongement trivial de σ_B de \mathbb{R}^3 à \mathbb{R}^4). Les première et troisième lois disent alors simplement que

$$d\mathfrak{F} = 0$$

et donc que, localement, $\mathfrak{F} = d\alpha$ où $\alpha = \omega_A - Vdt$ est une 1-forme sur \mathbb{R}^4 (avec ω_A définie sur \mathbb{R}^3 et prolongée trivialement à \mathbb{R}^4). Quant aux deuxième et quatrième lois, elles disent que si l'on considère la 2-forme sur \mathbb{R}^4 qu'est la duale de Hodge de \mathfrak{F} ,

$$*\mathfrak{F} = *(\sigma_B + \omega_E \wedge dt) ,$$

alors la 3-forme $d(*\mathfrak{F})$ est donnée comme

$$d(*\mathfrak{F}) = *J$$

où la 3-forme $*J$ est la duale de Hodge de la 1-forme de quadricourant $J = \rho dt - \omega_j$ (ω_j définie sur \mathbb{R}^3 étant prolongée trivialement à \mathbb{R}^4).

Pour faire le lien avec ce qui précède en évitant les confusions de notations, affublons d'un $\hat{}$ les symboles A et F de la 1-forme de connexion et de sa courbure. Le potentiel vecteur est ici $\hat{A} = \omega_A - Vdt$ et

$$\hat{F} = \nabla \hat{A} = d\hat{A} + \hat{A} \wedge \hat{A}$$

se réduit à $\hat{F} = d\hat{A}$ car $\hat{A} \wedge \hat{A} = 0$. Donc

$$\hat{F} = \mathfrak{F} = \sigma_B + \omega_E \wedge dt$$

donne le champ électromagnétique dérivant du potentiel vecteur $\hat{A} = \omega_A - Vdt$. L'identité de Bianchi

$$d\hat{F} = \hat{F} \wedge \hat{A} - \hat{A} \wedge \hat{F} = [\hat{F} \wedge \hat{A}]$$

devient, toujours à cause de la commutativité du groupe de symétrie interne,

$$d\hat{F} = 0,$$

c'est-à-dire $d\mathfrak{F} = 0$ ce qui donne la moitié des équations de Maxwell, celle indépendante des sources. L'autre moitié est donnée par dualité de Hodge.

4.2. Théories de jauge non abéliennes

Comme nous le disions, les connexions sont également omniprésentes en théorie quantique des champs. Le lagrangien de Yang-Mills servant de base à ces théories est la *norme de la courbure des connexions*. Il est invariant sous l'action du groupe de jauge et l'espace-temps y contribue comme champ de jauge à travers la courbure scalaire de sa connexion. Les dérivations covariantes permettent d'exprimer

géométriquement les *interactions*. Cette possibilité est si remarquable et essentielle qu'elle mérite d'être commentée ici.

Pour formaliser les interactions, les théories de jauge font dépendre les symétries *internes* des systèmes (qui sont des symétries globales apparemment non spatio-temporelles associées aux nombres quantiques des particules) de la position spatio-temporelle *externe*. Si on *localise* ainsi ces symétries internes globales et si l'on exige que les théories demeurent invariantes, on doit évidemment introduire des termes correctifs. On constate alors que ceux-ci redonnent exactement les termes d'*interaction*. Les forces et les interactions apparaissent ainsi de façon générale comme dérivables de principes de conservation *locaux*.

Le cas le plus simple (découvert par Hermann Weyl pour les changements d'échelle) est celui du "couplage minimal" entre un électron et un champ électromagnétique F .³² Soit $\psi(x)$ la fonction d'onde (à valeurs spineurs) de l'électron sur l'espace-temps M . Son évolution est régie par l'équation de Dirac. Le lagrangien de Dirac \mathcal{L}_D est invariant sous la symétrie interne globale $\psi \rightarrow e^{-ie\theta}\psi$ (où le e en exposant est la charge de l'électron et θ une phase) car l'équation de Dirac a la forme $\bar{\psi}D\psi$ où $\bar{\psi}$ est la conjuguée de ψ et D l'opérateur de Dirac. Le groupe des symétries internes est donc le groupe commutatif $U(1)$ des rotations du cercle.³³ D'autre part, le lagrangien de Maxwell du champ électro-magnétique \mathcal{L}_M est quant à lui invariant sous une transformation de jauge $A \rightsquigarrow A + d\Lambda$, où A est le potentiel vecteur du champ électro-magnétique F , Λ étant une fonction sur M et $d\Lambda$ la différentielle de Λ .³⁴

Si l'on admet que le facteur de phase θ peut dépendre de la position spatio-temporelle $x \in M$ et être donc une fonction $\theta(x)$ sur M , alors le lagrangien de Dirac \mathcal{L}_D n'est plus invariant. Mais Weyl a découvert que le terme de correction qui apparaît si l'on veut conserver l'invariance peut être *exactement compensé* par la transformation de jauge $A \rightsquigarrow A + d\theta$, i.e., puisque $A = A_\mu dx^\mu$ et $d\theta = (\partial_\mu\theta) dx^\mu$, $A_\mu \rightsquigarrow A_\mu + \partial_\mu\theta$. Il s'agit là d'une sorte de "miracle" dont la signification tant physico-mathématique que philosophique est profonde.

Cette formalisation remarquable des interactions électromagnétiques entre les bosons du champ (photons) et les fermions que sont les électrons fut généralisée au

32. Nous en avons parlé dans la section précédente 3.2.2 consacrée au *Raumproblem*. L'histoire est un peu compliquée. Initialement, Weyl avait introduit la notion de jauge pour introduire des changements d'échelle locaux en RG. Cela ne fonctionna pas et ce n'est que plus tard que l'idée fut reprise (avec Vladimir Fock et Fritz London), pour localiser les symétries internes en théorie quantique des champs. Les changements d'échelle locaux y sont remplacés par des changements locaux de phase.

33. U pour transformations "unitaires".

34. Nous reprenons les symboles \hat{A} et \hat{F} que nous avons reliés aux équations de Maxwell dans la section précédente par $\hat{A} = \omega_A - V dt$ et $\hat{F} = \mathfrak{F} = \sigma_B + \omega_E \wedge dt$ et nous les renommons A et F comme il est traditionnel.

doublet d'isospin proton/neutron pour les interactions nucléaires fortes puis d'autres doublets d'isospin pour les interactions nucléaires faibles. Le groupe des symétries internes était alors $G = SU(2)$. Cela ouvrit la voie aux théories de jauge non abéliennes où le groupe des symétries internes G ($G = SU(2)$, $SU(3)$, etc.) n'est plus commutatif, ce qui introduit des difficultés notables. Mais les idées principales subsistent. Les potentiels-vecteurs A sont maintenant des 1-formes de connexion sur M à valeurs dans l'algèbre de Lie \mathcal{G} de G , le champ considéré étant la courbure de cette connexion. Si les \mathfrak{g}_a constituent une base de \mathcal{G} , le lagrangien est maintenant invariant par les symétries internes globales $U(g) = \exp(-i\theta^a \mathfrak{g}_a)$. Si on localise les symétries internes en faisant dépendre les $\theta^a = \theta^a(x)$ de la position x dans l'espace-temps M , alors le lagrangien n'est plus invariant mais sa variation peut être exactement compensée par des transformations de jauge appropriées. Celles-ci sont plus compliquées que les $A_\mu \rightsquigarrow A_\mu + \partial_\mu \theta$ et, avec $A = A_\mu^\alpha dx^\mu \otimes \mathfrak{g}_\alpha$, sont de la forme

$$A_\mu^\alpha \rightsquigarrow c_{\gamma\beta}^\alpha A_\mu^\gamma \theta^\beta + \partial_\mu \theta^\alpha$$

où les $c_{\gamma\beta}^\alpha$ sont les constantes de structure de \mathcal{G} . Le défaut de commutativité induit des non linéarités et des auto-couplages des champs de jauge, ce qui complique considérablement la situation.

5. La dialectique “interne/externe”

5.1. Connexions sur le fibré des repères

Une classe de fibrés géométriquement essentiels est celle des fibrés vectoriels associés au fibré principal des *repères* $\mathfrak{E} \rightarrow M$ de l'espace de base M , ce fibré étant muni du groupe structural $G \subset GL(n)$ qui est le sous-groupe du groupe linéaire $GL(n)$ caractérisant le niveau de structure de M (métrique, etc.) ajouté à sa structure différentiable. Le groupe de base est le groupe linéaire $GL(n)$. Notre structure de contact de référence \mathbb{V}_J vue comme une connexion sur la base des (x, p) en est un exemple typique. L'intuition clé est celle de la méthode des “repères mobiles” de Cartan.

Les éléments des fibres sont donc les repères $\mathfrak{e} = \{\mathfrak{e}_\mu\}$ des espaces tangents $T_x M$ (notés $\mathfrak{e}_x = \{\mathfrak{e}_{\mu,x}\}$ si la référence à x est nécessaire), les sections locales trivialisantes $\sigma : U \rightarrow \mathfrak{P}$ consistant à se donner en chaque point $x \in U$ un repère “de référence” $\sigma(x) = \mathfrak{e}_x^0$. Les déplacements verticaux dans la fibre \mathfrak{E}_x reviennent à faire un changement de repère en x . Relativement au repère de référence \mathfrak{e}_x^0 , ce changement sera décrit par un élément g de G , par exemple une matrice inversible $n \times n$ si $G = GL(n)$, une matrice de rotation si $G = SO(n)$, etc. Mais pour relier entre eux des espaces tangents voisins, il faut changer x et donc *l'origine* des repères. On passe ainsi du contexte vectoriel au contexte *affine* (cf. plus bas).

Comme nous l’avons noté plus haut à la section 3.1.2, ce qui est très spécifique des fibrés de repères et des fibrés vectoriels associés qui sont fondés sur le fibré tangent TM est la dialectique entre l’“interne” et l’“externe”, les éléments “verticaux” dans les fibres représentant en fait des déplacements “horizontaux” infinitésimaux dans la base. Les indices “verticaux” i, j, \dots sont donc *également* des indices “horizontaux” μ, ν, \dots ce qui permet de nouvelles opérations *mixant* les types d’indice !

Cette possibilité est cruciale en relativité générale (RG) où, notons-le, les connexions A se notent traditionnellement Γ . On trouve donc dans la littérature

$$\Gamma = A_{\mu}^{\alpha} \mathbf{g}_{\alpha} dx^{\mu}$$

où les \mathbf{g}_{α} sont maintenant des matrices sur les vecteurs tangents à M , $(\mathbf{g}_{\alpha})_i^j$ pouvant être noté dans ce contexte $(\mathbf{g}_{\alpha})_{\rho}^{\nu}$ puisque les indices “internes” s’identifient à des indices “externes”. Γ est une 1-forme sur M à valeurs dans $\text{End}(TM)$. Les matrices de connexion A_i^j deviennent des

$$\Gamma_{\rho}^{\sigma} = \Gamma_{\mu}^{\alpha} (\mathbf{g}_{\alpha})_{\rho}^{\sigma} dx^{\mu} = \Gamma_{\rho\mu}^{\sigma} dx^{\mu}$$

avec

$$\Gamma_{\rho\mu}^{\sigma} = \Gamma_{\mu}^{\alpha} (\mathbf{g}_{\alpha})_{\rho}^{\sigma}.$$

Les $\Gamma_{\rho\mu}^{\nu}$ sont appelés les *symboles de Christoffel*³⁵ et l’on insiste toujours dans les introductions à la relativité générale sur leur nature *non tensorielle* un peu spéciale venant du fait que les indices ρ et ν sont “verticaux” (matrices dans $\text{End}(TM)$) alors que μ est “horizontal” (1-forme sur M).

La dérivation covariante ∇ définie par la connexion Γ se définit de la même manière que précédemment. Maintenant ∇ opère sur $\mathfrak{X}(M) = \Gamma(TM)$ ³⁶, c’est-à-dire sur les champs de vecteurs tangents Y sur M (considérés toutefois comme “internes” en tant que sections de TM). Les Y remplacent les sections \mathbf{v} du cas général et ∇Y est maintenant une 1-forme à valeurs dans TM . Autrement dit, $\nabla Y(X) = Z \in \mathfrak{X}(M)$ où Y et Z sont conçus comme “internes” et X comme “externe”. Si l’on fixe le champ externe X , on obtient l’opérateur de dérivation $\nabla_X Y$ qui dérive le long de X le champ Y en tant qu’entité paramétrée par M (le fait que $Y(x)$ appartienne à $T_x M$ n’étant ici pas pertinent). La formule

$$\nabla_X \mathbf{v} = (v^j A_{j,\mu}^i + \partial_{\mu} v^i) X^{\mu} \otimes \mathbf{f}_i$$

35. Élève de Dirichlet et de Kummer (thèse en 1856), successeur de Dedekind à Zurich (1862) puis professeur à Strasbourg (alors prussienne) à partir de 1872, Erwin Bruno Christoffel (1829-1900) fut l’un des fondateurs du calcul tensoriel et l’introducteur, dans son article [120] de 1869, de la dérivation covariante associée à la connexion des variétés riemanniennes. Il fut suivi par Ricci puis Levi-Civita.

36. Attention à l’ambiguïté de Γ dénotant traditionnellement à la fois des espaces de sections et une connexion. Nous préférons garder les notations traditionnelles.

devient

$$(\nabla_X Y)^\sigma = (Y^\rho \Gamma_{\rho\mu}^\sigma + \partial_\mu Y^\sigma) X^\mu$$

et les propriétés

$$\nabla_{fX} \mathbf{v} = f \nabla_X \mathbf{v}$$

et

$$\nabla_X (f\mathbf{v}) = f \nabla_X \mathbf{v} + df(X) \otimes \mathbf{v}$$

s'écrivent maintenant

$$\begin{cases} \nabla_{fX} Y = f \nabla_X Y & (\mathcal{C}^\infty(M)\text{-linéarité}) \\ \nabla_X (fY) = f \nabla_X Y + df(X) Y & (\text{Leibniz}) \end{cases}$$

On peut alors définir le transport parallèle des vecteurs tangents le long de courbes $\gamma(t) : x_0 \rightarrow y$ dans M . En effet, soit $\eta = Y_{x_0} \in T_{x_0}M$. Pour transporter η parallèlement le long de γ on considère un champ X le long de γ tel que $X_{x_0} = \eta$ et $\nabla_{\dot{\gamma}(t)} X = 0$ le long de γ . On voit bien ici l'opération supplémentaire que permet l'identification de l'interne et de l'externe : η est un élément "interne" de la fibre au-dessus de x_0 mais il peut être aussi un vecteur tangent "externe" X .

Remarque. Il existe des champs parallèles X le long des courbes mais pas forcément le long de sous-variétés de M de dimension > 1 . Cela n'est possible que si des conditions d'intégrabilité de Frobenius sont satisfaites. \square

Le tenseur de courbure F de la connexion Γ , noté aussi traditionnellement R ³⁷, se définit comme plus haut par

$$R = d\Gamma + \Gamma \wedge \Gamma .$$

Ses composantes $R_{\rho\mu\nu}^\sigma$ dépendent de quatre indices, deux horizontaux μ, ν et deux verticaux σ, ρ .³⁸ Il satisfait

$$R_{X,Y} = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}$$

lorsqu'on l'applique à deux champs de vecteurs tangents X, Y sur M . Il satisfait aussi l'identité de Bianchi (pour R)

$$dR = [R \wedge \Gamma], \text{ i.e. } \nabla R \equiv 0 .$$

37. Ces différentes notations sont traditionnelles mais diffèrent entre les traditions géométriques et les traditions physiques. C'est pourquoi nous préférons les alterner suivant le contexte.

38. Il faudrait positionner plus précisément les indices avec des espaces, mais nous n'en avons pas besoin ici.

5.2. Le problème de la torsion

Les connexions Γ sur le fibré vectoriel TM associé au fibré des repères ont une propriété très spéciale, dite de “torsion”, qui vient de la possibilité d’identifier l’“interne” et l’“externe”. En effet, sur le fibré des repères, il existe une 1-forme Θ très particulière à valeurs dans \mathbb{R}^n . Soit $\mathbf{e}_x = \{x, \{\mathbf{e}_\mu\}\}$ un élément de \mathfrak{E}_x , c’est-à-dire un repère en $x \in M$. Si $\delta\mathbf{e}_x$ est un vecteur tangent de $T_{\mathbf{e}_x}\mathfrak{E}$ (représenté comme une variation infinitésimale de \mathbf{e}_x), $\delta\mathbf{e}_x$ est une variation *affine* $\{\delta x, \delta\{\mathbf{e}_\mu\}\}$ de \mathbf{e}_x . Comme nous l’avons vu à la section 3.2, non seulement $\{\mathbf{e}_\mu\}$ varie de $\delta\{\mathbf{e}_\mu\}$ en gardant son origine fixe, mais en plus cette origine va subir, en général, un petit déplacement δx . Les déplacements “verticaux” $\delta\mathbf{e}_x$ dans les fibres \mathfrak{E}_x *internalisent* donc des déplacements “horizontaux” δx . Ces δx “verticaux” et “internes” sont à l’origine de la torsion qui traduit la contrainte que le δx “interne” soit identique au δx “externe” ce qui, comme on dit, “soude” l’“interne” et l’“externe”. On définit Θ par

$$\Theta : \delta\mathbf{e}_x \in T_{\mathbf{e}_x}\mathfrak{E} \mapsto \{\delta x^\mu\} \in \mathbb{R}^n$$

où $\{\delta x^\mu\}$ est δx dans le repère $\{\mathbf{e}_\mu\}$ de $T_x M$ intervenant dans \mathbf{e}_x . Θ est une 1-forme équivariante et horizontale (“tensorielle”) remontant une 1-forme θ définie sur la base M . Avec une connexion linéaire Γ et la 1-forme Θ on construit une connexion *affine* sur le fibré des repères \mathfrak{E} , c’est-à-dire une connexion de \mathfrak{E} considéré comme muni du groupe structural $\text{Aff}(n)$ produit semi-direct de $GL(n)$ par les translations.

Cette 1-forme Θ est équivariante et l’on peut donc la descendre en une 1-forme θ sur M à valeurs dans TM une fois choisies des jauges locales (des sections trivialisantes). Par construction, $\theta(X) = X$ et θ s’appelle la *forme canonique* (ou “tautologique”). Elle est l’identité mais une identité en quelque sorte “dialectisée” puisqu’elle transforme X considéré comme entité “externe” en ce même X considéré comme entité “interne”.

Étant donnée alors une connexion linéaire Γ et la dérivation covariante associée ∇ , il n’y a pas a priori de raison pour que $\nabla\theta$ soit nulle. Cartan a appelé $T = \nabla\theta$, qui est une 2-forme sur M à valeurs dans TM , la *torsion* de Γ . Comme ci-dessus, on trouve

$$\left\{ \begin{array}{l} T = \nabla\theta = d\theta + \Gamma \wedge \theta \\ T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] \\ T(fX, Y) = T(X, fY) = fT(X, Y) \quad (\mathcal{C}^\infty(M)\text{-linéarité ou “tensorialité”}) \\ dT = R \wedge \theta - \Gamma \wedge T \quad (\text{identité de Bianchi pour } T) \\ \nabla T = \nabla^2\theta = dT + \Gamma \wedge T = R \wedge \theta \end{array} \right.$$

On vérifie la cohérence.

– Pour la première équation, comme $\theta \in \Lambda^1(M) \otimes TM$, on a

$$d\theta \in \Lambda^2(M) \otimes TM ,$$

et par ailleurs, comme $\Gamma \in \Lambda^1(M) \otimes \text{End}(TM)$, on a

$$\Gamma \wedge \theta \in \Lambda^2(M) \otimes \text{End}(TM)(TM) = \Lambda^2(M) \otimes TM .$$

– Pour la seconde (et troisième) équation, comme $T \in \Lambda^2(M) \otimes TM$, on a

$$T(X, Y) \in \mathfrak{X}(M) ,$$

ce qui est bien le cas des trois termes de droite.

– Pour la quatrième équation, on a

$$dT \in \Lambda^3(M) \otimes TM$$

et

$$\Gamma \wedge T \in \Lambda^3(M) \otimes \text{End}(TM)(TM) = \Lambda^3(M) \otimes TM$$

et, comme $R \in \Lambda^2(M) \otimes \text{End}(TM)$, on a aussi

$$R \wedge \theta \in \Lambda^3(M) \otimes \text{End}(TM)(TM) = \Lambda^3(M) \otimes TM .$$

– Enfin, pour la cinquième équation les deux termes sont des éléments de $\Lambda^3(M) \otimes TM$.

Remarque. Quand on dualise cette construction et que l'on considère des connexions linéaires sur le fibré cotangent T^*M , la définition de la torsion devient plus évidente car elle est simplement $T(X, Y) = \Omega(\nabla_X, \nabla_Y)$, Ω étant la 2-forme symplectique canonique de T^*M (cf. chapitre 3, section 3.1). Comme on appelle *lagrangiennes* les sous-variétés de T^*M de dimension moitié n sur lesquelles Ω est identiquement nulle,³⁹ la torsion mesure par conséquent le caractère non lagrangien des espaces horizontaux. Comme le notent Andrei Agrachev, Davide Barilari et Ugo Boscain :

“The torsion computes how much the horizontal distribution is far from being Lagrangian.” ([5], p. 556)

6. Géométrie riemannienne

6.1. Métriques et courbures

Les connexions purement géométriques où l’“interne” se dialectise avec l’“externe” et où on peut donc “mixer” les indices⁴⁰ se rencontrent en particulier en *géométrie riemannienne* avec ce que l’on appelle la connexion de Levi-Civita.

Intuitivement, une variété M de dimension n est dotée d’une structure métrique riemannienne lorsque chaque espace tangent $T_x M$ est muni d’une structure métrique, c’est-à-dire d’une forme bilinéaire symétrique définie positive, celle-ci variant différemment avec le point x . Cela se formalise au moyen d’un tenseur $g = g_{\mu\nu} dx^\mu \otimes$

39. Nous reviendrons sur les sous-variétés lagrangiennes au chapitre 12, section 6 (en particulier à la sous-section 6.1 sur les caustiques en optique).

40. Ce qui rend d’ailleurs les déluges d’indices assez pénibles à bien maîtriser.

dx^ν dans les h -repères (ou dualement $g = g^{\mu\nu}\partial_\mu \otimes \partial_\nu$), ou encore $g = g_{\mu\nu}\mathbf{e}^\mu \otimes \mathbf{e}^\nu$ (ou dualement $g = g^{\mu\nu}\mathbf{e}_\mu \otimes \mathbf{e}_\nu$) dans les nh -repères, ce qui permet de réduire le groupe structural de M au groupe $G = SO(n)$ des automorphismes euclidiens de \mathbb{R}^n .⁴¹ La métrique g définit un produit scalaire $g(X, Y)$, noté $X.Y$ ou $\langle X, Y \rangle$, sur les vecteurs tangents de $T_x M$. On peut alors déduire de g , en prenant ses dérivées et en tenant compte des constantes de structure de \mathcal{G} , une connexion métrique ∇ dite de *Levi-Civita*⁴² qui est de *torsion nulle* ce qui introduit des *symétries* dans le tenseur de courbure. ∇ est la seule connexion de torsion nulle ($\nabla_X Y - \nabla_Y X = [X, Y]$) pour laquelle la métrique est “parallèle” au sens où $\nabla g \equiv 0$. On montre que les symboles de Christoffel de ∇ définis par $\nabla_\mu \mathbf{e}_\nu = \Gamma_{\mu\nu}^\rho \mathbf{e}_\rho$ sont donnés en fonction du tenseur métrique par les formules

$$\Gamma_{\mu\nu}^\rho = \frac{1}{2} g^{\rho\sigma} (\partial_\nu g_{\sigma\mu} + \partial_\mu g_{\sigma\nu} - \partial_\sigma g_{\mu\nu})$$

et comme, d’après les formules générales, on a

$$\nabla_\nu g_{\rho\mu} = \partial_\nu g_{\rho\mu} - g_{\sigma\mu} \Gamma_{\rho\nu}^\sigma - g_{\rho\sigma} \Gamma_{\mu\nu}^\sigma,$$

il est trivial de vérifier que $\nabla_\nu g_{\rho\mu} = 0$ en utilisant la symétrie de g et le fait que $g^{\rho\sigma} g_{\rho\sigma} = 1$.⁴³ Cette formule peut s’écrire plus intrinsèquement sous la forme

$$\partial_X g((Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$

41. Il y a une ambiguïté de notation entre le tenseur métrique g et les éléments d’un groupe G , mais elle est traditionnelle.

42. Assistant de Ricci, Tullio Levi-Civita (1873-1941) fut, avec lui et Christoffel, l’un des fondateurs du calcul tensoriel et fut très proche d’Einstein dans l’usage de ces formalismes en relativité générale.

43. Dans le cas de métriques L -invariantes sur les groupes de Lie, nous verrons à la section 4.1 du chapitre 7 une autre formule donnant les symboles de Christoffel (qui dans ce cas sont constants) en fonction des constantes de structures $c_{\mu\nu}^\rho$ de bases orthonormées \mathbf{e}_μ définies par $[\mathbf{e}_\mu, \mathbf{e}_\nu] = c_{\mu\nu}^\rho \mathbf{e}_\rho$. Comme la base est orthonormée, $\langle [\mathbf{e}_\mu, \mathbf{e}_\nu], \mathbf{e}_\rho \rangle = c_{\mu\nu}^\rho$ et donc

$$\begin{aligned} \langle \nabla_\mu \mathbf{e}_\nu, \mathbf{e}_\rho \rangle &= \frac{1}{2} (\langle [\mathbf{e}_\mu, \mathbf{e}_\nu], \mathbf{e}_\rho \rangle - \langle [\mathbf{e}_\nu, \mathbf{e}_\rho], \mathbf{e}_\mu \rangle + \langle [\mathbf{e}_\rho, \mathbf{e}_\mu], \mathbf{e}_\nu \rangle) = \frac{1}{2} (c_{\mu\nu}^\rho - c_{\nu\rho}^\mu + c_{\rho\mu}^\nu) \\ \nabla_\mu \mathbf{e}_\nu &= \frac{1}{2} (c_{\mu\nu}^\rho - c_{\nu\rho}^\mu + c_{\rho\mu}^\nu) \mathbf{e}_\rho = \Gamma_{\mu\nu}^\rho \mathbf{e}_\rho \\ \Gamma_{\mu\nu}^\rho &= \frac{1}{2} (c_{\mu\nu}^\rho - c_{\nu\rho}^\mu + c_{\rho\mu}^\nu). \end{aligned}$$

qui n'est rien d'autre que la formule de Leibniz pour le produit scalaire.⁴⁴ En termes de composantes, $\nabla_X Y$ s'écrit comme une somme de deux termes

$$\begin{aligned}(\nabla_X Y)^\rho &= X^\mu (\nabla_\mu Y)^\rho = X^\mu \partial_\mu Y^\rho + X^\mu \Gamma_{\mu\nu}{}^\rho Y^\nu, \\(\nabla_\mu Y)^\rho &= \partial_\mu Y^\rho + \Gamma_{\mu\nu}{}^\rho Y^\nu \text{ si } X = \mathbf{e}_\mu\end{aligned}$$

Le premier terme explique la façon dont Y varie sous l'action de X lors d'un déplacement infinitésimal δx et la seconde comment Y varie parce que le repère mobile varie lui-même lors d'un tel déplacement infinitésimal.

Le lien entre ∇ et g , l'absence de torsion et la symétrie de g impliquent des symétries. Par exemple, l'absence de torsion $\nabla_X Y - \nabla_Y X = [X, Y]$ implique

$$\begin{aligned}&(X^\mu \partial_\mu Y^\rho + X^\mu \Gamma_{\mu\nu}{}^\rho Y^\nu) - (Y^\mu \partial_\mu X^\rho + Y^\mu \Gamma_{\mu\nu}{}^\rho X^\nu) \\&= [X, Y]^\rho + X^\mu \Gamma_{\mu\nu}{}^\rho Y^\nu - Y^\mu \Gamma_{\mu\nu}{}^\rho X^\nu \\&= [X, Y]^\rho + X^\mu \Gamma_{\mu\nu}{}^\rho Y^\nu - Y^\nu \Gamma_{\nu\mu}{}^\rho X^\mu \\&= [X, Y]^\rho + X^\mu (\Gamma_{\mu\nu}{}^\rho - \Gamma_{\nu\mu}{}^\rho) Y^\nu\end{aligned}$$

pour tout X, Y et donc $\Gamma_{\mu\nu}{}^\rho = \Gamma_{\nu\mu}{}^\rho$.

Le tenseur de courbure

$$R_{\mu\nu} = R(\mathbf{e}_\mu, \mathbf{e}_\nu) = [\nabla_\mu, \nabla_\nu] - \nabla_{[\mu, \nu]}$$

(où μ dénote \mathbf{e}_μ) est une 2-forme à valeurs dans $\text{End}(\mathfrak{X}(M))$.⁴⁵ Avec les symboles de Christoffel, la formule $R = d\Gamma + \Gamma \wedge \Gamma$ s'écrit (avec déluge d'indices)

$$R_{\sigma\mu\nu}^\rho = R(\mathbf{e}_\mu, \mathbf{e}_\nu) \partial_\sigma \otimes dx^\rho = \partial_\mu \Gamma_{\nu\sigma}{}^\rho - \partial_\nu \Gamma_{\mu\sigma}{}^\rho + \Gamma_{\mu\tau}{}^\rho \Gamma_{\nu\sigma}{}^\tau - \Gamma_{\nu\tau}{}^\rho \Gamma_{\mu\sigma}{}^\tau.$$

En termes de la matrice de connexion $\omega_\nu^\sigma(\mathbf{e}_\mu) = g(\nabla_\mu \mathbf{e}_\nu, \mathbf{e}_\sigma)$ et de la 1-forme tautologique $\theta^\mu(X) = g(\mathbf{e}_\mu, X)$, on réobtient les formules déjà connues. Mais, antisymétrique en μ, ν puisqu'il s'agit d'une 2-forme, le tenseur R possède par ailleurs de nombreuses symétries. Par exemple

$$\begin{cases} g(R(X, Y)V, W) = -g(R(X, Y)W, V) \\ g(R(X, Y)V, W) = -g(R(Y, X)V, W) \\ g(R(X, Y)V, W) = g(R(V, W)X, Y) \end{cases}$$

ou encore la première identité de Bianchi qui s'écrit maintenant

$$R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0.$$

ou encore la deuxième identité de Bianchi $\nabla R = 0$ qui s'écrit maintenant

$$\nabla_X R(Y, Z) + \nabla_Y R(Z, X) + \nabla_Z R(X, Y) = 0.$$

44. La formule de Leibniz naturelle serait $\nabla_X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$ mais pour les fonctions $\nabla_X f = \partial_X f$ et donc $\nabla_X(g(Y, Z)) = \partial_X(g(Y, Z))$.

45. Donc, si $X, Y, V \in \mathfrak{X}(M)$, alors $R(X, Y)V \in \mathfrak{X}(M)$.

Ces symétries font que le tenseur de courbure possède $\frac{n^2(n^2-1)}{12}$ composantes indépendantes.

Une certaine moyenne de la courbure, dite *courbure de Ricci*⁴⁶ est essentielle. Si les $R_{\rho\mu\nu}^\sigma$ sont les composantes du tenseur de courbure R , elle s'obtient en contractant l'indice vertical σ avec l'indice horizontal μ : $\text{Ric}_{\sigma\nu} = R_{\sigma\mu\nu}^\mu$. Une conséquence des symétries de R est que la courbure de Ricci est un tenseur symétrique : $\text{Ric}_{\mu\nu} = \text{Ric}_{\nu\mu}$ et définit donc une forme quadratique sur TM . En termes de la métrique g , le tenseur de Ricci est défini par

$$\text{Ric}(X) = \sum_{\mu} R(X, \mathbf{e}_{\mu}) \mathbf{e}_{\mu} .$$

On peut alors définir la *courbure scalaire* r en contractant les $\text{Ric}_{\mu\nu}$ avec les $g^{\mu\nu}$, i.e.

$$r = g^{\mu\nu} \text{Ric}_{\mu\nu} .$$

C'est la trace du tenseur de Ricci par rapport à la métrique.

6.2. Géodésiques riemanniennes

Nous utiliserons de façon essentielle par la suite la notion de géodésique en géométrie sous-riemannienne en relation avec le calcul variationnel, le contrôle optimal, les notions de lagrangien et d'hamiltonien. Mais profitons de cette section sur la géométrie riemannienne pour anticiper sur certains aspects et donner leur définition classique.

Une géodésique est une courbe $\gamma(t)$ qui minimise la longueur $\int_{\gamma} ds = \int_{\gamma} \dot{\gamma} dt$ et aussi l'énergie $\int_{\gamma} |\dot{\gamma}|^2 dt$. Nous y reviendrons. Lorsqu'on la paramétrise par la longueur d'arc, c'est aussi une courbe $\gamma(s)$ qui transporte son vecteur tangent $\dot{\gamma}$ parallèlement à lui-même c'est-à-dire qui satisfait l'équation du second ordre

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0 .$$

En termes de coordonnées locales, l'équation s'explique en

$$\ddot{x}^{\rho} + \Gamma_{\mu\nu}^{\rho} \dot{x}^{\mu} \dot{x}^{\nu} = 0 .$$

Si $\zeta(s)$ est une courbe régulière quelconque, $k_{\zeta}(s) = \left\| \nabla_{\dot{\zeta}} \dot{\zeta} \right\|$ est la courbure géodésique de $\zeta(s)$ et donc $\gamma(s)$ est une géodésique si sa courbure géodésique est nulle.

Les éléments métriques standard de courbure (courbure de Riemann, courbure sectionnelle, courbure scalaire, courbure de Ricci) permettent donc de calculer les géodésiques. Si l'on considère ce calcul comme un problème "direct" (on connaît la

46. Assistant de Ulisse Dini à la Scuola Normale Superiore de Pise où il devint ensuite professeur, Gregorio Ricci-Curbastro (1853-1925) fut, avec Elwin Bruno Christoffel et son assistant Tullio Levi-Civita, l'un des fondateurs du calcul tensoriel.

métrique et on en déduit les géodésiques), on peut alors s'interroger sur le problème "inverse" (on connaît les courbes géodésiques et on en déduit des informations sur la métrique). C'est ce qu'on fait en physique en observant les trajectoires gravitationnelles et en postulant qu'elles sont géodésiques pour en déduire des informations sur la métrique. Autrement dit, la distribution des géodésiques encode des éléments de métrique.

Quand on considère plusieurs géodésiques, un certain nombre de scénarios peuvent se produire. Un cas particulièrement intéressant est lorsqu'une géodésique γ issue de a au temps $s = 0$ perd sa minimalité globale à la traversée d'un point b parce qu'une autre géodésique $\tilde{\gamma}$ issue de a entre en compétition avec elle au point b . Le point b est appelé un *point de coupure* ("cut point") et ces points engendrent ce que l'on appelle le "cut locus" de a . L'exemple standard est celui des méridiens allant du pôle Nord au pôle Sud sur une sphère. Il y a aussi la possibilité que γ perde sa minimalité locale en un point c . parce que les géodésiques ont une enveloppe et que γ devient tangente à cette enveloppe en c . On dit alors que le point c est *conjugué* de a . Le pôle Sud d'une sphère est conjugué du pôle Nord mais l'enveloppe est dégénérée en un point.

Un outil important dans l'analyse des variations de géodésiques est celui des *champs de Jacobi*.⁴⁷ Intuitivement un champ de Jacobi $J(s)$ est un champ de vecteurs le long d'une géodésique $\gamma(s)$ qui la déforme infinitésimalement en une autre géodésique. Autrement dit, si $\gamma_\alpha(s)$ est une famille de géodésiques paramétrées par $\alpha \in [\alpha_0, \alpha_1]$, $J(s) = \frac{\partial \gamma_\alpha(s)}{\partial \alpha} \upharpoonright_{\alpha=0}$ est un champ de Jacobi. Si les géodésiques $\gamma_\alpha(s)$ ont des extrémités fixes et vont toutes d'un point a à un point b , alors $J(a) = J(b) = 0$.

L'équation que satisfont les champs de Jacobi se déduit immédiatement de l'équation des géodésiques

$$\ddot{x}^\rho + \Gamma_{\mu\nu}{}^\rho(x) \dot{x}^\mu \dot{x}^\nu = 0 .$$

Soit J^μ les composantes de J , les coordonnées x^μ étant maintenant paramétrées par α . On intervertit les dérivations en s et en α dans

$$J(s) = \frac{\partial}{\partial \alpha} (\ddot{x}_\alpha^\rho + \Gamma_{\mu\nu}{}^\rho(x_\alpha) \dot{x}_\alpha^\mu \dot{x}_\alpha^\nu) \upharpoonright_{\alpha=0}$$

et on obtient

$$\ddot{J}^\rho + \Gamma_{\mu\nu}{}^\rho \left(J^\mu \dot{x}_\alpha^\nu + \dot{J}^\nu \dot{x}_\alpha^\mu \right) + \frac{\partial \Gamma_{\mu\nu}{}^\rho(x_\alpha)}{\partial x_\alpha^\sigma} J^\sigma \dot{x}_\alpha^\mu \dot{x}_\alpha^\nu .^{48}$$

47. Cf. Barilari-Rizzi [33].

48. $\Gamma_{\mu\nu}{}^\rho \left(\dot{J}^\mu \dot{x}_\alpha^\nu + \dot{J}^\nu \dot{x}_\alpha^\mu \right) = 2\Gamma_{\mu\nu}{}^\rho \dot{J}^\mu \dot{x}_\alpha^\nu$ à cause de la symétrie $\Gamma_{\mu\nu}{}^\rho = \Gamma_{\nu\mu}{}^\rho$.

Nous reviendrons plusieurs fois sur ces notions, en particulier au chapitre 12 section 6 à propos des caustiques et de leur interprétation lagrangienne et hamiltonienne, ainsi qu'au chapitre 14 section 6.1 à propos des géodésiques sous-riemanniennes.

6.3. L'exemple de la sphère \mathbb{S}^2

Pour prendre un exemple simple considérons la sphère euclidienne \mathbb{S}^2 de rayon 1 dans \mathbb{R}^3 . En fait l'exemple est moins simple qu'il n'y paraît. Nous avons appris à l'école que les géodésiques sont les grands cercles, mais il n'est pas évident de le voir directement à partir des formules générales de la géométrie riemannienne.⁴⁹

Soient (θ, φ) les coordonnées sphériques de \mathbb{S}^2 . L'angle $\theta \in [0, 2\pi]$ est la longitude, et l'angle $\varphi \in [0, \pi]$ est la latitude, comptée de 0 (le pôle Nord $z = 1$) jusqu'à π (le pôle Sud $z = -1$). Le plongement j de \mathbb{S}^2 dans \mathbb{R}^3 est de coordonnées

$$(x = \cos(\theta) \sin(\varphi), y = \sin(\theta) \sin(\varphi), z = \cos(\varphi))$$

et son application linéaire tangente est

$$Tj = \begin{pmatrix} \partial_\theta x & \partial_\varphi x \\ \partial_\theta y & \partial_\varphi y \\ \partial_\theta z & \partial_\varphi z \end{pmatrix} = \begin{pmatrix} -\sin(\theta) \sin(\varphi) & \cos(\theta) \cos(\varphi) \\ \cos(\theta) \sin(\varphi) & \sin(\theta) \cos(\varphi) \\ 0 & -\sin(\varphi) \end{pmatrix}.$$

On a donc les vecteurs colonnes

$$\begin{cases} Tj(\partial_\theta) = Tj((1, 0)) = (-\sin(\theta) \sin(\varphi), \cos(\theta) \sin(\varphi), 0) \\ Tj(\partial_\varphi) = Tj((0, 1)) = (\cos(\theta) \cos(\varphi), \sin(\theta) \cos(\varphi), -\sin(\varphi)) \end{cases}$$

La métrique euclidienne sur \mathbb{R}^3 est la forme quadratique

$$g_R = dx \otimes dx + dy \otimes dy + dz \otimes dz$$

et elle induit la métrique g de \mathbb{S}^2 par les formules

$$\begin{cases} g_{11} = g(\partial_\theta, \partial_\theta) = g_R(Tj(\partial_\theta), Tj(\partial_\theta)) = \sin^2(\varphi) \\ g_{12} = g_{21} = g(\partial_\theta, \partial_\varphi) = g_R(Tj(\partial_\theta), Tj(\partial_\varphi)) = 0 \\ g_{22} = g(\partial_\varphi, \partial_\varphi) = g_R(Tj(\partial_\varphi), Tj(\partial_\varphi)) = 1 \end{cases}$$

On en tire immédiatement les symboles de Christoffel

$$\begin{cases} \Gamma_{11}^1 = \Gamma_{22}^1 = \Gamma_{22}^2 = \Gamma_{12}^2 = \Gamma_{21}^2 = 0 \\ \Gamma_{21}^1 = \Gamma_{12}^1 = \cot(\varphi) \\ \Gamma_{11}^2 = -\sin(\varphi) \cos(\varphi) \end{cases}$$

et les équations des géodésiques

49. Nous reviendrons sur les géodésiques de la sphère et plus généralement des ellipsoïdes à la section 6.2 du chapitre 12.

$$\begin{cases} \ddot{\theta} + 2 \cot(\varphi) \dot{\theta} \dot{\varphi} = 0 \\ \ddot{\varphi} - \sin(\varphi) \cos(\varphi) (\dot{\theta})^2 = 0 \end{cases}$$

Ces équations ne sont pas si simples mais le deviennent si on fait un changement de variable convenable. Nous y reviendrons plus bas dans la section 6.2 du chapitre 12.

6.4. Note sur la relativité générale

Profitons de cette incursion dans la géométrie riemannienne pour faire quelques brèves remarques sur la relativité générale (RG). Ce que nous venons de voir se généralise aux variétés pseudo-riemanniennes où la métrique n'est plus forcément définie > 0 . Cela a permis à Einstein de définir son célèbre tenseur

$$G_{\mu\nu} = \text{Ric}_{\mu\nu} - \frac{1}{2} r g_{\mu\nu}$$

et d'écrire sa célébriissime équation $G = \frac{8\pi\mathfrak{G}}{c^4} T$ où T est le tenseur d'impulsion-énergie des champs occupant l'espace-temps, \mathfrak{G} la constante gravitationnelle et c la vitesse de la lumière. La genèse de cette idée est bien connue. L'identité de la masse grave et de la masse inertielle conduisit Einstein à l'hypothèse que tous les mouvements sont inertiels et décrivent donc des géodésiques de l'espace-temps, le mouvement de la matière manifestant la courbure de l'espace-temps. Cela exige que la géométrie de l'espace-temps soit pseudo-riemannienne avec une métrique variable localement minkowskienne ayant pour source le tenseur d'impulsion-énergie T . Mais, côté physique, ce tenseur satisfait un *principe de conservation* $\nabla T \equiv 0$ où ∇ est la dérivation covariante de la connexion métrique de Levi-Civita. Le problème est donc de trouver un tenseur géométrique G de même type que T et satisfaisant pour des raisons purement géométriques (synthétiques a priori) l'identité $\nabla G \equiv 0$. On posera alors une relation de proportionnalité $G = \alpha T$. Si l'on impose d'une part que $G \equiv 0$ dans un espace-temps plat⁵⁰ et d'autre part que G soit constructible linéairement à partir du tenseur métrique et de ses dérivées premières, alors le tenseur d'Einstein est la seule solution, l'identité $\nabla G \equiv 0$ venant de l'identité de Bianchi. L'analyse de la limite newtonienne donne alors pour la constante de proportionnalité α la valeur, $\alpha = \frac{8\pi\mathfrak{G}}{c^4}$.⁵¹

50. En effet, la courbure venant de T , si $T \equiv 0$ l'espace-temps doit être sans courbure et $G = \alpha T \equiv 0$. En fait, cette hypothèse est loin d'être évidente car elle renvoie au problème de la "constante cosmologique" qui mesure la densité moyenne d'énergie de l'état "vide" de l'espace-temps. Les doutes d'Einstein à son sujet sont bien connus.

51. Avec la constante cosmologique Λ , l'équation devient $G + \Lambda g = \frac{8\pi\mathfrak{G}}{c^4} T$. Λ permet des modèles d'univers statiques mais comme l'univers est en expansion ce n'est pas une bonne hypothèse.

Dans *Gravitation* [369] John Archibald Wheeler (avec Charles Misner et Kip Thorne) insiste sur le fait que, dans la mesure où l'identité de Bianchi vient de la cohomologie des formes différentielles ($d^2 = 0$, $\partial^2 = 0$ et théorème de Stokes), elle rend *a priori* les lois de conservation en RG. Wheeler est très explicite (presque lyrique) à ce propos :

“This conservation is not an accident. According to Einstein and Cartan, it is ‘automatic’; and automatic, moreover, as a consequence of exact equality between energy-momentum and an automatically conserved feature of the geometry”.

“Thus simply is all of general relativity tied to the principle that the boundary of a boundary is zero. No one has ever discovered a more compelling foundation for the principle of conservation of momentum and energy. No one has ever seen more deeply into that action of matter on space, and space on matter, which one calls gravitation. In summary, *the Einstein theory realizes the conservation of energy-momentum as the identity, ‘the boundary of a boundary is zero’.*” ([369], p. 380)⁵²

6.5. Note sur parallélisme “absolu” (Cartan-Einstein)

Même si les connexions métriques sont sans torsion, nous avons vu cependant que les connexions équivariantes sur le fibré des repères de l'espace de base M auront en général de la torsion. Élie Cartan a généralisé la RG en investiguant cette remarquable possibilité (cf. en particulier le long mémoire “Sur les variétés à connexion affine, et la théorie de la relativité généralisée” [95] de 1923-1925) (théorie d'Einstein-Cartan).⁵³

Entre 1928 et 1931, Einstein tenta à son tour une *unification* de la gravitation (RG) et de l'électro-magnétisme (EM)⁵⁴ en introduisant un “parallélisme à distance” ou “téléparallélisme” (“Fern-Parallelismus”) permettant de comparer les *directions* en des points différents de l'espace-temps. L'annonce d'un enrichissement de la RG de 1915 eut un écho retentissant, y compris médiatique. Le lecteur intéressé pourra consulter par exemple les articles de Tilman Sauer [487], de Friedrich Hehl, Yuri Obukhov [244], ou d'Erhard Scholz [491].

52. Dans ses *Philosophical Problems of Space and Time* ([233], chapitre 22 : *General relativity, Geometrodynamics and Ontology*, Adolf Grünbaum critique l'a priorisme de la géométrie dynamique de Wheeler pour défendre un conventionnalisme à la Poincaré. Le lecteur intéressé pourra se référer à notre étude [419] ainsi qu'à sa bibliographie.

53. Le sujet est pour nous très éclairant et pas du tout hors sujet car il montre encore une fois à quel point les modèles neurogéométriques fondés sur la notion de parallélisme infinitésimal sont en résonance avec des problèmes de fondements géométriques déjà rencontrés dans d'autres disciplines.

54. Ce que Weyl avait déjà essayé de faire, mais sans succès (cf. section 3.2.2).

Il y eut plusieurs articles avant qu'Einstein n'abandonne cette perspective suite à la reformulation de l'EM par la mécanique quantique. Les articles initiaux de juin 1928 mettaient en place la théorie. L'article de synthèse 1930 aux *Mathematische Annalen* [163] ainsi que sa reprise (issue de conférences données à l'I.H.P. en novembre 1929) "Théorie unitaire du champ physique" [164] sont particulièrement intéressants. Einstein veut unifier la RG et l'EM sur la base d'un enrichissement de la structure métrique de l'espace-temps mais sans introduire pour autant "deux états" indépendants de l'espace-temps, l'un qui serait défini par la métrique de la RG et l'autre par le potentiel vecteur du champ EM. Il note alors qu'une métrique riemannienne permet de comparer les longueurs en des points différents, mais que

"on ne peut pas faire la même chose pour leurs *directions* ; dans la géométrie riemannienne il n'existe pas de parallélisme à distance." ([164], p. 3)

L'idée est par conséquent d'introduire en plus, en chaque point x de l'espace-temps, un repère orthonormé $\mathbf{e}(x)$ variant différentiablement avec x et permettant de repérer les directions.⁵⁵ Autrement dit, Einstein considère, en plus de la métrique, une section $\mathbf{e}(x)$ du fibré des nh -repères orthonormés (au sens de la section 3.1.3 ci-dessus), c'est-à-dire n champs de vecteurs linéairement indépendants $\mathbf{e}_s(x)$ ⁵⁶ et il décide que les $\mathbf{e}(x)$ sont parallèles, autrement dit que deux vecteurs tangents $X \in T_x M$ et $Y \in T_y M$ sont parallèles s'ils ont les mêmes composantes dans $\mathbf{e}(x)$ et $\mathbf{e}(y)$. Ces repères ne sont bien sûr définis qu'à une rotation globale et commune près.

Mais maintenant, *dans l'autre sens*, on peut considérer une section $\mathbf{e}(x)$ du fibré des nh -repères et décider qu'ils sont *à la fois* orthonormaux *et* parallèles. On obtient ainsi une géométrie sur l'espace-temps comme variété différentiable de dimension 4. Le fibré des repères possède alors comme groupe structural le sous-groupe $SE(n)$ (translations et rotations) du groupe affine $\text{Aff}(n)$.⁵⁷

Il est intéressant de revenir aux sources. Einstein distingue bien les bases holonomes des plans tangents $T_x M$ (coordonnées locales x^μ et h -repères $\{\partial_\mu\}$, cf. section 3.1.3) et les bases non holonomes $\mathbf{e}(x)$ (nh -repères). En termes des $\{\partial_\mu\}$, les $\mathbf{e}(x)$ seront définis par les composantes h_s^μ de leurs vecteurs unitaires (le h -indice s numérotant les axes de $\mathbf{e}(x)$ et le nh -exposant μ numérotant les composantes du vecteur unitaire). Autrement dit $\mathbf{e}_s(x) = h_s^\mu \partial_\mu$ et les h_s^μ forment la matrice de changement de base $\{\partial_\mu\} \rightsquigarrow \mathbf{e}(x)$. On récupère les $g_{\mu\nu}$ de la métrique par la formule

55. Cela s'appelle un *Vierbein* en dimension 4 (n -Bein en dimension n). On parle aussi de "tétrade" pour $n = 4$.

56. Pour que ces champs soient globaux il faut supposer l'espace-temps parallélisable (cf. chapitre 5, section 2.3). Si ce n'est pas le cas il faut recoller des morceaux parallélisables.

57. Comme il s'agit de la physique de l'espace-temps, on est en dimension 4 et la métrique est minkowskienne : les rotations ne sont pas celles de $SO(4)$ mais celles de $SO(3,1)$. Mais nous faisons comme si on était en dimension n avec $SO(n)$.

$g_{\mu\nu} = h_{s\mu}h_{s\nu}$ que l'on déduit de $g(\mathbf{e}_s, \mathbf{e}_t) = \delta_{st}$ (ou η_{st} dans le cas minkowskien) en abaissant les indices des h_s^μ .⁵⁸ Il y a 16 h_s^ν alors qu'il n'y a que 10 $g_{\mu\nu}$ à cause de la symétrie. Ces degrés de liberté supplémentaires permettent d'enrichir la physique associée à cette géométrie.

“En résumé, la seule hypothèse nouvelle à introduire pour arriver à une géométrie plus complète que celle de Riemann, concerne l'existence de “directions” dans l'espace et de relations entre ces directions. Cette notion de “direction” n'est pas contenue, ni dans la notion de continu, ni dans celle d'espace. Il faut donc une hypothèse supplémentaire pour admettre qu'il y a dans l'espace quelque chose comme les relations de direction, exprimées par l'existence d'un parallélisme à distance finie.” ([164], p. 4)

À partir de cette hypothèse, Einstein va généraliser le formalisme de la RG (dérivation covariante, etc.), trouver avec virtuosité des équations du champ (c'est l'essentiel de l'article) et montrer “qu'en première approximation” on retrouve les équations du champ gravitationnel de la RG et les équations de Maxwell de l'EM. Dans ces géométries, il y a donc deux connexions, la connexion affine ∇ de coefficients de Christoffel Δ associée au champ $\mathbf{e}(x)$ et la connexion Γ de Levi-Civita de la métrique riemannienne (compatible avec Δ). Δ est donnée par

$$\Delta_{\alpha\beta}^\mu = h_s^\mu \partial_\beta h_{s\alpha} ,$$

formule qui montre que, contrairement au cas des connexions métriques, les coefficients de Christoffel *ne sont pas symétriques* en (α, β) . Le téléparallélisme signifie que $\nabla_{\mathbf{e}_s} \mathbf{e}_t = 0$ pour tout couple d'indice.

Einstein écrit en termes des h_s^μ et de leurs dérivées partielles la dérivation covariante, la courbure et la torsion de la connexion Δ ainsi que la courbure de la métrique g avec l'équivalent des formules vues ci-dessus. Mais le parallélisme à distance signifie que la courbure de Δ donnée par $R_\Delta = d\Delta + \Delta \wedge \Delta$ est *identiquement nulle* alors que sa torsion donnée par $T_\Delta = d\theta + \Delta \wedge \theta$ n'est pas nulle en général.

Dans les débats de l'époque, on expliquait les choses de façon à la fois formelle et intuitive. Dans une connexion affine “euclidienne”, quand on déplace parallèlement un repère $\mathbf{e}(x)$ d'un point x à un point infiniment voisin $x + dx$, l'origine du repère subira une petite translation et le repère tournera un peu. Si l'on se déplace le long d'un circuit infinitésimal, on obtiendra ainsi une translation infinitésimale et une rotation infinitésimale, ces éléments ne dépendant pas du lacet. D'où une 1-forme sur M à valeurs dans l'algèbre de Lie \mathbb{R}^n du groupe des translations ainsi qu'une 1-forme sur M à valeurs dans l'algèbre de Lie $\mathfrak{so}(n)$ du groupe des rotations, bref

58. $\mathbf{e}_s(x) = h_s^\mu \partial_\mu$ équivaut à $\partial_\mu = h_{s\mu} \mathbf{e}_s$. Donc

$$g_{\mu\nu} = g(\partial_\mu, \partial_\nu) = g(h_{s\mu} \mathbf{e}_s, h_{t\nu} \mathbf{e}_t) = h_{s\mu} h_{t\nu} g(\mathbf{e}_s, \mathbf{e}_t) = h_{s\mu} h_{t\nu} \delta_{st} = h_{s\mu} h_{s\nu} .$$

une 1-forme sur M à valeurs dans l'algèbre de Lie $\mathfrak{se}(n)$ du groupe des déplacements $SE(n)$. Il en va de même pour la version minkowskienne.

La donnée d'un parallélisme à *distance* signifie que le déplacement parallèle infinitésimal est *intégrable* au sens où quand on transporte parallèlement à lui-même un vecteur tangent le long d'un circuit il revient au point de départ identique à lui-même. En revanche le fait que la torsion soit non nulle signifie que les parallélogrammes infinitésimaux ne se referment pas. Intuitivement, si l'on considère deux déplacements infinitésimaux δx et $\mathfrak{d}x$ à partir de x et si l'on transporte parallèlement $\mathfrak{d}x$ en $x + \delta x$ et δx en $x + \mathfrak{d}x$, $(x + \delta x) + \mathfrak{d}x$ ne s'identifie pas à $(x + \mathfrak{d}x) + \delta x$. Il sont séparés par une translation infinitésimale. En revanche pour une métrique riemannienne, le déplacement parallèle infinitésimal n'est pas intégrable (quand on transporte parallèlement à lui-même un vecteur tangent le long d'un circuit, il revient au point de départ avec une variation) et la courbure mesure justement cette non intégrabilité. En revanche les parallélogrammes se ferment et il n'y a pas de torsion.

Bien que Δ et Γ ne soient pas des tenseurs, $\Gamma - \Delta$ et $\Lambda = \frac{1}{2}(\Delta - \hat{\Delta})$ (où $\hat{\Delta}_{\alpha\beta}^{\mu} = \Delta_{\beta\alpha}^{\mu}$) le sont en revanche. Le tenseur Λ mesure l'asymétrie de Δ et il est de composantes $\Lambda_{\alpha\beta}^{\mu} = \frac{1}{2}h_s^{\mu}(\partial_{\beta}h_{s\alpha} - \partial_{\alpha}h_{s\beta})$. Einstein propose alors d'identifier le tenseur *contracté* $\Lambda_{\alpha\beta}^{\alpha} = \frac{1}{2}h_s^{\alpha}(\partial_{\beta}h_{s\alpha} - \partial_{\alpha}h_{s\beta})$ au champ électro-magnétique. La virtuosité d'Einstein est saisissante. Guidé par des principes analogues à ceux qui l'avaient conduit en 1915 à la RG ainsi que par de puissantes heuristiques, il prospecta des possibilités d'équations des champs gravitationnel et EM à partir de lagrangiens appropriés et de principes variationnels. Les idées directrices étaient

1. l'obtention à partir de la connexion d'un tenseur de même type que celui du tenseur énergie-impulsion,
2. la covariance générale,
3. la conservation de l'énergie-impulsion,
4. la limite newtonienne de la RG et la limite maxwellienne de l'EM,
5. l'existence de solutions interprétables comme des particules, les équations du champ donnant dans ce cas les équations du mouvement de ces particules.

De telles connexions de courbure nulle et ne possédant donc que de la torsion avaient déjà été introduites par Roland Weitzenböck.⁵⁹ En revanche, la connexion de Levi-Civita Γ de g est de torsion T_{Γ} identiquement nulle mais de courbure R_{Γ} en général non nulle. Il y a donc une complémentarité entre les deux connexions, celle exprimant le parallélisme des directions et celle exprimant les longueurs.

59. Le mathématicien autrichien Roland Weitzenböck (1885-1955) était un spécialiste de géométrie différentielle. Mais, traumatisé par la première guerre mondiale il restait ennemi de la France et n'intégra pas Cartan dans ses ouvrages.

La théorie unifiée d'Einstein fondée sur le téléparallélisme suscita de nombreuses discussions et correspondances, par exemple avec Herman Müntz (1884-1956), ses collègues à Berlin Jakob Grommer et Cornelius Lanczos (1893-1974, son assistant en 1928-29), ou Walther Meyer (1887-1948, également son assistant en 1929 et qui le suivit à Princeton). Le débat avec Hermann Weyl fut serré. Weyl estimait pour sa part que le caractère “à distance” du “Fernparallelismus” était incompatible avec le principe imposant la construction de la géométrie uniquement à partir de transformations infinitésimales, alors qu'Einstein estimait de son côté que la première théorie de jauge de Weyl n'était guère convaincante.

Mais ce furent les échanges avec Élie Cartan qui eurent une importance particulière. En effet, dès qu'il prit connaissance de ces travaux, Cartan écrivit à Einstein (lettre du 8 mai 1929) que son modèle était un exemple de géométrie qu'il avait développé précédemment en particulier dans ses articles de 1922 [94] et 1923 [95], précisément sur les connexions affines, la possibilité de torsion et l'extension de la RG. Einstein en convint et lui proposa d'ajouter une note historique à son article [163] des *Mathematische Annalen*.

Bien qu'il ait pu retrouver au premier ordre ses équations de la RG et les équations de Maxwell, Einstein abandonna par la suite ce projet d'unification. La mécanique quantique, et en particulier l'équation de Dirac pour l'électron, conduisait en effet à revoir profondément les choses. C'est ce que fit d'ailleurs Weyl en repensant complètement sa théorie de jauge comme nous l'avons vu dans la section 4.2.

7. Géométrie sous-riemannienne (prélude)

Revenons à la neurogéométrie. Cet excursus dans le riche univers de la géométrie des connexions était justifié par le fait que les plans de contact “verticaux” de la structure de contact \mathcal{K} de \mathbb{V}_J sont interprétables comme les plans “horizontaux” d'une connexion. Nous allons maintenant introduire un autre concept fondamental de la neurogéométrie, celui de *géométrie sous-riemannienne*. Nous le ferons ici de façon très sommaire, à titre de prélude, car nous y reviendrons très longuement dans les chapitres suivants. Présentons-le d'abord en dimension 3 pour \mathcal{K} et \mathbb{V}_J , l'algèbre de Lie \mathcal{V}_J étant nilpotente avec un seul commutateur non nul

$$\begin{cases} [\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3 \\ \mathbf{t}_1 = \frac{\partial}{\partial x} + p \frac{\partial}{\partial y} \\ \mathbf{t}_2 = \frac{\partial}{\partial p} \\ \mathbf{t}_3 = -\frac{\partial}{\partial y} \end{cases}$$

et \mathcal{K} étant le noyau de la 1-forme de contact

$$\omega_J = dy - p dx$$

(invariante par les translations à gauche de \mathbb{V}_J et telle que $\omega_J \wedge d\omega_J \neq 0$). Rappelons que le champ de vecteurs $-\mathbf{t}_3 = \frac{\partial}{\partial y} = \chi$ est le champ caractéristique – le champ de Reeb – de la distribution \mathcal{K} de plans de contact $\text{Span}\{\mathbf{t}_1, \mathbf{t}_2\}$. Il est partout transverse à \mathcal{K} et satisfait $\omega_J(\chi) = 1$. Nous avons vu à la section 3.3 du chapitre 3 comment on pouvait introduire différentes métriques sur \mathbb{V}_J en modifiant ω_J en $f\omega_J$ ($f \neq 0$) (ce qui ne change pas \mathcal{K} mais change le champ de Reeb χ) et en prenant une base de \mathcal{K} augmentée de χ comme base orthogonale de \mathbb{V}_J . Nous allons approfondir cette idée.

On peut définir sur la distribution de contact \mathcal{K} des métriques très spéciales appelées par Gromov, Lafontaine et Pansu [397] des métriques de “Carnot-Carathéodory”,⁶⁰ métriques qui sont des cas particuliers de géométries sous-riemanniennes sur certains groupes de Lie. L’idée de géométrie sous-riemannienne est simple. Dès que l’on dispose d’une distribution \mathcal{K} de plans dans le fibré tangent TV de la variété V , on peut considérer des métriques $g_{\mathcal{K}}$ définies seulement sur les plans de \mathcal{K} , et non pas sur tout le fibré tangent TV comme c’est le cas pour les métriques riemanniennes, et se restreindre à la considération de courbes Γ dans V qui sont tangentes à \mathcal{K} , autrement dit qui sont des courbes intégrales de \mathcal{K} .⁶¹ On distingue par conséquent une classe spéciale de courbes qualifiées d’“admissibles” (qui, dans le cas des fibrations $\pi : V \rightarrow M$ envisagées jusqu’ici, sont les relevées legendriennes des courbes régulières dans M). Soient v et v' deux points de V . Pour tenir compte de la contrainte d’intégrabilité, il est naturel de définir leur distance $d_{\mathcal{K}}(v, v')$ comme l’*inf* des $g_{\mathcal{K}}$ -longueurs des courbes intégrales de \mathcal{K} joignant v à v' . Mais cette définition n’est consistante que si l’on peut montrer que, étant donnés deux points *quelconques* v et v' de V , il existe toujours des courbes *admissibles* les joignant. Un célèbre théorème, le *théorème de Chow*, dit que de telles courbes existent toujours lorsque les crochets de Lie de \mathcal{K} engendrent le fibré tangent complet TV . Il s’agit de la condition d’Hörmander (de “bracket generating”) que nous avons explicitée à la section 3.3.

Si la condition d’Hörmander est vérifiée, on peut alors définir la distance $d_{\mathcal{K}}$ par la formule :

$$d_{\mathcal{K}}(v, v') = \inf_{\Gamma \in \mathfrak{J}(v, v')} \int_I \|\Gamma'(s)\| ds$$

où $I = [0, 1]$ et $\mathfrak{J}(v, v')$ est l’ensemble des courbes intégrales $\Gamma(s)$ de \mathcal{K} joignant $\Gamma(0) = v$ à $\Gamma(1) = v'$. Une *géodésique* entre v et v' pour la métrique de Carnot-Carathéodory est alors une courbe intégrale de \mathcal{K} qui réalise la distance $d_{\mathcal{K}}(v, v')$.

60. Comme nous l’avons souligné à la section 2.8 du chapitre 3, ce nom vient de la théorie des processus adiabatiques en thermodynamique.

61. Par exemple pour \mathbb{V}_J on peut prendre la métrique pour laquelle $\{t_1, t_2\}$ est une base orthonormale. Par construction, elle est invariante à gauche.

Une métrique de Carnot-Carathéodory $g_{\mathcal{K}}$ est une “path-metric” appelée *sous-riemannienne* car elle n’est définie que sur le *sous-fibré* \mathcal{K} du fibré tangent TV . Bien sûr, on peut l’étendre à une métrique riemannienne g en posant que le champ caractéristique χ de \mathcal{K} (défini par $\omega(\chi) = 1$) est unitaire et orthogonal à \mathcal{K} . La structure de contact s’appelle alors une *polarisation* de g . Mais elle est fondamentalement différente de g . On peut la penser comme une *limite* de métriques riemanniennes g_{ε} sur V qui pénalisent de plus en plus le défaut d’intégrabilité des courbes, c’est-à-dire le défaut de tangence à \mathcal{K} , les courbes non intégrales devenant interdites à la limite. Plus précisément, on considère la décomposition $TV = \mathcal{K} \oplus \chi$, une base orthonormale $B_{\mathcal{K}}$ de \mathcal{K} , une base B_{χ} de χ et on prend $B_{\mathcal{K}} \cup_{\varepsilon} B_{\chi}$ comme base orthonormale d’une métrique riemannienne g_{ε} . Le ds_{ε}^2 de g_{ε} est alors $ds_{\mathcal{K}}^2 + \frac{1}{\varepsilon^2} ds_{\chi}^2$ et lorsque $\varepsilon \rightarrow 0$ les seules courbes admissibles (c’est-à-dire de longueur localement finie) sont de $ds_{\chi}^2 = 0$ et sont donc des courbes intégrales de \mathcal{K} .

La “path-metric” $g_{\mathcal{K}}$ est hautement *anisotrope* et non homogène, *singulière* et même *fractale*. En effet, on peut montrer que, même si V est topologiquement de dimension 3, sa dimension de Hausdorff relativement à $g_{\mathcal{K}}$ est 4.

Pour notre propos ici, il est très important que la métrique sous-riemannienne $g_{\mathcal{K}}$ soit une limite de métriques riemanniennes g_{ε} de plus en plus anisotropes définies en introduisant des différences de “poids” entre les directions respectivement tangentes et orthogonales aux plans de contact K_v . En effet, nous pouvons facilement implémenter cette différence en posant que les connexions connectant les neurones (a, p) correspondant aux plans de contact K_v sont “fortes”, alors que celles connectant des neurones (a, p) correspondant à la direction caractéristique χ_v sont très “faibles”, voire même *inhibitrices*. Lorsque que des connexions neuronales implémentent des structures géométriques, leurs “poids synaptiques” (positifs ou négatifs, et plus ou moins importants) permettent de moduler les propriétés de cette géométrie.

8. Groupes de Carnot et variétés de Heisenberg

La géométrie sous-riemannienne du groupe des déplacements $G = SE(2) = \mathbb{V}_G$ est plus complexe que celles de $J^1\mathbb{R}^2$ ou de \mathbb{H} et cela peut se lire directement sur les algèbres de Lie, celle \mathcal{H} de \mathbb{H} étant nilpotente alors que celle \mathcal{G} de G ne l’est pas. Nous voudrions maintenant éclairer un peu ce point. Pour la géométrie sous-riemannienne des groupes de Lie en dimension 3 munis d’une structure de contact, le groupe de Heisenberg joue le rôle de “structure tangente” universelle. De même que les variétés riemanniennes sont *localement*, mais non globalement, des espaces euclidiens et s’identifient dans un voisinage infinitésimal de chaque point à leur espace tangent muni de sa géométrie euclidienne, de même les groupes de Lie que nous considérons sont des “variétés de Heisenberg” qui s’identifient dans un voisinage infinitésimal de chacun de leurs points à leur “cône tangent” qui est isomorphe

au groupe de Heisenberg muni de sa géométrie sous-riemannienne. Les références fondamentales que nous avons consultées dans ce domaine sont Mikhaïl Gromov, Pierre Pansu, André Bellaïche et Andrei Agrachev. Une bonne synthèse est le cours de Marius Buliga [83] *Sub-Riemannian Geometry and Lie Groups*.

8.1. Les groupes de Carnot et leur propriétés d'homogénéité

Le groupe de Heisenberg est le plus simple des groupes dits “de Carnot”. Les groupes de Carnot sont les groupes de Lie (simplement connexes) nilpotents stratifiés \mathfrak{N} de dimension n ,⁶² dont l'algèbre de Lie \mathcal{N} admet une décomposition $\mathcal{N} = \bigoplus_{i=1}^{i=m} \mathcal{V}_i$ – ce que l'on appelle une *graduation* – en somme directe de sous-espaces \mathcal{V}_i de dimension $n_i = \dim(\mathcal{V}_i)$ tels que $\mathcal{V}_{i+1} = [\mathcal{V}_1, \mathcal{V}_i]$. La nilpotence signifie que cette suite s'arrête à l'ordre m , c'est-à-dire $[\mathcal{V}_1, \mathcal{V}_m] = 0$. Par exemple, le groupe de Heisenberg \mathbb{H} de distribution \mathcal{H} a pour $\mathcal{V}_1 \subset \mathcal{H}$ le plan de générateurs (X_1, X_2) (qui par exponentiation engendrent la base de coordonnées $z = (x_1, x_2)$) et pour \mathcal{V}_2 l'axe de générateur $T = -\frac{1}{4}[X_1, X_2]$ (qui par exponentiation engendre le centre Z de coordonnée t). On a bien $[\mathcal{V}_1, \mathcal{V}_2] = 0$. Le nombre m est l'ordre ou le niveau (“step” ou degré de non-holonomie) de \mathcal{N} . La formule, dite formule de Mitchell,

$$\begin{aligned} D_h &= \sum_{i=1}^{i=m} i \dim(\mathcal{V}_i) = \sum_{i=1}^{i=m} i n_i \\ &\neq n = \sum_{i=1}^{i=m} \dim(\mathcal{V}_i) = \sum_{i=1}^{i=m} n_i \end{aligned}$$

définit ce que l'on appelle la *dimension homogène* de \mathfrak{N} (elle n'est pas la dimension vectorielle n de \mathcal{N} égale à $\sum_{i=1}^{i=m} \dim(\mathcal{V}_i)$). Par exemple \mathbb{H} est de niveau 2 et de dimension homogène $D_h = 1 \times 2 + 2 \times 1 = 4$ alors que sa dimension comme variété est 3. Dans la formule donnant D_h , ce qui est essentiel est le facteur i devant $\dim(\mathcal{V}_i)$ qui montre que les composantes \mathcal{V}_i sont en quelque sorte de “*poinds*” i . Ce sont ces poids, ces “degrés d'homogénéité”, qui expriment l'extrême anisotropie de ces groupes de Lie nilpotents \mathfrak{N} .

Dans le cas du groupe de Heisenberg polarisé, on a $n = 3$, $m = 2$, $\dim(\mathcal{V}_1) = 2$ et $\dim(\mathcal{V}_2) = 1$, la géométrie sous-riemannienne de \mathfrak{N} consistant à prendre comme champ de plans de contact \mathcal{K} dans $\mathfrak{N} = \exp(\mathcal{N})$ les translatés à gauche de \mathcal{V}_1 muni d'un produit scalaire. Les variables x, p sont de poids 1, et la variable y de poids 2, pdx est de poids $1 + 1 = 2$, dy est de poids 2, et la 1-forme de contact $\omega = dy - pdx$ est donc homogène de poids 2. Les intégrales de \mathcal{K} qui sont géodésiques définissent alors une distance d .

62. Nous utilisons la notation gothique \mathfrak{N} pour “nilpotent”.

La structure d'homogénéité essentielle des groupes de Carnot, qui est l'équivalent de l'invariance par le groupe affine pour les espaces euclidiens, est donnée par les dilatations $\bar{\delta}$. On définit celles-ci d'abord pour l'algèbre de Lie $\mathcal{N} = \bigoplus_{i=1}^{i=m} \mathcal{V}_i$. Soit $\mathcal{X}_i = (X_{i,1}, \dots, X_{i,n_i})$ une base de \mathcal{V}_i et $X = (\xi_i) = \sum_{i=1}^{i=m} \xi_i \mathcal{X}_i$ un vecteur X de \mathcal{N} de composantes $\xi_i = (\xi_{i,1}, \dots, \xi_{i,n_i})$ dans la base (\mathcal{X}_i) .⁶³ Étant donnée une échelle δ , on associe à X le vecteur $\bar{\delta}(X) = (\delta^i \xi_i)$. Si maintenant $x \in \mathfrak{N}$, on considère ses coordonnées canoniques (dites aussi "exponentielles") définies par $x = \exp\left(\sum_{i=1}^{i=m} \xi_i \mathcal{X}_i\right) = \exp(X)$ et on pose $\bar{\delta}(x) = \exp(\bar{\delta}(X))$. On a évidemment $\bar{\delta}' \circ \bar{\delta} = \bar{\delta}' \bar{\delta}$. Pour le groupe de Heisenberg \mathbb{H} non polarisé ci-dessus on a donc les dilatations $\bar{\delta}(z, t) = (\delta z, \delta^2 t)$. En fait, l'existence de telles dilatations est caractéristique des groupes de Carnot. Comme le note Mikhaïl Gromov ([228], p. 92) leur conséquence fondamentale est que

"all geometry of V [i.e. \mathfrak{N}] can be seen in an arbitrary small neighborhood of v_0 [i.e. 0]."

Autrement dit, ces espaces sont à la fois *anisotropes* et *auto-similaires*, ce qui explique leur géométrie très spéciale.

Gromov et Pansu ont profondément étudié les "path metric cones" que sont les groupes de Lie (\mathfrak{N}, d) munis de leur distance sous-riemannienne d et ils ont démontré des résultats métriques assez généraux à leur propos. Par exemple, la boule $B(0, r)$ de centre 0 (l'origine de \mathfrak{N}) et de rayon r est encadrée par des boîtes $\text{Box}(r)$ dont les côtés successifs sont égaux à r^i : il existe deux constantes c et C telles que pour tout $r > 0$ on ait le "ball-box theorem" :

Théorème BB ("ball-box").

$$\text{Box}(cr) \subset B(0, r) \subset \text{Box}(Cr) \quad \diamond$$

C'est pour cette raison que la dimension de Hausdorff de \mathfrak{N} est donnée par la formule de Mitchell $D_h = \sum_1^m i \dim(\mathcal{V}_i)$. En effet, le volume euclidien de $B(0, r)$ est de l'ordre de r^{D_h} .

Comme le dit Mikhaïl Gromov, il est essentiel de comprendre "from within" la structure de ces espaces sous-riemanniens. En effet ([228], p. 88)

"If we live inside a Carnot-Carathéodory metric space V we may know nothing whatsoever about the (external) infinitesimal structures (i.e. the smooth structure on V , the subbundle $H \subset T(V)$ and the metric g on H) which were involved in the construction of the Carnot-Carathéodory metric."⁶⁴

C'est pourquoi il est nécessaire

63. Pour \mathbb{H} , le (X_1, X_2) ci-dessus est donc plus précisément $\mathcal{X}_1 = (X_{1,1}, X_{1,2})$.

64. Avec nos notations dans cette section $V = \mathfrak{N}$ et $H = \mathcal{K}$.

“to develop a sufficiently rich and robust internal Carnot-Carathéodory language which would enable us to capture the essential external characteristics of our Carnot-Carathéodory spaces.” (ibid., p. 89)

Remarque. Le théorème BB permet de démontrer facilement le théorème de Chow. Si x et y sont deux points et si γ est un chemin C^∞ les connectant, on peut, à cause de sa compacité, recouvrir γ par un nombre fini de boules $B(x_j, r_j)$, $j = 1, \dots, m$, $x_j \in \gamma$, $x_1 = x$, $x_m = y$, $B(x_j, r_j) \cap B(x_{j+1}, r_{j+1}) \neq \emptyset$. En prenant des z_j dans les $B(x_j, r_j) \cap B(x_{j+1}, r_{j+1})$ et des boîtes appropriées contenues dans les boules on peut construire des chemins des x_j aux z_j et des z_j aux x_{j+1} qui sont des courbes intégrales de $\mathcal{V}_1 = \mathcal{K}$. □

Pierre Pansu [397] a réussi à définir la *différentiabilité* pour les groupes de Carnot. On définit d’abord les applications dites *linéaires* $F : \mathfrak{N} \rightarrow \mathfrak{N}$ qui sont les morphismes de groupe commutant avec les dilatations $\bar{\delta}$. Une application $f : \mathfrak{N} \rightarrow \mathfrak{N}$ sera alors dite Pansu-différentiable si elle admet une F linéaire “tangente”, c’est-à-dire s’il existe $F = Df$ linéaire telle que

$$\text{Sup} \{d(F_\delta(y), Df(x)y) \mid y \in B(0, 1)\} \xrightarrow{\delta \rightarrow 0} 0$$

où $F_\delta(y) = \bar{\delta}^{-1} (f(x)^{-1} f(x\bar{\delta}(y)))$, ce qui est une version multiplicative de la formule classique

$$\frac{f(x + tX) - f(x)}{t} \xrightarrow{t \rightarrow 0} Df(x).$$

On généralise aux applications $f : \mathfrak{N} \rightarrow \mathfrak{P}$ entre groupes de Carnot différents.

Pour le groupe de Heisenberg \mathbb{H} , cela donne les résultats suivants (cf. Buliga [83]). Soit $\Gamma : [0, 1] \rightarrow \mathbb{H}$ une courbe différentiable dans \mathbb{H} paramétrée par $s : \Gamma(s) = (z(s), t(s))$. Elle sera différentiable si

$$F_\delta(s)(u) = \bar{\delta}^{-1} (\Gamma(s)^{-1} \Gamma(s + \delta u))$$

converge pour $\delta \rightarrow 0$. Mais ⁶⁵

$$F_\delta(s)(u) = \left(\frac{z(s + \delta u) - z(s)}{\delta}, \frac{t(s + \delta u) - t(s)}{\delta^2} - 2\Omega \left(z(s), \frac{z(s + \delta u) - z(s)}{\delta^2} \right) \right)$$

et il n’y a convergence que si $\dot{t}(s) = 2\Omega(z(s), \dot{z}(s))$. Mais cette condition signifie que Γ est horizontale (qu’elle est une intégrale de \mathcal{K}). Il n’y a donc que les courbes horizontales Γ qui sont différentiables au sens de Pansu et leur dérivée est

$$D\Gamma(s)u = u(\dot{z}(s), 0) .$$

65. Rappelons que $\Omega(z, z')$ est la 2-forme symplectique standard sur le plan des z .

Considérons maintenant une fonction différentiable $f : \mathbb{H} \rightarrow \mathbb{R}$ (c'est ce à quoi nous avons à faire lorsque nous considérons un état d'activité "différentiable" de l'aire $V1$). Son application linéaire tangente, si elle existe, sera la limite pour $\delta \rightarrow 0$ de

$$F_\delta(z, t)(u, r) = \frac{f(z + \delta u, t + \delta^2 r + 2\delta\Omega(z, u)) - f(z, t)}{\delta}.$$

Comme f est différentiable, elle est Pansu-différentiable et son application linéaire tangente est donnée par :

$$Df(z, t)(u, r) = u \frac{\partial f}{\partial z}(z, t) + 2\Omega(z, u) \frac{\partial f}{\partial t}(z, t).$$

Notons que, réciproquement, l'existence de dilatations implique trivialement la nilpotence (cf. [524]). Soit en effet \mathcal{G} une algèbre de Lie quelconque munie d'un groupe à 1-paramètre d'automorphismes $A(t)$. On peut alors considérer la dérivée en $t = 0$,

$$D(X) = \left. \frac{d}{dt} A(t) X \right|_{t=0}.$$

$D(X)$ est une dérivation de \mathcal{G} car

$$A(t)[X, Y] = [A(t)X, A(t)Y]$$

(puisque les $A(t)$ sont des automorphismes d'algèbre) implique (puisque $A(0) = Id$)

$$\begin{aligned} D([X, Y]) &= \left. \frac{d}{dt} A(t)[X, Y] \right|_{t=0} = \left. \frac{d}{dt} A(t)XA(t)Y - A(t)YA(t)X \right|_{t=0} \\ &= \left. \frac{dA(t)X}{dt} \right|_{t=0} A(0)Y + A(0)X \left. \frac{dA(t)Y}{dt} \right|_{t=0} \\ &\quad - \left. \frac{dA(t)Y}{dt} \right|_{t=0} A(0)X - A(0)Y \left. \frac{dA(t)X}{dt} \right|_{t=0} \\ &= X \left. \frac{dA(t)Y}{dt} \right|_{t=0} - \left. \frac{dA(t)Y}{dt} \right|_{t=0} X + \left. \frac{dA(t)X}{dt} \right|_{t=0} Y - Y \left. \frac{dA(t)X}{dt} \right|_{t=0} \\ &= [X, D(Y)] + [D(X), Y]. \end{aligned}$$

On considère alors le spectre $\text{Spec}(D)$ de D . Si α et β sont deux valeurs propres d'espaces propres \mathcal{G}_α et \mathcal{G}_β et si $X \in \mathcal{G}_\alpha$ et $Y \in \mathcal{G}_\beta$, alors

$$D([X, Y]) = [X, \beta Y] + [\alpha X, Y] = (\alpha + \beta)[X, Y]$$

et donc

$$[\mathcal{G}_\alpha, \mathcal{G}_\beta] \subset \mathcal{G}_{\alpha+\beta}. X \in \mathcal{G}_\alpha$$

si et seulement si $A(t)X$ est de la forme

$$A(t)X = e^{\alpha t} P_\alpha(t)$$

où $P_\alpha(t)$ est un polynôme

$$P_\alpha(t) = X + \sum_{j=2}^{j=N} t^j U_j .$$

Le groupe à 1-paramètre $A(t)$ est un groupe de *dilatations* si $\text{Spec}(D) \subset \mathbb{R}^+$. Cela implique bien que \mathcal{G} soit nilpotente car, plus on commute, plus $\alpha + \beta$ augmente et, comme le spectre est fini et donc borné, on arrive nécessairement à un $\mathcal{G}_{\alpha+\beta} = \{0\}$ et donc à $[\mathcal{G}_\alpha, \mathcal{G}_\beta] = \{0\}$, ce qui montre la nilpotence.

8.2. Groupes de Carnot et champ d'association

Intuitivement, le théorème BB signifie que les boules infinitésimales $B(0, \varepsilon)$ sont écrasées sur le plan de contact $K_0 \simeq \mathcal{V}_1$ (où elles sont d'ordre ε) et d'ordre ε^i dans les autres directions de degré d'homogénéité $i > 1$, dans notre cas d'ordre ε^2 dans la direction transverse à K_0 . Cela est dû au fait que, même si tout point est joignable à tout point par une courbe intégrale de \mathcal{K} d'après le théorème de Chow, ces courbes sont tangentes à \mathcal{K} et ne peuvent aller dans les autres directions qu'à travers des crochets de Lie, les directions de \mathcal{V}_i exigeant l'utilisation de $i - 1$ crochets. Mais cela correspond exactement à la condition du champ d'association implémentée dans les connexions horizontales, condition que nous avons modélisée au chapitre 5, section 5.2. du Vol I. On peut donc dire que la géométrie sous-riemannienne du groupe $\mathbb{V} = \mathbb{V}_J$ ou $\mathbb{V} = \mathbb{V}_S$ implémentée dans l'aire visuelle $V1$ est tout simplement la géométrie du champ d'association. C'est pourquoi son approche à la Gromov "from within" est centrale pour la neurogéométrie de la vision.

En fait, pour être plus précis, il faudrait introduire des connexions horizontales "fortes" et des connexions "faibles" dans les directions transverses. Cela revient, nous l'avons vu, à considérer la métrique sous-riemannienne $g_{\mathcal{K}}$ sur \mathcal{K} comme une limite de métriques riemanniennes g_ε qui tendent vers $g_{\mathcal{K}}$ lorsque $\varepsilon \rightarrow 0$ (cf. chapitre 5, section 7).

8.3. Variétés de Heisenberg, nilpotentisation et cônes tangents

Le cas du groupe $G = SE(2) = \mathbb{V}_S$ des déplacements du plan est très différent car il n'est pas nilpotent. Rappelons encore une fois que le champ des plans de contact \mathcal{K} est engendré par translations à gauche à partir de

$$\mathcal{V}_1 = K_0 = \text{Span} \{X_1, X_2\}$$

avec

$$\left\{ \begin{array}{l} X_1 = \cos(\theta) \partial_x + \sin(\theta) \partial_y \\ X_2 = \partial_\theta \\ X_3 = \sin(\theta) \partial_x - \cos(\theta) \partial_y \\ [X_1, X_2] = X_3 \\ [X_2, X_3] = X_1 \\ [X_1, X_3] = 0 \\ \omega = -\sin(\theta) dx + \cos(\theta) dy \end{array} \right.$$

Le groupe n'est plus nilpotent et on ne peut plus définir des degrés d'homogénéité pour les différentes coordonnées. En effet, on a bien

$$\mathcal{V}_2 = [\mathcal{V}_1, \mathcal{V}_1] = \text{Span} \{X_3\} ,$$

mais

$$\left\{ \begin{array}{l} \mathcal{V}_3 = [\mathcal{V}_1, \mathcal{V}_2] = \text{Span} \{X_1\} \neq 0 \\ \mathcal{V}_4 = [\mathcal{V}_1, \mathcal{V}_3] = \mathcal{V}_1 , \text{ etc.} \end{array} \right.$$

Cela est dû au fait que les coefficients de X_1 et X_3 n'étant pas polynomiaux mais trigonométriques, leurs dérivées successives ne s'annulent jamais.

Mais on peut néanmoins construire un groupe de Carnot "tangent" à G en 0 au moyen de ce que l'on appelle la procédure de *nilpotentisation* (cf. par exemple André Bellaïche [36] et Richard Montgomery [375]). Posons

$$\left\{ \begin{array}{l} \mathcal{W}_0 = \{0\} \\ \mathcal{W}_1 = \mathcal{V}_1 \\ \mathcal{W}_{i+1} = \mathcal{W}_i + [\mathcal{V}_1, \mathcal{W}_i] \end{array} \right.$$

Cette suite croissante de sous-espaces est une *filtration* de l'algèbre de Lie \mathcal{G} de G . Elle satisfait la propriété de compatibilité $[\xi, \eta] \in \mathcal{W}_{i+j}$ pour tous les $\xi \in \mathcal{W}_i$ et $\eta \in \mathcal{W}_j$.⁶⁶ À partir de \mathcal{W}_2 on a $\mathcal{W}_i = \mathcal{G}$ et la filtration devient stationnaire. On construit alors à partir de \mathcal{G} muni de cette filtration une algèbre de Lie *nilpotente* \mathcal{N} (de même espace vectoriel sous-jacent que \mathcal{G}) munie d'une *graduation* de la façon suivante. On pose $\mathcal{N} = \bigoplus \mathcal{N}_i$ avec $\mathcal{N}_i = \mathcal{W}_i / \mathcal{W}_{i-1}$ et on définit par quotientage le crochet de Lie $[\xi, \eta]_{\mathcal{N}}$ dans \mathcal{N} comme le crochet de \mathcal{G} mais avec la condition que si $\xi \in \mathcal{N}_i$, ξ est seulement considéré *modulo* \mathcal{W}_{i-1} . On montre que $\mathfrak{N} = \exp(\mathcal{N})$ est un groupe de Carnot. C'est le groupe de Heisenberg \mathbb{H} .

On peut aussi obtenir la nilpotentisation en utilisant les dilatations caractéristiques des groupes de Carnot. Comme on a $\mathcal{G} = \mathcal{N} = \bigoplus \mathcal{N}_i$ en tant qu'espace vectoriel, on peut considérer des "poids" (les degrés d'homogénéité) pour les $\xi_i \in \mathcal{N}_i$. Si $\xi \in \mathcal{G}$ s'écrit $\xi = \sum_i \xi_i \mathcal{X}_i$, on définit, nous l'avons vu, sa dilatation par $\bar{\delta} = \sum_i \delta^i \xi_i$ et on pose $[\xi, \eta]_{\mathcal{N}} = \lim_{\delta \rightarrow 0} \bar{\delta}^{-1}([\bar{\delta}(\xi), \bar{\delta}(\eta)])$.

66. Dans ce qui suit, pour simplifier les notations, nous notons ξ, η des vecteurs de \mathcal{G} . Si $\xi = (\xi_i)$, ξ est donc une abréviation pour le vecteur $X = \sum_{i=1}^{i=m} \xi_i \mathcal{X}_i$.

Dans notre cas, on obtient

$$\begin{cases} \mathcal{N}_2 = \mathcal{W}_2/\mathcal{W}_1 = \mathcal{G}/\mathcal{V}_1 \\ \mathcal{N}_3 = \mathcal{W}_3/\mathcal{W}_2 = \mathcal{G}/\mathcal{G} = \{0\} \end{cases}$$

ce qui assure la nilpotence. On a, comme dans \mathcal{G} ,

$$\begin{cases} [X_1, X_2]_{\mathcal{N}} = X_3 \\ [X_1, X_3]_{\mathcal{N}} = 0 \end{cases}$$

mais en revanche

$$[X_2, X_3]_{\mathcal{N}} = X_1 \bmod (\mathcal{W}_3) = 0$$

alors que

$$[X_2, X_3] = X_1 \neq 0$$

dans \mathcal{G} . Cela se comprend très facilement avec les dilatations :

$$\begin{aligned} [X_1, X_2]_{\mathcal{N}} &= \lim_{\delta \rightarrow 0} \bar{\delta}^{-1}([\delta X_1, \delta X_2]) = \lim_{\delta \rightarrow 0} \bar{\delta}^{-1}(\delta^2 [X_1, X_2]) \\ &= \lim_{\delta \rightarrow 0} \frac{\delta^2}{\delta^2} (X_3) = X_3 \end{aligned}$$

car X_3 est de poids 2. En revanche

$$\begin{aligned} [X_2, X_3]_{\mathcal{N}} &= \lim_{\delta \rightarrow 0} \bar{\delta}^{-1}([\delta X_2, \delta^2 X_3]) = \lim_{\delta \rightarrow 0} \bar{\delta}^{-1}(\delta^3 X_1) \\ &= \lim_{\delta \rightarrow 0} \frac{\delta^3}{\delta} (X_1) = 0. \end{aligned}$$

Autrement dit, le commutateur $[X_2, X_3]$ de \mathcal{G} est annulé, ce qui assure la nilpotentisation. On obtient donc bien ainsi le groupe de Heisenberg $\mathfrak{N} = \exp(\mathcal{N}) = \mathbb{H}$.

La partie la plus technique de ce genre d'approche consiste à montrer que \mathfrak{N} est bien un cône métrique tangent à l'espace métrique G en 0. Pour cela il faut utiliser les profonds travaux de Mikhaïl Gromov (cf. par exemple [228] et [229]) sur les espaces métriques et les suites convergentes d'espaces métriques. La difficulté est que les variétés sous-riemanniennes ne sont euclidiennes à *aucune* échelle et ne ressemblent donc pas du tout à des variétés riemanniennes. Leurs "espaces métriques tangents" ne sont pas du tout des hyperplans euclidiens mais des groupes de Carnot que l'on peut considérer selon André Bellaïche comme des sortes d'espaces vectoriels "non abéliens".

Ce que nous venons de voir pour G est un cas très particulier du *théorème de Mitchell* [370] disant que pour toute variété sous-riemannienne M de distribution \mathcal{K} , le cône tangent existe en tout point $x \in M$ et est le groupe de Carnot \mathfrak{N}_x obtenu

par nilpotentisation.⁶⁷ Margulis et Mostow [351] ont montré que l'on peut recoller le \mathfrak{N}_x en un "fibré tangent" de cônes tangents au-dessus de M .

Pour simplifier, supposons que la distribution \mathcal{K} soit complètement non holonome, autrement dit que le drapeau de Lie des

$$\mathcal{K}^j = \mathcal{K}^{j-1} + [\mathcal{K}, \mathcal{K}^{j-1}],$$

qui, notons-le, satisfait

$$[\mathcal{K}^j, \mathcal{K}^s] \subset \mathcal{K}^{j+s},$$

aille de $\mathcal{K} = \mathcal{K}^1$ jusqu'à $TM = \mathcal{K}^r$ pour un certain r . Le r minimal est, répétons-le, le "degré de non-holonomie" ou le "niveau" ("step") de \mathcal{K} . Nous le verrons plus en détail à la section 1 du chapitre 15.⁶⁸ Supposons également que \mathcal{K} et son drapeau soient *réguliers*, c'est-à-dire que tous les K_x soient de dimension constante. La suite des $n_j = \dim(\mathcal{K}^j)$ (avec $n_1 = k$ et $n_r = n$) s'appelle, nous le verrons également en détail à la section 1 du chapitre 15, le "vecteur de croissance" ("growth vector") de \mathcal{K} . Par exemple le groupe de Heisenberg est de vecteur de croissance $(2, 3)$.⁶⁹ Pour construire la nilpotentisation \mathfrak{N}_x on effectue la même construction que pour les groupes de Lie. On part du drapeau

$$\mathcal{K} = \mathcal{K}^1 \subset \mathcal{K}^2 \subset \dots \subset \mathcal{K}^r = TM$$

et on construit d'abord les quotients $\mathcal{V}_j = \mathcal{K}^j / \mathcal{K}^{j-1}$ puis ensuite le gradué associé qui en est la somme directe :

$$gr(\mathcal{K}) = \bigoplus_{j=1}^{j=r} \mathcal{N}_j.$$

Le crochet induit un crochet sur $gr(\mathcal{K})$ satisfaisant $[\mathcal{N}_j, \mathcal{N}_s] \subset \mathcal{N}_{j+s}$. En effet, soient $\tilde{X}_j \in \mathcal{N}_j$ de représentant $X_j \in \mathcal{K}^j$ et $\tilde{Y}_s \in \mathcal{N}_s$ de représentant $Y_s \in \mathcal{K}^s$. Comme

67. La problématique de la nilpotentisation a été introduite par Noboru Tanaka en 1970 dans [523]. \mathfrak{N}_x y était associée à la "symbol algebra" de \mathcal{K} en x .

68. Dans cette section 1 les notations seront plus générales car elles concerneront toutes les variétés sous-riemanniennes (M, \mathfrak{D}) . La filtration des sous-algèbres de Lie de $T_x M$ (le drapeau de Lie)

$$\mathfrak{D}_x^0 = \mathfrak{D}_x, \mathfrak{D}_x^j = \mathfrak{D}_x^{j-1} + [\mathfrak{D}_x, \mathfrak{D}_x^{j-1}]$$

engendre en chaque point x la graduation (nilpotentisation)

$$gr_x(\mathfrak{D}) = \mathfrak{D}_x \oplus \mathfrak{D}_x^1 / \mathfrak{D}_x \oplus \dots \oplus \mathfrak{D}_x^r / \mathfrak{D}_x^{r-1}.$$

Ici $\mathcal{K}^{j+1} = \mathfrak{D}^j$.

69. Lorsque les $n_j(x)$ ne sont pas localement constants, il s'introduit alors des points *singuliers* de la distribution \mathcal{K} . L'exemple sans doute le plus connu est la distribution de Martinet sur \mathbb{R}^3 où \mathcal{K} est le noyau de la 1-forme $\omega = dy - p^2 dx$ (et pas $\omega = dy - p dx$ comme pour Heisenberg). Les plans \mathcal{K}_x sont engendrés par les vecteurs $X = \partial_x + p^2 \partial_y$ et $Y = \partial_p$. On a $[X, Y] = -2p \partial_y$ et donc si $p = 0$, $[X, Y] = 0$ et il faut un autre crochet pour engendrer \mathbb{R}^3 , en l'occurrence $[[X, Y], Y] = 2 \partial_y$. En dehors du plan $p = 0$ le vecteur de croissance est $(2, 3)$ mais sur le plan $p = 0$ il est $(2, 2, 3)$.

$[\mathcal{K}^j, \mathcal{K}^s] \subset \mathcal{K}^{j+s}$, on a $[X_j, Y_s] \in \mathcal{K}^{j+s}$. Pour d'autres représentants $X_j + \xi_j$ et $Y_s + \eta_s$, il faut $\xi_j \in \mathcal{K}^{j-1}$ et $\eta_s \in \mathcal{K}^{s-1}$. Mais dans ce cas,

$$\begin{cases} [X_j, \eta_s] \in \mathcal{K}^{j+s-1} \\ [Y_s, \xi_j] \in \mathcal{K}^{s+j-1} \\ [\xi_j, \eta_s] \in \mathcal{K}^{j+s-2} \end{cases}$$

et par conséquent $[\widetilde{X}_j, \widetilde{Y}_s]$ est bien défini modulo \mathcal{K}^{j+s-1} et appartient à \mathcal{N}_{j+s} . Le gradué $gr_x(\mathcal{K})$ est alors la nilpotentisation \mathfrak{N}_x .

Dans le cas de niveau 2, $\mathcal{K}^2 = TM = \mathcal{K} + [\mathcal{K}, \mathcal{K}]$ et on peut considérer la 2-forme

$$\mathcal{L} : \mathcal{K} \times \mathcal{K} \rightarrow TM // \mathcal{K}$$

définie par le crochet $[\mathcal{K}, \mathcal{K}]$. Si

$$\xi = X + X', \eta = Y + Y', X, Y \in \mathcal{K}, X', Y' \in [\mathcal{K}, \mathcal{K}] ,$$

le produit dans le fibré de groupes de Carnot \mathfrak{N} est, en vertu de la formule de Campbell-Hausdorff,

$$\xi\eta = X + Y + \frac{1}{2}\mathcal{L}(X, Y) + X' + Y' .$$

Le quotient $TM // \mathcal{K} = [\mathcal{K}, \mathcal{K}]$ est contenu dans son centre puisque si $\xi = 0 + X' \in [\mathcal{K}, \mathcal{K}]$, $\mathcal{L}(0, Y) = 0$ pour tout Y et ξ commute avec tous les η .

Géométrie des groupes et algèbres de Lie

Après ce chapitre consacré à la géométrie des connexions et avant que d'en venir au thème des modèles géodésiques de contours illusoires, nous allons un peu approfondir la géométrie des groupes et algèbres de Lie réels que nous avons esquissée dans le chapitre 5.¹

1. Géométrie de la nilpotence

Nous avons introduit à la section 4 du chapitre 5 l'idéal dérivé d'une algèbre de Lie $\mathcal{G} : \mathcal{D}(\mathcal{G}) = [\mathcal{G}, \mathcal{G}]$ ainsi que la suite décroissante, dite *suite centrale descendante*, d'idéaux $\mathcal{G}_{(j)}$ définis par

$$\mathcal{G}_{(0)} = \mathcal{G}, \mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}_{(0)}] = \mathcal{D}(\mathcal{G}), \mathcal{G}_{(2)} = [\mathcal{G}, \mathcal{G}_{(1)}], \dots$$

Les algèbres nilpotentes sont celles dont la suite $\mathcal{G}_{(j)}$ stationne en $\{0\}$ à partir d'un certain "niveau" fini et donc où, par formation de commutateurs, on arrive toujours à la commutativité. Les plus simples sont, comme l'algèbre de Heisenberg, celles de niveau 2 telles que

$$\mathcal{G}_{(0)} = \mathcal{G}, \mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}_{(0)}] \neq 0, \mathcal{G}_{(2)} = [\mathcal{G}, \mathcal{G}_{(1)}] = 0.$$

Pour Heisenberg polarisé, on a

$$\mathcal{G} = \text{Span} \{ \mathbf{t}_1 = \partial_x + p\partial_y, \mathbf{t}_2 = \partial_p, \mathbf{t}_3 = -\partial_y \}$$

satisfaisant $[\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3, [\mathbf{t}_1, \mathbf{t}_3] = [\mathbf{t}_2, \mathbf{t}_3] = 0$. Donc

$$\mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}_{(0)}] = \text{Span} \{ \mathbf{t}_3 \}, \mathcal{G}_{(2)} = 0.$$

Pour Engel, on a

$$\mathcal{G} = \text{Span} \{ \mathbf{t}_1 = \partial_x + p\partial_y + \kappa\partial_p, \mathbf{t}_2 = \partial_p, \mathbf{t}_3 = -\partial_y, \mathbf{t}_4 = \partial_\kappa \}$$

satisfaisant en plus

$$[\mathbf{t}_1, \mathbf{t}_4] = -\mathbf{t}_2, [\mathbf{t}_2, \mathbf{t}_4] = [\mathbf{t}_3, \mathbf{t}_4] = 0$$

et donc

$$[[\mathbf{t}_1, \mathbf{t}_4], \mathbf{t}_1] = \mathbf{t}_3.$$

1. Par défaut, et donc sauf mention expresse du contraire, nous ne considérons ici que des groupes et algèbres sur \mathbb{R} . Parfois il convient de se placer sur \mathbb{C} qui est algébriquement clos. La théorie sur d'autres corps de base \mathbb{K} est plus compliquée.

Par conséquent,

$$\mathcal{G}_{(1)} = \text{Span} \{t_3, t_2\}, \quad \mathcal{G}_{(2)} = \text{Span} \{t_3\}, \quad \mathcal{G}_{(3)} = 0.$$

En utilisant le fait que, si \mathfrak{a} est un idéal de \mathcal{G} , le quotient \mathcal{G}/\mathfrak{a} est commutatif si et seulement si $[\mathcal{G}, \mathcal{G}] \subseteq \mathfrak{a}$ et que, si $\mathfrak{b} \supseteq \mathfrak{a}$ est un autre idéal, $\mathfrak{b}/\mathfrak{a}$ est *central* dans \mathcal{G}/\mathfrak{a} (i.e. inclus dans le centre de \mathcal{G}/\mathfrak{a}) si et seulement si $[\mathcal{G}, \mathfrak{b}] \subseteq \mathfrak{a}$, on voit que, puisque $[\mathcal{G}, \mathcal{G}_{(j)}] = \mathcal{G}_{(j+1)}$, le quotient $\mathcal{G}_{(j)}/\mathcal{G}_{(j+1)}$ est central dans $\mathcal{G}/\mathcal{G}_{(j+1)}$ pour tout j . On voit également que, pour tout i, j , $[\mathcal{G}_{(i)}, \mathcal{G}_{(j)}] \subset \mathcal{G}_{(i+j)}$.

La classe des algèbres nilpotentes est stable par sous-algèbres, algèbres quotients et extensions centrales, i.e. les extensions (suites exactes)

$$0 \rightarrow \mathcal{K} \rightarrow \mathcal{G} \rightarrow \mathcal{H} \rightarrow 0$$

telles que \mathcal{K} et \mathcal{H} soient nilpotentes et que \mathcal{K} soit incluse dans le centre de \mathcal{G} . C'est même la plus petite classe possédant ces propriétés qui contient les algèbres abéliennes. C'est en ce sens qu'elle est la généralisation directe des algèbres abéliennes.

Nous avons évoqué (section 4 du chapitre 5) à propos des algèbres nilpotentes le

Théorème de structure. Si \mathcal{G} est nilpotente, il existe une suite *croissante* d'idéaux \mathcal{G}_j (ce ne sont pas les $\mathcal{G}_{(j)}$) de dimension $j = 0, 1, \dots, n = \dim \mathcal{G}$ tels que $[\mathcal{G}, \mathcal{G}_{j+1}] \subset \mathcal{G}_j$. ◇

En particulier $[\mathcal{G}, \mathcal{G}_1] = \mathcal{G}_0 = \{0\}$ et donc \mathcal{G} et \mathcal{G}_1 commutent et \mathcal{G}_1 est inclus dans le centre Z de \mathcal{G} . Celui-ci n'est donc pas nul car il est de dimension $\geq \dim \mathcal{G}_1 = 1$. À l'autre extrême, $\mathcal{G}_n = \mathcal{G}$ et donc $[\mathcal{G}, \mathcal{G}] \subset \mathcal{G}_{n-1}$.

Nous allons maintenant préciser un peu cette structure "en drapeau", un drapeau² dans un espace vectoriel V de dimension n étant justement une suite de sous espaces emboîtés

$$V_0 = \{0\} \subset V_1 \subset \dots \subset V_n = V$$

de dimensions successives $j = 0, 1, \dots, n$. Pour démontrer qu'un tel drapeau existe nécessairement pour \mathcal{G} , on utilise le lemme sur la représentation adjointe ad_X disant que, si l'endomorphisme ad_X de \mathcal{G} est nilpotent pour tout $X \in \mathcal{G}$, alors il existe un $U \in \mathcal{G}$, $U \neq 0$, tel que $ad_X(U) = 0$ pour tout $X \in \mathcal{G}$. Autrement dit l'espace intersection $\bigcap_{X \in \mathcal{G}} \ker(ad_X)$ des noyaux des ad_X n'est pas réduite à $\{0\}$. Cela implique

l'existence d'un drapeau \mathcal{G}_j tel que $ad_X(\mathcal{G}_j) \subset \mathcal{G}_{j-1}$ pour tout $X \in \mathcal{G}$, i.e. $[\mathcal{G}, \mathcal{G}_j] \subset \mathcal{G}_{j-1}$. On le démontre par récurrence. Le cas de la dimension 0 étant trivial, on suppose qu'il existe un $U \neq 0$ dans \mathcal{G} tel que $ad_X(U) = 0$ pour tout $X \in \mathcal{G}$. On prend $\mathcal{G}_1 = \mathbb{R}U$ et on considère le quotient $\mathcal{G}' = \mathcal{G}/\mathcal{G}_1$. Par hypothèse de récurrence,

2. L'image du drapeau vient du fait qu'un drapeau est constitué de la hampe qui est une droite de dimension 1 à laquelle est attaché le tissu du drapeau qui est un plan de dimension 2.

il existe un drapeau \mathcal{G}'_k dans \mathcal{G}' et on construit un drapeau dans \mathcal{G} en prenant $\mathcal{G}_j = \mathcal{G}_1 + \mathcal{G}'_j$. On démontre ainsi le célèbre

Théorème d'Engel. Une algèbre de Lie \mathcal{G} est nilpotente si et seulement si ad_X est nilpotent pour tout $X \in \mathcal{G}$. \diamond

Notons que comme la dimension des \mathcal{G}_j est $j = 0, 1, \dots, n = \dim \mathcal{G}$, on peut considérer une base e_k de \mathcal{G} adaptée à ce drapeau, i.e. telle que $e_1 \in \mathcal{G}_1$, $e_2 \in$ supplémentaire de \mathcal{G}_1 dans \mathcal{G}_2 , etc. Dans cette base, les matrices des ad_X sont *triangulaires supérieures strictes* (i.e. tous les termes sur et sous la diagonale sont nuls) car $ad_X(\mathcal{G}_j) \subset \mathcal{G}_{j-1}$. Autrement dit, la représentation adjointe ad envoie \mathcal{G} sur une sous-algèbre de Lie de l'algèbre de Lie Δ_n des matrices triangulaires supérieures strictes sur \mathbb{R}^n .

À partir de ces théorèmes de structure, on montre qu'un groupe de Lie G (connexe et simplement connexe) est nilpotent si et seulement si il est égal à \mathbb{R}^n muni d'une loi de groupe *polynomiale*.

Dans le cas de l'algèbre de Heisenberg (chapitre 5), le drapeau est évident. Pour les éléments $\mathfrak{t} = (\xi, \eta, \pi) \in \mathcal{G}$ on a $[\mathfrak{t}, \mathfrak{t}'] = (0, \pi\xi' - \pi'\xi, 0)$. \mathcal{G}_1 est le centre $\mathbb{R}\eta'$, \mathcal{G}_2 est par exemple $\mathbb{R}\eta' \oplus \mathbb{R}\xi'$ avec $[\mathfrak{t}, \mathfrak{t}'] = (0, \pi\xi', 0) \in \mathcal{G}_1$ et $\mathcal{G}_3 = \mathcal{G}$. Dans la base adaptée les \mathfrak{t} doivent s'écrire (η, ξ, π) et donc la matrice de $ad_{\mathfrak{t}}$ est donnée par

$$\begin{pmatrix} 0 & \pi & -\xi \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta' \\ \xi' \\ \pi' \end{pmatrix} = \begin{pmatrix} \pi\xi' - \pi'\xi \\ 0 \\ 0 \end{pmatrix}$$

qui est bien triangulaire supérieure stricte.³ Par ailleurs la loi de groupe est bien polynomiale.

2. Algèbres résolubles

Une généralisation des algèbres nilpotentes qui garde un lien étroit avec une commutativité partielle est celle des algèbres de Lie *résolubles*. Ce sont celles pour lesquelles il existe un p tel que l'algèbre dérivée $\mathcal{D}^p(\mathcal{G}) = 0$.⁴

On comparera

$$\mathcal{D}^{j+1}(\mathcal{G}) = [\mathcal{D}^j(\mathcal{G}), \mathcal{D}^j(\mathcal{G})]$$

à la suite centrale descendante $\mathcal{G}_{(j+1)} = [\mathcal{G}, \mathcal{G}_{(j)}]$. Comme

$$\mathcal{G}_{(0)} = \mathcal{G}, \mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}_{(0)}] = \mathcal{D}(\mathcal{G}),$$

les deux suites commencent par le même terme et donc, par récurrence, $\mathcal{D}^{j+1}(\mathcal{G}) \subset \mathcal{G}_{(j)}$.

3. C'est bien la même matrice que celle de la section 6.2 du chapitre 5 avec l'ordre spécifique de la base adaptée.

4. Les $\mathcal{D}^j(\mathcal{G})$ sont souvent notées $\mathcal{G}^{(j)}$.

Comme nous l'avons déjà évoqué à propos de la théorie de Galois dans la section 1.5 du chapitre 1, être résoluble est équivalent au fait qu'il existe une suite décroissante de sous-algèbres

$$\mathcal{G} = \mathfrak{A}_1 \supseteq \mathfrak{A}_2 \supseteq \cdots \supseteq \mathfrak{A}_k \supseteq \mathfrak{A}_{k+1} = \{0\}$$

telle que \mathfrak{A}_{j+1} soit un idéal de \mathfrak{A}_j et que les quotients successifs $\mathfrak{A}_j/\mathfrak{A}_{j+1}$ soient *abéliens*, autrement dit, telle que \mathcal{G} s'obtienne par une suite d'extensions (pas forcément centrales)

$$0 \rightarrow \mathfrak{A}_{j+1} \rightarrow \mathfrak{A}_j \rightarrow \mathfrak{A}_j/\mathfrak{A}_{j+1} \rightarrow 0$$

avec $\mathfrak{A}_j/\mathfrak{A}_{j+1}$ abélien.

Par exemple notre algèbre \mathcal{V}_S (chapitre 5, section 7) avec le crochet

$$[\mathfrak{t}, \mathfrak{t}'] = [(\xi, \eta, \tau), (\xi', \eta', \tau')] = (-\tau\eta' + \eta\tau', \tau\xi' - \xi\tau', 0)$$

est résoluble. En effet $\mathfrak{A}_2 = \{(\xi, \eta, 0)\}$ est une sous-algèbre (commutative car $[\mathfrak{t}, \mathfrak{t}'] = 0$ si $\tau = \tau' = 0$) et un idéal de \mathcal{V}_S car $[\mathfrak{t}, \mathfrak{t}'] \in \mathfrak{A}_2$ pour tous les $\mathfrak{t}, \mathfrak{t}'$. Le quotient $\mathcal{V}_S/\mathfrak{A}_2$ est isomorphe à la sous-algèbre commutative $\{(0, 0, \tau)\}$. L'idéal $\mathfrak{A}_3 = \{0\}$ termine la suite puisque \mathfrak{A}_2 est commutative et $\mathfrak{A}_2/\mathfrak{A}_3 = \mathfrak{A}_2$.

La classe des algèbres résolubles est stable par sous-algèbres, algèbres quotients et extensions, i.e. les extensions (suites exactes)

$$0 \rightarrow \mathcal{K} \rightarrow \mathcal{G} \rightarrow \mathcal{H} \rightarrow 0$$

telles que \mathcal{K} et \mathcal{H} soient résolubles.⁵ C'est même la plus petite classe possédant ces propriétés qui contienne les algèbres abéliennes.

Il existe aussi un théorème de structure (dû à Lie) pour les algèbres résolubles. Mais il est plus subtil et exige que le corps de base \mathbb{K} soit *algébriquement clos* ($\mathbb{K} = \mathbb{C}$). La contrainte pour le drapeau V_j n'est plus $ad_X(V_j) \subset V_{j-1}$ mais

$$ad_X(V_j) \subset V_j$$

et donc les matrices des ad_X sont seulement triangulaires supérieures (pas forcément strictes) dans une base adaptée. Les matrices triangulaires supérieures T de diagonale (d_1, \dots, d_n) non nulle ne peuvent pas être nilpotentes car la diagonale de T^k est (d_1^k, \dots, d_n^k) et ne peut donc jamais s'annuler si l'un des d_j est $\neq 0$. En revanche si T et T' sont triangulaires supérieures, la diagonale de $[T, T']$ est

$$(d_1 d'_1 - d'_1 d_1, \dots, d_n d'_n - d'_n d_n) = (0, \dots, 0)$$

et par conséquent $[T, T']$ est nilpotente. On en tire le

Théorème de Lie. Sur \mathbb{C} , \mathcal{G} est résoluble si et seulement si $[\mathcal{G}, \mathcal{G}]$ est nilpotente. Cette propriété est équivalente au fait que \mathcal{G} soit trigonalisable au moyen d'un drapeau. \diamond

5. Soit $0 \rightarrow \mathcal{K} \xrightarrow{j} \mathcal{G} \xrightarrow{p} \mathcal{H} \rightarrow 0$ une extension. Si $\mathcal{D}^r(\mathcal{H}) = 0$ (\mathcal{H} résoluble), $p(\mathcal{D}^r(\mathcal{G})) = 0$. Donc $\mathcal{D}^r(\mathcal{G}) \subset j(\mathcal{K})$. Mais si $\mathcal{D}^s(\mathcal{K}) = 0$ (\mathcal{K} résoluble) alors $\mathcal{D}^{r+s}(\mathcal{G}) = 0$ et \mathcal{G} est résoluble.

La plus simple algèbre de Lie résoluble non nilpotente est celle du groupe *affine* A que nous discuterons à propos des transformées en ondelettes à la section 2.1.3 du chapitre 17. A est le produit semi-direct du groupe des translations spatiales $a \in (\mathbb{R}, +, 0)$ par le groupe multiplicatif des échelles $\sigma \in (\mathbb{R}_+^*, \times, 1)$ avec la loi de composition

$$(a', \sigma') \cdot (a, \sigma) = (a' + \sigma' a, \sigma' \sigma) .$$

L'élément neutre est $(0, 1)$ et l'inverse de $g = (a, \sigma)$ est $g^{-1} = (-\frac{a}{\sigma}, \frac{1}{\sigma})$. Une représentation matricielle commode est

$$g = (a, \sigma) = \begin{pmatrix} \sigma & a \\ 0 & 1 \end{pmatrix} .$$

Son algèbre de Lie \mathcal{A} est l'ensemble des $X = \begin{pmatrix} \varsigma & \alpha \\ 0 & 0 \end{pmatrix}$ (non nilpotents car de matrice triangulaire supérieure non stricte et de trace ς en général non nulle) muni de l'addition et du crochet nilpotent (car de matrice triangulaire supérieure stricte)

$$[X, X'] = \begin{pmatrix} 0 & \varsigma \alpha' - \varsigma' \alpha \\ 0 & 0 \end{pmatrix} .$$

Comme le produit de telles matrices est nul, a fortiori leur crochet est nul et donc

$$\begin{cases} \mathcal{D}(\mathcal{A}) = [\mathcal{A}, \mathcal{A}] = \begin{pmatrix} 0 & \mathbb{R} \\ 0 & 0 \end{pmatrix} \\ \mathcal{D}^2(\mathcal{A}) = 0 \end{cases}$$

3. Structure des algèbres de Lie générales

3.1. Idéaux et forme de Killing

Pour montrer comment la commutativité partielle liée à la résolubilité s'intrique avec la non-commutativité totale qu'est la simplicité, on peut indiquer un théorème du à Eugenio Levi⁶ (cf. Taylor [524]). Il est facile de montrer que si \mathcal{G} est une algèbre de Lie quelconque, ses idéaux résolubles \mathfrak{a} , \mathfrak{b} , etc. sont stables par addition ($\mathfrak{a} + \mathfrak{b} = \mathfrak{c}$ est résoluble) et que $\mathfrak{c}/\mathfrak{a} = \mathfrak{b}/(\mathfrak{a} \cap \mathfrak{b})$ est également résoluble. Cela implique qu'il existe un idéal résoluble *maximal* \mathcal{R} de \mathcal{G} , appelé *radical* de \mathcal{G} , tel que \mathcal{G}/\mathcal{R} soit *semi-simple*. \mathcal{G} est donc semi-simple si et seulement si son radical \mathcal{R} est nul. \mathcal{R} épuise en quelque sorte la résolubilité de \mathcal{G} et ce qui reste est pleinement non commutatif (semi-simple). Le *théorème de décomposition* de Levi (1905) qui résout une conjecture de Wilhelm Killing dit alors qu'il existe un complémentaire

6. Eugenio Levi (1883-1917) était le frère de Beppo Levi. Remarquable mathématicien, il fut le disciple de Luigi Bianchi et Ulisse Dini à la Scuola Normale Superiore de Pise. Il fut tué lors de la dramatique défaite italienne de Caporetto le 14 octobre 1917.

(le facteur de Levi) \mathcal{M} de \mathcal{R} dans \mathcal{G} (i.e. $\mathcal{R} + \mathcal{M} = \mathcal{G}$ et $\mathcal{R} \cap \mathcal{M} = \{0\}$), $\mathcal{M} \simeq \mathcal{G}/\mathcal{R}$, tel que \mathcal{G} devienne un produit semi-direct $\mathcal{G} \simeq \mathcal{R} \rtimes \mathcal{M}$.

Une autre façon d'éclairer la structure d'une algèbre de Lie générale est de considérer ce que l'on appelle sa *forme de Killing*. Il s'agit d'une forme bilinéaire symétrique sur \mathcal{G} entretenant les représentations adjointes et co-adjointes et qui mesure en quelque sorte la non-commutativité. Elle est définie par

$$K(X, Y) = \text{tr}(ad_X ad_Y)$$

et elle est bien symétrique car, pour deux matrices $n \times n$ quelconques A et B , $\text{tr}(AB) = \text{tr}(BA)$. Elle est invariante par les automorphismes de \mathcal{G} . Elle induit une application linéaire κ entre \mathcal{G} et \mathcal{G}^* qui associe à $X \in \mathcal{G}$ la forme linéaire $\kappa_X(Y) = K(X, Y)$. On note que

$$K(ad_Z X, Y) = K(Z, ad_X Y)$$

car

$$\begin{aligned} K(ad_Z X, Y) &= K([Z, X], Y) = \text{tr}(ad_{[Z, X]} ad_Y) = \text{tr}([ad_Z, ad_X] ad_Y) \\ &= \text{tr}(ad_Z ad_X ad_Y) - \text{tr}(ad_X ad_Z ad_Y) \\ K(Z, ad_X Y) &= K(Z, [X, Y]) = \text{tr}(ad_Z ad_{[X, Y]}) = \text{tr}(ad_Z [ad_X, ad_Y]) \\ &= \text{tr}(ad_Z ad_X ad_Y) - \text{tr}(ad_Z ad_Y ad_X) . \end{aligned}$$

On en tire, puisque

$$ad_Z X = [Z, X] = -[X, Z] = -ad_X Z ,$$

la relation

$$K(ad_Z X, Y) = -K(ad_X Z, Y) = -K(X, ad_Z Y) ,$$

ce qui montre que K entrelace les représentations adjointe et co-adjointe puisqu'il s'agit de la version infinitésimale dans \mathcal{G} de la relation

$$K(Ad_g X, Y) = K(X, Ad_{g^{-1}} Y)$$

avec $g = e^Z$ dans le groupe de Lie G intégrant \mathcal{G} .

$K(X, Y)$ étant bilinéaire symétrique elle définit une forme quadratique

$$Q(X) = K(X, X)$$

sur \mathcal{G} et, en développant $Q(X + Y)$ on montre immédiatement la relation classique

$$K(X, Y) = \frac{1}{2} (Q(X + Y) - Q(X) - Q(Y)) .$$

Par ailleurs elle est définie par une *trace*. Rappelons que la trace $\text{tr}(u)$ d'un endomorphisme u d'un \mathbb{K} -espace vectoriel V ($\mathbb{K} = \mathbb{R}$ ou \mathbb{C}) de dimension finie n , u étant de matrice A dans une base donnée, est la somme $\text{tr}(A)$ des termes diagonaux de A .

Il est trivial de voir que c'est une forme linéaire sur les u . Elle est évidemment invariante par transposition (i.e. dualité)⁷. et satisfait, répétons-le, $\text{tr}(AB) = \text{tr}(BA)$ (i.e. $\text{tr}([A, B]) = 0$). Le fait que la trace $\text{tr}(u)$ possède un sens géométrique *intrinsèque* indépendant de la base choisie (autrement dit que si C est un changement de base, $\text{tr}(C^{-1}AC) = \text{tr}(A)$) découle de ce qu'elle est égale à la somme des n valeurs propres λ_j (comptées avec multiplicité) de A i.e. à la somme $\sum_{j=1}^{j=n} \lambda_j$ des racines du polynôme caractéristique⁸

$$\begin{aligned} \Pi_A(\lambda) &= \det(\lambda I_n - A) = \lambda^n - \sum \lambda_j + \cdots + (-1)^n \prod \lambda_j \\ &= \lambda^n - \text{tr}(A) + \cdots + (-1)^n \det(A) . \end{aligned}$$

(On peut prendre $\Pi_A(\lambda) = \det(A - \lambda I_n)$ si l'on préfère : les racines sont les mêmes Or ce dernier est indépendant de la base choisie. La trace est d'ailleurs la dérivée du déterminant au sens où $\det(I_n + \varepsilon A) \sim I_n + \varepsilon \text{tr}(A)$ au premier ordre ou, ce qui revient au même, $\det(e^A) = e^{\text{tr}(A)}$.

Sur \mathbb{C} on peut toujours trigonaliser une matrice par un changement de base bien choisi de façon à avoir les λ_j sur la diagonale. Cela montre que si $P(x)$ est un polynôme, $\text{tr}(P(A)) = \sum P(\lambda_j)$ puisque $P(A)$ est trigonale avec les $P(\lambda_j)$ pour diagonale.

Si \mathcal{G} est abélienne alors $ad_X \equiv 0$ et donc $K \equiv 0$ et $\kappa \equiv 0$. Mais on peut aller plus loin. La forme de Killing est une trace de produits de ad_X qui, d'après le théorème d'Engel, sont nilpotentes lorsque l'algèbre de Lie \mathcal{G} est nilpotente. Il est donc pertinent de noter que si A est nilpotente son polynôme caractéristique est λ^n et que donc à la fois $\text{tr}(A)$, $\det(A)$ et toutes autres fonctions symétriques des racines sont nulles. Si A et B sont nilpotentes et *commutent* alors leurs produits et combinaisons linéaires sont nilpotentes. En effet si $A^r = 0$ et $B^s = 0$ et si $m = \max(r, s)$, $(AB)^m = A^m B^m$ (par commutation) = 0 (car $m \geq r, s$). Par ailleurs le développement de $(\alpha A + \beta B)^{2p}$ ne contient que des termes en $A^j B^{2m-j}$ et l'un des j ou $2m - j$ est $\geq m$ et donc tous ces termes sont nuls.

Si \mathcal{G} est nilpotente, alors

$$Q(X) = K(X, X) = \text{tr}((ad_X)^2) = 0$$

pour tout X et donc

$$K(X, Y) = \frac{1}{2} (Q(X + Y) - Q(X) - Q(Y)) = 0$$

7. A^T s'obtient en échangeant les lignes et les colonnes et les éléments diagonaux sont les intersections des lignes et colonnes de même rang.

8. Notons que si $\mathbb{K} = \mathbb{R}$ certaines λ_j peuvent être complexes. Mais leur somme est réelle car les λ_j complexes vont alors par paires de complexes conjugués $(\lambda_j, \overline{\lambda_j})$ et $\lambda_j + \overline{\lambda_j} = 2\Re(\lambda_j)$. De même $\lambda_j \overline{\lambda_j} = |\lambda_j|^2$ et donc $\det(A) = \prod_{j=1}^{j=n} \lambda_j$ est également réel.

pour tous X, Y . Donc $K \equiv 0$ équivaut à \mathcal{G} nilpotente, ce qui montre encore une fois que les algèbres nilpotentes sont les généralisations directes des algèbres abéliennes. À l'autre extrême, κ est un isomorphisme si et seulement si κ_X est injective et surjective, autrement dit si $\kappa_X(Y) (= K(X, Y)) = 0$ pour tout Y implique $X = 0$ et si pour tout $u(Y) \in \mathcal{G}^*$ il existe un $X \in \mathcal{G}$ tel que $u(Y) = K(X, Y)$.

En fait, on peut caractériser au moyen de la forme de Killing les deux propriétés complémentaires de résolubilité et de semi-simplicité qui, d'après le théorème de décomposition de Levi, permettent par produits semi-directs de construire toutes les algèbres de Lie. Il s'agit des *critères de Cartan*.

Théorème de Cartan (caractérisation par la forme de Killing).

- \mathcal{G} est résoluble si et seulement si $K(\mathcal{G}, [\mathcal{G}, \mathcal{G}]) \equiv 0$.
- \mathcal{G} est semi-simple si et seulement si $K(X, Y)$ est non dégénérée. \diamond

La forme de Killing est donc particulièrement utile pour comprendre la structure des \mathcal{G} semi-simples car alors $Q(X) = K(X, X)$ est une forme quadratique non dégénérée dont on peut déduire une métrique sur \mathcal{G} . On montre que \mathcal{G} semi-simple est l'algèbre de Lie d'un G semi-simple compact si et seulement si $K(X, Y)$ est non dégénérée définie négative.

Remarque. Revenons à nos algèbres de base \mathcal{V}_J et \mathcal{V}_S dont les représentations adjointes ont été explicitées aux sections 6.2 et 7 du chapitre 5. Pour \mathcal{V}_J , si $\mathfrak{t} = (\xi, \eta, \pi) \in \mathcal{V}_J$, la matrice de $ad_{\mathfrak{t}}$ est

$$ad_{\mathfrak{t}} = \begin{pmatrix} 0 & 0 & 0 \\ \pi & 0 & -\xi \\ 0 & 0 & 0 \end{pmatrix}$$

et comme $ad_{\mathfrak{t}}ad_{\mathfrak{t}'} = 0$, on vérifie que $K(\mathfrak{t}, \mathfrak{t}') \equiv 0$, ce qui correspond bien au fait que \mathcal{V}_J soit nilpotente. Pour \mathcal{V}_S qui n'est pas nilpotente, on a en revanche pour les $X = (\xi, \eta, \tau)$

$$ad_X = \begin{pmatrix} 0 & -\tau & \eta \\ \tau & 0 & -\xi \\ 0 & 0 & 0 \end{pmatrix},$$

$$ad_X ad_{X'} = \begin{pmatrix} -\tau\tau' & 0 & \tau\xi' \\ 0 & -\tau\tau' & \tau\eta' \\ 0 & 0 & 0 \end{pmatrix}.$$

et donc $K(X, X') = -2\tau\tau'$. On voit que $K(X, X') = 0$ dès que l'un des vecteurs a une composante rotationnelle nulle. Et comme

$$[X, X'] = (-\tau\eta' + \eta\tau', \tau\xi' - \xi\tau', 0),$$

on voit que

$$K(\mathcal{V}_S, [\mathcal{V}_S, \mathcal{V}_S]) \equiv 0,$$

ce qui correspond bien au fait que \mathcal{V}_S soit résoluble comme nous l'avons vu plus haut à la section 2.

3.2. Mesure de Haar et unimodularité

Sur tout groupe de Lie G de dimension finie n , il existe l'équivalent de la mesure de Lebesgue pour les espaces \mathbb{R}^n . Cette mesure μ est essentiellement unique à un facteur multiplicatif près et s'appelle une mesure de Haar.⁹ Il suffit de prendre une n -forme volume ω_0 sur \mathcal{G} (par exemple celle donnant la mesure de Lebesgue) et la translater sur tout G par translations à gauche. On obtient ainsi une n -forme volume ω sur G . Comme ω_0 est définie à un facteur multiplicatif près et que ce facteur reste constant par translation cela définit une mesure de Haar μ invariante à gauche. Si f est une fonction mesurable sur G on définit son intégrale $\int_G f d\mu$ (ou plus précisément $\int_G f(g) d\mu(g)$) par $\int_G f\omega$.

Remarque. En fait une mesure de Haar invariante à gauche existe plus généralement pour tout groupe topologique G séparable et localement compact. Tel est le vrai contenu du théorème (non trivial) d'existence de Alfred Haar (1933), le théorème d'unicité ayant été établi en 1935 par John von Neumann. \square

Mais si μ est une mesure de Haar invariante à gauche, sa translaturée à droite μ_h ($h \in G$) est aussi une mesure de Haar à gauche puisque les translations à gauche et à droite de G commutent. D'après le théorème d'unicité, il existe donc un morphisme de groupe $\Delta : G \rightarrow \mathbb{R}_{>0}^*$ (le groupe multiplicatif des réels > 0) tels que $\mu_h = \Delta(h)\mu$. Δ s'appelle la *fonction modulaire* et G est dit *unimodulaire* si $\Delta(g) \equiv 1$. Par exemple, nous avons vu aux sections 6.1 et 7 du chapitre 5 que les mesures de Lebesgue de \mathbb{V}_J et de \mathbb{V}_S sont des mesures de Haar à gauche et à droite et que donc \mathbb{V}_J et \mathbb{V}_S sont des groupes de Lie unimodulaires.

En fait $\Delta(h) = |\det(Ad_h)|$ car c'est Ad_h qui donne le jacobien du passage de μ à μ_h . L'unimodularité s'exprime ainsi au niveau infinitésimal de l'algèbre de Lie \mathcal{G} par le fait que les ad_X sont de *trace nulle*.

Des cas évidents de groupes de Lie unimodulaires sont les groupes nilpotents connexes, les groupes compacts, les groupes semi-simples. Si G est nilpotent ad_X est nilpotent et $Ad_{\exp(X)}$ est de déterminant 1 et, si G est connexe, on obtient G tout entier avec les $\exp(X)$. Si G est compact, $\Delta(G)$ doit être un sous-groupe compact de $\mathbb{R}_{>0}^*$ et donc est égal à $\{1\}$. Enfin si G est semi-simple, les Ad_h sont orthogonaux pour la forme de Killing (qui est non dégénérée) et donc $|\det(Ad_h)| = 1$.

9. Nous renvoyons à notre petit "vademecum" d'analyse fonctionnelle (chapitre 16 ci-dessous) pour une définition plus abstraite de la notion de mesure.

3.3. La classification des algèbres de Lie réelles de dimension ≤ 3

Nos deux algèbres de Lie de référence \mathcal{V}_J et \mathcal{V}_S sont des algèbres de Lie réelles de dimension 3. Notre excursus dans la théorie générale de leur structure nous incite à présenter la classification de ces algèbres. Elle constitue un résultat remarquable dû à Luigi Bianchi¹⁰ à la fin du XIX^e siècle et illustre de façon pédagogique comment des propriétés de structure exprimables en termes d'algèbre linéaire et de métrique permettent d'effectuer une classification à isomorphisme près. La référence [56] renvoie au *Mémoire* original et à sa traduction anglaise. Les présentations de cette classification, qui complète dans le cas réel la classification déjà effectuée par Lie dans le cas complexe et qui fut ensuite précisée par Cartan, sont nombreuses. Le lecteur intéressé par son contexte général pourra consulter par exemple, parmi beaucoup d'autres, l'ouvrage introductif de William Fulton et Joe Harris [202], ainsi que la thèse d'Allegra Fowler-Wright [190].

Nous présenterons plus bas dans deux autres sections les métriques riemanniennes et sous-riemanniennes sur ces algèbres de dimension 3.

Soit \mathcal{G} une algèbre de Lie réelle de dimension ≤ 3 . Nous avons vu à la section précédente 7 l'importance des suites $\mathcal{D}^j(\mathcal{G}) = \mathcal{G}^{(j)} = [\mathcal{G}^{(j-1)}, \mathcal{G}^{(j-1)}]$ et $\mathcal{G}_{(j)} = [\mathcal{G}, \mathcal{G}_{(j-1)}]$ ($\mathcal{G}^{(0)} = \mathcal{G}_{(0)} = \mathcal{G}$ et $\mathcal{G}^{(1)} = \mathcal{G}_{(1)}$). \mathcal{G} est résoluble si la suite $\mathcal{G}^{(j)}$ stationne en $\{0\}$ à partir d'un certain n et \mathcal{G} est nilpotent si c'est la suite $\mathcal{G}_{(j)}$ qui stationne en $\{0\}$ à partir d'un certain n . Deux idéaux de \mathcal{G} sont particulièrement importants : son centre $\mathcal{Z} = \mathcal{Z}(\mathcal{G})$ et son radical $\mathcal{R} = \mathcal{R}(\mathcal{G})$ qui est son plus grand idéal résoluble.

Si \mathcal{G} est de dimension 1, elle est commutative et isomorphe à $(\mathbb{R}, +)$.

Si \mathcal{G} est de dimension 2 et si $\{X_1, X_2\}$ en est une base vectorielle, alors

$$[X_1, X_2] = \alpha X_1 + \beta X_2 .$$

Si $\alpha = \beta = 0$, \mathcal{G} est abélienne et isomorphe à $(\mathbb{R}^2, +)$. Si α ou $\beta \neq 0$, par exemple $\alpha \neq 0$, l'algèbre dérivée $\mathcal{G}^{(1)} = \mathcal{D}(\mathcal{G}) = [\mathcal{G}, \mathcal{G}]$ est de dimension 1 et engendrée par $[X_1, X_2]$ puisque

$$[a_1 X_1 + a_2 X_2, b_1 X_1 + b_2 X_2] = (a_1 b_2 - a_2 b_1) [X_1, X_2] .$$

(On notera que $\mathcal{G}^{(1)} = \mathcal{G}$ est donc impossible pour une raison dimensionnelle). Comme $\alpha \neq 0$, on peut, pour normaliser, faire le changement de base

$$\{e_1, e_2\} = \left\{ X_1 + \frac{\beta}{\alpha} X_2, \frac{1}{\alpha} X_2 \right\} .$$

10. Nous avons déjà rencontré Luigi Bianchi à propos de la géométrie des connexions au chapitre 6.

Il s'agit bien d'un changement de base car son déterminant est $\frac{1}{\alpha} \neq 0$. Le commutateur devient alors

$$\begin{aligned} [e_1, e_2] &= \left[X_1 + \frac{\beta}{\alpha} X_2, \frac{1}{\alpha} X_2 \right] = \frac{1}{\alpha} [X_1, X_2] = \frac{1}{\alpha} (\alpha X_1 + \beta X_2) = X_1 + \frac{\beta}{\alpha} X_2 \\ &= e_1 . \end{aligned}$$

Cette algèbre \mathfrak{l}_2 est résoluble mais non nilpotente et de centre $\{0\}$.¹¹ Elle est, à isomorphisme près, la seule algèbre de Lie réelle non abélienne de dimension 2. Sa représentation adjointe est $ad_{e_1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ et $ad_{e_2} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$. La sous-algèbre $\mathcal{G}^{(1)} = \mathbb{R}e_1$ est un idéal et on montre que toutes les dérivations D de \mathfrak{l}_2 sont internes, i.e. de la forme ad_X (cf. chapitre 5, section 4).

L'algèbre \mathfrak{l}_2 est importante parce que toute algèbre de Lie \mathfrak{l} admettant \mathfrak{l}_2 comme idéal est décomposable en somme directe d'idéaux $\mathfrak{l} = \mathfrak{l}_2 \oplus \mathfrak{m}$, l'idéal \mathfrak{m} étant le centre de \mathfrak{l} relatif à \mathfrak{l}_2 , c'est-à-dire l'idéal des $x \in \mathfrak{l}$ qui commutent avec tous les y de \mathfrak{l}_2 .¹²

En dimension 3, la classification commence à devenir non triviale et il convient d'utiliser des propriétés de structure de \mathcal{G} et d'abord la dimension de l'algèbre dérivée $\mathcal{G}^{(1)} = [\mathcal{G}, \mathcal{G}]$.

3.3.1. $\dim \mathcal{G}^{(1)} = 0$.

Si $\mathcal{G}^{(1)} = \{0\}$ alors \mathcal{G} est commutative et isomorphe à $(\mathbb{R}^3, +)$. C'est la CLASSE DE BIANCHI I.

3.3.2. $\dim \mathcal{G}^{(1)} = 1$.

Si $\dim \mathcal{G}^{(1)} = 1$ on prend un X_1 engendrant $\mathcal{G}^{(1)}$ et on le complète en une base $\{X_1, X_2, X_3\}$ de \mathcal{G} . Par hypothèse, on a

$$[X_1, X_2] = \alpha X_1, \quad [X_1, X_3] = \beta X_1, \quad [X_2, X_3] = \gamma X_1$$

avec des α, β, γ non tous nuls.

(1) Supposons que α ou β (ils s'interchangent lorsqu'on interchange X_2 et X_3), par exemple α , soit $\neq 0$. On peut alors, pour normaliser, faire le changement de base

11. Comme $\mathcal{G}^{(1)} = \mathbb{R}e_1$, $\mathcal{G}^{(2)} = [\mathcal{G}^{(1)}, \mathcal{G}^{(1)}] = 0$ et \mathfrak{l}_2 est résoluble et $\mathcal{R}(\mathfrak{l}_2) = \mathfrak{l}_2$. Mais par ailleurs $\mathcal{G}^{(2)} = [\mathcal{G}^{(1)}, \mathcal{G}] = \mathbb{R}e_1 = \mathcal{G}^{(1)}$ et donc \mathfrak{l}_2 n'est pas nilpotente puisque la suite des $\mathcal{G}^{(i)}$ ne se stabilise pas sur $\{0\}$. Enfin, si $[\alpha e_1 + \beta e_2, x e_1 + y e_2] = (\alpha y - \beta x) e_1 = 0$ pour tout x, y alors $\alpha = \beta = 0$ et donc $\mathcal{Z} = \{0\}$.

12. Il s'agit bien d'un idéal car si $x \in \mathfrak{m}$ et $z \in \mathfrak{l}$ alors pour tout $y \in \mathfrak{l}_2$, d'après l'identité de Jacobi, $[[x, z], y] + [[z, y], x] + [[y, x], z] = 0$. Mais comme $[z, y] \in \mathfrak{l}_2$ car $y \in \mathfrak{l}_2$ et \mathfrak{l}_2 est un idéal et comme $x \in \mathfrak{m}$, $[[z, y], x] = 0$. Par ailleurs, comme $y \in \mathfrak{l}_2$ et $x \in \mathfrak{m}$, $[y, x] = 0$ et donc $[[y, x], z] = 0$. Par conséquent $[[x, z], y] = 0$ et donc $[x, z] \in \mathfrak{m}$. La somme $\mathfrak{l}_2 \oplus \mathfrak{m}$ est bien directe car si $x \in \mathfrak{l}_2 \cap \mathfrak{m}$ alors x est dans le centre de \mathfrak{l}_2 et celui-ci est nul. En utilisant le fait que les ad_X de \mathfrak{l} induisent des dérivations de \mathfrak{l}_2 et que toutes les dérivations de \mathfrak{l}_2 sont internes, on montre alors que cette \oplus est bien \mathfrak{l} .

$e_1 = X_1$, $e_2 = \frac{1}{\alpha}X_2$, $e_3 = \alpha X_3 - \beta X_2 + \gamma X_1$. C'est bien un changement de base car le déterminant est 1. On obtient alors

$$\begin{aligned} [e_1, e_2] &= \left[X_1, \frac{1}{\alpha}X_2 \right] = \frac{1}{\alpha} [X_1, X_2] = \frac{1}{\alpha} \alpha X_1 = X_1 \\ &= e_1 \\ [e_1, e_3] &= [X_1, \alpha X_3 - \beta X_2 + \gamma X_1] = \alpha (\beta X_1) - \beta (\alpha X_1) \\ &= 0 \\ [e_2, e_3] &= \left[\frac{1}{\alpha}X_2, \alpha X_3 - \beta X_2 + \gamma X_1 \right] = \gamma X_1 - \frac{\gamma}{\alpha} \alpha X_1 \\ &= 0 \end{aligned}$$

à savoir l'algèbre résoluble $\mathfrak{l}_2 \oplus (\mathbb{R}, +)$ représentable par les matrices 2×2 triangulaires supérieures avec la base

$$e_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Sa représentation adjointe dans la base $\{e_1, e_2, e_3\}$ est

$$ad_{e_1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad ad_{e_2} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad ad_{e_3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Comme \mathfrak{l}_2 , cette algèbre est résoluble mais non nilpotente. $\mathcal{G}^{(1)} = \mathbb{R}e_1$ et $\mathcal{Z} = \mathbb{R}e_3$ et donc $\mathcal{G}^{(1)} \subsetneq \mathcal{Z}$. Elle n'est pas unimodulaire (ad_{e_2} n'est pas de trace nulle). Sa forme de Killing se réduit à $\|e_2\|^2 = 1$, tous les autres $\langle e_i, e_j \rangle$ étant nuls. C'est la CLASSE DE BIANCHI III.

(2) Si $\alpha = \beta = 0$ mais $\gamma \neq 0$, on a $[X_1, X_2] = 0$, $[X_1, X_3] = 0$, et, pour normaliser, on peut utiliser le changement de base $e_1 = X_2$, $e_2 = X_3$, $e_3 = \gamma X_1$. On obtient alors

$$\begin{aligned} [e_1, e_2] &= [X_2, X_3] = \gamma X_1 = e_3 \\ [e_1, e_3] &= 0 \\ [e_2, e_3] &= 0 \end{aligned}$$

C'est l'algèbre nilpotente de Heisenberg \mathfrak{h} réalisable par les matrices strictement triangulaires supérieures.¹³ $\mathcal{G}^{(1)} = \mathbb{R}e_3$ et $\mathcal{Z} = \mathbb{R}e_3$ et donc $\mathcal{G}^{(1)} = \mathcal{Z}$. Sa représentation adjointe dans la base $\{e_1, e_2, e_3\}$ est

$$ad_{e_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad ad_{e_2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad ad_{e_3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

13. "Strictement" signifie, rappelons-le, que les éléments diagonaux sont tous nuls.

Elle est unimodulaire. Sa forme de Killing est identiquement nulle à cause de la nilpotence. C'est la CLASSE DE BIANCHI II. Nous l'avons traitée avec \mathbb{V}_J muni de la base $\{\mathfrak{t}_1, \mathfrak{t}_2, \mathfrak{t}_3\}$ avec $[\mathfrak{t}_1, \mathfrak{t}_2] = \mathfrak{t}_3$ et $[\mathfrak{t}_1, \mathfrak{t}_3] = [\mathfrak{t}_2, \mathfrak{t}_3] = 0$, c'est-à-dire

$$e_1 = \mathfrak{t}_1, e_2 = \mathfrak{t}_2, e_3 = \mathfrak{t}_3 .$$

Nous voyons ainsi comment notre *premier modèle* $(\mathbb{V}_J, \mathcal{V}_J)$ se situe dans la classification. Nous avons déjà noté à la section 6 du chapitre 5 que \mathbb{V}_J est unimodulaire.

3.3.3. $\dim \mathcal{G}^{(1)} = 2$.

Si $\dim \mathcal{G}^{(1)} = 2$, soit $\{X_1, X_2\}$ une base de $\mathcal{G}^{(1)}$ que l'on étend à une base $\{X_1, X_2, X_3\}$ de \mathcal{G} . On montre d'abord que X_1 et X_2 commutent nécessairement et que donc la sous-algèbre $\mathcal{G}^{(1)}$ ne peut pas être isomorphe à \mathfrak{l}_2 . En effet, l'application adjointe ad de \mathcal{G} prend par définition ses valeurs dans $\mathcal{G}^{(1)}$ et donc sa restriction à $\mathcal{G}^{(1)}$ est bien définie. Considérons $ad_{X_1} \upharpoonright_{\mathcal{G}^{(1)}}$. Si $[X_1, X_2] = \alpha X_1 + \beta X_2$, sa matrice est $\begin{pmatrix} 0 & \alpha \\ 0 & \beta \end{pmatrix}$. Mais comme $X_1 \in \mathcal{G}^{(1)}$, la trace de cette matrice est nulle. En effet, si $X = [Y, Z]$, $ad_X = ad_{[Y, Z]} = [ad_Y, ad_Z]$ car ad est un morphisme d'algèbres de Lie et la trace d'un commutateur de matrices est toujours nulle. Donc $\beta = 0$. Il en va de même pour α car $X_2 \in \mathcal{G}^{(1)}$.

Donc $[X_1, X_2] = 0$. Mais cela implique, puisque $\dim \mathcal{G}^{(1)} = 2$, que $[X_1, X_3] \neq 0$ et $[X_2, X_3] \neq 0$ et donc que $ad_{X_3} \upharpoonright_{\mathcal{G}^{(1)}}$ est un automorphisme. Toutes les algèbres de Lie de cette classe sont résolubles, non nilpotentes et de centre nul.¹⁴

Il est alors facile de montrer que les isomorphismes d'algèbres de Lie $\mathcal{G} \simeq \mathcal{H}$ se traduisent par la *similitude* des matrices $A = ad_{X_3} \upharpoonright_{\mathcal{G}^{(1)}}$ et $B = ad_{X'_3} \upharpoonright_{\mathcal{H}^{(1)}}$. Le problème de la classification se ramène par conséquent à celui des matrices 2×2 inversibles à similitude près.

Mais celle-ci est bien connue et dépend des valeurs propres λ_1, λ_2 de A qui doivent être toutes deux $\neq 0$ puisque A est inversible. Trois cas sont possibles : $\lambda_1, \lambda_2 \in \mathbb{R}$ et $\lambda_1 \neq \lambda_2$; $\lambda_1, \lambda_2 \in \mathbb{R}$ et $\lambda_1 = \lambda_2$; $\lambda_1, \lambda_2 \in \mathbb{C} - \mathbb{R}$ et $\lambda_1 = \lambda \in \mathbb{C}$, $\lambda_2 = \bar{\lambda}$.

(1) Si $\lambda_1, \lambda_2 \in \mathbb{R} - \{0\}$ et $\lambda_1 \neq \lambda_2$, la matrice A est diagonalisable et il existe donc une base $\{e_1, e_2\}$ de $\mathcal{G}^{(1)}$ telle que

$$\begin{cases} [e_1, e_2] = 0 \\ [X_3, e_1] = ad_{X_3}(e_1) = A(e_1) = \lambda_1 e_1 \\ [X_3, e_2] = ad_{X_3}(e_2) = A(e_2) = \lambda_2 e_2 \end{cases}$$

14. En effet comme $\mathcal{G}^{(1)}$ est abélienne, $\mathcal{G}^{(2)} = [\mathcal{G}^{(1)}, \mathcal{G}^{(1)}] = 0$ et \mathcal{G} est donc résoluble. Mais $\mathcal{G}_{(2)} = [\mathcal{G}_{(1)}, \mathcal{G}] = \mathcal{G}_{(1)}$ et \mathcal{G} n'est donc pas nilpotente. Enfin si $z \in \mathcal{Z}$, alors $ad_{X_3}(z) = 0$ et comme ad_{X_3} est un automorphisme de $\mathcal{G}^{(1)}$, il faut que $z = \alpha X_3$. Donc $\alpha ad_{X_3}(x) = [z, x]$ et comme $z \in \mathcal{Z}$, $[z, x] = 0$ et donc $\alpha ad_{X_3}(x) = 0$ pour tout x ce qui implique $\alpha = 0$ et $z = 0$.

En utilisant, pour normaliser, le changement de variable $e_3 = -\frac{1}{\lambda_1} X_3$ (qui est bien défini puisque $\lambda_1 \neq 0$) et en posant $\ell = \frac{\lambda_2}{\lambda_1}$ on obtient

$$\begin{aligned} [e_1, e_2] &= 0 \\ [e_1, e_3] &= -[e_3, e_1] = \frac{1}{\lambda_1} \lambda_1 e_1 \\ &= e_1 \\ [e_2, e_3] &= -[e_3, e_2] = \frac{1}{\lambda_1} \lambda_2 e_2 \\ &= \ell e_2 . \end{aligned}$$

Évidemment ℓ et $\frac{1}{\ell}$ donnent le même groupe puisqu'ils correspondent à l'échange de λ_1 et λ_2 . On obtient ainsi une famille à 1-paramètre d'algèbres de Lie non isomorphes entre elles. Leur représentation adjointe dans la base $\{e_1, e_2, e_3\}$ est

$$ad_{e_1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad ad_{e_2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \ell \\ 0 & 0 & 0 \end{pmatrix}, \quad ad_{e_3} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -\ell & 0 \\ 0 & 0 & 0 \end{pmatrix} .$$

Elles ne sont pas unimodulaires si $\ell \neq -1$ (ad_{e_3} n'est pas de trace nulle). Leur forme de Killing se réduit à $\|e_3\|^2 = 1 + \ell^2 > 0$, tous les autres $\langle e_i, e_j \rangle$ étant nuls. Elles constituent la CLASSE DE BIANCHI VI.

Une réalisation physique bien connue de ce cas pour $\ell = -1$ (i.e. A de trace nulle) est l'algèbre de Lie $\mathfrak{p}(1, 1)$ du groupe de Poincaré $P(1, 1)$ des matrices 3×3 $(M, v) = \begin{pmatrix} M & v \\ 0 & 1 \end{pmatrix}$ où $M \in O(1, 1)$ (le groupe des isométries du plan de Minkowski doté de la métrique hyperbolique $ds^2 = dx^2 - dy^2$) et $v \in R^2$. On peut noter ce dernier également $SH(2)$, l'équivalent "hyperbolique" de $SO(2)$. L'algèbre $\mathfrak{p}(1, 1) = \mathfrak{sh}(2)$ représente la CLASSE DE BIANCHI VI₀. Les matrices \mathfrak{m} sont symétriques et de diagonale nulle $\mathfrak{m}(m) = \begin{pmatrix} 0 & m \\ m & 0 \end{pmatrix}$ et leur algèbre est donc de dimension 1, ce qui avec les deux dimensions du vecteur tangent X donne bien une algèbre de dimension 3. Le crochet est

$$[(\mathfrak{m}, X), (\mathfrak{m}', X')] = ([\mathfrak{m}, \mathfrak{m}'], \mathfrak{m}X' - \mathfrak{m}'X) .$$

Il est naturel de prendre pour base

$$e_1 = (0, X_1), \quad e_2 = (0, X_2), \quad e_3 = (\mathfrak{m}(1), 0) .$$

On obtient alors $[e_1, e_2] = 0$, $[e_1, e_3] = (0, -\mathfrak{m}(1)X_1)$, $[e_2, e_3] = (0, -\mathfrak{m}(1)X_2)$. Pour obtenir la base précédente, il faut donc

$$-\mathfrak{m}(1)X_1 = X_1, \quad -\mathfrak{m}(1)X_2 = \ell_2 .$$

Les vecteurs X_1 et X_2 sont donc les vecteurs propres de $\mathfrak{m}(1)$ (définis à une homothétie près). Or les valeurs propres sont ± 1 et le vecteur propre pour -1 est $X_1 = (1, -1)$. Le vecteur X_2 doit donc être le vecteur propre $(1, 1)$ pour la valeur propre 1 , ce qui impose $\ell = -1$. La base est alors

$$e_1 = (0, (1, -1)), \quad e_2 = (0, (1, 1)), \quad e_3 = (\mathfrak{m}(1), 0) .$$

(2-1) Si $\lambda_1, \lambda_2 \in \mathbb{R} - \{0\}$ et $\lambda_1 = \lambda_2 = \lambda \neq 0$, et si $A = ad_{X_3}$ est diagonalisable, on obtient le cas $\ell = 1$, i.e.

$$[e_1, e_2] = 0, \quad [e_1, e_3] = e_1, \quad [e_2, e_3] = e_2 .$$

C'est la CLASSE DE BIANCHI V.

(2-2) Si A n'est pas diagonalisable, il y a toujours un X_3 complétant la base $\{e_1, e_2\}$ telle que $A = ad_{X_3}$ puisse se mettre sous la forme $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$. On aura alors

$$\begin{cases} [X_3, e_1] = A(e_1) = \lambda e_1 \\ [X_3, e_2] = A(e_2) = e_1 + \lambda e_2 \end{cases}$$

En normalisant en prenant comme plus haut $e_3 = -\frac{1}{\lambda}X_3$ et en remplaçant e_1 par $\frac{1}{\lambda}e_1$, on obtient les commutateurs

$$[e_1, e_2] = 0, \quad [e_1, e_3] = e_1, \quad [e_2, e_3] = e_1 + e_2 .$$

Sa représentation adjointe dans la base $\{e_1, e_2, e_3\}$ est

$$ad_{e_1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad ad_{e_2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad ad_{e_3} = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} .$$

Elle n'est pas unimodulaire (ad_{e_3} est de trace -2). Sa forme de Killing se réduit à $\|e_3\|^2 = 2 > 0$, tous les autres $\langle e_i, e_j \rangle$ étant nuls. C'est la CLASSE DE BIANCHI IV.

(3) Enfin, si $\lambda_1 = \lambda = a + ib \in \mathbb{C}$ avec $b \neq 0$, $\lambda_2 = \bar{\lambda} = a - ib$, a et b de même signe (sinon on échange λ et $\bar{\lambda}$), on peut choisir une base $\{e_1, e_2\}$ telle que $ad_{X_3}(e_1 + ie_2) = \lambda(e_1 + ie_2)$. Cela implique

$$ad_{X_3}(e_1) + iad_{X_3}(e_2) = (a + ib)(e_1 + ie_2) = ae_1 - be_2 + i(ae_2 + be_1)$$

et donc $ad_{X_3}(e_1) = ae_1 - be_2$ et $ad_{X_3}(e_2) = ae_2 + be_1$. Si l'on pose, pour normaliser, $e_3 = -\frac{1}{b}X_3$ et $\ell = \frac{a}{b} \geq 0$ (qui sont bien définis puisque $b \neq 0$), on obtient

$$\begin{aligned}
[e_1, e_2] &= 0 \\
[e_1, e_3] &= -[e_3, e_1] = \frac{1}{b}(ae_1 - be_2) \\
&= \ell e_1 - e_2 \\
[e_2, e_3] &= -[e_3, e_2] = \frac{1}{b}(ae_2 + be_1) \\
&= e_1 + \ell e_2 .
\end{aligned}$$

On obtient ainsi une autre famille à 1-paramètre d'algèbres de Lie non isomorphes entre elles. Leur représentation adjointe dans la base $\{e_1, e_2, e_3\}$ est

$$ad_{e_1} = \begin{pmatrix} 0 & 0 & \ell \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}, \quad ad_{e_2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & \ell \\ 0 & 0 & 0 \end{pmatrix}, \quad ad_{e_3} = \begin{pmatrix} -\ell & -1 & 0 \\ 1 & -\ell & 0 \\ 0 & 0 & 0 \end{pmatrix} .$$

Elles ne sont pas unimodulaires si $\ell \neq 0$ (ad_{e_3} est de trace -2ℓ). Leur forme de Killing se réduit à $\|e_3\|^2 = -2 + 2\ell^2$, tous les autres $\langle e_i, e_j \rangle$ étant nuls. Pour $\ell = 1$, elle est identiquement nulle. C'est la CLASSE DE BIANCHI VII.

On note que pour $\ell = 0$, $a = 0$ et $\lambda = ib$ imaginaire pur, on obtient

$$[e_1, e_2] = 0, \quad [e_1, e_3] = -e_2, \quad [e_2, e_3] = e_1 ,$$

c'est-à-dire l'algèbre de Lie $\mathfrak{se}(2)$ de $SE(2)$ que nous avons traitée avec la base

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = 0, \quad [X_2, X_3] = X_1 ,$$

c'est-à-dire

$$e_1 = X_1, \quad e_2 = X_3, \quad e_3 = -X_2 .$$

Nous voyons ainsi comment notre *second modèle* $(\mathbb{V}_S, \mathcal{V}_S)$ se situe dans la classification. Il est unimodulaire. Sa forme de Killing se réduit à $\|e_3\|^2 = -2$, tous les autres $\langle e_i, e_j \rangle$ étant nuls.¹⁵ C'est la CLASSE DE BIANCHI VII₀.

3.3.4. $\dim \mathcal{G}^{(1)} = 3$.

Si $\dim \mathcal{G}^{(1)} = 3$, $\mathcal{G}^{(1)} = \mathcal{G}$ et \mathcal{G} ne peut donc pas être nilpotente puisque

$$\mathcal{G}_{(2)} = [\mathcal{G}_{(1)}, \mathcal{G}] = [\mathcal{G}, \mathcal{G}] = \mathcal{G}_{(1)} = \mathcal{G}$$

et donc $\mathcal{G}_{(n)} = \mathcal{G}$ pour tout n . Elle ne peut pas non plus être résoluble puisque

$$\mathcal{G}^{(2)} = [\mathcal{G}^{(1)}, \mathcal{G}^{(1)}] = [\mathcal{G}, \mathcal{G}] = \mathcal{G}$$

et donc $\mathcal{G}^{(n)} = \mathcal{G}$ pour tout n . En fait \mathcal{G} est *simple*. En effet soit \mathfrak{a} un idéal de \mathcal{G} . S'il était différent de $\{0\}$ ou \mathcal{G} , il serait de dimension 1 ou 2 et donc résoluble et

15. La forme de Killing n'est pas forcément ≥ 0 et des vecteurs peuvent être de norme < 0 (comme les vecteurs de genre espace en relativité restreinte).

il en irait de même de \mathcal{G}/\mathfrak{a} qui serait de dimension 2 ou 1. Mais alors, comme \mathfrak{a} et \mathcal{G}/\mathfrak{a} seraient résolubles, \mathcal{G} serait lui-même résoluble, ce qui n'est pas le cas. Une conséquence en est que le centre $\mathcal{Z}(\mathcal{G})$ et le radical $\mathcal{R}(\mathcal{G})$ sont nuls car ce sont des idéaux et $\mathcal{Z}(\mathcal{G}) = \mathcal{G}$ est impossible puisque \mathcal{G} n'est pas abélienne et $\mathcal{R}(\mathcal{G}) = \mathcal{G}$ est impossible puisque \mathcal{G} n'est pas résoluble.

D'après le théorème d'Engel il existe donc $X \in \mathcal{G}$ tel que ad_X soit non nilpotente. Cela peut se vérifier facilement à la main en construisant à partir de ad_X nilpotente une $ad_{X'}$ non nilpotente. En effet si $X \neq 0$, ad_X est de rang 2 puisque $ad_X(X) = [X, X] = 0$ et $ad_X(Y) \neq 0$, $ad_X(Z) \neq 0$ si $\{X, Y, Z\}$ est une base (car sinon $[X, Y] = 0$ ou $[X, Z] = 0$ et $\dim \mathcal{G}^{(1)} < 3$). Donc ad_X a comme valeurs propres 0 et soit $\lambda_1, \lambda_2 \in \mathbb{R}$, soit $\lambda, \bar{\lambda} \in \mathbb{C}$. Si ad_X est nilpotente ses valeurs propres sont toutes nulles et elle peut se mettre sous forme triangulaire supérieure stricte dans une

certaine base $B : ad_X = \begin{pmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix}$ et comme ad_X est de rang 2, nécessairement

$a, c \neq 0$. Dans cette base, $X = (u, v, w)$ doit satisfaire $av + bw = 0$ et $cw = 0$ et, puisque $a, c \neq 0$, $w = 0$ et $v = 0$. On peut prendre par exemple $X = (1, 0, 0)$. Mais la forme de ad_X implique qu'il existe un $X' = (u', v', w')$ tel que $ad_X(X') = aX$. Il suffit que $v' = 1, w' = 0$. On peut prendre par exemple $X' = (0, 1, 0)$ Mais alors, comme $-ad_{X'}(X) = ad_X(X') = aX$, X est un vecteur propre de $ad_{X'}$ de valeur propre $-a \neq 0$ et $ad_{X'}$ n'est pas nilpotente.

On peut donc considérer un $X \in \mathcal{G}$ tel que ad_X soit de rang 2 et non nilpotente et de valeurs propres $\lambda_1, \lambda_2 \in \mathbb{R}$ ou $\lambda, \bar{\lambda} \in \mathbb{C}$ avec λ_1 ou $\lambda_2 \neq 0$ et $\lambda \neq 0$. Mais en fait, comme $\mathcal{G}^{(1)} = \mathcal{G}$, X est un commutateur et donc la trace $\text{Tr}(ad_X) = 0$. Par conséquent $\lambda_2 = -\lambda_1$ dans le premier cas et $\lambda = -\bar{\lambda}$ dans le second ce qui signifie que λ est imaginaire pur.

(1) Cas $\lambda_1 = \lambda \neq 0, \lambda_2 = -\lambda$. Soit X_1 un vecteur propre pour λ et X_2 un vecteur propre pour $-\lambda$. On a donc $[X, X_1] = \lambda X_1$ et $[X, X_2] = -\lambda X_2$. L'identité de Jacobi donne alors immédiatement $[X_1, X_2]$. En effet

$$\begin{aligned} [X, [X_1, X_2]] &= -[X_1, [X_2, X]] - [X_2, [X, X_1]] \\ &= -[X_1, \lambda X_2] - [X_2, \lambda X_1] = 0. \end{aligned}$$

Mais comme ad_X est de rang 2, son noyau se réduit à $\mathbb{R}X$ et par conséquent $[X_1, X_2] = \mu X$ avec $\mu \neq 0$. En faisant, pour normaliser, le changement de base

$$e_1 = X_2, \quad e_2 = \frac{a}{\lambda} X, \quad e_3 = \frac{b}{\lambda \mu} X_1$$

(qui est bien défini puisque $\lambda, \mu \neq 0$), on obtient

$$\begin{aligned}
[e_1, e_2] &= \frac{a}{\lambda} [X_2, X] = \frac{a}{\lambda} \lambda X_2 = aX_2 \\
&= ae_1 \\
[e_1, e_3] &= \left[X_2, \frac{b}{\lambda\mu} X_1 \right] = -\frac{b}{\lambda\mu} \mu X = -\frac{b}{\lambda} \frac{\lambda}{a} e_2 \\
&= -\frac{b}{a} e_2 \\
[e_2, e_3] &= \left[\frac{a}{\lambda} X, \frac{b}{\lambda\mu} X_1 \right] = \frac{ab}{\lambda^2\mu} \lambda X_1 = \frac{ab}{\lambda\mu} \frac{\lambda\mu}{b} e_3 \\
&= ae_3 .
\end{aligned}$$

Pour $a = b = 2$, on obtient

$$[e_1, e_2] = 2e_1, [e_1, e_3] = -e_2, [e_2, e_3] = 2e_3 .$$

Sa représentation adjointe dans la base $\{e_1, e_2, e_3\}$ est

$$ad_{e_1} = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}, ad_{e_2} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}, ad_{e_3} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -2 & 0 \end{pmatrix} .$$

Il est unimodulaire. Sa forme de Killing se réduit à $\|e_2\|^2 = 8$, et

$$\langle e_1, e_3 \rangle = \langle e_3, e_1 \rangle = 4,$$

tous les autres $\langle e_i, e_j \rangle$ étant nuls. Cela signifie que le produit scalaire est

$$\langle x_1e_1 + x_2e_2 + x_3e_3, y_1e_1 + y_2e_2 + y_3e_3 \rangle = 4(x_1y_3 + x_3y_1) + 8x_2y_2$$

et que la norme d'un vecteur est $\|x\|^2 = 8(x_1x_3 + x_2^2)$. On voit que e_1 et e_3 sont de norme nulle, que $e_1 + e_3$ est de norme 1, $e_1 - e_3$ de norme -1 et que ces deux vecteurs sont orthogonaux. C'est la CLASSE DE BIANCHI VIII.

La représentation la plus connue de cette algèbre est l'algèbre $\mathfrak{sl}(2)$ des matrices réelles 2×2 de trace nulle $\begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ avec la base

$$e_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, e_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, e_3 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} .$$

L'algèbre $\mathfrak{sl}(2)$ est l'algèbre de Lie du groupe spécial linéaire $SL(2, \mathbb{R})$ des matrices 2×2 réelles $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ de déterminant $\alpha\delta - \beta\gamma = 1$, c'est-à-dire des automorphismes linéaires du plan qui préservent les aires orientées puisque c'est le déterminant (le Jacobien) qui définit le facteur de transformation de la mesure des

aires.¹⁶ Ce groupe défini sur une sous variété de \mathbb{R}^4 (dont les intersections avec les hyperplans de coordonnées $\alpha = \text{cste}$, etc., sont des paraboloides hyperboliques) est d'une grande richesse et intervient dans énormément de domaines mathématiques et physiques. Serge Lang lui a consacré un ouvrage [314] qui est un grand classique.¹⁷ C'est un groupe connexe, non simplement connexe (son groupe d'homotopie π_1 est isomorphe à \mathbb{Z}), non compact, simple, de centre $\{\pm I\}$ et il admet le groupe des rotations $SO(2)$ comme sous-groupe compact maximal. Ses éléments sont de plusieurs types. Comme $\det(A) = 1$, le polynôme caractéristique de A est $\lambda^2 - \text{tr}(A)\lambda + 1$ de discriminant $\Delta = \text{tr}(A)^2 - 4$.

(i) Si $\text{tr}(A) < 2$, A est dit "elliptique", les 2 valeurs propres sont complexes conjuguées et A est conjuguée à une rotation qui laisse les cercles $x^2 + y^2 = r^2$ invariants.

(ii) Si $\text{tr}(A) > 2$, A est dit "hyperbolique", les 2 valeurs propres sont réelles et A est conjuguée à une rotation hyperbolique comme $\begin{pmatrix} \alpha & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix}$ qui laisse les hyperboles $xy = c$ invariants.

(iii) Si $\text{tr}(A) = 2$, A est conjuguée à une transvection de type $\begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}$.

(2) Cas $\lambda, \bar{\lambda} \in \mathbb{C}$, $\lambda \neq 0$ imaginaire pur, $\lambda = i\nu$, $\bar{\lambda} = -i\nu = -\lambda$, $\nu > 0$. Il existe alors un $X_1 + iX_2$ tel que $ad_{X'}(X_1 + iX_2) = i\nu(X_1 + iX_2)$, ce qui implique

$$\begin{aligned} ad_{X'}(X_1) + iad_{X'}(X_2) &= [X, X_1] + i[X, X_2] = -\nu X_2 + i\nu X_1 \\ [X, X_1] &= -\nu X_2, \quad [X, X_2] = \nu X_1. \end{aligned}$$

On notera la différence avec le cas précédent $[X, X_1] = \lambda X_1$ et $[X, X_2] = -\lambda X_2$. L'identité de Jacobi donne alors immédiatement $[X_1, X_2]$. En effet

$$[X, [X_1, X_2]] = [X_1, \nu X_1] + [X_2, \nu X_2] = 0$$

16. $SL_2(\mathbb{R})$ est bien de dimension 3 car il y a 4 dimensions pour les coefficients des matrices et une condition $\det(A) = 1$.

17. $SL_2(\mathbb{R})$ est un sous groupe de $SL_2(\mathbb{C})$ qui est encore beaucoup plus riche et contient aussi le sous-groupe $SU(2)$.

et donc $[X_1, X_2] \in \ker(ad_X) = \mathbb{R}X$, i.e. $[X_1, X_2] = \mu X$ avec $\mu \neq 0$. On peut normaliser en posant $X' = \frac{1}{\nu}X$, $X'_1 = \frac{1}{\sqrt{|\mu\nu|}}X_1$, $X'_2 = -\frac{1}{\sqrt{|\mu\nu|}}X_2$ et l'on obtient

$$\begin{aligned} [X', X'_1] &= \frac{1}{\nu\sqrt{|\mu\nu|}} [X, X_1] = -\frac{1}{\sqrt{|\mu\nu|}} X_2 = X'_2 \\ [X', X'_2] &= -\frac{1}{\sqrt{|\mu\nu|}} \frac{1}{\nu} [X, X_2] = -\frac{1}{\sqrt{|\mu\nu|}} X_1 = -X'_1 \\ [X'_1, X'_2] &= -\frac{1}{|\mu\nu|} [X_1, X_2] = -\frac{\mu}{|\mu\nu|} X = -\frac{\mu\nu}{|\mu\nu|} X' = -\frac{\mu}{|\mu|} X' \text{ car } \nu > 0. \end{aligned}$$

On obtient deux cas différents suivant le signe de μ . Si $\mu > 0$, $[X'_1, X'_2] = -X'$. Si l'on pose

$$e_1 = X'_1 - X', \quad e_2 = 2X'_2, \quad e_3 = X'_1 + X',$$

on obtient

$$[e_1, e_2] = 2e_1, [e_1, e_3] = -e_2, [e_2, e_3] = 2e_3$$

c'est-à-dire de nouveau \mathfrak{sl}_2 . C'est donc le cas $\mu < 0$, $[X'_1, X'_2] = X'$ qui apporte de la nouveauté. En prenant

$$e_1 = X'_1, \quad e_2 = X'_2, \quad e_3 = X',$$

on obtient

$$[e_1, e_2] = e_3, [e_2, e_3] = e_1, [e_3, e_1] = e_2.$$

Sa représentation adjointe dans la base $\{e_1, e_2, e_3\}$ est

$$ad_{e_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad ad_{e_2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad ad_{e_3} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Il est unimodulaire et sa forme de Killing est définie par $\|e_i\|^2 = -2$, et $\langle e_i, e_j \rangle = 0$ si $i \neq j$ (la base des e_i est orthogonale). Elle est non dégénérée définie ≤ 0 . C'est la CLASSE DE BIANCHI IX.

La réalisation la plus connue de cette algèbre est $\mathfrak{so}(3)$, l'algèbre de Lie du groupe compact $G = SO(3)$ des rotations de \mathbb{R}^3 , la compacité de G étant en accord avec le fait que la forme de Killing soit non dégénérée définie ≤ 0 comme nous l'avons vu plus haut. Les éléments A de $SO(3)$ sont les matrices réelles 3×3 orthogonales de déterminant 1, A étant dite orthogonale si sa transposée est égale à son inverse : $A^t = A^{-1}$. Les rotations sont des produits de rotations $R_x(\theta)$, $R_y(\varphi)$, $R_z(\psi)$ autour des axes x, y, z de \mathbb{R}^3 . L'algèbre de Lie $\mathfrak{so}(3)$ est l'algèbre des matrices réelles 3×3 antisymétriques. Une base naturelle est

$$e_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Le groupe des rotations est déjà notablement compliqué comme on peut s'en convaincre en jouant avec un Rubik cube (voir l'*Introduction*) où en observant les mouvements étranges de corps en rotation comme les gyroscopes ou les toupies. Mais il cache une profonde subtilité supplémentaire lié au fait qu'il *n'est pas simplement connexe*. Son revêtement universel est le groupe spécial unitaire $SU(2)$ que nous avons rencontré à la section 2.3 du chapitre 5 comme groupe des matrices 2×2 sur \mathbb{C} de la forme $\begin{pmatrix} u & v \\ -\bar{v} & \bar{u} \end{pmatrix}$ de déterminant $u\bar{u} + v\bar{v} = |u|^2 + |v|^2 = 1$. Nous avons alors vu que $SU(2)$ est isomorphe au groupe des quaternions unitaires qu'est la sphère \mathbb{S}^3 . $SU(2)$ est un groupe réel de dimension 3, compact et simplement connexe en tant que variété et simple en tant que groupe. Le lien entre $SO(3)$ et $SU(2)$ se fait à travers les quaternions. Nous reviendrons plus en détail sur le corps \mathbb{H}_Q des quaternions au chapitre 15 section 4.6.2. Les quaternions s'écrivent¹⁸ $q = w + xi + yj + zk$ avec

$$\begin{aligned} i^2 = j^2 = k^2 &= -1, \\ ij = k, \quad jk = i, \quad ki = j, \quad ji = -ij, \quad kj = -jk, \quad ik = -ki. \end{aligned}$$

On a donc

$$q = (w + xi) + (y + zi)j = u + jv$$

avec $u = w + xi$, $v = y + zi \in \mathbb{C}$. On associe à q unitaire (i.e. tel que $|q|^2 = w^2 + x^2 + y^2 + z^2 = |u|^2 + |v|^2 = 1$) l'élément $\begin{pmatrix} u & v \\ -\bar{v} & \bar{u} \end{pmatrix}$ de $SU(2)$. Si l'on identifie alors \mathbb{R}^3 à $\text{Span}\{i, j, k\}$ dans \mathbb{H}_Q , on vérifie que les q unitaires agissant par conjugaison $r \mapsto qrq^{-1}$ laissent ce \mathbb{R}^3 globalement invariant et y opèrent comme rotations. On constate alors que q et $-q$ donnent la même rotation ce qui montre que $SU(2)$ est un revêtement d'ordre 2 de $SO(3)$, la projection étant un morphisme de groupes de Lie. Comme $SU(2)$ est simplement connexe, c'est le revêtement universel de $SO(3)$ et les deux groupes ont la même algèbre de Lie. Mais il y a ici une subtilité liée au fait que les matrices de $SO(3)$ sont des matrices 3×3 à coefficients réels alors que celles de $SU(2)$ sont des matrices 2×2 à coefficients complexes.

Le lien entre $SU(2)$ et $SO(3)$ (entre un quaternion unitaire q et une rotation R_q) est un peu compliqué à expliciter. Bornons-nous à la simple correspondance suivante. Soit q_x unitaire tel que $y = z = 0$. On a donc $w^2 + x^2 = 1$ et l'on peut poser

$$w = \cos\left(\frac{\theta}{2}\right), \quad x = \sin\left(\frac{\theta}{2}\right).$$

18. Étant donnée l'anticommutativité on écrit les vecteurs de base $\{i, j, k\}$ à droite de leurs coefficients.

L'élément de $SU(2)$ associé à $q_x = \cos\left(\frac{\theta}{2}\right) + i \sin\left(\frac{\theta}{2}\right)$ est donc $u = \cos\left(\frac{\theta}{2}\right) + i \sin\left(\frac{\theta}{2}\right)$, $v = 0$, i.e.

$$U_x = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) + i \sin\left(\frac{\theta}{2}\right) & 0 \\ 0 & \cos\left(\frac{\theta}{2}\right) - i \sin\left(\frac{\theta}{2}\right) \end{pmatrix}$$

et les formules générales montrent que la rotation associée est

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

De même, si $x = z = 0$ et $w = \cos\left(\frac{\theta}{2}\right)$, $y = \sin\left(\frac{\theta}{2}\right)$, l'élément de $SU(2)$ associé à $q_y = \cos\left(\frac{\theta}{2}\right) + j \sin\left(\frac{\theta}{2}\right)$ est donc $u = \cos\left(\frac{\theta}{2}\right)$, $v = \sin\left(\frac{\theta}{2}\right)$, i.e.

$$U_y = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) & \sin\left(\frac{\theta}{2}\right) \\ -\sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{pmatrix}$$

et les formules générales montrent que la rotation associée est

$$R_y = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}.$$

Enfin, si $x = y = 0$ et $w = \cos\left(\frac{\theta}{2}\right)$, $z = \sin\left(\frac{\theta}{2}\right)$, l'élément de $SU(2)$ associé à $q_z = \cos\left(\frac{\theta}{2}\right) + k \sin\left(\frac{\theta}{2}\right)$ est donc $u = \cos\left(\frac{\theta}{2}\right)$, $v = i \sin\left(\frac{\theta}{2}\right)$, i.e.

$$U_z = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) & i \sin\left(\frac{\theta}{2}\right) \\ i \sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{pmatrix}$$

et les formules générales montrent que la rotation associée est

$$R_z = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

C'est le passage de $\frac{\theta}{2}$ à θ qui explique que $SU(2)$ est un revêtement d'ordre 2 de $SO(3)$. En effet quand on change $\frac{\theta}{2}$ en $\frac{\theta}{2} + \pi$, q change de signe, mais, comme θ change de 2π , la rotation R_q reste la même.

Les versions infinitésimales sont

$$\begin{aligned}\mathfrak{r}_x &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = e_2 , \\ \mathfrak{r}_y &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} = e_1 , \\ \mathfrak{r}_z &= \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = -e_3 .\end{aligned}$$

et

$$\begin{aligned}\mathbf{u}_x &= \frac{i}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \\ \mathbf{u}_y &= \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} , \\ \mathbf{u}_z &= \frac{i}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .\end{aligned}$$

$SU(2)$ et son algèbre de Lie $\mathfrak{su}(2)$ jouent un rôle déterminant en mécanique quantique car ils permettent de formaliser la notion de spin $\frac{1}{2}$ pour des particules comme l'électron¹⁹, ainsi que la notion d'isospin²⁰. Pour les particules de spin $\frac{1}{2}$ il existe pour chaque direction choisie par l'appareil de mesure deux états de spin "haut" et "bas" ou $+\frac{1}{2}$, $-\frac{1}{2}$. Dans les traités de physique, $SU(2)$ est défini comme le groupe des matrices 2×2 sur \mathbb{C} qui sont unitaires (groupe $U(2)$) et de déterminant 1. Dire que A est unitaire, c'est dire que $AA^* = A^*A = I$, où l'adjointe $A^* = \overline{A}^t$ de A est la transposée de sa conjuguée. Cela implique que A est normale (i.e. commute avec son adjointe), que $|\det A| = 1$, que A est diagonalisable et possède des vecteurs propres orthogonaux de valeurs propres de module 1. Si $A = \begin{pmatrix} u & v \\ w & t \end{pmatrix}$, $u, v, w, t \in$

19. La notion de spin a été introduite par Pauli en 1927 comme "moment cinétique intrinsèque" d'une particule. C'est une notion purement quantique.

20. L'isospin a été introduit par Heisenberg en 1932 pour expliquer que par rapport aux interactions nucléaires le proton et le neutron sont deux états d'une même particule, le nucléon, même s'ils sont différents pour les interactions électro-magnétiques puisqu'un seul (le proton) possède une charge électrique.

\mathbb{C} , $A^* = \begin{pmatrix} \bar{u} & \bar{w} \\ \bar{v} & \bar{t} \end{pmatrix}$ et les conditions sont donc

$$\begin{aligned} u\bar{u} + v\bar{v} &= 1, u\bar{w} + v\bar{t} = 0, w\bar{u} + t\bar{v} = 0, w\bar{w} + t\bar{t} = 1 \text{ (pour } U(2)) \\ ut - vw &= 1 \text{ (pour } SU(2)) \end{aligned}$$

conditions qui sont bien toutes satisfaites pour les matrices $\begin{pmatrix} u & v \\ -\bar{v} & \bar{u} \end{pmatrix}$ ci-dessus.²¹

La version infinitésimale des matrices unitaires de déterminant 1 conduit à définir l'algèbre de Lie $\mathfrak{su}(2)$ comme l'algèbre des matrices 2×2 sur \mathbb{C} anti-hermitiennes A de trace nulle, A étant anti-hermitienne si elle est égale à son adjointe. Cela signifie que A est de la forme $A = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & -\alpha \end{pmatrix}$ avec $\alpha \in \mathbb{R}$ ²² et est donc une combinaison linéaire à coefficients réels de la base standard qui, en physique, est celle des *matrices de Pauli*, $\sigma_x, \sigma_y, \sigma_z$, correspondant aux trois mesures d'un spin $\frac{1}{2}$ par rapport à un repère orthonormé (x, y, z) . Ces matrices de base sont données par

$$\sigma_x = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Elles ont comme valeurs propres ± 1 qui correspondent aux valeurs "haut"/"bas" du spin. Elles satisfont $\sigma_j^2 = I$, $\sigma_k \sigma_j = -\sigma_j \sigma_k$ si $j \neq k$ et donc $[\sigma_j, \sigma_k] = 2\sigma_j \sigma_k$, ainsi que $\sigma_j \sigma_k = i\sigma_l$, et donc $[\sigma_j, \sigma_k] = 2i\sigma_l$, si (j, k, l) est une permutation circulaire de $(1, 2, 3)$. En posant $e_j = -\frac{i}{2}\sigma_j$, on réobtient les relations de commutations de $\mathfrak{so}(3)$,

$$[e_1, e_2] = e_3, [e_2, e_3] = e_1, [e_3, e_1] = e_2.$$

Comme

$$\mathbf{u}_x = \frac{i}{2}\sigma_z, \mathbf{u}_y = \frac{i}{2}\sigma_y, \mathbf{u}_z = \frac{i}{2}\sigma_x$$

et que le système des $\{e_i\}$ est équivalent au système des $\{-e_i\}$, on retrouve la version de $\mathfrak{so}(3)$ vue plus haut.

21. On obtient bien ainsi toutes les matrices de $SU(2)$. En effet si $u = 0$, alors $|v| = 1$, $t = 0$, $|w| = 1$ et $-vw = 1$, ce qui implique $w = -\bar{v}$. Si $u \neq 0$, $ut - vw = 1$ implique $t = \frac{1+vw}{u}$ et $u\bar{w} + v\bar{t} = 0$ implique alors $\bar{u}u\bar{w} + v + v\bar{v}\bar{w} = 0$ et, comme $u\bar{u} + v\bar{v} = 1$, on doit avoir $\bar{w} + v = 0$, soit $w = -\bar{v}$. En partant de v au lieu de u , on trouve de même $t = \bar{u}$. Ainsi défini, $SU(2)$ est bien de dimension réelle 3 car u, v donnent 4 dimensions sur \mathbb{R} et $|\det A| = 1$ donne une condition sur \mathbb{R} .

22. Ces A forment bien une algèbre de dimension 3 sur \mathbb{R} puisque $\alpha \in \mathbb{R}$ et $\beta \in \mathbb{C} = \mathbb{R}^2$.

Pour rester dans le domaine réel, les physiciens utilisent parfois la base

$$\begin{aligned}\sigma^+ &= \frac{1}{2}(\sigma_x + i\sigma_y) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\ \sigma^- &= \frac{1}{2}(\sigma_x - i\sigma_y) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \\ \sigma_z &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.\end{aligned}$$

Les relations de commutation sont dans ce cas

$$[\sigma^+, \sigma^-] = \sigma_z, \quad [\sigma^+, \sigma_z] = -2\sigma^+, \quad [\sigma^-, \sigma_z] = 2\sigma^-.$$

Mais alors une combinaison linéaire à coefficients *réels* $\{a, b, c\}$ de $\{\sigma_x, \sigma_y, \sigma_z\}$ donne une combinaison de $\{\sigma^+, \sigma^-, \sigma_z\}$ à coefficients $\{a - ib, a + ib, c\}$ dont les deux premiers sont *complexes conjugués*. Les relations de $\{\sigma^+, \sigma^-, \sigma_z\}$ redonnent celles de $\mathfrak{sl}(2)$ avec

$$\sigma^+ = e_3, \quad \sigma^- = e_1, \quad \sigma_z = e_2.$$

Mais cela n'est pas aussi étrange que cela peut paraître. C'est un problème de relations entre bases et coefficients. Pour $\mathfrak{sl}(2)$ on a des combinaisons de $\{\sigma^+, \sigma^-, \sigma_z\}$ à coefficients *réels* $\{\alpha, \beta, \gamma\}$ ce qui correspond à des coefficients

$$\left\{ \frac{1}{2}(\alpha + \beta), \frac{i}{2}(\alpha - \beta), \gamma \right\},$$

dont le second est *imaginaire pur*, par rapport à la base $\{\sigma_x, \sigma_y, \sigma_z\}$. En revanche $\mathfrak{su}(2)$ correspond, répétons-le, à des combinaisons de $\{\sigma_x, \sigma_y, \sigma_z\}$ à coefficients *réels* $\{a, b, c\}$. En fait $\mathfrak{sl}(2)$ et $\mathfrak{su}(2)$ sont deux sous-algèbres de $\mathfrak{sl}(2, \mathbb{C})$ qui admettent toutes deux $\mathfrak{sl}(2, \mathbb{C})$ comme complexifiée.

Ce point mériterait d'être approfondi avec soin car $SL(2, \mathbb{C})$ est le revêtement universel du groupe de Lorentz restreint et il existe une intrication très subtile entre les représentations de ses divers sous-groupes. Mais cela dépasse notre propos.

3.4. Algèbre enveloppante et algèbre symétrique

Les algèbres standard \mathcal{A} possèdent un produit AB et donc automatiquement une structure d'algèbre de Lie dont le crochet est donné par $[A, B] = AB - BA$. Mais beaucoup d'algèbres de Lie sur des espaces vectoriels ne viennent pas d'un produit de vecteurs. Il est donc intéressant d'essayer de comprendre leur lien avec les algèbres standard. Nous allons profiter de la section précédente consacrée à la structure générale des algèbres de Lie (de dimension finie) pour dire quelques mots sur ce que l'on appelle leurs "algèbres enveloppantes" qui résolvent le problème. Il s'agit d'un sujet certes plus technique mais qui montre un aspect plus "mathématiques pures". Il comprend deux aspects, l'un algébrique et l'autre fonctionnel.

3.4.1. *Aspect algébrique.*

Étant donnée une algèbre de Lie \mathcal{G} sur \mathbb{K} (\mathbb{K} est le corps des scalaires, \mathbb{R} ou \mathbb{C}), on peut construire ce que l'on appelle son *algèbre enveloppante* $\mathfrak{U}(\mathcal{G})$, qui est l'algèbre standard des *polynômes non commutatifs* sur \mathcal{G} , à savoir celle des sommes symboliques de monômes symboliques non commutatifs $X_1 \cdots X_k$ de degrés k , $X_j \in \mathcal{G}$, où

- (i) les scalaires (polynômes de degré 0) commutent avec les X_j et
- (ii) les polynômes $X_1 X_2 - X_2 X_1$ de degré 2 sont identifiés aux polynômes de degré 1 que sont les commutateurs $[X_1, X_2] = X_3$.

Si $\{X_1, \dots, X_n\}$ est une base de \mathcal{G} , $\mathfrak{U}(\mathcal{G})$ est donc en général plus grosse que l'algèbre standard des polynômes commutatifs $\mathbb{K}[X_1, \dots, X_n]$ et plus petite que l'algèbre standard des polynômes non commutatifs dont elle est un quotient. L'algèbre $\mathfrak{U}(\mathcal{G})$ peut être rigoureusement définie en prenant l'algèbre tensorielle $\mathfrak{T}(\mathcal{G}) = \bigoplus_{k=0}^{k=\infty} \mathcal{G}^{\otimes k}$ et en la quotientant par l'idéal bilatère (i.e. à droite et à gauche) \mathcal{I} des

$$X_1 \otimes_{\mathbb{K}} X_2 - X_2 \otimes_{\mathbb{K}} X_1 - [X_1, X_2]$$

($\mathcal{G}^{\otimes 0} = \mathbb{K}$, $\mathcal{G}^{\otimes 1} = \mathcal{G}$, et $X_1 \otimes_{\mathbb{K}} X_2$ est la forme rigoureuse de $X_1 X_2$). \mathcal{G} s'injecte trivialement dans $\mathfrak{U}(\mathcal{G})$ par $J : \mathcal{G} = \mathcal{G}^{\otimes 1} \hookrightarrow \mathfrak{U}(\mathcal{G})$. La propriété essentielle et caractéristique de $\mathfrak{U}(\mathcal{G})$ est d'être *universelle* pour les morphismes d'algèbres de Lie φ de \mathcal{G} dans des \mathbb{K} -algèbres standard \mathcal{A} avec élément unité $1_{\mathcal{A}}$, c'est-à-dire des φ tels que

$$\varphi([X, Y]_{\mathcal{G}}) = [\varphi(X), \varphi(Y)]_{\mathcal{A}} = \varphi(X)\varphi(Y) - \varphi(Y)\varphi(X) ,$$

la première égalité exprimant que φ est un morphisme d'algèbres de Lie et la seconde découlant du fait que \mathcal{A} est une algèbre standard où $[A, B] = AB - BA$.

L'universalité signifie que tout morphisme d'algèbres de Lie φ , satisfaisant donc $\varphi([X, Y]_{\mathcal{G}}) = [\varphi(X), \varphi(Y)]_{\mathcal{A}}$, se factorise de façon unique par un morphisme $\tilde{\varphi}$ d'algèbres standard avec élément unité tel que $\varphi = \tilde{\varphi} \circ J$. Toute représentation de \mathcal{G} se prolonge donc automatiquement à une représentation de $\mathfrak{U}(\mathcal{G})$, et en particulier la représentation adjointe qui agit par $ad_X(U) = X \otimes U - U \otimes X$.

Évidemment, comme toute algèbre de polynômes, $\mathfrak{U}(\mathcal{G})$ est naturellement filtrée par les $\mathfrak{U}^n(\mathcal{G})$ venant des $\mathfrak{T}^n(\mathcal{G}) = \bigoplus_{k=0}^{k=n} \mathcal{G}^{\otimes k}$ (polynômes de degré $\leq n$). Un théorème classique, dit de *Poincaré-Birkhoff-Witt*²³, dit que si X_1, \dots, X_n est une base totalement ordonnée de \mathcal{G} , alors les monômes $\prod_{k=1}^{k=n} X_k^{m_k}$ forment une base de $\mathfrak{U}(\mathcal{G})$. Si $\mathcal{G} \simeq \mathbb{K}$ avec un générateur X , $\mathfrak{U}(\mathbb{K}) \simeq \mathbb{K}[X]$ car l'idéal \mathcal{I} par lequel on quotiente est nul et si $\mathcal{G} \simeq \mathbb{K}^n$ de générateurs X_1, \dots, X_n , $\mathfrak{U}(\mathbb{K}^n) \simeq \mathbb{K}[X_1, \dots, X_n]$ car l'idéal \mathcal{I} rend les variables X_j commutatives.

On notera que l'algèbre tensorielle $\mathfrak{T}(\mathcal{G})$ ne fait intervenir que la structure d'espace vectoriel de \mathcal{G} et non pas sa structure d'algèbre de Lie. Un autre quotient

23. Henri Poincaré 1900, Garrett Birkhoff 1937, Ernst Witt 1937.

de $\mathfrak{T}(\mathcal{G})$ qui ne fait intervenir que la structure d'espace vectoriel de \mathcal{G} est l'*algèbre symétrique* $\mathfrak{S}(\mathcal{G})$ quotient de $\mathfrak{T}(\mathcal{G})$ par l'idéal bilatère \mathcal{J} des $X_1 \otimes_{\mathbb{K}} X_2 - X_2 \otimes_{\mathbb{K}} X_1$.²⁴ $\mathfrak{S}(\mathcal{G})$ est la solution du problème universel des morphismes d'espaces vectoriels (et pas d'algèbres de Lie) φ de \mathcal{G} dans des \mathbb{K} -algèbres *commutatives* \mathcal{A} avec élément unité.

On a évidemment une application d'espaces vectoriels $\beta : \mathfrak{S}(\mathcal{G}) \rightarrow \mathfrak{U}(\mathcal{G})$ qui envoie simplement les $X^{\otimes k}$ sur les $X^{\otimes k}$. Pour les monômes $\prod_{j=1}^{j=k} X_j$ il faut symétriser et

$$\beta(X_1 \dots X_k) = \frac{1}{k!} \sum_{\sigma \in S_k} X_{\sigma(1)} \dots X_{\sigma(k)}$$

où les $\sigma \in S_k$ sont les permutations de $\{1, \dots, k\}$. Le théorème de Poincaré-Birkhoff-Witt dit essentiellement que β est un isomorphisme *d'espaces vectoriels* (mais pas du tout d'algèbres en général).

La représentation adjointe ad_X de \mathcal{G} dans \mathcal{G} (qui donne les dérivations internes de \mathcal{G}) se prolonge évidemment à l'algèbre tensorielle $\mathfrak{T}(\mathcal{G})$ et, par passage au quotient, à $\mathfrak{U}(\mathcal{G})$ et $\mathfrak{S}(\mathcal{G})$. L'application β commutant à ces dérivations, on peut donc considérer les sous-algèbres $\mathfrak{U}(\mathcal{G})_{inv}$ et $\mathfrak{S}(\mathcal{G})_{inv}$ des éléments *ad*-invariants.

Il existe tout un ensemble de relations subtiles entre $\mathfrak{U}(\mathcal{G})$ et $\mathfrak{S}(\mathcal{G})$ lorsque l'on tient compte de la structure d'algèbre de $\mathfrak{U}(\mathcal{G})$ et de l'exponentielle $\exp : \mathcal{G} \rightarrow G$.

3.4.2. Aspect fonctionnel.

Poussons un peu plus loin l'analyse. On sait que si \mathcal{G} est l'algèbre de Lie du groupe de Lie G , les éléments X de \mathcal{G} agissent sur $\mathcal{C}^\infty(G)$ (l'algèbre des fonctions f sur G à valeurs dans \mathbb{C}) comme des opérateurs différentiels G -invariants (à gauche) puisque ce sont en fait des champs de vecteurs tangents G -invariants (à gauche) sur G . Plus explicitement, on fait agir X sur $\mathcal{C}^\infty(G)$ en prenant la dérivée en $t = 0$ de la façon dont le groupe à 1-paramètre e^{tX} agit à droite sur les $f(x) \in \mathcal{C}^\infty(G)$. Cette action est automatiquement invariante à gauche et la formule est

$$\tilde{X}f(g) = \left. \frac{d}{dt} \right|_{t=0} f(g \exp(tX)) .$$

On montre que $\mathfrak{U}(\mathcal{G}_{\mathbb{C}})$ (où $\mathcal{G}_{\mathbb{C}}$ est la *complexifiée* de \mathcal{G} si $\mathbb{K} = \mathbb{R}$) est isomorphe à l'algèbre des opérateurs différentiels à coefficients \mathcal{C}^∞ G -invariants sur $\mathcal{C}^\infty(G)$ (la contrainte de G -invariance est une contrainte très forte). Ce fait est assez facile à intuitionner, mais l'intuition est piègeuse : dans des coordonnées x_1, \dots, x_n associées à une base X_1, \dots, X_n de $\mathcal{G} = T_e G$, les X_j conçus comme des opérateurs différentiels G -invariants deviennent les champs translatsés des $\frac{\partial}{\partial x_j}$ en 0 et un monôme $\prod_{k=1}^{k=n} X_k^{m_k}$ avec $\sum m_k = M$ devient le translatsé de l'opérateur différentiel $\frac{\partial^M}{\partial x_1^{m_1} \dots \partial x_n^{m_n}}$ en 0. La non-commutativité est essentielle, l'idéal \mathcal{I} correspondant à la non-commutativité

24. $X_1 \otimes_{\mathbb{K}} X_2 - X_2 \otimes_{\mathbb{K}} X_1$ et non pas $X_1 \otimes_{\mathbb{K}} X_2 - X_2 \otimes_{\mathbb{K}} X_1 - [X_1, X_2]$ comme pour $\mathfrak{U}(\mathcal{G})$.

des dérivées secondes $\frac{\partial^2}{\partial x \partial y} \neq \frac{\partial^2}{\partial y \partial x}$ (non-holonomie) lorsque $[X, Y] \neq 0$ (i.e. les translatés des $\frac{\partial}{\partial x_j}$ en 0 ne sont pas des $\frac{\partial}{\partial x_j}$ pour des coordonnées locales au voisinage de 0, cf. section 6.1 du chapitre 5). Le résultat n'est donc pas trivial.

L'exemple le plus élémentaire est celui de $\mathcal{G} = (\mathbb{R}, +)$, G étant alors $G = (\mathbb{R}^*, \times)$, le groupe multiplicatif (commutatif) des réels > 0 et $\mathcal{G}_{\mathbb{C}}$ étant tout simplement $(\mathbb{C}, +)$. Si $X \in \mathcal{G} = \mathbb{R}$ et si $f(x) \in \mathcal{C}^\infty(G = \mathbb{R}^*)$,

$$\tilde{X}f(x) = \left. \frac{d}{dt} \right|_{t=0} f(x \exp(tX)) = xX \frac{d}{dx} f(x)$$

qu'il faut voir comme une section $(X, x \frac{d}{dx} f(x))$ du fibré tangent TG de G qui, G étant parallélisable, est trivial et identifiable à $\mathcal{G} \times G$. Il est bien G -invariant puisque si $g \in \mathbb{R}^*$, $gx \frac{d}{d(gx)} = gx \frac{d}{dx} \frac{dx}{d(gx)} = x \frac{d}{dx}$. Considérons alors les opérateurs différentiels $D = \sum_{j=1}^{j=n} a_j(x) \frac{d^j}{dx^j}$. Ils sont G -invariants (à gauche) si et seulement si ils sont de la forme $\sum_{j=1}^{j=n} \left(a_j, x^j \frac{d^j}{dx^j} \right)$. Donc l'algèbre \mathcal{D}_G des D G -invariants est isomorphe à l'algèbre de polynômes $\mathbb{C}[T]$. Mais nous avons vu que $\mathfrak{U}(\mathbb{C}) \simeq \mathbb{C}[T]$.

Outre son interprétation comme l'algèbre des opérateurs différentiels à coefficients \mathcal{C}^∞ G -invariants sur $\mathcal{C}^\infty(G)$, on peut aussi interpréter $\mathfrak{U}(\mathcal{G})$ comme l'algèbre de convolution des distributions²⁵ sur G de support ponctuel e_G : si δ_{e_G} est la distribution de Dirac en e_G , toutes les dérivées $D\delta_{e_G}$ sont de support e_G . On peut également interpréter $\mathfrak{S}(\mathcal{G})$ comme l'algèbre des polynômes sur \mathcal{G}^* .

Le centre $\mathfrak{Z}(\mathcal{G})$ de l'algèbre enveloppante $\mathfrak{U}(\mathcal{G})$ est l'ensemble des $U \in \mathfrak{U}(\mathcal{G})$ qui commutent avec tous les autres $U \in \mathfrak{U}(\mathcal{G})$. Et comme les $X \in \mathcal{G}$ engendrent $\mathfrak{U}(\mathcal{G})$, il suffit que U commute avec tous les X , autrement dit que U appartienne à tous les noyaux des ad_X opérant sur $\mathfrak{U}(\mathcal{G})$. On a donc $\mathfrak{Z}(\mathcal{G}) = \mathfrak{U}(\mathcal{G})_{inv}$. Si alors ρ est une représentation²⁶ de G considérée aussi comme représentation de \mathcal{G} et si $Z \in \mathfrak{Z}(\mathcal{G})$ alors $\rho(Z)$ commute avec tous les $\rho(X)$, $X \in \mathcal{G}$. Si ρ est irréductible, alors, d'après le lemme de Schur,²⁷ $\rho(Z)$ agit de façon scalaire et l'on obtient ainsi un caractère du centre $\mathfrak{Z}(\mathcal{G})$. On l'appelle le "caractère infinitésimal" de ρ . Il généralise en dimension infinie à des opérateurs unitaires pas forcément traçables la définition du caractère $\chi_\rho(g) = \text{tr}(\rho(g))$ lorsque ρ est une représentation de dimension finie $\rho : G \rightarrow GL(V)$.

Lorsque G est nilpotent, un théorème de Jacques Dixmier montre alors que toute représentation unitaire irréductible ρ correspondant à une orbite co-adjointe \mathcal{O} (au

25. Nous reviendrons rigoureusement sur ce concept de distribution à la section 5 du chapitre 16.

26. La notion de représentation générale sera précisée dans la section 1 du chapitre 17. Jusqu'ici elle n'intervient que comme morphisme de groupes $\rho : G \rightarrow GL(V)$ de G dans le groupe des automorphismes $GL(V)$ d'un espace vectoriel.

27. Cf. la section 5.1 du chapitre 17.

sens que nous expliciterons plus bas à la section 6.4 du chapitre 17) peut être réalisée dans un Hilbert $L^2\left(\mathbb{R}^{\frac{1}{2}\dim\mathcal{O}}\right)$ de façon à ce que l'algèbre enveloppante $\mathfrak{U}(\mathcal{G})$ deviennent représentée par l'algèbre des opérateurs différentiels à coefficients *polynomiaux*. La démonstration se fait par une induction sur la dimension qui est rendue possible par la nilpotence (cf. Guichardet [235]). Et si l'on interprète à travers l'isomorphisme β^{-1} les éléments de $\mathfrak{U}(\mathcal{G})$ comme des éléments de $\mathfrak{S}(\mathcal{G})$ et donc comme des polynômes sur \mathcal{G}^* , alors les éléments du centre $\mathfrak{Z}(\mathcal{G})$ s'identifient aux polynômes invariants par la représentation co-adjointe Ad_G^* . Si P est un tel polynôme de $\mathfrak{Z}(\mathcal{G})$, la valeur du caractère infinitésimal de ρ en P est $P(i\mathcal{O})$ (qui est bien défini puisque P est Ad_G^* -invariant).

De façon plus générale, un théorème de Michel Duflo (1977) généralise le théorème de Dixmier à toutes les algèbres de Lie de dimension finie en affirmant que le centre $\mathfrak{Z}(\mathcal{G}) = \mathfrak{U}(\mathcal{G})_{inv}$ de $\mathfrak{U}(\mathcal{G})$ est isomorphe *comme algèbre* à l'algèbre $\mathfrak{S}(\mathcal{G})_{inv}$ des éléments de $\mathfrak{S}(\mathcal{G})$ invariants, c'est-à-dire à l'algèbre des polynômes sur \mathcal{G}^* qui sont invariants. Ce résultat de Dixmier pour les \mathcal{G} nilpotentes avait été également montré pour les \mathcal{G} semi-simples par Harish-Chandra et Claude Chevalley. Le théorème de Duflo le démontre en général. Tous ces résultats sont techniques et hautement non triviaux.

Un outil essentiel de la preuve est l'utilisation de la dérivée de l'exponentielle $\exp : \mathcal{G} \rightarrow G$ non seulement en 0 où elle est l'identité mais également, et c'est une toute autre affaire, pour les $X \neq 0$. Nous avons vu à la section 9 du chapitre 5 que l'application tangente $T_0 \exp$ en $0 \in \mathcal{G}$ est l'identité entre $T_0\mathcal{G} \simeq \mathcal{G}$ et $T_0G = \mathcal{G}$. Le calcul de $T_X \exp$ en $X \neq 0$ est beaucoup plus difficile. Pour le mener à bien, on considère des chemins différentiables $X(t)$ dans \mathcal{G} dont les dérivées $X'(t)$ vont donner des vecteurs tangents de $T_{X(t)}\mathcal{G}$ (identifiables à des vecteurs de \mathcal{G} car \mathcal{G} est un espace vectoriel). En appliquant \exp , on obtient un chemin différentiable $\exp(X(t)) = g(t)$ dans G et l'on veut calculer la dérivée $g'(t) \in T_{g(t)}G$, $T_{g(t)}G$ étant $T_0G = \mathcal{G}$ translaté (à gauche) par $g(t)$. La non-commutativité fait que l'on ne peut pas appliquer la formule classique

$$g'(t) = (\exp(X(t)))' = X'(t) \exp(X(t))$$

car, quand on bouge un peu t , $X(t)$ et $X(t+dt)$ ne commutent pas en général. Pour calculer correctement $g'(t)$, on utilise alors une méthode standard consistant à introduire une fonction appropriée dépendant non seulement de t mais d'une *nouvelle variable* s et satisfaisant une équation différentielle que l'on sait intégrer et dont l'intégrale donne la formule cherchée pour $s = 1$. En fait, on utilise la remarque qu'en remplaçant les X par des sX on peut toujours *déformer* une algèbre de Lie de façon à la rendre triviale pour $s = 0$. C'est un peu comme une "homotopie" de sa structure.

Ici, la fonction astucieuse est la déformation

$$\Phi(s, t) = \exp(-sX(t)) \frac{\partial}{\partial t} (\exp(sX(t))) .$$

On a

$$\begin{aligned} \Phi(0, t) &= 1 \frac{\partial}{\partial t} (1) = 0 , \\ \Phi(1, t) &= \exp(-X(t)) \frac{\partial}{\partial t} (\exp(X(t))) \end{aligned}$$

et donc

$$g'(t) = \exp(X(t)) \Phi(1, t) .$$

Calculons alors $\frac{\partial}{\partial s} (\Phi(s, t))$. On obtient formellement

$$\begin{aligned} \frac{\partial}{\partial s} (\Phi(s, t)) &= -X(t) \exp(-sX(t)) \frac{\partial}{\partial t} (\exp(sX(t))) \\ &\quad + \exp(-sX(t)) \left(\frac{\partial}{\partial t} ((X(t) \exp(sX(t)))) \right) \\ &= -X(t) \exp(-sX(t)) \frac{\partial}{\partial t} (\exp(sX(t))) \\ &\quad + \exp(-sX(t)) \left(X'(t) \exp(sX(t)) + X(t) \frac{\partial}{\partial t} (\exp(sX(t))) \right) . \end{aligned}$$

La fonction Φ est astucieuse parce que dans $\frac{\partial}{\partial s} (\Phi(s, t))$ les termes contenant

$$\frac{\partial}{\partial t} (\exp(sX(t)))$$

(que l'on cherche à calculer) s'éliminent et il ne reste que

$$\begin{aligned} \frac{\partial}{\partial s} (\Phi(s, t)) &= \exp(-sX(t)) X'(t) \exp(sX(t)) \\ &= Ad_{\exp(-sX)} X'(t) = \exp(-ad_{sX}) X'(t) . \end{aligned}$$

On peut alors faire un développement en série formelle de $\exp(-ad_{sX})$ et comme $\Phi(0, t) = 0$, on obtient

$$\begin{aligned} \Phi(1, t) &= \int_0^1 \frac{\partial}{\partial s} (\Phi(s, t)) ds = \int_0^1 \left(\sum_{k \geq 0} (-1)^k \frac{s^k}{k!} (ad_{sX})^k X'(t) \right) ds \\ &= \sum_{k \geq 0} (-1)^k \frac{1}{(k+1)!} (ad_{sX})^k X'(t) = \frac{1 - \exp(-ad_X)}{ad_X} X'(t) \end{aligned}$$

et donc

$$\frac{d}{dt}(\exp(X(t))) = \exp(X(t)) \frac{1 - \exp(-ad_X)}{ad_X} X'(t)$$

où l'expression $\frac{1 - \exp(-ad_X)}{ad_X}$ dénote simplement une série formelle.

Si G est nilpotent, tous les ad_X sont nilpotents d'après le théorème d'Engel et les séries sont en fait des polynômes ne posant plus de problèmes de convergence.

Si l'on note $J(X)$ le déterminant jacobien de \exp en X , on a donc

$$J(X) = \det \left(\frac{1 - \exp(-ad_X)}{ad_X} \right).$$

Le théorème de Duflo explique ce que devient le produit de l'algèbre $\mathfrak{Z}(\mathcal{G}) = \mathfrak{U}(\mathcal{G})_{inv}$ une fois qu'on l'a transférée à $\mathfrak{S}(\mathcal{G})_{inv}$ au moyen de l'inverse β^{-1} de l'isomorphisme d'espaces vectoriels $\beta : \mathfrak{S}(\mathcal{G}) \rightarrow \mathfrak{U}(\mathcal{G})$. Il dit que si l'on interprète les éléments P, Q de $\mathfrak{S}(\mathcal{G})_{inv}$ comme des distributions de support $0 \in \mathcal{G}$ *invariantes* (munies du produit de convolution $P * Q$) et que l'on traduit J dans cette interprétation, alors β qui ne satisfait pas du tout $\beta(PQ) = \beta(P)\beta(Q)$ (c'est un isomorphisme *d'espaces vectoriels* et pas d'algèbres) satisfait en revanche un isomorphisme *d'algèbres* entre $\mathfrak{S}(\mathcal{G})_{inv}$ et $\mathfrak{U}(\mathcal{G})_{inv}$ sous la forme

$$\beta \left(J^{\frac{1}{2}}(P * Q) \right) = \beta \left(J^{\frac{1}{2}}(P) \right) * \beta \left(J^{\frac{1}{2}}(Q) \right).$$

3.5. Retour sur la formule de Baker-Campbell-Hausdorff

Nous avons plusieurs fois utilisé la formule fondamentale de Baker-Campbell-Hausdorff (BCH, sections 5.1 et 5.3 du chapitre 5) et nous l'utiliserons encore (par exemple à la section 2.2 du chapitre 14, aux sections 1.3, 1.3.2 et 5.2 du chapitre 17). Élaborée après des travaux de Schur (1890) par John Edward Campbell (1897), Henri Poincaré (1899), Henry Frederick Baker (1902) et Felix Hausdorff (1906), elle dit que, étant donné le difféomorphisme \exp de $(\mathcal{G}, 0)$ dans (G, e_G) (G connexe et simplement connexe de dimension finie) on a, pour tous $X, Y \in \mathcal{G}$,

$$e^X e^Y = e^{Z(X, Y)}$$

où $Z(X, Y)$ est une série convergente de polynômes de Lie à valeurs dans \mathcal{G} définie de façon universelle par des crochets itérés.

Si l'on est dans la situation où l'on peut représenter les éléments de \mathcal{G} et de G par des *matrices* (ce qui nous suffit ici et est d'ailleurs toujours le cas pour les \mathcal{G} de dimension finie sur \mathbb{R} ou \mathbb{C} d'après le classique théorème d'Ado) on peut alors utiliser, en tenant compte de la non-commutativité et en faisant attention aux problèmes de convergence des séries formelles que l'on manipule, l'analogie des séries classiques

$$e^X = 1 + X + \frac{X^2}{2} + \cdots + \frac{X^n}{n!} + \cdots$$

et si $x \in \mathbb{R}^*$,

$$\ln(1+x) = x - \frac{x^2}{2} + \dots (-1)^{n+1} \frac{x^n}{n} + \dots$$

ainsi que la formule $[X, Y] = XY - YX$. En calculant $e^X e^Y = e^{Z(X,Y)}$ et en écrivant $Z(X, Y) = \ln(e^X e^Y)$, on obtient facilement une expression formelle pour $Z(X, Y)$:

$$\begin{aligned} e^X e^Y &= \left(1 + \sum_{\mu \geq 1} \frac{X^\mu}{\mu!}\right) \left(1 + \sum_{\nu \geq 1} \frac{Y^\nu}{\nu!}\right) = 1 + \sum_{\mu+\nu \geq 1} \frac{X^\mu Y^\nu}{\mu! \nu!} = 1 + S \\ Z(X, Y) &= \ln(e^X e^Y) = \sum_{n \geq 1} \frac{(-1)^{n+1}}{n} S^n = \sum_{n \geq 1} \frac{(-1)^{n+1}}{n} \left(\sum_{\mu+\nu \geq 1} \frac{X^\mu Y^\nu}{\mu! \nu!}\right)^n \\ &= \sum_{n \geq 1} \frac{(-1)^{n+1}}{n} \left(\sum_{\mu_j+\nu_j \geq 1, j=1, \dots, n} \left(\prod_j \frac{X^{\mu_j} Y^{\nu_j}}{\mu_j! \nu_j!}\right)\right). \end{aligned}$$

Mais ce n'est pas du tout la formule BCH qui dit qu'il existe un développement en série ne faisant intervenir que des *commutateurs* itérés, ce qui implique que le produit de G qui n'est que C^∞ est calculable uniquement à partir de la structure *algébrique* des crochets de Lie.

En 1947, Eugene Dynkin a proposé une telle formule. Notons

$$\left[\prod_{j=1}^{j=n} X^{\mu_j} Y^{\nu_j} \right]$$

le multicrochet enchâssé de gauche à droite²⁸ enchâssant successivement μ_1 fois X puis ν_1 fois Y , puis \dots , puis μ_n fois X et enfin ν_n fois Y .²⁹ La remarquable formule, plus compliquée, de Dynkin dit que

$$Z(X, Y) = X + Y + \sum_{n \geq 2} \frac{(-1)^{n+1}}{n} \left(\sum_{\mu_j+\nu_j \geq 1, j=1, \dots, n} \left(\frac{[\prod_j X^{\mu_j} Y^{\nu_j}]}{(\sum_j (\mu_j + \nu_j)) \prod_j \mu_j! \nu_j!}\right)\right).$$

Le début du développement en série est

$$\begin{aligned} Z(X, Y) &= X + Y + \frac{1}{2} [X, Y] + \frac{1}{12} ([X, [X, Y]] + [Y, [Y, X]]) \\ &\quad - \frac{1}{48} ([Y, [X, [X, Y]]] + [X, [Y, [X, Y]]]) + \dots \end{aligned}$$

28. C'est-à-dire de type $[X, Y], [X, [Y, Z]], [X, [Y, [Z, T]]], \dots$

29. Si $\nu_n \geq 2$, le dernier enchâssement est $[Y, Y]$ et le multicrochet est nul. Si $\nu_n = 0$ et $\mu_n \geq 2$, le dernier enchâssement est $[X, X]$ et le multicrochet est également nul.

(le début $Z(X, Y) = X + Y$ correspondant au cas abélien : tous les commutateurs sont nuls) et comme pour le groupe nilpotent de Heisenberg X et Y commutent avec leur crochet, nous avons la formule $Z(X, Y) = X + Y + \frac{1}{2}[X, Y]$ que nous avons utilisée. La complexité de $Z(X, Y)$ est considérable.

Autour de la formule BCH, de très nombreux calculs de séries formelles ont été développés dans des cas particuliers où les crochets ont une forme particulière, avec des considérations de convergence sur des voisinages assez petits de l'origine $0 \in \mathcal{G}$. Ils sont d'une grande virtuosité et rappellent la grande époque d'Euler. Si par exemple $[X, Y] = sY$,

$$Z(X, Y) = X + Y + \frac{se^s - e^s + 1}{s(e^s - 1)} [X, Y] = X + \frac{s}{1 - e^{-s}} Y .$$

Il existe une double *formule intégrale* de BCH qui exprime $Z(X, Y)$ comme une intégrale de séries des opérateurs $Ad_{\exp(X)} = \exp(ad_X)$ et $Ad_{\exp(Y)} = \exp(ad_Y)$ appliquée soit à X soit à Y . Elle dit que

$$\begin{aligned} Z(X, Y) &= X + Y + \int_0^1 \left(\sum_{n \geq 1} \frac{(Id - \exp(tad_X) \exp(ad_Y))^n}{n+1} (X) \right) dt \\ &= X + Y - \int_0^1 \left(\sum_{n \geq 1} \frac{(Id - \exp(ad_X) \exp(tad_Y))^n}{n(n+1)} (Y) \right) dt . \end{aligned}$$

Cela peut se voir à partir de la formule

$$\frac{d}{dt} (\exp(X(t))) = \exp(X(t)) \frac{1 - \exp(-ad_X)}{ad_X} X'(t)$$

de la section précédente appliquée à $\tilde{Z}(t) = Z(X, tY)$, $\tilde{Z}(1)$ donnant $Z(X, Y)$. On obtient

$$\exp(-\tilde{Z}(t)) \frac{d}{dt} (\exp(\tilde{Z}(t))) = \frac{1 - \exp(-ad_{\tilde{Z}})}{ad_{\tilde{Z}}} \tilde{Z}'(t) .$$

Mais, vu la forme spéciale de $\tilde{Z}(t)$,

$$\exp(-\tilde{Z}(t)) \frac{d}{dt} (\exp(\tilde{Z}(t))) = Y$$

et par conséquent

$$\tilde{Z}'(t) = \frac{ad_{\tilde{Z}}}{1 - \exp(-ad_{\tilde{Z}})} Y .$$

Donc, toujours en termes de séries formelles,

$$\tilde{Z}(1) = Z(X, Y) = \int_0^1 \left(\frac{ad_{Z(X, tY)}}{1 - \exp(-ad_{Z(X, tY)})} Y \right) dt .$$

Mais en posant $\exp(ad_{Z(X,tY)}) = w$, $\frac{ad_{Z(X,tY)}}{1-\exp(-ad_{Z(X,tY)})}$ se développe en

$$\frac{\log(w)}{1-\frac{1}{w}} = \frac{w \log(w)}{w-1} = 1 - \sum_{n \geq 1} \frac{(1-w)^n}{n(n+1)}$$

et comme $w = \exp(ad_{Z(X,tY)}) = \exp(ad_x) \exp(tad_Y)$ on obtient la seconde des formules intégrales.

Là encore, si G est nilpotent les ad_X sont nilpotents d'après le théorème d'Engel et les séries en ad_X deviennent des polynômes.

De nombreux travaux plus approfondis ont été consacrés à BCH, de ceux de Masaki Kashiwara et Michèle Vergne (1978) jusqu'à ceux plus récents de Maxim Kontsevich (2003). La conjecture de Kashiwara-Vergne a été démontrée en 2005 par Anton Alekseev et Eckhard Meinrenken [13] (cf. le séminaire Bourbaki de Charles Torossian [537]).

4. Métriques invariantes sur les groupes de Lie (prélude)

À partir du prochain chapitre nous allons introduire des modèles utilisant des géodésiques pour différentes métriques sur nos groupes de Lie de référence, métriques non invariantes ou invariantes, riemanniennes ou sous-riemanniennes. Le contexte de cette problématique est très vaste et nous ne pourrions pas l'aborder dans sa généralité. À titre de prélude, signalons toutefois qu'il existe de nombreuses relations entre la structure d'un groupe de Lie G et les métriques riemanniennes *L-invariantes* que l'on peut y définir. Ces propriétés "mixent" ce qui relève de la structure de G (sous-groupes à 1 paramètre et exponentielle, représentations adjointes et co-adjointes, nilpotence, résolubilité, etc.) et ce qui relève de la métrique (connexion de Levi-Civita, courbure de Riemann, courbure de Ricci, courbure scalaire, géodésiques, etc.).

4.1. Métriques riemanniennes *L-invariantes* en D3 (Milnor)

Nous supposons dans cette section que G est un groupe de Lie *connexe de dimension 3*. Une excellente introduction à ce sujet est l'article classique "Curvatures of Left Invariant Metrics on Lie Groups" (1976) de John Milnor³⁰ [368]. Nous allons le résumer en utilisant les concepts de la section 6. du chapitre 6.

30. Né en 1931, Médaille Fields 1962, Prix Abel 2011, John Milnor est l'un des grands spécialistes de topologie différentielle (théorie de Morse, cobordisme) et de la théorie des singularités dans le domaine complexe. Il a montré qu'il existe des sphères "exotiques" (15 en dimension 7) qui sont homéomorphes à la sphère standard sans lui être pour autant difféomorphes.

Soit \check{g} une métrique L -invariante (i.e. G -invariante à gauche) sur G .³¹ Elle provient d'une métrique sur \mathcal{G} , encore notée \check{g} , qui est L -translatée sur G . Tous les éléments métriques faisant intervenir des champs de vecteurs L -invariants deviennent alors interprétables comme des éléments métriques sur l'algèbre de Lie \mathcal{G} et on peut étudier leur interaction avec la structure d'algèbre de Lie.³²

Nous noterons comme d'habitude $\langle X, Y \rangle$ le produit scalaire $\check{g}(X, Y)$ défini par \check{g} . La métrique \check{g} permet de définir les *adjoints* A^* des opérateurs linéaires A de \mathcal{G} dans \mathcal{G} par la formule $\langle A(Y), Z \rangle = \langle Y, A^*(Z) \rangle$. A est auto-adjoint si $A^* = A$ et anti-auto-adjoint si $A^* = -A$.³³

La connexion de Levi-Civita ∇ opère maintenant aussi sur \mathcal{G} , $\nabla_X Y$ étant un élément de \mathcal{G} pour $X, Y \in \mathcal{G}$. ∇_X est linéaire en X et en Y , i.e. $\nabla_{\alpha X} Y = \alpha \nabla_X Y$ et $\nabla_X \beta Y = \beta \nabla_X Y$.³⁴ L'opérateur ∇ est une application linéaire de \mathcal{G} dans $\text{End}(\mathcal{G})$ mais *n'est pas* a priori un morphisme d'algèbres de Lie (cela dépend de la courbure, voir plus bas).

On a (nullité de la torsion de la connexion de Levi-Civita)

$$\nabla_X Y - \nabla_Y X = [X, Y] .$$

On a par ailleurs (cf. chapitre 6, section 6)

$$\partial_X \check{g}(Y, Z) = \check{g}(\nabla_X Y, Z) + \check{g}(Y, \nabla_X Z) ,$$

soit, puisque $\partial_X \check{g}(Y, Z) = 0$ à cause de la L -invariance, la formule

$$\langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle = 0 ,$$

qui montre que ∇_X est *anti-auto-adjoint*.³⁵ Ces deux formules permettent de calculer facilement la connexion en termes du crochet de Lie et du produit scalaire. En effet, d'un côté,

$$\begin{cases} \langle \nabla_X Y, Z \rangle - \langle \nabla_Y X, Z \rangle = \langle [X, Y], Z \rangle \\ \langle \nabla_Y Z, X \rangle - \langle \nabla_Z Y, X \rangle = \langle [Y, Z], X \rangle \\ \langle \nabla_Z X, Y \rangle - \langle \nabla_X Z, Y \rangle = \langle [Z, X], Y \rangle \end{cases}$$

31. Nous notons ici la métrique \check{g} pour la distinguer des éléments de G et d'une algèbre de Lie \mathcal{G} .

32. Ces éléments dépendent de \mathcal{G} et de \check{g} mais nous ne l'indiquerons pas dans les notations sauf lorsque plusieurs \mathcal{G} et/ou \check{g} interviendront.

33. Il s'agit d'adjoints au sens de la métrique, à ne pas confondre avec la représentation adjointe $ad_X(Y) = [X, Y]$.

34. Pour une métrique quelconque on a la règle de Leibniz $\nabla_X(fY) = f\nabla_X Y + df(X)Y$ mais si f L -translate β , $df = 0$.

35. On a donc $\langle \nabla_X Y, Z \rangle = -\langle Y, \nabla_X Z \rangle$, mais on a $\langle \nabla_X Y, Z \rangle = \langle Z, \nabla_X Y \rangle$ à cause de la symétrie du produit scalaire.

et, d'un autre côté,

$$\begin{cases} \langle \nabla_X Y, Z \rangle = -\langle Y, \nabla_X Z \rangle = -\langle \nabla_X Z, Y \rangle \\ \langle \nabla_Y Z, X \rangle = -\langle Z, \nabla_Y X \rangle = -\langle \nabla_Y X, Z \rangle \\ \langle \nabla_Z X, Y \rangle = -\langle X, \nabla_Z Y \rangle = -\langle \nabla_Z Y, X \rangle \end{cases}$$

Donc, si l'on additionne la première et la troisième ligne en leur retranchant la seconde, on obtient $2 \langle \nabla_X Y, Z \rangle$, les autres termes s'éliminant deux à deux. D'où la formule de base

$$\langle \nabla_X Y, Z \rangle = \frac{1}{2} (\langle [X, Y], Z \rangle - \langle [Y, Z], X \rangle + \langle [Z, X], Y \rangle) . \quad ((F\nabla))$$

On peut alors relier explicitement les constantes de structure $c_{\mu\nu}{}^\rho$ de \mathcal{G} (chapitre 6) avec les symboles de Christoffel $\Gamma_{\mu\nu}{}^\rho$ en considérant des bases *orthonormées* pour \check{g} . On a en effet pour toute base $\{\mathbf{e}_\mu\}$ de \mathcal{G} $[\mathbf{e}_\mu, \mathbf{e}_\nu] = c_{\mu\nu}{}^\rho \mathbf{e}_\rho$. Mais si la base est orthonormée, alors $\langle [\mathbf{e}_\mu, \mathbf{e}_\nu], \mathbf{e}_\rho \rangle = c_{\mu\nu}{}^\rho$ et donc, si on note $\nabla_\mu = \nabla_{\mathbf{e}_\mu}$,

$$\begin{aligned} \langle \nabla_\mu \mathbf{e}_\nu, \mathbf{e}_\rho \rangle &= \frac{1}{2} (\langle [\mathbf{e}_\mu, \mathbf{e}_\nu], \mathbf{e}_\rho \rangle - \langle [\mathbf{e}_\nu, \mathbf{e}_\rho], \mathbf{e}_\mu \rangle + \langle [\mathbf{e}_\rho, \mathbf{e}_\mu], \mathbf{e}_\nu \rangle) = \frac{1}{2} (c_{\mu\nu}{}^\rho - c_{\nu\rho}{}^\mu + c_{\rho\mu}{}^\nu) \\ \nabla_\mu \mathbf{e}_\nu &= \frac{1}{2} (c_{\mu\nu}{}^\rho - c_{\nu\rho}{}^\mu + c_{\rho\mu}{}^\nu) \mathbf{e}_\rho = \Gamma_{\mu\nu}{}^\rho \mathbf{e}_\rho \\ \Gamma_{\mu\nu}{}^\rho &= \frac{1}{2} (c_{\mu\nu}{}^\rho - c_{\nu\rho}{}^\mu + c_{\rho\mu}{}^\nu) . \end{aligned}$$

formule à comparer à celle du chapitre 6, section 6 donnant les $\Gamma_{\mu\nu}{}^\rho$ en fonction des dérivées du tenseur métrique. On voit ainsi que le choix d'une base comme orthonormée définit facilement la métrique L -invariante associée.

Milnor définit la courbure de \check{g} par

$$R_{X,Y} = \nabla_{[X,Y]} - [\nabla_X, \nabla_Y] .$$

Il s'agit de l'*opposé* de la formule habituelle, mais nous gardons sa convention car, comme nous allons le voir, cela ne change rien aux courbures sectionnelles qui sont les seules utilisées. Par définition $R_{X,Y}$ est antisymétrique comme 2-forme à valeurs dans $\text{End}(\mathcal{G})$. On notera que le fait que ∇ soit un morphisme *d'algèbres de Lie* de \mathcal{G} dans $\text{End}(\mathcal{G})$ signifie que $\nabla_{[X,Y]} = [\nabla_X, \nabla_Y]$, autrement dit que $R_{X,Y} \equiv 0$. À partir de $R_{X,Y}$ on peut construire les produits scalaires $\langle R_{X,Y}(U), V \rangle$ qui donnent les quadri-indices du tenseur de courbure. La *courbure sectionnelle* du plan $\text{Span}(X, Y)$ pour $\{X, Y\}$ orthonormés est alors donnée par la forme biquadratique $\mathfrak{K}(X, Y) = \langle R_{X,Y}(X), Y \rangle$. Évidemment $\mathfrak{K}(X, X) = 0$ puisque $R_{X,Y}$ est anti-symétrique. La métrique \check{g} est dite *plate* si $\mathfrak{K}(X, Y) \equiv 0$.

Remarque. La définition est bien la définition habituelle. Pour $\{X, Y\}$ une base de $\text{Span}(X, Y)$, la définition habituelle de $\mathfrak{K}(X, Y)$ est $\mathfrak{K}(X, Y) = \frac{\langle R_{X,Y}(Y), X \rangle}{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2}$ avec $R = -R^{\text{Milnor}}$ et donc $\mathfrak{K}(X, Y) = \langle R_{X,Y}(Y), X \rangle$ pour une base orthonormale.

Il est facile de vérifier que le passage de $\langle R_{X,Y}^{\text{Milnor}}(X), Y \rangle$ à $\langle R_{X,Y}(Y), X \rangle$ corrige l'inversion de signe. Il suffit de remarquer que

$$\begin{aligned} \langle \nabla_U(\nabla_X(Y)), Z \rangle &= -\langle \nabla_X(Y), \nabla_U(Z) \rangle = -\langle \nabla_U(Z), \nabla_X(Y) \rangle \\ &= \langle \nabla_X(\nabla_U(Z)), Y \rangle \end{aligned}$$

et de comparer les deux définitions. □

Dans ce qui suit, nous allons supposer que G est *connexe de dimension 3* et, afin d'avoir les mêmes notations (un peu plus simples) que dans la section 3.3, nous noterons $\{e_1, e_2, e_3\}$ la base de \mathcal{G} choisie comme base orthonormée pour définir \check{g} et c_{ij}^k les constantes de structures. Il est alors facile de calculer les courbures sectionnelles $\mathfrak{K}(e_i, e_j)$ à partir des c_{ij}^k de cette base : ces constantes définissent la structure d'algèbre de Lie mais, dans la mesure où la base est par hypothèse orthonormée, elles encodent aussi la métrique. On a puisque $[e_i, e_j] = c_{ij}^k e_k$ (somme d'Einstein) et vu les propriétés de linéarité de ∇ :

$$\begin{aligned} \mathfrak{K}(e_i, e_j) &= \langle R_{e_i e_j}(e_i), e_j \rangle = \langle \nabla_{[e_i, e_j]} e_i - [\nabla_{e_i}, \nabla_{e_j}] e_i, e_j \rangle \\ &= \langle \nabla_{[e_i, e_j]} e_i - [\nabla_{e_i}, \nabla_{e_j}] e_i, e_j \rangle \\ &= \langle c_{ij}^k \nabla_{e_k} e_i, e_j \rangle - \langle \nabla_{e_i}(\nabla_{e_j} e_i), e_j \rangle + \langle \nabla_{e_j}(\nabla_{e_i} e_i), e_j \rangle . \end{aligned}$$

Mais d'après la formule $\nabla_\mu e_\nu = \frac{1}{2}(c_{\mu\nu}^\rho - c_{\nu\rho}^\mu + c_{\rho\mu}^\nu) e_\rho$, on a

$$\begin{aligned} \langle c_{ij}^k \nabla_{e_k} e_i, e_j \rangle &= \frac{1}{2} \langle c_{ij}^k (c_{ki}^\rho - c_{i\rho}^k + c_{\rho k}^i) e_\rho, e_j \rangle = \frac{1}{2} c_{ij}^k (c_{ki}^j - c_{ij}^k + c_{jk}^i) \\ - \langle \nabla_{e_i}(\nabla_{e_j} e_i), e_j \rangle &= -\frac{1}{2} \langle \nabla_{e_i} ((c_{ji}^\rho - c_{i\rho}^j + c_{\rho j}^i) e_\rho), e_j \rangle \\ &= -\frac{1}{2} (c_{ji}^\rho - c_{i\rho}^j + c_{\rho j}^i) \langle \nabla_{e_i} e_\rho, e_j \rangle \\ &= -\frac{1}{4} (c_{ji}^\rho - c_{i\rho}^j + c_{\rho j}^i) (c_{i\rho}^\sigma - c_{\rho\sigma}^i + c_{\sigma i}^\rho) \langle e_\sigma, e_j \rangle \\ &= -\frac{1}{4} (c_{ji}^\rho - c_{i\rho}^j + c_{\rho j}^i) (c_{i\rho}^j - c_{\rho j}^i + c_{ji}^\rho) \\ \langle \nabla_{e_j}(\nabla_{e_i} e_i), e_j \rangle &= \frac{1}{4} (c_{ii}^\rho - c_{i\rho}^i + c_{\rho i}^i) (c_{j\rho}^j - c_{\rho j}^j + c_{jj}^\rho) . \end{aligned}$$

Mais les c_{ij}^k sont antisymétriques en i, j et donc

$$c_{ii}^\rho = c_{jj}^\rho = 0, \quad -c_{i\rho}^i + c_{\rho i}^i = 2c_{\rho i}^i, \quad c_{j\rho}^j - c_{\rho j}^j = -2c_{\rho j}^j$$

et donc

$$\langle \nabla_{e_j}(\nabla_{e_i} e_i), e_j \rangle = -c_{\rho i}^i c_{\rho j}^j .$$

En faisant la somme des termes, on trouve les courbures sectionnelles

$$\mathfrak{K}(e_i, e_j) = \frac{1}{2}c_{ij}^k (c_{ki}^j - c_{ij}^k + c_{jk}^i) - \frac{1}{4}(c_{ji}^\rho - c_{i\rho}^j + c_{\rho j}^i) (c_{i\rho}^j - c_{\rho j}^i + c_{ji}^\rho) - c_{\rho i}^i c_{\rho j}^j \quad (\text{K})$$

et on vérifie à la main la symétrie $\mathfrak{K}(e_i, e_j) = \mathfrak{K}(e_j, e_i)$.

On peut alors considérer la *courbure de Ricci*

$$\mathfrak{R}(X) = \sum_j \mathfrak{K}(X, e_j) = \sum_j \langle R_{X, e_j}(X), e_j \rangle$$

qui est une forme quadratique représentant une sorte de “moyenne” des courbures sectionnelles des plans contenant X . On peut l’identifier à

$$-\sum_j \langle R_{X, e_j}(e_j), X \rangle = \langle \mathfrak{R}^*(X), X \rangle$$

avec

$$\mathfrak{R}^*(X) = -\sum_j R_{X, e_j}(e_j) .$$

$\mathfrak{R}^*(X) \in \text{End}(\mathcal{G})$ et on peut en faire la décomposition spectrale. Ses valeurs propres sont les courbures de Ricci *principales*. Si la base orthonormée $\{e_j\}$ est choisie de façon à diagonaliser \mathfrak{R}^* , alors les $\mathfrak{R}(e_j)$ sont les courbures de Ricci principales et, si $X = \sum_j x_j e_j$, on a

$$\mathfrak{R}(X) = \sum_j x_j^2 \mathfrak{R}(e_j) .$$

Les signes des $\mathfrak{R}(e_j)$ définissent la signature de la forme quadratique \mathfrak{R} .

Quant à la courbure scalaire \mathfrak{r} , elle est la somme des courbures de Ricci principales.

Il existe des liens étroits entre le signe de $\mathfrak{K}(X, Y)$ et la représentation adjointe de G car on peut considérer les adjoints des opérateurs de la représentation adjointe $ad_X(Y) = [X, Y]$ puisque l’on dispose d’une métrique.³⁶ La définition est

$$\langle ad_X(Y), Z \rangle = \langle [X, Y], Z \rangle = \langle Y, ad_X^*(Z) \rangle .$$

Dire que ad_X est *anti-auto-adjoint*, i.e. $ad_X^* = -ad_X$, c’est dire que

$$\langle [X, Y], Z \rangle = -\langle Y, [X, Z] \rangle = \langle Y, [Z, X] \rangle = \langle [Z, X], Y \rangle .$$

Il est alors facile de montrer que, pour toute métrique pour laquelle $X = e_1$, $\mathfrak{K}(X, Y) \geq 0$ pour tout Y et que $\mathfrak{K}(X, Y) = 0$ si et seulement si $X \perp [Y, \mathcal{G}]$ (\perp signifiant l’orthogonalité), et donc que $\mathfrak{R}(X) = \sum_j \mathfrak{K}(X, e_j) \geq 0$ et que $\mathfrak{R}(X) = 0$

36. Comme nous l’avons déjà noté, il ne faut pas confondre les deux types d’adjonction, l’une relevant de la structure d’algèbre de Lie et l’autre de la métrique.

si et seulement si $X \perp [\mathcal{G}, \mathcal{G}]$. En effet, $ad_{e_1}^* = -ad_{e_1}$ signifie que c_{1j}^k est antisymétrique en j, k , ce qui implique

$$c_{k1}^j = -c_{1k}^j = c_{1j}^k, \quad c_{\rho 1}^1 = -c_{1\rho}^1 = 0, \quad c_{1j}^k c_{jk}^1 = -c_{1k}^j c_{jk}^1 = c_{1k}^j c_{kj}^1 = c_{\rho j}^1 c_{1\rho}^j$$

et par conséquent

$$\begin{aligned} \mathfrak{K}(e_1, e_j) &= \frac{1}{2} c_{1j}^k c_{jk}^1 - \frac{1}{4} c_{\rho j}^1 (2c_{1\rho}^j - c_{\rho j}^1) - c_{\rho 1}^1 c_{\rho j}^j \\ &= \frac{1}{2} c_{1j}^k c_{jk}^1 - \frac{1}{2} c_{\rho j}^1 c_{1\rho}^j + \frac{1}{4} (c_{\rho j}^1)^2 \\ &= \frac{1}{4} (c_{\rho j}^1)^2 \geq 0. \end{aligned}$$

On voit donc que $\mathfrak{K}(e_1, e_j) = 0$ si et seulement si $c_{\rho j}^1 = 0$ pour tout ρ . Mais si $Y = \sum_{\rho} y_{\rho} e_{\rho}$, le crochet

$$[Y, e_j] = \sum_{\rho} y_{\rho} [e_{\rho}, e_j] = \sum_{\rho, k} y_{\rho} c_{\rho j}^k e_k$$

et donc $c_{\rho j}^1 = 0$ implique $[Y, e_j] \perp e_1$. Cela permet de conclure. En particulier, si X appartient au *centre* \mathcal{Z} de \mathcal{G} , $ad_X = 0$ est automatiquement anti-auto-adjoint et donc $\mathfrak{K}(X, Y) \geq 0$ pour tout Y quelle que soit la métrique.

On voit ainsi le lien entre l'anti-auto-adjonction de la représentation adjointe et la *positivité* de la courbure \mathfrak{K} .

Nous venons de voir que si ad_X est anti-auto-adjoint, alors $\mathfrak{K}(X) = 0$ si et seulement si $X \perp [\mathcal{G}, \mathcal{G}]$. Mais réciproquement, si $X \perp [\mathcal{G}, \mathcal{G}]$ on a seulement $\mathfrak{K}(X) \leq 0$. Si de plus ad_X est anti-auto-adjoint alors $\mathfrak{K}(X) \geq 0$ et donc nécessairement $\mathfrak{K}(X) = 0$. Milnor démontre ce lemme en considérant le supplément orthogonal X^{\perp} de X qui est un *idéal* de \mathcal{G} (puisque $[\mathcal{G}, \mathcal{G}] \subset X^{\perp}$ par hypothèse) muni de sa métrique induite. Les calculs sont très intéressants.

Soit $L = ad_X$ restreint à X^{\perp} et $S = \frac{1}{2}(L + L^*)$ sa partie auto-adjointe sur X^{\perp} , $A = \frac{1}{2}(L - L^*)$ étant sa partie anti-auto-adjointe. On peut calculer facilement la connexion ∇ à partir de celle ∇^{\perp} sur X^{\perp} ainsi que de L et de S en utilisant la formule de base (F ∇). En effet si $U, V \in X^{\perp}$ (on suppose tous les vecteurs unitaires pour simplifier),

$$\begin{aligned} \langle \nabla_U V, X \rangle &= \frac{1}{2} (\langle [U, V], X \rangle - \langle [V, X], U \rangle + \langle [X, U], V \rangle) \\ &= \frac{1}{2} (\langle L(V), U \rangle + \langle L(U), V \rangle) \quad \text{car } [U, V] \in X^{\perp} \\ &= \langle S(U), V \rangle \end{aligned}$$

et donc $\nabla_U V = \nabla_U^\perp V + \langle S(U), V \rangle X$. Par ailleurs

$$\langle \nabla_X X, Z \rangle = \frac{1}{2} (\langle [X, X], Z \rangle - \langle [X, Z], X \rangle + \langle [Z, X], X \rangle) = \langle [Z, X], X \rangle .$$

Mais comme $[Z, X] \in [\mathcal{G}, \mathcal{G}] \subset X^\perp$, $\langle \nabla_X X, Z \rangle = 0$ pour tout Z et donc $\nabla_X X = 0$. Pour $\nabla_X U$, $U \in X^\perp$, on trouve pour tout Z

$$\begin{aligned} \langle \nabla_X U, Z \rangle &= \frac{1}{2} (\langle [X, U], Z \rangle - \langle [U, Z], X \rangle + \langle [Z, X], U \rangle) \\ &= \frac{1}{2} (\langle L(U), Z \rangle - \langle L(Z), U \rangle) = \frac{1}{2} (\langle L(U), Z \rangle - \langle Z, L^*(U) \rangle) \\ &= \frac{1}{2} (\langle L(U), Z \rangle - \langle L^*(U), Z \rangle) = \frac{1}{2} (\langle L(U) - L^*(U), Z \rangle) \\ \nabla_X U &= \frac{1}{2} (L(U) - L^*(U)) = A(U) . \end{aligned}$$

Enfin, pour $\nabla_U X$, $U \in X^\perp$, on trouve

$$\begin{aligned} \langle \nabla_U X, Z \rangle &= \frac{1}{2} (\langle [U, X], Z \rangle - \langle [X, Z], U \rangle + \langle [Z, U], X \rangle) \\ &= -\frac{1}{2} (\langle L(U), Z \rangle + \langle L(Z), U \rangle) = -\frac{1}{2} (\langle L(U), Z \rangle + \langle Z, L^*(U) \rangle) \\ &= -\frac{1}{2} (\langle L(U), Z \rangle + \langle L^*(U), Z \rangle) = -\langle S(U), Z \rangle \\ \nabla_U X &= -S(U) . \end{aligned}$$

On considère alors une base orthonormée de X^\perp $\{u_1, u_2\}$ diagonalisant S , i.e. $Su_i = \sigma_i u_i$, $i = 1, 2$. On a pour $U \in X^\perp$

$$\begin{aligned} \mathfrak{K}(X, U) &= \langle \nabla_{[X, U]} X - \nabla_X \nabla_U X + \nabla_U \nabla_X X, U \rangle \\ &= \langle \nabla_{[X, U]} X - \nabla_X \nabla_U X, U \rangle \text{ car } \nabla_X X = 0 \\ &= \langle \nabla_{[X, U]} X + \nabla_X S(U), U \rangle = \langle -S([X, U]) + AS(U), U \rangle \\ &= \langle -SL(U) + AS(U), U \rangle . \end{aligned}$$

En particulier

$$\mathfrak{K}(X, u_i) = \langle -SL(u_i) + AS(u_i), u_i \rangle .$$

Mais par définition $L = S + A$ et donc $L(u_i) = S(u_i) + A(u_i)$ et donc

$$\begin{aligned}\mathfrak{K}(X, u_i) &= \langle -S^2(u_i) - SA(u_i) + AS(u_i), u_i \rangle \\ &= -\sigma_i^2 - \langle SA(u_i), u_i \rangle + \sigma_i \langle A(u_i), u_i \rangle \\ &= -\sigma_i^2 - \langle u_i, A^*S^*(u_i) \rangle + \sigma_i \langle A(u_i), u_i \rangle \\ &= -\sigma_i^2 - \sigma_i \langle u_i, A^*(u_i) \rangle + \sigma_i \langle A(u_i), u_i \rangle \\ &= -\sigma_i^2\end{aligned}$$

car A est anti-auto-adjoint. On en déduit d'abord que

$$\mathfrak{K}(X) = \mathfrak{K}(X, u_1) + \mathfrak{K}(X, u_2) = -(\sigma_1^2 + \sigma_2^2) \leq 0$$

et ensuite que $\mathfrak{K}(X) = 0$ si et seulement si $\sigma_1 = \sigma_2 = 0$, i.e. $S = 0$ et $L = A$ anti-auto-adjoint.

Le caractère anti-auto-adjoint de ad_X est lié aux propriétés d'invariance supplémentaires de la métrique \check{g} . En effet il y a équivalence entre le fait que ad_X soit anti-auto-adjoint pour tout X et le fait que \check{g} soit *bi-invariante*.³⁷ Dans ce cas, d'après ce qui précède, $\mathfrak{K}(X, Y) \geq 0$ pour tout X et Y car $\mathfrak{K}(X, Y) = \frac{1}{4} \langle [X, Y], [X, Y] \rangle$. En effet, dans la formule (FV)

$$\langle \nabla_X Y, Z \rangle = \frac{1}{2} (\langle [X, Y], Z \rangle - \langle [Y, Z], X \rangle + \langle [Z, X], Y \rangle) ,$$

on a

$$\begin{aligned}-\langle [Y, Z], X \rangle + \langle [Z, X], Y \rangle &= \langle ad_Z(Y), X \rangle + \langle ad_Z(X), Y \rangle \\ &= \langle ad_Z(Y), X \rangle + \langle X, ad_Z^*(Y) \rangle \\ &= \langle ad_Z(Y), X \rangle + \langle ad_Z^*(Y), X \rangle \\ &= \langle ad_Z(Y), X \rangle \langle ad_Z(Y), X \rangle \\ &= 0\end{aligned}$$

et donc

$$\begin{aligned}\langle \nabla_X Y, Z \rangle &= \frac{1}{2} \langle [X, Y], Z \rangle \\ \nabla_X Y &= \frac{1}{2} ad_X(Y) .\end{aligned}$$

37. C'est la version infinitésimale du fait que Ad_g soit une isométrie pour tout $g \in G$. En effet la matrice de Ad_g est orthogonale si et seulement si $Ad_g^{-1} = Ad_g^*$ et comme $Ad_g = e^{ad_X}$ avec $g = e^X$, cela implique $ad_X = -ad_X^*$.

On en déduit $R_{X,Y} = \frac{1}{4}ad_{[X,Y]}$ et par conséquent

$$\begin{aligned}\mathfrak{K}(X, Y) &= \langle R_{X,Y}(X), Y \rangle = \frac{1}{4} \langle [[X, Y], X], Y \rangle = -\frac{1}{4} \langle [X, [X, Y]], Y \rangle \\ &= -\frac{1}{4} \langle ad_X([X, Y]), Y \rangle = \frac{1}{4} \langle [X, Y], ad_X(Y) \rangle = \frac{1}{4} \langle [X, Y], [X, Y] \rangle .\end{aligned}$$

Cette propriété implique que G soit isomorphe au produit d'un groupe de Lie K compact et d'un groupe abélien A .

Ceci dit, la contrainte que $\mathfrak{K}(X, Y) = 0$ si et seulement si $X \perp [Y, \mathcal{G}]$ implique qu'il est difficile d'avoir la positivité stricte $\mathfrak{K}(X, Y) > 0$. Un théorème de Nolan Wallach dit que $SU(2)$ (le revêtement universel de $SO(3)$) est le seul groupe de Lie simplement connexe de dimension 3 possédant cette propriété de positivité stricte de la courbure.

De façon générale, il existe une métrique L -invariante sur G de courbure de Ricci $\mathfrak{R}(X) > 0$ pour tout X , si et seulement si G est compact avec un groupe fondamental $\pi_1(G)$ fini. Il s'agit de la réciproque d'un célèbre théorème dû à Pierre-Ossian Bonnet (1935) et Summer Byron Myers (1941) disant que si M est une variété riemannienne complète³⁸ de dimension n de courbure de Ricci partout $\geq (n-1)\delta > 0$ alors le diamètre $D(M)$ de M est $\leq \frac{\pi}{\sqrt{\delta}}$ avec égalité si et seulement si M est isométrique à la sphère de rayon $\frac{1}{\sqrt{\delta}}$ (Shiu-Yuen Cheng, 1975)³⁹. Cela implique que M soit compacte de $\pi_1(G)$ fini.

En fait il y a souvent un mixte de positivité et de négativité. Par exemple Wolf et Milnor démontrent le théorème que si G est nilpotent non commutatif, alors il existe $X, Y \in \mathcal{G}$ tels que $\mathfrak{R}(X) > 0$ et $\mathfrak{R}(Y) < 0$. On considère la suite centrale descendante

$$\mathcal{G} \supset [\mathcal{G}, \mathcal{G}] \supset [\mathcal{G}, [\mathcal{G}, \mathcal{G}]] \supset \dots \supset 0 .$$

Comme \mathcal{G} est nilpotente cette suite atteint 0 en un nombre fini de pas. Comme \mathcal{G} est non abélienne, cette suite comprend au moins 3 termes. Soit $X \neq 0$ un élément non nul de l'avant-dernier terme (le dernier non nul). On a $X \neq 0$ et $X \in [\mathcal{G}, \mathcal{G}]$. Et comme X appartient à l'avant-dernier terme de la suite, $[\mathcal{G}, X] = 0$ et donc $ad_X = 0$ est anti-auto-adjoint. D'après ce qui précède, cela implique $\mathfrak{R}(X) > 0$ puisque $\mathfrak{R}(X) \geq 0$ et qu'on ne peut avoir $\mathfrak{R}(X) = 0$ puisque la propriété $X \in [\mathcal{G}, \mathcal{G}]$ est l'opposé de la propriété $X \perp [\mathcal{G}, \mathcal{G}]$. Par ailleurs, il existe un Y tel que $Y \perp [\mathcal{G}, \mathcal{G}]$ et $Y \notin \mathcal{Z}$. Sinon on aurait $\mathcal{G} = [\mathcal{G}, \mathcal{G}] + \mathcal{Z}$ et donc

$$[\mathcal{G}, \mathcal{G}] = [\mathcal{G}, [\mathcal{G}, \mathcal{G}] + \mathcal{Z}] = [\mathcal{G}, [\mathcal{G}, \mathcal{G}]]$$

38. M est complète si les suites de Cauchy convergent, i.e. s'il n'y a pas de points limites de telles suites "qui manquent".

39. Ici, "diamètre" signifie la longueur maximale d'une géodésique entre deux points et donc la moitié πR du périmètre d'un grand cercle pour une sphère de rayon R .

et la suite centrale s'arrêterait à $[\mathcal{G}, \mathcal{G}] = 0$. Donc $ad_Y \neq 0$ (car $Y \notin \mathcal{Z}$) et ad_Y est nilpotente car \mathcal{G} est nilpotente. Cela implique que ad_Y ne peut pas être anti-auto-adjoint. En effet, si elle l'était et si U était tel que $ad_Y(U) \neq 0$ et $ad_Y^k(U) \neq 0$, on aurait alors

$$\langle ad_Y^{2k}(U), U \rangle = \pm \langle ad_Y^k(U), ad_Y^k(U) \rangle \neq 0$$

et on ne pourrait jamais arriver à la nilpotence $ad_Y^{2k}(U) = 0$. Donc, comme $Y \perp [\mathcal{G}, \mathcal{G}]$, $\mathfrak{K}(Y) \leq 0$ et comme ad_Y n'est pas anti-auto-adjoint $\mathfrak{K}(Y) \neq 0$ et donc $\mathfrak{K}(Y) < 0$.

Un autre théorème dit que si G est *résoluble* (et a fortiori nilpotent) alors toute métrique L -invariante est soit plate soit de courbure scalaire $\mathfrak{r} < 0$ et si \mathcal{G} est non abélienne il existe toujours au moins une métrique L -invariante \check{g} de $\mathfrak{r} < 0$. C'est le cas de nos groupes de référence \mathbb{H}_{pol} et $SE(2)$.

Les métriques *plates* sont celles de courbure $\mathfrak{K}(X, Y) \equiv 0$. C'est évidemment le cas pour les algèbres commutatives. Plus généralement, c'est encore le cas pour les $\mathcal{G} = \mathcal{A} \oplus^\perp \mathcal{J}$ avec \mathcal{A} une sous-algèbre abélienne et \mathcal{J} un idéal abélien tel que ad_X soit anti-auto-adjoint pour tout $X \in \mathcal{J}$. Ces G sont résolubles comme $SE(2)$.

Regardons par conséquent les possibilités de courbure sectionnelle $\mathfrak{K}(X, Y) \leq 0$. Milnor cite plusieurs théorèmes. L'un de Robert Azencott et Wilson (1976) dit que si $\mathfrak{K}(Y, Z) \leq 0$ pour tout (Y, Z) alors G est *résoluble* et que si de plus G est *unimodulaire* alors la métrique est nécessairement *plate* (ce qui implique que si G est unimodulaire et si la métrique n'est pas plate alors il existe au moins une courbure sectionnelle $\mathfrak{K}(X, Y) > 0$). L'autre est un théorème de Ernest Heintze (1974) disant que si $\mathfrak{K}(Y, Z) < 0$ pour tout (Y, Z) , alors $\mathcal{G} = [\mathcal{G}, \mathcal{G}] \oplus^\perp \mathbb{R}X$ avec $L = ad_X$ restreint à $[\mathcal{G}, \mathcal{G}] = X^\perp$ ayant toutes ses valeurs propres de partie réelle > 0 (nous venons de voir que $\mathfrak{K}(Y, Z) \geq 0$ lorsque L est anti-auto-adjointe et donc de valeurs propres $= 0$).

Milnor donne l'exemple intéressant de courbure < 0 pour les \mathcal{G} ayant la propriété que les crochets $[Y, Z]$ appartiennent toujours à l'espace $\text{Span}\{Y, Z\}$. On montre que cela implique en fait

$$[Y, Z] = \varphi(Y)Z - \varphi(Z)Y$$

où φ est une application linéaire $\varphi : \mathcal{G} \rightarrow \mathbb{R}$ et $\varphi \neq 0$ si \mathcal{G} n'est pas commutative. Soit alors \mathfrak{n} le noyau de φ . C'est trivialement un idéal commutatif de codimension 1 de \mathcal{G} . On peut donc choisir un X unitaire tel que $X \perp \mathfrak{n}$ (i.e. $\mathfrak{n} \subset X^\perp$) ce qui fait que $\varphi(X) = \|\varphi\|$. En considérant la restriction L de ad_X à \mathfrak{n} et en utilisant les mêmes arguments que plus haut, on montre que pour tout $Y \in \mathfrak{n}$ la courbure sectionnelle est donnée par

$$\mathfrak{K}(X, Y) = \|\varphi\|^2 (\langle X, Y \rangle^2 - \langle X, X \rangle \langle Y, Y \rangle)$$

et donc par

$$\mathfrak{K}(X, Y) = -\|\varphi\|^2 < 0$$

si $\{X, Y\}$ est orthonormé.

Remarque. En revanche, si $[Y, Z]$ est linéairement indépendant de $\{Y, Z\}$, alors il existe une métrique L -invariante \check{g} telle que $\mathfrak{R}(Y) < 0$ mais $\mathfrak{R}(Z) > 0$. \square

Pour obtenir des formules plus explicites pour G connexe de dimension 3, Milnor utilise les formes normales de la classification de Bianchi que nous avons présentées à la section 3.3 et en donne la variante suivante.

1. Si G est unimodulaire, il existe une base orthonormée $\{e_1, e_2, e_3\}$ de \mathcal{G} pour \check{g} telle que

$$[e_1, e_2] = \lambda_3 e_3, [e_2, e_3] = \lambda_1 e_1, [e_3, e_1] = \lambda_2 e_2$$

et qui diagonalise la courbure de Ricci. Les courbures principales de Ricci sont alors (avec \circlearrowleft symbolisant une permutation circulaire) $\mathfrak{R}(e_1) = 2\mu_2\mu_3$ (\circlearrowleft) avec $\mu_i = \frac{1}{2} \left(\sum_j \lambda_j \right) - \lambda_i$ et la courbure scalaire est $\mathfrak{r} = 2 \sum_{\circlearrowleft} \mu_2\mu_3$.

2. Si en revanche G n'est pas unimodulaire, il existe une base orthogonale $\{e_1, e_2, e_3\}$ de \mathcal{G} pour \check{g} telle que

$$[e_1, e_2] = \alpha e_2 + \beta e_3, [e_2, e_3] = 0, [e_1, e_3] = \gamma e_2 + \delta e_3$$

avec $\alpha + \delta = 2$ et $\alpha\gamma + \beta\delta = 0$ et qui diagonalise la courbure de Ricci. Les courbures principales de Ricci sont alors

$$\begin{cases} \mathfrak{R}(e_1) = -(\alpha^2 + \delta^2) - \frac{1}{2}(\beta + \gamma)^2 \\ \mathfrak{R}(e_2) = -\alpha(\alpha + \delta) + \frac{\gamma^2 - \beta^2}{2} \\ \mathfrak{R}(e_3) = -\delta(\alpha + \delta) + \frac{\beta^2 - \gamma^2}{2} \end{cases}$$

et si l'on pose

$$\alpha = 1 + \xi, \delta = 1 - \xi, \beta = (1 + \xi)\eta, \gamma = -(1 - \xi)\eta,$$

ce qui est une façon simple de satisfaire les contraintes, on obtient

$$\begin{cases} \mathfrak{R}(e_1) = -2(1 + \xi^2(1 + \eta^2)) \\ \mathfrak{R}(e_2) = -2(1 + \xi(1 + \eta^2)) \\ \mathfrak{R}(e_3) = -2(1 - \xi(1 + \eta^2)) \end{cases}$$

Comme $\mathfrak{R}(X) = \sum_j x_j^2 \mathfrak{R}(e_j)$, on peut ainsi calculer facilement toutes les courbures de Ricci.⁴⁰

40. Avec des changements de base évidents, la classe B I (B pour CLASSE DE BIANCHI) correspond à $\lambda_1 = 0$; la classe B II à $\lambda_3 = 1, \lambda_1 = \lambda_2 = 0$; la classe B III à $\alpha = \beta = \gamma = 0, \delta = 2$; la classe B IV demande un changement de base moins simple ; la classe B VI à $\beta = \gamma = 0, \alpha = -1, \delta = -\ell$ (pour avoir $\alpha + \delta = 2$ il faut un changement de base moins simple) ; la classe B V à $\ell = 1$; la classe B VI₀ à $\ell = -1$; la classe B VII à $\alpha = -\ell, \delta = -\ell, \beta = 1, \gamma = -1$ (on a bien $\alpha\gamma + \beta\delta = 0$ et pour avoir $\alpha + \delta = 2$ il faut un changement de base moins simple) ; la classe B VII₀ à $\lambda_1 = \lambda_2 = 1, \lambda_3 = 0$; la classe B VIII à un changement de base moins simple ; la classe B IX à $\lambda_1 = \lambda_2 = \lambda_3 = 1$.

Une façon très simple d'obtenir ces formes normales est de tenir compte du fait que \mathcal{G} étant de dimension 3 et muni d'une métrique \check{g} , il y existe un *produit vectoriel* $X \times Y$ au sens de l'analyse vectorielle classique que nous avons évoquée à la section 8.1 du chapitre 2. Cela permet de corréler très facilement le crochet de Lie et la métrique au moyen de \times . Choisissons une base orthonormale orientée $\{e_1, e_2, e_3\}$ de \mathcal{G} . On a par définition $e_1 \times e_2 = e_3$ ($\check{\circ}$) et l'on peut considérer l'application linéaire $W : \mathcal{G} \longrightarrow \mathcal{G}$ définie par $W(e_1) = [e_2, e_3]$ ($\check{\circ}$). Elle satisfait par construction $W(e_i \times e_j) = [e_i, e_j]$ et donc $W(X \times Y) = [X, Y]$. Il est alors facile de calculer la matrice des ad_{e_i} à partir de celle de $W = (w_{ij})$.⁴¹

Or nous avons vu plus haut que G est unimodulaire si et seulement si les ad_X sont de trace nulle. Donc, dans ce cas, $w_{32} - w_{23} = 0$ ($\check{\circ}$) et par conséquent W est symétrique (auto-adjointe). En choisissant la base $\{e_1, e_2, e_3\}$ de façon à diagonaliser W , on obtient

$$W(e_1) = [e_2, e_3] = \lambda_1 e_1 \text{ ($\check{\circ}$)}$$

et donc la forme normale

$$[e_1, e_2] = \lambda_3 e_3 \text{ ($\check{\circ}$) .}$$

On en déduit les constantes c_{ij}^k et, en appliquant soigneusement la formule (K), on calcule aisément les $\mathfrak{K}(e_i, e_j)$, ce qui donne les formules ci-dessus.⁴²

Si en revanche G n'est pas unimodulaire, on utilise le fait que d'après Jacobi $ad_{[X, Y]} = [ad_X, ad_Y]$ et que donc $T : X \longrightarrow \text{tr}(ad_X)$ est un morphisme d'algèbres de Lie de \mathcal{G} dans $(\mathbb{R}, +)$ avec $T \neq 0$ puisque G n'est pas unimodulaire. Le noyau \mathfrak{n} de T est donc un idéal *propre* de \mathcal{G} , idéal de dimension ≤ 2 qui contient $[\mathcal{G}, \mathcal{G}]$ car $\text{tr}(ad_{[X, Y]}) = \text{tr}([ad_X, ad_Y]) = 0$ puisque $\text{tr}(AB) = \text{tr}(BA)$ pour toute paire d'endomorphismes. Si $[\mathcal{G}, \mathcal{G}]$ est de dimension 1, le seul G non unimodulaire est d'après la classification 3.3 de classe B III avec $\alpha = \beta = \gamma = 0$, $\delta = 2$ et satisfait bien

41. Par exemple, à la main,

$$\begin{aligned} ad_{e_1}(Y) &= [e_1, Y] = W(e_1 \times Y) = \sum_j Y_j W(e_1 \times e_j) = Y_2 W(e_3) - Y_3 W(e_2) \\ &= Y_2 \left(\sum_j w_{3j} e_j \right) - Y_3 \left(\sum_j w_{2j} e_j \right) = \begin{pmatrix} 0 & w_{31} & -w_{21} \\ 0 & w_{32} & -w_{22} \\ 0 & w_{33} & -w_{23} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} . \end{aligned}$$

42. Par exemple

$$[e_1, e_2] = \lambda_3 e_3 = c_{12}^1 e_1 + c_{12}^2 e_2 + c_{12}^3 e_3$$

et donc

$$c_{12}^1 = 0, c_{12}^2 = 0, c_{12}^3 = \lambda_3, \text{ etc.}$$

et aussi

$$\mathfrak{K}(e_1, e_2) = \frac{1}{2} \lambda_3 (\lambda_2 - \lambda_3 + \lambda_1) - \frac{1}{4} (-\lambda_3 + \lambda_2 - \lambda_1) (-\lambda_2 + \lambda_1 - \lambda_3),$$

etc., ce qui donne bien $\mathfrak{K}(e_1) = \mathfrak{K}(e_1, e_2) + \mathfrak{K}(e_1, e_3) = 2\mu_2 \mu_3$.

les conditions. On peut donc supposer que $[\mathcal{G}, \mathcal{G}]$ est de dimension 2 et par conséquent que \mathfrak{n} est aussi de dimension 2. L'idéal \mathfrak{n} est par construction unimodulaire en tant qu'algèbre de Lie et, puisqu'il est de dimension 2, est commutatif d'après la classification 3.3. Soit $e_1 \in \mathcal{G}$ tel que $\text{tr}(ad_{e_1}) = 2$ et considérons comme plus haut $L = ad_{e_1}$ restreint à \mathfrak{n} de matrice $\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$. Si L est l'identité de \mathfrak{n} , on a alors la forme normale $\alpha = \delta = 1, \beta = \gamma = 0$ qui satisfait bien les conditions. Si L n'est pas l'identité, on choisit une base orthonormée $\{e_2, e_3\}$ de \mathfrak{n} et on obtient la forme normale déclarée, $\alpha\gamma + \beta\delta = 0$ exprimant l'orthogonalité $e_2 \perp e_3$. Le déterminant $\alpha\delta - \beta\gamma$ classe les différentes algèbres possibles.

Pour les courbures sectionnelles, en appliquant la formule (K) avec

$$c_{12}^1 = 0, c_{12}^2 = \alpha, c_{12}^3 = \beta, c_{13}^1 = 0, c_{13}^2 = \gamma, c_{13}^3 = \delta, c_{23}^1 = 0, c_{23}^2 = 0, c_{23}^3 = 0$$

et l'antisymétrie $c_{ji}^k = -c_{ij}^k$, on obtient

$$\begin{aligned} \mathfrak{R}(e_1, e_2) &= -\alpha^2 + \frac{1}{4}(\gamma^2 - 3\beta^2 - 2\beta\gamma) \\ \mathfrak{R}(e_1, e_3) &= -\delta^2 + \frac{1}{4}(\beta^2 - 3\gamma^2 - 2\gamma\beta) \\ \mathfrak{R}(e_2, e_3) &= \frac{1}{4}(\beta + \gamma)^2 - \alpha\delta \end{aligned}$$

ce qui donne les courbures principales de Ricci $\mathfrak{R}(e_i)$ déclarées.

Ces valeurs explicites des courbures principales $\mathfrak{R}(e_i)$ ont des corollaires immédiats. Par exemple si G est unimodulaire, le produit des courbures principales est ≥ 0 car il est égal à $8(\mu_1\mu_2\mu_3)^2$. S'il est nul, c'est que l'un des $\mu_i = 0$ et par conséquent les deux courbures principales $\mathfrak{R}(e_j)$ $j \neq i$ sont nulles.

- Pour le groupe de Heisenberg $\lambda_3 = 1, \lambda_1 = \lambda_2 = 0, \mu_i = \frac{1}{2} - \lambda_i, \mu_1 = \mu_2 = \frac{1}{2}, \mu_3 = -\frac{1}{2}, \mathfrak{R}(e_1) = \mathfrak{R}(e_2) = -\frac{1}{2}, \mathfrak{R}(e_3) = \frac{1}{2}$ et $\mathfrak{r} = -\frac{1}{2} < 0$.
- Pour $SE(2)$ $\lambda_3 = 0, \lambda_1 = \lambda_2 = 1, \mu_i = 1 - \lambda_i, \mu_1 = \mu_2 = 0, \mu_3 = 1, \mathfrak{R}(e_1) = \mathfrak{R}(e_2) = \mathfrak{R}(e_3) = 0$ et la métrique est plate.

Pour les G non unimodulaires, le déterminant

$$D = \alpha\delta - \beta\gamma = (1 - \xi^2)(1 + \eta^2)$$

et

$$\begin{cases} \mathfrak{R}(e_1) = -2(1 + \xi^2(1 + \eta^2)) \leq -2 \\ \mathfrak{R}(e_2) = -2(1 + \xi(1 + \eta^2)) \\ \mathfrak{R}(e_3) = -2(1 - \xi(1 + \eta^2)) \end{cases}$$

Quitte à échanger e_2 et e_3 , on peut supposer $\xi, \eta \geq 0$ et construire un "diagramme de phase" dans ce quadrant ≥ 0 du plan (ξ, η) . Les différentes possibilités sont représentées à la figure 1. La séparatrice S pour les courbures est la courbe d'équation $\xi(1 + \eta^2) = 1$ qui est décroissante sur $\xi = [0, 1]$, asymptotique à l'axe η pour $\xi \rightarrow 0^+$

et arrivant en $(1, 0)$ avec une tangente verticale après avoir inversé sa courbure pour $\xi = \frac{3}{4}$. Sur S , $D = \frac{1-\xi^2}{\xi} = \frac{1}{\xi} - \xi$.

1. Pour $\eta = 0$, $D = 1 - \xi^2$,

$$\begin{cases} \mathfrak{R}(e_1) = -2(1 + \xi^2) \\ \mathfrak{R}(e_2) = -2(1 + \xi) \\ \mathfrak{R}(e_3) = -2(1 - \xi) \end{cases}$$

et donc pour $\xi > 1$, par exemple pour $\xi = \frac{3}{2}$, $D = -\frac{5}{4}$,

$$\begin{cases} \mathfrak{R}(e_1) = -\frac{13}{2} \\ \mathfrak{R}(e_2) = -2(1 + \xi) = -5 \\ \mathfrak{R}(e_3) = 1 \end{cases}$$

2. Pour $\eta > 0$, $\xi > 1$, par exemple $\xi = \frac{3}{2}$, $\eta = \frac{1}{2}$, $D = -\frac{25}{16}$ et

$$\begin{cases} \mathfrak{R}(e_1) = -\frac{61}{8} \\ \mathfrak{R}(e_2) = -\frac{23}{4} \\ \mathfrak{R}(e_3) = \frac{7}{4} \end{cases}$$

3. Pour $\eta = 0$, $\xi = 1$, $D = 0$ et

$$\begin{cases} \mathfrak{R}(e_1) = -4 \\ \mathfrak{R}(e_2) = -4 \\ \mathfrak{R}(e_3) = 0 \end{cases}$$

4. Pour $\eta > 0$, par exemple $\xi = 1$, $\eta = \frac{1}{2}$, $D = 0$ et

$$\begin{cases} \mathfrak{R}(e_1) = -\frac{9}{5} \\ \mathfrak{R}(e_2) = -\frac{9}{2} \\ \mathfrak{R}(e_3) = \frac{1}{2} \end{cases}$$

5. Pour $\eta = 0$, $\xi < 1$, D devient $D = 1 - \xi^2 > 0$ et toutes les courbures sont < 0 . Par exemple pour $\eta = 0$, $\xi = \frac{1}{2}$, $D = \frac{3}{4}$,

$$\begin{cases} \mathfrak{R}(e_1) = -\frac{5}{2} \\ \mathfrak{R}(e_2) = -3 \\ \mathfrak{R}(e_3) = -1 \end{cases}$$

6. Mais pour $\xi < 1$ et η assez grand ((ξ, η) au-dessus de S), D reste < 0 mais $\mathfrak{R}(e_3)$ devient > 0 . Par exemple pour $\xi = \frac{1}{2}$, $\eta = \frac{3}{2}$, $D = \frac{39}{16}$ et

$$\begin{cases} \mathfrak{R}(e_1) = -\frac{29}{8} \\ \mathfrak{R}(e_2) = -\frac{21}{4} \\ \mathfrak{R}(e_3) = \frac{5}{4} \end{cases}$$

7. À l'origine, $D = 1$ et $\mathfrak{R}(e_1) = \mathfrak{R}(e_2) = \mathfrak{R}(e_3) = -2$.

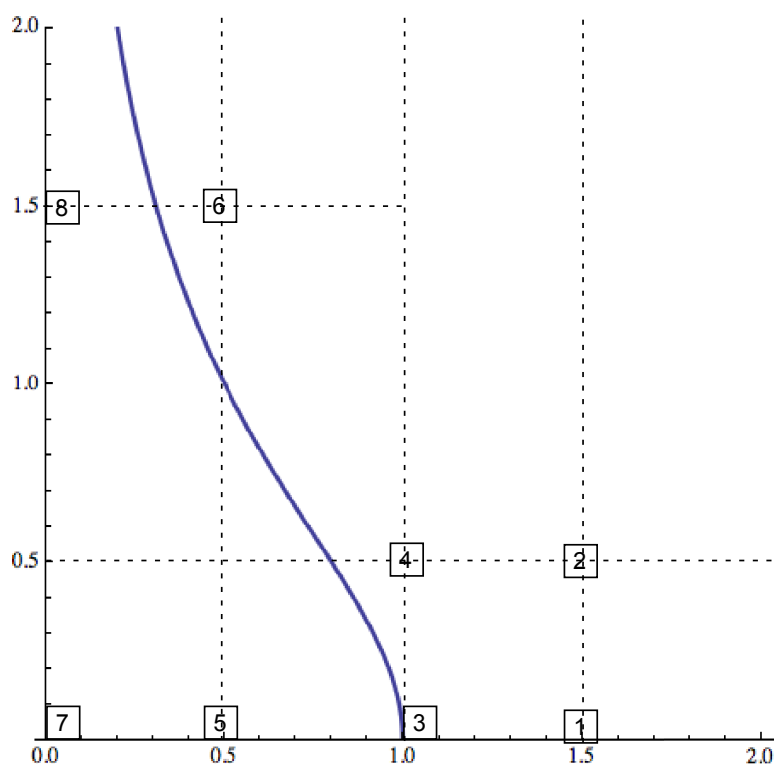


FIGURE 1. Le “diagramme de phase” des algèbres de Lie connexes $3D$ non unimodulaires. Les coordonnées sont les paramètres (ξ, η) donnant $D = (1 - \xi^2)(1 + \eta^2)$ et $\Re(e_1) = -2(1 + \xi^2(1 + \eta^2))$, $\Re(e_2) = -2(1 + \xi(1 + \eta^2))$, $\Re(e_3) = -2(1 - \xi(1 + \eta^2))$. Voir le texte pour les différents cas.

8. Il en va de même pour $\xi = 0$, $\eta > 0$, seul D variant comme $(1 + \eta^2)$.

4.2. Géodésiques homogènes

Si G est muni d’une métrique L -invariante, il y a deux types de courbes de G démarrant d’un $X \neq 0$ de \mathcal{G} . D’une part les orbites d’un groupe à 1-paramètre e^{tX} et d’autre part les géodésiques. Les géodésiques qui sont aussi des sous-groupes à 1-paramètre s’appellent des *géodésiques homogènes* (g.h.) et il est intéressant de savoir les calculer.

La théorie générale concerne les “espaces homogènes”, c’est-à-dire les variétés riemanniennes sur lesquelles un groupe de Lie opère isométriquement et transitivement. Elle est assez subtile et a été développée en grande partie d’abord par Valerius Kajzer [276], puis par Oldřich Kowalski et János Szenthe (cf. par exemple [306]) qui ont montré qu’il existe toujours au moins une g.h. La relation avec la classification de Milnor en dimension 3 a été effectuée par Rosa Anna Marinosci dans [352]. Le lecteur intéressé pourra se référer au survey [161] de Zdeněk Dušek.

Si l'on s'en tient aux résultats de base, les calculs sont explicites et simples. Les géodésiques sont définies par l'équation $\nabla_X X = 0$. D'après la formule (FV) cela équivaut à $\langle [X, Z], X \rangle = 0$ pour tout Z . Si G est unimodulaire, la forme normale $[e_1, e_2] = \lambda_3 e_3$ (\odot) montre que

$$\begin{aligned} \left\langle \left[e_1, Z = \sum_j z_j e_j \right], e_1 \right\rangle &= \langle z_2 [e_1, e_2], e_1 \rangle + \langle z_3 [e_1, e_3], e_1 \rangle \\ &= z_2 \lambda_3 \langle e_3, e_1 \rangle - z_2 \lambda_2 \langle e_2, e_1 \rangle = 0 \quad (\odot) \end{aligned}$$

et donc que les e_i sont homogènes. En fait ce sont les *seuls* (à un facteur près) si les λ_i sont tous distincts car si $X = \sum_j x_j e_j$ est homogène, alors

$$\langle [X, Z], X \rangle = (\lambda_1 - \lambda_2) x_1 x_2 z_3 + (\lambda_2 - \lambda_3) x_2 x_3 z_1 + (\lambda_3 - \lambda_1) x_3 x_1 z_2$$

pour tout Z et donc $x_1 x_2 = x_2 x_3 = x_3 x_1 = 0$ et donc deux des x_i doivent s'annuler.

À l'autre extrême, si tous λ_i sont *égaux*, alors *tous* les $X \in \mathcal{G}$ sont homogènes. Si deux des λ_i sont égaux, par exemple $\lambda_1 = \lambda_2$, la condition d'homogénéité se réduit à

$$(\lambda_2 - \lambda_3) x_2 x_3 z_1 + (\lambda_3 - \lambda_1) x_3 x_1 z_2$$

c'est-à-dire à $x_2 x_3 = x_3 x_1 = 0$. Cela signifie que e_3 (à un facteur près) est homogène ainsi que tous les X du plan orthogonal $x_3 = 0$.

Si G n'est pas unimodulaire, on utilise la forme normale

$$[e_1, e_2] = \alpha e_2 + \beta e_3, \quad [e_2, e_3] = 0, \quad [e_1, e_3] = \gamma e_2 + \delta e_3$$

avec $\alpha + \delta = 2$ et $\alpha\gamma + \beta\delta = 0$ qui diagonalise la courbure de Ricci et l'on pose comme plus haut $\alpha = 1 + \xi$, $\delta = 1 - \xi$, $\beta = (1 + \xi)\eta$, $\gamma = -(1 - \xi)\eta$. Comme $[e_1, Z] \in \text{Span}\{e_2, e_3\} \perp e_1$, on a $\langle [e_1, Z], e_1 \rangle = 0$ pour tout Z et donc e_1 est homogène. Regardons les conditions pour qu'il existe d'autres g.h. En annulant les coefficients des z_j dans $\langle [X, Z], X \rangle = 0$ on obtient les équations en les x_i

$$\begin{cases} (1 + \xi) x_2^2 + 2\xi\eta x_2 x_3 + (1 - \xi) x_3^2 = 0 \\ (1 + \xi) x_2 x_1 + (1 + \xi) \eta x_3 x_1 = 0 \\ (1 - \xi) x_3 x_1 - (1 - \xi) \eta x_2 x_1 = 0 \end{cases}$$

Commençons par les cas particuliers (nous avons vu plus haut qu'on peut supposer $\xi, \eta \geq 0$). Si $\xi = 0$, les équations se réduisent à

$$\begin{cases} x_2^2 + x_3^2 = 0 \\ x_2 x_1 + \eta x_3 x_1 = 0 \\ x_3 x_1 - \eta x_2 x_1 = 0 \end{cases}$$

dont la seule solution est $x_2 = x_3 = 0$, c'est-à-dire $\text{Span}\{e_1\}$ déjà compté. Si $\xi = 1$, les équations se réduisent à

$$\begin{cases} x_2(x_2 + \eta x_3) = 0 \\ x_1(x_2 + \eta x_3) = 0 \end{cases}$$

Tous les X du plan $x_2 + \eta x_3 = 0$ sont homogènes ainsi que $\text{Span}\{e_3\}$ ⁴³. Si $\xi \neq 0, 1$, alors la première équation est une équation du second degré et on peut supposer que x_2 et x_3 ne sont pas tous les deux nuls car $\text{Span}\{e_1\}$ est déjà compté. Supposons $x_3 \neq 0$ et soit $x = \frac{x_2}{x_3}$. L'équation

$$(1 + \xi)x^2 + 2\xi\eta x + (1 - \xi) = 0$$

a pour discriminant

$$\Delta = 4\xi^2\eta^2 - 4(1 - \xi^2) = 4\xi^2(\eta^2 + 1) - 4$$

et $\Delta = 0$ pour $\xi^2(\eta^2 + 1) = 1$. Si $\Delta < 0$, i.e. $\xi^2(\eta^2 + 1) < 1$, il n'y a pas de solution réelle et donc il faut $x_2 = x_3 = 0$ et on retrouve $\text{Span}\{e_1\}$. Si $\Delta = 0$, i.e. $\xi^2 = \frac{1}{\eta^2 + 1}$, il y a une racine double $(1 + \xi)x_2 = -\xi\eta x_3$. Il faut $x_1 = 0$ car si $x_1 \neq 0$, les deux autres équations donnent

$$\begin{cases} x_2 + \eta x_3 = 0 \\ x_3 - \eta x_2 = 0 \end{cases}$$

soit $x_2(\eta^2 + 1) = 0$ ce qui implique $x_2 = 0$, donc $x_3 = 0$. On a donc une droite de X homogènes $(1 + \xi)x_2 = -\xi\eta x_3$ dans le plan $x_1 = 0$. Si $\Delta > 0$, i.e. $\xi^2(\eta^2 + 1) > 1$, il y a deux racines et l'on obtient deux droites (non orthogonales) de X homogènes dans le plan $x_1 = 0$.

4.3. Métriques sous-riemanniennes L -invariantes en D3 (Agrachev)

4.3.1. Les invariants métriques δ et κ .

On peut également classer les structures de contact munies d'une métrique *sous-riemannienne* L -invariante sur les groupes de Lie en dimension 3. Nous allons reprendre ici quelques éléments du chapitre 17 de l'opus [5] d'Andrei Agrachev, Davide Barilari, Ugo Boscain

Soit G avec sa distribution de plans de contact $\mathcal{K} = \ker(\omega)$, χ son champ de Reeb (donc $TG = \mathcal{K} \oplus \mathbb{R}\chi$) et $g_{\mathcal{K}}$ une métrique sous-riemannienne L -invariante sur \mathcal{K} . Soit $\{X_1, X_2\}$ une base *orthonormée* L -invariante de \mathcal{K} , $\{X_1, X_2, \chi = X_0\}$ la base L -invariante de TG associée et $\{\omega_1, \omega_2, \omega = \omega_0\}$ la base duale. Il s'agit de coupler la classification des algèbres de Lie de dimension 3 que nous avons vu plus haut à la section 3.3 avec le choix de la structure de contact \mathcal{K} (un plan de \mathcal{G} satisfaisant la condition d'Hörmander) et de la métrique $g_{\mathcal{K}}$, puis de calculer ensuite des formes

43. Pour $\xi = -1$ on échange x_2 et x_3 et on change η en $-\eta$.

normales de métriques en termes des constantes de structures c_{ij}^k , $i, j, k = 0, 1, 2$, données par

$$[X_j, X_i] = \sum_{i,j,k} c_{ij}^k X_k$$

ou, dualement, par

$$d\omega_k = \sum_{i<j,k} c_{ij}^k \omega_i \wedge \omega_j .$$

La stratégie n'est pas la même que celle de la section 3.3 qui ne tenait pas compte de ces données supplémentaires et, pour distinguer les deux perspectives nous utiliserons des indices romains plutôt que grecs. Notons aussi que, les 1-formes étant privilégiées par rapport aux champs de vecteurs puisque ω définit \mathcal{K} , ce sont les formules $d\omega_k = \sum_{i<j,k} c_{ij}^k \omega_i \wedge \omega_j$ qui priment et que donc, par rapport à la section 3.3, $[X_j, X_i] = \sum_{i,j,k} c_{ij}^k X_k$ l'ordre des i, j étant inverse à gauche et à droite.

On peut aussi normaliser ω relativement à la métrique par la condition

$$d\omega (X_1, X_2) = 1 ,$$

i.e. $d\omega = \omega_1 \wedge \omega_2$ sur \mathcal{K} . On a évidemment $c_{ii}^k = 0$ à cause de l'antisymétrie du commutateur et du produit extérieur et aussi $c_{01}^0 = c_{02}^0 = 0$ (i.e. $[X_1, X_0]_0 = [X_2, X_0]_0 = 0$) et $c_{12}^0 = 1$ (i.e. $[X_2, X_1]_0 = 1$) à cause de

$$d\omega (X_0, X_1) = d\omega (X_0, X_2) = 0, \quad d\omega (X_1, X_2) = 1 .$$

On a aussi $c_{01}^1 + c_{02}^2 = 0$ (i.e; $[X_1, X_0]_1 + [X_2, X_0]_2 = 0$). Pour le voir, on écrit que $d^2\omega = 0$ en tenant compte du fait que, puisque $d\omega = \omega_1 \wedge \omega_2$, on a

$$\begin{aligned} d^2\omega &= d\omega_1 \wedge \omega_2 - \omega_1 \wedge d\omega_2 = c_{01}^1 \omega_0 \wedge \omega_1 \wedge \omega_2 - \omega_1 \wedge c_{02}^2 \omega_0 \wedge \omega_1 \\ &= (c_{01}^1 + c_{02}^2) \omega_0 \wedge \omega_1 \wedge \omega_2 = 0 . \end{aligned}$$

Comme $\omega_0 \wedge \omega_1 \wedge \omega_2$ est une 3-forme volume, donc partout non nulle, $d^2\omega = 0$ implique

$$c_{01}^1 + c_{02}^2 = 0 .$$

De même, $d^2\omega_1 = 0$ et $d^2\omega_2 = 0$ impliquent

$$c_{01}^1 c_{12}^1 + c_{02}^1 c_{12}^2 = 0$$

(i.e; $[X_1, X_0]_1 [X_2, X_1]_1 + [X_2, X_0]_1 [X_2, X_1]_2 = 0$) et

$$c_{01}^2 c_{12}^1 + c_{02}^2 c_{12}^2 = 0$$

(i.e; $[X_1, X_0]_2 [X_2, X_1]_1 + [X_2, X_0]_2 [X_2, X_1]_2 = 0$).⁴⁴

Comme nous l'avons vu à la section 4 du chapitre 3 en utilisant la "formule d'homotopie" de Cartan

$$\mathcal{L}_\chi \omega = i_\chi d\omega + d(i_\chi \omega)$$

de la section 7 du chapitre 2, le champ de Reeb préserve ω car, d'une part, $i_\chi \omega = \omega(\chi) = 1$ implique $d(i_\chi \omega) = 0$, alors que, d'autre part,

$$i_\chi d\omega = d\omega(\chi, \bullet) = 0 .$$

Mais en général il ne préserve pas la métrique. Dans [5] deux invariants métriques sont alors définis, δ et κ ,⁴⁵ qui permettent de définir des bases orthonormées privilégiées. L'idée est que, si l'on considère les coordonnées λ_i des T^*G associées à la base $\{\omega_1, \omega_2, \omega = \omega_0\}$ données pour $\lambda \in T^*G$ par $\lambda_i = \lambda(X_i)$,⁴⁶ et si l'on considère l'hamiltonien quadratique

$$H = \frac{1}{2} ((\lambda_1)^2 + (\lambda_2)^2)$$

sur T^*G qui exprime la métrique,⁴⁷ alors le crochet de Poisson⁴⁸ $Q(\lambda) = \{H, \lambda_0\}$, qui calcule la dérivée de H le long du flot de Reeb χ , est la forme quadratique

$$c_{01}^1 (\lambda_1)^2 + (c_{01}^2 + c_{02}^1) \lambda_1 \lambda_2 + c_{02}^2 (\lambda_2)^2$$

sur T^*G . En effet,

$$\{H, \lambda_0\} = \lambda_1 \{\lambda_1, \lambda_0\} + \lambda_2 \{\lambda_2, \lambda_0\}$$

et comme $\{\lambda_1, \lambda_0\}$ et $\{\lambda_2, \lambda_0\}$ dualisent $[X_1, \chi] = c_{01}^1 X_1 + c_{01}^2 X_2$ et $[X_2, \chi] = c_{02}^1 X_1 + c_{02}^2 X_2$, on a

$$\begin{aligned} Q(\lambda) &= \{H, \lambda_0\} = \lambda_1 (c_{01}^1 \lambda_1 + c_{01}^2 \lambda_2) + \lambda_2 (c_{02}^1 \lambda_1 + c_{02}^2 \lambda_2) \\ &= c_{01}^1 (\lambda_1)^2 + (c_{01}^2 + c_{02}^1) \lambda_1 \lambda_2 + c_{02}^2 (\lambda_2)^2 . \end{aligned}$$

44. On a

$$\begin{aligned} d^2 \omega_1 &= c_{01}^1 (d\omega_0 \wedge \omega_1 - \omega_0 \wedge d\omega_1) + c_{02}^1 (d\omega_0 \wedge \omega_2 - \omega_0 \wedge d\omega_2) \\ &\quad + c_{12}^1 (d\omega_1 \wedge \omega_2 - \omega_1 \wedge d\omega_2) \\ &= (-c_{01}^1 c_{12}^1 - c_{02}^1 c_{12}^2 + c_{12}^1 c_{01}^1 + c_{12}^1 c_{02}^2) \omega_0 \wedge \omega_1 \wedge \omega_2 \\ &= (-c_{01}^1 c_{12}^1 - c_{02}^1 c_{12}^2) \omega_0 \wedge \omega_1 \wedge \omega_2 = 0 . \end{aligned}$$

45. δ est noté χ dans [5] mais ici χ est déjà le champ de Reeb et, par ailleurs, cet invariant étant associé à un déterminant, δ est une notation acceptable.

46. Les $\lambda \in T^*G$ sont des couples (g, λ_g) d'un élément $g \in G$ et d'un covecteur $\lambda_g \in T_g^*G$. Souvent nous identifions λ à λ_g lorsque le contexte est clair.

47. Nous reviendrons très longuement plus bas, en particulier au chapitre 12, sur les interprétations hamiltoniennes des métriques et de leurs géodésiques.

48. Cf. la section 2 du chapitre 4.

Mais en fait $Q(\lambda)$ s'annule pour les $\lambda = (0, 0, \lambda_0)$, autrement dit sur les sous-espaces de T^*G multiples de ω et $\ker Q_g(\lambda) = K_g^\perp$. Donc, par passage au quotient, $Q(\lambda)$ est en fait définie sur les plans $K_g^* = T_g^*G/K_g^\perp \simeq \text{Span}\{\lambda_1, \lambda_2\}$, plans eux-mêmes identifiables aux plans de contact d'aux K_g à travers le produit scalaire. Il s'ensuit que $Q(\lambda) = \{H, \lambda_0\}$ peut être associé à la matrice symétrique de trace nulle

$$S = \begin{pmatrix} c_{01}^1 & \frac{c_{01}^2 + c_{02}^1}{2} \\ \frac{c_{01}^2 + c_{02}^1}{2} & c_{02}^2 = -c_{01}^1 \end{pmatrix}$$

telle que

$$Q(\lambda) = \{H, \lambda_0\} = (\lambda_1, \lambda_2) S \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$

La propriété essentielle de nullité de la trace signifie que le champ de Reeb χ préserve non seulement la forme de contact ω mais également l'aire $d\omega$ des plans de contact. $\{H, \lambda_0\}$ mesure alors le fait que χ ne préserve pas la métrique en général. Comme le disent les auteurs de [5] :

“ $\{H, \lambda_0\}$ computes [...] the obstruction to the flow of the Reeb field λ_0 (which preserves the distribution and the volume form on it) to be metric-preserving.” ([5], p. 579)

D'où l'idée de considérer le déterminant

$$\Delta = -(c_{01}^1)^2 - \left(\frac{c_{01}^2 + c_{02}^1}{2}\right)^2 \leq 0$$

de $Q(\lambda) = \{H, \lambda_0\}$. En effet, le flot de χ est un flot d'isométries si et seulement si $\Delta = 0$. Dans ce cas, $S = 0$ car si une matrice $S = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$ symétrique de trace nulle est de déterminant $\Delta = -a^2 - b^2 = 0$, alors nécessairement $a = b = 0$ et $S = 0$.

Le premier invariant est $\delta = \sqrt{-\Delta}$. Un second invariant apparemment moins évident est défini par

$$\kappa = \frac{c_{01}^2 - c_{02}^1}{2} - (c_{12}^1)^2 - (c_{12}^2)^2,$$

nous allons voir pourquoi. Il est lié à la *courbure* de la métrique L -invariante g_K . Soit $\vec{H} = \left(\frac{\partial H}{\partial \lambda}, -\frac{\partial H}{\partial g}\right)$ le champ hamiltonien défini par H , champ qui, par définition, dualise la différentielle de H :

$$dH = \frac{\partial H}{\partial \lambda} d\lambda + \frac{\partial H}{\partial g} dg \in T^*(T^*G).$$

Si $H = \frac{1}{2}$, on a $(\lambda_1)^2 + (\lambda_2)^2 = 1$ et l'on peut écrire $\lambda_1 = \cos(\theta)$, $\lambda_2 = \sin(\theta)$ puis considérer le champ \vec{H}' sur T^*G défini par le crochet $\vec{H}' = \left[\partial_\theta, \vec{H}\right]$. Cela permet

de définir un type de courbure r_λ , qui est une forme quadratique sur T^*G , par la formule

$$r_\lambda = \Omega_\lambda \left(\left[\vec{H}, \vec{H}' \right], \vec{H}' \right)$$

où Ω est la 2-forme symplectique canonique sur T^*G . Des calculs longs et complexes montrent alors

(i) que $r_\lambda \upharpoonright_{\mathcal{K}^\perp = \text{Span}(\omega)}$ est égal à $(\lambda_0)^2$ et donc que $r_\omega \equiv 1$ sur la forme de contact $\omega = (0, 0, 1)$,

(ii) que $r_\lambda \upharpoonright_{\mathcal{K}^\perp = \text{Span}(\omega_1, \omega_2)}$ est égal à $(\kappa + 3\delta)(\lambda_1)^2 + (\kappa - 3\delta)(\lambda_2)^2$.

Cela permet de donner une interprétation simple du second invariant métrique puisque 2κ est simplement la trace de r_λ . On retrouve également δ car le discriminant de r_λ , défini comme $\text{trace}(r_\lambda)^2 - 4 \det(r_\lambda)$, est égal à $4\kappa^2 - 4(\kappa^2 - 9\delta^2) = 36\delta^2$.

On peut alors entreprendre la classification. Pour faire le lien avec les formes normales de Bianchi de la section 3.3 on pourra utiliser des transformations de \mathcal{G} induisant des isométries sur $K = K_e$ (le plan de contact dans l'espace tangent $\mathcal{G} = T_e G$ en l'élément neutre e de G).

Résumons la situation :

Contraintes générales
$c_{ii}^k = 0$ (antisymétrie), $c_{01}^0 = c_{02}^0 = 0$, $c_{12}^0 = 1$ ($d\omega = \omega_1 \wedge \omega_2$)
$c_{01}^1 + c_{02}^2 = 0$, $c_{01}^1 c_{12}^1 + c_{02}^2 c_{12}^2 = 0$, $c_{01}^2 c_{12}^1 + c_{02}^1 c_{12}^2 = 0$ ($d^2 \omega_i = 0$)
Invariants métriques
$\delta = \sqrt{(c_{01}^1)^2 + \left(\frac{c_{01}^2 + c_{02}^1}{2}\right)^2}$, $\kappa = \frac{c_{01}^2 - c_{02}^1}{2} - (c_{12}^1)^2 - (c_{12}^2)^2$
Relations de commutation
$\begin{cases} [X_1, \chi] = c_{01}^1 X_1 + c_{01}^2 X_2 \\ [X_2, \chi] = c_{02}^1 X_1 + c_{02}^2 X_2 = c_{02}^1 X_1 - c_{01}^1 X_2 \\ [X_2, X_1] = c_{12}^1 X_1 + c_{12}^2 X_2 + \chi \end{cases}$

Les ad_{X_i} sont de matrices

$$ad_{X_1} = \begin{pmatrix} 0 & -c_{12}^1 & c_{01}^1 \\ 0 & -c_{12}^2 & c_{01}^2 \\ 0 & -1 & 0 \end{pmatrix}, \quad ad_{X_2} = \begin{pmatrix} c_{12}^1 & 0 & c_{02}^1 \\ c_{12}^2 & 0 & c_{02}^2 \\ 1 & 0 & 0 \end{pmatrix}, \quad ad_\chi = \begin{pmatrix} -c_{01}^1 & -c_{02}^1 & 0 \\ -c_{01}^2 & -c_{02}^2 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

cette dernière étant de trace nulle. La forme de Killing

$$\tilde{K}(X_i, X_i) = \text{tr}(ad_{X_i} \cdot ad_{X_j})$$

est donc donnée par⁴⁹

49. Nous la notons ici \tilde{K} car K dénote déjà les plans de contact.

$$\left\{ \begin{array}{l} \widetilde{K}(X_1, X_1) = -2c_{01}^2 + (c_{12}^2)^2 \\ \widetilde{K}(X_1, X_2) = c_{01}^1 - c_{02}^2 - c_{12}^1 c_{12}^2 \\ \widetilde{K}(X_1, \chi) = c_{01}^2 c_{12}^1 + c_{02}^2 c_{12}^2 \\ \widetilde{K}(X_2, X_2) = 2c_{02}^1 + (c_{12}^1)^2 \\ \widetilde{K}(X_2, \chi) = -c_{02}^1 c_{12}^2 - c_{01}^1 c_{12}^1 \\ \widetilde{K}(\chi, \chi) = 2c_{01}^2 c_{02}^1 + (c_{01}^1)^2 + (c_{02}^2)^2 \end{array} \right.$$

4.3.2. *Cas* $\delta \neq 0$.

Si δ est $\neq 0$, il existe essentiellement (au signe près) une structure sous-riemannienne L -invariante.

On peut par changement de base orthonormale ramener $S = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$ à $S' = \begin{pmatrix} 0 & \delta \\ \delta & 0 \end{pmatrix}$, de trace nulle, de déterminant $-\delta^2 = \Delta$ et de valeurs propres $\pm\delta$.⁵⁰ Dans cette nouvelle base $\{X_1, X_2\}$ on a $c_{01}^1 = c_{02}^2 = 0$ (qui est plus fort que $c_{01}^1 + c_{02}^2 = 0$) et le premier invariant est $\delta = \frac{c_{01}^2 + c_{02}^1}{2}$. Les relations de commutation sont par conséquent⁵¹ :

$$\left\{ \begin{array}{l} [X_1, \chi] = c_{01}^2 X_2 \\ [X_2, \chi] = c_{02}^1 X_1 \\ [X_2, X_1] = c_{12}^1 X_1 + c_{12}^2 X_2 + \chi \end{array} \right.$$

1. Le cas $c_{12}^1 = c_{12}^2 = 0$ correspond au fait que G soit *unimodulaire*. En effet, nous avons vu (cf. par exemple chapitre 5, section 6.2) que G est unimodulaire si

50. Il faut trouver une rotation $R_\varphi = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$ (R_φ est orthogonale, i.e. sa transposée R_φ^t est égale à son inverse R_φ^{-1}) telle que $R_\varphi^t S R_\varphi = S'$. Les équations donnent $a \cos(2\varphi) - b \sin(2\varphi) = 0$ et $b \cos(2\varphi) + a \sin(2\varphi) = \delta$. Autrement dit, φ est l'angle tel que $R_{2\varphi}$ envoie le vecteur (a, b) sur le vecteur $(0, \delta)$ (vecteur qui est bien de même norme puisque $\delta^2 = -\Delta = a^2 + b^2$). Il est bien défini puisque $\delta \neq 0$.

51. Rappelons que par définition $[X_j, X_i] = \sum_{i,j,k} c_{ij}^k X_k$ l'ordre des i, j étant inverse à gauche et à droite.

seulement si les ad_{X_i} sont de trace nulle. Or ici les ad_{X_i} sont de matrices

$$\begin{aligned} ad_{X_1} &= \begin{pmatrix} 0 & -c_{12}^1 & 0 \\ 0 & -c_{12}^2 & c_{01}^2 \\ 0 & -1 & 0 \end{pmatrix}, \\ ad_{X_2} &= \begin{pmatrix} c_{12}^1 & 0 & c_{02}^1 \\ c_{12}^2 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \\ ad_{\chi} &= \begin{pmatrix} 0 & -c_{02}^1 & 0 \\ -c_{01}^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

et la nullité des traces implique bien $c_{12}^1 = c_{12}^2 = 0$, condition qui rend d'ailleurs les conditions $c_{01}^1 c_{12}^1 + c_{02}^1 c_{12}^2 = 0$ et $c_{01}^2 c_{12}^1 + c_{02}^2 c_{12}^2 = 0$ automatiquement satisfaites. Par conséquent $\kappa = \frac{c_{01}^2 - c_{02}^1}{2}$, et, comme $\delta = \frac{c_{01}^2 + c_{02}^1}{2}$, on a $c_{01}^2 = \delta + \kappa$ et $c_{02}^1 = \delta - \kappa$. D'où la forme normale

$$\begin{cases} [X_1, \chi] = c_{01}^2 X_2 = (\delta + \kappa) X_2 \\ [X_2, \chi] = c_{02}^1 X_1 = (\delta - \kappa) X_1 \\ [X_2, X_1] = \chi \end{cases}$$

et les matrices ad_{X_i} se réduisent à des endomorphismes du sous-espace complémentaire de X_i :

$$\begin{aligned} ad_{X_1} &= \begin{pmatrix} 0 & c_{01}^2 \\ -1 & 0 \end{pmatrix} \text{ sur } \text{Span} \{X_2, \chi\}, \\ ad_{X_2} &= \begin{pmatrix} 0 & c_{02}^1 \\ 1 & 0 \end{pmatrix} \text{ sur } \text{Span} \{X_1, \chi\}, \\ ad_{\chi} &= \begin{pmatrix} 0 & -c_{02}^1 \\ -c_{01}^2 & 0 \end{pmatrix} \text{ sur } \text{Span} \{X_1, X_2\}. \end{aligned}$$

La forme de Killing est

$$\begin{cases} \tilde{K}(X_1, X_1) = -2c_{01}^2, \\ \tilde{K}(X_2, X_2) = 2c_{02}^1, \\ \tilde{K}(\chi, \chi) = 2c_{01}^2 c_{02}^1 \\ \tilde{K}(X_i, X_j) = 0 \text{ si } i \neq j. \end{cases}$$

Plusieurs sous-cas sont alors possibles.

(i) $c_{02}^1 = 0$ et donc $\delta = \kappa = \frac{c_{01}^2}{2} > 0$. Les relations deviennent

$$\begin{cases} [X_1, \chi] = 2\delta X_2 \\ [X_2, \chi] = 0 \\ [X_2, X_1] = \chi \end{cases}$$

et montrent que $[\mathcal{G}, \mathcal{G}] = \text{Span}\{X_2, \chi\}$ est de dimension 2. La matrice de ad_{X_1} sur $\text{Span}\{X_2, \chi\}$ est $\begin{pmatrix} 0 & 2\delta \\ -1 & 0 \end{pmatrix}$ dont les valeurs propres imaginaires pures conjuguées sont $\mu, \bar{\mu} = \pm i2\delta$. Ce sont les relations de l'algèbre $\mathfrak{se}(2)$ **Bianchi_VII₀** de 3.3 à travers le changement de base

$$e_1 = \chi, \quad e_2 = \sqrt{2\delta}X_2, \quad e_3 = \frac{1}{\sqrt{2\delta}}X_1.$$

La forme de Killing se réduit à $\tilde{K}(X_1, X_1) = -2c_{01}^2 < 0$, tous les autres $\tilde{K}(X_i, X_j)$ s'annulent.

(ii) $c_{01}^2 = 0$ et donc $\delta = -\kappa = \frac{c_{02}^1}{2} > 0$. Les relations de commutation deviennent

$$\begin{cases} [X_1, \chi] = 0 \\ [X_2, \chi] = 2\delta X_1 \\ [X_2, X_1] = \chi \end{cases}$$

avec $[\mathcal{G}, \mathcal{G}] = \text{Span}\{X_1, \chi\}$ de dimension 2. Ce sont celles de l'algèbre $\mathfrak{sh}(2) = \mathfrak{p}(1, 1)$ de l'algèbre de **Bianchi_VI₀** de 3.3.⁵² La forme de Killing se réduit à $\tilde{K}(X_1, X_1) = 2c_{02}^1 > 0$ tous les autres $\tilde{K}(X_i, X_j)$ s'annulent.

(iii) $c_{01}^2 > 0$, $c_{02}^1 < 0$, $c_{01}^2 > -c_{02}^1$ (car $\delta > 0$), $\delta - \kappa < 0$ (donc $\kappa > 0$ et $\delta + \kappa > 0$). Les relations de commutation

$$\begin{cases} [X_1, \chi] = c_{01}^2 X_2 = (\delta + \kappa) X_2 \\ [X_2, \chi] = c_{02}^1 X_1 = (\delta - \kappa) X_1 \\ [X_2, X_1] = \chi \end{cases}$$

52. Si on échange X_1 et X_2 on obtient $[X_2, \chi] = 0$, $[X_1, \chi] = 2\delta X_2$, $[X_1, X_2] = \chi$, c'est-à-dire le cas précédent $\mathfrak{se}(2)$ mais avec $[X_1, X_2]$ changé de signe. C'est la différence entre les métriques euclidiennes et hyperboliques de \mathbb{R}^2 .

montrent que $[\mathcal{G}, \mathcal{G}] = \mathcal{G}$ est de dimension 3 et que les 3 matrices ad_{X_i} sont inversibles et de valeurs propres

$$\mu_{1\pm} = \pm\sqrt{-c_{01}^2} \text{ pour } \text{ad}_{X_1} ,$$

$$\mu_{2\pm} = \pm\sqrt{c_{02}^1} \text{ pour } \text{ad}_{X_2} ,$$

$$\mu_{\chi\pm} = \pm\sqrt{c_{02}^1 c_{01}^2} \text{ pour } \text{ad}_{\chi} .$$

Comme $c_{01}^2 > 0$, $c_{02}^1 < 0$, toutes ces racines sont imaginaires pures et l'on retrouve l'algèbre $\mathfrak{su}(2)$ **Bianchi_IX** de 3.3. La forme de Killing donne trois $\tilde{K}(X_i, X_i) < 0$:

$$\begin{cases} \tilde{K}(X_1, X_1) = -2c_{01}^2 \\ \tilde{K}(X_2, X_2) = 2c_{02}^1 \\ \tilde{K}(\chi, \chi) = 2c_{01}^2 c_{02}^1. \end{cases}$$

Pour que $-\tilde{K}$ fasse de $\{X_1, X_2\}$ une base orthonormée, il faudrait $c_{01}^2 = 1$ et $c_{02}^1 = -1$ mais cela exigerait $\delta = 0$ (cf. plus bas).

- (iv) $c_{01}^2 < 0$, $c_{02}^1 > 0$, $c_{02}^1 > -c_{01}^2$ (car $\delta > 0$), $\kappa < 0 < \delta$, $\delta - \kappa > 0$, $\delta + \kappa < 0$. Les valeurs propres $\mu_{1,\pm} = \pm\sqrt{-c_{01}^2}$, $\mu_{2,\pm} = \pm\sqrt{c_{02}^1}$ sont réelles, $\mu_{\chi,\pm} = \pm\sqrt{c_{02}^1 c_{01}^2}$ sont imaginaires pures. En prenant celles réelles, on retrouve l'algèbre $\mathfrak{sl}(2)$ **Bianchi_V** de 3.3. La forme de Killing donne

$$\begin{cases} \tilde{K}(X_1, X_1) = -2c_{01}^2 > 0 \\ \tilde{K}(X_2, X_2) = 2c_{02}^1 > 0 \\ \tilde{K}(\chi, \chi) = 2c_{01}^2 c_{02}^1 < 0 \end{cases}$$

Pour que \tilde{K} fasse de $\{X_1, X_2\}$ une base orthonormée, il faudrait $c_{01}^2 = -1$ et $c_{02}^1 = 1$ mais cela exigerait $\delta = 0$ (cf. plus bas).

- (v) $c_{01}^2 > 0$, $c_{02}^1 > 0$, $\delta + \kappa > 0$ et $\delta - \kappa > 0$, les valeurs propres $\mu_{2,\pm} = \pm\sqrt{c_{02}^1}$ et $\mu_{\chi,\pm} = \pm\sqrt{c_{02}^1 c_{01}^2}$ sont réelles tandis que les $\mu_{1,\pm} = \pm\sqrt{-c_{01}^2}$ sont imaginaires pures. En prenant celles réelles, on retrouve encore l'algèbre $\mathfrak{sl}(2)$ **Bianchi_V** de 3.3. La forme de Killing donne

$$\begin{cases} \tilde{K}(X_1, X_1) = -2c_{01}^2 < 0 \\ \tilde{K}(X_2, X_2) = 2c_{02}^1 > 0 \\ \tilde{K}(\chi, \chi) = 2c_{01}^2 c_{02}^1 > 0 \end{cases}$$

2. On peut aussi satisfaire les contraintes

$$c_{01}^1 c_{12}^1 + c_{02}^1 c_{12}^2 = c_{02}^1 c_{12}^2 = 0,$$

$$c_{01}^2 c_{12}^1 + c_{02}^2 c_{12}^2 = c_{01}^2 c_{12}^1 = 0$$

non plus avec $c_{12}^1 = c_{12}^2 = 0$ mais avec $c_{12}^1 = 0$ et $c_{02}^1 = 0$ ou, symétriquement, $c_{12}^2 = 0$ et $c_{01}^2 = 0$, $c_{02}^1 = 0$ et $c_{01}^2 = 0$ étant impossible car $\delta = \frac{c_{01}^2 + c_{02}^1}{2} \neq 0$.

- (i) Si $c_{12}^1 = 0$ et $c_{02}^1 = 0$ (et donc $c_{01}^2 \neq 0$), alors $\delta = \frac{c_{01}^2}{2} > 0$, $\kappa = \frac{c_{01}^2}{2} - (c_{12}^2)^2$ et $\delta - \kappa = (c_{12}^2)^2 \geq 0$. Les relations de commutation deviennent

$$\begin{cases} [X_1, \chi] = c_{01}^2 X_2 \\ [X_2, \chi] = 0 \\ [X_2, X_1] = c_{12}^2 X_2 + \chi \end{cases}$$

Elles impliquent que $[\mathcal{G}, \mathcal{G}] = \text{Span}\{X_2, \chi\}$ est de dimension 2 et que $\text{ad } X_1$ agit sur $\text{Span}\{X_2, \chi\}$ avec la matrice $\begin{pmatrix} -c_{12}^2 & c_{01}^2 \\ -1 & 0 \end{pmatrix}$ de trace $-c_{12}^2$ et de déterminant $c_{01}^2 \neq 0$. Les ad_{X_i} complets sont

$$\text{ad}_{X_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -c_{12}^2 & c_{01}^2 \\ 0 & -1 & 0 \end{pmatrix} \text{ de trace } -c_{12}^2 ,$$

$$\text{ad}_{X_2} = \begin{pmatrix} 0 & 0 & 0 \\ c_{12}^2 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ de trace } 0 ,$$

$$\text{ad}_{\chi} = \begin{pmatrix} 0 & 0 & 0 \\ -c_{01}^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ de trace } 0 .$$

Il n'y a donc pas unimodularité si $c_{12}^2 \neq 0$.

La forme de Killing est $\tilde{K}(X_1, X_1) = -2c_{01}^2 + (c_{12}^2)^2$, tous les autres $\tilde{K}(X_i, X_j)$ s'annulant, ce qui correspond à la classe **Bianchi VII** de la section 3.3 avec $c_{01}^2 = 1$ et $c_{12}^2 = \ell$.

On peut alors faire, exactement comme dans la section 3.3, l'analyse spectrale de ad_{X_1} et changer de base dans $[\mathcal{G}, \mathcal{G}] = \text{Span}\{X_2, \chi\}$ de façon à mettre ad_{X_1} sous forme normale.

Si de plus $c_{12}^2 = 0$, et donc $\delta = \kappa$ et $\text{trace}(\text{ad } X_1) = 0$, il y a unimodularité et l'on retrouve l'algèbre $\mathfrak{se}(2)$ du groupe unimodulaire $SE(2)$.

- (ii) Si $c_{12}^2 = 0$ et $c_{01}^2 = 0$ (et donc $c_{02}^1 \neq 0$), alors $\delta = \frac{c_{02}^1}{2} > 0$, $\kappa = -\frac{c_{02}^1}{2} - (c_{12}^1)^2$ et $\delta + \kappa = -(c_{12}^1)^2 \leq 0$. Les relations de commutation deviennent

$$\begin{cases} [X_1, \chi] = 0 \\ [X_2, \chi] = c_{02}^1 X_1 \\ [X_2, X_1] = c_{12}^1 X_1 + \chi \end{cases}$$

Elles impliquent que $\text{ad } X_2$ agit sur $[\mathcal{G}, \mathcal{G}] = \text{Span} \{X_1, \chi\}$ avec la matrice $\begin{pmatrix} c_{12}^1 & c_{02}^1 \\ 1 & 0 \end{pmatrix}$ de trace c_{12}^1 et de déterminant $-c_{02}^1$.

Si de plus $c_{12}^1 = 0$ et donc $\delta = -\kappa$ et $\text{trace}(\text{ad } X_2) = 0$, on retrouve l'algèbre $\mathfrak{sh}(2)$ du groupe unimodulaire $SH(2)$.

4.3.3. Cas $\delta = 0$.

Si en revanche $\delta = 0$, i.e. $c_{01}^2 = -c_{02}^1 = c$, on peut prendre $c_{01}^1 = c_{02}^2 = 0$ et il reste $\kappa = \frac{c_{01}^2 - c_{02}^1}{2} - (c_{12}^1)^2 - (c_{12}^2)^2 = c - (c_{12}^1)^2 - (c_{12}^2)^2$. En fait on peut également se ramener à $c_{12}^1 = c_{12}^2 = 0$, et donc à la forme normale

$$\begin{cases} [X_1, \chi] = c_{01}^2 X_2 = \kappa X_2 \\ [X_2, \chi] = c_{02}^1 X_1 = -\kappa X_1 \\ [X_2, X_1] = \chi \end{cases}$$

1. Si $\kappa = 0$, \mathcal{G} est isomorphe à l'algèbre de Heisenberg \mathfrak{h} avec les relations de commutation

$$[X_2, X_1] = \chi, [X_1, \chi] = 0, [X_2, \chi] = 0$$

redonnant celle de 3.3 **Bianchi I**

$$[e_1, e_2] = e_3, [e_1, e_3] = 0, [e_2, e_3] = 0$$

à travers l'identification $X_1 \rightsquigarrow e_2, X_2 \rightsquigarrow e_1, \chi \rightsquigarrow e_3$. G est donc isomorphe au groupe de Heisenberg avec sa métrique sous-riemannienne standard.

2. Si $\kappa > 0$, $[\mathcal{G}, \mathcal{G}] = \mathcal{G}$. On retrouve le $\mathfrak{su}(2)$ de 1.(iii) mais, comme nous l'avons vu plus haut, avec $\delta = 0$ et avec l'opposé de sa forme de Killing.
3. Si $\kappa < 0$, $[\mathcal{G}, \mathcal{G}] = \mathcal{G}$. On retrouve le $\mathfrak{sl}(2)$ de 1.(iv) mais, comme nous l'avons vu plus haut, avec $\delta = 0$ et avec sa forme de Killing. Le lien avec 3.3 peut se faire en prenant

$$\chi = e_1 - e_3, X_1 = \frac{1}{\sqrt{2}}(e_1 + e_3), X_2 = -\frac{1}{\sqrt{2}}e_2.$$

En effet les relations de commutations

$$[e_1, e_2] = 2e_1, [e_1, e_3] = -e_2, [e_2, e_3] = 2e_3$$

impliquent

$$\begin{cases} [X_1, \chi] = \frac{1}{\sqrt{2}}[e_1 + e_3, e_1 - e_3] = \frac{1}{\sqrt{2}}(2e_2) = -2X_2 = c_{01}^2 X_2 \\ [X_2, \chi] = -\frac{1}{\sqrt{2}}[e_2, e_1 - e_3] = -\frac{1}{\sqrt{2}}(-2e_1 - 2e_3) = 2X_1 = c_{02}^1 X_1 \\ [X_2, X_1] = -\frac{1}{2}[e_2, e_1 + e_3] = -\frac{1}{2}(-2e_1 + 2e_3) = \chi \end{cases}$$

correspondant à $c_{01}^2 = -2 < 0$ et $c_{02}^1 = 2 > 0$ qui est un rescaling.

4. Le cas 1.(v) est incompatible avec $\delta = 0$ si $\kappa \neq 0$ car alors $\kappa > 0$ et $-\kappa > 0$.
5. Avec $\delta = 0$ et $\kappa = -1 < 0$ on trouve aussi l'algèbre $\mathcal{G} = \mathfrak{l}_2 \oplus (\mathbb{R}, +)$ de 3.3 Bianchi_III pour laquelle $[\mathcal{G}, \mathcal{G}]$ est de dimension 1.

Il peut exister jusqu'à trois structures non isométriques possédant les mêmes invariants (et donc (δ, κ) n'est pas un système complet d'invariants). Mais si G est unimodulaire, alors il n'existe qu'une seule structure sous-riemannienne pour (δ, κ) donnés.

*
* *

En conclusion, ces deux dernières sections sur la structure et les métriques des algèbres de Lie de dimension 3 montrent comment nos deux modèles de base \mathbb{V}_J et \mathbb{V}_S s'insèrent dans un ensemble de possibilités. Leur vertu pédagogique est aussi de montrer que, si l'on comprend suffisamment bien la structure abstraite d'une classe particulière d'objets, on peut dans les bons cas en fournir une classification exhaustive au moyen de calculs certes nombreux mais relativement simples.

Partie II : Géométrie sous-riemannienne

Contours illusoires modaux et modèles variationnels

Nous allons maintenant aborder dans ce chapitre, le second exemple invoqué dans l'*Introduction* du Vol I et qui concerne un phénomène gestaltiste beaucoup plus intrigant que l'intégration de contours que nous avons modélisée jusqu'ici. En un mot, il s'agit d'appliquer la structure de contact de *V1* à l'éclaircissement du problème des *contours subjectifs* à la Kanizsa, contours qui sont des cas typiques d'une complétion ("filling-in") de données sensorielles lacunaires fondée sur la loi gestaltiste de "bonne continuation".

Nous donnerons des indications bibliographiques au fur et à mesure. Mais citons d'emblée par exemple le recueil [446] dirigé par Susan Petry et Glenn Meyer en 1987 où le lecteur trouvera présentées beaucoup de recherches de base.

1. Complétion modale *VS* complétion amodale

En psychophysique, tous les contours complétés par le sujet percevant en l'absence de différences de contraste objectives dans les stimuli sont appelés subjectifs ou "anomaux". On distingue alors parmi eux (cf. figure 1) :

- Les contours qui sont effectivement perçus : le sujet voit une différence de contraste. Ces contours illusoires sont également appelés *modaux*. Ils sont de nature perceptive. Ils s'accompagnent en général de la perception d'une brillance accrue de la *surface* délimitée par le contour. Le rapport des deux phénomènes de contraste et de surbrillance n'est pas complètement élucidé, mais il semble qu'ils résultent de processus distincts (cf. Dresp et Bonnet [152] et la dualité "Boundary Contour System / Featural Contour System" chez Grossberg et Mingolla [230], [231], [232]). Dans ce qui suit, nous nous concentrons sur les contours eux-mêmes, en ne retenant du phénomène de "surbrillance" que la preuve que les contours illusoires, bien que sensoriellement inexistantes, sont néanmoins perceptivement "réels" puisqu'ils peuvent fonctionner comme des *bords* de surfaces.
- Les contours qui sont seulement conçus et imaginés : le sujet ne perçoit pas de différence de contraste mais il extrapole mentalement le contour. C'est le cas en général lors de l'occlusion (ou masquage) d'un objet par un autre. On parle alors de contours *amodaux*. Ils sont de nature plus cognitive.

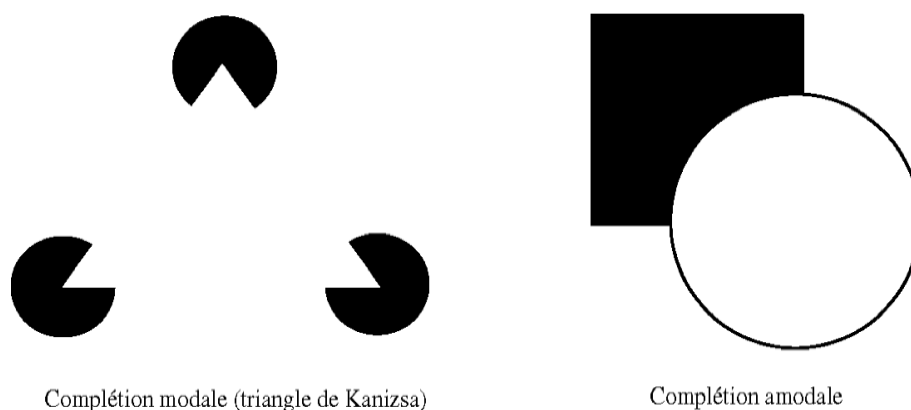


FIGURE 1. Contours subjectifs modaux et amodaux.

Une question est de savoir si ces deux types de contours illusoires doivent être considérés comme résultant ou non des mêmes mécanismes perceptifs, qu'ils soient de bas ou de haut niveau, les contours amodaux ajoutant une dimension cognitive. Cependant, il semble plutôt y avoir continuité que rupture entre complétions modales et amodales (voir les expériences de Shipley et Kellman [498]), ce qui a amené plusieurs spécialistes à les traiter comme un seul et même phénomène (Finkel et Edelman [183], Peterhans et von der Heydt [409]). Un argument assez convaincant est donné par Kellman et Shipley [286] à travers un exemple qui semble être intermédiaire entre les deux modes de complétion, celui des figures “à scission spontanée”. Bien que leur surface soit homogène, elles tendent à être perçues comme composées de deux éléments distincts : le premier est vu comme placé au-dessus de l'autre, et complété de façon *modale*, alors que le second est également vu comme une unité figurale mais la partie cachée n'est complétée que de façon *amodale*. Cependant, après un certain temps de fixation, les positions s'inversent : le second objet est alors complété modalement et le premier amodalement. À partir de ce moment, le basculement se produit régulièrement (cf. figure 2). Selon Kellman et Shipley, c'est la façon dont le sujet ordonne les différentes surfaces selon la profondeur qui déciderait de leur mode de complétion.

La facilité avec laquelle se produisent les transitions d'un mode de complétion à l'autre plaide en faveur d'une proximité de leurs mécanismes, au moins dans les contextes géométriques simples.

2. Les contours illusoires modaux

Nous nous focaliserons à partir de maintenant sur les contours illusoires *modaux*. Ils constituent une fenêtre d'un intérêt considérable sur les mécanismes neuronaux de la vision. Comme le note Janine Mendola et ses co-auteurs dans [361],

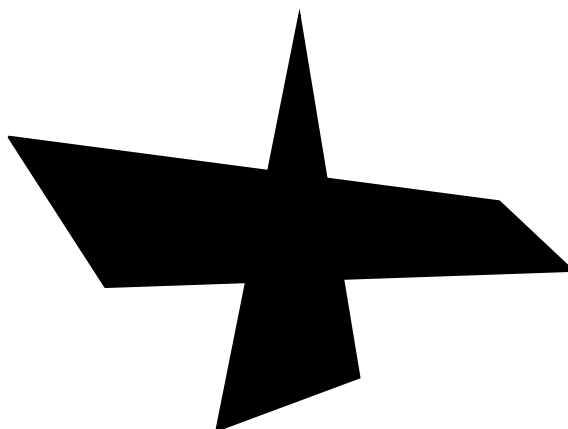


FIGURE 2. Un exemple de figure à scission spontanée.

“They may reveal brain mechanisms that contribute critically to object perception.”

De même, selon Peter Kok et Floris de Lange [302],

“Illusory shape perception provides an ideal test bed, as the illusory shape results in both unexpected absence of visual input [at the location where the shape is perceived but retinal input is absent] and expected presence of visual input [at the location where the shape provides an explanation for the bottom-up input]” ([302], p. 1531)

Ils peuvent être étonnamment complexes. Un exemple omniprésent dans la littérature est celui du chien dalmatien de la figure 3 (Richard Gregory [222], 1970) que l’on reconstruit facilement alors que presque tous les bords sont absents du stimulus.

Cette capacité de reconstruction de formes très largement lacunaires est fondamentale pour la survie de nombreuses espèces dont les proies sont camouflées par exemple dans des feuillages.

2.1. Exemples de contours subjectifs modaux

2.1.1. *Contours illusoires du premier ordre.*

Commençons par les contours illusoires (CI) modaux simples comme le triangle de Kanizsa à gauche de la figure 1. Ils sont l’une des manifestations les plus spectaculaires des propriétés gestaltistes de complétion à très longue portée de données sensorielles lacunaires. Les conditions aux limites définies par les pacmen¹ engendrent

1. Le terme “pacman” vient d’un célèbre jeu d’arcade japonais où un disque jaune (le “pacman”, initialement nommé “puck-man”) avec un secteur ouvert variable (la “bouche”) doit naviguer dans un labyrinthe en y “mangeant” des “pac-dots” et en y évitant divers fantômes “ennemis”.



FIGURE 3. Image d'un chien dalmatien dont on reconstruit les bords absents du stimulus grâce à des contours illusoires modaux. Cf. Gregory [222].

des CIs qui interpolent entre elles. Ces contours sont donc de type “prolongement” et “bonne continuation” au sens de la théorie de la Gestalt. On pourrait dire qu'ils sont du “premier ordre”. Leur netteté et leur vivacité – leur “saillance” perceptive – peuvent être accrues en utilisant par exemple (comme Diego Varin dès 1971) des pacmen constitués d'arcs de cercles concentriques ou encore l'effet “néon” qui consiste à colorer les pacmen et à observer la diffusion de couleur à l'intérieur de la surface définie par les contours illusoires. Cette diffusion chromatique confirme la surbrillance de la surface dans le cas achromatique.² La figure 4 donne un exemple de carré de Kanizsa avec effet Varin et néon.

L'effet néon (aussi appelé “watercolor illusion”) est particulièrement saisissant. Comme le soulignent Baingio Pinna, John Werner et Lothar Spillmann, dans [448], il concerne un

“long-range color spreading from the inner edge of an outline figure onto the enclosed surface area”

et renforce le principe gestaltiste (remontant à Wertheimer comme nous l'avons vu dans la section 4.3 du chapitre 2) de l'appartenance des bords à la figure et non pas au fond. La figure 5 en donne un exemple frappant. Il est remarquable que cet aspect de la ségrégation figure/fond puisse continuer à s'appliquer à des CIs et à des

2. Nous verrons à la section 2.3 que l'on peut encore accentuer la saillance en faisant varier dynamiquement l'angle d'ouverture des pacmen.

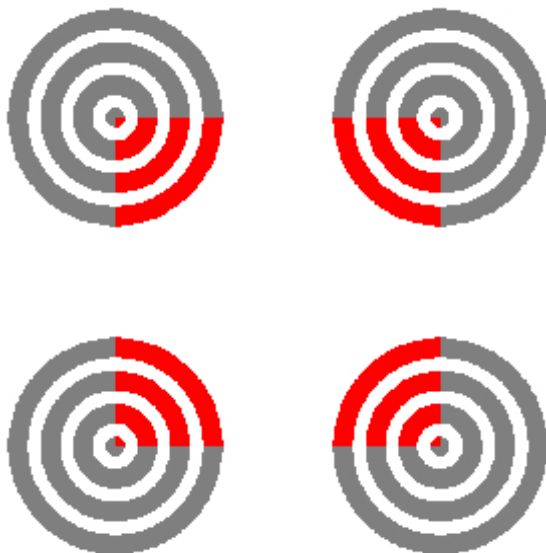


FIGURE 4. Un exemple de contours illusoires modaux à la Kanizsa avec “effet néon” : les contours subjectifs fonctionnent comme des bords pour la diffusion de couleur à l’intérieur du carré.

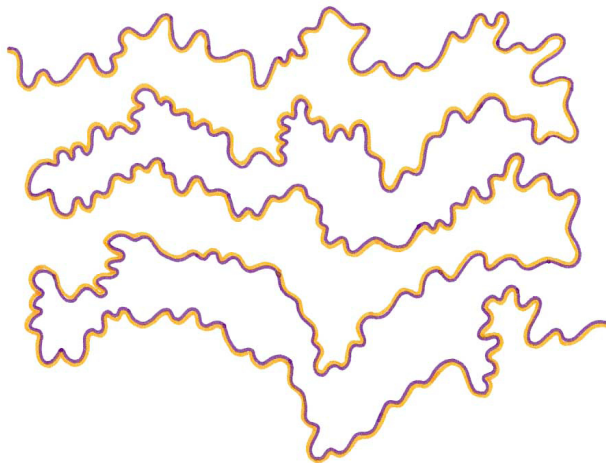


FIGURE 5. L’effet néon ou la “watercolor illusion”. (D’après Pinna *et al.* [448]).

surfaces virtuelles. On trouvera une synthèse de cet effet dans l’article [76] de Paola Bressan, Ennio Mingolla, Lothar Spillmann et Takeo Watanabe.

Les CIs peuvent même être *courbes*. La figure 6 en donne un exemple.

Ce fait est épistémologiquement important car il montre qu’il existe des déformations *continues* de CIs et qu’on ne peut donc pas les expliquer au moyen d’un catalogue fini (et a fortiori restreint) de formes stockées dans quelque mémoire de formes,

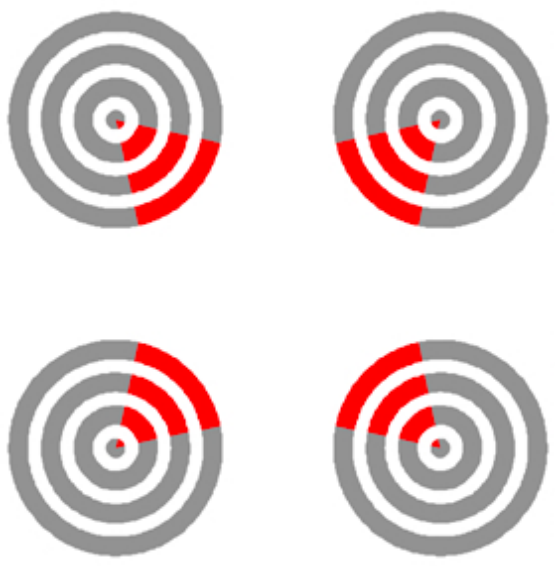


FIGURE 6. Un autre exemple de contours illusoires modaux à la Kanizsa avec “effet néon”. Les contours illusoires y sont courbes.

ni même d’un catalogue fini de routines de construction de formes géométriques simples (triangles, rectangles, cercles, etc.). En fait, toute courbe lacunaire, aussi sinueuse soit elle, peut-être complétée par des CIs. Il faut donc que les mécanismes de complétion soient très puissants. Nous allons voir qu’une possibilité de répondre à ce problème repose sur le recours à des *principes variationnels*.

Les contours illusoires sont encore plus nets en vision *stéréoscopique*. On sait depuis les travaux pionniers de Béla Julesz que la stéréopsie vient du fait que les petites disparités binoculaires sont interprétées par le système visuel comme des indices de profondeur. Un exemple typique et standard est celui de la figure 7 où, dans une grille de points aléatoires, on décale d’un cran les points d’un petit carré central. Le fusionnement des deux images interprète cette disparité comme le fait que le carré central flotte au-dessus du fond. On peut même, comme dans la figure 8, mêler cette disparité de distributions aléatoires de points avec une déformation de figure de façon à faire émerger une structure tridimensionnelle.

La stéréopsie peut s’appliquer aux CIs comme le montre la figure 9).

Certains effets sont remarquables comme dans la figure 10. La perception stéréoscopique de deux CIs courbes bien calculés les fait apparaître comme les bords droits d’une surface illusoire courbe.

Dans tous ces exemples, les protocoles expérimentaux peuvent faire varier de nombreux paramètres comme la taille des pacmen, la vitesse de leur dynamique



FIGURE 7. Un exemple de stéréopsie pure dû à Béla Julesz. Dans une grille de points aléatoires, on décale d'un cran les pixels d'un petit carré central. Le fusionnement des deux images interprète cette disparité comme le fait que le carré central flotte au-dessus du fond.

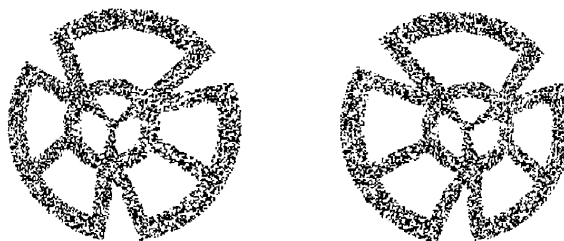


FIGURE 8. Un exemple de stéréopsie mêlant une disparité binoculaire entre distributions aléatoires de points avec une déformation de figure.



FIGURE 9. Contours illusoires en vision stéréoscopique. Si l'on fusionne les deux images, le triangle illusoire semble flotter au-dessus des trois pacmen.

d'ouverture, le contraste des figures, etc. D'où un nombre considérable de résultats expérimentaux.

Le fait que les CIs puissent être des *bords* de surfaces illusoires est fondamental puisqu'il montre au niveau de l'expérience perceptive la plus immédiate que les CIs sont *fonctionnellement réels*. Cela n'est pas étonnant car la détection de bord est à la base de la perception. Il est donc compréhensible que les mécanismes perceptifs

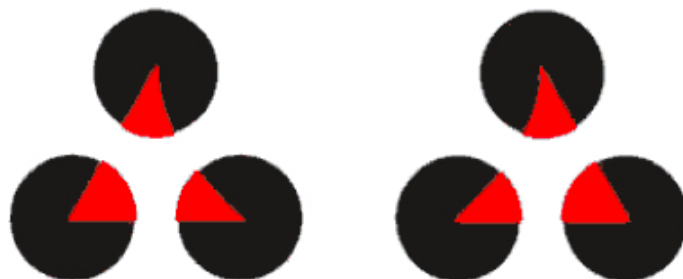


FIGURE 10. La perception stéréoscopique de deux CIs courbes bien calculés les fait apparaître comme les bords droits d’une surface illusoire courbe.

de construction de bords puissent être activés même par des stimuli très lacunaires. Des résultats d’imagerie par électroencéphalographie (EEG) confirment bien que la reconstruction neuronale de contours illusoires correspond à la *même* activité cérébrale que la détection de contours réels. La figure 11, due à Catherine Tallon-Baudry [522], en témoigne. On présente aux sujets des stimuli cohérents (triangles illusoires ou réels)

“leading to a coherent percept through a bottom-up feature binding process”.

Si l’on considère la

“time–frequency power of the EEG at electrode Cz (overall average of 8 subjects), in response to the illusory triangle (top) and to the no-triangle stimulus (bottom)”,

on observe

“two successive bursts of oscillatory activities [...]. A first burst at about 100ms and 40Hz. It showed no difference between stimulus types. A second burst around 280ms and 30-60Hz. It is most prominent in response to coherent stimuli.”

Or la seconde bouffée d’activation, qui correspond à l’intégration des données locales en une forme globale, est la même pour les deux types de triangles.

2.1.2. Les deux types d’intégration.

Dans les contours illusoires à la Kanizsa, le passage du local au global s’opère sur de très longues distances à l’échelle neuronale et leur construction est véritablement *globale*, holistique au sens méréologique. La saillance (le pop-out des CIs) vient d’une *intégration* des inputs sensoriels locaux réels en une unité figurale perceptuelle globale. Si les inputs locaux réels sont “complets”, la saillance est celle d’une forme

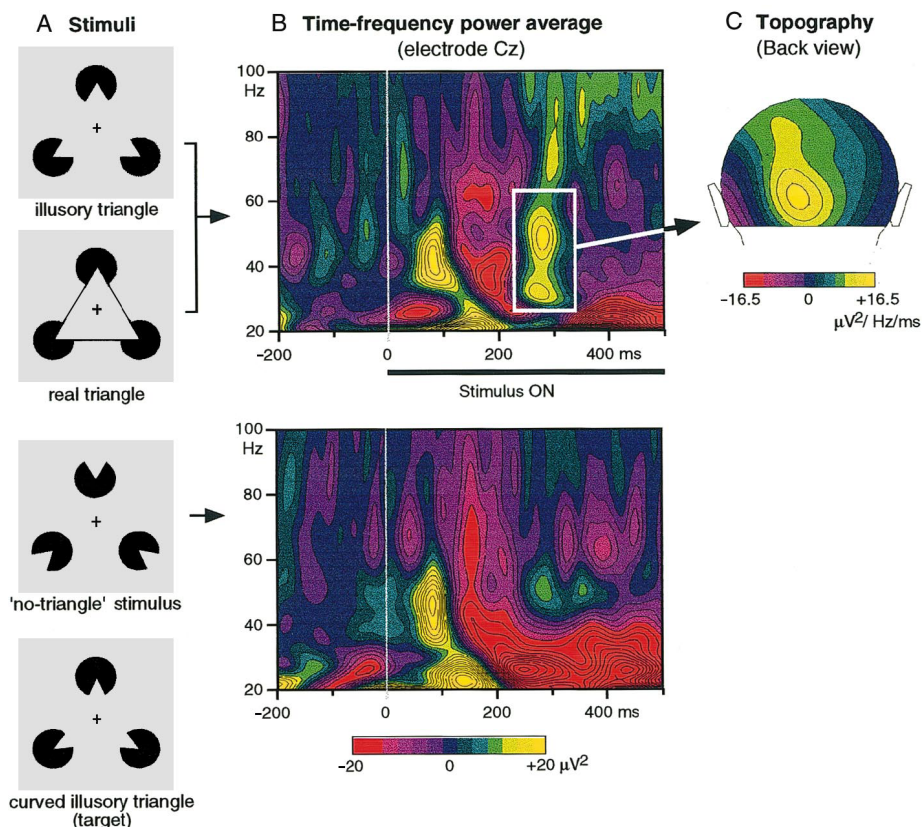


FIGURE 11. *En bas*. Réponse à des stimuli composés de trois pacmen sans cohérence globale. On observe un pic d'activité autour de 100ms. *En haut*. Réponse à des stimuli composés de trois pacmen avec cohérence globale (triangle de Kanizsa et triangle réel). On observe en plus du pic d'activité autour de 100ms un second pic autour de 280ms : il correspond à la constitution des contours illusoires dans le triangle de Kanizsa. (D'après Tallon-Baudry [522]).

globale réelle. S'ils sont "incomplets" (lacunaires), elle est celle d'une forme globale virtuelle. Sur le plan explicatif, toute la question est donc de savoir ce que signifie ici le terme d'"intégration" au niveau neuronal.

Il y a essentiellement deux types d'intégration que l'on peut envisager.

1. D'une part il y a les mécanismes d'intégration de type "équation différentielle" qui consistent à considérer les inputs comme des conditions initiales et les CIs comme des résultats de l'intégration. Nous appellerons ce type "intégration fonctionnelle". Il est omniprésent en mathématiques et en physique. En neurosciences, il présuppose des architectures fonctionnelles appropriées comme celles que nous avons étudiées dans les chapitres précédents. Il est de bas niveau, s'effectue au niveau des aires rétinotopiques primaires comme V1 et

V2 et envoie ses outputs aux autres aires post-V2 de façon “feedforward” et ascendante (“bottom-up”)

2. Mais d’autre part il y a aussi des mécanismes d’intégration qui consistent à traiter globalement une forme par des neurones visuels d’aires hiérarchiquement supérieures possédant de larges champs récepteurs.³ Nous appellerons ce type “intégration inférentielle”. Il est de plus haut niveau, présuppose des mémoires de formes et rétro-agit par “feedback” de façon descendante (“top-down”) et “inférentielle” sur les aires inférieures. Les inputs lacunaires élicitent une forme globale qui permet de les regrouper (“grouping”) comme autant de fragments d’une unité figurale. La forme globale n’est donc pas construite par une intégration au sens classique à partir de conditions initiales mais activée comme une hypothèse à confirmer. Elle module alors les activités sous-jacentes des aires primaires comme V1 et V2. Si ces activités confirment l’hypothèse figurale elles se trouvent renforcées mais elles peuvent être en partie inhibées si elles ne confirment pas l’hypothèse figurale et si le mismatch reste au-dessous d’un certain seuil. Si le mismatch est trop grand, alors il élicite une autre hypothèse. Il s’agit de théories dites “génératives” et “prédictives”. Certains résultats vont dans ce sens mais de nombreux autres vont dans le sens de mécanismes ascendants. On pourra consulter par exemple le travail de Peter Kok et Floris de Lange [302] qui présente les deux perspectives pour V1.

Le problème du caractère global (holistique) des CIs implique donc l’ensemble de l’épistémologie de la perception. On y retrouve l’opposition classique entre les théories de “bonne continuation” et les théories à la Helmholtz faisant des CIs des “hypothèses” pour interpréter les données sensorielles par des inférences inconscientes. Comme l’expliquent bien Mark Haiko, Ennio Mingolla et David Somers dans [240], trois mécanismes dominent les théories des CIs (nous en donnerons plus bas des exemples concrets) :

1. les mécanismes d’interpolation (Dresp, Grossberg, Mingolla, Kellmann, Shipley, etc.) selon lesquels un CI connecte une paire de pacmen inducteurs ;
2. les mécanismes d’extrapolation (Field, Hayes, Hess, Hoffman, Kellman, Shipley, etc.) selon lesquels un inducteur est prolongé de façon aléatoire ;
3. les mécanismes de “feedback figural” (Hoffman, Gregory, Grossberg, Kourtzi, Lee, Mendola, Rubin, Stanley, etc.) selon lesquels les fragments de contour donnés par l’input sensoriel induisent une perception de surface qui, par feedback, induit la complétion modale.

3. On les appelle parfois de façon imagée des “neurones grand-mère” car ils ont été introduits pour les aires où s’effectue la reconnaissance des visages, comme le visage de sa grand-mère.

Mais en fait, les mécanismes feed-forward de coopération locale 1 et 2 sont très proches, une interpolation entre deux inducteurs consistant en la jonction entre deux extrapolations de directions inverses émanant des deux inducteurs.

Dans la section 5.2 du chapitre 5 du Vol I nous avons exposé la théorie du champ d'association de David Field, Anthony Hayes et Robert Hess pour l'intégration des contours et nous l'avons également évoquée dans ce volume lors de nos rappels introductifs et à la section 2.3 du chapitre 3. Il est pertinent d'y revenir ici encore une fois. En effet, ces expériences portent sur le pop-out, la saillance perceptive, de chaînes (a_i, p_i) de petits segments de centres a_i et d'orientations p_i alignés mais séparés par des vides (en quelque sorte une courbe "en pointillé"). La saillance vient d'une complétion modale connectant les segments le long d'une courbe globale au moyen de morceaux de contours illusoires. Il y a donc regroupement ("grouping") et intégration. Mais les auteurs insistent sur le fait qu'il s'agit d'un mécanisme d'intégration de bas niveau qui ne s'effectue pas au moyen d'un regroupement dû à de grands champs récepteurs. Rappelons leur thèse :

"(...) the points along the length of a curved edge can be linked together according to a set of local rules that allow the edge to be seen as a whole, even though different components of the edge are detected by independent mechanisms." ([182], p. 174)

Et ils ajoutent

"In our stimuli, there does not exist any 'global' feature that allows the path to be segregated from the background. (...) Our results imply that the path segregation is based on local processes which group features locally." ([182], p. 191)

Ces mécanismes d'intégration locaux et de bas niveau du "champ d'association" reposent sur l'implémentation neuronale d'une condition d'intégrabilité par une corrélation entre position et orientation. Rappelons les citations clé :

"Elements are associated according to *joint constraints* of position and orientation." ([182], p. 187)

"There is a *unique link* between the relative positions of the elements and their relative orientations." ([182], p. 181)

La chaîne d'éléments de contact $c_i = (a_i, p_i)$ engendre une courbe perceptivement saillante si et seulement si les orientations p_i sont tangentes à une courbe γ interpolant entre les positions a_i . Il s'agit d'une interpolation par "intégration fonctionnelle" dont les CIs du premier ordre fournissent une version beaucoup plus spectaculaire avec une complétion modale à très longue distance.

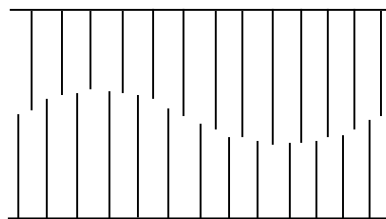


FIGURE 12. Effet de peigne (“comb effect” et “abutting grating illusion”) produit par un alignement de points d’arrêt. Il fait intervenir l’aire V2.

2.1.3. *Contours illusoires du second ordre.*

Il existe un autre type de CIs également très étudiés, que l’on pourrait appeler de “second ordre”. Il s’agit d’“effets de peigne” (“comb effect”) où des lignes se terminent par des points d’arrêt (“end-points”) alignés et où le CI interpole entre ces extrémités comme si les directions orthogonales aux lignes étaient activées, ce qui exige des cellules “end-stopped”. La figure 12 (reprenant la figure 5.31 de la section 5.8. du Vol I, chapitre 5) en donne un exemple. On parle de “abutting grating illusion” (cf. par exemple l’article [507] de Manuel Soriano, Lothar Spillmann et Michael Bach). Là encore, on peut faire varier plusieurs paramètres comme l’épaisseur des lignes, leur contraste ou leur périodicité, ce qui conduit à de très nombreuses expériences.

Pour de nombreux auteurs, comme par exemple Anna Roe, il s’agit là d’exemples de contours inférés obtenus par regroupement (“grouping”), descendants (“top-down”), opérant par feedback, et de nature plus cognitive :

“Inferred contours are not defined by luminance contrast but rather by more global features that are perceived only by grouping multiple cues across space. They include occluded contours, texture element borders, and have been referred to as higher-order contours, illusory contours, cognitive contours, and anomalous contours.” ([469], p. 121)

Mais on peut aussi penser qu’ils sont, bien qu’un peu plus complexes, de même nature, d’aussi bas niveau et tout aussi feedforward et ascendants (“bottom-up”) que les CIs du premier ordre. Si l’on considère en effet, comme dans la figure 13, un triangle de Kanizsa dont on renforce les contours virtuels par la partie des bords réels d’un triangle qui serait sous-jacent, on constate que le renforcement est induit par les extrémités de ces bords. Les cellules “end-stopped” qui détectent ces extrémités activent localement la direction orthogonale. Mais la cohérence globale du stimulus impose en fait l’activation d’une direction oblique non orthogonale.⁴

4. Nous reviendrons à la section 2.4 sur l’utilisation de détection d’extrémités pour le positionnement de contours illusoires.

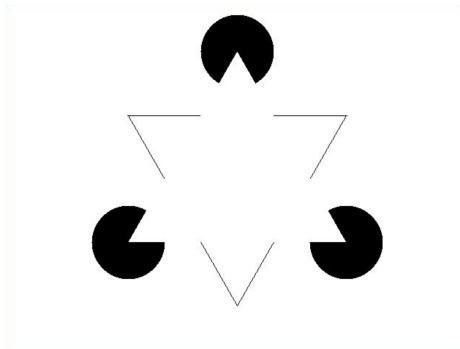


FIGURE 13. Les contours virtuels d'un triangle de Kanizsa sont renforcés par les extrémités des bords réels d'un triangle vu comme sous-jacent. Localement, c'est la direction orthogonale qui devrait être activée aux extrémités. Mais la cohérence globale du stimulus impose l'activation d'une direction oblique non orthogonale.

2.1.4. Certains corrélats neuronaux des CIs.

On voit ainsi la richesse expérimentale des CIs et le nombre de résultats de psychophysique et d'imagerie auxquels ils ont donné lieu. De nombreuses expériences, soit par enregistrements électrophysiologiques d'activités neuronales, soit par IRMf, ont également été menées pour découvrir et modéliser leurs corrélats neuronaux. Nous allons en donner un très bref aperçu en renvoyant au Vol 1 pour les concepts neurophysiologiques que nous utilisons.

L'influence des divers paramètres contrôlant la formation des CIs a été peu explorée jusqu'à la fin des années 1980, comme le déploraient Petry et Gannon [445]. Certains auteurs ont ensuite entrepris de combler cette lacune, mais les résultats disponibles dans les années 1990 n'étaient pas encore très nombreux, comme en témoignent des panoramas comme ceux de Peterhans et von der Heydt [409] ou Spillman et Dresch [510].

On a commencé par des enregistrements de l'activité de neurones des aires rétiniotopiques $V1$ et $V2$ dont les champs récepteurs sont petits. Puis on a étudié les aires post- $V2$ qui sont de moins en moins rétiniotopiques et dont les neurones possèdent des champs récepteurs de plus en plus larges.

À partir de 1995 les études par IRMf ont débuté. Elles ont permis d'observer chez l'homme l'activité des aires concernées, de $V1$ à $V8$ et au LOC (cortex latéral occipital) dont nous avons déjà parlé à la section 4.3 du chapitre 2 à propos de la dialectique fond/forme. Par exemple, la figure 14 tirée de Mendola *et al.* [361] montre les aires post- $V2$ qui sont plus activées par un carré de Kanizsa cohérent que par un carré dont les pacmen sont incohérents (les seuils choisis ne permettent pas d'observer l'activité de $V1$ et $V2$). C'est une version IRMf des résultats de Catherine Tallon-Baudry explicités à la figure 11.

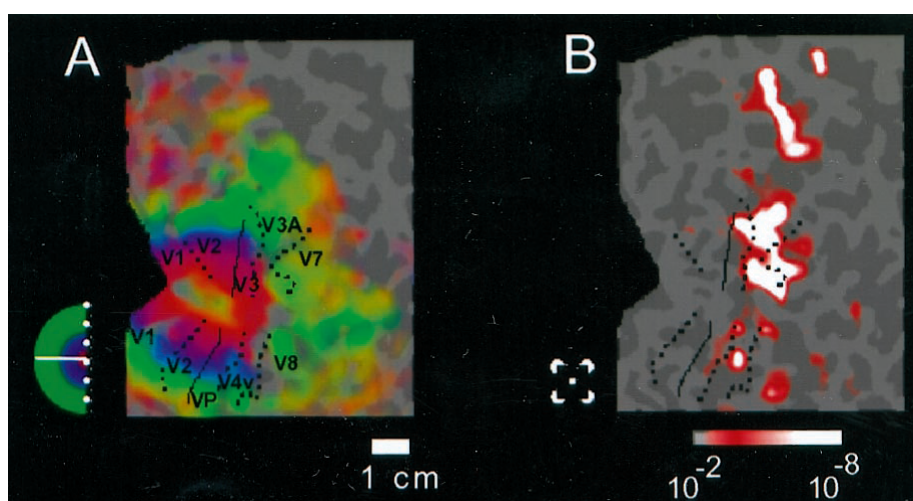


FIGURE 14. (A) Délimitations des aires de V1 à V8 sur une reconstruction plane de la surface du cortex. Les couleurs sont celles de la cible hémirétinienne en bas à gauche et montrent la projection rétinotopique. (B) Les différences d'activation des aires post-V2 stimulées par un carré de Kanizsa cohérent ou par quatre pacmen globalement incohérents. Les seuils choisis (barre en bas) ne permettent pas d'observer l'activité de V1 et V2.

Pour les CIs du second ordre que nous venons d'évoquer, rappelons que les recherches pionnières ont été celles d'Esther Peterhans et Rüdiger von der Heydt qui ont montré le rôle fondamental de l'aire V2. Cf. le Vol I, section 5.8. du chapitre 5 pour leurs expériences montrant que certains neurones de V2 détectent des orientations virtuelles qui sont orthogonales à des extrémités de lignes comme le montrent les figures 15 et 16 présentant des enregistrements chez le singe. Cela explique neuralemement les effets de “peigne” de la figure 12 ou les contours illusoires bistables de la figure 17, du moins si l'on peut définir une *intégration fonctionnelle* de ces orientations virtuelles. On peut en effet faire l'hypothèse que les réponses de V1 aux segments réellement contenus dans le stimulus (les “dents” du peigne) sont supprimées aux extrémités (“end points”) par une *inversion* de réponse, V2 activant la direction orthogonale. La bistabilité dans la figure 17 est particulièrement importante car elle montre que le système visuel peut utiliser *différentes stratégies* de complétion. Lorsque l'on observe un cercle, c'est que le CI est de courbure constante et donc maximale diffusée le long du contour. En revanche, lorsque l'on observe un carré c'est que le CI est partout rectiligne (de courbure nulle) sauf en des points singuliers anguleux où toute la courbure est concentrée. Il y a donc une stratégie de diffusion et une stratégie de concentration-trivialisation qui sont en compétition.

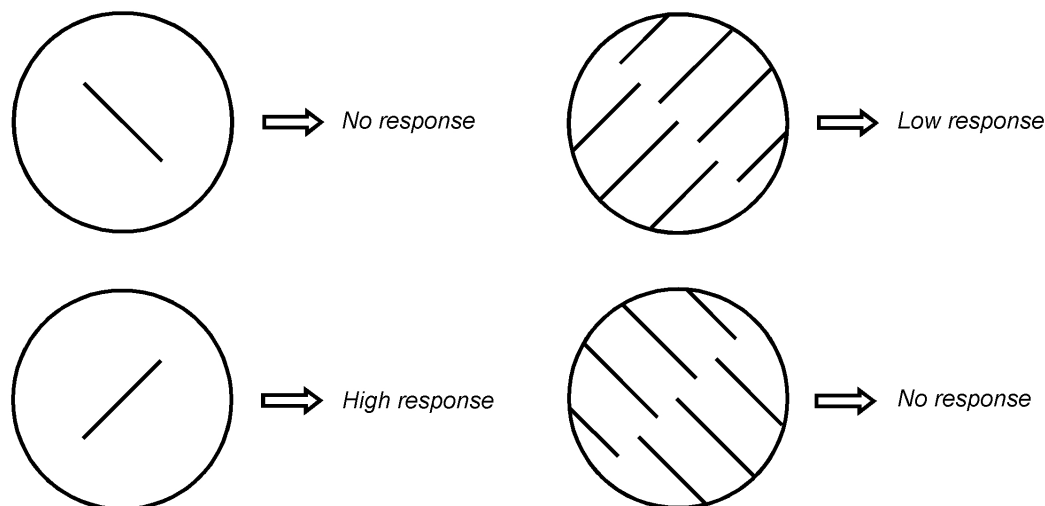


FIGURE 15. Dans cette figure et la suivante les deux disques représentent le stimulus (barre ou grille) avec une orientation de 0° et 90° . Les lignes des fenêtres de résultats représentent les réponses d'un neurone de l'aire $V2$ chez le singe (temps sur l'axe des abscisses) au stimulus tourné progressivement d'un angle variant de 0° à 180° (axe des ordonnées). *Gauche* : un neurone de $V1$ répondant à une orientation préférentielle p . *Droite* : lorsque le stimulus est un peigne dont les dents sont d'orientation p le neurone répond à cette orientation et ne répond pas au bord virtuel d'orientation orthogonale p^\perp . (Redessiné d'après Peterhans, von der Heydt [409] et von der Heydt, Peterhans [556]).

Toujours chez le primate, des travaux techniques comme ceux de Benjamin Ramden, Chou Hung et Anna Roe [460] ont précisé comment des “abutting lines gratings” activant des cellules de $V2$ détectant l'orientation de leur CI inhibent la détection de cette orientation par des cellules de $V1$ dont c'est pourtant l'orientation préférée. Pour étudier cette “inversion d'activation” (“activation reversal”), les auteurs utilisent des méthodes à la fois d'électrophysiologie et d'imagerie optique in vivo. La figure 18 montre par exemple les réponses d'une cellule de $V1$ d'orientation préférée $p = 135^\circ$ et d'orientation non préférée (orthogonale) $p^\perp = 45^\circ$. À gauche, on présente alternativement un segment d'orientation p^\perp et d'orientation p et on observe, comme il se doit, une suite de non-activations et d'activations. À droite, on présente des “peignes” de “dents” verticales dont le CI est alternativement d'orientation p^\perp et d'orientation p et l'on observe que la cellule de $V1$ est activée par le CI d'orientation non préférée p^\perp et inhibée par le CI d'orientation p . Les auteurs concluent

“we find that illusory contour stimulation leads to an activation map that, after appropriate subtraction of real line signal, is inversely related to the real

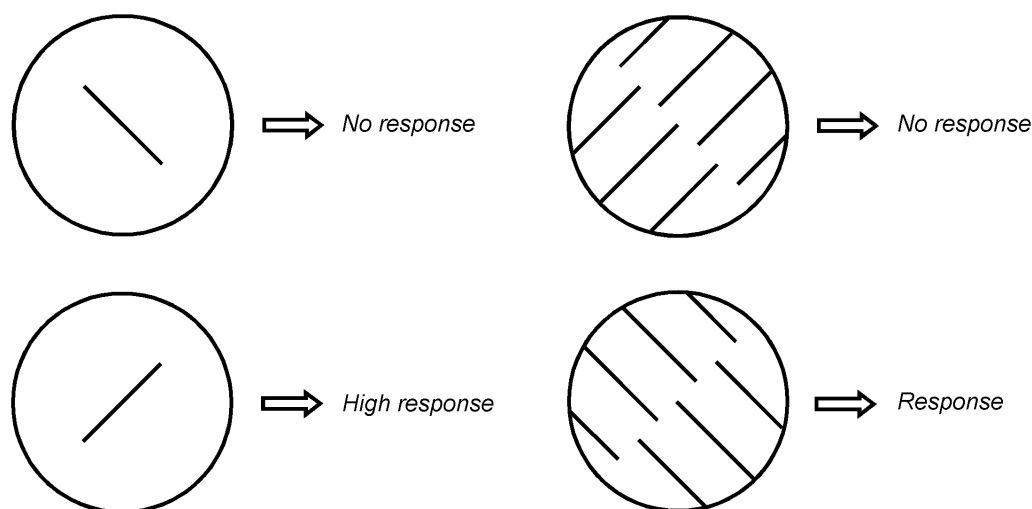


FIGURE 16. *Gauche* : un neurone de $V2$ chez le singe répondant à une orientation préférentielle p . *Droite* : lorsque le stimulus est un peigne dont les dents sont d'orientation p , le neurone ne répond pas à cette orientation mais répond en revanche au bord virtuel d'orientation orthogonale p^\perp . (Redessiné d'après Peterhans, von der Heydt [409] et von der Heydt, Peterhans [556]).

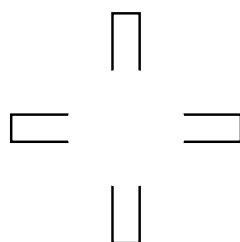


FIGURE 17. Contours illusoires bistables faisant intervenir l'aire $V2$. Il existe une compétition entre la forme "carré" et la forme "cercle" des contours subjectifs.

orientation map. The illusory contour orientation is thus negatively signaled or de-emphasized in $V1$. This 'activation reversal' is robust." ([460], p.648)

Bref, $V2$ intervient de façon dominante dans les réponses aux CIs du second ordre, la différence entre $V1$ et $V2$ résidant dans la proportion supérieure de neurones de $V2$ répondant à ces CIs (60% contre 42%) et dans le fait que la réponse des neurones de $V1$ à ces CIs "peigne" est souvent masquée par leur réponse aux lignes inductrices que sont les "dents", alors que celle des neurones de $V2$ est plus "pure". C'est un tel masquage qui expliquerait que les réponses dans $V1$ n'aient pas été tout de suite observées. Il est par conséquent justifié de s'interroger d'abord

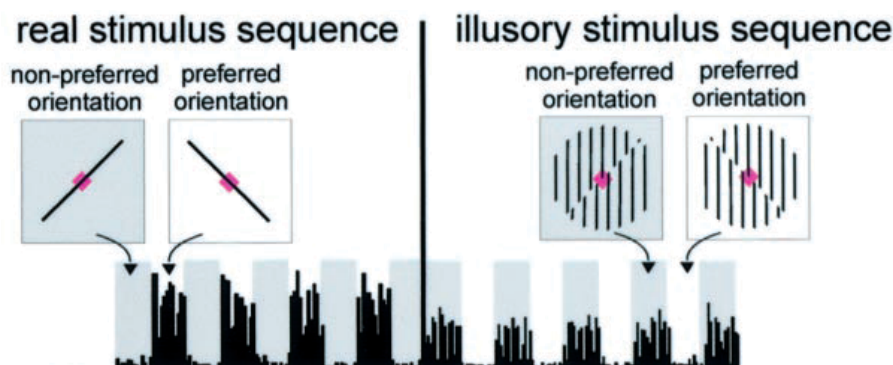


FIGURE 18. Réponses d'une cellule de V1 d'orientation préférée $p = 135^\circ$ et d'orientation non préférée (orthogonale) $p^\perp = 45^\circ$. À gauche, suite de non-activations par p^\perp et d'activations par p . À droite, présentation alternative de "peignes" de "dents" verticales dont le CI est d'orientation p^\perp puis d'orientation p . On observe que la cellule est activée par le CI d'orientation non préférée p^\perp et inhibée par le CI d'orientation p . (Adapté de la figure 9 de [460])

sur V1 et de voir comment ce que l'on sait de son architecture fonctionnelle peut permettre de comprendre les mécanismes d'intégration "fonctionnelle" en jeu. En effet, conformément à la "high-resolution buffer hypothesis" de Mumford-Lee [320] expliquée dans l'Introduction générale, et l'Introduction du chapitre 4 du Vol I, V1 se trouve fortement impliqué et l'on peut faire l'hypothèse d'une construction progressive des CIs partant du niveau de V1 pour devenir de plus en plus riche dans les aires successives avec feedbacks descendants ("top down").

2.1.5. Présence des CIs chez de nombreuses espèces.

Beaucoup d'expériences ont été menées sur des modèles animaux comme le singe ou le chat. Par exemple, rappelons que nous avons vu dans le Vol I que les CIs de type "effet peigne" ont été observés par imagerie chez le chat entre autres par Shavin Sheth, Jitendra Sharma, Chenchal Rao et Mriganka Sur [497].

Mais en fait la détection de CIs existe chez de nombreuses autres espèces d'animaux comme certains oiseaux et même certains insectes (abeilles). Cela est important car on a souvent avancé l'argument que les stimuli utilisés pour étudier les CIs étaient artificiels et n'existaient pas dans la nature. Telle était la position de James Gibson et de son "approche écologique de la perception" [212] fondée sur la notion de "perception directe". Mais l'argument n'est pas valide car les contours lacunaires sont innombrables en vision naturelle et la reconstruction de bords globaux permettant d'identifier des objets comme des proies est vitale pour nombre de prédateurs par exemple nocturnes (au crépuscule l'illumination est très faible) ou dont les proies peuvent se camoufler dans le décor. La complétion modale est donc cruciale pour la survie de certaines espèces.

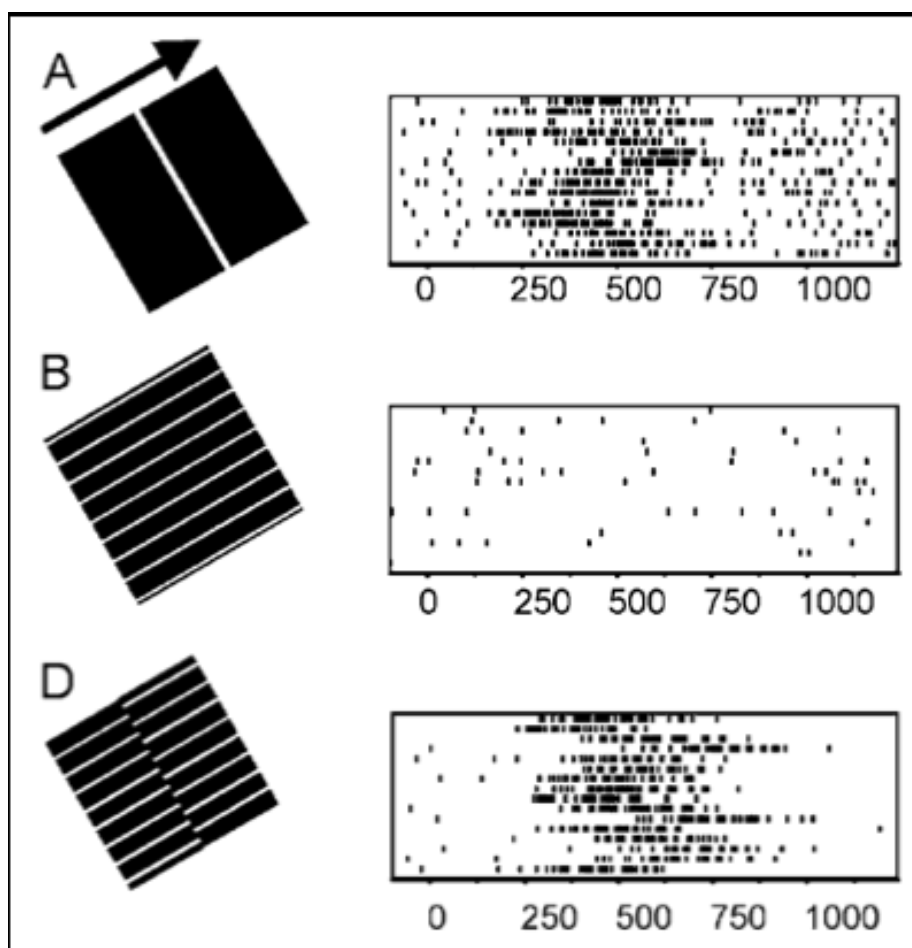


FIGURE 19. Réponse d'un neurone du Wulst de la chouette effraie. (A) Réponse à l'orientation préférentielle p . (B) Absence de réponse à un grating d'orientation p^\perp . (D) Réponse au CI d'orientation p d'un "abutting lines grating" d'orientation p . (Repris de [389], figure 8)

La figure 19 due à Andreas Nieder et Hermann Wagner montre les enregistrements d'un neurone du Wulst de la chouette effraie ("owl barn").⁵ On y voit (A) la réponse à une barre d'orientation préférentielle p traversant son champ récepteur (à gauche). Cette réponse (à droite) est une suite de spikes sur un intervalle de $1250ms$ répétées sur 16 essais. Comme il se doit, la réponse à un grating d'orientation p^\perp est inexistante. En revanche, en (D) la cellule répond très significativement à un CI d'orientation p .

5. Le Wulst est la terminaison de la voie visuelle thalamofuge, l'équivalent du cortex visuel primaire.

Les travaux sur $V2$ ont permis de mieux comprendre la façon dont $V2$ participe au calcul des relations de ségrégation figure/fond, calcul qui, comme le note Anna Roe, consiste à déterminer

“how boundaries and surfaces relate to each other, thereby forming the initial stage of figure/ground segregation.” ([469], p. 127).

Le problème de l’attribution des bords à l’un des deux domaines qu’ils séparent (“border ownership”) n’est pas résoluble au niveau de $V1$. En revanche $V2$ répond à des bords appartenant à une surface car ses champs récepteurs répondent non seulement à des fragments de lignes mais également à un segment constituant le bord de l’un des deux demi-espaces qu’il sépare. La ségrégation fond/forme résulte alors d’un calcul global et fortement contextuel permettant d’intégrer ces données locales d’assignation de bords.

Bref, des travaux convergents ont conduit de nombreux spécialistes comme Andreas Nieder [389] à considérer que les CIs relevaient de processus essentiellement de bas niveau et bottom-up même si des feed-back top down venant d’aires post- $V2$ jouent aussi certainement un rôle. Comme Manuel Soriano et ses co-auteurs l’affirment dans [507]

“It appears that bottom-up processing is sufficient to explain the occurrence of subjective contours and that a higher cognitive function (i.e. top-down processing) is not warranted.” ([507], p.109)

Notre propos va donc être désormais de modéliser neurogéométriquement les CIs dans ce contexte expérimental. Ces modèles seront des modèles géométriques d’intégration fonctionnelle. Mais avant que d’y venir nous allons donner quelques très brèves indications sur des modèles neuronaux classiques qui ont été proposés.

2.2. Quelques modèles classiques de CIs

Commençons par des modèles inspirés par les travaux classiques de Grossberg et Mingolla [230] ainsi que de von der Heydt et Peterhans [556].⁶

Pour le premier modèle, le CI prototypique est du type Kanizsa et il est formé par interaction entre des bords (approximativement) colinéaires, le contour étant engendré dans leur direction commune. A leur propos, Heitger parle de “paragrouping”. Pour expliquer les contours produits par des extrémités de ligne, Grossberg et Mingolla supposent que ces terminaisons induisent un bord élémentaire perpendiculaire à la ligne, ces éléments induits pouvant alors coopérer entre eux et avec les bords réels. Il n’est pas besoin de discontinuités pour enclencher ce processus car le phénomène des CIs résulte des mécanismes “normaux” d’intégration des contours

6. Cette sous-section vient d’une compilation effectuée autrefois (1998) par notre élève Yannick Tondut.

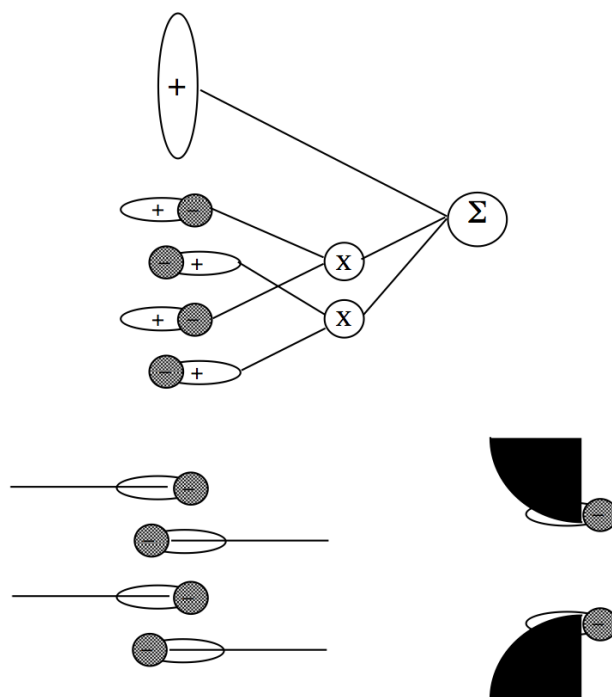


FIGURE 20. Le mécanisme de formation des CIs de Peterhans et von der Heydt : les cellules à accord de distance détectent les discontinuités. Les réponses obtenues sont multipliées par couples de cellules parallèles (ceci modélise la non-linéarité du processus) et les résultats de tous les couples sont sommés le long de la direction perpendiculaire commune. Les éventuels contours réels viennent s'y ajouter. La cellule détectrice du contour (notée sigma) décharge si la somme totale atteint ou dépasse son seuil. Ce mécanisme s'applique aussi bien aux contours "à terminaisons" qu'aux contours "à bords".

réels et de regroupement perceptif ("grouping") plutôt que d'un processus général de prolongation sous occlusion.

À l'inverse, pour von der Heydt et Peterhans, ainsi que pour ceux qui s'inspirent de leur modèle (Kellman et Shipley [286]), les CIs résultent de l'interaction entre des détecteurs de discontinuités (cellules "end-stopped" ou à accord de distance, voir Heitger *et al.* [247]) et sont formés perpendiculairement à l'orientation des bords (cf. figure 20). À leur propos, Heitger parle de "orthogrouping". Les contours qui prolongent les bords sont alors un cas particulier de ce mécanisme (Peterhans et von der Heydt [408]).

Un défaut commun à ces modèles est qu'ils sont conçus pour expliquer les contours rectilignes. Leur adaptation aux contours courbes pose par conséquent de sérieux problèmes :

- Chez Peterhans et von der Heydt, il faudrait prendre en compte tous les couples de cellules approximativement parallèles, puis sommer les produits de leurs

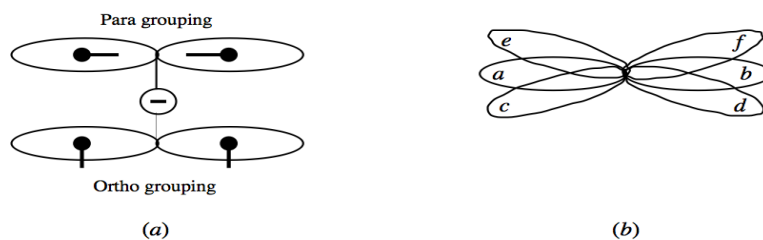


FIGURE 21. Le champ de regroupement de Heitger et von der Heydt génère localement un segment orienté lorsque ses deux lobes sont simultanément activés. (a) Para- et ortho-grouping (pour l'orthogrouping on tolère une orthogonalité approximative). (b) Le champ de regroupement prend en compte les alignements courbes. La contribution des chemins de courbure constante (ab , cd , cf , ed et ef) est double de celle des autres (ad , af , bc , be).

activités selon tous les contours possibles. Ceci nécessiterait un très grand nombre d'unités (explosion combinatoire) et un câblage sophistiqué.

- Cette critique vaut aussi pour le modèle de Grossberg et Mingolla, qui entend résoudre le problème des contours courbes en tolérant des alignements approximatifs. Ces alignements (réels ou illusoires) sont détectés par des cellules “dipôles” de grand champ récepteur. Le risque d'explosion combinatoire guette également un tel mécanisme.

Un autre modèle intéressant est celui de Heitger et von der Heydt [248]. Il se situe en position intermédiaire entre les deux familles que nous venons d'exposer, puisqu'il admet des regroupements aussi bien dans la direction des éléments orientés (“paragrouping”) que dans la direction perpendiculaire (“orthogrouping”). De plus, tous les mécanismes qu'il utilise sont “bottom-up”, alors que dans les modèles de Grossberg et Mingolla et de Finkel et Edelman les boucles de rétroaction pourraient se révéler peu efficaces dans le traitement d'images naturelles complexes. Enfin, ce modèle fonctionne aussi bien avec les contours courbes qu'avec les contours droits (cf. figure 21).

Heitger et von der Heydt traitent simultanément les contours sous occlusion et les contours illusoires. Le principe fondateur est la détection de certaines singularités ou points-clés (Heitger *et al.* [247]) et leur appariement. Ce sont des cellules orientées dotées de deux lobes de grande taille, comme les cellules dipôles de Grossberg et Mingolla, qui constituent ce champ de regroupement (“grouping field”). Le principe est le suivant : lorsque ses lobes contiennent tous les deux un point-clé, la cellule envoie un signal local de formation d'un contour subjectif (c'est-à-dire qu'elle crée un petit segment orienté). Ce signal vient s'ajouter aux sorties des détecteurs de contours réels de même localisation. Le tracé du contour est ensuite déterminé par les maxima locaux d'activation. Le champ de regroupement est, par sa forme, proche du “champ d'association” de Field, Hayes et Hess (que nous avons expliqué

et modélisé dans la section 5.2. du Vol I et rappelé plus haut) ainsi que du “champ de coopération” de Grossberg et Mingolla.

La principale question que soulève ce modèle concerne sa vraisemblance biologique. En effet, un même champ de regroupement doit contenir les deux extrémités inductrices d’un contour subjectif. Or certains CIs, sans parler des contours sous occlusion, peuvent combler des vides importants. Ceci supposerait donc des champs récepteurs de très grande taille, qui n’ont pas été observés dans $V1$. Le modèle semble davantage inspiré de $V2$, mais aucune contrepartie neuronale précise n’est proposée. Il semble en tous cas que les auteurs aient, eux aussi, délaissé la question de l’échelle des processus proposés.

Tous ces modèles peuvent facilement être interprétés géométriquement en termes de fibrations $\pi : M \times P \rightarrow M$ si l’on considère des fibres dont les éléments ne sont plus de simples orientations p mais des paires d’orientations orthogonales (p, p^\perp) , c’est-à-dire des repères dans l’espace de base M . Le fibré projectif $\pi : M \times P \rightarrow M$ est alors remplacé par le fibré des repères (“frames”) $\rho : M \times F \rightarrow M$. Nous reviendrons plus bas sur ce point important dans la section 2 du chapitre 11 à propos de la méthode du *repère mobile* d’Élie Cartan.

L’hypothèse de détecteurs de configurations globales est insatisfaisante car ceux-ci devraient être très nombreux et couvrir de larges zones. De plus, les mécanismes proposés pour la formation des CIs semblent être trop souvent ad hoc. C’est pourquoi nous proposons un modèle neurogéométrique alternatif qui fonctionne uniquement sur la base d’inductions locales, la spécificité de cette induction locale étant décrite par le modèle du champ d’association. Un même mécanisme général de traitement des contours serait alors responsable aussi bien de l’intégration des contours “réels” que de l’extrapolation de CIs.

2.3. Méthodes de mesure des CIs

Nous allons nous focaliser sur des contours de Kanizsa *courbes* – que nous appellerons *K-contours* – où les côtés des angles internes des pacmen ne sont pas vraiment alignés (cf. les figures 22 et 23 qui reprennent en les simplifiant les figures 6 et 10). En effet le type *précis* de courbes que peuvent réaliser des contours subjectifs modaux courbes fournit un indice de premier ordre sur les mécanismes neuronaux d’intégration des contours.

Le problème est évidemment que ce type précis est difficile à évaluer car il faut arriver à mesurer avec une précision suffisante la position du CI, en particulier son extremum médian lorsqu’il s’agit d’un CI symétrique du premier ordre. Il faut le faire avec une précision suffisante pour départager entre les différents types de courbes interpolant entre les pacmen inducteurs. Mais cela est très délicat car toutes les courbes de faible courbure qui courbent un même segment de droite sont très proches.

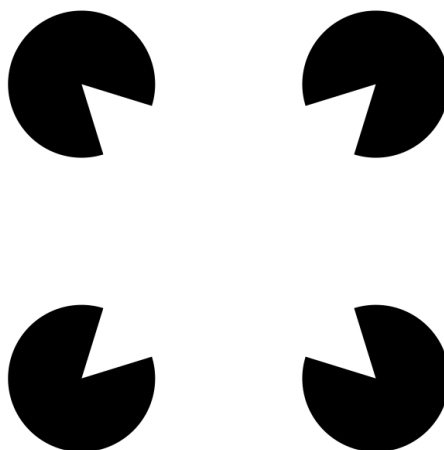


FIGURE 22. Un carré de Kanizsa aux bords illusoires courbes.

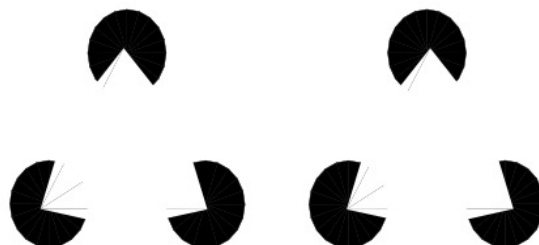


FIGURE 23. Un triangle de Kanizsa aux bords illusoires courbes en vision stéréoscopique. Si l'on fusionne les deux images, le triangle illusoire semble flotter au-dessus des trois pacmen.

À la fin des années 1980 une question aussi essentielle que celle de l'influence de la longueur des bords inducteurs et des lacunes à combler sur la netteté du contour restait controversée (Kellman et Shipley [286], Shipley et Kellman [499], Spillman et Dresp [510]). Au début des années 1990, Dresp et Bonnet [151], [152], [153] et Dresp et Grossberg [154] ont rapproché le phénomène de formation des CIs de la modification des seuils dans les tâches de détection de cibles. Cela a fourni une méthode intéressante pour mesurer l'extremum d'un K -contour, celle de "subthreshold summation" : le seuil de détection d'un petit segment sous-liminaire parallèle au K -contour s'abaisse brutalement quand le segment se superpose exactement au K -contour (Dresp et Bonnet [153], p. 1071) :

"When both lines are strictly superimposed, the neural response to these stimuli add together and a lower contrast is needed for a threshold response."

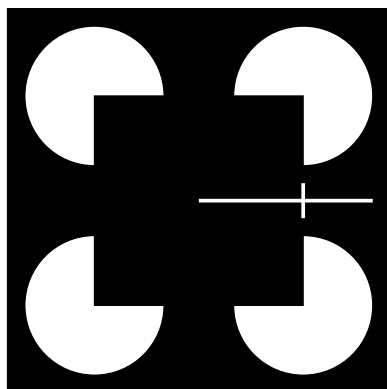


FIGURE 24. La configuration inductrice. On déplace le petit segment vertical de gauche à droite et on mesure son seuil de détection. (D'après Dresp, Bonnet [152], p. 762).

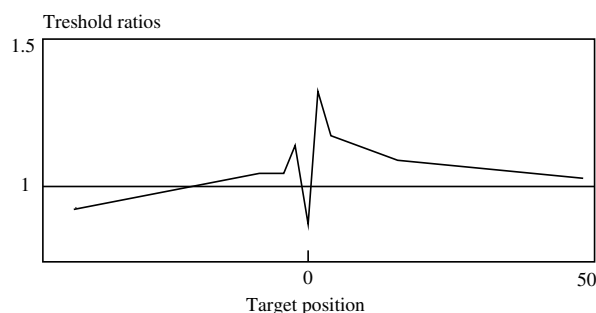


FIGURE 25. Courbe typique de réponse. On voit que lorsque le segment est juste sur le contour virtuel son seuil de détection s'abaisse brusquement. On notera aussi que le seuil de détection est légèrement supérieur lorsque le segment est à l'extérieur du carré de Kanizsa. (D'après Dresp, Bonnet [152], p. 763).

L'interprétation la plus plausible est qu'il y a une activité additive dans le cortex et que le CI préactive les neurones détectant le segment (cf. figures 24 et 25).

Une autre méthode permettant de mesurer avec une assez bonne résolution l'extremum d'un K -contour s'inspire du fait, illustré à la figure 13, que les CIs sont notablement renforcés lorsque le point d'arrêt d'un segment transversal est positionné juste sur le contour. On peut donc demander aux sujets de positionner au mieux le segment et son extrémité fournit alors une bonne mesure. Ce sont des variantes de cette méthode que nous avons utilisées dans une expérience de 1998 que nous allons présenter dans la section suivante.

Dans [240], Haiko, Mingolla et Somers ont introduit une nouvelle méthode, plus performante, de mesure des CIs. Dite avec humour de "chomping pacman" (pacman mâchant bruyamment), elle consiste à faire osciller l'ouverture des pacmen inducteurs avec une certaine fréquence. On obtient ainsi des configurations dynamiques

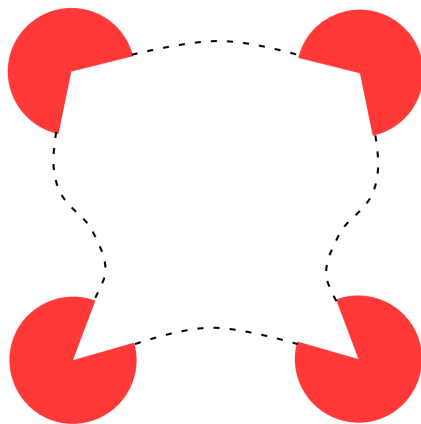


FIGURE 26. La méthode de “chomping pacman” de Haiko, Mingolla et Somers. On fait osciller les ouvertures des pacmen avec des phases différentes.

induisant des *déformations dynamiques* de CIs. Avec des anneaux de Varin et un effet néon on peut encore augmenter la saillance. Les auteurs ont trouvé que la fréquence optimale est autour de 2Hz, ce qui donne une indication sur la vitesse de propagation des spikes le long des connexions horizontales. Un aspect particulièrement intéressant de cette méthode est que l’on peut faire osciller les ouvertures des pacmen avec des phases différentes, de tels déphasages induisant des contours non seulement courbes mais avec des changements de signe de la courbure (et donc des points d’inflexions) que l’on ne pourrait pas observer avec des configurations statiques. La figure 26 en donne un exemple (un peu exagéré).

Évidemment toutes ces méthodes possèdent d’importantes limites. Des résultats quantitatifs plus fins et plus précis sur la position des extrema des CIs modaux seraient donc du plus haut intérêt pour mieux comprendre, en permettant de sélectionner un modèle plus vraisemblable que les autres, quels sont les mécanismes géométriques d’intégration neuralemement implémentés.

2.4. Une expérience sur les K -contours courbes (avec J. Ninio)

En collaboration avec Jacques Ninio (École Normale Supérieure, Paris) qui en a programmé l’interface et le déroulement, nous avons effectué en 1998 une expérience sur les K -contours courbes. Notre but était de mesurer la position précise du point extrémal du contour et de le comparer avec les prédictions de modèles simples. Nous avons utilisé des familles de K -contours avec

- 2 configurations : triangle, carré ;
- 2 tailles de configurations ;

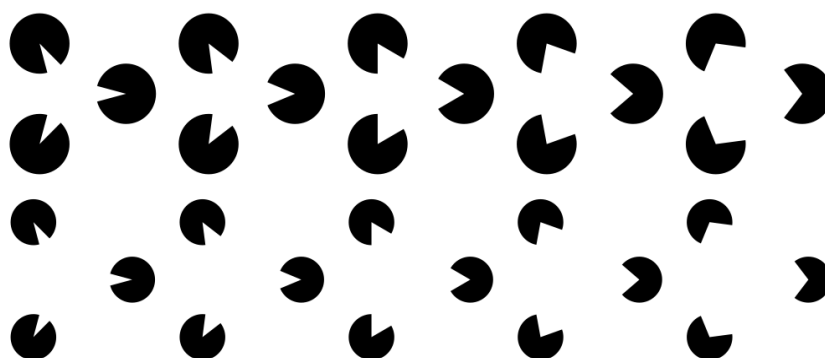


FIGURE 27. Triangles de Kanizsa courbes utilisés pour l'expérience avec Jacques Ninio.

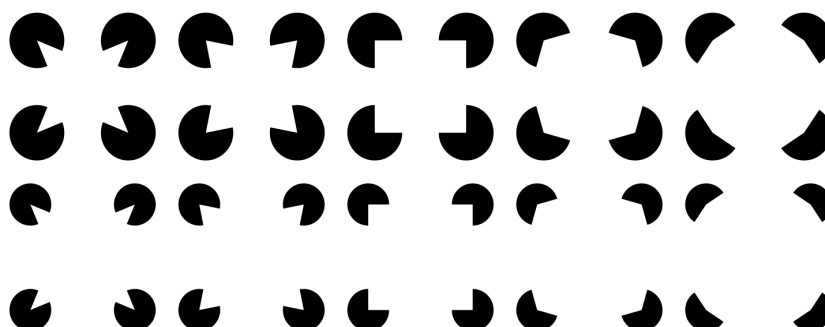


FIGURE 28. Carrés de Kanizsa courbes utilisés pour l'expérience avec J. Ninio.

- 2 tailles de pacmen ;
- 4 orientations ;
- 5 angles (cf. figures 27 et 28).

Pour la détection, nous avons demandé aux sujets de positionner correctement une marque (l'extrémité d'une ligne orthogonale, un petit segment, l'axe d'une petite bande) sur l'extrémum du contour (cf. figure 29).

Pour différents cas (triangle / carré et taille des pacmen petite / grande) nous avons comparé 3 positions (cf. figure 30) :

- la position linéaire par morceaux (intersection des côtés des deux pacmen inducteurs) ;
- la position choisie par les sujets ;
- la position donnée par l'extrémum de l'arc de cercle tangent aux deux côtés des pacmen.

Nous montrons à la figure 31 les résultats de l'expérience pour le cas des carrés avec petits pacmen. On constate que le modèle de l'arc de cercle devient rapidement

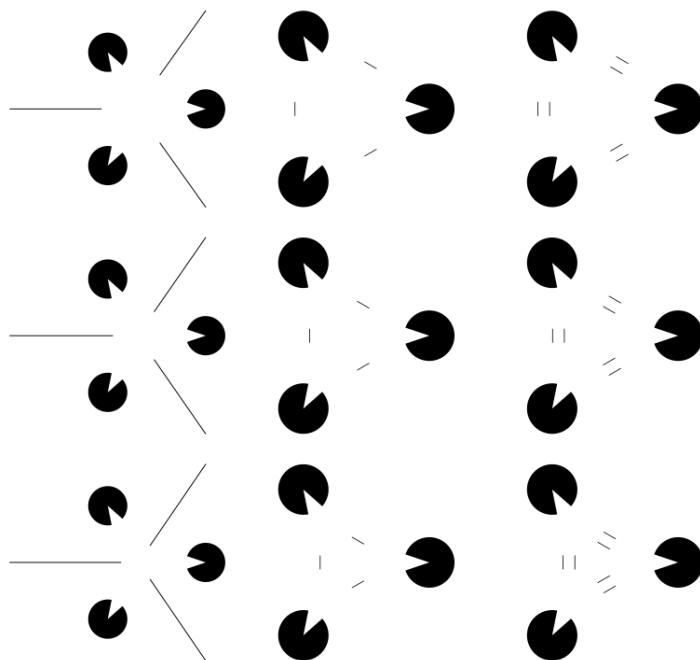


FIGURE 29. La méthode de détection de l'extrémum d'un K -contour courbe. Le sujet doit placer une marque (l'extrémité d'une ligne orthogonale, un petit segment, l'axe d'une petite bande) aussi précisément que possible sur l'extrémum.

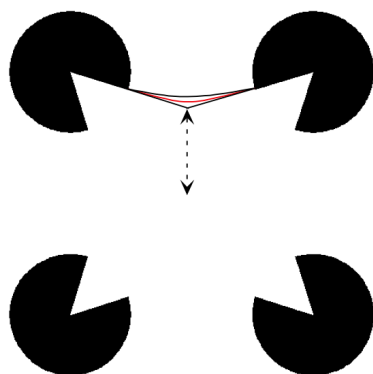


FIGURE 30. Comparaison de trois K -contours : le contour linéaire par morceaux (intersection des côtés des deux pacmen), le contour choisi par les sujets, le contour circulaire (arc de cercle tangent aux deux côtés des pacmen).

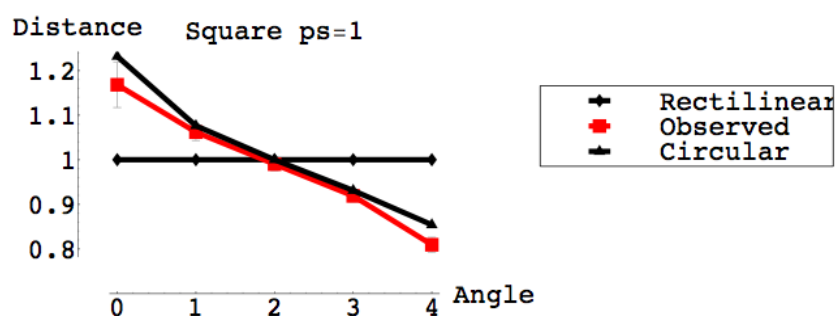


FIGURE 31. Les résultats de l'expérience pour le cas des carrés avec petits pacmen (paramètre $ps = \text{"pacmen size"} = 1$). Le graphique représente la distance d de l'extrémum du K -contour au centre de la configuration comme fonction de l'angle d'ouverture. La distance d est mesurée par son rapport au cas linéaire par morceaux. 5 angles d'ouverture $\#i$ sont considérés en abscisse : $\#2$ correspond au cas classique d'un K -contour droit ($d_2 = 1$), $\#1$ à un K -contour légèrement concave ($d_1 > d_2 = 1$), $\#0$ à un K -contour plus concave ($d_0 > d_1 > d_2 = 1$), $\#3$ à un K -contour légèrement convexe ($d_3 < d_2 = 1$), $\#4$ à un K -contour plus convexe ($d_4 < d_3 < d_2 = 1$). On voit que le K -contour détecté empiriquement se situe entre le cas linéaire par morceaux et le cas circulaire.

assez mauvais et que le K -contour détecté empiriquement se situe entre le cas linéaire par morceaux et le cas circulaire.

2.5. Psychophysique et traitement d'image

Les CIs constituent un phénomène majeur étudié par la Gestalttheorie et celle-ci a été relayée à l'époque moderne par les algorithmes de vision computationnelle et de traitement d'images. Ce domaine de recherche est immense et intimement solidaire des expériences de psychophysique. Parmi tous les travaux qui lui sont consacrés nous voudrions signaler en particulier le programme de recherche de Jean-Michel Morel et de son équipe qui s'est explicitement proposé de transformer la Gestalttheorie de Wertheimer à Kanizsa en un "computer vision programme". Ce n'est pas du tout évident car il faut trouver des algorithmes à même de fournir une "automatic computation of gestalts in digital image". On pourra consulter [141] et [142].

Ces approches rejoignent les données expérimentales, répétons-le, à travers la psychophysique. Quant à nous, nous voudrions faire le lien avec le connectome des aires visuelles. en utilisant les ressources de la neurogéométrie. Pour cela, il nous faut commencer par les phénomènes les plus primitifs.

2.6. Neurogéométrie de la contribution de V1

Nous allons donc désormais nous focaliser dans les prochains chapitres sur la possibilité de géométriser, au sens de la neurogéométrie, la contribution possible de V1 et de V2 aux CIs et donc les fondations de bas niveau, feedforward et ascendantes de leur construction. Notre principe théorique de base (introduit dès les années 1990 dans nos premiers travaux de neurogéométrie) sera en accord avec le point de vue “constructiviste” développé plus tard par Andreas Nieder en 2002 : la reconstruction des CIs doit être expliquée

“according to the wiring principles of the visual system” ([389], p. 250).

Autrement dit, on doit partir de *l’architecture fonctionnelle* des aires primaires. Et dans la mesure où ces architectures fonctionnelles sont modélisées par un certain type de structures géométriques, c’est dans le cadre de cette géométrie que l’on doit mathématiser les CIs.

Dans cette optique, les CIs du premier ordre interpolent entre leurs conditions aux limites et l’on peut chercher, parmi toutes les interpolations possibles, celles qui sont “optimales” en un certain sens. Une telle option modélisatrice est justifiée puisque nous avons vu dans les précédents chapitres que l’architecture fonctionnelle de V1 implémente des algorithmes d’intégration au sens “fonctionnel”. On peut par conséquent appliquer des *principes variationnels* à condition de pouvoir définir des “fonctions de coût” à optimiser. Les fonctions de coût les plus classiques sont définies en introduisant une métrique et en minimisant la longueur des interpolations, ce qui conduit tout naturellement à des modèles *géodésiques* des CIs.

Évidemment, de tels modèles neurogéométriques ne constituent en rien une explication exclusive et exhaustive des CIs. Ils se bornent à contextualiser, dans l’immense univers mathématique des modèles variationnels, une géométrisation de la possible *contribution* de V1 et de son architecture fonctionnelle.

3. Les premiers modèles variationnels (Ullman, Horn, Mumford)

3.1. “L’intelligence des lignes courbes”

L’hypothèse que l’architecture fonctionnelle de V1 implémente la structure de contact de $\mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R}$ conçue comme espace des 1-jets des courbes régulières de \mathbb{R}^2 va nous permettre de commencer à comprendre la formation énigmatique de contours illusoires (CIs) modaux du premier ordre de longue portée comme ceux étudiés par Kanizsa. En effet, en première approximation (le modèle doit être complexifié en tenant compte au moins de V2), les pacmen définissent des conditions au bord (a_1, θ_1) et (a_2, θ_2) et l’hypothèse de base est que le contour est solution d’un *problème variationnel*.

Nous reviendrons longuement au chapitre 9 sur l'histoire des modèles variationnels, en particulier ceux qui ont été élaborés pour accéder à ce que Guillaume de l'Hopital appelait joliment en 1696 "l'intelligence des lignes courbes" ; mais faisons d'emblée quelques rappels permettant d'en justifier ici l'usage. Résumons ce que nous avons vu :

- (i) Dans les perceptions visuelles de bas niveau que nous modélisons à une échelle mésoscopique, il y a d'abord les inputs sensoriels périphériques qui activent des éléments du plan de base \mathbb{R}^2 de la rétine, puis, en suivant la voie rétino-géniculo-corticale, certaines cellules de $V1$ et d'autres aires visuelles primaires.
- (ii) Ensuite la perception *complète* les données sensorielles. Les inputs fonctionnent comme des conditions au bord ou des conditions aux limites lacunaires et, à partir d'eux, la perception ajoute des interpolations. Les CIs en sont un exemple typique.
- (iii) Le phénomène fondamental de complétion perceptive résulte de propagations d'activités (spikes) le long des connexions corticales. C'est donc le design de la connectivité qui doit expliquer la structure (en particulier la morphologie géométrique) de l'état perceptif observé.
- (iv) Mais il existe un nombre considérable (à la limite infinie dans les modèles continus) de complétions possibles. Il doit donc exister un principe de *choix* qui *actualise* une complétion en *virtualisant* toutes les autres qui seraient a priori possibles. Dans le cas des CIs, il en existe une infinité interpolant entre les conditions aux limites. Mais le système visuel en choisit un. C'est cette dialectique entre le virtuel et l'actuel dans le processus de sélection qui justifie le recours à des modèles variationnels puisque ces derniers sont par définition faits pour cela. La complétion sélectionnée devient *visible* : elle émerge de façon saillante par pop-out, les autres complétions possibles restant invisibles (non perçues).
- (v) Le principe de choix va consister à optimiser le "coût" des complétions possibles, i.e. des états d'activation de l'aire considérée. Supposons que nous investissons $V1$. Tout repose sur son architecture fonctionnelle (AF).
- (vi) Nous avons vu que l'AF possède deux aspects : (a) la structure de fibration implémentant l'espace des jets et les relèvements legendriens des courbes du plan de base, (b) la structure de contact.
- (vii) En général les "coûts" à optimiser font intervenir une métrique des espaces utilisés. Par exemple les rayons lumineux minimisent le chemin optique et, dans un milieu homogène, ils se propagent en ligne droite parce que les droites sont les géodésiques de la métrique euclidienne. Dans les questions que nous traitons ici, nous trouvons plusieurs métriques naturelles : la métrique (qu'on peut supposer euclidienne) du plan de base \mathbb{R}^2 ; la métrique euclidienne de

$\mathbb{V}_J \simeq \mathbb{R}^3$; la métrique de $SE(2)$; des métriques sous-riemanniennes de \mathbb{V}_J et de $SE(2)$. D'où de nombreuses possibilités dont il nous faudra présenter la théorie en laissant aux résultats expérimentaux la tâche de les départager.

- (viii) La théorie des CIs modaux comme problème de complétion en vision de bas niveau se trouve ainsi plongée dans un univers mathématique dont l'histoire et la richesse sont tout à fait prodigieuses : celui des problèmes variationnels qui, depuis Leibniz puis Euler et Lagrange, englobe une partie considérable de la physique mathématique. Nous ne répèterons jamais assez que, dans ces domaines, comme nous l'avons déjà noté dans l'*Introduction* du Vol I, section 2.5., les premiers grands inventeurs et promoteurs du calcul différentiel et intégral classique ont dépensé une énergie et une virtuosité sans pareilles pour résoudre des problèmes apparemment élémentaires au niveau de l'observation macro mais dont l'explication en termes de processus physiques sous-jacents micro était d'une complexité insoupçonnée. Ils firent l'hypothèse que cette explication pouvait reposer sur le calcul intégral-différentiel de niveau méso si l'on supposait que le niveau micro (disons "atomique" même si cela était spéculatif à l'époque) était moyennable par une approximation continue et localement représentable par des données différentielles. Ils ont ainsi dû mener à bien des calculs extrêmement techniques (et infiniment riches d'avenir) pour calculer des choses apparemment aussi simples que la longueur d'un arc d'ellipse, la forme d'une chaînette, le mouvement d'un pendule ou la forme d'une lame élastique pliée (*elastica*). Ce sont les équations différentielles méso traduisant des propriétés mécaniques locales qui établissent un pont entre les faits observés au niveau macro (par exemple la courbure de la lame élastique) et la physique fondamentale sous-jacente (les forces de liaison et les propriétés atomiques du réseau métallique expliquant l'élasticité de la lame). La complexité de leur résolution n'a donc rien d'étonnant si l'on tient compte de cette nécessité de "bridging the scales". Il en va exactement de même dans notre cas et on ne voit donc vraiment pas pourquoi il ne faudrait pas faire de même pour calculer la forme de contours illusoires courbes : le cerveau vaut bien une chaînette ou une lame élastique ! L'explication neuro-physiologique micro (neurones et potentiels d'action) des CIs observés au niveau perceptif macro est d'une complexité inouïe et les modèles neurogéométriques méso que nous allons présenter constituent un pont jeté au dessus de l'abîme reliant ces deux niveaux.
- (ix) Beaucoup de modèles variationnels classiques concernaient la dynamique de systèmes mécaniques S d'espace de configurations M . Celle-ci pouvait être approchée soit de façon causaliste à la Newton en termes de forces s'exerçant sur les composants élémentaires de S , soit de façon finaliste à la Euler-Lagrange en appliquant un principe de moindre action. On cherchait les trajectoires

$q(t) \in M$ de S , un point q de M représentant un état instantané de S et la dimension de M étant donc égale au nombre de degrés de liberté de S . Mais dans notre cas, il s'agit plus de géométrie que de dynamique et nous considérons par exemple une fonction $y = f(x)$ où x est une variable spatiale et non pas le temps t . Mais les formalismes sont les mêmes. D'ailleurs, nous le verrons, pour les courbes dans des "bons" espaces métriques comme ceux que nous considérons, on peut facilement relier dynamique et géométrie en paramétrant les courbes par leur élément d'arc de longueur ds .

- (x) Dans l'univers des problèmes variationnels dans lequel nous allons nous aventurer, il existe une autre dialectique, celle entre géométrie et analyse. En général on impose aux solutions des équations différentielles que l'on souhaite résoudre des contraintes d'"admissibilité" (par exemple de régularité) qui s'expriment en termes compliqués de topologie sur les espaces fonctionnels des solutions possibles. Par exemple, pour une lame élastique, on imposera que la lame se courbe mais sans se briser, ce qui relaxerait l'énergie élastique et remplacerait une forme courbe par une courbe linéaire par morceaux dont la courbure est partout nulle, sauf en un point singulier anguleux où la courbure diverge. Nous y reviendrons entre autres à la section 5.4 du chapitre 14. Dans cet ouvrage nous privilégierons l'aspect "géométrie" qui nous est plus familier, mais la bibliographie est assez fournie pour permettre au lecteur d'approfondir s'il le souhaite l'aspect "analyse".
- (xi) Une dernière remarque, particulièrement importante. Dans les modèles variationnels classiques dont nous nous inspirerons, les fonctions "coût" F à optimiser font intervenir des éléments différentiels ν comme, par exemple, la tangente ou la courbure d'une courbe. Cela signifie que les problèmes sont formulés dans des espaces de *jets* où ces éléments sont considérés de deux façons : (a) comme des variables indépendantes, (b) comme des variables devenant elles-mêmes des fonctions du temps t ou d'une variable spatiale plus primitive x . La fonction $F(\nu)$ devient ainsi une *fonction composée* de t (ou de x) à travers la variable $\nu : F(\nu(t))$. Il y a donc deux façons de *dérivée* F : soit comme fonction de ν considérée comme variable indépendante, ce qui donne

$$\frac{\partial F(\nu)}{\partial \nu}$$

soit comme fonction de t à travers ν , ce qui donne

$$\frac{dF(\nu(t))}{dt} = \frac{\partial F(\nu)}{\partial \nu} \frac{d\nu(t)}{dt} .$$

Ce double niveau, introduit génialement par Euler, est la clé des équations d'Euler-Lagrange. Il est le plus difficile à comprendre sur le plan pédagogique.

Par exemple, comme nous allons le voir, pour trouver les équations différentielles d'un problème d'optimisation décrit par une fonction $y = f(x)$, sa dérivée première $y'(x)$ et sa dérivée seconde $y''(x)$, Euler traite dans un premier temps explicitement y' et y'' comme des variables *indépendantes* p et q pour, seulement dans un second temps, tenir compte du fait qu'il s'agit de dérivées. Mais, pour l'approche neurogéométrique, cette stratégie calculatoire est tout à fait naturelle. En effet, traiter ν comme une variable indépendante c'est l'implémenter dans une famille de *modules encapsulés* (comme par exemple les hypercolonnes d'orientation) dont on ne regarde pas la structure interne alors que la traiter comme une fonction, c'est tenir compte à la fois de l'architecture fonctionnelle reliant entre eux les modules et de la structure interne des modules. Autrement dit, les architectures fonctionnelles neuro-physiologiques, avec leurs modules et leurs propriétés d'encapsulation, implémentent dans un hardware neuronal une architecture formelle analogue à celle des modèles variationnels classiques. Il s'agit là d'une confirmation éclatante du bien fondé de la notion de *simplicité* introduite par Alain Berthoz dans [54] et que nous avons déjà discutée à la section 5.4.6. du chapitre 5 du Vol I. En quelque sorte, modularité et simplicité montrent que l'évolution a inventé un hardware neuronal pour le calcul différentiel à la Euler-Lagrange (l'univers mécanique de la matière inerte l'avait déjà fait bien avant, avec des particules, des atomes, des molécules, des forces, des champs, etc.). La neurogéométrie se situe donc bien au-delà de simples simulations numériques. Elle est non seulement computationnelle mais également *théorique* et *structurale*.

3.2. Les classes de modèles

Dans les sections et les chapitres qui suivent, nous allons présenter plusieurs classes de modèles variationnels.

1. D'abord, très brièvement, les modèles pionniers de Ullman et Horn.
2. Ensuite, le modèle remarquable des "elasticæ" introduit par David Mumford qui repose sur un principe de minimisation de la courbure dans l'espace de base M .
3. Puis, notre première classe de modèles de "géodésiques legendriennes" qui reposent sur un principe "géodésique" de minimisation de la longueur dans le fibré $\mathbb{V} = J^1M$ ou $\mathbb{V} = CM$. Ils utilisent une métrique riemannienne sur \mathbb{V} et des équations d'Euler-Lagrange contraintes, et donc des *multiplicateurs* de Lagrange.
4. Enfin notre seconde classe de modèles géodésiques dans $\mathbb{V}_J = J^1M$ et CM ou encore $\mathbb{V}_S = M \times \mathbb{S}^1 = SE(2)$ qui reposent sur la géométrie *sous-riemannienne* de ces fibrés de contact.

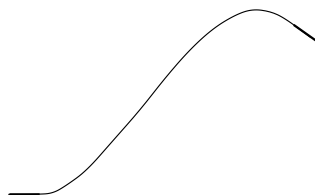


FIGURE 32. Contours illusoires d'énergie minimale de Horn.

Le voyage sera long et ardu et nous permettra de rencontrer un nombre important de concepts fondamentaux de géométrie différentielle.

3.3. Les modèles de Ullman et Horn

À notre connaissance, c'est Shimon Ullman [549] qui introduisit le premier en 1976 l'idée de modèles variationnels pour les contours de Kanizsa courbes dans son papier "Filling-in the gaps : the shape of subjective contours and a model for their generation". Il remarqua que

"An important but hitherto neglected problem posed by the filling-in phenomena concerns the *shape* of the filled-in contours and trajectories"

et développa l'idée que

"A network with the local property of trying to keep the contours 'as straight as possible' can produce curves possessing the global property of minimizing total curvature."

Chez Ullman, l'hypothèse de localité est assez problématique. Elle stipule que la continuation d'un contour en un point ne dépend que de l'orientation du contour en ce point, et non pas de la forme globale du contour. Brady, Grimson et Langridge [75] l'ont abandonnée et ont introduit des contours minimisant la courbure totale. En approximant la courbure κ de la courbe $y = f(x)$

$$\kappa = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}}$$

par $f''(x)$ quand les contours ne sont pas trop incurvés (i.e. $f'(x)^2$ assez petit), ils ont retrouvé les courbes d'interpolation classiques que sont les splines cubiques.

Ensuite Berthold Horn, un disciple d'Ullman, raffina son modèle et introduisit en 1983 [257] des courbes "d'énergie minimale" (cf. figure 32).

3.4. Le modèle des elasticæ de Mumford

Puis en 1992 David Mumford introduisit dans son célèbre article "Elastica and Computer Vision" (Mumford [381], cf. aussi Nitzberg, Mumford, Shiota [390] et

Williams, Jacobs [568]) les *elastica*⁷ qui sont des courbes minimisant à la fois la longueur et l'intégrale du carré de la courbure κ des courbes, c'est-à-dire une énergie du type

$$E = \int_{\gamma} (\alpha\kappa^2 + \beta) ds$$

où γ est une courbe dans \mathbb{R}^2 d'élément de longueur d'arc ds . On peut justifier ce modèle de la façon suivante. Le contour virtuel correspondrait à une chaîne d'éléments de contact (a_i, p_i) le long de laquelle la perte d'activité est la plus faible possible. Mais les pertes ont une double origine :

- une perte proportionnelle au nombre N d'éléments de la chaîne, avec un facteur constant β ;
- une perte due à la courbure, égale à la somme des déflexions d'orientation entre éléments consécutifs, avec un facteur constant α . Si θ_i est l'angle de la pente p_i , on peut prendre par exemple $\sum_{i=1}^{i=N-1} (\theta_{i+1} - \theta_i)^2$.

À la limite, le nombre d'éléments dans le premier terme tend vers la longueur $\int_{\gamma} ds$, et la somme des déviations $\Delta\theta$ dans le second terme tend vers l'intégrale de la courbure $\int_{\gamma} \kappa^2 ds$ puisque par définition $\kappa = \frac{d\theta}{ds}$. La minimisation des pertes "en ligne" conduit par conséquent au problème variationnel :

$$\min \left(\int_{\gamma} (\alpha\kappa^2 + \beta) ds \right) \quad (1)$$

avec les conditions au bord :

$$\begin{cases} f(x_A) = y_A, f'(x_A) = \tan(\theta_A) \\ f(x_B) = y_B, f'(x_B) = \tan(\theta_B) \end{cases}$$

où A et B sont les extrémités de γ .

Or ce problème variationnel est bien connu en théorie de l'élasticité et remonte à Euler. Ses solutions sont des courbes transcendantes (i.e. non algébriques) appelées *elastica* que l'on peut représenter explicitement au moyen de fonctions elliptiques. Nous allons y revenir longuement. La figure 33 en donne quelques exemples.

David Mumford a pour sa part développé une explication *stochastique* de son modèle. On suppose que la courbure $\kappa(s)$ de la courbe γ (paramétrée par sa longueur d'arc s) est un bruit blanc. Comme $\kappa(s) = \dot{\theta}(s)$ (où $\theta(s)$ est la pente de γ et $\dot{\theta}(s)$ la dérivée $\frac{d\theta(s)}{ds}$ de $\theta(s)$ par rapport à s), cela signifie que $\theta(s)$ est un *mouvement*

7. Comme l'introduction de la déclinaison latine serait bizarre, nous utiliserons le terme "elastica" comme un nom féminin se substituant par métonymie à "curva elastica" ou "courbe du modèle de l'elastica". Nous utiliserons "elasticae" pour le pluriel mais nous n'utiliserons pas les cas latins (nominatif, accusatif, datif, etc.). Comme beaucoup d'autres auteurs, Mumford utilise "elastica" comme un nom invariable.

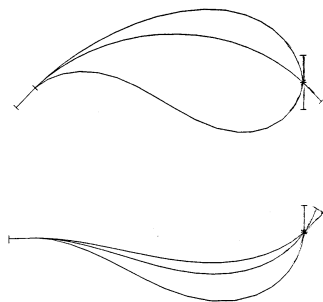


FIGURE 33. Quelques exemples d'elasticæ.

brownien, autrement dit que, en chaque instant, le mouvement est une variable aléatoire gaussienne de moyenne nulle et de variance σ .⁸ Si l'on suppose de plus que la longueur ℓ de γ est une variable aléatoire qui suit une loi exponentielle $\lambda e^{-\lambda \ell} d\ell$ (ℓ est donc constante pour $\lambda = 0$), alors la probabilité d'une courbe γ particulière est donnée par :

$$Pr(\gamma) = e^{-\int (\alpha \kappa^2 + \beta) ds} \quad (2)$$

avec $\alpha = \frac{1}{2\sigma^2}$ et $\beta = \lambda$. Ce sont par conséquent les elasticæ qui sont les courbes les plus probables. Comme l'affirme Mumford ([381], p. 496) :

“Thus we see that elastica have the interpretation of being the *mode* of the probability distribution underlying this stochastic process restricted to curves with prescribed boundary behavior, e.g., the maximum likelihood curve with which to reconstruct hidden contours.”

Les elasticæ sont solutions d'une équation différentielle du second ordre en la courbure κ . On en trouvera une dérivation “à la main” dans l'article de Mumford. On part du fait bien connu (méthode du repère mobile, cf. aussi plus bas les sections 5 et 5.1) que si $t(s)$ et $n(s)$ sont respectivement la tangente unitaire et la normale unitaire à γ au point $a(s)$, alors

$$\begin{cases} \dot{a}(s) = t(s) \\ \dot{t}(s) = \kappa(s)n(s) \\ \dot{n}(s) = -\kappa(s)t(s) . \end{cases}$$

On considère ensuite une déformation $a_\delta(s) = a(s) + \delta(s).n(s)$ de $a(s)$ (*i.e.* de γ). En dérivant et en explicitant la condition de préservation des longueurs on obtient une estimation au premier ordre de la courbure $\kappa_\delta(s)$ de $a_\delta(s)$:

8. Nous reviendrons techniquement sur le mouvement brownien à propos des processus de diffusion dans la section 9.1.5 du chapitre 17.

$$\kappa_\delta(s) = \kappa(s) + \ddot{\delta}(s) + \delta(s)\kappa(s)^2. \quad (3)$$

En recalculant l'intégrale $\int \kappa(s)^2 ds$, en développant au premier ordre et en faisant une intégration par parties en tenant compte de la condition $\int \delta(s)\kappa(s)ds = 0$ (exprimant que la longueur $\int ds$ reste constante), on obtient la contrainte :

$$\int (2\ddot{\kappa}(s) + \kappa(s)^3) \delta(s)ds = 0 \quad \text{si} \quad \int \delta(s)\kappa(s)ds = 0.$$

Comme $\delta(s)$ est une perturbation arbitraire, cela implique que $2\ddot{\kappa} + \kappa^3$ soit proportionnel à κ . D'où l'équation différentielle :

$$2\ddot{\kappa}(s) + \kappa(s)^3 = b\kappa(s).$$

En multipliant par $\dot{\kappa}(s)$ et en intégrant, on obtient :

$$\dot{\kappa}(s)^2 + \frac{1}{4}\kappa(s)^4 = \frac{b}{2}\kappa(s)^2 + c \quad (4)$$

où c est une constante d'intégration.

La relation entre κ et s est donc donnée par une intégrale elliptique :

$$s = \int \frac{2d\kappa}{\sqrt{-\kappa^4 + 2b\kappa^2 + 4c}}.$$

Si E est la courbe elliptique d'équation :

$$v^2 = -u^4 + 2bu^2 + 4c,$$

l'application $s \mapsto (\kappa(s), 2\dot{\kappa}(s))$ envoie l'elastica γ sur E . En utilisant la théorie classique des courbes elliptiques, Mumford montre alors comment un elastica peut être paramétré par ce que l'on appelle des "fonctions théta".

4. Des elasticæ dans \mathbb{R}^2 aux géodésiques dans \mathbb{V}

4.1. Relèvements dans les espaces de jets

Les modèles d'Ullman, Horn et Mumford conçoivent les contours illusoires comme des courbes planes dans l'espace de base $M = \mathbb{R}^2$. Mais il est plus naturel de se placer dans $V1$ et de chercher des courbes solutions d'un problème variationnel dépendant de son architecture fonctionnelle.

Dans le cadre géométrique d'une fibration $\pi : \mathbb{V} = M \times P \rightarrow M$ idéalissant $V1$, nous pouvons expliquer l'effet Kanizsa de façon principielle. Deux pacmen de centres respectifs A et B avec un certain angle d'ouverture définissent deux éléments de contact $(a_1, p_1) = v_1$ et $(a_2, p_2) = v_2$ de \mathbb{V} . Un K -contour interpolant entre (a_1, p_1) et (a_2, p_2) est

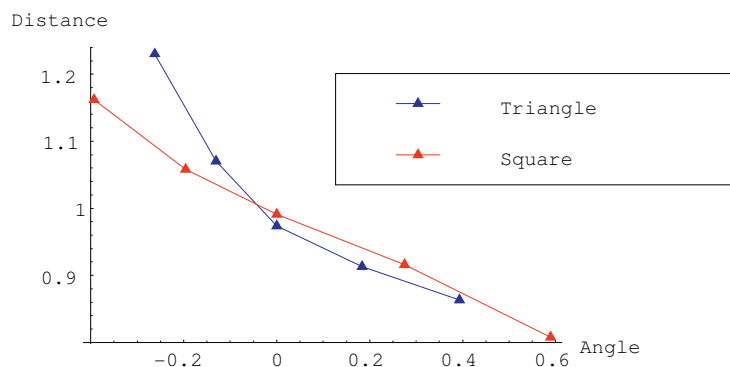


FIGURE 34. Comparaison du point extrémal de K -contours ayant les mêmes conditions au bord mais immergés dans des triangles ou des carrés. Leur différence montre que le système visuel calcule plutôt des surfaces minimales que des arcs minimaux.

1. une courbe γ joignant a_1 à a_2 dans M avec une tangente p_1 en a_1 et une tangente p_2 en a_2 (on peut supposer pour simplifier que γ est définie par une équation $y = f(x)$);
2. une courbe minimisant un lagrangien, c'est-à-dire une fonctionnelle définissant une sorte d'“énergie” pour γ (problème variationnel).

Si nous relevons le problème dans \mathbb{V} , nous devons trouver dans \mathbb{V} une courbe Γ de la forme $(x, y = f(x), p = g(x))$ interpolant entre (a_1, p_1) et (a_2, p_2) dans \mathbb{V} , et qui est dans le même temps :

1. “as straight as possible” comme le disait Ullman, c'est-à-dire “géodésique” dans \mathbb{V} – et comme la variation de p mesure la courbure κ de γ , il s'agit de minimiser d'une certaine façon la courbure;
2. une courbe intégrale de la structure de contact, c'est-à-dire une courbe satisfaisant la condition d'intégrabilité $g(x) = f'(x)$.

Nous allons décrire toute une chaîne de modèles, partant de modèles riemanniens de géodésiques “legendriennes” dans \mathbb{V}_J pour aboutir à des modèles géodésiques sous-riemanniens dans \mathbb{V}_J et \mathbb{V}_S fondés sur des métriques sous-riemanniennes $d_{\mathcal{K}}$ (cf. le chapitre 5).

4.2. Contours illusoires et surfaces minimales

En fait, le problème de la modélisation des CIs est encore plus complexe. En effet, l'expérience avec Jacques Ninio (section 2.4) montre que les déflexions ne sont pas les mêmes pour les mêmes conditions aux limites suivant qu'elles sont immergées dans un triangle ou un carré (cf. figure 34). Leur différence montre que le système visuel calcule plutôt des *surfaces minimales* que des arcs minimaux.

Il s'agit d'un effet contextuel global très intéressant car des surfaces virtuelles sont solutions d'un problème variationnel plus complexe sur le groupe de Lie \mathbb{V} muni de sa structure de contact \mathcal{K} et de sa métrique sous-riemannienne $d_{\mathcal{K}}$. Or le problème des surfaces minimales pour des métriques sous-riemanniennes est difficile et encore presque inexploré. Quelques résultats sont disponibles, notamment ceux de Scott Pauls [401]. Le premier point est de définir une mesure appropriée des surfaces S dans \mathbb{V} . Mikhaïl Gromov [228] a proposé de prendre la mesure de Hausdorff 3-dimensionnelle H^3 de S associée à $d_{\mathcal{K}}$.⁹ Pansu [397] a montré que (avec les notations de la section 2 du chapitre 5)

$$H^3(S) = \int \|\nabla_{\mathcal{K}}(f)\| d\sigma$$

où $f = 0$ est une équation de S et $d\sigma$ est l'élément d'aire riemannien induit par la métrique invariante à gauche g rendant orthonormale la base invariante à gauche $\{t_1, t_2, t_3\}$ de l'algèbre de Lie de \mathbb{V} .

Le lecteur trouvera dans les articles [401], [252] de Scott Pauls et Robert Hladky un cadre variationnel pour les surfaces minimales dans \mathbb{V} qui utilise comme lagrangien la projection $n_{\mathcal{K}}^0$ du vecteur normal n de S sur \mathcal{K} (le vecteur $n_{\mathcal{K}}$ de la section 7 du chapitre 5 est la renormalisation de $n_{\mathcal{K}}^0$).

Nous ne tiendrons pas compte ici de ces effets de contexte globaux et nous nous restreindrons pour simplifier à des *courbes* minimales, le problème simplifié étant déjà, nous allons le voir, d'une certaine complexité mathématique.

4.3. Lagrangiens et métriques

Nous devons donc définir maintenant des lagrangiens appropriés sur \mathbb{V} et étudier les courbes Γ qui sont solutions des équations d'Euler-Lagrange associées. Dans notre premier modèle nous allons présenter trois possibilités, l'une très naturelle mais analytiquement complexe, les deux autres moins naturelles mais plus facile à formuler analytiquement. Nous verrons que les projections de ces courbes sur l'espace de base M présentent de fortes analogies avec les elasticæ.

Nous utiliserons ensuite les ressources de la géométrie sous-riemannienne pour améliorer cette première classe de modèles et en faire de vrais modèles géodésiques. Ce sera ensuite, répétons-le, à l'expérience de décider entre les modèles.

4.4. "Géodésiques legendriennes" dans le fibré de contact \mathbb{V}_J

Pour définir des lagrangiens dans le fibré de contact $\mathbb{V} = CM$ ou l'espace des 1-jets $\mathbb{V} = J^1M$ et plus précisément $\mathbb{V}_J = J^1\mathbb{R}^2$, nous devons d'abord définir des métriques dans \mathbb{V} . Une première idée (que nous améliorerons par la suite) est de prendre des métriques riemanniennes et d'y refléter l'*affaiblissement* des connexions

9. Rappelons que la dimension de Hausdorff de \mathbb{V} pour $d_{\mathcal{K}}$ est 4 et non pas 3.

horizontales cortico-corticales quand l'écart entre les valeurs aux bords θ_A et θ_B augmente. Si θ est mesuré relativement à l'axe AB (ce qui lui donne un sens géométrique intrinsèque), l'affaiblissement doit être nul pour $\theta = 0$ et $\theta = \pi$ (force maximale des connexions) et diverger pour $\theta = \frac{\pi}{2}$. La fonction $p = f' = \tan \theta$ étant une fonction très simple partageant ces propriétés, il semble justifié de tester d'abord la métrique euclidienne de $\mathbb{V}_J = J^1\mathbb{R}^2$ mais en se restreignant à des géodésiques entre deux éléments de contact (x_1, y_1, p_1) et (x_2, y_2, p_2) tels que $y_1 = y_2$. Nous choisissons donc un repère Oxy de M où l'axe des x est identifié à AB . L'invariance par changement de repère est alors exprimée par l'action du groupe euclidien $SE(2)$ sur \mathbb{V}_J (cf. la section 2.4 du chapitre 3).

Nous allons donc étudier dans un premier temps les courbes Γ de longueur minimale dans \mathbb{V}_J pour sa métrique euclidienne, mais, évidemment, uniquement celles qui sont des relevées legendriennes, c'est-à-dire qui satisfont la condition d'intégrabilité et sont des courbes intégrales de la structure de contact \mathcal{K} . Nous les appellerons des "géodésiques legendriennes". Mais avant d'effectuer quelques calculs, nous allons faire un assez long détour par l'histoire des modèles variationnels classiques et leur univers foisonnant.

CHAPITRE 9

L'univers des modèles variationnels classiques

Nous présentons dans ce chapitre quelques éléments de l'histoire du calcul des variations classique pour quatre raisons :

1. D'abord il s'agit du contexte naturel des algorithmes d'optimisation. Les modèles comme ceux des élastiques ou des géodésiques réactivent ce qui a été mis au point au XVIII^e siècle pour résoudre les équations différentielles et élaborer les modèles de nombreux systèmes mécaniques auxquels les contours illusoires ressemblent beaucoup. Il s'agit de modèles à la Euler-Lagrange-Hamilton. Notre problème appartient donc à l'immense univers mathématique qui s'est mis en place à l'aube du calcul différentiel et intégral et qui, entre le milieu du XVII^e et le milieu du XIX^e, a conduit à la théorie moderne des fonctions elliptiques.
2. Ensuite cette réactivation du calcul variationnel classique permet de préciser la façon dont ses algorithmes sont neuralemement implémentés, cette implémentation étant l'apport principal de la neurogéométrie.
3. Ensuite également parce que, en retour, cette implémentation neuronale focalisée sur des hardware modularisés qui sont dédiés à des tâches computationnelles précises réactive tous les principaux problèmes fondationnels du calcul différentiel, intégral et variationnel classique.
4. Enfin parce que les nombreux calculs que ces premiers spécialistes ont dû mener à bien leur ont révélé l'écart abyssal qui existe entre des courbes simples comme les coniques et des courbes que, comme le disait joliment Jacques Bernoulli en 1694, "la nature produit elle-même sans aucun artifice, d'un mouvement rapide et pour ainsi dire instantané".¹ Or les contours illusoires que nous étudions sont des courbes *naturelles*, elles aussi produites spontanément par des mécanismes naturels d'optimisation et il est normal qu'ils relèvent de calculs analogues. Il s'agit là d'un point méthodologique et philosophique important. Dans les cas qui nous intéressent ici, la complexité des calculs n'est pas liée à une complexité intrinsèque de la nature mais au fait que la nature n'est pas régie par des formules algébriques simples mais par des *équations différentielles* simples. Or les solutions d'équations différentielles très simples

1. Cf. [51], p. 188. Cité dans Ivahn Smadja [505], p. 8.

peuvent être très compliquées à calculer. C'est la grande leçon de la mécanique newtonienne-leibnizienne.

Nous allons commencer par quelques brefs repères biographiques pour ensuite résumer les principales étapes ayant conduit aux modèles dont nous nous inspirons.

1. Quelques repères biographiques

Ces repères biographiques sont minimaux et ne retiennent de la pléiade des génies cités (dont les travaux sont souvent immenses) que quelques éléments pertinents pour notre propos. Le lecteur pourra trouver tous les compléments souhaitables en ligne, dans des sources comme Wikipedia.

1. **René Descartes** (1596-1650). Introduteur de la géométrie analytique qui a permis de développer le calcul différentiel et intégral comme calcul. *La géométrie* de 1637 est un Appendice au *Discours de la Méthode*. Elle unifie algèbre et géométrie et présente algébriquement des éléments de géométrie classique remontant à l'Antiquité.

2. **Gilles Personne de Roberval** (1602-1675). Bien connu pour ses découvertes expérimentales, il fut un précurseur de Newton et du calcul des fluxions comme méthode de construction des tangentes à une courbe. Le problème, dit à l'époque "problème de la tangente", n'est pas trivial car, si la courbe C est d'équation $y = f(x)$, il ne s'agit pas seulement de calculer $f'(x)$ mais de *construire* géométriquement la tangente en un point quelconque. Il inventa avant Cavalieri les "indivisibles" précurseurs des infinitésimales. Spécialiste des quadratures, il réussit en 1637 la quadrature d'un arc de cycloïde dont l'équation paramétrique en fonction de l'angle θ des coordonnées polaires (ρ, θ) est $x(\theta) = R(\theta - \sin(\theta))$, $y(\theta) = R(1 - \cos(\theta))$.

3. **Pierre de Fermat** (1605/1608-1665). Il introduisit en 1636 la méthode d'*adégalisation*, précurseur du calcul différentiel et qui conduira plus tard à la formule classique des dérivées $f'(a) = \lim_{\varepsilon \rightarrow 0} \frac{f(a+\varepsilon) - f(a)}{\varepsilon}$. On développe $f(a + \varepsilon)$ puis on soustrait $f(a)$ puis on divise par ε et on fait enfin $\varepsilon = 0$ dans le résultat. En fait, cela revient à développer $f(a + \varepsilon)$ en supposant que ε est une infinitésimale *nilpotente* : $\varepsilon \neq 0$ mais $\varepsilon^2 = 0$. L'histoire de cette idée dont la cohérence a fait longtemps question est fascinante. Nous avons vu à la section 6.2.2 du chapitre 2 traitant de la notion d'infinitésimale symbolique qu'elle fut explicitement défendue par Bernard Nieuwentijt (1654-1718), un contemporain de Leibniz, mais fut violemment accusée d'inconsistance par le même Leibniz, dont le calcul différentiel symbolique était lui-même violemment accusé d'inconsistance par Berkeley et d'autres philosophes. Et pourtant c'est sur cette base que, deux siècles et demi plus tard, Alexandre Grothendieck pourra, à la suite de travaux précurseurs de Cartan et de Weil, développer

un calcul différentiel pour une géométrie algébrique où la notion de limite $\varepsilon \rightarrow 0$ n'a pas de sens.² Comme le dit Jean Dieudonné :

“D’une façon générale, c’est dans les éléments nilpotents que la théorie des schémas³ trouve l’équivalent algébrique des phénomènes infinitésimaux”.

([144], I, p.202)

4. **Evangelista Torricelli** (1608-1647). Il croise notre propos parce qu’il a réussi en 1640 la rectification géométrique de la spirale logarithmique (d’équation $\rho = a\theta$ en coordonnées polaires).

5. **John Wallis** (1616-1703). Précurseur de Newton. En 1655-1656 il réussit la rectification géométrique de l’ellipse (d’ailleurs sans soulever un grand intérêt). Dans *Arithmetica Infinitorum* il introduisit le symbole ∞ et l’infinitésimale $\frac{1}{\infty}$.

6. **Christian Huygens** (1629-1695). En dehors de ses travaux bien connus, il réussit la rectification géométrique de certaines courbes, par exemple en ramenant en 1657 la rectification de la parabole à la quadrature de l’hyperbole. Il montra en particulier l’isochronisme de la cycloïde, à savoir que si un pendule est contraint de parcourir un arc de cycloïde alors sa période d’oscillation devient constante et indépendante de l’amplitude des oscillations (le pendule normal décrivant des arcs de cercle n’est pas isochrone). Il utilisa cette propriété pour faire faire de grands progrès en horlogerie. Mais ses méthodes étaient encore géométriques, pré-calcul différentiel.

7. **Isaac Barrow** (1630-1677). Il thématiza, relativement aux procédures traditionnelles de constructions géométriques, le calcul intégral et le calcul différentiel comme des problèmes inverses l’un de l’autre. Les dérivées correspondent aux problèmes traditionnels de “tangentes” (construire géométriquement la tangente en un point d’une courbe donnée par une procédure de construction) et les intégrales correspondent aux problèmes traditionnels de “quadrature”. Il fut le premier à démontrer que $(x^n)' = nx^{n-1}$, ce qui n’a rien d’évident en termes de courbes et de constructions géométriques.

8. **Christopher Wren** (1632-1723). En 1658 il réussit la rectification géométrique de la cycloïde. Mais il est surtout connu comme l’architecte de la cathédrale Saint Paul de Londres et la reconstruction de la capitale après le tragique incendie de 1666.

9. **William Neil** (ou Neile, 1637-1670). Elève de John Wallis, il réussit la rectification géométrique de la parabole semi-cubique $y^2 = x^3$ en 1657.

10. **Isaac Newton** (1643-1727). Le père fondateur, avec Leibniz, du calcul différentiel et intégral systématique et général. Dès 1669⁴, dans *De analysi per aequationes numero terminorum infinitas* (publié seulement en 1711), s’inspirant

2. Le lecteur intéressé par cette histoire pourra consulter notre étude [430].

3. Le nom donné par Grothendieck à la géométrie algébrique moderne.

4. Les fameux “Principia” – *Philosophiae naturalis principia mathematica* – sont de 1687.

de Descartes, Wallis et Barrow, il pensa les fonctions $f(t)$ comme des “fluentes” et leurs dérivées (les vitesses) comme des “fluxions”.⁵ Mais la notion de limite n’est pas claire chez Newton. Le calcul différentiel était chez lui inséparable de la Mécanique. Sa seconde loi $f = m\gamma$ deviendra une source inépuisable d’équations différentielles naturelles mais de plus en plus difficiles à résoudre.

11. **Gottfried Wilhelm Leibniz** (1646-1716). Le père fondateur, avec Newton, son exact contemporain, du calcul différentiel et intégral systématique et général. En 1674, il introduisit les notations symboliques et les formules de base pour les différentielles df : d est une opération linéaire, $da = 0$ si a est une constante, pour les produits $d(fg) = fdg + gdf$, pour les quotients $d\left(\frac{f}{g}\right) = \frac{gdf - f dg}{g^2}$. Toutes les généralisations des concepts de dérivée et de différentielle (jusqu’aux plus sophistiquées d’aujourd’hui) ont continué à utiliser ces formules leibniziennes universelles. C’est avec Leibniz que le *Calculus* est devenu véritablement un calcul. Ce calcul symbolique fut violemment attaqué comme contradictoire (entre autres par Berkeley) parce que, comme nous l’avons déjà évoqué à la section 6.3 du chapitre 2, les symboles dx ne peuvent pas avoir de référent numérique. Il faudra l’analyse non standard d’Abraham Robinson pour résoudre ce paradoxe. Très jeune, dès 1666, Leibniz commença à s’intéresser aux courbes “transcendantes”, c’est-à-dire qui ne sont pas des graphes de fonctions algébriques.

12. **Jacques (Jakob) Bernoulli** (1654-1705), inséparable de son frère Jean (Johann) Bernoulli (1667-1748). En 1690, il résolut le problème de la courbe “isochrone” étudié par Huygens en 1687 et Leibniz en 1689. En 1691 il développa la théorie mécanique des elasticæ qui deviendra l’un des chefs-d’oeuvre du calcul variationnel avec Euler. La différentielle associée au problème étant $dy = \frac{x^2 dx}{\sqrt{a^4 - x^4}}$, il s’agit de calculer l’intégrale inconnue $y(x) = \int_0^x \frac{t^2 dt}{\sqrt{a^4 - t^4}}$. Il pensait que seuls des développements en série permettaient de le faire par approximations successives. Pour calculer les longueurs d’arcs, l’élément de longueur d’arc est $ds = \sqrt{1 + \frac{x^4}{a^4 - x^4}} dx = \frac{a^2}{\sqrt{a^4 - x^4}} dx$. La rectification se ramène ainsi à la quadrature d’une courbe dite isochrone paracentrique. En 1694 Leibniz et Jakob Bernoulli résolurent le problème. Puis Jakob étudia la lemniscate. Il étudia aussi les développées des courbes (lieu des centres de courbures) et les caustiques en optique. Ce fut un grand probabiliste. Les frères Bernoulli résolurent de nombreux problèmes avec la conception systématique et générale du calcul différentiel et intégral introduite par Leibniz.⁶

5. Pour l’histoire des fluxions et de l’analyse différentielle chez Newton on pourra se référer à l’ouvrage [398] de Marco Panza.

6. Pour l’histoire compliquée de la relation entre les deux frères, on pourra consulter l’étude [404] de Jeanne Peiffer.

13. **Pierre Varignon** (1654-1722). Il traduisit la mécanique de Newton basée sur les “fluxions” en problèmes à la Leibniz en utilisant des expressions symboliques comme $y' = \frac{dy}{dx}$, $y'' = \frac{d^2y}{dx^2}$, etc. Mais il se heurta à l’hostilité philosophique envers l’usage de ces symboles différentiels sans référents numériques possibles.

14. **Guillaume de l’Hospital** (1661-1704). Il publia en 1696 son *Analyse des infiniment petits pour l’intelligence des lignes courbes*.

15. **Jean (Johann) Bernoulli I** (1667-1748, père de Daniel, Jean et Nicolas). En 1690 il rectifia la chaînette.⁷ Il résolut le problème de la courbe brachistochrone. Il développa le calcul intégral des fonctions rationnelles et montra que le logarithme est la primitive de $\frac{1}{x}$. À partir de 1693 il poursuivit une vaste correspondance avec Leibniz qu’il soutenait, avec son frère Jacques, dans la querelle avec Newton. En 1702, il donna les formules des intégrales de toutes les fonctions rationnelles comme combinaisons linéaires de fonctions rationnelles et de fonctions logarithmiques. Il fut le maître d’Euler.

16. **Abraham De Moivre** (1667-1754). Ami de Newton, Leibniz (et aussi de Halley). Il fut choisi en 1710 par la Royal Society (dont il était membre depuis 1697) pour arbitrer leur querelle. Il publia en 1707 sa fameuse formule de trigonométrie rendant immédiates les formules d’addition des fonctions circulaires par l’utilisation des nombres complexes.

17. **Giulio Carlo Conte di Fagnano Marchese dé Toschi** (1682-1766). Spécialiste de la géométrie du triangle. Il démontra que $\pi = 2i \log \frac{1-i}{1+i}$. Ses travaux (vers 1718) sur la division des arcs de lemniscate et sur les rapports géométriques remarquables entre lemniscate, ellipse et hyperbole équilatère – travaux qu’il envoya à Euler pour devenir membre de l’Académie de Berlin et que ce dernier évalua en 1751 – jouèrent un grand rôle dans l’élaboration initiale des fonctions elliptiques.

18. **Brook Taylor** (1685-1731). Il participa en 1712 au comité scientifique jugeant de la controverse entre Newton et Leibniz. Il publia en 1715 son *Methodus incrementorum directa et inversa*. C’est le spécialiste des développements en série qui nous intéresse ici.

19. **Colin MacLaurin** (1698-1746). Il publia en 1742 son monumental *Treatise of Fluxions* systématisant le calcul de Newton. On lui doit la fameuse série de Taylor-MacLaurin pour le développement d’une fonction indéfiniment différentiable au voisinage de 0 :

$$f(x) \rightarrow f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n + \dots$$

La question de savoir si la série converge et si, dans ce cas, elle converge vers $f(x)$ est une question délicate.

7. Rappelons que la chaînette est la courbe (transcendante) que fait une chaîne suspendue par ses extrémités sous l’action de la gravitation.

20. **Daniel Bernouilli** (1700-1782). Fils de Jean et neveu de Jacques. Mathématicien, physicien (hydrodynamique, astronomie), naturaliste, médecin. Membre de nombreuses Académies. Ami d'Euler il lui proposa de résoudre le problème de l'élasticité au moyen du calcul des variations.

21. **Leonhard Euler** (1707-1783). L'un des pères fondateurs du calcul variationnel. Calculateur génial et universel. Il étudia la lemniscate, les *elasticæ*, publia en 1748 *Introductio in analysin infinitorum*. C'est un maître des développements en série permettant de résoudre des équations différentielles que l'on ne sait pas intégrer au moyen de fonctions classiques. Nous lui consacrerons plusieurs sections.

22. **Jean Le Rond D'Alembert** (1717-1783). Il rédigea l'entrée "Calcul différentiel" de l'Encyclopédie. Il y expliqua que les symboles leibniziens comme dx sont des "signes" pour des quantités infiniment petites que l'on ne peut pas connaître numériquement, seuls les rapports $\frac{dy}{dx}$ pouvant avoir un sens numérique.

23. **Joseph-Louis Lagrange** (1736-1813). Il publia en 1797 sa monumentale *Théorie des fonctions analytiques*. Il y allait au-delà des symboles leibniziens et des fluxions newtoniennes pour développer une approche algébrique. Comme Euler, il utilisa des développements en séries entières (en particulier des séries de Taylor). Mais, comme Euler et tant d'autres, il rencontra des problèmes de convergence des séries.

24. **Adrien-Marie Legendre** (1752-1833). Il classa les fonctions elliptiques dès 1786 et leur donna leur nom. En 1794 il publia son *Mémoire sur les transcendentes elliptiques*. Son monumental *Traité des fonctions elliptiques et des intégrales Eulériennes* parut en trois tomes en 1825, 1826, 1828. Nous y reviendrons longuement.

25. **Carl Friedrich Gauss** (1777-1855). Nous croiserons l'œuvre gigantesque du "Mathematicorum Principi" à propos des fonctions elliptiques associées à la lemniscate qu'il étudia dans sa jeunesse en 1796. Il fut l'un des premiers après Lagrange et Legendre à étudier les intégrales elliptiques.

26. **Augustin-Louis Cauchy** (1789-1857). De son *Cours d'Analyse* de 1821 à ses *Leçons sur le calcul différentiel et intégral* de 1829, il fut l'un des fondateurs de la théorie des fonctions d'une variable complexe et l'un des promoteurs de la rigueur en analyse en fondant le calcul différentiel et intégral sur la notion de limite. Cela lui permit de préciser les notions très délicates de continuité, de dérivabilité, d'intégrabilité des fonctions et de sommabilité des séries. Cette rigueur sera parachevée par Weierstrass.

27. **Niels-Henrik Abel** (1802-1829). Dans sa fulgurante carrière, il découvre dans ses articles au *Journal de Crelle* la double périodicité des fonctions elliptiques et leur formule d'addition (l'analogue des formules d'addition pour les fonctions trigonométriques). En 1826, lors d'un voyage à Paris, il n'arriva pas à communiquer

avec Legendre, Poisson et Cauchy. Mais soutenu par Jacobi, il reçut à titre posthume (1830) le grand prix de mathématiques de l'Institut de France.

28. **Karl Gustav Jacob Jacobi** (1804-1851). Pour pouvoir calculer les intégrales elliptiques, il introduisit les *fonctions* elliptiques sn , cn , dn qui portent son nom et développa pour elles l'équivalent de ce qu'est la trigonométrie pour les fonctions circulaires. Il publia son grand traité *Fundamenta nova theoriae functionum ellipticarum* en 1829. Nous y reviendrons longuement.

29. **Joseph Liouville** (1809-1882). Il intervient dans notre histoire comme spécialiste des fonctions d'une variable complexe et par son théorème disant qu'une fonction holomorphe sur une surface de Riemann compacte (en particulier une courbe elliptique) est nécessairement constante. Une fonction analytique non constante sur une telle surface ne peut être que méromorphe (i.e. posséder des pôles).

30. **Karl Weierstrass** (1815-1897). L'un des réformateurs de l'analyse, il y introduisit la "rigueur" nécessaire à la résolution de nombreuses difficultés. Il croise notre problématique à travers la fonction elliptique dite de Weierstrass associée à une courbe elliptique. Avec lui, nous sommes de plain pied dans la théorie moderne des fonctions elliptiques.⁸

31. **Bernhard Riemann** (1826-1866). Ce "prodigieux génie" (dixit Jean Dieudonné) croise notre histoire comme spécialiste des fonctions d'une variable complexe (conditions de Cauchy-Riemann) et fondateur de la théorie moderne des courbes et fonctions elliptiques.

2. Quadratures et rectifications

2.1. Quadratures

Étant donné une courbe Γ construite au-dessus d'un axe et un arc MN de Γ au-dessus d'un intervalle AB de l'axe, le terme classique de "quadrature" réfère à l'aire de la surface située entre MN et AB . Nous apprenons aujourd'hui à l'école que si $y = f(x)$ est l'équation de Γ et si A et B correspondent aux valeurs a et b de x , cette aire est l'intégrale $\int_a^b f(x) dx$. Mais depuis la géométrie euclidienne dans l'Antiquité jusqu'à l'introduction par Descartes de la géométrie analytique et l'émergence du calcul différentiel et intégral dans la seconde moitié du XVII^e, on ne savait utiliser que des méthodes géométriques, souvent *ad hoc*, mais souvent, aussi, incroyablement ingénieuses, pour calculer des quadratures. Avec l'avènement du *Calculus* on découvrit des règles générales de calcul et des formules remarquables explicites et on mit au point des algorithmes. Cela permit d'intégrer toutes les fonctions rationnelles $\frac{P(x)}{Q(x)}$, P et Q étant deux polynômes, Q de degré ≥ 2 . La formule

8. Le lecteur intéressé pourra consulter notre compilation [421] de la résolution par Andrew Wiles de la conjecture de Shimura-Tanyama-Weil au moyen des développements profonds de cette théorie.

de base est $\int x^n dx = \frac{x^{n+1}}{n+1}$. On obtenait ainsi des fonctions rationnelles. D'où une première classe bien définie de fonctions pour lesquelles dérivation et intégration obéissent à un calcul bien réglé. Quant à l'intégrale de $\frac{1}{x}$, on découvrit qu'elle était le logarithme $\log(x)$. Cela permit à Jean Bernoulli de donner en 1702 les formules intégrales de toutes les fonctions rationnelles.

2.2. Rectifications

Ce que l'on a appelé pendant longtemps la rectification d'une courbe Γ consistait à calculer la longueur d'un de ses arcs MN au-dessus du segment AB . "Rectifier" signifiait tendre l'arc MN pour en faire un segment rectiligne et en mesurer la longueur. Le problème remonte à l'Antiquité et Archimède savait déjà qu'il existait une *même* constante universelle π telle que, pour un cercle de rayon R , le périmètre soit $2\pi R$ et l'aire πR^2 . Il savait donc, à la valeur exacte de π près, effectuer la quadrature et rectifier les arcs de cercle.⁹

Mais le problème de la rectification est beaucoup plus compliqué que celui de la quadrature. Cela peut facilement se comprendre en formulant le problème comme on le fait aujourd'hui à l'école. Soit $y = f(x)$ l'équation de Γ et $[a, b]$ l'intervalle de x correspondant à MN . La quadrature est, nous venons de le voir, $\int_a^b f(x) dx$. Mais pour la rectification, il faut intégrer l'élément de longueur d'arc ds de Γ le long de MN . D'après le théorème de Pythagore infinitésimal, $ds^2 = dx^2 + dy^2$ et donc $ds = \sqrt{1 + f'(x)^2} dx$. On doit donc calculer

$$L = \int_A^B ds = \int_a^b \sqrt{1 + f'(x)^2} dx .$$

Mais même si la fonction f est simple, la $\sqrt{\quad}$ rend l'intégrale compliquée. Et les spécialistes du XVII^e découvrirent avec une certaine surprise la monumentalité du problème. Certains, tels Torricelli, Roberval, Huygens, Neil, Wren, réussirent le tour de force de rectifier des courbes particulières comme la parabole, la spirale logarithmique, la parabole semi-cubique ou la cycloïde, mais il ne s'agissait que de cas particuliers. Et tous se heurtèrent à l'impossibilité de rectifier des courbes aussi simples que des arcs d'ellipse au moyen des fonctions connues à l'époque comme les fonctions rationnelles, les logarithmes ou les exponentielles.

Parallèlement à des problèmes d'origine purement géométrique, les Bernoulli, de l'Hopital, Varignon, et leurs collègues aimaient aussi à se lancer des défis pour déterminer des trajectoires mécaniques en résolvant des équations différentielles issues de la loi de Newton. De nombreux problèmes mécaniques se ramenaient à des

9. Le problème dit de la quadrature du cercle est de savoir si π est ou non un nombre algébrique. Il n'a été résolu (par la négative) qu'en 1882 par von Lindemann qui démontra que π est transcendant.

problèmes géométriques d'“intelligence de lignes courbes” à résoudre au moyen des méthodes systématiques et générales du calcul différentiel et intégral. Ils se mirent à étudier par exemple

1. La courbe “isochrone” (Leibniz lance le défi en 1687) étudiée par Huygens en 1687, par Leibniz en 1689 et résolu par Jacques Bernoulli en 1690. “Isochrone” signifie ici courbe de descente constante : quel que soit le point de départ, une particule descendant par gravité le long de la courbe atteint l'extrémité de la courbe en un temps constant.
2. La courbe “brachistochrone” (Jean Bernoulli lance le défi en 1696). Il s'agit de la courbe décrite entre deux points dans le temps le plus bref par un point matériel soumis à la pesanteur. C'est un arc de cycloïde.
3. La caténaire (chaînette). Cf. plus haut.
4. La tractrice ou courbe équitangentielle. C'est la courbe que décrit un point du plan de façon à ce que la longueur du segment de tangente entre un point de la courbe et l'axe des x soit constante.

3. La rectification de l'ellipse

Un des problèmes centraux fut celui de la rectification de l'ellipse. Il avait l'air pourtant bien simple, mais il conduisit à un enrichissement considérable de l'analyse fonctionnelle. John Wallis commença à s'en occuper en 1655. Formulons-le comme on le fait aujourd'hui à l'école. Soit

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (a > b > 0)$$

l'équation cartésienne de l'ellipse E (si $a = b$, c'est un cercle de rayon a). À l'époque on utilisait surtout la définition géométrique : une ellipse est le lieu des points M tels que $F_1M + MF_2 = \text{cste}$ ($> F_1F_2$) où F_1 et F_2 sont les deux foyers. L'aire S de E , $S = \pi ab$, s'obtient trivialement à partir de celle du cercle πR^2 en faisant des changements d'échelle sur les axes. Mais il en va tout autrement pour la longueur d'arc. Un symptôme de la difficulté du problème est que, contrairement au cas du cercle, il n'existe pas une même constante (comme π) reliant les aires et les arcs.

Pour calculer l'élément de longueur d'arc $ds^2 = dx^2 + dy^2$ de E , plaçons nous dans le premier quadrant $x \in [0, a]$ et $y \in [0, b]$ ou $y = \frac{b}{a}\sqrt{a^2 - x^2}$. On a :

$$\begin{aligned}
 ds^2 &= dx^2 + dy^2 = \left(1 + \left(\frac{b}{a} \frac{1}{2} \frac{(-2x)}{\sqrt{a^2 - x^2}} \right)^2 \right) dx^2 \\
 &= \frac{a^4 - x^2(a^2 - b^2)}{a^2(a^2 - x^2)} dx^2 \\
 ds &= \sqrt{\frac{a^4 - x^2(a^2 - b^2)}{a^2(a^2 - x^2)}} dx .
 \end{aligned}$$

La formule est donc compliquée. Il est traditionnel de diviser numérateur et dénominateur par a^2 et d'introduire l'excentricité k définie par $k^2 = \frac{a^2 - b^2}{b^2} \geq 0$ ($k = 0$ correspond au cercle). On peut aussi, en utilisant une homothétie, supposer $a = 1$. On obtient alors comme longueur de l'arc de E au-dessus de l'intervalle $[0, x]$:

$$L = \int_0^x \sqrt{\frac{1 - k^2 t^2}{1 - t^2}} dt .$$

Il s'agit de l'intégrale princeps dont l'"intelligence" conduira à la théorie des intégrales et des fonctions elliptiques.

Remarque. Pour tenir compte des autres quadrants, on utilise les symétries de E . □

On peut formuler les choses en utilisant d'autres coordonnées et en posant $x = a \sin(\theta)$, $y = b \cos(\theta)$, l'équation de E devenant alors simplement l'identité remarquable $\sin^2(\theta) + \cos^2(\theta) = 1$.¹⁰ L'élément de longueur d'arc est alors donné par

$$ds^2 = (a^2 \cos^2(\theta) + b^2 \sin^2(\theta)) d\theta^2$$

Mais $a^2 \cos^2(\theta) = a^2(1 - \sin^2(\theta))$ et donc

$$ds^2 = a^2(1 - k^2 \sin^2(\theta)) d\theta^2$$

la longueur L devenant, si $a = 1$ et si $[0, \varphi]$ correspond à $[0, x]$,

$$L = \int_0^\varphi \sqrt{1 - k^2 \sin^2(\theta)} d\theta$$

qui est l'intégrale précédente avec le changement de variable $t = \sin(\theta)$.

10. L'angle θ est donc compté dans le sens direct à partir du demi-axe $y \geq 0$. Il pourrait sembler plus naturel d'écrire $x = a \cos(\theta)$, $y = b \sin(\theta)$, mais la convention choisie conduit ci-dessous au changement de variable $t = \sin(\theta)$ qui a l'avantage que $t = 0$ correspond à $\theta = 0$.

4. Développements en séries

Devant l'impossibilité d'intégrer les intégrales donnant un arc d'ellipse, les mathématiciens eurent recours aux développements en séries. L'un des grands maîtres de cette stratégie fut Euler. Elle est particulièrement opératoire car elle fournit de puissantes procédures d'approximation et, parfois, conduit à des formules exactes. En effet, supposons qu'une fonction $f(x)$ soit développable en série dans un intervalle $[a, b]$ avec des coefficients $c(n)$ donnés par des formules explicites :

$$f(x) = \sum_{n=0}^{n=\infty} c(n) x^n .$$

Disposer d'une *loi* pour les coefficients, c'est connaître f implicitement et pouvoir la calculer par approximations successives.

Supposons qu'il n'y ait pas de problèmes de convergence et que l'on puisse commuter \sum et \int . Alors, dans un intervalle $[a_0, x] \subseteq [a, b]$, on pourra intégrer terme à terme et utiliser le développement

$$\int_{a_0}^x f(t) dt = \sum_{n=0}^{n=\infty} \frac{c(n)}{n+1} (x^{n+1} - a_0^{n+1})$$

Or pour les intégrales elliptiques ci-dessus, il est facile de faire des développements en série des $\sqrt{\quad}$. On a donc une loi pour les coefficients.

Mais au-delà de ces développements en série, émergea progressivement l'hypothèse que, derrière ces intégrales "elliptiques" se cachait un calcul qui ressemblait à la trigonométrie, tant il est vrai que la trigonométrie n'est rien d'autre que la rectification des arcs de cercles, l'intégrale princeps étant, puisque $k = 0$,

$$L = \int_0^x \frac{1}{\sqrt{1-t^2}} dt .$$

Il est indispensable de comprendre ce lien entre d'un côté la trigonométrie et les fonctions "circulaires" sin et cos et d'un autre côté les intégrales "circulaires" si l'on veut comprendre l'histoire des fonctions elliptiques et des intégrales elliptiques introduites par Abel et Jacobi.

5. Fonctions circulaires et trigonométrie

La trigonométrie repose sur le lien entre les fonctions circulaires sin et cos et la rectification des arcs de cercle. Avec les mêmes conventions qu'à la section précédente, la longueur de l'arc du cercle unité dans le premier quadrant au-dessus de l'intervalle $[0, x]$ est l'intégrale "circulaire" $L = \int_0^x \frac{1}{\sqrt{1-t^2}} dt$. C'est la mesure en radians de l'angle φ correspondant à x .

5.1. Des intégrales circulaires aux fonctions circulaires

Commençons ici par quelques remarques qui pourront paraître triviales mais ne le seront plus du tout lorsque nous passerons aux intégrales elliptiques. Si nous disposons déjà des fonctions circulaires, alors nous pouvons faire le changement de variable $t = \sin(\theta)$ et nous obtenons

$$L = \int_0^{\varphi = \text{Arcsin}(x)} \frac{\cos(\theta)}{\sqrt{1 - \sin^2(\theta)}} d\theta = \int_0^{\varphi = \text{Arcsin}(x)} d\theta = \varphi.$$

Mais cela montre que l'on peut *en sens inverse* définir la fonction circulaire sin directement à partir de l'intégrale "circulaire". En effet, notons $u(x)$ l'intégrale

$$u(x) = \int_0^x \frac{1}{\sqrt{1-t^2}} dt$$

considérée comme fonction de sa borne supérieure d'intégration. Nous pouvons alors considérer x comme la fonction *réciproque* de $u(x) = \varphi$. Notons $\sin(\varphi)$ cette fonction dont nous ne disposons pas encore et dont nous ne connaissons pas déjà les propriétés. Par définition, $x = \sin(\varphi)$ et nous retrouvons l'intégrale initiale $u(x) = \varphi$ comme fonction réciproque de la nouvelle fonction, i.e. $\varphi = \text{Arcsin}(x)$ (la réciproque de la réciproque est l'identité). L'intégrale circulaire $L(x) = \int_0^x \frac{1}{\sqrt{1-t^2}} dt$ rectifiant le cercle définit ainsi, par fonction réciproque, une "fonction circulaire". C'est exactement cette stratégie de définition de nouvelles fonctions qu'introduiront Abel et Jacobi à partir des intégrales elliptiques.

Toute la trigonométrie découle de cette définition du sin comme fonction circulaire. D'abord on a évidemment $\sin(0) = \int_0^0 = 0$. On a aussi $\sin\left(\frac{\pi}{2}\right) = 1$ puisque $x = 1$ correspond au quart de cercle $\frac{2\pi}{4} = \frac{\pi}{2}$. Mais montrer formellement, sans se référer à la géométrie du cercle, que

$$L(1) = \int_0^1 \frac{1}{\sqrt{1-t^2}} dt = \frac{\pi}{2}$$

est un très beau résultat dû à Euler.¹¹ Par ailleurs, si l'on change t en $t' = -t$, on obtient

$$\int_0^x \frac{1}{\sqrt{1-t^2}} dt = - \int_0^{-x} \frac{1}{\sqrt{1-t'^2}} dt'$$

et donc $u(-x) = -u(x) = -\varphi$, i.e. $\sin(-\varphi) = -x = -\sin(\varphi)$ et la fonction sin peut être directement prolongée au second quadrant $x \leq 0, y \geq 0$ en tant que fonction impaire. On peut la prolonger aux autres quadrants en utilisant les symétries du cercle mais il est plus intéressant d'utiliser les formules d'addition que nous allons définir.

En remplaçant x par y et en faisant les mêmes opérations, on peut définir une nouvelle fonction $y = \cos(\varphi)$ correspondant à $M(y) = \int_y^1 \frac{1}{\sqrt{1-t^2}} dt$. Cette évidence géométrique se vérifie aisément de façon purement algébrique en faisant le changement de variable $t = \sqrt{1-t'^2}$, $t, t' \in [0, 1]$.¹² En effet, on a

$$dt = \frac{1}{2} \frac{(-2t')}{\sqrt{1-t'^2}} dt', \quad \sqrt{1-t^2} = \sqrt{1-(1-t'^2)} = t'$$

et donc

$$\frac{1}{\sqrt{1-t^2}} dt = - \frac{t'}{\sqrt{1-t'^2}} \frac{dt'}{t'} = - \frac{1}{\sqrt{1-t'^2}} dt' .$$

On a donc bien

$$M(y) = - \int_1^{y=\sqrt{1-x^2}} \frac{1}{\sqrt{1-t^2}} dt = \int_{y=\sqrt{1-x^2}}^1 \frac{1}{\sqrt{1-t^2}} dt = \varphi .$$

D'où la nouvelle fonction $v(y) = \varphi$ dont la fonction réciproque est définie par $y = \cos(\varphi)$. Pour $y = 1$ on a $v(1) = \varphi = \int_1^1 = 0$ et donc $\cos(0) = 1$. Pour $y = 0$, on a $v(0) = \varphi = \int_0^1 = \frac{\pi}{2}$ et donc $\cos(\frac{\pi}{2}) = 0$. Contrairement à ce qui se passe pour le sin, le changement de t en $-t$ ne donne pas directement une symétrie du cos car la borne 1 devient -1 (alors que pour le sin la borne 0 reste 0). Par ailleurs, comme $x = \sin(\varphi)$, $y = \cos(\varphi)$ et $x^2 + y^2 = 1$ on a automatiquement

11. C'est un cas particulier simple de tout un ensemble d'intégrales dites eulériennes :

$$B(x, y) = \int_0^1 u^{x-1} (1-u)^{y-1} du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

où la fonction $\Gamma(z) = \int_0^\infty u^{z-1} e^{-u} du$ est la fonction Gamma étendant au domaine complexe la fonction factorielle $\Gamma(n) = (n-1)!$ (car $\Gamma(0) = 1$ et $\Gamma(z+1) = z\Gamma(z)$). Les valeurs de la fonction $\Gamma(z)$ sont reliées par la relation $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$. Dans le cas de l'intégrale $L(1)$, on fait le changement de variable $u = t^2$ et l'on obtient

$$\frac{1}{2} \int_0^1 u^{\frac{1}{2}} (1-u)^{\frac{1}{2}} du = \frac{1}{2} B\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{1}{2}\right) = \frac{1}{2} \frac{\pi}{\sin(\frac{\pi}{2})} = \frac{\pi}{2} .$$

12. Notons que la fonction $R(s) = \sqrt{1-s^2}$ est involutive : $R(R(s)) = s$.

$$\cos^2(\varphi) + \sin^2(\varphi) = 1 .$$

Il est facile de calculer les dérivées des fonctions circulaires de façon purement algébrique. En utilisant la règle du calcul différentiel disant que

$$\frac{d}{dx} \left(\int_a^x f(t) dt \right) = f(x) ,$$

on obtient

$$u'(x) = \frac{du}{dx} = \frac{d\varphi}{dx} = \frac{1}{\sqrt{1-x^2}} ,$$

$$\sin'(\varphi) = \frac{dx}{d\varphi} = \sqrt{1 - \sin^2(\varphi)} = \cos(\varphi) .$$

En tant que dérivée d'une fonction impaire, $\cos(\varphi)$ est une fonction paire.

De même,

$$v'(y) = \frac{dv}{dy} = \frac{d\varphi}{dy} = -\frac{1}{\sqrt{1-y^2}} ,$$

$$\cos'(\varphi) = \frac{dt}{d\varphi} = -\sqrt{1 - \cos^2(\varphi)} = -\sin(\varphi) .$$

On a également $\int_0^x + \int_x^1 = \int_0^1 = \frac{\pi}{2}$ et donc, avec $\varphi = \int_0^x$, $\psi = \int_x^1$, $\psi = \frac{\pi}{2} - \varphi$ et $x = \sin(\varphi) = \cos(\psi)$, on a

$$\begin{cases} \sin(\varphi) = \cos\left(\frac{\pi}{2} - \varphi\right) \\ \cos(\varphi) = \sin\left(\frac{\pi}{2} - \varphi\right) \end{cases}$$

ce qui est évident géométriquement.

5.2. Les formules d'addition

Une des propriétés fondamentales des fonctions circulaires concerne leurs formules d'*addition*. Étant donnés les angles φ , ψ , $\varphi + \psi \in [0, \frac{\pi}{2}]$, il s'agit de savoir s'il existe des formules générales donnant $\sin(\varphi + \psi)$ et $\cos(\varphi + \psi)$ en fonction de $\sin(\varphi)$, $\sin(\psi)$, $\cos(\varphi)$ et $\cos(\psi)$. Aujourd'hui, à l'école, on donne la formule

$$\sin(\varphi + \psi) = \sin(\varphi)\cos(\psi) + \sin(\psi)\cos(\varphi) = S(\varphi, \psi)$$

et l'on vérifie qu'elle marche. En effet, il est facile de constater que les dérivées partielles $\frac{\partial S}{\partial \varphi}$ et $\frac{\partial S}{\partial \psi}$ sont égales et égales toutes deux à $\cos(\varphi)\cos(\psi) - \sin(\varphi)\sin(\psi)$. Cela implique que la fonction à deux variables $S(\varphi, \psi)$ est en fait une fonction de la seule variable $\varphi + \psi$. Mais alors, $S(\theta) = S(\theta + 0) = S(\theta, 0) = \sin(\theta)$, d'où la formule d'addition. On montre de même que

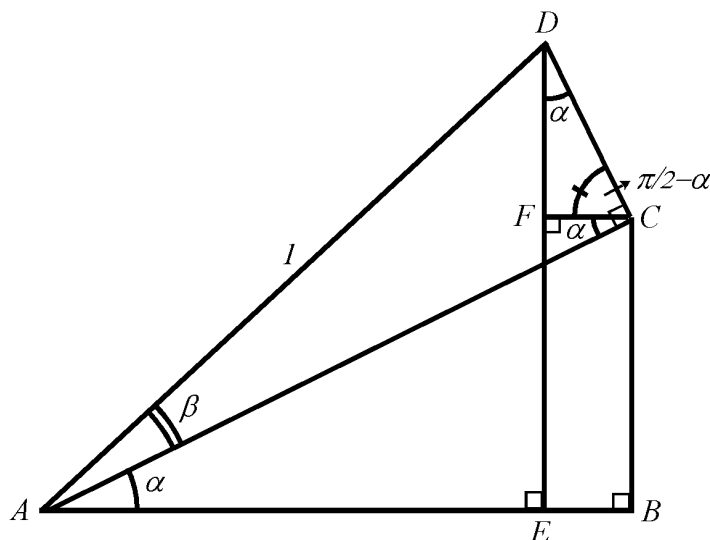


FIGURE 1. Preuve géométrique de la formule d'addition du sin. On construit un triangle rectangle ABC avec un angle $\widehat{CAB} = \varphi$. Sur son hypoténuse AC on construit un autre triangle rectangle ACB d'angle $\widehat{DAC} = \psi$ et d'hypoténuse $AD = 1$. Donc $AC = \cos(\psi)$ et $DC = \sin(\psi)$. Dans le triangle rectangle AED , on a $DE = \sin(\varphi + \psi)$ car l'hypoténuse AD est 1. On écrit alors $DE = DF + FE$. Mais $FE = CB = \sin(\varphi) AC = \sin(\varphi) \cos(\psi)$ et, comme $\widehat{FDC} = \widehat{FCA} = \widehat{CAB} = \varphi$, on a par ailleurs $\cos(\varphi) = \frac{DF}{DC} = \frac{DF}{\sin(\psi)}$, soit $DF = \sin(\psi) \cos(\varphi)$. $DE = DF + FE$ donne alors la formule d'addition.

$$\cos(\varphi + \psi) = \cos(\varphi) \cos(\psi) - \sin(\varphi) \sin(\psi)$$

(= $\sin'(\varphi + \psi)$ comme il se doit). En particulier $\sin(2\varphi) = 2 \sin(\varphi) \cos(\varphi)$ et $\cos(2\varphi) = \cos^2(\varphi) - \sin^2(\varphi)$. Ces formules ne sont pas évidentes géométriquement et il faut un peu d'astuce pour les démontrer. Une jolie présentation pédagogique est celle de la Khan Academy (fr.khanacademy.org). La construction de base est représentée à la figure 1 ainsi que les pas de la preuve.

Grâce à la formule magique de De Moivre (1707)

$$e^{i\varphi} = \cos(\varphi) + i \sin(\varphi)$$

exprimant la rotation d'angle φ dans le plan \mathbb{C} , les formules d'addition deviennent immédiates : il suffit en effet d'écrire $e^{i(\varphi+\psi)} = e^{i\varphi} e^{i\psi}$

$$\begin{aligned} & \cos(\varphi + \psi) + i \sin(\varphi + \psi) \\ &= (\cos(\varphi) + i \sin(\varphi)) (\cos(\psi) + i \sin(\psi)) \\ &= (\cos(\varphi) \cos(\psi) - \sin(\varphi) \sin(\psi)) + i (\sin(\varphi) \cos(\psi) + \sin(\psi) \cos(\varphi)) . \end{aligned}$$

En particulier $\cos(n\varphi) + i \sin(n\varphi) = (\cos(\varphi) + i \sin(\varphi))^n$ donne immédiatement les formules (non triviales) de n -plication.

Si l'on utilise la fonction réciproque Arcsin, la formule d'addition pour les sinus implique une relation remarquable entre intégrales. Elle s'écrit en effet (dans le premier quadrant)

$$\sin(\varphi + \psi) = \sin(\varphi) \sqrt{1 - \sin^2(\psi)} + \sin(\psi) \sqrt{1 - \sin^2(\varphi)}$$

et par conséquent, si l'on introduit $u = \sin(\varphi)$ et $v = \sin(\psi)$ et si l'on pose

$$r = u\sqrt{1 - v^2} + v\sqrt{1 - u^2} ,$$

l'identité $\text{Arcsin}(\varphi + \psi) = \varphi + \psi$ s'écrit pour $r = u\sqrt{1 - v^2} + v\sqrt{1 - u^2}$

$$\int_0^r \frac{dt}{\sqrt{1 - t^2}} = \int_0^u \frac{dt}{\sqrt{1 - t^2}} + \int_0^v \frac{dt}{\sqrt{1 - t^2}} .$$

On voit ainsi s'intriquer étroitement, jusqu'à s'unifier, des constructions géométriques, des calculs d'intégrales et des formules remarquables de nouvelles fonctions.

5.3. Extension à \mathbb{R}

Il est facile d'étendre les fonctions circulaires définies sur le premier quadrant à partir de l'hypothèse que les formules d'addition sont universelles. Si par exemple $\varphi \in [\frac{\pi}{2}, \pi]$, comme $\varphi - \frac{\pi}{2} \in [0, \frac{\pi}{2}]$, on écrit

$$\begin{aligned} \sin(\varphi) &= \sin\left(\left(\varphi - \frac{\pi}{2}\right) + \frac{\pi}{2}\right) = \sin\left(\varphi - \frac{\pi}{2}\right) \cos\left(\frac{\pi}{2}\right) + \sin\left(\frac{\pi}{2}\right) \cos\left(\varphi - \frac{\pi}{2}\right) \\ &= \cos\left(\varphi - \frac{\pi}{2}\right) = \cos\left(\frac{\pi}{2} - \varphi\right) \end{aligned}$$

car $\cos(\frac{\pi}{2}) = 0$, $\sin(\frac{\pi}{2}) = 1$, et \cos est paire. De même

$$\left\{ \begin{array}{l} \cos(\varphi) = -\sin\left(\varphi - \frac{\pi}{2}\right) = \sin\left(\frac{\pi}{2} - \varphi\right) , \\ \sin\left(\varphi + \frac{\pi}{2}\right) = \cos(\varphi) \\ \cos\left(\varphi + \frac{\pi}{2}\right) = -\sin(\varphi) \\ \sin(\varphi + \pi) = -\sin(\varphi) \\ \cos(\varphi + \pi) = -\cos(\varphi) \\ \sin(\varphi + 2\pi) = \sin(\varphi) \\ \cos(\varphi + 2\pi) = \cos(\varphi) \end{array} \right.$$

Toutes ces propriétés sont évidentes géométriquement à cause des symétries du cercle.

On constate ainsi que les fonctions circulaires définies sur \mathbb{R} à partir des intégrales circulaires sont des fonctions *périodiques* de période 2π . Ce point géométriquement (mais non pas analytiquement) évident est essentiel pour la compréhension des fonctions elliptiques.

5.4. Extension à \mathbb{C}

On étend ensuite les fonctions circulaires à \mathbb{C} par prolongement analytique, et d'abord à $i\mathbb{R}$. Pour ce faire, on effectue le changement de variable $t = -it'$, i.e. $t' = it$ et $t^2 = -t'^2$. On obtient

$$\begin{aligned}\varphi &= \int_0^x \frac{1}{\sqrt{1-t^2}} dt = - \int_0^{ix} \frac{i}{\sqrt{1+t'^2}} dt' \\ i\varphi &= \int_0^{ix} \frac{1}{\sqrt{1+t'^2}} dt' .\end{aligned}$$

Mais, de même que les intégrales circulaires $\int \frac{1}{\sqrt{1-t^2}} dt$ rectifient les arcs du cercle $x^2 + y^2 = 1$ et déterminent la trigonométrie classique, de même les intégrales $\int \frac{1}{\sqrt{1+t'^2}} dt'$ rectifient les arcs de l'hyperbole équilatère $x^2 - y^2 = 1$ et déterminent la trigonométrie hyperbolique. Par définition, si

$$w(s) = \int_0^s \frac{1}{\sqrt{1+t'^2}} dt' = h$$

est considérée comme une nouvelle fonction, la fonction réciproque $s = \sinh(h)$ s'appelle le *sinus hyperbolique* de h . Toute une trigonométrie hyperbolique peut alors être développée sur cette base. Dans ce contexte, $ix = i \sin(\varphi) = \sinh(i\varphi)$ et donc $\sin(\varphi) = -i \sinh(i\varphi)$ et donc $\sin(i\varphi) = i \sinh(\varphi)$. En fait, toutes les formules de la trigonométrie hyperbolique se déduisent des définitions classiques :

$$\begin{cases} \sinh(h) = \frac{1}{2}(e^h - e^{-h}) \\ \cosh(h) = \frac{1}{2}(e^h + e^{-h}) \end{cases}$$

qui sont absolument analogues à celles pour la trigonométrie normale que l'on peut dériver de la formule de De Moivre :

$$\begin{cases} \sin(\varphi) = \frac{1}{2i}(e^{i\varphi} - e^{-i\varphi}) \\ \cos(\varphi) = \frac{1}{2}(e^{i\varphi} + e^{-i\varphi}) \end{cases}$$

On vérifie par exemple que

$$\sin(i\varphi) = \frac{1}{2i}(e^{-\varphi} - e^{\varphi}) = -\frac{1}{i} \sinh(\varphi) = i \sinh(\varphi) .$$

De même,

$$\cos(i\varphi) = \frac{1}{2}(e^{-\varphi} + e^{\varphi}) = \cosh(\varphi) .$$

On obtient ainsi deux trigonométries, l'une circulaire et l'autre hyperbolique, qui s'échangent en passant de \mathbb{R} à $i\mathbb{R}$ et peuvent être toutes les deux prolongées au plan complexe tout entier en utilisant l'universalité des formules d'addition. Mais, le prolongement analytique au plan complexe ne fait pas apparaître de nouvelle période car les fonctions hyperboliques sont *sans période*. En effet, supposons que $\theta + i\alpha$, $\theta, \alpha \in \mathbb{R}$, soit une période du sinus. Il faudra alors que $\sin(\theta + i\alpha) = \sin(0) = 0$. Mais

$$\begin{aligned} \sin(\theta + i\alpha) &= \sin(\theta) \cos(i\alpha) + \sin(i\alpha) \cos(\theta) \\ &= \sin(\theta) \cosh(\alpha) + i \sinh(\alpha) \cos(\theta) . \end{aligned}$$

Il faut donc $\sin(\theta) \cosh(\alpha) = 0$ et $\sinh(\alpha) \cos(\theta) = 0$. Comme $\cosh(\alpha) \neq 0$, il faut $\sin(\theta) = 0$. Mais alors $\cos(\theta) = \pm 1$ et il faut donc $\sinh(\alpha) = 0$. Mais cela implique $\alpha = 0$ et $\theta = k\pi$. Il n'y a donc que la période réelle.

La grande découverte des fonctions elliptiques sera qu'elles possèdent *deux* périodes.

5.5. Fonctions analytiques et conditions de Cauchy-Riemann

Nous venons de considérer un prolongement des fonctions circulaires et des fonctions hyperboliques au plan complexe. Cela est une bonne occasion de rappeler ce qu'est une fonction de la variable complexe $f : \mathbb{C} \rightarrow \mathbb{C}$, $z \mapsto f(z)$ qui est dérivable en tant que fonction de z . A priori, si $z = x + iy$, f est une fonction des deux variables x, y et s'écrit

$$f(x, y) = X(x, y) + iY(x, y) .$$

Pour savoir si f est vraiment une fonction de z seul, il faut d'abord exprimer x et y en relation avec z . Cela est possible si l'on introduit le conjugué de z , $\bar{z} = x - iy$. En effet, on a alors

$$x = \frac{1}{2}(z + \bar{z}), \quad y = \frac{1}{2i}(z - \bar{z}) .$$

Dans les coordonnées (z, \bar{z}) de \mathbb{C} , f devient une fonction $f(z, \bar{z})$ et dire qu'elle est une fonction de z seul, c'est dire qu'elle ne dépend pas de \bar{z} . Si f est dérivable, cela s'exprime par $\frac{\partial f}{\partial \bar{z}} = 0$. Cette condition s'écrit

$$\frac{\partial X}{\partial \bar{z}} + i \frac{\partial Y}{\partial \bar{z}} = 0 .$$

Mais

$$\begin{cases} \frac{\partial X}{\partial \bar{z}} = \frac{\partial X}{\partial x} \frac{\partial x}{\partial \bar{z}} + \frac{\partial X}{\partial y} \frac{\partial y}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial X}{\partial x} + i \frac{\partial X}{\partial y} \right) \\ \frac{\partial Y}{\partial \bar{z}} = \frac{\partial Y}{\partial x} \frac{\partial x}{\partial \bar{z}} + \frac{\partial Y}{\partial y} \frac{\partial y}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial Y}{\partial x} + i \frac{\partial Y}{\partial y} \right) \end{cases}$$

Il faut donc

$$\left(\frac{\partial X}{\partial x} + i \frac{\partial X}{\partial y} \right) + i \left(\frac{\partial Y}{\partial x} + i \frac{\partial Y}{\partial y} \right) = 0 = \frac{\partial X}{\partial x} - \frac{\partial Y}{\partial y} + i \left(\frac{\partial X}{\partial y} + \frac{\partial Y}{\partial x} \right).$$

D'où les deux conditions

$$\frac{\partial X}{\partial x} = \frac{\partial Y}{\partial y}, \quad \frac{\partial X}{\partial y} = -\frac{\partial Y}{\partial x}.$$

Ces conditions, dites de Cauchy-Riemann, caractérisent les fonctions $f(z)$ dérivables par rapport à z , c'est-à-dire admettant une dérivée $f'(z)$ par rapport à z qui est bien définie. En effet on a

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = \left(\frac{\partial X}{\partial x} + i \frac{\partial Y}{\partial x} \right) dx + \left(\frac{\partial X}{\partial y} + i \frac{\partial Y}{\partial y} \right) dy.$$

Pour pouvoir écrire $df = f'(z) dz = f'(z) (dx + i dy)$, il faut avoir en même temps $f'(z) = \frac{\partial X}{\partial x} + i \frac{\partial Y}{\partial x}$ et $if'(z) = \frac{\partial X}{\partial y} + i \frac{\partial Y}{\partial y}$, c'est-à-dire les conditions de C-R.

Remarque. Nous avons déjà rencontré ces conditions à la section 4.6. du Vol I quand nous avons traité les pinwheels comme champs de phases. \square

Les fonctions $f(z)$ définies dans un domaine de \mathbb{C} qui sont dérivables en z sont appelées analytiques. Elles sont appelées méromorphes quand elles n'ont pour singularités que des pôles et holomorphes quand elles sont sans singularités. Elles possèdent des propriétés remarquables que Cauchy et Riemann ont commencé à étudier de façon systématique. Ce sont des applications dites *conformes* qui préservent les angles. Elles peuvent être très compliquées.

Il est facile de vérifier que les formules de prolongement des fonctions circulaires au plan complexe satisfont aux conditions de C-R. Reprenons par exemple

$$\begin{aligned} f(\theta, \alpha) &= \sin(\theta + i\alpha) = \sin(\theta) \cosh(\alpha) + i \sinh(\alpha) \cos(\theta) \\ &= F(\theta, \alpha) + iG(\theta, \alpha). \end{aligned}$$

On a

$$\begin{cases} \frac{\partial F}{\partial \theta} = \cos(\theta) \cosh(\alpha) \\ \frac{\partial F}{\partial \alpha} = \sin(\theta) \sinh(\alpha) \\ \frac{\partial G}{\partial \theta} = -\sin(\theta) \sinh(\alpha) \\ \frac{\partial G}{\partial \alpha} = \cos(\theta) \cosh(\alpha) \end{cases}$$

soit les conditions de C-R

$$\frac{\partial F}{\partial \theta} = \frac{\partial G}{\partial \alpha}, \quad \frac{\partial F}{\partial \alpha} = -\frac{\partial G}{\partial \theta}.$$

6. La lemniscate

6.1. La géométrie de la lemniscate

Avec l'avènement du calcul intégral-différentiel systématique à partir de Newton et Leibniz, on se posa de nombreux nouveaux problèmes, soit géométriques soit mécaniques (i.e. issus de la mise en équation de problèmes mécaniques) et beaucoup d'entre eux conduisirent à des obstructions analogues à celles rencontrées dans la rectification de l'ellipse. Des virtuoses se mirent à calculer, au moyen de changements de variables astucieux, nombre de relations entre des intégrales que l'on ne savait pas calculer en termes de fonctions rationnelles, logarithmiques, exponentielles ou trigonométriques. On se mit ainsi à emmagasiner une foule de données sur des identités remarquables reliant entre elles des intégrales de forme précise mais non exprimables explicitement. Ces relations et identités remarquables ressemblant à celles qui existent entre fonctions circulaires, on se mit à soupçonner l'existence d'une sorte de "trigonométrie elliptique" déformant la trigonométrie circulaire. Après une somme d'efforts prodigieux, ce mouvement historique a abouti à la théorie des fonctions elliptiques dans laquelle nos "géodésiques legendriennes" modélisant les contours illusoire à la Kanizsa se trouvent naturellement immergées.

Dans les sections suivantes nous discuterons en particulier le problème mécanique du pendule et le problème géométrique des *elasticæ*. Mais nous allons d'abord dire quelques mots de la trigonométrie "lemniscatique".

Connue depuis l'Antiquité et redécouverte en particulier par Jacques¹³ et Jean Bernouilli, la lemniscate \mathfrak{L} (dite de Bernouilli) est une courbe fascinante. On l'obtient géométriquement en effectuant sur l'hyperbole équilatère \mathfrak{H} d'équation $x'^2 - y'^2 = 1$ l'inversion géométrique de pôle 0 et de cercle d'inversion le cercle unité (cf. plus bas la figure 2). Si l'on identifie \mathbb{R}^2 à \mathbb{C} à travers $z = x + iy$, cette inversion est l'application $z \rightarrow \frac{1}{\bar{z}}$. En coordonnées polaires, l'équation de \mathfrak{L} est

$$\rho^2 = \cos(2\theta) = \cos^2(\theta) - \sin^2(\theta) ,$$

avec $\theta \in [0, \frac{\pi}{4}] \cup [\frac{3\pi}{4}, \pi] \cup [\pi, \frac{5\pi}{4}] \cup [\frac{7\pi}{4}, 0]$. En effet si un rayon vecteur d'origine O et d'angle θ (par exemple dans le premier quadrant) coupe \mathfrak{L} en $P = (x = \rho \cos(\theta), y = \rho \sin(\theta))$ et \mathfrak{H} en $P' = (x', y')$, on doit avoir $OP \cdot OP' = 1$ et donc $x' = \frac{\cos(\theta)}{\rho}$ et $y' = \frac{\sin(\theta)}{\rho}$. L'équation $x'^2 - y'^2 = 1$ équivaut par conséquent à

$$\frac{\cos^2(\theta) - \sin^2(\theta)}{\rho^2} = 1$$

qui est bien l'équation de \mathfrak{L} . Lorsque P' s'éloigne asymptotiquement vers l'_{∞} sur \mathfrak{H} , P doit tendre vers 0.

13. En relation avec les *elasticæ*, cf. plus bas.

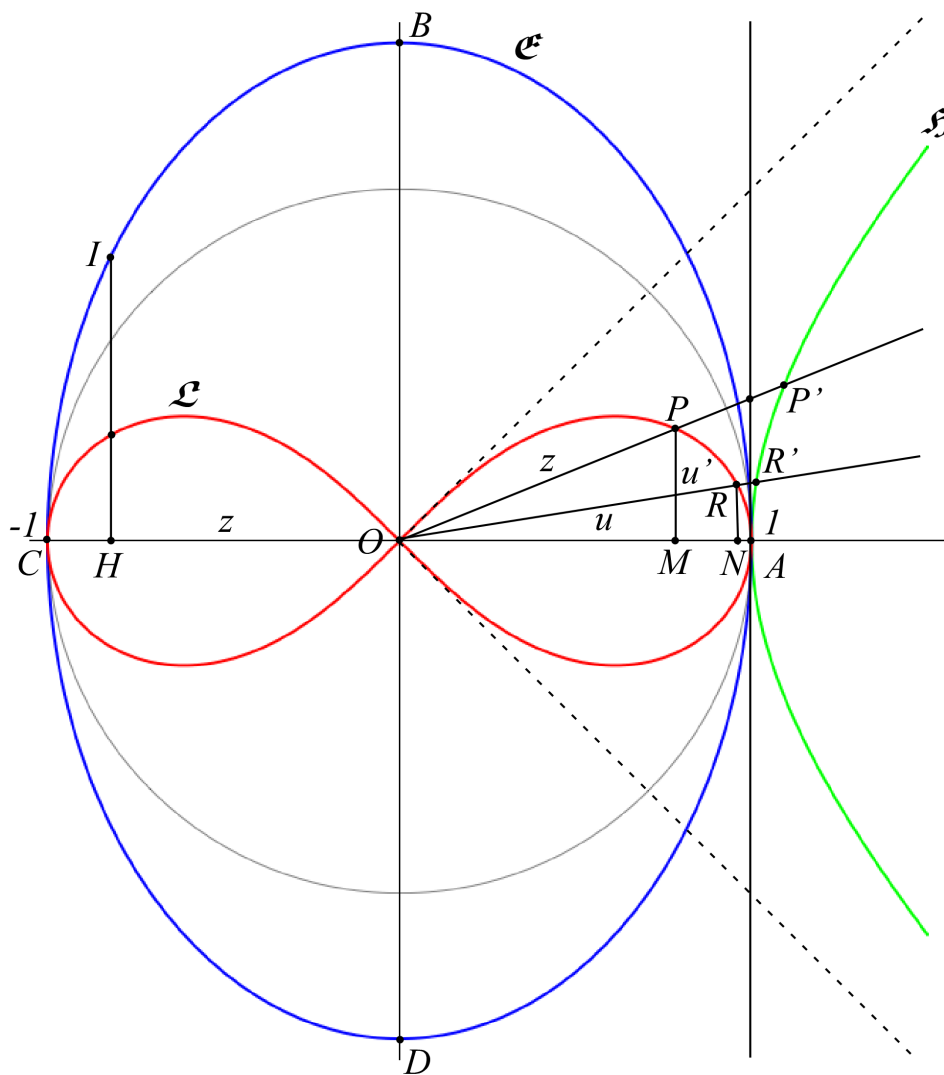


FIGURE 2. La lemniscate unité \mathcal{L} de Bernouilli (en rouge). L'hyperbole équilatère \mathcal{H} d'équation $x'^2 - y'^2 = 1$ (en vert) avec ses asymptotes en pointillés. L'ellipse \mathcal{E} de demi-axes 1 et $\sqrt{2}$ (en bleu). On a représenté les cordes $z = OP$, $u = OR$, $u' = OR'$. Pour les autres notations et pour les arcs, voir le texte.

L'équation en coordonnées cartésiennes $x = \rho \cos(\theta)$, $y = \rho \sin(\theta)$ est évidente, puisqu'en multipliant l'équation par ρ^2 , on obtient

$$\rho^4 = (x^2 + y^2)^2 = \rho^2 (\cos^2(\theta) - \sin^2(\theta)) = x^2 - y^2$$

qui, notons-le, est une équation algébrique. Les bornes de \mathfrak{L} sur l'axe des x sont ± 1 . On peut par homothétie les prendre égales à $\pm a$, ce qui introduit un paramètre d'échelle.

Il est trivial de vérifier que \mathfrak{L} est le lieu des points $P = (x, y)$ dont le produit $\rho_1 \rho_2$ des distances aux foyers $(\pm \frac{1}{\sqrt{2}})$ est égal à $\frac{1}{2}$. En effet, comme

$$\rho_1^2 = \left(x - \frac{1}{\sqrt{2}}\right)^2 + y^2, \quad \rho_2^2 = \left(x + \frac{1}{\sqrt{2}}\right)^2 + y^2,$$

on doit avoir

$$\begin{aligned} \rho_1^2 \rho_2^2 &= \frac{1}{4} = \left(\left(x - \frac{1}{\sqrt{2}}\right)^2 + y^2\right) \left(\left(x + \frac{1}{\sqrt{2}}\right)^2 + y^2\right) \\ &= (x^2 + y^2)^2 - (x^2 - y^2) + \frac{1}{4}. \end{aligned}$$

Posons $t = \cos(2\theta) = \rho^2 \in [0, 1]$ et donc $\rho = \sqrt{t}$. On a $\cos^2(\theta) = \frac{1+\cos(2\theta)}{2}$ et donc, dans le premier quadrant $\theta \in [0, \frac{\pi}{4}]$, $\cos(\theta) = \sqrt{\frac{1+t}{2}}$ et $\sin(\theta) = \sqrt{\frac{1-t}{2}}$, d'où la paramétrisation de \mathfrak{L} :

$$x = \rho \cos(\theta) = \sqrt{\frac{t+t^2}{2}}; \quad y = \rho \sin(\theta) = \sqrt{\frac{t-t^2}{2}}.$$

On en tire l'élément de longueur d'arc ds et la longueur d'arc $s(u)$ correspondant à un angle θ_0 :

$$\begin{aligned} ds^2 &= dx^2 + dy^2 = \left[\frac{1}{\sqrt{2}} \frac{1}{2} \left(\frac{1+2t}{\sqrt{t+t^2}}\right)\right]^2 dt^2 + \left[\frac{1}{\sqrt{2}} \frac{1}{2} \left(\frac{1-2t}{\sqrt{t-t^2}}\right)\right]^2 dt^2 \\ ds &= \frac{1}{2} \frac{dt}{\sqrt{t-t^3}} \\ s(u) &= \frac{1}{2} \int_0^u \frac{dt}{\sqrt{t-t^3}} \\ u &= \cos(2\theta_0). \end{aligned}$$

En coordonnées polaires, on déduit de $\rho^2 = \cos(2\theta)$ la différentielle

$$2\rho d\rho = -2 \sin(2\theta) d\theta$$

et donc, puisque $\sin(2\theta) = \sqrt{1 - \cos^2(2\theta)} = \sqrt{1 - \rho^4}$,

$$d\theta = -\frac{\rho d\rho}{\sqrt{1 - \rho^4}},$$

soit

$$ds^2 = d\rho^2 + \rho^2 d\theta^2 = \left(1 + \frac{\rho^4}{1 - \rho^4}\right) d\rho^2 = \frac{1}{1 - \rho^4} d\rho^2$$

$$ds = \frac{d\rho}{\sqrt{1 - \rho^4}}.$$

On notera que

$$\frac{d\theta}{\sqrt{\cos(2\theta)}} = \frac{d\theta}{\rho} = -\frac{d\rho}{\sqrt{1 - \rho^4}} = \frac{1}{2} \frac{dt}{\sqrt{t - t^3}}.$$

Pour les autres quadrants, on modifie les formules en utilisant les symétries de \mathcal{L} .

6.2. Le Comte de Fagnano et Euler

Des découvertes remarquables sur la lemniscate furent faites par le Comte de Fagnano dans deux mémoires de 1718, *Metodo per misurare la lemniscata* (cf. [178] et [505]), travaux qu'il communiqua plus tard à Euler comme pièces pour une candidature à l'Académie de Berlin. On sait qu'Euler examina ces documents le 23 décembre 1751 et en fut très intéressé. Selon Jacobi, cette date est la date de naissance de la théorie des fonctions elliptiques. Le comte de Fagnano avait découvert une construction géométrique, "avec la règle et le compas" comme on disait à l'époque, permettant de diviser en deux un arc de lemniscate, autrement dit, l'équivalent pour la lemniscate de la division d'un arc de cercle au moyen des formules d'addition des fonctions circulaires.

Rappelons la formule d'addition des sinus en termes d'Arcsin :

$$\int_0^r \frac{dt}{\sqrt{1-t^2}} = \int_0^u \frac{dt}{\sqrt{1-t^2}} + \int_0^v \frac{dt}{\sqrt{1-t^2}}$$

pour $r = u\sqrt{1-v^2} + v\sqrt{1-u^2}$ Cela donne pour $u = v$ (division des angles)

$$\int_0^{r(u)} \frac{dt}{\sqrt{1-t^2}} = 2 \int_0^u \frac{dt}{\sqrt{1-t^2}}$$

pour $r(u) = u\sqrt{1-u^2}$.

Fagnano trouva que si l'on utilise le changement de variable involutif

$$t \mapsto s = r(t) = \frac{1-t}{1+t},$$

et donc

$$dt = -\frac{1}{1+s} - \frac{1-s}{(1+s)^2} = -\frac{2}{(1+s)^2} ds,$$

on obtient

$$s(u) = \frac{1}{2} \int_0^u \frac{dt}{\sqrt{t-t^3}} = -\frac{1}{2} \int_1^{r(u)} \frac{ds}{\sqrt{s-s^3}}$$

et donc, si $u = r(u)$ (i.e. $u = \sqrt{2} - 1$), deux arcs de même longueur.¹⁴

Pour les calculs à partir de $\frac{d\rho}{\sqrt{1-\rho^4}}$ et non pas de $\frac{1}{2} \frac{dt}{\sqrt{t-t^3}}$, le changement de variable involutif est $z \mapsto u(z) = \frac{\sqrt{1-z^2}}{\sqrt{1+z^2}}$ et

$$\int_{O=0}^{P=z} \frac{dr}{\sqrt{1-r^4}} = \int_{R=u}^{A=1} \frac{dr}{\sqrt{1-r^4}}$$

autrement dit, si les cordes $OP = z$ et $OR = u$ de la figure 2 ci-dessus sont reliées par cette relation, alors les arcs \widetilde{OP} et \widetilde{RA} sont égaux. Il suffit par conséquent de faire $P = R$ i.e. $u = z$ pour diviser en deux le premier quadrant de \mathfrak{L} , ce qui est, remarquons le, beaucoup plus simple que de diviser un arc quelconque. L'équation donnant z est l'équation bicarrée $z^4 + 2z^2 - 1$, qui a pour solution positive, comme il se doit, $z = \sqrt{\sqrt{2} - 1}$. Nous allons voir que le changement de variable $u(z)$ est en fait la relation entre les sinus et cosinus lemniscatiques.

Il est intéressant de remarquer que le lien entre les arcs \widetilde{OP} et \widetilde{RA} provient de la possibilité d'utiliser ρ ou θ comme variable de base. Si l'on utilise ρ , alors $\int_0^{\rho_P}$ correspond à \widetilde{OP} . Si l'on utilise θ , alors $\int_0^{\theta_R}$ correspond à \widetilde{AR} . L'égalité $\widetilde{OP} = \widetilde{AR}$ est donc une égalité entre $\int_0^{\rho_P}$ et $\int_0^{\theta_R}$ avec $\rho_P = \sqrt{t_P}$ et $t_R = \cos(2\theta_R)$. Pour que l'égalité soit effective, il faut une relation entre t_P et t_R et celle-ci est

$$t_P + t_R + t_P t_R = 1 \tag{F}$$

Le changement de variable de Fagnano $u = \frac{\sqrt{1-z^2}}{\sqrt{1+z^2}}$ s'écrit $u^2 = \frac{1-z^2}{1+z^2}$, i.e.

$$z^2 + u^2 + z^2 u^2 = 1,$$

autrement dit comme la relation ci-dessus avec $t_P = z^2$ et $t_R = u^2$, ce qui est normal puisque $t = \rho^2$.

Fagnano trouva également (cf. encore la figure 2) des relations remarquables entre des arcs de différentes courbes sous la condition de relations algébriques bien définies

14. Dans les calculs qui suivent, nous rencontrerons quelques problèmes de notation. D'abord il est naturel d'utiliser des notations traditionnelles comme par exemple ρ en coordonnées polaires. Par ailleurs beaucoup d'intégrales sont de la forme $\int_{a(w_0)}^{b(w_0)} f(w) dw$, le w_0 qui intervient dans les bornes fonctionnant lui-même comme une variable. D'où l'intérêt de distinguer pour une même variable son rôle de variable courante dans l'intégrande et son rôle dans les limites, par exemple t et u dans $\int_0^u \frac{dt}{\sqrt{t-t^3}}$ (l'écriture habituelle $\int_0^u \frac{du}{\sqrt{u-u^3}}$ prête à confusion). Enfin, par exemple dans les formules d'addition, il est pertinent de traiter plusieurs cordes ρ_1, ρ_2 , etc. comme des variables z, u , etc. intervenant dans des relations algébriques. Nous avons essayé de satisfaire au mieux ces contraintes.

entre les bornes des intégrales mesurant ces arcs. Considérons par exemple dans la figure 2 la lemniscate \mathfrak{L} et l'ellipse \mathfrak{E} de demi-axes $OA = (0, 1)$ et $OB = (0, \sqrt{2})$. Considérons un P sur \mathfrak{L} (dans le premier quadrant) et notons z la corde OP . Nous savons que la longueur de l'arc $\widetilde{OP}_{\mathfrak{L}}$ de \mathfrak{L} est donnée par l'intégrale

$$\int_0^z \frac{d\rho}{\sqrt{1-\rho^4}}.$$

Quant à l'arc inverse $\widetilde{PA}_{\mathfrak{L}}$ il est donné par

$$\int_z^1 \frac{d\rho}{\sqrt{1-\rho^4}}.$$

Reportons alors z sur le segment OH et considérons l'arc $\widetilde{BI}_{\mathfrak{E}}$ de l'ellipse \mathfrak{E} au-dessus de OH . Nous savons que sa longueur est donnée par l'intégrale

$$\int_0^z \frac{\sqrt{1+t^2}}{\sqrt{1-t^2}} dt$$

et que l'arc réciproque $\widetilde{IC}_{\mathfrak{E}}$ est donné par l'intégrale $\int_z^1 \frac{\sqrt{1+t^2}}{\sqrt{1-t^2}} dt$.¹⁵ Enfin, dans une hyperbole équilatère comme \mathfrak{H} considérons un point R' et notons u' le rayon OR' . La longueur de l'arc $\widetilde{AR}'_{\mathfrak{H}}$ est donnée par l'intégrale $\int_1^u \frac{v^2 dv}{\sqrt{v^4-1}}$. Le théorème de Fagnano dit que si les cordes z et u' entretiennent une certaine relation algébrique, et plus précisément, si

$$u' = \frac{\sqrt{1+z^2}}{\sqrt{1-z^2}}$$

(l'inverse du changement de variable $u(z) = \sqrt{\frac{1-z^2}{1+z^2}}$ ci-dessus), alors on a la relation

$$\begin{aligned} \widetilde{OP}_{\mathfrak{L}} &= \widetilde{BI}_{\mathfrak{E}} + \widetilde{AR}'_{\mathfrak{H}} - zu' \\ \int_0^z \frac{d\rho}{\sqrt{1-\rho^4}} &= \int_0^z \frac{\sqrt{1+t^2}}{\sqrt{1-t^2}} dt + \int_1^{u'} \frac{v^2 dv}{\sqrt{v^4-1}} - zu'. \end{aligned}$$

On remarquera que si $z = 0$, alors $u' = 1$ et tous les termes s'annulent. Si en revanche $z = 1$ alors des divergences apparaissent. Elles correspondent aux asymptotes de \mathfrak{H} .

Avec Fagnano, c'est ainsi tout un univers de changements de variables possédant une signification géométrique et analytique profonde qui s'est ouvert et qui, à travers Euler, puis Legendre, Gauss, Abel, Jacobi et bien d'autres, conduira à la théorie des fonctions elliptiques.

15. Ce n'est pas l'intégrale $\int_0^z \sqrt{\frac{1-k^2t^2}{1-t^2}} dt$ que nous avons rencontrée plus haut avec $a \geq b$ et l'excentricité $k^2 = \frac{a^2-b^2}{b^2} \geq 0$ car ici $b = \sqrt{2} > a = 1$ et l'intégrale est $\int_0^z \sqrt{\frac{1+k^2t^2}{1-t^2}} dt$ avec $k^2 = \frac{b^2-a^2}{a^2} \geq 0$ ($k^2 = 1$).

Entre autres, Fagnano introduisit des changements de variable transformant la différentielle $\frac{dz}{\sqrt{1-z^4}}$ en $n\frac{du}{\sqrt{1-u^4}}$. Par exemple si

$$\frac{z\sqrt{2}}{\sqrt{1-z^4}} = \frac{1}{u}\sqrt{1-\sqrt{1-u^4}},$$

relation déjà assez compliquée mais constructible “par la règle et le compas”, on a

$$\frac{du}{\sqrt{1-u^4}} = 2\frac{dz}{\sqrt{1-z^4}}$$

et donc l'arc \widetilde{OR} correspondant à la corde u est le double de l'arc \widetilde{OP} correspondant à la corde z . On obtient donc la formule pour la division d'un arc \widetilde{OP} quelconque de \mathfrak{L} (et pas seulement de \widetilde{OA} comme ci-dessus).

En termes des intégrales correspondant à θ et non plus à ρ , on trouve que le changement de variable $t = \frac{2s}{1+s^2}$ transforme la différentielle $\frac{dt}{\sqrt{t-t^3}}$ en $\frac{\sqrt{2}ds}{\sqrt{s+s^3}}$ et que le second changement de variable $s = \frac{2v}{1-v^2}$ transforme la différentielle $\frac{ds}{\sqrt{s+s^3}}$ en $\frac{\sqrt{2}dv}{\sqrt{v-v^3}}$. En enchaînant les deux, on obtient la duplication

$$\frac{dt}{\sqrt{t-t^3}} = \frac{2dv}{\sqrt{v-v^3}}.$$

Le changement est donc

$$t = 4v\frac{1-v^2}{1+v^2}.$$

Des formules de ce genre se vérifient facilement en différenciant.¹⁶ Mais il n'en reste pas moins que les trouver ne va pas de soi et que l'on ne sait pas très bien comment Fagnano les a découvertes.

De même, si

$$\frac{1}{u\sqrt{2}}\sqrt{1-u^4} = \frac{1}{z}\sqrt{1-\sqrt{1-z^4}},$$

on a $\frac{dz}{\sqrt{1-z^4}} = -2\frac{du}{\sqrt{1-u^4}}$ et donc l'arc \widetilde{OP} correspondant à la corde z est le double de l'arc inverse \widetilde{RA} . Si donc $P = R$, i.e. $z = u$, on obtient une trisection du quadrant \widetilde{OA} . L'équation donnant z est alors l'équation en z^4 : $z^8 + 6z^4 - 3 = 0$ donnant $z = \sqrt[4]{2\sqrt{3}-3}$.

On peut implémenter ainsi des formules de division des arcs de \mathfrak{L} au moyen de relations *algébriques* entre les cordes.

En méditant et calculant sur le résultat de Fagnano, Euler commença à se convaincre qu'il devait exister une trigonométrie “lemniscatique” qui serait aux

16. Par exemple, de $t = \frac{2s}{1+s^2}$ on tire $dt = 2\frac{1-s^2}{(1+s^2)^2}ds$. Mais on en tire aussi $1-t^2 = \frac{(1-s^2)^2}{(1+s^2)^2}$ et donc $t-t^3 = \frac{2s}{1+s^2} \cdot \frac{(1-s^2)^2}{(1+s^2)^2}$ et donc $\frac{dt}{\sqrt{t-t^3}} = 2\frac{1-s^2}{(1+s^2)^2} \frac{\sqrt{1+s^2}}{\sqrt{2s}} \frac{1+s^2}{1-s^2} ds$.

intégrales du type $\frac{1}{2} \int \frac{dt}{\sqrt{t-t^3}}$ et $\int \frac{d\rho}{\sqrt{1-\rho^4}}$ ce qu'est la trigonométrie classique pour les intégrales de type $\int \frac{dt}{\sqrt{1-t^2}}$. Il publia deux articles à ce sujet en 1756/1757. Tout de suite, avec son incomparable génie, il généralisa, systématisa et chercha les raisons "occultes" sous-jacentes à la réussite des constructions géométriques et des relations algébriques apparaissant dans les calculs d'intégrales de Fagnano. Un bon résumé de cette histoire étonnante est l'article d'Ivahn Smadja [505].¹⁷ Par exemple, Euler chercha au moyen de quels changements de variable on pouvait obtenir des égalités de différentielles

$$m \frac{dx}{\sqrt{1-x^4}} = n \frac{dy}{\sqrt{1-y^4}}$$

généralisant les formules de duplication. Pour la formule d'addition

$$\widetilde{OP} = \widetilde{OM} + \widetilde{ON}$$

entre trois arcs de \mathfrak{L} correspondant à trois points successifs M, N, P de cordes respectives x, y, z , il trouva la relation

$$z = \frac{x\sqrt{1-y^4} + y\sqrt{1-x^4}}{1+x^2y^2}.$$

On peut vérifier que cette formule peut s'écrire

$$z = \frac{x\sqrt{\frac{1-y^2}{1+y^2}} + y\sqrt{\frac{1-x^2}{1+x^2}}}{1 - xy\sqrt{\frac{1-x^2}{1+x^2}}\sqrt{\frac{1-y^2}{1+y^2}}} \quad (\text{E})$$

qui, nous allons le voir, est la formule d'addition du sinus "lemniscatique".¹⁸

6.3. Une trigonométrie "lemniscatique" (Gauss, Abel)

En 1796, le jeune Gauss (19 ans) reprit les calculs de Fagnano et d'Euler et fit à son tour de profonds calculs sur d'éventuels sinus et cosinus "lemniscatiques" sl et cl . Il ne les publia pas mais on les retrouva dans ses papiers. Pour des détails, on pourra se référer au site [405] de Jean-Claude Pénin. Abel développera des calculs analogues dans un contexte plus large (cf. par exemple [1]).

Pour les intégrales rectifiant la lemniscate (dans le premier quadrant)

17. Cf. la bibliographie détaillée de [505], les *Mathematische Werke* d'Euler [175], ainsi que les *Archives Euler* à l'URL eulerarchive.maa.org.

18. Pour vérifier l'égalité des deux expressions de $z = \frac{A}{B} = \frac{C}{D}$, on vérifie que $AD = BC$. Les termes comme $x^3y^2\sqrt{\frac{1-y^2}{1+y^2}}$ s'éliminent et les termes en $\sqrt{1-y^2}$ ont un facteur $\frac{x}{\sqrt{1+y^2}}$ à gauche et un facteur $x\sqrt{1+y^2} - \frac{xy^2}{\sqrt{1+y^2}}$ à droite. Ces deux facteurs sont égaux. Il en va de même pour les termes obtenus en échangeant x et y .

$$F(\rho) = \int_0^\rho \frac{dr}{\sqrt{1-r^4}} = \varphi,$$

Gauss introduisit la *fonction réciproque* $sl(\varphi)$ (i.e. $sl(\int_0^\rho) = \rho$ et $F(\rho) = \int_0^{sl(\varphi)}$). $F(\rho)$ est une fonction impaire. En effet si $r' = -r$, $F(\rho) = \int_0^{-\rho} \frac{-dr'}{\sqrt{1-r'^4}} = -F(-\rho)$. La fonction réciproque $sl(\varphi)$ est donc également impaire. Par définition $sl(0) = 0$. Soit $\frac{1}{2}K = F(1)$ la longueur d'un quart de \mathfrak{L} . On a $sl\left(\frac{K}{2}\right) = 1$. Posons

$$cl(\theta) = sl\left(\frac{K}{2} - \theta\right),$$

i.e. $\theta = \int_{cl(\theta)}^1 \frac{dr}{\sqrt{1-r^4}}$. On a

$$cl\left(\frac{K}{2}\right) = sl(0) = 0, \quad cl(0) = sl\left(\frac{K}{2}\right) = 1.$$

En utilisant les symétries de \mathfrak{L} pour prolonger sl et cl aux autres quadrants, on obtient

$$sl(nK) = 0, \quad sl\left(\left(n + \frac{1}{2}\right)K\right) = \pm 1,$$

+ si n est pair, - si n est impair. La période de sl et cl est donc de $2K$. Traditionnellement, on écrit $K = \pi G$ où

$$G = \frac{2}{\pi} \int_0^1 \frac{dr}{\sqrt{1-r^4}}$$

est la constante de Gauss. C'est un nombre transcendant que l'on sait calculer au-delà de 15 milliards de décimales.

En termes de ces fonctions "lemniscatiques", la formule F de Fagnano s'écrit

$$sl^2 + cl^2 + sl^2 cl^2 = 1,$$

soit $cl^2 = \frac{1-sl^2}{1+sl^2}$ et dualement $sl^2 = \frac{1-cl^2}{1+cl^2}$, $cl^2 = \frac{1-sl^2}{1+sl^2}$ étant le changement de variable $u^2 = \frac{1-z^2}{1+z^2}$ découvert par Fagnano. Gauss démontra alors les formules d'addition

$$\begin{aligned} sl(\varphi \pm \theta) &= \frac{sl(\varphi) cl(\theta) \pm sl(\theta) cl(\varphi)}{1 \mp sl(\varphi) sl(\theta) cl(\varphi) cl(\theta)} \\ cl(\varphi \pm \theta) &= \frac{cl(\varphi) cl(\theta) \mp sl(\theta) sl(\varphi)}{1 \pm sl(\varphi) sl(\theta) cl(\varphi) cl(\theta)} \end{aligned}$$

d'où

$$\begin{aligned} sl(2\varphi) &= \frac{2sl(\varphi)cl(\varphi)}{1 - sl^2(\varphi)cl^2(\varphi)} = 2sl(\varphi)cl(\varphi) \frac{1 + sl^2(\varphi)}{1 + sl^4(\varphi)} \\ &= 2sl(\varphi)cl(\varphi) \frac{1 + cl^2(\varphi)}{1 + cl^4(\varphi)}. \end{aligned}$$

La formule d'addition des sl est la formule (E) d'Euler. En effet si $\widetilde{OP} = \widetilde{OM} + \widetilde{ON}$ avec $\widetilde{OM} = \int_0^x \frac{dr}{\sqrt{1-r^4}} = \varphi$, $\widetilde{ON} = \int_0^y \frac{dr}{\sqrt{1-r^4}} = \theta$ et $\widetilde{OP} = \int_0^z \frac{dr}{\sqrt{1-r^4}} = \varphi + \theta$, on a $x = sl(\varphi)$, $y = sl(\theta)$, $z = sl(\varphi + \theta)$, $cl = \sqrt{\frac{1-sl^2}{1+sl^2}}$ et la formule de Gauss devient la formule d'Euler.

Il semble que Gauss ait dès cette époque compris que l'on pouvait prolonger sl et cl au plan complexe et qu'il émergeait alors, en plus de la période réelle $2K$, une *nouvelle* période, imaginaire pure, $2iK$. Pour ce faire, Gauss changea φ en $i\varphi$ et trouva $sl(i\varphi) = i(sl(\varphi))$ et $cl(i\varphi) = \frac{1}{cl(\varphi)}$.

Pour des prolongements récents de la trigonométrie lemniscatique on pourra consulter par exemple l'étude [387] d'Edward Neumann.

7. Le pendule et les fonctions elliptiques

7.1. Les équations du mouvement

Un autre exemple qui joua un rôle décisif dans l'élaboration de la théorie des fonctions elliptiques fut non pas un exemple de rectification de courbe géométrique, mais un exemple de courbe mécanique, simple et naturel, "produit par la nature sans aucun artifice" comme le disait Jacques Bernoulli, celui du mouvement du pendule. Autant le mouvement est facile à décrire pour les "petites" oscillations pour lesquelles l'approximation au premier ordre $\sin(\theta) \sim \theta$ est valable, autant la description des grandes oscillations est compliquée. Nous en esquissons ici la théorie car nous en rencontrerons des avatars purement géométriques dans nos modèles de contours illusoire. La figure 3 présentée plus bas décrit la situation.

Soient m la masse du pendule, O son centre de rotation et l sa longueur. Le mouvement s'effectue le long du cercle C de centre O et de rayon l . Le pendule est soumis à l'action de la pesanteur dont l'accélération est g . Soit θ l'angle du pendule P par rapport au demi-axe OA vertical descendant. L'énergie, qui est une constante du mouvement en l'absence de friction, est la somme de l'énergie cinétique $\frac{1}{2}m\left(\frac{d\theta}{dt}\right)^2$ et de l'énergie potentielle $mgl(1 - \cos(\theta))$ qui est le travail de mg le long de la différence de hauteur HA . Pour simplifier, on pose $m = g = l = 1$ et, pour des raisons explicitées plus bas,

$$E = 2k^2 = \frac{1}{2}\dot{\theta}^2 + 1 - \cos(\theta) .$$

Si on la dérive, on obtient la célèbre équation $\ddot{\theta} = -\sin(\theta)$ qui est une équation de Newton. Dans le régime des petites oscillations, $\sin(\theta) \sim \theta$ et l'équation devient celle de l'oscillateur harmonique $\ddot{\theta} = -\theta$.

Comme $1 - \cos(\theta) = 2 \sin^2\left(\frac{\theta}{2}\right)$, l'équation s'écrit

$$\dot{\theta}^2 = 4 \left(k^2 - \sin^2\left(\frac{\theta}{2}\right) \right) .$$

Le fait que ce soit une équation en $\dot{\theta}^2$ et donc avec des solutions en $\dot{\theta} = \pm\sqrt{\quad}$ correspond à la symétrie $\theta \rightarrow -\theta$ de la situation. Étant donnée cette symétrie, on peut supposer $\theta \in [0, \pi]$. On remarquera par ailleurs le fait que l'équation donne $\dot{\theta}(t)$ comme fonction de θ et non pas de t et que donc, il sera pertinent de considérer à un certain moment la fonction réciproque $t(\theta)$ plutôt que la fonction directe $\theta(t)$.

Pour $k = 0$, $\cos(\theta) = 1$, $\dot{\theta} = 0$ et le pendule est au repos dans sa position basse A (équilibre stable). Pour $k = 1$, l'angle maximal est donné par $\cos(\theta_0) = -1$, $\dot{\theta}_0 = 0$ et le pendule a juste l'énergie lui permettant de rejoindre sa position haute B (équilibre instable). Pour $k < 1$ le régime du pendule est oscillatoire avec une vitesse de passage en A de $2k$ et une hauteur maximale où $\dot{\theta}_0 = 0$ et $\sin\left(\frac{\theta_0}{2}\right) = k$. Pour $k > 1$, le pendule accomplit des tours complets avec une vitesse de passage en A de $2k$ et une vitesse de passage en B de $2\sqrt{k^2 - 1}$.

Le pendule est géométriquement trivial puisque que son espace de configuration est le cercle C . Géométriquement parlant, sa trigonométrie est donc simplement la trigonométrie circulaire des angles. C'est la *loi temporelle* $t(\theta)$ sur C , autrement dit la durée du mouvement effectif, qui est hautement non triviale. Le pendule est un "convertisseur" mécanique d'espace en temps qui convertit des arcs de cercles en intervalles temporels. D'où l'idée de convertir la trigonométrie "circulaire" des arcs en une trigonométrie en quelque sorte "chronométrique" des intervalles de temps.¹⁹ Esquissons les étapes de cette conversion.

Considérons le régime oscillatoire ($k < 1$) et soit P_0 d'angle θ_0 la position de hauteur maximale et (P, θ) la position courante. Il est habituel de *normaliser* la situation en ramenant le cas $\theta_0 < \pi$ au cas limite $\theta_0 = \pi$. Pour cela (voir figure 3) on considère le point H_0 de même hauteur que P_0 sur l'axe vertical AB , on construit le cercle C_0 de diamètre $AH_0 = 2k^2$, donc de rayon k^2 , on considère le point d'intersection Q de ce C_0 avec l'horizontale HP . On a

$$HA = 1 - \cos(\theta) = 2 \sin^2\left(\frac{\theta}{2}\right) .$$

19. Nous introduisons ce terme non conventionnel dans cette section parce que le terme conventionnel d'"elliptique" vient d'une antonomase (la rectification d'un arc d'ellipse est un représentant typique d'un type d'intégrales) et n'a lexicalement rien à voir avec le pendule.

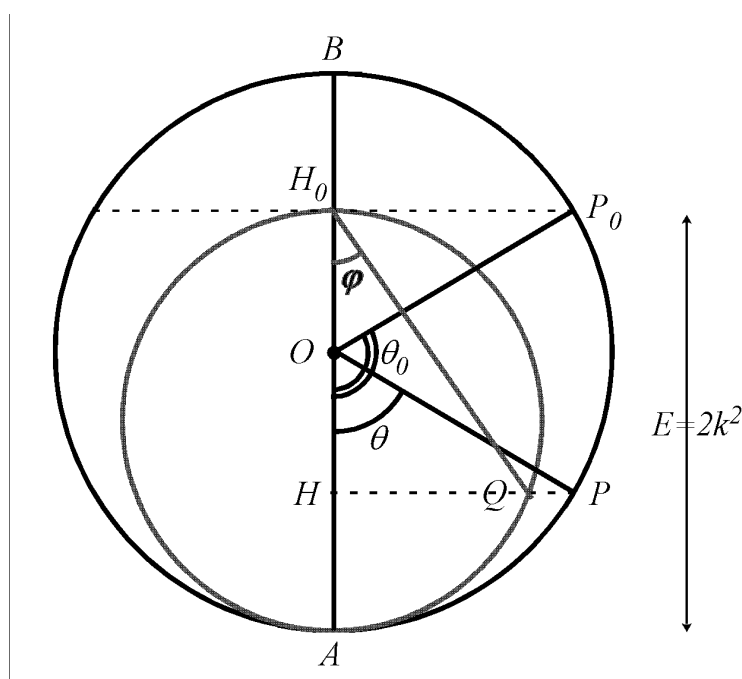


FIGURE 3. La normalisation du mouvement oscillatoire d'un pendule (voir texte).

Soit alors φ l'angle $\widehat{AH_0Q}$. On voit que lorsque $\theta = \theta_0$ est maximal la valeur maximale de φ est toujours $\frac{\pi}{2}$ et que sur C_0 le point Q décrit un demi-tour complet jusqu'à sa valeur maximale $Q_0 = H_0$. Lorsque $k \sim 0$ et que l'on est dans le régime des petites oscillations, H_0 devient très proche de A , le cercle C_0 devient très petit et les petites oscillations qui, sur le cercle C s'effectuent linéairement sur la tangente en A , s'enroulent sur la circonférence de C_0 . À la limite, lorsque $k = 0$, le cercle C_0 se contracte sur le point A et la durée du mouvement sur C_0 s'annule. Au contraire, lorsque $k = 1$, H_0 est en B , C_0 s'identifie à C et la position du pendule n'est plus paramétrée par θ mais par $\varphi = \frac{\theta}{2}$.

À travers cette normalisation, la durée du mouvement entre A et P_0 est le temps mis à parcourir le (demi-) cercle C_0 : elle varie de 0 à ∞ lorsque k varie de 0 à 1, le cercle C_0 variant du cercle ponctuel A au cercle C .

7.2. Les fonctions elliptiques $\text{sn}(u)$, $\text{cn}(u)$ et $\text{dn}(u)$

Expliquons maintenant l'apparition d'intégrales elliptiques. Le demi-axe positif à partir duquel sont comptés les angles est l'axe OA donc $OH = \cos(\theta)$ et $OH_0 =$

$\cos(\theta_0)$. Compté à partir de θ ,

$$HA = 1 - \cos(\theta) = 2 \sin^2\left(\frac{\theta}{2}\right).$$

Mais compté à partir de φ , comme le triangle AH_0Q est rectangle par construction puisque Q est sur le cercle de diamètre AH_0 , HA est donné par

$$\begin{aligned} HA &= HH_0 + H_0A = H_0A - H_0Q \cos(\varphi) = H_0A - (H_0A \cos(\varphi)) \cos(\varphi) \\ &= H_0A (1 - \cos^2(\varphi)) = H_0A \sin^2(\varphi) \end{aligned}$$

Et comme $H_0A = 2 \sin^2\left(\frac{\theta_0}{2}\right)$, on obtient

$$HA = 2 \sin^2\left(\frac{\theta_0}{2}\right) \sin^2(\varphi),$$

et donc en définitive, puisque par ailleurs $HA = 2 \sin^2\left(\frac{\theta}{2}\right)$ et $k = \sin\left(\frac{\theta_0}{2}\right)$, la relation

$$\sin\left(\frac{\theta}{2}\right) = \sin\left(\frac{\theta_0}{2}\right) \sin(\varphi) = k \sin(\varphi)$$

avec $\varphi = 0$ si $\theta = 0$ et $\varphi = \frac{\pi}{2}$ si $\theta = \theta_0$. Évidemment, si $k = 1$, on retrouve l'égalité $\sin\left(\frac{\theta}{2}\right) = \sin(\varphi)$.

Il est donc naturel de faire le changement de variable un peu sophistiqué

$$\varphi = \text{Arcsin}\left(\frac{1}{k} \sin\left(\frac{\theta}{2}\right)\right).$$

Au niveau des différentielles on en tire

$$\frac{1}{2} \cos\left(\frac{\theta}{2}\right) \dot{\theta} = k \cos(\varphi) \dot{\varphi}$$

et par conséquent

$$\begin{aligned} \dot{\theta}^2 &= \frac{4k^2 \cos^2(\varphi)}{\cos^2\left(\frac{\theta}{2}\right)} \dot{\varphi}^2 = \frac{4k^2 (1 - \sin^2(\varphi))}{1 - \sin^2\left(\frac{\theta}{2}\right)} \dot{\varphi}^2 \\ &= \frac{4k^2 (1 - \sin^2(\varphi))}{1 - k^2 \sin^2(\varphi)} \dot{\varphi}^2. \end{aligned}$$

Mais comme $\dot{\theta}^2 = 4(k^2 - \sin^2\left(\frac{\theta}{2}\right)) = 4k^2(1 - \sin^2(\varphi))$, on obtient l'équation différentielle pour φ

$$\dot{\varphi}^2 = 1 - k^2 \sin^2(\varphi)$$

soit, en inversant,

$$dt = \frac{d\varphi}{\sqrt{1 - k^2 \sin^2(\varphi)}}$$

autrement dit, comme nous le précisons à la section 11, une intégrale elliptique de *module* k . Puisque, étant donnée la symétrie $\theta \rightarrow -\theta$, on peut se restreindre à $\theta \in [0, \pi]$, on peut aussi se restreindre à $\varphi \in [0, \frac{\pi}{2}]$.

Nous obtenons ainsi la *durée* du mouvement $t = u(\varphi)$ en fonction de la position φ et non l'inverse comme on en a l'habitude. Cela permet de *mesurer le temps* et c'est effectivement à cette "chronométrie" que sert entre autres choses le pendule. Calculer t par l'intégrale

$$t = u(\varphi) = \int_0^\varphi \frac{d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}} ,$$

revient à prendre pour $t = 0$ le passage du pendule en A où $\varphi = 0$. L'angle φ considéré comme fonction réciproque de $t = u(\varphi)$ s'appelle traditionnellement l'*amplitude* de $t = u(\varphi)$ et est notée $\text{am}(u)$. L'amplitude φ est une observable *spatiale* qui permet de chronométrer des *durées*. Elle est par définition le changement de variable $\varphi = \text{Arcsin}\left(\frac{1}{k} \sin\left(\frac{\theta}{2}\right)\right)$ dont nous sommes partis. On a donc (il est bon de s'exercer à cette gymnastique de changements de variable)

$$u(\text{am}(u)) = \int_0^{\text{am}(u)} \frac{d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}} = \int_0^\varphi \frac{d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}} = u(\varphi) = t .$$

On définit alors, depuis Jacobi, les cosinus et sinus elliptiques de $t = u(\varphi)$, notés $\text{cn}(u)$ et $\text{sn}(u)$, comme les cosinus et sinus classiques (circulaires) de l'amplitude φ . On a donc par définition

$$\text{cn}^2(u) + \text{sn}^2(u) = 1 .$$

Le pendule convertit l'espace (amplitudes φ) en temps (durées de mouvement $t = u(\varphi)$) et la trigonométrie classique des angles (des oscillations spatiales normalisées) se convertit ainsi en trigonométrie "chronométrique". On notera que pour $k = 0$ (pendule en équilibre stable en A) $t = u = \varphi$ et que donc $\text{sn}(u) = \sin(\varphi)$ devient tautologique. Lorsque le pendule est immobile à l'équilibre stable, $\varphi = 0$ géométriquement et $t = 0$ mécaniquement car, comme nous l'avons vu, puisqu'il n'y a pas de mouvement, la durée du mouvement sur le cercle C_0 , qui est alors contracté sur le point A , est nulle. En revanche, pour $k = 1$, la durée du mouvement est infinie puisque le pendule met un temps infini pour rejoindre son équilibre instable en B .

Les fonctions elliptiques $\text{sn}(u)$ et $\text{cn}(u)$ liées aux fonctions $\sin(\varphi)$ et $\cos(\varphi)$ ne suffisent cependant pas à élaborer la trigonométrie "chronométrique". En plus de $\text{cn}(u) = \sqrt{1 - \text{sn}^2(u)}$, il faut introduire une *troisième* fonction elliptique

$$\text{dn}(u) = \sqrt{1 - k^2 \text{sn}^2(u)} ,$$

i.e. $\text{dn}^2(u) + k^2 \text{sn}^2(u) = 1$. Pour $k = 0$, $\text{dn}(u) \equiv 1$ et pour $k = 1$, $\text{dn}(u) \equiv \text{cn}(u)$. En utilisant $\text{cn}^2(u) + \text{sn}^2(u) = 1$, et en définissant ce que l'on appelle le *comodule* k' par

$k'^2 = 1 - k^2$, on tire de cette définition de $\operatorname{dn}(u)$ les autres relations remarquables

$$k^2 \operatorname{cn}^2(u) + k'^2 = \operatorname{dn}^2(u) \quad , \quad \operatorname{cn}^2(u) + k'^2 \operatorname{sn}^2(u) = \operatorname{dn}^2(u) \quad .$$

En ce qui concerne leur parité, $\operatorname{cn}(u)$ et $\operatorname{dn}(u)$ sont des fonctions paires, tandis que $\operatorname{sn}(u)$ est une fonction impaire.

Une des difficultés pédagogiques soulevées par ces intégrales elliptiques et qu'elles jouent non pas entre deux variables mais entre *trois* variables : $t = u(\varphi)$, $\cos(\varphi)$ et $\sin(\varphi)$, $\operatorname{cn}(u)$ et $\operatorname{sn}(u)$. Nous venons de voir l'intégrale chronométrant t en fonction de la différentielle associée à φ , c'est-à-dire à partir de $\varphi = \operatorname{Arcsin}\left(\frac{1}{k} \sin\left(\frac{\vartheta}{2}\right)\right)$ (conversion de l'espace en temps). Mais, dans la mesure où l'on a par définition $\varphi = \operatorname{Arcsin}(\sin(\varphi))$ on peut aussi calculer t en fonction de la différentielle associée à la variable $\sigma = \sin(\psi)$. On obtient alors une autre intégrale elliptique à partir de laquelle t s'obtient par inversion :

$$\begin{aligned} t = u(\varphi) &= \int_0^{\sin(\varphi)} \frac{d\sigma}{\cos(\psi) \sqrt{1 - k^2 \sin^2(\psi)}} \\ &= \int_0^{\operatorname{sn}(u)} \frac{d\sigma}{\sqrt{1 - \sigma^2} \sqrt{1 - k^2 \sigma^2}} . \end{aligned}$$

On pourra comparer cette dernière formule à celle que nous avons rencontrée pour les fonctions circulaires donnant la longueur d'un arc de cercle par la formule

$$\int_0^{\sin(\varphi)} \frac{1}{\sqrt{1 - t^2}} dt .$$

C'est à cause de la forme des intégrales en $\sqrt{1 - \sigma^2} \sqrt{1 - k^2 \sigma^2}$ qu'on doit introduire $\operatorname{dn}(u)$.

On note traditionnellement K l'intégrale "complète"

$$K = \int_0^1 \frac{d\sigma}{\sqrt{1 - \sigma^2} \sqrt{1 - k^2 \sigma^2}} = \int_0^{\frac{\pi}{2}} \frac{d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}} = u\left(\frac{\pi}{2}\right) .$$

Il s'agit d'un *quart* de période pour les fonctions elliptiques, l'amplitude φ augmentant de $\frac{\pi}{2}$ lorsque $t = u(\varphi)$ augmente de K . Il est trivial de vérifier les valeurs particulières, pour tout k ,

$$\operatorname{cn}(0) = 1, \quad \operatorname{sn}(0) = 0, \quad \operatorname{dn}(0) = 1, \quad \operatorname{cn}(K) = 0, \quad \operatorname{sn}(K) = 1, \quad \operatorname{dn}(K) = \sqrt{1 - k^2} = k',$$

où k' est le comodule. On a également

$$\operatorname{cn}\left(\frac{K}{2}\right) = \sqrt{\frac{k'}{1 + k'}}, \quad \operatorname{sn}\left(\frac{K}{2}\right) = \frac{1}{\sqrt{1 + k'}}, \quad \operatorname{dn}\left(\frac{K}{2}\right) = \sqrt{k'} ,$$

ce qui, pour $k = 0$, et donc $\frac{K}{2} = \frac{\pi}{4}$ et $k' = 1$, redonne bien $\cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$, $\sin\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$ et $\operatorname{dn}\left(\frac{\pi}{4}\right) = 1$.

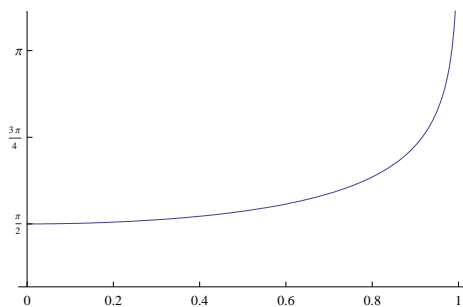


FIGURE 4. Le graphe de la fonction période $K(k)$ en fonction du module k .

On peut résumer la situation dans le diagramme suivant (où ψ est la variable courante associée à φ).

$$\begin{array}{ccc}
 \text{sn}(u) \in [0, 1] & \begin{array}{c} \xleftarrow{\text{sn}} \\ \xrightarrow{\text{sn}^{-1}} \\ \nearrow \end{array} & \begin{array}{l} u \in [0, K] \\ u(\varphi) = t \\ u \uparrow \downarrow u^{-1} = \text{am}(u) \end{array} \\
 \parallel & & \\
 \sin(\varphi) \in [0, 1] & \begin{array}{c} \xleftarrow{\sin} \\ \xleftarrow{\text{Arcsin}} \end{array} & \begin{array}{l} \varphi \in [0, \frac{\pi}{2}] \\ \varphi = \text{am}(u) \end{array} \\
 \sigma = \sin(\psi) & &
 \end{array}$$

Il est intéressant de représenter le graphe de la fonction $K(k)$ (cf. figure 4). Pour $k = 0$ (pendule à l'équilibre), $K = \int_0^1 \frac{d\sigma}{\sqrt{1-\sigma^2}} = \frac{\pi}{2}$, ce qui va de soi puisque K est un quart de période de $\text{sn}(u)$, que $u = \varphi$, $\text{sn}(u) = \sin(\varphi)$ et que $\frac{\pi}{2}$ est le quart de période de $\sin(\varphi)$. En revanche, pour $k = 1$, $K(k)$ diverge. En effet $K(1)$ est la limite

$$\int_0^{a \rightarrow \infty} \frac{d\sigma}{1-\sigma^2} = \lim_{a \rightarrow \infty} \frac{1}{2} \log \left(\frac{1+a}{1-a} \right) .$$

Donc $K(1^-)$ diverge comme $-\log(0^+)$ et plus précisément comme $4 \log \left(\frac{4}{k'} \right)$ avec $k' = \sqrt{1-k^2}$. Le graphe de $K(k)$ montre que la fonction est très plate et ne diverge (logarithmiquement) que pour k très proche de 1.

Nous allons voir que le comportement des fonctions elliptiques de Jacobi en fonction de k est analogue : sur presque tout l'intervalle $[0, 1[$ elles évoluent lentement à partir de $k = 0$ où cn et sn sont les fonctions trigonométriques et où $\text{dn}(u) \equiv 1$ pour se transformer très rapidement et très brutalement au voisinage de $k = 1$. On peut montrer que l'intégrale $\int_0^1 K(k) dk$ est le double de la constante de Catalan. Cette constante est donnée par la série $\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2}$ et se trouve reliée à d'autres fonctions comme la fonction $\beta(s) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^s}$ de Dirichlet (elle vaut $\beta(2)$) ou la fonction $\chi_v(z) = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)^v}$ de Legendre (elle vaut $-i\chi_2(i)$).

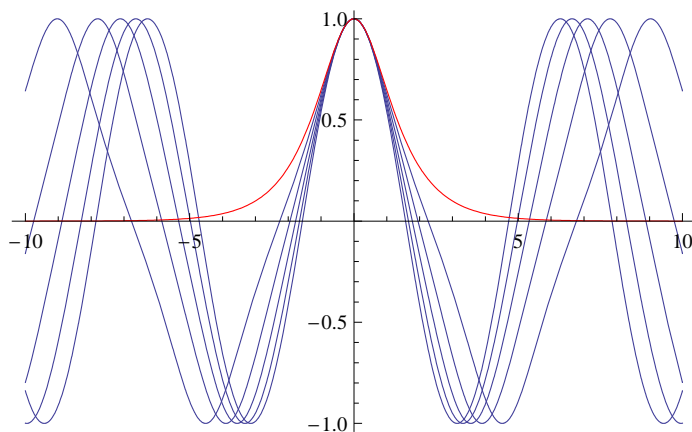


FIGURE 5. La famille des $\text{cn}(u, k)$ pour k^2 allant de 0 à 1 par pas de 0.2 et $u \in [-10, 10]$. Le cosinus initial $\text{cn}(u, 0) = \cos(u)$ se déforme peu. Pour $k^2 = 0.8$, $\text{cn}(u, k)$ est encore très cosinusoidal, alors que pour $k = 1$, $\text{cn}(u, 1) = \frac{1}{\cosh(u)}$ (courbe en rouge).

7.3. La limite $k \rightarrow 1$

Pour $k = 0$, on retrouve évidemment la trigonométrie “circulaire” classique. À l’autre extrême, à la limite $k = 1$, on a $\varphi = \frac{\theta}{2}$ et la trigonométrie “chronométrique” du pendule déformant la trigonométrie “circulaire” devient une sorte de trigonométrie “hyperbolique inversée”. En effet, pour $k = 1$,

$$\text{cn}(u, 1) = \frac{1}{\cosh(u)} = \text{dn}(u, 1), \quad \text{sn}(u, 1) = \tanh(u).$$

Cette sorte d’*interpolation* entre ces deux trigonométries est en quelque sorte mesurée par $\text{dn}(u, k)$ qui pour $k = 0$ ne compte pas parce qu’il est égal à 1 et qui pour $k = 1$ s’identifie à $\text{cn}(u, k)$.

La figure 5 montre la famille des $\text{cn}(u, k)$ pour k^2 allant de 0 à 1 par pas de 0.2. On voit que le cosinus initial $\text{cn}(u, 0) = \cos(u)$ se déforme peu, que, pour $k^2 = 0.8$, $\text{cn}(u, k)$ est encore très cosinusoidal, alors que pour $k = 1$, il devient $\frac{1}{\cosh(u)}$ (courbe en rouge).

La figure 6 montre $\text{cn}(u, 0.95)$ en bleu, $\text{cn}(u, 0.9995)$ en orange et $\text{cn}(u, 1)$ (légèrement décalée pour la visibilité) en rouge. On voit que $\text{cn}(u, 0.95)$ ressemble encore assez à un cosinus puis que les tangentes des points d’intersection avec l’axe des u s’horizontalisent comme en des points d’inflexion en même temps que K augmente rapidement. À la limite $k = 1$ où K diverge, les oscillations, devenues très espacées, disparaissent à l’infini et il ne reste plus que la partie centrale $u \in [-K, K]$ s’étendant sur l’axe entier $(-\infty, \infty)$.

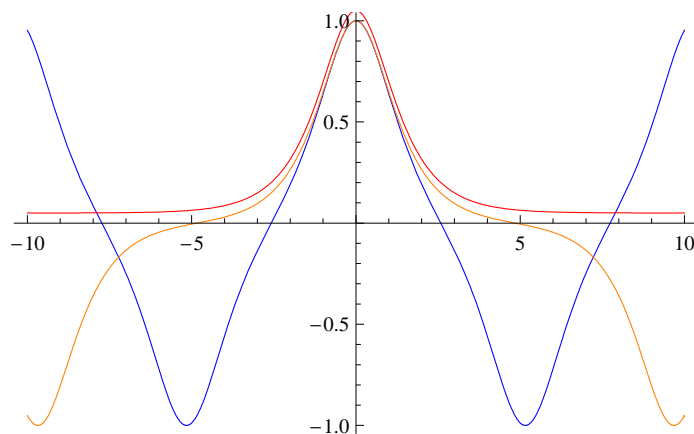


FIGURE 6. $\text{cn}(u, 0.95)$ en bleu, $\text{cn}(u, 0.9995)$ en orange et $\text{cn}(u, 1) = \frac{1}{\cosh(u)}$ en rouge pour $u \in [-10, 10]$. $\text{cn}(u, 1)$ est décalée vers le haut de 0.05 pour la visibilité.

Il en va de même pour $\text{sn}(u, k)$. Pour $k = 1$, on a

$$\begin{aligned} \text{sn}^2(u, 1) &= 1 - \text{cn}^2(u, 1) = 1 - \frac{1}{\cosh^2(u)} = \frac{\cosh^2(u) - 1}{\cosh^2(u)} \\ &= \frac{\sinh^2(u)}{\cosh^2(u)} = \tanh^2(u) . \end{aligned}$$

La figure 7 montre en haut à gauche la famille $\text{sn}(u, k)$ pour k^2 allant de 0 à 0.8 par pas de 0.2. On voit que le sinus initial $\text{sn}(u, 0) = \sin(u)$ se déforme peu. En haut à droite est représentée la limite $\tanh(u) = \text{sn}(u, 1)$ de $\text{sn}(u, k)$ pour $k = 1$. En bas à gauche, est représenté $\text{sn}(u, 0.9995)$ et en bas à droite sa comparaison avec $\text{sn}(u, 1)$. On voit que, alors que $\text{cn}(u, k)$ s'étranglait et s'aplatissait sur l'axe horizontal, $\text{sn}(u, k)$ se met à présenter des plateaux périodiques de plus en plus larges de hauteur approchant 1 et -1 (ce qui est obligatoire puisque $\text{cn}^2 + \text{sn}^2 = 1$), proches de 1 dans les intervalles $[4nK, 4(n + \frac{1}{2})K]$ ($n \in \mathbb{Z}$) et proches de -1 dans les intervalles $[4(n + \frac{1}{2})K, 4(n + 1)K]$, et que, à la limite $k = 1$, il ne reste plus que les deux plateaux infinis de $\tanh(u)$.

Enfin, en ce qui concerne la troisième fonction elliptique de Jacobi $\text{dn}(u)$, la figure 8 montre, à gauche, la façon dont $\text{dn}(u, k)$ évolue à partir de $\text{dn}(u, 0) \equiv 1$ pour se rapprocher des demi-oscillations positives de $\text{cn}(u, k)$ sur les intervalles $[(4n - 1)K, (4n + 1)K]$ ($n \in \mathbb{Z}$) et au contraire s'opposer aux demi-oscillations négatives de $\text{cn}(u, k)$ dans les intervalles $[(4n + 1)K, (4n + 3)K]$. On voit que pour k proche de 1, $\text{dn}(u, k)$ n'est pas du tout globalement voisine de $\text{cn}(u, k)$, mais seulement sur une demi-période sur 2. À droite, on voit que l'identification à la limite $k = 1$ de $\text{dn}(u, k)$ à $\text{cn}(u, k)$ (toutes deux égales à $\frac{1}{\cosh(u)}$) n'est donc pas

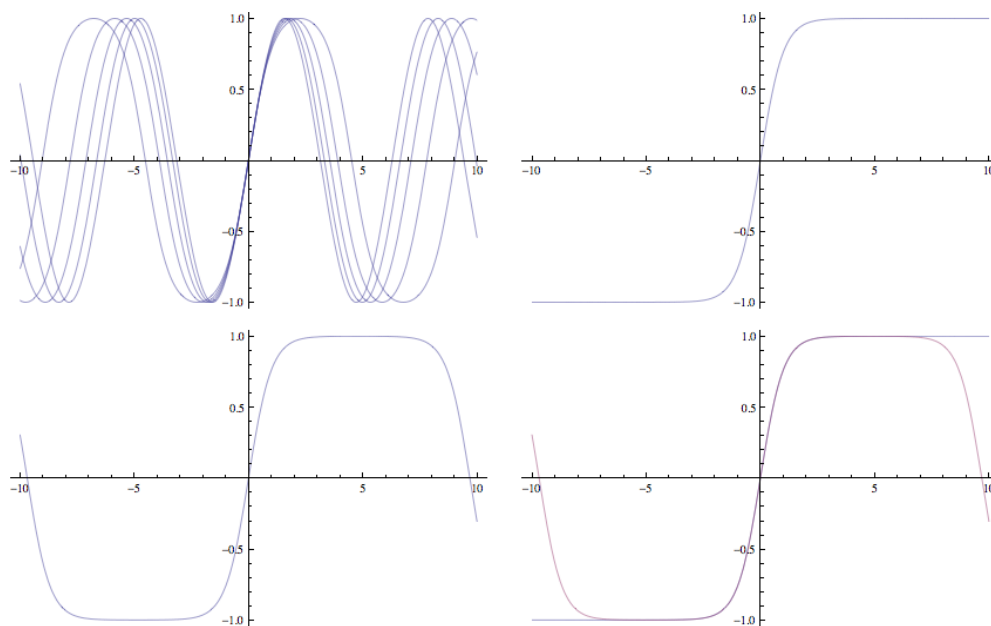


FIGURE 7. En haut à gauche, la famille $\text{sn}(u, k)$ pour $u \in [-10, 10]$ et k^2 allant de 0 à 0.8 par pas de 0.2. Le sinus initial $\text{sn}(u, 0) = \sin(u)$ se déforme peu. En haut à droite, la limite $\tanh(u) = \text{sn}(u, 1)$ de $\text{sn}(u, k)$ pour $k = 1$. En bas à gauche, $\text{sn}(u, 0.9995)$. En bas à droite la comparaison avec $\text{sn}(u, 1) = \tanh(u)$.

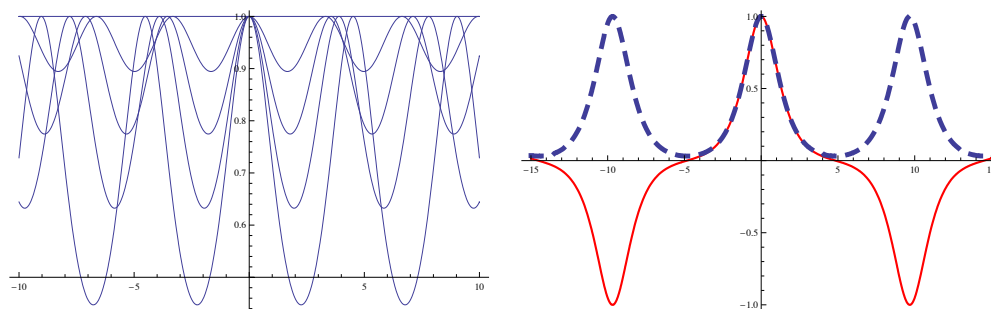


FIGURE 8. À gauche, l'évolution de $\text{dn}(u, k)$ vers $\text{cn}(u, k)$ à partir de $\text{dn}(u, 0) \equiv 1$ pour $u \in [-10, 10]$ et k^2 allant de 0 à 0.8 par pas de 0.2. À droite comparaison de $\text{dn}(u, 0.9995)$ avec $\text{cn}(u, 0.9995)$. À la limite $k = 1$ $\text{dn}(u, k)$ et $\text{cn}(u, k)$ deviennent toutes deux égales à $\frac{1}{\cosh(u)}$ représentée en rouge à la figure 5.

intuitive. Elle est liée à la divergence de K envoyant à l'infini toutes les oscillations et ne gardant que la partie centrale.

7.4. Translations, dérivations et formules d'addition

On démontre facilement pour la trigonométrie “chronométrique” des formules en tous points analogues à celle de la trigonométrie classique sauf que l'on doit tenir compte des trois fonctions elliptiques. Par exemple, nous allons voir que l'ajout du quart de période K donne

$$\operatorname{sn}(u + K) = \frac{\operatorname{cn}(u)}{\operatorname{dn}(u)}, \quad \sin\left(\varphi + \frac{\pi}{2}\right) = \cos(\varphi) \quad \text{pour } k = 0$$

$$\operatorname{cn}(u + K) = -k' \frac{\operatorname{sn}(u)}{\operatorname{dn}(u)}, \quad \cos\left(\varphi + \frac{\pi}{2}\right) = -\sin(\varphi) \quad \text{pour } k = 0$$

$$\operatorname{dn}(u + K) = \frac{k'}{\operatorname{dn}(u)}$$

$$\operatorname{sn}(u + 2K) = \frac{\operatorname{cn}(u + K)}{\operatorname{dn}(u + K)} = -k' \frac{\operatorname{sn}(u)}{\operatorname{dn}(u)} \frac{\operatorname{dn}(u)}{k'} = -\operatorname{sn}(u)$$

$$\operatorname{cn}(u + 2K) = -k' \frac{\operatorname{sn}(u + K)}{\operatorname{dn}(u + K)} = -k' \frac{\operatorname{cn}(u)}{\operatorname{dn}(u)} \frac{\operatorname{dn}(u)}{k'} = -\operatorname{cn}(u)$$

$$\operatorname{dn}(u + 2K) = \frac{k'}{\operatorname{dn}(u + K)} = \operatorname{dn}(u) .$$

Donc $\operatorname{cn}(u)$ et $\operatorname{sn}(u)$ sont de périodes $4K$, mais $\operatorname{dn}(u)$ est seulement de période $2K$.

De même, on calcule facilement les formules de duplication

$$\operatorname{sn}^2(u) = \frac{1 - \operatorname{cn}(2u)}{1 + \operatorname{dn}(2u)}, \quad \cos(2\varphi) = 1 - 2\sin^2(\varphi) = \cos^2(\varphi) - \sin^2(\varphi) \quad \text{pour } k = 0$$

$$\operatorname{cn}^2(u) = \frac{\operatorname{dn}(2u) + \operatorname{cn}(2u)}{1 + \operatorname{dn}(2u)}$$

$$\operatorname{dn}^2(u) = \frac{\operatorname{dn}(2u) + \operatorname{cn}(2u)}{1 + \operatorname{cn}(2u)} .$$

Pour calculer les dérivées des fonctions elliptiques, on peut par exemple revenir à la définition $t = u(\varphi) = \int_0^\varphi \frac{d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}}$ et écrire

$$u'(\varphi) = \frac{du}{d\varphi} = \frac{1}{\sqrt{1 - k^2 \sin^2(\varphi)}} = \frac{1}{\sqrt{1 - k^2 \operatorname{sn}^2(u)}} = \frac{1}{\operatorname{dn}(u)}$$

$$\frac{d\varphi}{du} = \operatorname{dn}(u) .$$

Donc

$$\begin{aligned}
\frac{d(\operatorname{cn}(u))}{du} &= \frac{d(\cos(\varphi))}{du} = -\sin(\varphi) \frac{d\varphi}{du} = -\operatorname{sn}(u) \operatorname{dn}(u) \\
\frac{d(\operatorname{sn}(u))}{du} &= \frac{d\sin(\varphi)}{du} = \cos(\varphi) \frac{d\varphi}{du} = \operatorname{cn}(u) \operatorname{dn}(u) \\
\frac{d(\operatorname{dn}(u))}{du} &= \frac{d\sqrt{1-k^2\operatorname{sn}^2(u)}}{du} = \frac{-2k^2\operatorname{sn}(u)\operatorname{sn}'(u)}{2\sqrt{1-k^2\operatorname{sn}^2(u)}} \\
&= \frac{-k^2\operatorname{sn}(u)\operatorname{cn}(u)\operatorname{dn}(u)}{\operatorname{dn}(u)} = -k^2\operatorname{sn}(u)\operatorname{cn}(u) .
\end{aligned}$$

On vérifie que pour $k = 0$ on retrouve bien les dérivations de la trigonométrie classique et que pour $k = 1$ on retrouve bien les dérivées de $\frac{1}{\cosh(u)}$ et $\tanh(u)$.

Pour les formules d'addition, pour $u, v, u + v \in [0, K]$ on vérifie, par exemple avec la méthode déjà utilisée à la section 5.2 pour les fonctions circulaires, que

$$\begin{aligned}
\operatorname{sn}(u + v) &= \frac{\operatorname{sn}(u)\operatorname{cn}(v)\operatorname{dn}(v) + \operatorname{sn}(v)\operatorname{cn}(u)\operatorname{dn}(u)}{1 - k^2\operatorname{sn}^2(u)\operatorname{sn}^2(v)} \\
\operatorname{cn}(u + v) &= \frac{\operatorname{cn}(u)\operatorname{cn}(v) - \operatorname{sn}(u)\operatorname{sn}(v)\operatorname{dn}(u)\operatorname{dn}(v)}{1 - k^2\operatorname{sn}^2(u)\operatorname{sn}^2(v)} \\
\operatorname{dn}(u + v) &= \frac{\operatorname{dn}(u)\operatorname{dn}(v) - k^2\operatorname{sn}(u)\operatorname{sn}(v)\operatorname{dn}(u)\operatorname{dn}(v)}{1 - k^2\operatorname{sn}^2(u)\operatorname{sn}^2(v)} .
\end{aligned}$$

On en dérive immédiatement

$$\begin{aligned}
\operatorname{sn}(2K - v) &= -\operatorname{sn}(-v) = \operatorname{sn}(v) , \\
\operatorname{cn}(2K - v) &= -\operatorname{cn}(-v) = -\operatorname{cn}(v) , \\
\operatorname{dn}(2K - v) &= \operatorname{dn}(-v) = \operatorname{dn}(v) .
\end{aligned}$$

7.5. Extension à \mathbb{R} et à \mathbb{C}

Ensuite, grâce aux formules d'addition, on étend les fonctions elliptiques de Jacobi à \mathbb{R} . Supposons par exemple $u \in [K, 2K]$. On écrit

$$\begin{aligned} \operatorname{sn}(u) &= \operatorname{sn}((u - K) + K) \\ &= \frac{\operatorname{sn}(u - K) \operatorname{cn}(K) \operatorname{dn}(K) + \operatorname{sn}(K) \operatorname{cn}(u - K) \operatorname{dn}(u - K)}{1 - k^2 \operatorname{sn}^2(u - K) \operatorname{sn}^2(K)} \\ &= \frac{\operatorname{cn}(u - K) \operatorname{dn}(u - K)}{1 - k^2 \operatorname{sn}^2(u - K)} \\ &= \frac{\operatorname{cn}(u - K)}{\operatorname{dn}(u - K)}. \end{aligned}$$

De même

$$\operatorname{cn}(u) = -k' \frac{\operatorname{sn}(u - K)}{\operatorname{dn}(u - K)}, \quad \operatorname{dn}(u) = \frac{k'}{\operatorname{dn}(u - K)}.$$

On en déduit les formules introduites plus haut.

Pour prolonger à \mathbb{C} , on étend d'abord à $i\mathbb{R}$ en utilisant le comodule $k' = \sqrt{1 - k^2}$ et l'élément de période associé $K' = \int_0^1 \frac{d\sigma}{\sqrt{1 - \sigma^2} \sqrt{1 - k'^2 \sigma^2}}$. On obtient

$$\operatorname{sn}(iv, k) = i \frac{\operatorname{sn}(v, k')}{\operatorname{cn}(v, k')}, \quad \operatorname{cn}(iv, k) = \frac{1}{\operatorname{cn}(v, k')}, \quad \operatorname{dn}(iv, k) = \frac{\operatorname{dn}(v, k')}{\operatorname{cn}(v, k')}.$$

On définit ensuite $\operatorname{sn}(u + iv, k)$, etc. au moyen des formules d'addition et on montre que les fonctions prolongées à \mathbb{C} satisfont bien les équations de Cauchy-Riemann.

La grande découverte fut, en ce qui concerne les périodes, de montrer qu'il existe des périodes imaginaires pures, à savoir que, en plus de leur période réelle de $4K$, sn et dn sont $4iK'$ -périodiques alors que cn est $2iK'$ -périodique. Les fonctions circulaires ont une seule période car la partie imaginaire de leur argument fait intervenir la trigonométrie hyperbolique qui est sans période. Quand on les étend au plan complexe \mathbb{C} , elles sont définies sur le *cylindre* $\mathbb{S}^1 \times \mathbb{R}$. En revanche les fonctions elliptiques de Jacobi sont définies sur le *tore* quotient de \mathbb{C} par le parallélogramme des deux périodes $(4K, 4iK')$. Ce tore est *compact* et, comme elles sont des fonctions analytiques, cela implique de nombreuses propriétés. De même, les fonctions lemniscatiques sont définies sur le tore quotient de \mathbb{C} par le parallélogramme des périodes $(2\pi G, 2\pi iG)$ où G est la constante de Gauss.

C'est vraiment cette double périodicité qui fut la découverte majeure de la trigonométrie "chronométrique" du pendule.

8. Calcul des variations : le *Methodus* d'Euler (1744)

Nous allons maintenant revenir aux sources du calcul des variations permettant de dériver les équations de tout un ensemble de courbes devant satisfaire des propriétés d'optimalité sous contraintes. C'est Euler qui a systématisé cette méthode extraordinaire qui gouverne depuis tout un continent de la géométrie différentielle et pratiquement toute la physique lagrangienne. Ce qui nous intéresse dans ce retour aux sources est qu'on y rencontre une conception qui est parfaitement affine à ce que nous avons expliqué jusqu'ici. Nous avons souvent dit que les architectures fonctionnelles que modélise la neurogéométrie implémentent, en termes de jets et de formes différentielles associées, des formulations du calcul différentiel qui plongent jusqu'à ses racines conceptuelles. Mais le bref retour amont que nous allons effectuer montre que ces racines sont également les racines *historiques*.

C'est dans un texte admirable publié en 1744 (et édité par Constantin Carathéodory dont nous croiserons souvent le nom) [173] qu'Euler expose et exemplifie de façon éblouissante sa nouvelle *Méthode pour trouver les lignes courbes jouissant de propriétés de maximum et de minimum, ou solution de problèmes isopérimétriques au sens le plus large du terme*.²⁰ C'est dans ce texte qu'apparaissent les équations d'Euler qui deviendront plus tard les équations d'Euler-Lagrange, la méthode dite plus tard des multiplicateurs de Lagrange et le principe de moindre action comme un principe général valide bien au-delà du principe de Fermat en optique. Le texte sur les *elasticæ De curvis elastica* [174] qui nous motive particulièrement en constitue l'Annexe 1 (*Additamentum*). Nous ne commenterons ici que le *Methodus* et son Annexe.

On trouve le texte dans les *Opera Omnia* (Series 1, volume 24) et à la référence E65 du site *Euler Archive* (eulerarchive.maa.org [176]) de la Mathematical Association of America avec les ouvrages et commentaires édités par Edward Sandifer. Une traduction allemande du *Methodus* a été proposée par Alexander Aycok et Artur Diener (toujours référence E65 dans *Euler Archive*). Sa traduction anglaise se trouve sur le site de Ian Bruce [81]. Pour l'histoire générale du Calcul des variations on pourra se référer par exemple au classique ouvrage de Herman Heine Goldstine [216]. Quant à l'œuvre d'Euler, elle est si monumentale que nous renvoyons au site *Euler Archive* où l'on trouvera les références de plusieurs ouvrages célébrant le tricentenaire de sa naissance (2007) et les remarquables notes "How Euler did it" d'Edward Sandifer. On pourra aussi consulter l'ouvrage d'Emil Fellmann [181]. De nombreuses traductions des textes latins d'Euler (et d'autres mathématiciens, y compris Newton et Leibniz) sont disponibles sur le site [81] de Ian Bruce.

Remarque. Dans ce qui suit, nous suivons de façon un peu précise le texte d'Euler pour saisir l'acte de naissance des principes variationnels. Notre souci est,

20. *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici lattissimo sensu accepti.*

conformément à nos choix pédagogiques et épistémologiques, un retour amont aux sources historiques, mais également une mise en évidence de la façon dont de tels principes peuvent être neuralemement implémentés. Nous respecterons presque à la lettre le texte d'Euler tout en nous permettant d'utiliser parfois des notations un petit peu plus modernes, d'autant plus qu'il y a dans le texte de nombreuses ambiguïtés de notation. □

Euler considère des courbes dans le plan avec un axe des abscisses (*abscissa*) x et un axe des ordonnées (*applicata*) y . Il les suppose suffisamment différentiables et il les schématise comme des courbes régulières au-dessus d'un segment AZ de l'axe des x correspondant à un intervalle que nous noterons si besoin est (x_A, x_Z) .²¹ Il considère alors une "formule" (une "fonctionnelle" comme on a dit ensuite) $W = \int Z dx$ fonction de la courbe²² et cherche les conditions pour que l'intégrale $\int_{x_A}^{x_Z} Z dx$ soit maximale ou minimale (nous dirions maintenant "extrémale" ou "optimale"). Euler n'utilise pas explicitement la notation d'intégrale définie (qui ne sera introduite que beaucoup plus tard par Fourier) mais la notation d'intégrale indéfinie $\int Z dx$ en précisant les bornes quand c'est nécessaire. Il parle de trouver une courbe "inter omnes omnino curvas eidem abscissæ respondentēs".

Ce qui est remarquable chez Euler est sa confiance absolue envers le calcul différentiel *symbolique* à la Leibniz. Il en va d'ailleurs de même pour ses calculs algébriques et ses calculs de séries (dont il était un incroyable virtuose) qu'il effectue *formellement* sans trop se soucier de "rigueur". Très inspiré par les "honorables frères Bernoulli" qui étaient radicalement leibniziens, il estimait qu'il fallait calculer formellement avec des symboles différentiels régis par certaines règles et qu'ensuite seulement on pouvait introduire des valeurs numériques permettant d'évaluer les symboles différentiels en termes de différences finies. Contrairement aux newtoniens qui pensaient en termes de "fluxions" (de limites, comme on le fera à nouveau au XIX^e), il pensait que ce que nous pourrions appeler "l'architecture symbolique du calcul" était primordiale.

Cela se voit d'emblée dans le fait qu'il introduit des symboles de variables *indépendantes* pour les dérivées. Il avait déjà le concept de fonction puisqu'il en était l'un des inventeurs. Il utilisait donc ce que nous notons aujourd'hui $y = f(x)$. Mais il introduisit de nouveaux symboles "littéraux" $p = \frac{dy}{dx}$, $q = \frac{dp}{dx}$, $r = \frac{dq}{dx}$, $s = \frac{dr}{dx}$, etc. qu'il traita comme de nouveaux symboles *de variables*. Il s'agit là du geste clé qui préfigure la notion fondamentale de jet au XX^e siècle. En effet, Euler peut alors poser que la "formule" Z est une fonction de plusieurs (à la limite une infinité) variables x, y, p, q, \dots et il peut considérer des dérivées partielles

21. La notation $A\dots Z$ pour les abscisses et $a\dots z$ pour les points des courbes lui permet de les énumérer alphabétiquement.

22. On notera l'ambiguïté de notation, Z référant à la fois à l'extrémité de l'intervalle AZ et à l'intégrande de W .

de Z (par exemple $\frac{\partial Z}{\partial p}$) par rapport à ces variables. C'est de cette possibilité que naîtront ses célèbres équations. Autrement dit, au lieu de travailler dans un *plan* de coordonnées (x, y) en y considérant des courbes avec leurs tangentes, courbures, inflexions, etc., Euler travaille déjà, de façon implicite et non rigoureuse, dans un espace de dimension *infinie* de coordonnées (x, y, p, q, r, \dots) mais en tenant compte du fait qu'une courbe plane va s'envoyer comme une courbe dans cet espace infini parce que toutes les coordonnées vont devenir des fonctions de x cohérentes entre elles. L'espace des (x, y, p, q, r, \dots) est bien implicitement de dimension infinie puisque les développements de fonctions régulières en séries de Taylor comprenant une infinité de dérivées étaient déjà bien connus et massivement utilisés dans des calculs formels.

Euler va adapter à ces intégrales (ces "fonctionnelles") qu'il s'agit d'optimiser ce qui se passe de façon élémentaire pour les fonctions : si $F(x)$ est une fonction (dérivable), la condition pour que $F(a)$ soit un maximum ou un minimum est que $F'(a) = 0$ (condition du premier ordre). Alors, $F(a)$ sera un maximum si $F''(a) < 0$ et un minimum si $F''(a) > 0$ (condition du deuxième ordre). Dès le chapitre 1, note (*Scholium*) 1, §6, il explique que ce n'est plus un point sur une courbe qui doit être un extremum, mais une courbe toute entière qui doit optimiser une intégrale.²³

Dans son *Methodus*, Euler donne des exemples très nombreux et très variés (une centaine) : de courbes spéciales comme la brachistochrone, les géodésiques sur des surfaces entre deux points fixés, des solutions de problèmes comprenant des contraintes isopérimétriques comme chez Jakob Bernouilli (cf. les *elasticæ*), des solutions plus faciles à calculer en changeant de coordonnées, des problèmes mécaniques relevant du principe de moindre action.

Dans le Corollaire I de l'Hypothèse I, §17, il explique qu'avec l'introduction des symboles p, q, r , etc., on peut, par substitution, éliminer des calculs toutes les différentielles autres que dx .²⁴ Il suppose alors dx "constante" comme différentielle de base servant à définir toutes les autres (§18)²⁵, ce qui signifie que la différentielle seconde d^2x satisfait $d^2x = 0$ et, dans le Corollaire III, §19, il définit symboliquement les différentielles successives de y en fonction de dx . Par définition de p , $dy = p dx$. Euler dérive alors $d^2y = q dx^2$, (i.e. $q = \frac{d^2y}{dx^2}$ comme on l'apprend encore aujourd'hui)

23. Hæc itaque Methodus maximorum & minimorum maxime discrepat ab illa, quam alibi exposimus. Ibi enim, pro data ac determinate linea curva, locum determinavimus, ubi proposita quædam quantitas variabilis ad curvam pertinens fiat maxima vel minima. Hic autem ipsa linea curva quæritur, in qua quantitas quædam proposita fiat maxima vel minima.

24. His igitur substitutionibus omnia differentialia ipsius y cujuscunque gradus ex expressionibus tollentur, atque præter differentiale dx nulla alia differentialia relinquentur.

25. Interim tamen, ob methodum infra adhibendam, necesse erit differentiale dx tanquam constans assumere.

en écrivant (à sa façon, nous changeons un peu la présentation)²⁶

$$\begin{aligned} d^2y &= d(dy) = d(pdx) = dpdx + pd^2x = dpdx = \frac{dp}{dx} dx^2 \\ &= qdx^2 . \end{aligned}$$

De même, $d^3y = rdx^3$, etc. Il est essentiel de souligner l'importance du concept de "substitution". La formule $p = \frac{dy}{dx}$ dit que le symbole p a vocation à être interprété comme une dérivée $f'(x)$ si y est donné par une fonction $y = f(x)$. Mais au niveau de l'"architecture symbolique" du calcul, p sert prioritairement à éliminer dy en lui substituant pdx . Toute expression différentielle comprenant des dérivées de y d'ordre aussi élevé que l'on veut pourra ainsi se ramener par élimination à une expression de plusieurs variables ne comprenant que dx comme différentielle. Il est vraiment fascinant de voir que, dès sa source eulérienne, le calcul des variations repose sur l'idée formalisée plus tard par les jets et les formes différentielles. L'idée force, longuement développée dans notre Vol I et dans la section 3, que $p = \frac{dy}{dx}$ signifie que la 1-forme différentielle $\omega = dy - pdx$ s'annule mais que cette 1-forme possède un sens symbolique indépendamment de cette condition d'intégrabilité et de toute courbe particulière parce que p fonctionne comme un symbole supplémentaire de variable indépendante et que l'on travaille implicitement dans l'espace de dimension infinie $(x, y, p, q, r \dots)$, cette idée force est donc déjà intuitivement présente chez Euler!

Dans les Corollaires suivant, (§§20-22), Euler exprime la longueur d'un arc de courbe, qu'il note w , par la formule $w = \int \sqrt{dx^2 + dy^2}$, la différentielle dw étant $\sqrt{dx^2 + dy^2}$ (théorème de Pythagore infinitésimal). Il applique aussitôt ses substitutions, ce qui donne la formule fondamentale $dw = \sqrt{1 + p^2} dx$. Puis il calcule

$$\begin{aligned} d^2w &= d\left(\sqrt{1 + p^2} dx\right) = \frac{2pdp}{\sqrt{1 + p^2}} dx = \frac{2pq}{\sqrt{1 + p^2}} dx^2 \\ d^3w &= \left(\frac{pr}{\sqrt{1 + p^2}} + \frac{q^2}{(1 + p^2)^{\frac{3}{2}}}\right) dx^3, \text{ etc.} \end{aligned}$$

Cela donne le rayon de courbure

$$-\frac{dw^3}{dx d^2y} = -\frac{(1 + p^2)^{\frac{3}{2}} dx^3}{dx q dx^2} = -\frac{(1 + p^2)^{\frac{3}{2}}}{q}$$

formule que nous apprenons toujours à l'école.

Dans ce chapitre I, après ces préliminaires, Euler commence à exposer sa méthode. Il considère un intervalle AZ de l'axe des x avec des points intermédiaires

26. d^2y est une différentielle du second ordre alors que dx^2 est un carré de différentielle du premier ordre.

M , N , etc., une courbe C au-dessus de AZ avec des points a , m , n , z et se donne une “formule” $W = \int Z dx$ (une fonctionnelle) dépendant de la “nature” de C et qu’il s’agit d’optimiser le long de C .²⁷ Il considère en fait des intégrales définies

$$W = \int_{x_A}^{x_Z} Z dx$$

qu’il appelle (§38) “valor formulæ $\int Z dx$ pro abscissa AZ ”. Il distingue le cas “absolu” où l’on cherche la courbe extrémale parmi “toutes” les courbes (évidemment il n’y a pas encore de notion d’espace fonctionnel chez Euler) et le cas “relatif” où l’on cherche la courbe extrémale seulement parmi les courbes satisfaisant certaines contraintes.

Il commence par les cas où l’intégrande Z de W est une fonction (en général algébrique) des variables x , y , p , etc. et ne contient pas elle-même d’intégrales.²⁸ Cela implique que Z est *locale*. Comme dit Euler, c’est une fonction “déterminée” dont la valeur ne dépend que des valeurs de x , y , p , ... en *un seul* point et non pas en plusieurs points.²⁹,³⁰ Il en déduit sa “Proposition I. Théorème” : si $W = \int_{x_A}^{x_Z} Z dx$ est extrémale, alors $\int_{x_M}^{x_N} Z dx$ est extrémale pour tout intervalle (M, N) dans (A, Z) .³¹

Pour trouver les conditions d’extrémalité, Euler découpe la courbe C en segments “infinitésimaux” de largeur dx à partir d’un point m d’abscisse M et considère les points voisins à droite d’abscisse N , O , P , ... et à gauche L , K , J , etc. En quelque sorte, il implémente numériquement les variations infinitésimales dans des variations *discrètes*, c’est-à-dire les différentielles par ce que l’on appelle aujourd’hui des “différences finies”. Si f est une fonction de x , il note $f = f(x_M)$, $f^i = f(x_N)$, $f^{ii} = f(x_O)$, etc. et $f_i = f(x_L)$, $f_{ii} = f(x_K)$, etc.³² et il donne des valeurs numériques aux symboles en écrivant $f^i = f + df$, $f^{ii} = f^i + df^i$, etc. Il en tire

27. Il y a souvent des ambiguïtés de notations dans le texte car les minuscules a , m , n , z sont tantôt des points de C et tantôt leurs abscisses.
28. Ut per maximi minimive formulam W , curva determinetur amz , quæ præ omnibus reliquis satisfaciat, formula W debet esse quantitas integralis indefinita, quæ, nisi data assumatur relation inter x & y , integrari nequeat.
29. Functio scilicet determinata ita est comparata, ut, pro quovis loco, a præsentibus valoribus litterarum x , y , p , q , &c. tantum pendent, neque valores earum anteriores in se compleatur.
30. Nous avons constamment insisté depuis le Vol. I sur cet aspect crucial de localité que l’on retrouve dans la notion de jet et qui permet une implémentation neuronale par des processeurs ponctuels.
31. Valor formulæ $\int Z dx$ pro abscissa AZ est aggregatum omnium valorum ejusdem formulæ, qui singulis abscissæ AZ portionibus respondent.
32. Nous changeons un tout petit peu les notations pour éviter les ambiguïtés.

(§§51-53) les formules mixtes (symboliques-numériques) que l'on utilise encore aujourd'hui³³ :

$$\begin{aligned} p &= \frac{dy}{dx} = \frac{Nn - Mm}{dx} = \frac{y^i - y}{dx}, \quad p^i = \frac{y^{ii} - y^i}{dx}, \quad p_i = \frac{y - y_i}{dx} \\ q &= \frac{dp}{dx} = \frac{p^i - p}{dx} = \frac{y^{ii} - y^i}{dx^2} - \frac{y^i - y}{dx^2} = \frac{y^{iii} - 2y^{ii} + y^i}{dx^2} \\ r &= \frac{dq}{dx} = \frac{y^{iiii} - 3y^{iii} + 3y^{ii} - y^i}{dx^3}, \text{ etc.} \end{aligned}$$

Dans le Corollaire IV (§54), il utilise le fait que $\int Zdx$ est une somme

$$\dots + Z_i dx + Z dx + Z^i dx + \dots$$

de dZ sur les intervalles infinitésimaux constituant AZ (où le Z de Zdx est maintenant la valeur de l'intégrande Z en x_M) et dans sa "Proposition IV. Théorème" (§58) il introduit le résultat fondamental généralisant aux fonctionnelles ce qui se passe pour les extrema des fonctions, à savoir que si la courbe C extrémalise l'intégrale $W = \int Zdx$ alors une perturbation ("mutatio") infinitésimale de C ("ab ista infinite parum discrepans") sur un segment d'arc infinitésimal ne change pas la valeur de W .³⁴ Aujourd'hui, cette condition d'extrémalité se noterait $\frac{\delta W[C]}{\delta C} = 0$. Et comme la position de M est quelconque et que Z est locale, cette condition de stationnarité doit être réalisée partout, ce qui implique une *équation différentielle* appelée depuis "équation d'Euler". Citons le texte latin :

"Posita igitur tali mutatione in curva, mutatio inde etiam in valore formulæ $\int Zdx$ orietur; quæ autem per demonstrationem erit evanescens. Atque hoc modo ex tali assumta mutatione orietur æquatio, quæ simul curvæ, quæsitæ naturam præbebit."

Suivant l'ordre de l'équation différentielle trouvée, il faut considérer les conditions aux limites permettant de déterminer les constantes d'intégration.

À partir de là, Euler va commencer à traiter des exemples, et d'abord des exemples relevant de la méthode "absolue". Le chapitre II commence par le calcul de la variation produite sur W par une petite variation nv ("particula infinite parva") de l'ordonnée Nn (i.e. de $y(x_N)$). L'abscisse x_N reste la même, mais les autres quantités y , p , q , etc. varient un peu. Cf. la figure 9.

Étant données les formules aux différences finies expliquées plus haut, la perturbation se propage aux points voisins de n jusqu'à un certain rang. Par exemple, le point de référence étant N d'abscisse x et d'ordonnée y , $Nn = y^i$ varie de nv et donc

33. Ambiguïté de notation : p , etc. dénotent à la fois la variable et sa valeur au point de référence M . Il serait mieux de noter cette dernière p_0 . Mais nous préférons rester proche des notations d'Euler.

34. [...] tum valor formula $\int Zdx$ pro utraque curva erit idem.

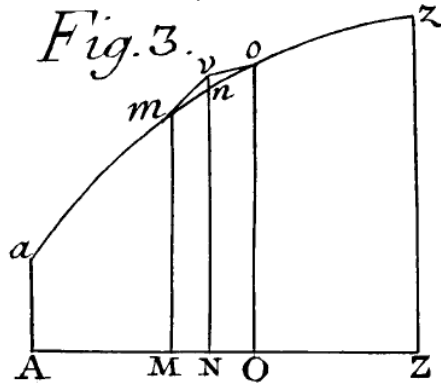


FIGURE 9. Une variation infinitésimale locale d'une courbe au voisinage d'un point n . (Fig. 3 du *Methodus* d'Euler).

y^i devient $y^i + nv$ (aujourd'hui, on écrirait plutôt $nv = \delta y^i$), les autres valeurs y^{ii} , y^{iii} , \dots , y , y_i , y_{ii} , etc. ne variant pas. Pour chaque variable y , p , q , etc. on regarde alors les formules aux différences finies contenant y^i , on les modifie en conséquence et on laisse le reste inchangé. Par exemple, comme la formule pour p fait intervenir deux valeurs voisines de y , p variera à deux positions voisines : $p = \frac{y^i - y}{dx}$ variera de $\frac{nv}{dx}$ et $p^i = \frac{y^{ii} - y^i}{dx}$ variera de $-\frac{nv}{dx}$. De même, q_i variera de $\frac{nv}{dx^2}$, q de $-2\frac{nv}{dx^2}$ et q^i de $\frac{nv}{dx^2}$; r_{ii} variera de $\frac{nv}{dx^3}$, r_i de $-3\frac{nv}{dx^3}$, r de $3\frac{nv}{dx^3}$, r^i de $-\frac{nv}{dx^3}$, etc. On calcule ainsi les *valeurs* des différentielles induites par la "mutatio" nv de y^i . Considérons par exemple l'expression $e = y^i \sqrt{1 + p^2}$. La différentielle formelle est $de = dy^i \sqrt{1 + p^2} + y^i \frac{p dp}{\sqrt{1 + p^2}}$. Mais l'évaluation donne $dy^i = nv$ et $dp = \frac{nv}{dx}$ et donc e varie de

$$de = nv \sqrt{1 + p^2} + y^i \frac{p nv}{dx \sqrt{1 + p^2}} .$$

Il suffit de calculer formellement. On voit que la rigueur n'est pas le souci majeur d'Euler.

Il faut bien comprendre qu'Euler joue sur deux types de différentielles, en accord avec le double statut des infinitésimales leibniziennes : d'une part des différentielles formelles "abstraites" comme dy et d'autre part des différentielles "concrètes" comme nv qui sont des différences finies δy très petites. Pour une courbe régulière C , il y a les relations comme $p = \frac{dy}{dx}$ pour la tangente en un point $(x, y(x))$. Mais si l'on modifie $y(x)$ de δy , alors p sera modifiée de $\frac{dy + \delta y}{dx} = p + \frac{\delta y}{dx}$, etc. Mais Euler n'utilise pas une notation comme δ et note uniformément toutes les différentielles.

Une fois calculées ces variations et donc évaluées les différentielles, il s'agit de formuler la condition de stationnarité. Pour ce faire, Euler va utiliser le fait que $W = \int Z dx = \int dZ$ et que, Z étant une fonction de plusieurs variables *formellement*

indépendantes, la différentielle dZ s'écrit en termes de dérivées partielles à partir de la formule universelle (en notation actuelle) :

$$\begin{aligned} dZ &= \frac{\partial Z}{\partial x} dx + \frac{\partial Z}{\partial y} dy + \frac{\partial Z}{\partial p} dp + \frac{\partial Z}{\partial q} dq + \dots \\ &= Mdx + Ndy + \dots \end{aligned}$$

Il s'agira alors *d'évaluer* ces variations et d'écrire la condition de stationnarité. Euler le fait en plusieurs étapes en appliquant une règle de calcul différentiel mixte *commutant* d et δ , règle que l'on pourrait écrire (il le fait à chaque étape mais sans introduire de formule générale) :

$$\begin{aligned} \delta(dZ) &= \delta(Zdx) = dx.\delta(Z) + Z\delta(dx) = dx.(M\delta x + N\delta y + \dots) \\ &= dx.(N\delta y + \dots) \text{ car } \delta(dx) = 0 \text{ et } \delta x = 0. \end{aligned}$$

Il commence par le cas le plus simple $Z = Z(x, y)$ et donc $dZ = Mdx + Ndy$ (valeur à l'abscisse M), $dZ^i = M^i dx + N^i dy^i$, etc. Comme seul y^i varie (de nv), seul dZ^i varie. Mais $dZ^i = Z^i dx$ et la variation de dZ^i est donc de $N^i.nv.dx$. Mais $N^i = N + dN$, dN est négligeable devant N (" dN praë N evanescit") et la contrainte de stationnarité s'écrit donc $N.nv.dx = 0$, soit, puisque $dx \neq 0$ et $nv \neq 0$ par hypothèse, $N = 0$ (i.e. $\frac{\partial Z}{\partial y} = 0$). C'est le premier exemple de l'équation d'Euler. Euler l'applique par exemple à $Z = y(ax - y^2)$ où $M = ay$ et $N = ax - 3y^2$ où la condition $N = 0$ définit une parabole.

Les choses commencent à devenir intéressantes lorsque Z dépend aussi de p et que donc $dZ = Mdx + Ndy + Pdp$. C'est la "Proposition III. Problème", §22. Les calculs sont faciles. Z (valeur en x_M) est fonction de x, y, p (valeurs en x_M), Z^i (valeur en x_N) est fonction de x^i, y^i, p^i (valeurs en x_M), etc. et les dZ, dZ^i , etc. contiennent des dy, dp, dy^i, dp^i , etc. Mais ces différentielles doivent être *évaluées* (Euler y insiste), et seuls $y^i, p = \frac{y^i - y}{dx}$ et $p^i = \frac{y^i - y^i}{dx}$ varient, respectivement de $nv, \frac{nv}{dx}$ et $-\frac{nv}{dx}$.³⁵ Donc Z (valeur en x_M) varie de $P.\frac{nv}{dx}$ et Z^i (valeur en x_N) varie de $N^i.nv - P^i.\frac{nv}{dx}$. L'intégrale $\int Zdx$ ne varie donc que sur les deux tranches infinitésimales Zdx et $Z^i dx$ et la somme des variations est $P.nv - P^i.nv + N^i.nv.dx$. Mais $P^i - P = dP$, d'où une variation (dN étant négligeable devant N) de $nv.(Ndx - dP)$. Mais cette condition doit être vérifiée pour toute petite perturbation nv et l'on obtient ainsi *l'équation*

35. Quamobrem hi termini debebunt differentiari, atque in eorum differentialibus loco dy^i, dp & dp^i scribi oportet valores supra indicatos $+nv$; $+\frac{nv}{dx}$ & $-\frac{nv}{dx}$.

différentielle d'Euler pour $Z(x, y, p)$ ³⁶

$$Ndx - dP = 0, \quad N = \frac{dP}{dx},$$

soit

$$\frac{\partial Z}{\partial y} - \frac{d}{dx} \left(\frac{\partial Z}{\partial p} \right) = 0$$

qui redonne évidemment $\frac{\partial Z}{\partial y} = 0$ si $\frac{\partial Z}{\partial p} = 0$. Si $P = \frac{\partial Z}{\partial p}$ contient p , alors cette équation différentielle est du deuxième ordre et il faut donc avoir deux conditions aux limites pour déterminer les deux constantes d'intégration.

Euler traite tout de suite des cas particuliers. Si par exemple $Z = Z(y, p)$ ne dépend pas de x , alors $M = 0$ et $N = \frac{dP}{dx}$ impliquent

$$dZ = \frac{dP}{dx} dy + P dp = dP \frac{dy}{dx} + P dp = p dP + P dp = d(Pp)$$

et donc $Z + c = Pp = p \frac{\partial Z}{\partial p}$ (où c est une constante d'intégration), ce qui est une équation différentielle en y et $p = y'$.

Si de plus, $Z = Z(p)$ ne dépend pas non plus de y et dépend donc seulement de p , alors l'équation $N = \frac{dP}{dx}$ devient $\frac{dP}{dx} = 0$, c'est-à-dire $P = \text{cste}$. Tel est le cas pour la longueur d'arc $\int \sqrt{1+p^2} dx$, cas pour lequel Euler s'amuse (Exemple II, §33) à démontrer que les géodésiques sont les droites. En effet, $P = \frac{p}{\sqrt{1+p^2}}$ et $P = \text{cste}$ implique $p = \text{cste} = a$ et donc $y = ax + b$ avec les deux constantes d'intégration a et b .

Dans l'Exemple III, §34, Euler traite un cas où Z ne dépend que de x et p et pas de y . Il s'agit de la courbe de descente la plus rapide donnée par la formule $W = \int \frac{\sqrt{1+p^2}}{\sqrt{x}} dx$. On a

$$dZ = Mdx + Pdp = -\frac{\sqrt{1+p^2}}{2x\sqrt{x}} dx + \frac{pdp}{\sqrt{x}\sqrt{1+p^2}}.$$

Comme $N = 0$, la condition se réduit à $\frac{dP}{dx} = 0$, c'est-à-dire $P = \frac{p}{\sqrt{x}\sqrt{1+p^2}} = \text{cste}$, soit $ap^2 = x + p^2x$, ou $p = \sqrt{\frac{x}{a-x}}$ et $y = \int \sqrt{\frac{x}{a-x}}$. Il s'agit d'une cycloïde.

Après plusieurs autres exemples, Euler considère dans sa "Proposition IV, Problème", §40, des fonctions $Z(x, y, p, q)$ dépendant aussi de q (i.e. de la dérivée seconde de y : jets d'ordre 2). Lorsque seulement y^i change (de nv), seulement y^i , p , p^i , q_i , q et q^i changent, et donc seulement $dZ_i = Z_i dx$, $dZ = Z dx$ (valeur en x_M) et $dZ^i = Z^i dx$ changent. Euler applique alors la règle que nous avons écrite

36. Quare cum formulæ $\int Z dx$ valor differentialis nihilo æqualis factus præbeat æquationem pro curva quæsitâ, hæc erit $0 = Ndx - dP$, vel $N - \frac{dP}{dx} = 0$, qua æquatione natura curvæ quæ sitæ exprimetur. Q.E.I.

$\delta(Zdx) = dx \cdot (N\delta y + P\delta p + Q\delta q)$ (car $\delta x = 0$), etc., avec les évaluations $\delta y^i = nv$, $\delta y = 0$, $\delta y_i = 0$, $\delta p^i = -\frac{nv}{dx}$, $\delta p = \frac{nv}{dx}$, $\delta p_i = 0$, $\delta q^i = \frac{nv}{dx^2}$, $\delta q = -2\frac{nv}{dx^2}$, $\delta q_i = \frac{nv}{dx^2}$. D'où la condition de stationnarité

$$nv \cdot dx \cdot \left(N^i - \frac{P^i}{dx} + \frac{P}{dx} + \frac{Q^i}{dx^2} - \frac{2Q}{dx^2} + \frac{Q_i}{dx^2} \right)$$

Mais $-\frac{P^i}{dx} + \frac{P}{dx} = -\frac{dP}{dx}$ et $\frac{Q^i}{dx^2} - \frac{2Q}{dx^2} + \frac{Q_i}{dx^2} = \frac{d^2Q_i}{dx^2}$. Et $d^2Q_i = d^2Q$ et on obtient donc en définitive l'équation d'Euler

$$N - \frac{dP}{dx} + \frac{d^2Q}{dx^2} = 0$$

soit

$$\frac{\partial Z}{\partial y} - \frac{d}{dx} \left(\frac{\partial Z}{\partial p} \right) + \frac{d^2}{dx^2} \left(\frac{\partial Z}{\partial q} \right) = 0 .$$

Dans son Exemple II (§51), Euler traite par cette méthode un exemple faisant intervenir l'évolute d'une courbe (le lieu de ses rayons de courbure). Si Am est un arc de C (A sur l'axe des x et m d'abscisse M) et AR son évolute, R est par définition le point sur la normale à C en m tel que mR soit le rayon de courbure. Euler pose $AM = x$, $Mm = y$, $mR = -\frac{(1+p^2)^{\frac{3}{2}}}{q}$ (voir plus haut) et il cherche à minimiser la surface AmR . La formule géométrique donnant cette aire est $\int \frac{1}{2}mRdw$ où dw est l'élément d'arc et donc la "formule" à minimiser est $W = \int Zdx = \int \frac{(1+p^2)^2}{q} dx$. On a, puisque $M = N = 0$,

$$dZ = \frac{4(1+p^2)pdp}{q} - \frac{(1+p^2)^2 dq}{q^2} = Pdp + Qdq .$$

La condition de stationnarité étant $\frac{dP}{dx} - \frac{d^2Q}{dx^2} = 0$ puisque $N = 0$, la solution est $P - \frac{dQ}{dx} = \text{cste} = C$, C étant une constante d'intégration. Et Euler calcule ... Si l'on multiplie cette équation par $dp = qdx$, on obtient $Pdp - qdQ = Cdp$. Comme $dZ = Pdp + Qdq$, on a $Cdp + qdQ = Pdp = dZ - Qdq$, soit $Cdp + d(qQ) = dZ$ et donc $Z = Cp + Qq + D$, D étant une seconde constante d'intégration. On remplace alors Z et Q par leurs expressions et on obtient l'équation

$$2(1+p^2)^2 = Dq + Cpq$$

soit, puisque $q = \frac{dp}{dx}$ et $p = \frac{dy}{dx}$ une équation différentielle reliant x à p mais en faisant de x une fonction $x(p)$ de p :

$$2dx = \frac{(D + Cp) dp}{(1+p^2)^2} .$$

Après l'intégration et une redéfinition des constantes on obtient pour x l'expression

$$x(p) = \frac{a + bp + cp^2}{1 + p^2} + b \operatorname{Arctan}(p)$$

et par conséquent, puisque, par intégration par parties, $y = \int p dx = px - \int x dp$:

$$y(p) = \frac{f + (c - a)p + (b + f)p^2}{1 + p^2} + (c - a) \operatorname{Arctan}(p) .$$

Cette équation *paramétrique* de la courbe C dépend de quatre constantes d'intégration et il faut donc quatre conditions aux limites pour déterminer une solution. Ces constantes d'intégration rendent difficile la reconnaissance de l'espèce géométrique de la courbe. Pour surmonter cette difficulté, Euler manipule de façon ingénieuse les équations. D'abord, il élimine $\operatorname{Arctan}(p)$ pour trouver une équation unique reliant x , y et p . En renommant les constantes, il obtient l'équation

$$\alpha x - \beta y = \frac{\beta^2 - \alpha^2 + 2\alpha\beta p}{1 + p^2} .$$

Ensuite, il fait un astucieux changement d'axes pour soustraire β^2 (i.e. $\beta^2 \frac{1+p^2}{1+p^2}$, ce qui transforme le numérateur en $-\beta^2 p^2 - \alpha^2 + 2\alpha\beta p = -(\beta p - \alpha)^2$. D'où

$$\sqrt{\beta y - \alpha x} = \frac{\beta p - \alpha}{\sqrt{1 + p^2}} = \frac{\beta dy - \alpha dx}{dx \sqrt{1 + p^2}} .$$

Mais $dx \sqrt{1 + p^2}$ est l'élément de longueur d'arc dw et donc, puisque

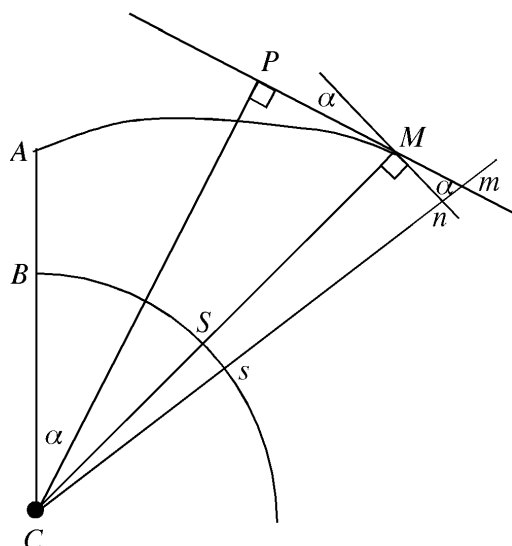
$$dw = \frac{\beta dy - \alpha dx}{\sqrt{\beta y - \alpha x}} = 2d\left(\sqrt{\beta y - \alpha x}\right) ,$$

$w = 2\sqrt{\beta y - \alpha x}$ ce qui caractérise une *cycloïde*.

Cet exemple remarquable montre toutes les techniques, changements de variables, fonctions inverses, élimination, que l'on peut utiliser pour intégrer les équations de la condition de stationnarité.

Dans son Exemple III (§52), Euler résout le problème d'extrémiser les fonctionnelles $\int \rho^n dw$ où dw est l'élément d'arc et ρ le rayon de courbure. Elles correspondent aux intégrandes $Z = \frac{(1+p^2)^{\frac{(3n+1)}{2}}}{q^n}$. Par des méthodes analogues à celles de l'exemple précédent, Euler trouve l'équation paramétrique qui, dans le cas $n = \frac{1}{2}$ qui est en résonance avec les modèles que nous développons, s'écrit

$$\begin{aligned} x &= \int \frac{a}{(1+p^2)^{\frac{5}{2}}} dp = \frac{ap(1 + \frac{2}{3}p^2)}{(1+p^2)^{\frac{3}{2}}} \\ y &= \int \frac{ap}{(1+p^2)^{\frac{5}{2}}} dp = -\frac{a}{3} \frac{1}{(1+p^2)^{\frac{3}{2}}} . \end{aligned}$$

FIGURE 10. L'exemple I d'Euler au §8 du chapitre IV du *Methodus*.

On en tire $(1 + p^2)^{\frac{3}{2}} = -\frac{a}{3y}$, donc $p^2 = \sqrt[3]{\frac{a^2}{9y^2}} - 1$ et ensuite l'équation

$$x = - \left(2\sqrt[3]{\frac{a^2 y}{9}} + y \right) \sqrt{\sqrt[3]{\frac{a^2}{9y^2}} - 1} .$$

Dans le Chapitre III, Euler aborde des problèmes plus compliqués où l'intégrande Z n'est plus une fonction locale des variables x, y, p, q , etc. et contient elle-même des intégrales. Nous n'en parlerons pas ici.

Dans le chapitre IV il traite un certain nombre d'applications du *Methodus* à des "questions variées". Son Exemple I, §8, est "splendide" ("luculentus"). Il redémontre le fait que les géodésiques du plan sont des droites mais en utilisant des coordonnées polaires.³⁷ La courbe AM cherchée n'est plus exprimée en coordonnées cartésiennes mais au moyen du cercle unité de centre C , le point A étant sur le rayon CB et le point M sur le rayon CS (cf. Figure 10).

Il s'agit alors de trouver la ligne AM la plus courte. Euler prend comme coordonnées la longueur $BS = x$ et $CM = y$, autrement dit l'angle $\theta = x$ en radians et le rayon vecteur $\rho = y$, et veut exprimer en leurs termes la condition de longueur minimale. Utilisons les notations d'aujourd'hui : θ est la variable de base et on cherche

37. Interim tamen hanc quæstionem secundum præcepta data resolvi conveniet, ut consensus Methodicum veritate luculentius perspiciatur.

une fonction $\rho(\theta)$. Soit $d\theta$ une petite variation de θ , Sur le cercle unité, S se déplace en un point très voisin s et, sur le cercle de rayon $CM = \rho$, M se déplace en un point très voisin n . Comme les triangles CSs et CMm sont semblables, on a $\frac{1}{d\theta} = \frac{\rho}{Mn}$ et donc $Mn = \rho d\theta$. Soit $Mm = dw$ l'élément d'arc sur la courbe que l'on cherche. Mm correspond à $d\rho = \rho d\theta$ et $Mn = \rho d\theta$ et donc $dw = \sqrt{\rho^2 + p^2} d\theta$ et la longueur à minimiser est $AM = \int \sqrt{\rho^2 + p^2} d\theta = \int Z(\rho, p) d\theta$. Pour les dérivées partielles de Z , on a $M = 0$, $N = \frac{\rho}{\sqrt{\rho^2 + p^2}}$, $P = \frac{p}{\sqrt{\rho^2 + p^2}}$. La condition de stationnarité $N - \frac{dP}{d\theta} = 0$, soit, en multipliant par $d\rho = \rho d\theta$ si $p \neq 0$,³⁸ $Nd\rho - \rho dP = 0$, donne donc, puisque $dZ = Nd\rho + Pdp$,

$$dZ = \rho dP + Pdp = d(P\rho)$$

(nous avons déjà rencontré ce type de calcul) soit $Z + C = P\rho$ où C est une constante d'intégration. On en tire que $\frac{\rho^2}{\sqrt{\rho^2 + p^2}}$ est constante. Mais la géométrie de la situation (cf. figure 10) montre que $\frac{\rho^2}{\sqrt{\rho^2 + p^2}}$ est la perpendiculaire CP du centre C du cercle à la tangente à la courbe cherchée en M . En effet, la relation $\frac{Mm}{Mn} = \frac{MC}{CP}$ qui montre que les triangles CPM et Mnm sont semblables est bien satisfaite : $\frac{\sqrt{\rho^2 + p^2} d\theta}{\rho d\theta} = \frac{\rho}{\left(\frac{\rho^2}{\sqrt{\rho^2 + p^2}}\right)}$.

Donc CP est constante et, comme le cas dégénéré du cercle (de rayon nul) est écarté, CP est fixe et la courbe cherchée est une droite.

Dans le Chapitre V, Euler aborde la méthode "relative" applicable aux cas où l'on cherche une solution extrémale seulement parmi les courbes satisfaisant certaines contraintes.³⁹ C'est le cas qui nous intéresse au premier chef. On y voit apparaître le premier exemple de "multiplicateur". Des exemples typiques sont connus depuis l'Antiquité : ce sont les célèbres problèmes *isopérimétriques* où la longueur de la courbe est fixée.⁴⁰ Mais Euler en considère d'autres où les propriétés communes s'expriment par une fonction "indéterminée" qui ne concerne pas un unique élément de la courbe mais la courbe dans son entier. Interviennent donc des conditions de constance de valeur = B qui ont la même forme $W = \int Z_W dx$ que la "formule" à

38. Si $p \equiv 0$, i.e. si $\rho = \text{cste}$, le cas est dégénéré : $P \equiv 0$, donc $N \equiv 0$, donc $\rho = \text{cste} = 0$. La courbe serait un cercle de rayon nul.

39. Nunc autem progredimur ad Methodum relativam, in qua unam lineam maximi minimive proprietate præditam determinare docebimus, non ex omnibus omnino lineis eidem abscissæ respondentibus, verum ex illis, innumerabilibus quidem, lineis curvis tantum, quibus una quædam proprietate proposita pluresve sint communes. (§2)

40. Hoc in genere inprimis celebre est *Problema Isoperimetricum*, initio hujus sæculi publice propositum, in quo, inter omnes curvas ejusdem longitudinis quæ quidem eidem abscissæ respondent, eam definiri oportebat, quæ contineret maximi minimive cujuspiam proprietatem.

optimiser, formule ou expression qu'Euler note V (et non plus W) de valeur optimale A (Z_V et Z_W sont nos notations).⁴¹

Dans les Corollaires II-V (§§4-7), Euler explique que l'on a donc deux formules, la formule de valeur A et la propriété commune de valeur B , B pouvant prendre une infinité de valeurs numériques mais chaque valeur autorisant une infinité de courbes. Autrement dit, chaque valeur b de B définit un "genre" de courbes et c'est à l'intérieur d'un genre que l'on applique la Méthode : il y a une infinité de "genera" ("infinite dentur genera", §7).⁴² Le problème est traité à la "Proposition II. Problème" (§14). Soit une courbe C allant de a à z au-dessus de l'intervalle AZ des abscisses. On veut optimiser $V = \int Z_V dx$ à B constant. Euler reprend sa méthode de découpage de C en tranches infinitésimales de largeur dx et fait varier un peu C dans l'intervalle mno , Nn variant de la "particula infinite parva" nv et Oo variant de ow . Les deux variations sont nécessaires car la condition de stationnarité "relative" doit comprendre deux équations, l'une pour l'optimisation, l'autre pour la condition. Avec une seule petite "mutatio" nv on en reste à la méthode "absolue".⁴³ Les deux équations seront du type $S.nv + T.ow = 0$ avec S et T des grandeurs liées à la courbe. En éliminant nv et ow on obtient alors l'équation de la courbe cherchée. La méthode est la même que dans le cas absolu mais les formules sont évidemment plus compliquées (Corollaire I, §15).

Dans le Corollaire V, §19, Euler explique alors que A et B étant exprimées par des intégrales de même type, l'optimisation et la condition sont *interchangeables* et "commutables".⁴⁴ Par exemple, un problème de type "surface optimale" avec une condition de type "longueur constante" sera équivalent à un problème de type "longueur optimale" avec une condition de type "surface constante". C'est la remarque qui conduit à la méthode des multiplicateurs.

Dans le §22, "Proposition III. Problème" Euler reprend l'évaluation des différentielles. Il commence par l'abscisse x de AI avec $Ii = y$ et, comme avant, note $y^i = Kk, \dots, y^{iv} = Nn, y^v = Oo$. Seuls y^{iv} et y^v varient, respectivement de nv et ow . Pour $p = \frac{y^i - y}{dx}$ (valeur en $x = x_I$), on aura⁴⁵ $\delta p = 0$, de même $\delta p^i = \delta p^{ii} = 0$.

41. Omnino igitur pari modo erit comparata quo ipsa maximi minimive formula, seu expressio.

42. Il y a donc plusieurs niveaux : celui des intégrandes Z , celui des intégrales indéfinies $\int Z dx$, celui de leurs valeurs numériques entre des bornes sur une courbe, celui de leurs valeurs sur des courbes particulières.

43. [...] quoniam enim hoc modo tota mutatio unica conditione determinatur, per eam efficitur nequit, ut tam proprietas communis B in ipsam curvam & immutatem æqualiter competat, quam maximi minimive expressio A .

44. In hac itaque operatione, ambæ expressiones A & B omnino pariter tractantur. [...] Ex quo perspicuum est, eandem solutionem prodire debere, si expressiones A & B inter se commutentur.

45. Pour éviter trop d'ambiguïtés dans les expressions différentielles mixtes symboliques-numériques nous utilisons la notation δ qu'Euler n'utilise pas.

Mais en revanche, comme $p^{iii} = \frac{y^{iv}-y^{iii}}{dx}$, $\delta p^{iii} = \frac{nv}{dx}$ et de même $\delta p^{iv} = \frac{ow}{dx} - \frac{nv}{dx}$, et $\delta p^v = -\frac{ow}{dx}$, etc. Pour q , on a les variations $\delta q^{ii} = \frac{nv}{dx^2}$, $\delta q^{iii} = -2\frac{nv}{dx^2} + \frac{ow}{dx^2}$, $\delta q^{iv} = \frac{nv}{dx^2} - 2\frac{ow}{dx^2}$, $\delta q^v = \frac{ow}{dx^2}$, etc. Ces formules sont universelles et sont régies par une combinatoire. Euler remarque par exemple que le $-2\frac{nv}{dx^2}$ de δq^{iii} devient le $-2\frac{ow}{dx^2}$ de δq^{iv} . Quand on évalue les variations de A et B on obtiendra donc des expressions du type $nv.I + ow.K$ avec $K = I^i = I + dI$ (il y a une confusion des notations avec les abscisses I, K).

La “Proposition IV. Problème”, §27, est cruciale. On veut qu’une intégrale W soit constante de valeur B et qu’une autre intégrale V soit extrémale de valeur A . La “mutatio” de Nn et Oo par une “particula infinite parva” nv et ow va entraîner des variations différentielles dA et dB des valeurs A et B des intégrales V et W : $\delta V = nv.dA + ow.dA^i$ et $\delta W = nv.dB + ow.dB^i$. Les *deux* doivent s’annuler. Mais alors, remarque Euler, on peut éliminer nv et ow . On tire nv de la seconde équation, on le reporte dans la première et comme l’annulation doit être valide pour tout ow , on en déduit $dB^i dA = dB dA^i$. Mais $dA^i = d(A + dA) = dA + d^2A$ et de même $dB^i = dB + d^2B$. On en tire l’égalité symbolique des dérivées logarithmiques $\frac{d^2A}{dA} = \frac{d^2B}{dB}$, puis par intégration $\log(dA) = \log(dB) + \log(C)$ (C est la constante d’intégration), soit $dA - CdB = 0$. On a ainsi ramené le problème “relatif” à un problème “absolu” portant sur l’expression $V - CW$. Cela rend évident la “commutabilité” des opérations d’optimisation et de valeur constante. La constante C est la première apparition de ce qu’on appelle maintenant un *multiplicateur de Lagrange*. L’idée est bien expliquée dans le Corollaire III, §30, dont nous donnons le texte latin :

“Proposita ergo proprietate communi W , & maximi minimive expressione V , utriusque expressionis valorem differentialem ex his præceptis quæri oportet : quibus inventis, & per constantes arbitrarias multiplicatis, eorum aggregatum nihilo æquale positum dabit æquationem pro curva quæsitâ.”

L’Exemple I, §40, résout le cas élémentaire $W = \int xy dx$ et $V = \int y^2 dx$. Comme

$$dZ_W = d(xy) = ydx + xdy = Mdx + Ndy,$$

on a $N_W = x$ et l’accroissement sera de $N_W.dx.nv = nv.dx.x$. Par ailleurs $dZ_V = d(y^2) = 2ydy$ et $N_V = 2y$, d’où un accroissement de $nv.dx.2y$. La solution est donc une combinaison linéaire $\alpha x + 2\beta y = 0$ ou $y = Cx$ (droite de pente C passant par l’origine). Pour connaître C on intègre $W = \int_0^a Cx^2 dx = C\frac{a^3}{3}$ et comme on veut $W = B$, on doit prendre $C = \frac{3B}{a^3}$.

Dans l’Exemple II, §41, Euler reprend le célèbre problème *isopérimétrique* de l’Antiquité. On veut que la longueur de la courbe C au-dessus de AZ , i.e. $W = \int_{x_A}^{x_Z} \sqrt{1+p^2} dx$, soit donnée et que la surface $AazZ$, i.e. $V = \int_{x_A}^{x_Z} y dx$ soit optimale. Les variations de W et V sont respectivement de $-nv.d\frac{p}{\sqrt{1+p^2}}$ et de $nv.dx$, d’où la

condition $dx = b \cdot d \frac{p}{\sqrt{1+p^2}}$ (b constante). En intégrant, on obtient $x + c = b \cdot \frac{p}{\sqrt{1+p^2}}$, soit

$$p = \frac{dy}{dx} = \frac{x + c}{\sqrt{b^2 - (x + c)^2}} .$$

L'intégration donne $y = f \pm \sqrt{b^2 - (x + c)^2}$, soit

$$b^2 = (y - f)^2 + (x + c)^2 .$$

Il s'agit, comme on le sait depuis l'Antiquité, de l'équation d'un cercle. L'arc de cercle concave donne le maximum, l'arc de cercle convexe donne le minimum.

Dans la "Proposition V. Problème", §52, Euler traite des situations beaucoup plus techniques où la condition est qu'une intégrale $\Pi = \int [Z] dx$ soit constante (par exemple la longueur de la courbe) tandis que la formule à optimiser $\int Z dx$ a une intégrande Z fonction non seulement de x, y, p, q , etc. mais également de Π . Mais nous ne commenterons pas ici ces avancées remarquables.

9. Mécanique et Lagrangiens

9.1. D'Euler à Lagrange

Inventées par Euler et approfondies par Lagrange, les équations d'Euler-Lagrange (E-L) pour le calcul des variations ont été une innovation extraordinaire permettant de complètement repenser la mécanique newtonienne en termes de *principes de moindre action*. Elles ont introduit une double façon de penser les phénomènes mécaniques, soit de façon "causaliste" à partir de forces et de l'équation de Newton $f = m\gamma$ (f = force appliquée à un point matériel, m = masse du point matériel, $\gamma = \ddot{x}(t)$ = accélération le long de la trajectoire du point), soit de façon "finaliste" à partir d'un principe de minimisation d'une action, comme en optique où les rayons lumineux minimisent un "chemin optique".

Pour le leibnizien Euler, elles étaient un exemple majeur du calcul des variations développé dans le *Methodus* des maxima et minima et, au-delà des calculs, exprimaient un principe métaphysique. Comme il l'affirmait dans le premier paragraphe de son étude [174] sur les elasticæ que nous allons présenter dans la section suivante 10 :

"Tous les plus grands Géomètres ont depuis longtemps reconnu que la Méthode présentée dans ce Livre est non seulement extrêmement utile en Analyse, mais qu'elle aide également beaucoup à la solution de problèmes physiques. Dans la mesure où la construction de l'univers est absolument parfaite et résulte du travail d'un Créateur très sage, rien ne peut advenir dans l'univers sans qu'une relation de maximum ou de minimum n'apparaisse clairement.

Il est donc absolument certain que chaque effet dans le Monde peut être expliqué de façon aussi satisfaisante par des causes finales, à l'aide de la méthode des maxima et minima, que par les seules causes efficientes.”⁴⁶

Les équations d'E-L sont si fondamentales et dominent la physique de façon si hégémonique, jusqu'aux développements les plus sophistiqués de la théorie quantique des champs, que la littérature à leur sujet est immense. Toute introduction à la Mécanique (cours universitaire, encyclopédie, ouvrage de vulgarisation) les expliquent en détail et cela justifie que nous ne les présentions que de façon très succincte. La *Mécanique analytique* de Lagrange lui-même (1788, [311]) est évidemment la référence absolue. Citons aussi le Colloque du bicentenaire [312] en 1988 à la Fondation Hugot du Collège de France ainsi que les articles [191], [192], [193] de Craig Fraser.

9.2. Équations d'E-L et relevées legendriennes

Aujourd'hui, les équations d'E-L se démontrent scolairement en considérant un lagrangien $L(t, x, p)$ où x est supposée être une fonction du temps $x(t)$ décrivant une trajectoire dans un espace de configurations M ⁴⁷ et où, dans ce cas, p est une dérivée “cachée” astreinte dans son interprétation à devenir la dérivée $p(t) = \dot{x}(t)$. Comme $\dot{x}(t)$ est un vecteur tangent à M , cela signifie que le lagrangien L est en fait une fonction (différentiable) $L(t, x, p)$ définie sur le fibré tangent TM indépendamment de toute trajectoire et que la structure de fibration $\pi : TM \rightarrow M$ est essentielle à la structure des équations d'E-L. Mais L devient une fonction composée de t le long des trajectoires $(x(t), p(t))$. Les équations d'E-L reposent sur les deux façons de dériver L : soit comme fonction de x et p , ce qui peut faire intervenir $\frac{\partial L}{\partial x}$ et $\frac{\partial L}{\partial p}$, soit comme fonction composée de t , ce qui peut faire intervenir des termes comme $\frac{\partial L}{\partial x} \frac{dx(t)}{dt}$, $\frac{\partial L}{\partial p} \frac{dp(t)}{dt}$, $\frac{d}{dt} \left(\frac{\partial L}{\partial x} \right)$ ou $\frac{d}{dt} \left(\frac{\partial L}{\partial p} \right)$. Nous retrouvons la grande perspective d'Euler commentée à la section précédente.

On veut trouver les trajectoires γ entre deux points donnés $x(t_0) = x_0$ et $x(t_1) = x_1$ optimisant une fonctionnelle

$$S(\gamma) := \int_{t_0}^{t_1} L(t, x(t), \dot{x}(t)) dt ,$$

46. Jam pridem summi quique Geometrae agnoverunt, Methodi in hoc Libro traditae non solum maximum esse usum in ipsa Analysis, sed etiam eam ad resolutionem Problematum physicorum amplissimum subsidium affere. Cum enim Mundi universi fabrica sit perfectissima, atque a Creatore sapientissimo absoluta, nihil omnino in mundo contingit, in quo non maximi minimive ration quaequam eluceat : quamobrem dubium prorsus est nullum, quin omnes Mundi effectus ex causis finalibus, ope Methodi maximorum & minimorum aequae feliciter determinari queant, atque ex ipsis causis efficientibus.

47. En mécanique (surtout hamiltonienne), une notation traditionnelle pour les positions dans l'espace de configuration est q plutôt que x .

c'est-à-dire l'intégrale de L le long de la trajectoire. Cette intégrale s'appelle l'*action*. Supposons que la variable x soit de dimension 1. On se situe donc bien dans l'espace des 1-jets de coordonnées (t, x, p) ⁴⁸ avec sa structure de contact canonique $\omega = dx - p dt$. Le lagrangien L est une fonction sur l'espace total des jets "précédant" les trajectoires $x(t)$ et la fonctionnelle d'action S est son intégrale le long des relevées legendriennes (des 1-jets) des trajectoires. On voit ainsi admirablement fonctionner la différence entre les dérivées partielles de L par rapport à ses variables (t, x, p) considérées comme indépendantes et les dérivées temporelles le long des 1-jets où x devient une fonction de t et p une fonction de t et de $x(t)$. C'est alors tout simplement la loi de dérivation des fonctions composées qui induit automatiquement les équations d'E-L.

Remarque. Ce fait suffit à comprendre pourquoi ce genre de formalisme variationnel est par essence adapté aux explications de bas niveau et bottom-up des contours illusoires modaux puisque ceux-ci sont précisément des relevées legendriennes satisfaisant une certaine contrainte d'optimisation. \square

Si $h(t)$ est une petite variation différentiable de $x(t)$ avec $h(t_0) = h(t_1) = 0$ (les extrémités de la trajectoire déformée $x(t) + h(t)$ restant fixes), on écrit la condition d'optimisation d'Euler

$$\left. \frac{\delta S(x + \varepsilon h)}{\delta \varepsilon} \right|_{\varepsilon=0} = 0 .$$

En supposant que les conditions nécessaires de différentiabilité et de commutation entre intégration et dérivation soient valides, on trouve

$$\int_{t_0}^{t_1} \left(\frac{\partial}{\partial x} L(t, x(t), \dot{x}(t)) h(t) + \frac{\partial}{\partial p} L(t, x(t), \dot{x}(t)) \dot{h}(t) \right) dt = 0 .$$

Mais une intégration par parties donne

$$\begin{aligned} & \int_{t_0}^{t_1} \frac{\partial}{\partial p} L(t, x(t), \dot{x}(t)) \dot{h}(t) dt \\ &= \left[\frac{\partial}{\partial p} L(t, x(t), \dot{x}(t)) h(t) \right]_{t_0}^{t_1} - \int_{t_0}^{t_1} \frac{d}{dt} \left(\frac{\partial}{\partial p} L(t, x(t), \dot{x}(t)) \right) h(t) dt \end{aligned}$$

48. En géométrie, les coordonnées sont traditionnellement (x, y, p) , y étant une fonction $y(x)$ (une courbe) et p la dérivée spatiale "cachée" qu'est la tangente $y'(x)$. En mécanique, les coordonnées sont traditionnellement (t, x, p) ou (t, q, p) , x ou q étant une fonction $x(t)$ ou $q(t)$ (une trajectoire) et p la dérivée temporelle "cachée" qu'est la vitesse $\dot{x}(t)$ ou $\dot{q}(t)$. Ces différences de notations sont héritées de l'histoire et nous en tiendrons compte lorsque le contexte les rendra pertinentes.

et comme $\left[\frac{\partial}{\partial p} L(t, x(t), \dot{x}(t)) h(t) \right]_{t_0}^{t_1} = 0$ puisque $h(t_0) = h(t_1) = 0$, la condition s'écrit

$$\int_{t_0}^{t_1} \left(\frac{\partial}{\partial x} L(t, x(t), \dot{x}(t)) - \frac{d}{dt} \left(\frac{\partial}{\partial p} L(t, x(t), \dot{x}(t)) \right) \right) h(t) dt = 0 .$$

Puisqu'elle doit être vérifiée pour toute perturbation h admissible, on obtient l'équation d'E-L :

$$\frac{\partial}{\partial x} L(t, x(t), \dot{x}(t)) - \frac{d}{dt} \left(\frac{\partial}{\partial p} L(t, x(t), \dot{x}(t)) \right) = 0 .$$

9.3. L'exemple newtonien

Pour comprendre pourquoi les équations d'E-L sont en mécanique l'équivalent de l'équation de Newton, le plus simple est de revenir à la trajectoire $x(t)$ d'un point matériel de masse m se déplaçant dans l'espace \mathbb{R}^3 sous l'action d'un champ de force dérivant d'un potentiel (différentiable avec éventuellement des singularités) $V(x)$. L'équation de Newton s'écrit dans ce cas $m\ddot{x}(t) = -\nabla V(x(t))$ (où ∇V est le gradient spatial de V) car la force est par définition $f = -\nabla V$. L'énergie cinétique du point matériel est $T = \frac{1}{2}m \|\dot{x}(t)\|^2$ ($\|\dot{x}(t)\|^2 = \sum_{i=1}^3 \dot{x}_i(t)^2$) et son énergie potentielle le long de sa trajectoire est $V(x(t))$. Introduisons alors le lagrangien $L = T - V$. Il est trivial de vérifier que l'équation de Newton équivaut au système d'équations d'E-L :

$$\frac{\partial L}{\partial x_i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) = 0 , \quad i = 1, 2, 3 .$$

En effet, $\frac{\partial L}{\partial x_i} = -\frac{\partial V}{\partial x_i}$, $\frac{\partial L}{\partial \dot{x}_i} = m\dot{x}_i(t)$ et $\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) = m\ddot{x}_i(t)$ et donc le système d'équations s'écrit $-\nabla V(x(t)) - m\ddot{x}(t) = 0$, soit $m\ddot{x}(t) = -\nabla V(x(t))$. Les $\frac{\partial L}{\partial x_i} = -(\nabla V)_i$ sont les forces, les $\frac{\partial L}{\partial \dot{x}_i} = m\dot{x}_i(t)$ sont les impulsions, les $\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right)$ sont les accélérations et on a bien réécrit l'équation de Newton.

Remarquons que si le potentiel $V(x) \equiv 0$, le mouvement est celui d'une particule *libre* dans l'espace et que le lagrangien se réduit à l'énergie cinétique $T = \frac{1}{2}m \|\dot{x}(t)\|^2$. Les trajectoires sont alors des mouvements rectilignes uniformes car les équations d'E-L donnent $\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_i} \right) = 0$, i.e. $\dot{x}_i = \text{cste}$. Ce sont également les géodésiques de

l'espace euclidien parce que la minimisation de la longueur $ds = \sqrt{\sum \dot{x}_i(t)^2} dt$ correspond au lagrangien $L = \sqrt{\sum \dot{x}_i(t)^2}$ et que les équations d'E-L sont

$$\frac{d}{dt} \left(\frac{\dot{x}_i}{\sqrt{\sum \dot{x}_i^2}} \right) = 0 .$$

Or $\frac{\dot{x}_i}{\sqrt{\sum \dot{x}_i^2}} = \text{cste}$ implique $\dot{x}_i = \text{cste}$. Ce lien entre géométrie (géodésiques) et mécanique (lagrangien = énergie cinétique) est fondamental.

Exactement la même situation se produit pour un lagrangien $L(q, \dot{q})$ défini sur le fibré tangent TM d'un espace de configuration M qui est une variété différentiable de dimension n dont les points q représentent les états instantanés d'un système à n degrés de liberté. On obtient n équations d'E-L qui, étant des équations du second ordre, ont des solutions dépendant de $2n$ constantes d'intégration, celles-ci étant définies par les $2n$ conditions au bord (q_0, q_1) ou (q_0, \dot{q}_0) .

Remarque. Le principe de moindre action a suscité depuis Leibniz de longs et profonds débats métaphysiques car on ne voyait pas comment la nature pouvait “choisir” de façon non pas causale mais “finaliste” un chemin global. La réponse est venue tardivement à travers l'optique et la compréhension du lien entre optique géométrique et optique ondulatoire. Il s'agit d'un sujet technique dont le lecteur trouvera une présentation pédagogique dans notre compilation [413]. La clé est le principe dit de la “phase stationnaire” dans les “intégrales oscillantes” : toutes les trajectoires possibles autres que celle minimisant la longueur optique (l'action) s'éliminent par des interférences destructives rendant leur probabilité nulle et seule celles minimisant l'action sont observables. De même, en mécanique, les trajectoires extrémales minimisant l'action sont les seules à être observables parce que les autres sont de probabilité nulle. Mais pour le comprendre il faut passer de la mécanique “géométrique” classique de Lagrange-Hamilton à la mécanique “ondulatoire”, c'est-à-dire à la mécanique quantique, ce qui dépasse notre propos. Nous reviendrons néanmoins plus en détail sur les intégrales oscillantes dans la section 9.2 du chapitre 16. □

Remarque. Nous avons insisté sur le fait que le lagrangien est une fonction qui *préexiste* aux trajectoires de points matériels. Il est défini sur le fibré TM , l'espace de configurations M étant considéré comme un cadre, comme une “background structure” (nous avons longuement expliqué ce concept à la section 1 du chapitre 4), et il agit dynamiquement et de façon *a priori* sur la matière distribuée dans cet espace. Cette différence de statut entre le lagrangien et les trajectoires qu'il régit est si fondamentale que, nous l'avons vu, la relativité générale intégrera les potentiels et les champs de force à la géométrie de l'espace, les traduira en termes de métriques

(pseudo)-riemanniennes et pourra alors interpréter les trajectoires comme de simples géodésiques dans ces espaces “courbes” (cf. chapitre 5, section 6.4).

Encore une fois, l’analogie est frappante avec les architectures fonctionnelles neuronales. L’architecture fonctionnelle préexiste aux perceptions ; les inputs sensoriels périphériques sont comme des conditions au bord ; ces inputs déclenchent des dynamiques de complétion modale conduisant à des percepts ; ces dynamiques sont régies par l’architecture fonctionnelle qui agit au moyen des connexions neuronales sur la propagation de potentiels d’action. La différence fondamentale est que, en neurogéométrie, les relations spatiales ne sont plus idéales mais implémentées dans les connexions neuronales, l’*idéalisme* géométrique se convertissant en *matérialisme* neuronal. Dans l’Introduction du Vol I nous avons beaucoup insisté sur cette analogie. On peut la résumer dans le tableau suivant :

Mécanique	Neurogéométrie
Espace “background” M et positions q .	Aire neuronale \mathbb{V} et méso-neurones.
Forces et potentiels structurant l’espace.	Architectures fonctionnelles de connexions.
Conditions aux bord $(q_0, q_1$ ou $q_0, \dot{q}_0)$.	Inputs (par exemple, un ensemble de (a_i, p_i)).
Trajectoire temporelle $q(t) \subset M$ interpolant entre les conditions au bord.	Complétion des inputs en états perceptifs de \mathbb{V} interpolant entre les inputs.
Les $q(t)$ sont dynamiquement régis par les forces et les potentiels.	Les percepts sont dynamiquement régis par les architectures fonctionnelles.

Conformément à cette analogie, on pourrait dire que la neurogéométrie est une sorte de “mécanique lagrangienne” des percepts. □

10. Les elasticæ : de Bernoulli et Euler à Kirchoff

Nous allons maintenant présenter la théorie des elasticæ qui est un splendide exemple de calcul des variations développé par Euler de façon magistrale. Comme nous l’avons vu à la section 3.4 du chapitre 8 ces elasticæ ont été utilisés par David Mumford en 1992 [381] pour modéliser les contours illusoires. Leur théorie hamiltonienne dans le cadre de la théorie du contrôle optimal a été entièrement décrite par Yuri Sachkov dans [476], [477], [480]. Nous y consacrerons l’ensemble du chapitre 13. Nous tendons donc un arc entre les origines du calcul variationnel et l’actualité neurogéométrique.

Les elasticæ forment une famille de courbes de formes très variées dont la classification complète a été donnée par Euler dans *Additamentum 1. De curvis elastica* de son *Methodus*. Une traduction anglaise [394] de l' *Additamentum 1* a été proposée en 1933 par W.A. Oldfather, Charles Ellis et Donald Brown. À titre d'introduction, on peut consulter les *Lectures on Elastic Curves and Rods* [504] de David Singer et le "technical report" historique [327] de Raph Levien, ainsi que les textes classiques de Clifford Truesdell [543] et [544] sur Leibniz, Bernouilli et Euler.

10.1. Le problème initial

La théorie de l'élasticité est ancienne et remonte au Moyen-Âge. Galilée s'en est occupé et Hooke en établit la première loi en 1678. Mais il faut attendre l'avènement du calcul différentiel pour que la théorie puisse commencer à se mettre en place en 1687 dans une correspondance entre Leibniz et Jacques Bernouilli. Bernouilli formule le problème mécanique en 1691 (et le publie en 1694 dans les *Acta Eruditorum*). On pourra consulter par exemple le texte de Raph Levien [327] pour une présentation.

Pour l'elastica dit "rectangulaire", la lame Γ est clampée horizontalement au point B d'un mur et courbée par la force F d'un poids suspendu à son autre extrémité A . La force F est telle que la tangente en A soit verticale (d'où le qualificatif de "rectangulaire"). Soit x l'abscisse du point courant M (dans un repère adapté). L'idée est que le moment Fx de la force F en M est équilibré par la courbure $\kappa(x)$ de la lame en M et que donc $\kappa(x)$ est proportionnelle à x , $\kappa(x) = \frac{2}{a^2}x$ (on verra pourquoi $\frac{2}{a^2}$). Mais l'on connaissait depuis Newton la formule de la courbure $\frac{d^2y}{dx ds}$ (ds la longueur d'arc) et on en dérivait immédiatement $\frac{dy}{ds} = \int_0^x \kappa(t) dt = Y(x) = \frac{x^2}{a^2}$ et donc

$$\frac{dy}{dx} = \frac{dy ds}{ds dx} = \frac{Y(x)}{\sqrt{1 - Y(x)^2}}$$

car $\frac{dx^2}{ds^2} = 1 - \frac{dy^2}{ds^2}$. D'où l'équation différentielle de l'elastica rectangulaire

$$\frac{dy}{dx} = \frac{x^2}{\sqrt{a^4 - x^4}}.$$

Comme le rappelle Levien, Huygens nota que cette équation n'était valable que pour l'elastica rectangulaire et la généralisa. Puis Bernouilli lui-même donna une formule plus générale. Mais c'est avec le texte d'Euler de 1744 que la théorie atteint un premier sommet.

10.2. L'explication variationnelle

C'est donc Daniel Bernouilli (le neveu de Jacques), "l'homme le plus illustre et le plus perspicace dans sa façon sublime d'étudier la nature", qui, dans une lettre

du 20 octobre 1742, proposa le problème des elasticæ à Euler. Le problème est de savoir quelle forme prend une lame élastique Γ de longueur donnée ℓ (droite au repos et de section et d'élasticité uniformes) lorsqu'on la plie et replie. L'explication causale trouvée par Bernoulli (ce qu'Euler appelle la méthode "directe") consiste à décomposer la lame en petits segments rigides articulés par des ressorts, à considérer les forces exercées sur les articulations, à calculer leur moment par rapport à une origine fixe et à appliquer les principes de la statique élémentaire pour exprimer l'équilibre des forces et en déduire la forme de la lame. Daniel Bernoulli obtiendra une équation qui sera retravaillée par de nombreux autres auteurs : Clebsch (1862), Thomson, Tait (1867), Kirchoff et E. et F. Cosserat (1909). Si \widehat{AB} est la courbe Γ cherchée (A et B sont deux points du plan) de point courant $M = (x, y)$, d'élément d'arc de longueur ds et de rayon de courbure R , ce qu'il s'agit de minimiser selon Daniel Bernoulli est l'intégrale (la fonctionnelle)

$$\int_{x_A}^{x_B} \frac{1}{R^2} ds = \int_{x_A}^{x_B} \kappa^2 ds ,$$

$\kappa = \frac{1}{R}$ étant la courbure.⁴⁹ Le problème est isopérimétrique puisque la longueur de Γ , $\ell = \int_{x_A}^{x_B} ds$ est fixée avec $\ell \geq \overline{AB}$ (la longueur du segment AB).

Pour Euler, nous l'avons vu, en plus des explications par la "méthode directe" des "causes efficientes", il y a toujours des explications par la "méthode indirecte" des "causes finales". C'est donc la méthode "indirecte" qu'il va utiliser. Comme la courbure intervient, les équations différentielles obtenues seront du deuxième ordre et il faudra quatre conditions aux limites pour déterminer les solutions, par exemple (A, θ_A) et (B, θ_B) où les θ sont les pentes. Fidèle à son *Methodus*, Euler introduit p et q avec $dy = p dx$ et $dp = q dx$, $ds = \sqrt{1 + p^2} dx$, $R = \frac{(1 + p^2)^{\frac{3}{2}}}{q}$ et il calcule ... Il faut extrémiser la "formule" (la fonctionnelle) $W = \int Z dx$ avec

$$Z = \frac{q^2}{(1 + p^2)^3} \sqrt{1 + p^2} = \frac{q^2}{(1 + p^2)^{\frac{5}{2}}} .$$

Comme Z ne dépend ni de x ni de y mais seulement de p et q , on a $dZ = P dp + Q dq$ avec

$$P = \frac{\partial Z}{\partial p} = -5 \frac{pq^2}{(1 + p^2)^{\frac{7}{2}}}, \quad Q = \frac{\partial Z}{\partial q} = \frac{2q}{(1 + p^2)^{\frac{5}{2}}} .$$

49. Si nous revenons à la formule de Mumford $E = \int_{\gamma} (\alpha \kappa^2 + \beta) ds$ de la section 3.4 du chapitre 8, on voit que si la longueur $\int_{\gamma} ds$ est fixée (contrainte isopérimétrique), alors le problème de Mumford devient celui de Bernoulli et Euler.

Les équations d'E-L pour l'optimisation ont pour terme de gauche

$$\frac{\partial Z}{\partial y} - \frac{d}{dx} \left(\frac{\partial Z}{\partial p} \right) + \frac{d^2}{dx^2} \left(\frac{\partial Z}{\partial q} \right) .$$

Pour la contrainte isopérimétrique (constance de la longueur ℓ de Γ), on a $\ell = \int_{x_A}^{x_B} L dx$ avec $L = \sqrt{1+p^2}$ et, puisque L ne dépend pas de y ,

$$\frac{\partial L}{\partial y} - \frac{d}{dx} \left(\frac{\partial L}{\partial p} \right) = - \frac{d}{dx} \left(\frac{\partial L}{\partial p} \right) = - \frac{d}{dx} \left(\frac{p}{\sqrt{1+p^2}} \right) .$$

D'après la méthode des multiplicateurs, la condition de stationnarité est donc

$$\alpha \frac{d}{dx} \left(\frac{p}{\sqrt{1+p^2}} \right) = \frac{dP}{dx} - \frac{d^2Q}{dx^2} .$$

qui est une différentielle exacte En intégrant on obtient par conséquent

$$\frac{\alpha p}{\sqrt{1+p^2}} + \beta = P - \frac{dQ}{dx} .$$

On multiplie alors par $dp = q dx$ en tenant compte de l'interprétation naturelle de la variable q , et l'on obtient

$$\frac{\alpha p dp}{\sqrt{1+p^2}} + \beta dp = \alpha d \left(\sqrt{1+p^2} \right) + \beta dp = P dp - q dQ .$$

Mais comme $dZ = P dp + Q dq$, $P dp = dZ - Q dq$ et l'on obtient donc en définitive

$$\alpha d \left(\sqrt{1+p^2} \right) + \beta dp = dZ - Q dq - q dQ = dZ - d(qQ)$$

qui est une différentielle exacte. On peut donc intégrer, ce qui donne

$$\alpha \sqrt{1+p^2} + \beta p + \gamma = Z - qQ ,$$

autrement dit

$$\alpha \sqrt{1+p^2} + \beta p + \gamma = - \frac{q^2}{(1+p^2)^{\frac{5}{2}}} .$$

Mais $q = \frac{dp}{dx}$ et, avec des constantes d'intégration convenables,

$$dx = \frac{dp}{(1+p^2)^{\frac{5}{4}} \sqrt{\alpha \sqrt{1+p^2} + \beta p + \gamma}} = \frac{dp}{(1+p^2)^{\frac{5}{4}} \sqrt{E}}$$

$$dy = p dx .$$

Ces intégrales n'étaient pas intégrables avec les fonctions connues à l'époque (ce sont des intégrales elliptiques) et Euler les ramène à des formes normales. Grand calculateur, il remarque subtilement que

$$d\left(\frac{2\sqrt{E}}{(1+p^2)^{\frac{1}{4}}}\right) = \frac{(\beta - \gamma p) dp}{(1+p^2)^{\frac{5}{4}} \sqrt{E}} = \beta dx - \gamma dy$$

ce qui est une différentielle exacte que l'on peut intégrer. D'où

$$\beta x - \gamma y + \delta = \frac{2\sqrt{\alpha\sqrt{1+p^2} + \beta p + \gamma}}{(1+p^2)^{\frac{1}{4}}}.$$

Euler simplifie ensuite cette équation en choisissant des coordonnées adaptées. Par translation, on peut éliminer δ . Par rotation et changement d'échelle, on peut éliminer γ . Il reste en définitive

$$\begin{aligned} 2\sqrt{\alpha\sqrt{1+p^2} + \beta p} &= \beta x (1+p^2)^{\frac{1}{4}} \\ 4\alpha\sqrt{1+p^2} + 4\beta p &= \beta^2 x^2 \sqrt{1+p^2} \\ 4\beta p &= (\beta^2 x^2 - 4\alpha) \sqrt{1+p^2}. \end{aligned}$$

Pour l'homogénéité, Euler pose $\alpha = \frac{4m}{a^2}$, $\beta = \frac{4n}{a^2}$, d'où, en élevant l'équation au carré,

$$p = \frac{n^2 x^2 - ma^2}{\sqrt{n^2 a^4 - (n^2 x^2 - ma^2)^2}}.$$

Et enfin, dernière étape, en refaisant une translation, et en redéfinissant les constantes, Euler arrive à l'intégrale elliptique des *elasticæ* (avec de nouvelles constantes α, β, γ) :

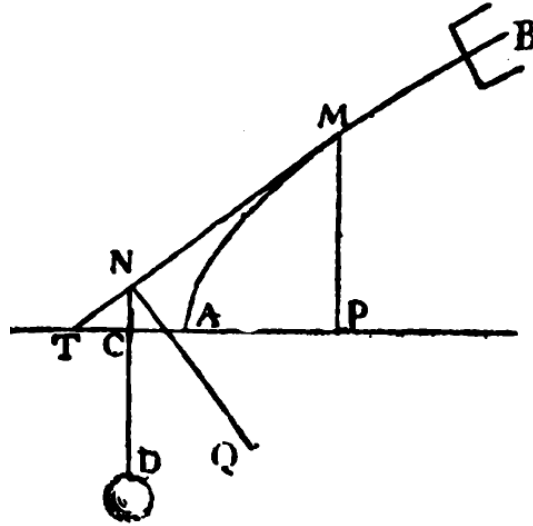
$$dy = \frac{\alpha + \beta x + \gamma x^2}{\sqrt{a^4 - (\alpha + \beta x + \gamma x^2)^2}} dx = \frac{P(x)}{\sqrt{a^4 - P(x)^2}} dx \quad (\text{EL } 1)$$

où $P(x)$ est un polynôme du second degré quelconque.⁵⁰

10.3. L'explication mécanique causale

Euler revient alors sur l'explication mécanique causale fournie par Bernoulli. Supposons, comme indiqué sur la figure 3 d'Euler (figure 11) que la lame Γ est clampée en son extrémité droite B avec une pente θ_B et soit M le point courant de Γ .

50. $dy = \frac{x^2}{\sqrt{a^4 - x^4}} dx$ de l'elastica "rectangulaire" correspond à $P(x) = x^2$, i.e. $\alpha = \beta = 0$ et $\gamma = 1$.

FIGURE 11. La figure 3 de l'*Additamentum 1* du *Methodus* d'Euler [174].

Soit N l'extrémité gauche de Γ lorsqu'on ne la plie pas. On exerce une force de pesanteur F en N , la lame se plie et N se déplace en A . On prend un axe des abscisses passant par A , M étant de projection P à droite de A , d'abscisse $AP = x > 0$ par rapport à A , et N de projection C à gauche de A , d'abscisse $CA = c > 0$ par rapport à A . Soit R le rayon de courbure de la lame en M . La force élastique F_{el} de la lame est proportionnelle à la courbure κ et donc inversement proportionnelle à R , $F_{el} = \frac{Ek^2}{R}$ (E et k sont des paramètres de la lame, Ek^2 étant son élasticité absolue). F_{el} doit être en équilibre avec le moment de F qui est $F(x+c)$. En développant $\frac{Ek^2}{R} = F(x+c)$, en multipliant par dx et en intégrant, Euler obtient $F\left(\frac{1}{2}x^2 + cx + f\right) = -Ek^2 \frac{dy}{ds}$ car $R = -\frac{ds^3}{dx dy}$. Il en tire l'équation

$$dy = -\frac{F\left(\frac{1}{2}x^2 + cx + f\right)}{\sqrt{E^2k^4 - F^2\left(\frac{1}{2}x^2 + cx + f\right)^2}}dx \quad (\text{EL } 2)$$

qu'il compare à l'équation précédente EL 1. Il en tire l'expression de la force et des constantes c et f : $F = -2\frac{Ek^2}{a^2}\gamma$, $c = \frac{\beta}{2\gamma}$, $f = \frac{\alpha}{2\gamma}$.

10.4. La classification (*enumeratio*) des solutions

La famille des elasticæ est intéressante à étudier en tant que telle. Comme le note Raph Levien ([327], p.1),

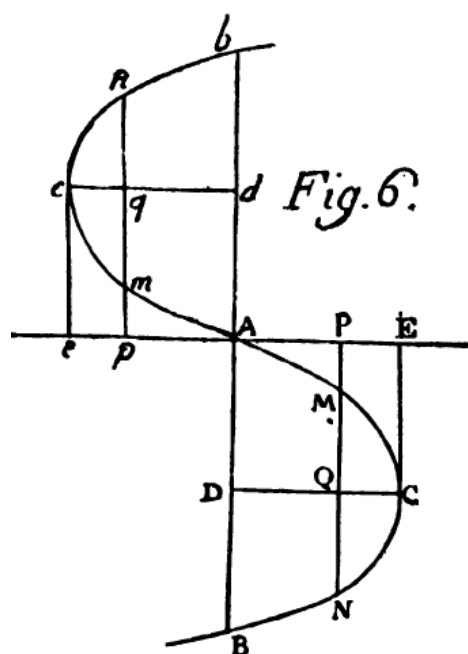


FIGURE 12. La figure 6 de l'*Additamentum 1* du *Methodus* d'Euler [174].

“It is a beautiful family of curves based on beautiful mathematics and a rich and fascinating history”.

Et un aspect particulièrement remarquable de l'*Additamentum* est la *classification* des solutions possibles en fonction de domaines bien définis des constantes d'intégration. L'équation différentielle (EL 1)

$$dy = \frac{\alpha + \beta x + \gamma x^2}{\sqrt{a^4 - (\alpha + \beta x + \gamma x^2)^2}} dx$$

dépend de trois paramètres. Euler déplace l'origine de $\frac{\beta}{2\gamma}$, ce qui permet d'annuler le nouveau β . Il remplace ensuite a^2 par $\frac{a^2}{\gamma}$, ce qui revient à faire $\gamma = 1$ et il remplace α par $a^2 - c^2$. D'où l'équation simplifiée

$$dy = \frac{(a^2 - c^2) + x^2}{\sqrt{(c^2 - x^2)(2a^2 - (c^2 - x^2))}} dx . \quad (\text{EL } 3)$$

Reprenons alors la figure 6 de l'*Additamentum* (figure 12).⁵¹

51. Les notations sont différentes de celles de la figure 3 (figure 11).

La courbe est symétrique par rapport à l'origine A et Euler note (C, c) , etc. les couples de points symétriques.⁵² Il explique que la force $2\frac{Ek^2}{a^2}$ ($\gamma = 1$) exercée en A d'abscisse $x = 0$ est perpendiculaire à l'axe AP car $\beta = 0$. La tangente en A est $\frac{dy}{dx} = \frac{(a^2 - c^2)}{c\sqrt{(2a^2 - c^2)}}$ (si on a choisi $c \geq 0$) qui est la tangente de l'angle $\theta = \widehat{PAM}$. On a

$$\sin(\theta) = \frac{a^2 - c^2}{a^2} \text{ et } \cos(\theta) = \frac{c\sqrt{(2a^2 - c^2)}}{a^2}.$$

1. Lorsque $a^2 \rightarrow \infty$, la force exercée $2\frac{Ek^2}{a^2}$ s'annule, il n'y a plus de courbure et la lame est droite et orthogonale à l'axe AP car $\theta \rightarrow \frac{\pi}{2}$ ($\sin(\theta) \rightarrow 1$ et $\cos(\theta) \rightarrow 0$).

2. Lorsque a^2 décroît, θ décroît également. Au début, $a^2 - c^2 > 0$ (et à fortiori $2a^2 - c^2 > 0$) et $\tan(\theta) > 0$. Comme $2a^2 - (c^2 - x^2) \geq 0$, il faut $x^2 < c^2$, i.e. $x \in]-c, c[$, pour que l'expression sous la $\sqrt{\quad}$ soit > 0 . Mais pour $x = \pm c$, $\frac{dy}{dx}$ diverge et l'on obtient les points C et c à tangente verticale de la figure 12. Pour $x > c$ et $x < -c$ la courbe est imaginaire. Elle rebrousse chemin en C , nous allons voir comment.

3. Lorsque a^2 continue à décroître, les ondulations s'accroissent et on arrive au cas limite $a^2 - c^2 = 0$ où la tangente en A devient horizontale ($\theta = 0$). L'équation devient

$$dy = \frac{x^2}{\sqrt{(c^2 - x^2)(c^2 + x^2)}} dx = \frac{x^2}{\sqrt{c^4 - x^4}} dx .$$

4. Ensuite, pour $a^2 < c^2$ mais $a^2 > \frac{c^2}{2}$ la pente de la tangente devient négative, les ondulations se creusent et apparaissent des points à tangentes horizontales d'abscisses $x = \pm\sqrt{c^2 - a^2}$, les points d'abscisses $x = \pm c$ continuant à être des points à tangente verticale comme l'illustre la figure 7 d'Euler (notre figure 13).

5. Lorsque $a^2 = \frac{c^2}{2}$ on obtient une autre valeur limite. L'équation devient

$$dy = \frac{-a^2 + x^2}{\sqrt{(2a^2 - x^2)x^2}} dx .$$

La tangente en A devient verticale mais A s'éloigne à l'infini, nous allons y revenir. Il y a toujours les points à tangente horizontale $x = a$ mais il n'y a plus qu'un seul point $x = c$ à tangente verticale car l'autre s'est éloigné à l'infini. La symétrique $Aomcnob$ de cette courbe limite $AOMCNOB$ avec A et b partant à l'infini est aussi une solution, solution pour laquelle c'est $x = -c$ qui est le point à tangente verticale et $x = -a$ qui donne les points à tangente horizontale.

6. Lorsque $a^2 < \frac{c^2}{2}$, la pente en A devient imaginaire et la courbe ne passe plus par A mais elle continue à exister pour $x \in [-c, -\sqrt{c^2 - 2a^2}] \cup [\sqrt{c^2 - 2a^2}, c]$. Aux extrémités de ces intervalles, les tangentes sont verticales et elles sont horizontales pour $x = \pm\sqrt{c^2 - a^2}$.

52. Il y a une confusion de notations pour c qui était aussi l'abscisse de C .

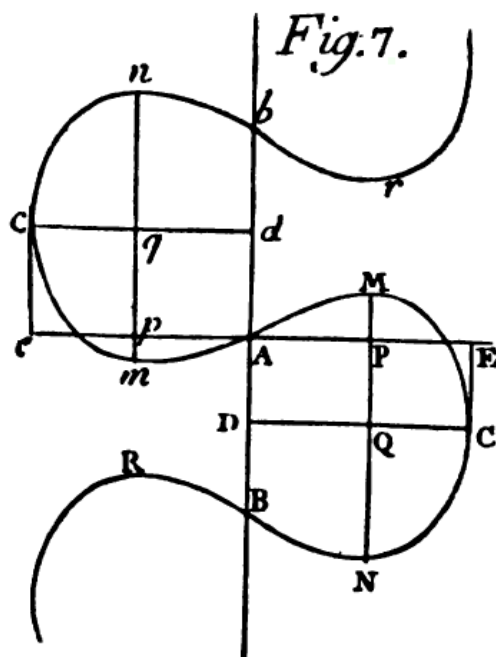


FIGURE 13. La figure 7 de l'*Additamentum 1* du *Methodus* d'Euler [174].

Pour expliquer comment, par exemple dans la figure 7 d'Euler (notre figure 13), la courbe se prolonge au-delà du point C à tangente verticale, Euler déplace l'origine en C . Avec ses notations pour le quadrant $AMECQD$, il pose $QC = t$, $QM = u$, $x + t = AE = CD = c$, i.e. $x = c - t$ et $y + u = CE = AD = b$, i.e. $y = b - u$. D'où une nouvelle équation remplaçant (EL 3) :

$$du = \frac{(a^2 - 2ct + t^2)}{\sqrt{t(2c - t)(2a^2 - 2ct + t^2)}} dt . \quad (\text{EL } 4)$$

En prenant t infinitésimal (i.e. Q très voisin de C), (EL 4) devient $du = \frac{a^2}{2\sqrt{tca^2}} dt$ et en sortant a^2 du radical, il y a une ambiguïté $\pm a$. D'où deux branches $\pm u$ symétriques par rapport à l'axe DC .

Dans sa classification des *elasticæ*, Euler dénombre 9 classes. Repartons de (EL 3),

$$dy = \frac{(a^2 - c^2) + x^2}{\sqrt{(c^2 - x^2)(2a^2 - (c^2 - x^2))}} dx .$$

Elles correspondent à différents intervalles du paramètre $h = \frac{c^2}{2a^2}$.

La classe I correspond à $c \sim 0$. Pour $c = 0$ (ou $a = \infty$, i.e. $h = 0$), on a

$$dy = \frac{a^2 + x^2}{\sqrt{-x^2(2a^2 + x^2)}} dx$$

et la condition $x \in [-c, c]$ implique $x = 0$. La solution est une droite verticale. Pour $c \sim 0$, la même condition implique $x \sim 0$ et donc

$$dy = \frac{a}{\sqrt{2(c^2 - x^2)}} dx ,$$

soit

$$y = \frac{a}{\sqrt{2}} \arcsin\left(\frac{x}{c}\right)$$

qui est l'équation d'une courbe sinusoïdale aussi dite *trochoïde* : la lame ondule légèrement à partir de la position verticale avec des arcs ACB de longueur $AB = \frac{\pi a}{\sqrt{2}}$. La force minimale pour qu'une telle déformation se produise est $\frac{2Ek^2}{a^2}$ (cf. plus haut).

La classe II correspond au cas $c^2 < a^2$ ($h < \frac{1}{2}$). Nous l'avons déjà étudiée.

La classe III correspond au cas limite $a = c$ ($h = \frac{1}{2}$) où la tangente en A est horizontale. C'est le cas de l'*elastica rectangulaire* qui, nous l'avons vu, a été à l'origine de toute la théorie. Dans ce cas, on a :

$$dy = \frac{x^2}{\sqrt{a^4 - x^4}} dx, \quad ds = \frac{a^2}{\sqrt{a^4 - x^4}} dx .$$

Soit $AC = f$ (longueur d'arc) et $AD = b$. Par des développements explicites en série, Euler réussit à les calculer et retrouve la relation remarquable

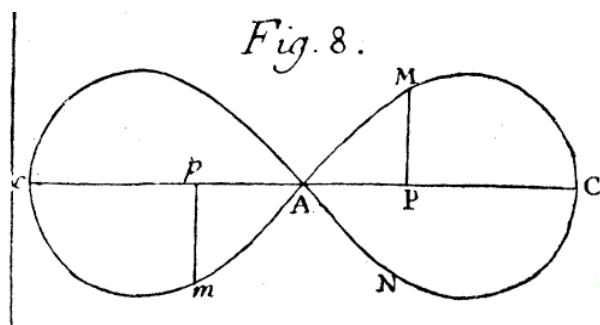
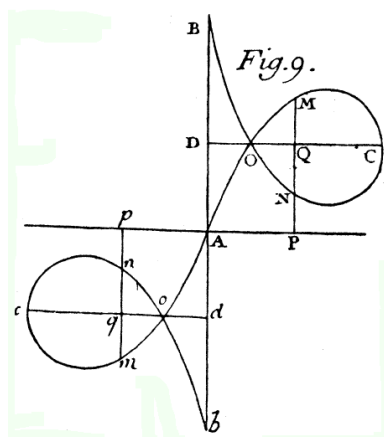
$$4bf = \pi a^2$$

(qu'il avait déjà démontrée en 1738) qui est un cas particulier anticipant la remarquable *relation de Legendre* que nous rencontrerons à la section suivante 11.1. Cela lui permet de calculer la force à exercer pour obtenir l'*elastica rectangulaire*. Le calcul montre que $a \sim \frac{12f}{5\pi}$ et donc la force est

$$\frac{2Ek^2 \cdot 25\pi^2}{(12f)^2} = \frac{25\pi^2 Ek^2}{72f^2} .$$

La classe IV correspond à ce qui se passe lorsque c^2 commence à devenir $> a^2$ ($\frac{1}{2} < h < \nu$ où ν est une valeur critique). On obtient des courbes "en Ω " comme dans la figure 7 d'Euler (notre figure 13), courbes dont le demi-écartement est $AD = b > 0$. Le sinus de l'angle MAP est $\frac{c^2 - a^2}{a^2}$. Euler calcule, grâce à ses développements en série, qu'il a une limite d'environ $40^\circ 41'$, limite à laquelle b s'annule.

La classe V correspond à ce cas limite $h = \nu$ (cf. la figure 8 d'Euler 14). C'est celui de la *lemniscate*. Il correspond à $\frac{c^2 - a^2}{a^2} \sim 0.651868\dots$ et $\frac{c^2}{2a^2} = h = \nu \sim 0.825934\dots$ (on appréciera la précision des calculs d'Euler). On se rendit compte plus tard qu'il

FIGURE 14. La figure 8 de l'*Additamentum 1* du *Methodus* d'Euler [174].FIGURE 15. La figure 9 de l'*Additamentum 1* du *Methodus* d'Euler [174].

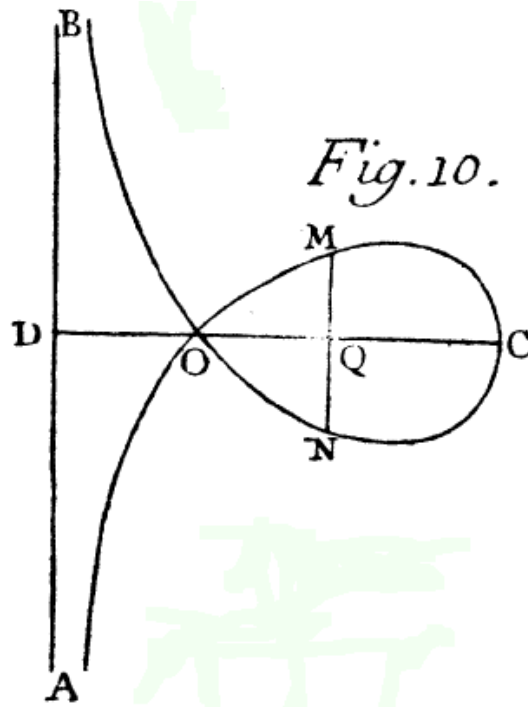
s'agit de la valeur du module k des fonctions elliptiques (cf. la section suivante 11) pour laquelle $2E(k) - K(k) = 0$. Comme l'angle $MAP \sim 40^\circ 41'$, la force verticale appliquée en A fait donc un angle d'environ $40^\circ 41' + 90^\circ = 130^\circ 41'$ avec la courbe.

La classe VI correspond à $h = \frac{c^2}{2a^2} > v$ mais $h < 1$, ce nouveau cas limite $c^2 = 2a^2$ ayant déjà été rencontré plus haut. La lemniscate se déploie verticalement en une série de boucles alternant des deux côtés de l'axe vertical en A (cf. la figure 9 d'Euler, notre figure 15).

La classe VII correspond au cas limite $c^2 = 2a^2$ ($h = 1$), autrement dit à

$$dy = \frac{x^2 - a^2}{x\sqrt{2a^2 - x^2}} dx, \quad ds = \frac{a^2}{x\sqrt{2a^2 - x^2}} dx.$$

Les boucles déployant la lemniscate sont devenues infiniment séparées et il n'en reste plus qu'une seule. Les points A et B d'abscisse $x = 0$ partent à l'infini et la tangente verticale devient une asymptote (cf. la figure 10 d'Euler, notre figure 16). Cela se

FIGURE 16. La figure 10 de l'*Additamentum 1* du *Methodus* d'Euler.[174].

voit bien car l'équation est intégrable et donne ($x = DQ, y = QM$)

$$y = \sqrt{c^2 - x^2} - \frac{c}{2} \log \left(\frac{c + \sqrt{c^2 - x^2}}{x} \right) .$$

On constate que y diverge logarithmiquement pour $x = 0$. Les points $x = a$ sont à tangente horizontale, Le nœud O de la boucle est donné par $y = 0$. Euler calcule et trouve un angle de $112^\circ 56' 48'' \dots$ La courbe était appelée à l'époque une "syntractrix" (ou "courbe des forçats").

La classe VIII correspond à $c^2 > 2a^2$, i.e. $c^2 = 2a^2 + g^2$, $g^2 \leq c^2$, soit $a^2 = \frac{c^2 - g^2}{2}$. L'équation (EL 3) devient alors

$$dy = \frac{x^2 - \frac{1}{2}(c^2 + g^2)}{\sqrt{(c^2 - x^2)(x^2 - g^2)}} dx . \quad (\text{EL } 5)$$

Avec les notations de la figure 11 d'Euler (notre figure 17), $x = DQ, y = QM$, il faut $x \in [g, c]$, les droites $x = g$ et $x = c$ étant des tangentes verticales. Les points

$$x = \sqrt{\frac{1}{2}(c^2 + g^2)}$$

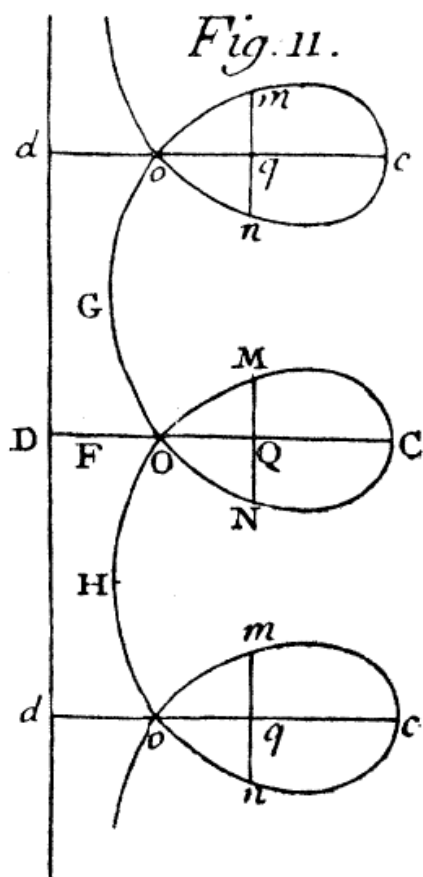


FIGURE 17. La figure 11 de l'*Additamentum 1* du *Methodus* d'Euler [174].

sont à tangente horizontale mais un seul point, $x = c$, est à tangente verticale car l'autre est parti à l'infini. La symétrique $Aomcnob$ de cette courbe limite $AOMCNOB$ avec A et b partant à l'infini est aussi une solution et alors c 'est $x = -c$ le point à tangente verticale et $x = -\sqrt{\frac{1}{2}(c^2 + g^2)}$ les points à tangente horizontale.

Enfin, la classe IX devrait correspondre au cas où $g = c$, i.e. $C = F$ dans la figure 11 d'Euler (notre figure 17). Mais dans ce cas, la courbe disparaît.puisque le radical $\sqrt{(c^2 - x^2)(x^2 - c^2)} = \sqrt{-(c^2 - x^2)^2}$ devient imaginaire. Euler introduit alors très subtilement un cas supplémentaire où $g = c$ signifie que g et c tendent tous deux vers l'infini leur différence restant toutefois *finie*. Cela implique que la courbe limite occupe un espace fini. Autrement dit, Euler interprète $g = c$ comme $\frac{c}{g} = 1$ et le considère comme le cas limite $\frac{c}{g} \rightarrow 1$ ce qui lui permet d'écrire $g = c - 2h$ avec $c = \infty$ et h fini! Mais il va plus loin et, en utilisant en quelque sorte une technique

de géométrie projective, il considère une courbe située au voisinage de l'infini en posant $x = c - h - t$ avec t fini. C'est ainsi l'introduction des deux grandeurs *finies* h et t qui lui permet de surmonter le problème de la négativité du radical. Il déduit en effet des approximations $\frac{1}{2}(c^2 + g^2) \sim c^2 - 2ch$ car $h^2 \sim 0$ par rapport à c et donc, puisque h et t sont finis alors que c est infini,

$$x^2 - \frac{1}{2}(c^2 + g^2) \sim (c - h - t)^2 - c^2 + 2ch \sim -2ct$$

D'où $(c^2 - x^2) \sim 2c(h + t)$ et $(x^2 - g^2) \sim 2c(h - t)$, ce qui conduit en définitive à une équation différentielle finie d'où c a disparu et où le radical est positif si $h^2 \geq t^2$

$$dy = \frac{t}{\sqrt{h^2 - t^2}} dt$$

Cette équation est connue et c'est tout simplement celle d'un *cercle*. Le cercle constitue ainsi la classe IX des elasticæ d'Euler.

Il y a beaucoup d'autres choses admirables dans le texte d'Euler, mais nous ne pouvons pas nous y arrêter et nous allons passer à la classification générale des intégrales elliptiques après avoir fait une remarque sur les relations entre elasticæ et pendule, relation que nous rencontrerons plusieurs fois.

10.5. Géométrie des elasticæ et mécanique du pendule

Il existe une équivalence entre la géométrie des elasticæ et la mécanique du pendule. Elle a été mise au jour par Kirchoff en 1859 (cf. [327]). Elle peut se voir en notant que la courbure $\kappa(s)$ d'une courbe est la déflexion $\frac{d\theta}{ds}$ de l'angle de sa tangente par rapport à un axe de référence (ds est l'élément d'arc). La minimisation de l'énergie élastique de la lame $\int_0^\ell \kappa(s)^2 ds$ avec des conditions aux limites fixées conduit à une équation différentielle du type

$$\frac{d^2\theta}{ds^2} + a \sin(\theta) + b \cos(\theta) = 0$$

que, en changeant d'axe pour annuler b et en posant $a = -\frac{g}{l}$, on peut ramener à l'équation du pendule. Dans cette équivalence, le décours temporel de l'angle du pendule est l'analogie de la variation de l'angle de la tangente le long de l'elastica.

Dans sa thèse de 1906 [66] *Stability of elastic lines in the plane and the space*, Max Born, plus tard co-fondateur de la mécanique quantique, a approfondi cette remarquable équivalence. Pour les elasticæ la vitesse est unitaire (la loi du mouvement est simple) mais la tangente et la courbure suivent des lois compliquées, alors que pour le pendule la géométrie est simple (cercle) mais la loi du mouvement est compliquée.

Étant donnée cette équivalence entre elasticæ et pendule, les elasticæ sont calculables en termes d'intégrales elliptiques et de fonctions elliptiques de Jacobi. Mais

nous différons cette présentation jusqu'à la section 13 du chapitre 14 après que nous ayons expliqué le cadre de la théorie du contrôle optimal.

11. Intégrales, fonctions et courbes elliptiques

11.1. La classification de Legendre

Récapitulons ce que nous avons vu à propos de la lemniscate et du pendule dans les sections 6 et 7 (IE = intégrale elliptique).

	Lemniscate (géométrie)	Pendule (mécanique)
Equation	$\rho^2 = \cos(2\theta)$	$(\dot{\varphi})^2 = 1 - k^2 \sin^2(\varphi)$
Variable 1	angle $t = \cos(2\theta)$ $\rho = \sqrt{t}$	angle $\varphi =$ amplitude $\text{am}(u)$
Différentielle 1	$ds = \frac{1}{2} \frac{dt}{\sqrt{t-t^3}}$ (élément d'arc)	$dt = \frac{d\varphi}{\sqrt{1-k^2 \sin^2(\varphi)}}$ (durée)
IE 1	$s(u) = \frac{1}{2} \int_0^u \frac{dt}{\sqrt{t-t^3}}$ (longueur d'arc)	$t = u(\varphi) =$ $\int_0^{\varphi=\text{am}(u)} \frac{d\psi}{\sqrt{1-k^2 \sin^2(\psi)}}$
Variable 2	corde ρ	$\sigma = \sin(\varphi) = \text{sn}(u)$
Différentielle 2	$ds = \frac{dr}{\sqrt{1-r^4}}$	$dt = \frac{d\sigma}{\sqrt{1-\sigma^2} \sqrt{1-k^2 \sigma^2}}$
IE 2	$s(\rho) = \int_0^\rho \frac{dr}{\sqrt{1-r^4}}$	$t = u(\varphi) =$ $\int_0^{\text{sn}(u)} \frac{d\sigma}{\sqrt{1-\sigma^2} \sqrt{1-k^2 \sigma^2}}$
Fonctions elliptiques	$sl(\varphi)$ avec $\varphi = \int_0^\rho \frac{dr}{\sqrt{1-r^4}}, cl(\varphi)$	$\text{sn}(u), \text{cn}(u), \text{dn}(u)$

Nous voyons apparaître des intégrales ayant un fort air de famille dans la mesure où elles sont du type $\int R(x, \sqrt{P(x)}) dx$ où R est une fraction rationnelle et P un polynôme de degré 3 ou 4 dont les racines sont simples (car s'il y a une racine double a et que $P(x) = (x-a)^2(x-c)$ on peut sortir $(x-a)$ du radical). Au cours du XVIII^e siècle on entreprit de les analyser, de les classer, de les structurer, de les décomposer, de les transformer par changements de variable, d'en trouver les cas élémentaires permettant de reconstruire toutes les autres à partir d'opérations bien réglées et calculables comme des combinaisons linéaires.

Il existe une foule de références plus ou moins classiques pour naviguer dans cet univers des intégrales et des fonctions elliptiques et découvrir leurs surprenantes relations avec d'autres fonctions transcendentes (Gamma, Zêta, Thêta, hypergéométrique, de Legendre, de Bessel, de Mathieu, etc.). Nous nous bornons à en indiquer quelques-unes. Par exemple le *Course of Modern Analysis* de Edmund Whittaker (1902) ([567], éditions successives avec G.N. Watson), le *Handbook* [85]

de Paul Byrd et Morris Friedman (1954), le traité de fonctions spéciales [89] de Carlson (1977), la *Digital Library of Mathematical Functions* [145], l'intéressant site de Jean-Claude Pénin [405] *Les fonctions elliptiques : une histoire*, mis à jour depuis 2007. Les entrées "Elliptic Integral" de *Wikipedia* et *Mathematica* (Wolfram MathWorld) sont excellentes.

En 1784, dans son texte "Une nouvelle méthode de Calcul Intégral" [310], Joseph-Louis Lagrange montra que les intégrales concernées $\int R(x, P(x)) dx$ se ramenaient par changements de variables successifs à une forme

$$\int Ldz + \int \frac{M}{\sqrt{(1 \pm h^2 z^2)(1 \pm k^2 z^2)}} dz$$

avec L et M des fonctions rationnelles de z^2 et $0 < k < h$. Ces travaux considérables d'"entomologiste" conduisirent à l'extraordinaire traité d'Adrien-Marie Legendre *Traité des fonctions elliptiques et des intégrales Eulériennes* [323] qui collectait ses travaux commencés dès 1786 et parut en trois tomes en 1825, 1826, 1828. Les trois "espèces" de base dégagées par Legendre et dont Liouville montra ensuite qu'elles sont irréductibles, c'est-à-dire inexprimables au moyen d'intégrales de fonctions déjà connues, sont :

(I) Première espèce :

$$u = \int_0^x \frac{d\sigma}{\sqrt{(1 - \sigma^2)(1 - k^2 \sigma^2)}},$$

$k \in]0, 1[$ (forme de Jacobi), ou

$$F(\varphi, k) = \int_0^\varphi \frac{d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}}$$

(forme de Legendre), avec le changement de variable $x = \sin(\varphi)$ et donc $\varphi = \text{Arcsin}(x)$. Le lien entre les deux formes est donné par

$$x = \sin(\varphi) = \sin(\text{am}(u)) = \text{sn}(u, k) .$$

Le comodule est $k' = \sqrt{1 - k^2} = \cos(\text{am}(u))$. Pour $x = 1$, $\varphi = \frac{\pi}{2}$, les intégrales complètes donnant le quart de période K sont

$$K = \int_0^1 \frac{d\sigma}{\sqrt{(1 - \sigma^2)(1 - k^2 \sigma^2)}} = \int_0^{\frac{\pi}{2}} \frac{d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}} .$$

(II) Deuxième espèce :

$$\int_0^x \sqrt{\frac{1 - k^2 \sigma^2}{1 - \sigma^2}} d\sigma$$

ou

$$E(\varphi, k) = \int_0^\varphi \sqrt{1 - k^2 \sin^2(\psi)} d\psi$$

C'est l'intégrale princeps rectifiant l'ellipse. Comme

$$\sqrt{\frac{1 - k^2 \sigma^2}{1 - \sigma^2}} = \frac{1 - k^2 \sigma^2}{\sqrt{(1 - \sigma^2)(1 - k^2 \sigma^2)}} ,$$

on a

$$k^2 \int_0^x \frac{\sigma^2 d\sigma}{\sqrt{(1 - \sigma^2)(1 - k^2 \sigma^2)}} = k^2 \int_0^{\frac{\pi}{2}} \frac{\sin^2(\psi) d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}} = (I) - (II) .$$

(III) Troisième espèce :

$$\int_0^x \frac{1}{(1 \pm n^2 \sigma^2) \sqrt{((1 - \sigma^2) 1 - k^2 \sigma^2)}} d\sigma$$

ou

$$\Pi(\varphi, k, n) = \int_0^\varphi \frac{1}{(1 \pm n^2 \sin^2(\psi)) \sqrt{1 - k^2 \sin^2(\psi)}} d\psi .$$

En ce qui concerne les quarts de période, notés traditionnellement K pour (I) et E pour (II), les spécialistes de l'époque ont trouvé, en calculant les équations différentielles (ED) que K satisfait, de remarquables développements en série (S) explicites :

$$(ED_I) \quad \frac{d}{dk} \left(k(1 - k^2) \frac{dK(k)}{dk} \right) = kK(k)$$

$$(ED_{II}) \quad (k^2 - 1) \frac{d}{dk} \left(k \frac{dE(k)}{dk} \right) = kE(k) ,$$

$$(S_I) \quad K(k) = \frac{\pi}{2} \sum_{n=0}^{n=\infty} \left(\frac{(2n)!}{2^{2n} (n!)^2} \right)^2 k^{2n}$$

$$(S_{II}) \quad E(k) = \frac{\pi}{2} \sum_{n=0}^{n=\infty} \left(\frac{(2n)!}{2^{2n} (n!)^2} \right)^2 \frac{k^{2n}}{1 - 2n} .$$

Par exemple, pour démontrer l'équation différentielle (ED_I), on remarque que

$$\begin{aligned} \frac{dE}{dk} &= - \int_0^{\frac{\pi}{2}} \frac{k \sin^2(\psi) d\psi}{\sqrt{1 - k^2 \sin^2(\psi)}} = \frac{E - K}{k} \\ \frac{dK}{dk} &= \frac{E - (1 - k^2) K}{k^2 (1 - k^2)} \\ \frac{d^2 K}{dk^2} &= \frac{(-1 + 3k^2) E + (1 - 3k^2 + 2k^4) K}{k^2 (k^2 - 1)^2} . \end{aligned}$$

Il est alors trivial de vérifier (ED_I) , c'est-à-dire l'équation différentielle, dite *hypergéométrique*,

$$k(1-k^2) \frac{d^2 K}{dk^2} + (1-3k^2) \frac{dK}{dk} - kK = 0.$$

Pour démontrer le développement en série S_I , on développe $\frac{1}{\sqrt{1-k^2 \sin^2(\psi)}}$ en série par rapport à $\sin^2(\psi)$, on utilise les intégrales de Wallis

$$\int_0^{\frac{\pi}{2}} \sin^{2n}(\psi) d\psi = \frac{\pi (2n-1)!!}{2^{2n} n!}$$

(!! est la double factorielle) et on réorganise les termes.

Les spécialistes ont découvert d'innombrables relations remarquables dans l'univers des intégrales elliptiques. L'une des plus remarquable est la relation (symétrique en k, k') de Legendre reliant les intégrales complètes de première et de deuxième espèce pour le module k et le comodule $k' = \sqrt{1-k^2}$:

$$K(k) E(k') + K(k') E(k) - K(k) K(k') = \frac{\pi}{2}.$$

La relation $4bf = \pi a^2$ découverte en 1738 par Euler pour les *elasticæ* rectangulaires (cf. section 10) en est un avatar.⁵³ En effet rappelons que dans ce cas

$$f = \int_0^a ds = \int_0^a \frac{a^2}{\sqrt{a^4 - x^4}} dx$$

alors que

$$b = \int_0^a dy = \int_0^a \frac{x^2}{\sqrt{a^4 - x^4}} dx.$$

Donc (par changement de variable $x = a \cos(\psi)$) $f = \frac{a}{\sqrt{2}} K\left(\frac{1}{2}\right)$ et $b = \frac{a}{\sqrt{2}} (2E - K)$. Comme $k = \frac{1}{2}$, $k' = 1 - k = \frac{1}{2}$, $K = K'$ et la relation de Legendre devient $(2E - K) K = \frac{\pi}{2}$, soit $\frac{b\sqrt{2}}{a} \frac{f\sqrt{2}}{a} = \frac{\pi}{2}$ qui est exactement la relation d'Euler.

Pour toutes ces intégrales on peut considérer les fonctions inverses comme l'ont effectué Gauss et Abel et obtenir les fonctions elliptiques doublement périodiques approfondies par Jacobi dès 1827. Les *Recherches sur les fonctions elliptiques* de Niels Abel [1] sont une référence absolue.

11.2. Vers les courbes elliptiques

Les fonctions elliptiques (FE) étudiées sont des fonctions méromorphes sur des tores. Les progrès décisifs de la théorie générale des fonctions d'une variable complexe, d'abord par Cauchy, puis par Riemann (théorie des surfaces de Riemann) et

53. Cf. [394].

d'autres grands mathématiciens comme Liouville ou Weierstrass, conduisit à renverser le point de vue et à développer une théorie des fonctions méromorphes sur des tores complexes J quotients de \mathbb{C} par un réseau $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ engendré par un parallélogramme (non dégénéré) de périodes (ω_1, ω_2) ⁵⁴. Un point essentiel est que Λ étant un sous-groupe additif de \mathbb{C} , l'addition de \mathbb{C} passe au quotient et que donc J est canoniquement muni d'une structure de *groupe additif*. Cela conduisit à l'extraordinaire théorie des *courbes elliptiques* dont les applications sont innombrables (par exemple, la démonstration du théorème de Fermat par Andrew Wiles⁵⁵).

On montra qu'à des FE de ce type étaient associées des intégrales dont l'intégrande comprenait des \sqrt{P} , P étant un polynôme de degré 3 ou 4. On montra que si $f(z)$ est une telle fonction sur $J = \mathbb{C}/\Lambda$, $f'(z)$ l'est également et que les FE sur J forment un corps $K(J)$. Étant donné un tore J , on construisit des FE sur J en utilisant des séries du type $\sum_{\lambda \in \Lambda} (z - \lambda)^{-k}$ périodiques par construction quitte à les modifier un peu pour satisfaire de bonnes conditions de convergence. La plus connue est la FE (paire et possédant des pôles doubles aux points de Λ) introduite par Weierstrass :

$$\mathfrak{P}(z) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda - \{0\}} \left((z - \lambda)^{-2} - \frac{1}{\lambda^2} \right)$$

$$\mathfrak{P}'(z) = -2 \sum_{\lambda \in \Lambda} (z - \lambda)^{-3} .$$

On montre que \mathfrak{P} et \mathfrak{P}' engendrent le corps $K(J)$ des FE et aussi, c'est fondamental, qu'elles sont reliées par une équation algébrique cubique (i.e. du 3^e degré) :

$$\mathfrak{P}'(z)^2 = 4\mathfrak{P}(z)^3 - g_2\mathfrak{P}(z) - g_3 = P(\mathfrak{P}(z))$$

dont les coefficients sont exprimables en termes de séries d'Eisenstein

$$G_m = \sum_{\lambda \in \Lambda - \{0\}} \frac{1}{\lambda^m}$$

au moyen de $g_2 = 60G_4$ et $g_3 = 140G_6$. Si l'on pose $x = \mathfrak{P}(z)$ et $y = \mathfrak{P}'(z)$, on obtient la cubique complexe E dans \mathbb{C}^2 d'équation $y^2 = 4x^3 - g_2x - g_3$, courbe que l'on projectivise car les pôles s'envoient à l'infini. E est dite *courbe elliptique*. À travers \mathfrak{P} et \mathfrak{P}' , le tore complexe $J = \mathbb{C}/\Lambda$ devient isomorphe à E , la loi d'addition

54. Les fonctions sont seulement méromorphes parce que, rappelons-le, d'après un théorème de Liouville (cf. section 1), toute fonction holomorphe sur une surface complexe compacte T est nécessairement constante. En effet les équations de Cauchy-Riemann les empêchent d'avoir des extrema alors que la compacité de T les force au contraire à en posséder.

55. Pour quelques informations pédagogiques sur ce grand théorème, voir notre compilation [421] et sa bibliographie.

se définissant alors géométriquement de façon très simple : si a et b sont deux points de E , la droite ab coupe E en un troisième point c puisque E est cubique. On pose alors $a + b + c = 0$.

Comme $\mathfrak{P}'(z)^2 = \left(\frac{d(\mathfrak{P}(z))}{dz}\right)^2 = P(\mathfrak{P}(z))$, on a $dz = \frac{d\mathfrak{P}}{\sqrt{P(\mathfrak{P})}}$ et l'on retrouve les intégrales elliptiques dont on était parti. Par exemple, pour la lemniscate de parallélogramme des périodes $(2\pi G, 2\pi iG)$, le calcul des séries d'Eisenstein donne $g_2 = 4$ et $g_3 = 0$, d'où la cubique $y^2 = 4x^3 - 4x = 4x(x+1)(x-1)$. On a donc

$$z = \frac{1}{2} \int \frac{d\mathfrak{P}}{\sqrt{\mathfrak{P}^3 - \mathfrak{P}}}$$

qui est bien l'intégrale

$$\frac{1}{2} \int \frac{dt}{\sqrt{t^3 - t}}$$

que nous avons rencontrée section 6.

La fascinante histoire des courbes elliptiques dépasse le cadre de cet ouvrage et nous allons donc en revenir, après l'avoir ainsi contextualisé historiquement et théoriquement, à notre modèle variationnel de géodésiques "legendriennes".

Contours illusoires modaux et géodésiques “legendriennes”

1. Les différents modèles

Nous allons maintenant commencer notre long trajet parcourant les différents modèles géodésiques possibles de contours illusoires modaux. Rappelons que le problème général est de modéliser ces contours en tant qu’engendrés par des processus *corticaux*, certes de bas niveau mais corticaux. Si l’on veut pouvoir modéliser la contribution de *V1* on doit donc partir d’un modèle \mathbb{V} de *V1*. Les modèles classiques (comme celui des *elasticæ*) sont des modèles de traitement d’image et se situent dans le plan de base \mathbb{R}^2 . Nous voulons quant à nous investiguer des modèles géodésiques dans \mathbb{V} . Plusieurs classes en sont envisageables.

1. Comme nous l’avons anticipé à la section 3.2 du chapitre 8, notre première classe est celle de “géodésiques legendriennes” qui reposent sur un principe de minimisation de la longueur dans le fibré $\mathbb{V}_J = J^1\mathbb{R}^2$.
 - (a) Le premier modèle utilise une métrique riemannienne sur \mathbb{V}_J et des équations d’Euler-Lagrange contraintes, et donc des multiplicateurs de Lagrange.
 - (b) Le second utilise une métrique sous-riemannienne sur les plans de contact de \mathbb{V}_J qui est invariante par la structure du groupe de Heisenberg.
2. Notre seconde classe de modèles géodésiques concerne $\mathbb{R}^2 \times \mathbb{S}^1$ mais de différentes façons.
 - (a) Le premier, que nous noterons \mathbb{V}_θ , est celui du “fibré en cercles”. Il utilise lui aussi une métrique riemannienne et des équations d’Euler-Lagrange contraintes, et donc des multiplicateurs de Lagrange.
 - (b) Le second, défini sur $\mathbb{V}_S = \mathbb{R}^2 \rtimes \mathbb{S}^1 = SE(2)$, repose sur la géométrie sous-riemannienne de la structure de contact de $SE(2)$.
 - (c) Le modèle \mathbb{V}_S possède des singularités, les géodésiques pouvant avoir des cusps. On peut alors le restreindre à sa partie régulière \mathbb{V}_S^\emptyset ($\emptyset =$ “sans cusps”).
 - (d) En fait, dans la mesure où les orientations ne sont définies que modulo π , il est plus naturel de considérer le quotient $\mathbb{V}_P = \mathbb{R}^2 \times \mathbb{P}^1 = SE(2) \pmod{\pi}$.

La première classe de modèles se situe sur \mathbb{V}_J qui, en tant que groupe de Heisenberg polarisé, est la nilpotentisation de $SE(2)$. Elle est donc *l'approximation nilpotente* de la seconde classe

Remarque sur le ds. Dans les chapitres qui suivent nous discuterons plusieurs modèles de géodésiques et deux types de courbes interviendront : (i) les courbes dans le plan de base \mathbb{R}^2 , que nous noterons en général γ , (ii) les courbes dans \mathbb{V} , que nous noterons en général Γ . Suivant les cas, des courbes γ seront liftées en des Γ dans \mathbb{V} (relevées legendriennes) ou au contraire des courbes Γ dans \mathbb{V} seront projetées sur des courbes γ dans \mathbb{R}^2 . Ces courbes seront paramétrées et on devra souvent faire usage dans les calculs de la paramétrisation par la longueur d'arc. Celle-ci est traditionnellement notée s , l'élément de longueur d'arc ds définissant la métrique par le ds^2 . Cette notation est si universelle que nous essayerons de l'utiliser autant que faire se peut. Mais comme il y aura le ds des γ et le ds des Γ , il pourra y avoir des ambiguïtés. Dans les contextes où il n'y aura pas d'ambiguïté nous garderons la notation classique ds en signalant de quel ds il s'agit. Ce n'est que lorsque les deux ds interviendront dans le même contexte que nous changerons de notation. —□

Revenons à la section 4.4 de la fin du chapitre 8. Nous allons calculer dans un style classique (à la Euler) notre premier modèle géodésique. Les calculs pourront paraître un peu fastidieux mais cela est le cas de tous les modèles variationnels de ce genre comme nous venons de le voir avec les *elasticæ*.

2. Les équations d'Euler-Lagrange (I)

2.1. L'essai de la métrique euclidienne

C'est en 1996 que nous avons introduit l'idée de ramener le problème des contours illusoires modaux à un problème de géodésiques dans le modèle choisi \mathbb{V} (\mathbb{V}_J ou \mathbb{V}_S) de $V1$. Le choix d'une métrique appropriée permet de minimiser la longueur des courbes Γ dans \mathbb{V} qui

- (i) satisfont des conditions au bord $(a_1, p_1) = v_1$ et $(a_2, p_2) = v_2$,
- (ii) sont des intégrales de la structure de contact.

Dans [428] et [431], le problème pour l'espace des 1-jets

$$\mathbb{V}_J = \mathbb{R}^2 \times \mathbb{R} = \{x, y, p = \tan(\theta)\}$$

muni de sa forme de contact

$$\omega_J = dy - pdx$$

est résolu de façon classique, à la Euler, au moyen des équations d'Euler-Lagrange associées à la métrique *euclidienne* $ds^2 = dx^2 + dy^2 + dp^2$ où ds est la différentielle

de longueur d'arc.¹ Le long de la relevée legendrienne Γ d'une courbe γ de \mathbb{R}^2 d'équation $y = f(x)$, on a par conséquent $p = f'(x) = \tan \theta(x)$.

Ce n'est qu'un peu plus tard, au début des années 2000, que nous nous sommes rendu compte qu'il s'agissait d'un problème typique de géométrie *sous-riemannienne*.

Tester la métrique euclidienne de \mathbb{V}_J peut sembler relever d'un choix hybride car, alors que ω_J est invariante pour la structure de groupe de \mathbb{V}_J , la métrique ds^2 est associée à la base holonome (x, y, p) qui n'est pas invariante. Mais cela est néanmoins pertinent et intéressant pour deux raisons.

- (i) D'abord, sur le plan purement géométrique, nous avons vu dans la section 3.4 du chapitre 3 que la métrique euclidienne g_E était compatible avec la 1-forme $\omega = \frac{1}{\sqrt{1+p^2}}\omega_J$, et donc la distribution des plans de contact, au sens où les sept conditions suivantes (non indépendantes) sont réalisées, J dénotant la projection orthogonale sur les plans de contact suivie d'une rotation de $\frac{\pi}{2}$ dans ces plans.

- (1) $\omega(X) = g_E(X, \chi)$;
- (2) $J(\chi) = 0$ où χ est le champ de Reeb;
- (3) $\omega \circ J = 0$;
- (4) $d\omega(X, J(Y)) = d\omega(Y, J(X))$;
- (5) $J^2(X) = -X + \omega(X)\chi$;
- (6) $g_E(J(X), J(Y)) = g_E(X, Y) - \omega(X)\omega(Y)$;
- (7) $d\omega(X, Y) = \frac{1}{1+p^2}g_E(X, J(Y))$.

- (ii) Ensuite, sur le plan de la modélisation, comme nous l'avons anticipé à la fin du chapitre 8 section 4.4, la métrique euclidienne est une façon simple d'encoder le fait expérimental de l'*affaiblissement* des connexions horizontales cortico-corticales, et donc des contours illusoirs, quand l'écart entre les valeurs aux bords augmente. On veut calculer des contours illusoirs connectant (a_1, p_1) et (a_2, p_2) . L'axe (a_1, a_2) est donc privilégié et on le prend naturellement comme axe des x . L'angle θ possède donc maintenant une signification *intrinsèque*. L'expérience montre que les connexions s'affaiblissent lorsque θ varie de 0 à $\pm\frac{\pi}{2}$ d'où la nécessité de pénaliser fortement la déviation par rapport au segment de droite (a_1, a_2) i.e. par rapport à $\theta = 0$. C'est l'alignement qui est la référence, i.e. les droites horizontales dans les plans de contact qui relèvent les géodésiques euclidiennes de la base, et les déviations doivent rester limitées. La fonction $p = \tan \theta$ semblant idoine pour modéliser de telles contraintes, il semble justifié de tester d'abord la métrique euclidienne de \mathbb{V}_J .

1. Le ds est donc ici la longueur d'arc dans le fibré \mathbb{V} (et non pas dans la base \mathbb{R}^2).

Remarque. Le choix de l’axe des x signifie que les conditions aux limites pour f aux positions a_1 et a_2 sont $f(x_1) = f(x_2) = 0$. En fait, on s’intéresse surtout au cas *symétrique* où $x_1 = -x_2$, $\theta_2 = \pi - \theta_1$ et où $f(x)$ est paire et $f'(x)$ impaire. \square

2.2. Le lagrangien non contraint

Nous considérons donc l’approximation nilpotente de $SE(2)$ muni de la métrique euclidienne. Nous choisissons un repère Oxy de M où l’axe des x est identifié à (a_1, a_2) . L’invariance par changement de repère est alors exprimée par l’action du groupe euclidien $SE(2)$ sur \mathbb{V}_J (cf. la section 2.4 du chapitre 3). Nous devons calculer les courbes Γ de longueur minimale dans \mathbb{V}_J pour sa métrique euclidienne sous la contrainte “cinématique” qu’elles soient des relevées legendriennes, c’est-à-dire satisfassent la condition d’intégrabilité d’être des courbes intégrales de la structure de contact \mathcal{K} . Nous appellerons ces géodésiques des “géodésiques legendriennes”.

Le long d’une relevée legendrienne $\Gamma = \{x, f(x), f'(x)\}$ de $(a_1, p_1) = v_1$ à $(a_2, p_2) = v_2$ paramétrée par x , on a $dy = f'(x) dx$ et $dp = f''(x) dx$, soit

$$ds = \sqrt{1 + f'(x)^2 + f''(x)^2} dx .$$

La fonctionnelle à minimiser est la longueur $\mathcal{E} = \int_{v_1}^{v_2} ds$ et il faut par conséquent minimiser l’intégrale $\mathcal{E} = \int_{x_1}^{x_2} L(x) dx$, où L est le lagrangien

$$L = \sqrt{1 + f'(x)^2 + f''(x)^2} .$$

À la Euler et avec ses notations, on remarquera que, relativement à l’espace des jets, le lagrangien

$$L = \sqrt{1 + p^2 + q^2}$$

avec $p = f'(x)$ et $q = f''(x)$ est définissable pour toutes les courbes Γ dans \mathbb{V}_J , qu’elles soient ou non des relevées legendriennes.

On peut exprimer la situation plus intrinsèquement. Si l’on note, comme dans les chapitres précédents, $\mathbf{t} = (\xi, \eta, \pi)$ les vecteurs tangents, on a

$$\{1, f'(x)^2, p'(x)^2\} = \{\xi^2 = 1, \eta^2, \pi^2\}$$

le long de Γ (avec $\xi = 1$ car x est prise comme variable de base). On peut donc se placer sur le fibré tangent $T\mathbb{V}_J$ de $\mathbb{V}_J \simeq \mathbb{R}^3$, espace de dimension 6 de coordonnées $(v; \mathbf{t}) = (x, y, p; \xi, \eta, \pi)$ et introduire sur $T\mathbb{V}_J$ le lagrangien

$$L(v; \mathbf{t}) = \sqrt{\xi^2 + \eta^2 + \pi^2} = \|\mathbf{t}\| ,$$

autrement dit la norme de \mathbf{t} pour la métrique euclidienne. Le long d’une courbe (régulière) quelconque Γ , $L(v; \mathbf{t})$ devient la fonction composée de x ,

$$L(x) = L(v(x); \mathbf{t}(x)) ,$$

et la fonctionnelle \mathcal{E} devient l'intégrale $\int_{\Gamma} L(v(x); \mathbf{t}(x)) dx$. Il faut insister encore une fois sur le fait que L est une fonction de x en tant que fonction composée à travers $(v; \mathbf{t})$ et qu'elle est donc dérivable d'une part comme fonction $L(v; \mathbf{t})$ avec ses 6 dérivées partielles $\left\{ \frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}, \frac{\partial L}{\partial p}, \frac{\partial L}{\partial \xi}, \frac{\partial L}{\partial \eta}, \frac{\partial L}{\partial \pi} \right\}$ et d'autre part comme fonction $L(x) = L(v(x); \mathbf{t}(x))$ avec

$$\frac{dL}{dx} = \frac{\partial L}{\partial x} \frac{dx}{dx} + \frac{\partial L}{\partial y} \frac{dy}{dx} + \frac{\partial L}{\partial p} \frac{dp}{dx} + \frac{\partial L}{\partial \xi} \frac{d\xi}{dx} + \frac{\partial L}{\partial \eta} \frac{d\eta}{dx} + \frac{\partial L}{\partial \pi} \frac{d\pi}{dx}.$$

Le long de $\Gamma = \{x, y = f(x), p(x)\}$, qu'elle soit ou non une relevée legendrienne, $\mathbf{t} = \{\xi = 1, \eta = f'(x), \pi = p'(x)\}$ et le lagrangien L est

$$L = \sqrt{\xi^2 + \eta^2 + \pi^2} = \sqrt{1 + f'(x)^2 + p'(x)^2} dx.$$

Les équations d'Euler-Lagrange *non contraintes* de ce système sont :

$$\begin{cases} \frac{\partial L}{\partial y} - \frac{d}{dx} \left(\frac{\partial L}{\partial \eta} \right) = 0 \\ \frac{\partial L}{\partial p} - \frac{d}{dx} \left(\frac{\partial L}{\partial \pi} \right) = 0 \end{cases}$$

Elles ne sont pas contraintes par la condition d'intégrabilité disant que les courbes recherchées doivent être des intégrales de la structure de contact et satisfaire par conséquent la relation $p = f'(x)$, autrement dit $p = \eta$ et $p'(x) = f''(x)$. Les solutions sont évidemment les droites de \mathbb{V}_J qui sont les géodésiques de sa métrique euclidienne. Cela se vérifie immédiatement. Comme $\frac{\partial L}{\partial y} = 0$ et $\frac{\partial L}{\partial p} = 0$ en absence de contrainte, les équations d'E-L s'intègrent immédiatement : il existe deux constantes d'intégration A et B avec $A^2 + B^2 < 1$ telles que

$$\begin{cases} \frac{\partial L}{\partial \eta} = \frac{f'}{\sqrt{1+f'(x)^2+p'(x)^2}} = A \\ \frac{\partial L}{\partial \pi} = \frac{p'}{\sqrt{1+f'(x)^2+p'(x)^2}} = B \end{cases}$$

et donc f' et p' sont constantes :

$$\begin{cases} f'(x)^2 = \frac{A^2}{1-(A^2+B^2)} \\ p'(x)^2 = \frac{B^2}{1-(A^2+B^2)} \end{cases}$$

Ces solutions ne sont pas en général des relevées legendriennes sauf si $B = 0$ car leurs projections sur \mathbb{R}^2 sont des droites et les relevées des droites sont "horizontales" puisque $p' = f'' = 0$.

Pour tenir compte de la contrainte d'intégrabilité "cinématique" $\Sigma = p - \eta = 0$, on doit considérer le lagrangien *contraint* $L + \lambda \Sigma$, où λ est un *multiplicateur de Lagrange*.

2.3. Le multiplicateur de Lagrange

2.3.1. La méthode.

Rappelons brièvement le principe de la résolution de ce genre de problèmes variationnels avec contrainte que nous avons vu apparaître section 8 dans le *Methodus* d’Euler². Comme nous l’avons vu au chapitre précédent (9), le calcul variationnel généralise aux fonctionnelles d’action $S(\gamma)$ qui sont définies sur des espaces fonctionnels (en général un espace de trajectoires γ dans un espace de configurations M) la recherche d’extrema de fonctions (suffisamment différentiables) $f(x)$ au moyen de l’équation $\nabla f = 0$ (pour que ∇f soit bien défini, il faut que M soit une variété riemannienne). Mais supposons que le système soit soumis à une contrainte formulable par une condition $h(x) = 0$ (h suffisamment différentiable) définissant une hypersurface H . Pour optimiser f en tenant compte de cette contrainte, on considère les hypersurfaces de niveau F_c d’équation $f(x) = \text{cste} = c$, et on regarde les valeurs de la constante c pour lesquelles il existe des points x_0 de F_c où F_c est tangente à H . La tangence s’exprime par le fait que les gradients ∇f et ∇h qui sont par définition normaux à F_c et à H sont *colinéaires* et donc par le fait qu’il existe un réel $\lambda \neq 0$ tel que $\nabla f(x_0) = \lambda \nabla h(x_0)$. Mais cela revient à considérer la fonction $G(x, \lambda) = f(x) - \lambda h(x)$ sur $M \times \mathbb{R}$ et à chercher un point x_0 tel que $\nabla_{(x, \lambda)} G(x_0) = 0$. Les composantes $\nabla_x G(x_0) = 0$ donnent $\nabla_x f(x_0) = \lambda \nabla_x h(x_0)$ et la composante $\nabla_\lambda G(x_0) = \frac{\partial F}{\partial \lambda} = 0$ donne la condition $h(x_0) = 0$ et par conséquent $G(x_0, \lambda) = f(x_0) = c$.

La valeur de λ est intéressante à calculer. Supposons que nous modifions un peu la contrainte et imposons $h = \varepsilon$ plutôt que $h = 0$. Une solution x_0 avec $\lambda = \lambda_0$ devient alors une fonction $x_0(\varepsilon)$ de ε avec avec $\lambda = \lambda_0(\varepsilon)$ et l’on calcule

$$\begin{aligned} \left. \frac{df(x_0(\varepsilon))}{d\varepsilon} \right|_{\varepsilon=0} &= \nabla f(x_0) \cdot \left. \frac{dx_0(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} \quad (\text{produit scalaire}) \\ &= \lambda_0 \nabla h(x_0) \cdot \left. \frac{dx_0(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = \lambda_0 \left. \frac{dh(x_0(\varepsilon))}{d\varepsilon} \right|_{\varepsilon=0}. \end{aligned}$$

Mais $\left. \frac{dh(x_0(\varepsilon))}{d\varepsilon} \right|_{\varepsilon=0} = 1$ puisque $h(x_0(\varepsilon)) = \varepsilon$ et donc $\lambda_0 = \left. \frac{df(x_0(\varepsilon))}{d\varepsilon} \right|_{\varepsilon=0}$. On voit qu’il existe une compétition entre l’optimisation de f et la contrainte $h = 0$. Il faut donc introduire des “forces” contraignant f à respecter la contrainte. C’est la valeur de λ qui implémente cette force.

La méthode des multiplicateurs de Lagrange consiste à étendre cette méthode d’optimisation sous contrainte aux fonctionnelles d’action.

2.3.2. L’équation différentielle pour la tangente $g = f'$.

2. Pour une introduction plus complète, voir par exemple Bourguignon [74].

Appliquons cette méthode à la contrainte $\Sigma = p - \eta = 0$, l'idée étant que les équations d'E-L pour le lagrangien L sous la contrainte $\Sigma = 0$ sont les mêmes que les équations d'E-L sans contrainte pour le lagrangien $L + \lambda\Sigma$. Les nouvelles équations d'E-L s'écrivent alors :

$$\begin{cases} \left(\frac{\partial}{\partial y} - \frac{d}{dx} \frac{\partial}{\partial \eta} \right) (L + \lambda\Sigma) = 0 \\ \left(\frac{\partial}{\partial p} - \frac{d}{dx} \frac{\partial}{\partial \pi} \right) (L + \lambda\Sigma) = 0 \\ \left(\frac{\partial}{\partial \lambda} - \frac{d}{dx} \frac{\partial}{\partial \lambda'} \right) (L + \lambda\Sigma) = 0 \end{cases}$$

où $\lambda(x)$ est le multiplicateur de Lagrange. La dernière équation n'est rien d'autre que la contrainte d'intégrabilité $\Sigma = 0$.

En substituant les expressions de L et de Σ dans les équations et en exprimant les variables y, p, η, π en fonction de x , il vient, en tenant compte du fait que ni L ni Σ ne dépendent de y (et donc $\frac{\partial L}{\partial y} = 0$ et $\frac{\partial \Sigma}{\partial y} = 0$), que $\frac{\partial \Sigma}{\partial \eta} = -1$, que L ne dépend pas de p (et donc $\frac{\partial L}{\partial p} = 0$), que $\frac{\partial \Sigma}{\partial p} = 1$, et que ni λ ni Σ ne dépendent de π (et donc $\frac{\partial \lambda}{\partial \pi} = 0$ et $\frac{\partial \Sigma}{\partial \pi} = 0$) :

$$\begin{cases} \frac{d}{dx} \left(\frac{\partial L}{\partial \eta} - \lambda(x) \right) = 0 \\ \lambda(x) - \frac{d}{dx} \frac{\partial L}{\partial \pi} = 0 \end{cases}$$

Remarque. Si la contrainte Σ est remplacée par la contrainte intégrale

$$\bar{\Sigma} = \int_{x_A}^{x_B} (p - f')^2 dx$$

qui lui est équivalente sous des conditions de bonne régularité de f et p , alors les équations d'E-L ont la même forme que les précédentes avec le multiplicateur $\lambda(x)$ remplacé par $\Lambda(x) = 2\lambda(x)(p - f')$. En éliminant $\lambda(x)$ et $\Lambda(x)$, on obtient donc les mêmes équations. □

Dans notre cas, étant donné que

$$\begin{cases} \frac{\partial L}{\partial \eta} = \frac{f'(x)}{\sqrt{1+f'(x)^2+p'(x)^2}} \\ \frac{\partial L}{\partial \pi} = \frac{p'(x)}{\sqrt{1+f'(x)^2+p'(x)^2}} \end{cases}$$

on obtient :

$$\begin{cases} \frac{d}{dx} \left(\frac{f'(x)}{\sqrt{1+f'(x)^2+p'(x)^2}} - \lambda(x) \right) = 0 \\ \lambda(x) - \frac{d}{dx} \left(\frac{p'(x)}{\sqrt{1+f'(x)^2+p'(x)^2}} \right) = 0 \end{cases}$$

Comme les équations d'E-L sont des identités, le système est trivialement intégrable : il existe une constante A telle que :

$$\frac{\partial L}{\partial \eta} = A + \lambda(x) = A + \frac{d}{dx} \frac{\partial L}{\partial \pi}, \quad (1)$$

soit, puisque la condition d'intégrabilité implique $p'(x) = f''(x)$,

$$\frac{f'(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}} = A + \frac{d}{dx} \left(\frac{f''(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}} \right).$$

autrement dit,

$$\frac{d}{dx} \left(\frac{f''(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}} \right) = \frac{f'(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}} - A.$$

Remarque. Dans le cas *symétrique* qui nous intéresse, on a $A = 0$. _____□

On développe alors la dérivée totale $\frac{dL}{dx}$ en tenant compte du fait que L ne dépend pas de x, y, p et que $\xi = 1, \eta = f'(x), \pi = p'(x)$:

$$\begin{aligned} \frac{dL}{dx} &= \frac{\partial L}{\partial x} \frac{dx}{dx} + \frac{\partial L}{\partial y} \frac{dy}{dx} + \frac{\partial L}{\partial p} \frac{dp}{dx} + \frac{\partial L}{\partial \xi} \frac{d\xi}{dx} + \frac{\partial L}{\partial \eta} \frac{d\eta}{dx} + \frac{\partial L}{\partial \pi} \frac{d\pi}{dx} \\ &= \frac{\partial L}{\partial \eta} \frac{d\eta}{dx} + \frac{\partial L}{\partial \pi} \frac{d\pi}{dx}, \\ \frac{dL}{dx} &= \frac{\partial L}{\partial f'} f'' + \frac{\partial L}{\partial f''} f''' , \end{aligned}$$

soit

$$\frac{dL}{dx} = \frac{\partial L}{\partial f'} f'' + \frac{\partial L}{\partial f''} f''' .$$

Comme par ailleurs,

$$\frac{d}{dx} \left(f'' \frac{\partial L}{\partial f''} \right) = f''' \frac{\partial L}{\partial f''} + f'' \frac{d}{dx} \left(\frac{\partial L}{\partial f''} \right) = f''' \frac{\partial L}{\partial f''} + f'' \left(\frac{\partial L}{\partial f'} - A \right)$$

d'après (1) on a donc :

$$\frac{dL}{dx} = \frac{d}{dx} \left(f'' \frac{\partial L}{\partial f''} \right) + A f''$$

ou encore :

$$\frac{d}{dx} \left(L - f'' \frac{\partial L}{\partial f''} \right) = A f'' , \quad (2)$$

ce qui donne par intégration :

$$L - f'' \frac{\partial L}{\partial f''} = A f' + B .$$

B étant une nouvelle constante d'intégration. Bref, on a :

$$\sqrt{1 + f'(x)^2 + f''(x)^2} = Af'(x) + B + \frac{f''(x)^2}{\sqrt{1 + f'(x)^2 + f''(x)^2}}$$

ou encore, en multipliant les deux membres par $\sqrt{1 + f'(x)^2 + f''(x)^2}$ (qui est toujours ≥ 1) :

$$1 + f'(x)^2 = (Af'(x) + B)\sqrt{1 + f'(x)^2 + f''(x)^2}$$

avec la contrainte $Af'(x) + B > 0$.

Il est impossible d'avoir en même temps $A = 0$ et $B = 0$, car on aurait alors $1 + f'(x)^2 = 0$, ce qui est impossible dans le domaine réel. En dehors de l'origine du plan (A, B) des paramètres, nous obtenons ainsi une équation différentielle du premier ordre en f' . En l'élevant au carré et en posant $g = f'$,³ on obtient la nouvelle équation :

$$(g')^2 = \frac{(1 + g^2)^2 - (1 + g^2)(Ag + B)^2}{(Ag + B)^2} \quad (3)$$

avec la contrainte $(1 + g^2) \geq (Ag + B)^2$, *i.e.* $0 \leq Ag + B \leq \sqrt{1 + g^2}$ puisque $Ag + B \geq 0$.

Faisons quelques remarques sur cette équation différentielle.

1. Contrairement aux équations différentielles simples de la forme $f'(x) = F(x)$ disant que la dérivée d'une fonction inconnue f de x est une fonction connue F de x , f s'obtenant donc par quadrature, elle exprime directement une relation entre g' et une fonction $G(g)$ de g elle-même où x n'intervient pas. Cela justifie une stratégie particulière qui consiste à considérer la *fonction réciproque* $x(g)$ de $g(x)$. En effet, si $g'(x) = \frac{dg}{dx}$, alors $x'(g) = \frac{dx}{dg}$ et l'équation s'écrit

$$x'(g) = \frac{1}{g'(x)} = \frac{1}{G(g)}$$

qui est une équation différentielle de la variable g .

2. L'équation est d'un type particulier correspondant à un calcul de longueur d'arc.
3. Si $g(x)$ était une fonction associée à un mouvement mécanique avec x comme variable temporelle, alors $g'(x)$ serait la vitesse du mouvement et $(g')^2$ serait, à un facteur de masse près, l'énergie cinétique. Comme souvent, notre problème géométrique est formulable comme un problème de mécanique.

3. Nous renommons f' par g et non pas par p car nous voulons pouvoir la décontextualiser et la considérer comme une fonction quelconque.

Si l'on réécrit l'équation (3) sous la forme réciproque :

$$dx^2 = \frac{(Ag + B)^2}{(1 + g^2)^2 - (1 + g^2)(Ag + B)^2} dg^2 ,$$

la solution est alors donnée par l'intégrale (exprimant x en fonction de g et non pas g en fonction de x) avec une constante C d'intégration :

$$x = C + \int_a^g \frac{At + B}{\sqrt{(1 + t^2) [(1 + t^2) - (At + B)^2]}} dt .$$

Il s'agit d'une intégrale elliptique.

2.3.3. Les classes de solutions.

Il faudrait ici faire comme Euler et ses successeurs pour les *elasticæ* :

1. classer les solutions réelles en faisant varier les paramètres dans le plan (A, B) ;
2. considérer les solutions dans le domaine complexe, la disparition de solutions réelles dans certains domaines pour certaines valeurs des paramètres signifiant que les solutions deviennent complexes à la traversée de frontières.

La tâche complète serait trop longue, trop fastidieuse et de peu d'intérêt pour la modélisation. Mais nous en mènerons néanmoins à bien une petite partie. Remarquons qu'il y a deux classes extrêmes : soit $A = 0$ et $B \neq 0$, soit $B = 0$ et $A \neq 0$. Nous traiterons d'abord dans la section suivante le cas $A = 0, B \neq 0$ qui est celui qui nous intéresse. Puis nous dirons quand même un mot du cas $B = 0, A \neq 0$, même s'il n'est pas pertinent pour la modélisation des contours illusoires.

3. Intégration numérique dans le cas symétrique $A = 0, B \neq 0$

Le cas qui nous intéresse est celui où f est *paire*, i.e. où il y a une symétrie $x \leftrightarrow -x$ (la dérivée $f' = g$ est alors impaire et donc $f'(0) = g(0) = 0$). Cette contrainte implique $A = 0$ et donc $0 \leq B \leq \sqrt{1 + g^2}$, d'où, en supposant $B \neq 0$ (puisque la double condition $A = 0$ et $B = 0$ est impossible dans le domaine réel) et en posant $k = 1/B \geq 0$, l'équation différentielle pour $g = f'$:

$$(g')^2 = (1 + g^2) [k^2 (1 + g^2) - 1] \tag{4}$$

avec $k^2 (1 + g^2) \geq 1$ et $g(0) = 0$. Le paramètre $k \geq 0$ est lié à la courbure. En fait (cf. plus bas), $k^2 - 1 = \kappa(0)^2$. Comme la symétrie impose $f'(0) = g(0) = 0$, on doit avoir $f''(0)^2 = g'(0)^2 = k^2 - 1 \geq 0$, et donc $k \geq 1$ (ce qui rend d'ailleurs automatique $k^2 (1 + g^2) \geq 1$). On obtient ainsi

$$x = x_1 + \int_{g_1}^g \frac{1}{\sqrt{(1+t^2)[k^2(1+t^2)-1]}} dt, \quad (5)$$

qui est une intégrale elliptique de première espèce.⁴

Notons que pour le cas limite $k = 1$, l'équation 4 devient $(g')^2 = (1+g^2)g^2$ et que 0 est un point d'inflexion de g puisque $g(0) = g'(0) = 0$. On peut l'intégrer explicitement et l'on obtient

$$g(x) = \frac{4e^{x+C}}{1-4e^{2(x+C)}}$$

où C est une constante d'intégration. Mais les conditions $g(0) = g'(0) = 0$ sont irréalisables dans le domaine réel (il faudrait $C = -\infty$) et il faut donc considérer la formule dans le domaine complexe. Les conditions y disent que les parties réelles

$$\Re(g(0)) = \Re(g'(0)) = 0.$$

Si la constante C est telle que $e^C = i$, autrement dit si

$$C = \text{Log}(i) = i\frac{\pi}{2},$$

alors

$$g(x) = \frac{4ie^x}{1+4e^{2x}},$$

les conditions sont réalisables et $\Re(g) \equiv 0$ puisque g est imaginaire pur. Il s'agit là d'un minuscule aperçu du prolongement analytique des solutions dans le domaine complexe.

À l'autre extrême, quand $k \rightarrow \infty$ (i.e. $B \rightarrow 0$), comme

$$1 + g(x)^2 = B\sqrt{1 + g(x)^2 + g'(x)^2}$$

et que $g(0) = 0$, au voisinage de 0, c'est g' qui domine avec $g'(0)^2 = k^2 - 1 \sim \frac{1}{B^2}$. La courbe g est donc très pentue, ce qui signifie que f monte quasi verticalement au-dessus de x_1 et x_2 et, très haut, forme une arche de pont entre les deux branches.

La figure 1 montre comment la solution $g = f'$ évolue lorsque k varie de 1 à 1.65 par pas de 0.05. Pour $k = 1$, $f''(0)^2 = g'(0)^2 = 0$ et la solution réelle est le segment de droite $[x_1, x_2]$. On voit la pente de la tangente aux extrémités $p_1 = \tan(\theta_1)$ et $p_2 = \tan(\theta_2)$ augmenter en module. La figure 2 montre l'évolution de l'intégrale f de g lorsque les tangentes au bord se redressent ainsi que celle des "géodésiques" legendriennes associées dans l'espace des 1-jets.

La figure 3 montre dans le plan de base \mathbb{R}^2 et dans le fibré \mathbb{V}_J comment la géodésique legendrienne correspondant à la valeur $k = 1.5$ s'insère entre l'arc de

4. Nous rencontrerons à nouveau les fonctions elliptiques au chapitre 14. On remarquera que g est périodique à cause de la périodicité des fonctions elliptiques et possède des pôles.

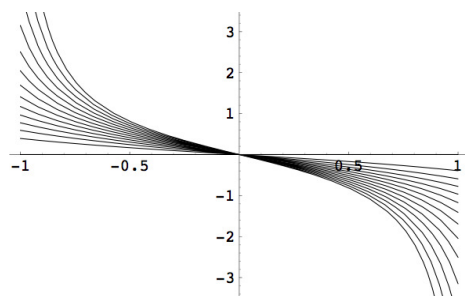


FIGURE 1. Évolution des solutions $g = f'$ lorsque le paramètre k varie de 1 à 1.65 par pas de 0.5.

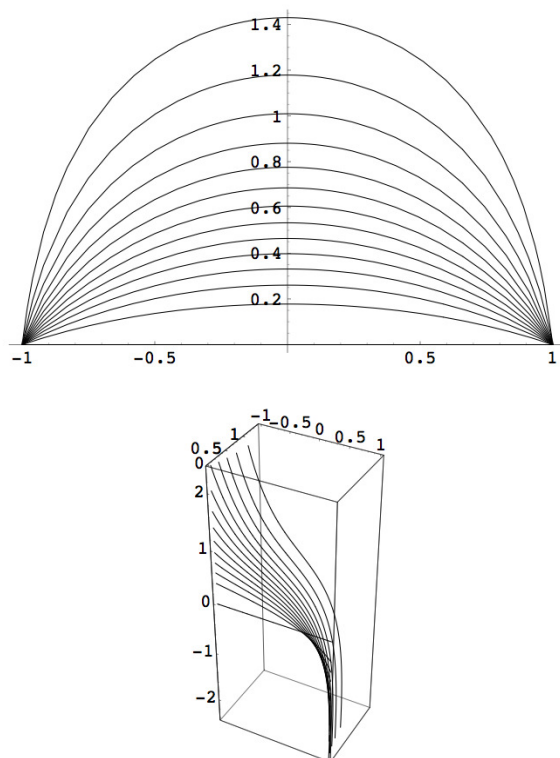


FIGURE 2. Évolution des courbes solutions f et de leurs relevées “géodésiques” pour des tangentes aux bords de plus en plus verticales.

cerle, l’arc de parabole et la solution linéaire par morceaux. Le tableau suivant montre bien la minimisation de la longueur de ces courbes gauches :

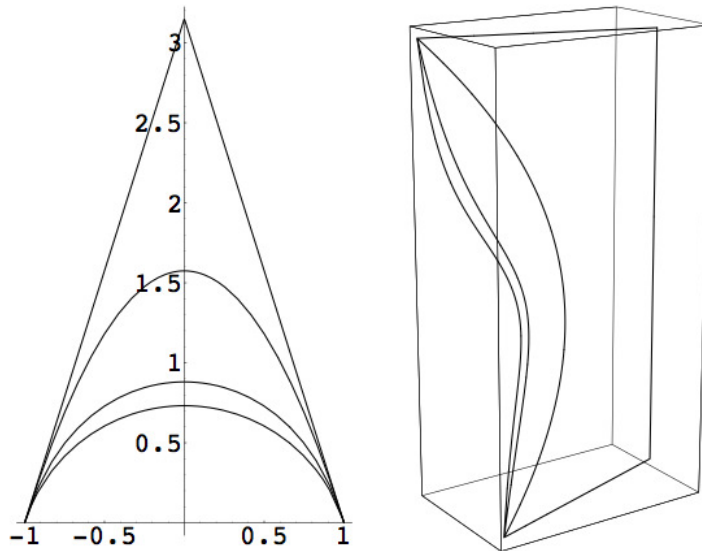


FIGURE 3. Dans cette configuration symétrique, la géodésique legendrienne est intermédiaire entre d'un côté l'arc de cercle et d'un autre côté la parabole et la solution linéaire par morceaux.

Type des courbes	Géodésique	Arc de cercle	Arc de parabole	Linéaire par morceaux
Longueur	7.02277	7.04481	7.50298	12.9054

4. Remarques sur le cas $A \neq 0, B = 0$

Disons maintenant quelques mots sur le cas particulier inverse où $A \neq 0$ et $B = 0$ même s'il n'est pas pertinent pour notre modélisation. Repartons de l'équation initiale 3 :

$$\sqrt{1 + g(x)^2 + g'(x)^2} = Ag(x) + \frac{g'(x)^2}{\sqrt{1 + g(x)^2 + g'(x)^2}} \quad (6)$$

$$1 + g(x)^2 = Ag(x)\sqrt{1 + g(x)^2 + g'(x)^2}$$

$$(g')^2 = (1 + g^2) \left(\frac{(1 + g^2)}{(Ag)^2} - 1 \right)$$

avec les conditions $Ag(x) \geq 0$ et $Ag(x) \leq \sqrt{1 + g(x)^2}$ qu'elle impose.

Pour prendre contact avec le problème, essayons $A = 1$, ce qui correspond à l'équation

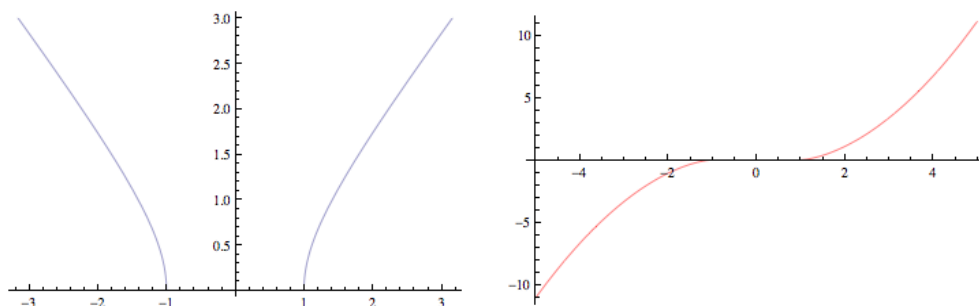


FIGURE 4. Les graphes de g et f pour les constantes d'intégrations $A = 1$, $B = 0$ et $C = 0$.

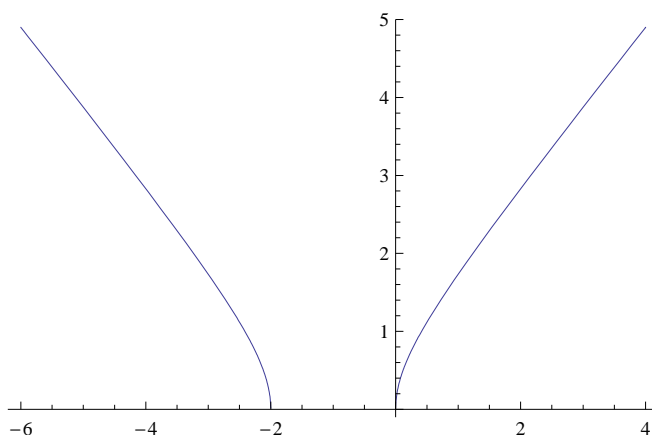


FIGURE 5. Le graphe de g pour les constantes d'intégrations $A = 1$, $B = 0$ et $C = 1$.

$$g'(x)^2 = \frac{1 + g(x)^2}{g(x)^2} .$$

Si g peut tendre vers ∞ alors $g'(x)^2 \rightarrow 1^+$ et $g'(x) \rightarrow \pm 1$. En revanche si g peut tendre vers 0, alors $g'(x) \rightarrow \pm\infty$ et g rejoint l'axe des x avec une pente verticale. On peut intégrer explicitement l'équation. On obtient, avec une constante d'intégration

$$g(x) = \pm\sqrt{-1 + x^2 \pm 2xC + C^2} .$$

Pour une branche comme $g(x) = \sqrt{-1 + (x - C)^2}$, il faut $(x - C)^2 \geq 1$ pour la partie réelle, autrement dit, $x \geq 1 + C$ ou $x \leq -1 + C$. Dans l'intervalle médian $]-1 + C, 1 + C[$ g devient imaginaire. La figure 4 montre les graphes de $g = f'$ et de f pour $C = 0$ et la figure 5 montre le graphe de g pour $C = 1$.

Remarque. L'équation (6) est intégrable explicitement dans le domaine complexe. Pour le fun, donnons la formule fournie par *Mathematica* :

$$\frac{\sqrt{-8A^2 e^{\frac{2\sqrt{1-A^2}x}{A} + 2i\sqrt{1-A^2}C} + 16e^{\frac{2\sqrt{1-A^2}x}{A} + 2i\sqrt{1-A^2}C} + e^{\frac{4\sqrt{1-A^2}x}{A}} + 16A^4 e^{4i\sqrt{1-A^2}C}}{4\sqrt{A^2-1}} \times \exp\left(\frac{1}{2}\left(-\frac{2\sqrt{1-A^2}x}{A} - 2i\sqrt{1-A^2}C\right)\right)$$

Sa complexité est digne d'attention. □

Pour $A < 1$, nous donnons quelques exemples dont les solutions ont une forme assez simple ne faisant intervenir que des sinus hyperboliques. Elles sont du type :

$$g(x) = E\sqrt{-\alpha + \beta \sinh^2(\gamma x + \delta)}.$$

Pour illustrer la situation numériquement, nous considérons les cas particuliers

$$A = \frac{m}{n}, \quad n = 8, \quad m = 1, \dots, 7.$$

Dans ce cas, $E = \frac{1}{\sqrt{n^2 - m^2}}$, $\alpha = n^2 - m^2$, $\beta = m^2$, $\gamma = \frac{\sqrt{n^2 - m^2}}{m}$, $\delta = \sqrt{n^2 - m^2}C$. La table des paramètres est donc la suivante :

A	E	α	β	γ	δ
$\frac{1}{8}$	$\frac{1}{\sqrt{63}} = \frac{1}{3\sqrt{7}}$	63	1	$3\sqrt{7}$	$3\sqrt{7}C$
$\frac{2}{8} = \frac{1}{4}$	$\frac{1}{\sqrt{60}} = \frac{1}{2\sqrt{15}}$	60	4	$\sqrt{15}$	$2\sqrt{15}C$
$\frac{3}{8}$	$\frac{1}{\sqrt{55}}$	55	9	$\frac{\sqrt{55}}{3}$	$\sqrt{55}C$
$\frac{4}{8} = \frac{1}{2}$	$\frac{1}{\sqrt{48}} = \frac{1}{4\sqrt{3}}$	48	16	$\sqrt{3}$	$4\sqrt{3}C$
$\frac{5}{8}$	$\frac{1}{\sqrt{39}}$	39	25	$\frac{\sqrt{39}}{5}$	$\sqrt{39}C$
$\frac{6}{8} = \frac{3}{4}$	$\frac{1}{\sqrt{28}} = \frac{1}{2\sqrt{7}}$	28	36	$\frac{\sqrt{7}}{3}$	$2\sqrt{7}C$
$\frac{7}{8}$	$\frac{1}{\sqrt{15}}$	15	49	$\frac{\sqrt{15}}{7}$	$\sqrt{15}C$

Pour déterminer les racines de $g(x) = 0$, savoir d'où partent les branches à tangentes verticales et quel est l'intervalle médian $]-x_0 + C, x_0 + C[$ dans lequel la solution est imaginaire, il faut considérer la contrainte $-\alpha + \beta \sinh^2(\gamma x + \delta) \geq 0$, i.e.

$$x + mC \geq \frac{m}{\sqrt{n^2 - m^2}} \operatorname{Arcsinh}\left(\frac{\sqrt{n^2 - m^2}}{m}\right)$$

$$x + mC \leq -\frac{m}{\sqrt{n^2 - m^2}} \operatorname{Arcsinh}\left(\frac{\sqrt{n^2 - m^2}}{m}\right)$$

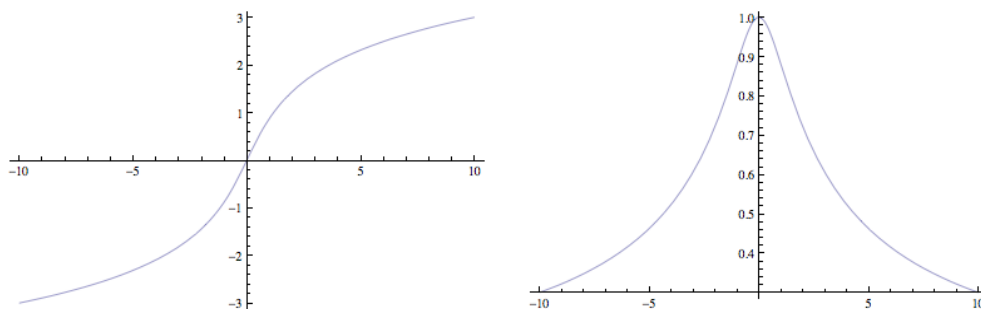


FIGURE 6. Les graphes des fonctions $\text{Arcsinh}(a)$ et $\frac{1}{a} \text{Arcsinh}(a)$.

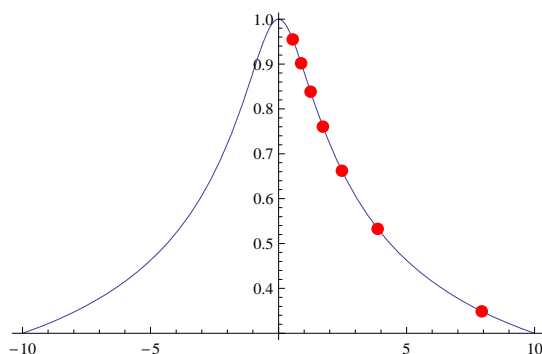


FIGURE 7. En ordonnée, les valeurs $x_0 = \frac{m}{\sqrt{n^2 - m^2}} \text{Arcsinh}\left(\frac{\sqrt{n^2 - m^2}}{m}\right)$ pour $n = 8$ et $m = 1, \dots, 7$.

La figure 6 montre les graphes de $\text{Arcsinh}(a)$ et $\frac{1}{a} \text{Arcsinh}(a)$ qui permettent de visualiser ces contraintes. Quant à la figure 7, elle montre la borne x_0 déterminant l'intervalle médian $[-x_0, x_0]$ dans lequel g est imaginaire. On voit que pour $A \rightarrow 1^-$, $x_0 \rightarrow 1^-$ et que pour $A \rightarrow 0^+$, $x_0 \rightarrow \infty$.

La figure 8 montre les solutions pour $C = 0$. On constate que, à partir de $m = 1$, l'intervalle médian s'élargit, que les branches s'évasent pour arriver au cas $A = 1$ de la figure 4.

La croissance des fonctions g augmente très rapidement lorsque A décroît vers 0 et devient linéaire de pente 1 lorsque A tend vers 1. La figure 9 montre la différence de croissance entre $A = \frac{1}{8}$ et $A = 1$, pour $x \in [0, 10]$.

5. Équation différentielle sur la courbure

Revenons au cas symétrique $A = 0$, $B \neq 0$. Par comparaison avec les elasticæ, on peut essayer d'obtenir une équation différentielle pour la courbure κ dans le plan de base. Rappelons que, si s est la longueur d'arc dans le plan de base et

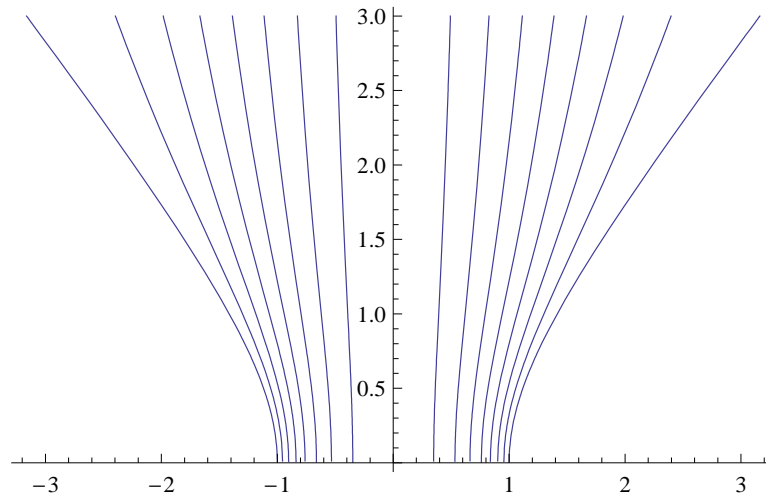


FIGURE 8. Les courbes g pour les valeurs $A = \frac{m}{n}$, $n = 8$ et $m = 1, \dots, 7$.

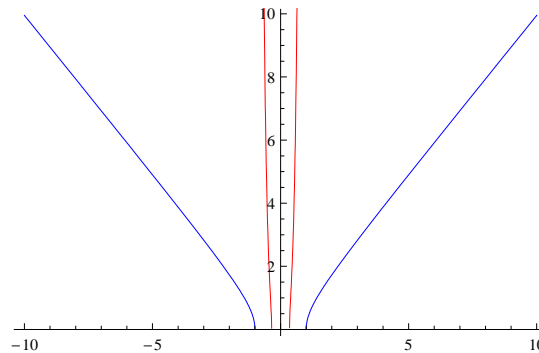


FIGURE 9. La différence de croissance entre $A = \frac{1}{8}$ et $A = 1$, pour $x \in [0, 10]$.

$a = (x, y = f(x))$ est un point sur la courbe, le vecteur tangent unitaire en a est donné par

$$t = \frac{da}{ds} = \dot{a} = \frac{da}{dx} \frac{dx}{ds} = \frac{(1, f')}{\sqrt{1 + (f')^2}}$$

(car $ds = \sqrt{1 + (f')^2} dx$).⁵ Le vecteur normal unitaire n en a est donc donné par

$$n = \frac{(-f', 1)}{\sqrt{1 + (f')^2}}$$

5. Nous notons \dot{u} les dérivées par rapport à s et u' les dérivées par rapport à x .

et, de la formule

$$\frac{dt}{ds} = \kappa n = \frac{dt}{dx} \frac{dx}{ds} = \frac{(-f'f'', f'')}{(1 + (f')^2)^{3/2}} \frac{1}{\sqrt{1 + (f')^2}} = \frac{(-f'f'', f'')}{(1 + (f')^2)^2},$$

on tire la formule bien connue de la courbure :

$$\kappa = \frac{f''}{(1 + (f')^2)^{3/2}}.$$

Si $x(t), y(t)$ est une paramétrisation quelconque, $\kappa = \frac{x\dot{y} - y\dot{x}}{(|\dot{x}|^2 + |\dot{y}|^2)^{3/2}}$.

Si $g = f'$, on obtient par conséquent le système d'équations :

$$\begin{cases} \kappa = \frac{g'}{(1+g^2)^{3/2}} \\ (g')^2 = (1+g^2)[k^2(1+g^2) - 1] \end{cases}.$$

On remarquera que, comme $g(0) = 0$ dans le cas symétrique, ces équations donnent $g'(0)^2 = k^2 - 1$ et donc la relation $\kappa(0)^2 = k^2 - 1$ évoquée plus haut.

Par élimination, on obtient une équation bicarrée en $\dot{\kappa} = \frac{d\kappa}{ds}$. Les calculs sont un peu fastidieux. On prend $\kappa^2 = \frac{(g')^2}{(1+g^2)^3}$ et on y remplace $(g')^2$ par sa valeur, d'où

$$\kappa^2 = \frac{k^2(1+g^2) - 1}{(1+g^2)^2}.$$

On dérive ensuite κ et la seconde équation. On obtient

$$\begin{cases} \kappa' = \frac{g''}{(1+g^2)^{3/2}} - \frac{3g(g')^2}{(1+g^2)^{5/2}} \\ \dot{\kappa} = \frac{g''}{(1+g^2)^2} - \frac{3g(g')^2}{(1+g^2)^3} \end{cases}$$

ainsi que

$$g'g'' = gg'[-1 + 2k^2(1+g^2)]$$

En calculant $g'\kappa'$ de façon à faire apparaître $g'g''$ et en remplaçant $g'g''$ par sa valeur, on peut exprimer $\kappa' = \frac{d\kappa}{dx}$ et $\dot{\kappa} = \frac{d\kappa}{ds}$ en fonction de g . On obtient

$$\dot{\kappa} = \frac{g[2 - k^2(1+g^2)]}{(1+g^2)^2}.$$

Enfin on élimine g entre κ^2 et $\dot{\kappa}$, ce qui donne l'équation bicarrée en $\dot{\kappa}$:

$$\begin{aligned} \dot{\kappa}^4 + \dot{\kappa}^2(2\kappa + k^2)(2\kappa - k^2)(2\kappa^2 - k^2(k^2 - 1)) \\ + \kappa^2(2\kappa + k^2)^2(2\kappa - k^2)^2(\kappa^2 - (k^2 - 1)) = 0 \end{aligned} \quad (7)$$

dont le discriminant Δ est

$$\Delta = (k^4 - 4\kappa^2)^3 (k^2 - 1)^2.$$

Pour que les solutions soient réelles, il faut que le discriminant Δ soit ≥ 0 , i.e. $-k^2 \leq 2\kappa \leq k^2$. Vu les conditions $k^2 - 1 \geq 0$ (cf. ci-dessus) et $\Delta \geq 0$, ces solutions s'écrivent :⁶

$$\kappa^2 = \frac{1}{2} (k^4 - 4\kappa^2) \left[2\kappa^2 - k^2 (k^2 - 1) \pm (k^2 - 1) \sqrt{k^4 - 4\kappa^2} \right]. \quad (8)$$

Mais κ^2 est réel ≥ 0 . Vu l'expression de l'équation 7, le signe du produit Π des racines est celui de $\kappa^2 - (k^2 - 1)$ et celui de la somme Σ est celui de $2\kappa^2 - k^2 (k^2 - 1)$. Il faut au moins une racine ≥ 0 .

On obtient ainsi une stratification⁷ de domaines d'admissibilité dans le quadrant positif \mathfrak{K} des $(X = k^2 \geq 0, Y = \kappa^2 \geq 0)$ avec les séparatrices

- (i) $k^2 = 1$ (droite verticale : $X = 1$),
- (ii) $4\kappa^2 = k^4$ (parabole P_1 : $Y = \frac{1}{4}X^2$),
- (iii) $\kappa^2 = k^2 - 1$ (droite D : $Y = X - 1$),
- (iv) $2\kappa^2 = k^2 (k^2 - 1)$ (parabole P_2 : $Y = \frac{1}{2}(X(X - 1))$).

Pour chaque valeur de $X \geq 1$ (i.e. de k puisque $k \geq 1$) on a les deux trajectoires $\kappa(s)$ solutions de 8 qui doivent rester dans le domaine d'admissibilité et subissent des accidents lorsqu'elles en atteignent les bords.

La figure 10 montre la stratification du quadrant \mathfrak{K} . Les points admissibles doivent être au-dessous de la parabole P_1 qui possède une tangente horizontale en 0, passe au point $(1, \frac{1}{4})$, et passe en $(2, 1)$ avec une pente 1. Comme la droite D passe aussi en $(2, 1)$, D est la tangente à P_1 en ce point et P_1 est située entièrement au-dessus de D . Si le produit Π est < 0 , alors il y a nécessairement une racine > 0 . Pour cela il faut $\kappa^2 - (k^2 - 1) < 0$ et le point $(X = k^2, Y = \kappa^2)$ doit être situé au-dessous de D . Il est alors entièrement au-dessous de P_1 . D'où la région admissible bleue de la figure. Mais des points admissibles peuvent également être situés au-dessus de D (mais au-dessous de P_1). Cela signifie que $\Pi > 0$ et donc que les deux racines sont de même signe. Pour qu'il y en ait une > 0 les deux doivent être > 0 et donc il faut $\Sigma > 0$, et donc $2\kappa^2 \geq k^2 (k^2 - 1)$. Le point (X, Y) doit par conséquent être au-dessus de la parabole P_2 qui passe en $(1, 0)$ avec la pente $-\frac{1}{2}$, en $(\frac{3}{2}, \frac{3}{8})$ avec la pente 1 et en $(2, 1)$ avec la pente $\frac{3}{2}$. Après ce point, P_2 passe au-dessus de P_1 , ce qui est interdit. Il y a donc des points admissibles pour $X \in (1, 2)$ situés au-dessus de P_2 et au-dessous de P_1 . Comme dans cette région D est au-dessus de P_2 , c'est la

6. Rappelons la bonne vieille formule $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ donnant les racines de l'équation du second degré $ax^2 + bx + c = 0$.

7. Nous précisons le terme de "stratification" dans le chapitre 15 à la section 1.4. Ici il s'agit simplement de la décomposition d'un domaine en sous-domaines par des séparatrices.

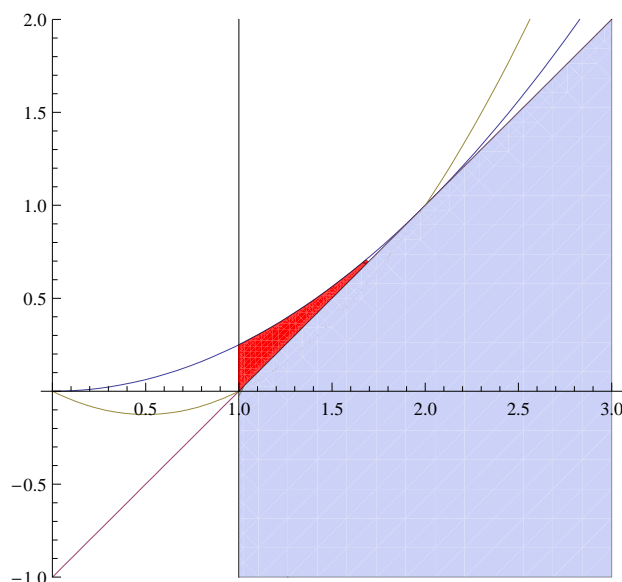


FIGURE 10. Les domaines d’admissibilité des solutions de l’équation (7) dans le quadrant positif du plan $(X = k^2, Y = \kappa^2)$. Voir le texte pour les détails.

contrainte d’être au-dessus de D qui l’emporte. D’où la région admissible rouge de la figure 10.

Pour le cas limite $k = 1$, on a $\Delta = 0$ et donc une racine double $\dot{\kappa}^2 = (1 - 4\kappa^2)\kappa^2$, κ^2 pouvant varier de 0 à $\frac{1}{4}$. Pour le cas limite P_1 , $4\kappa^2 = k^4$, $\Delta = 0$ et $\dot{\kappa}^2 = 0$. Pour le cas limite P_2 , $2\kappa^2 = k^2(k^2 - 1)$, $\Sigma = 0$ et les deux racines sont opposées et pour le cas limite D , $\kappa^2 = k^2 - 1$, $\Pi = 0$ et l’une des racines est nulle.

Pour $k^2 = 2$, on obtient l’équation simplifiée (avec $\kappa^2 \leq 1$) avec la solution positive :

$$\dot{\kappa}^2 = 4(1 - \kappa^2) \left[-(1 - \kappa^2) + \sqrt{1 - \kappa^2} \right] .$$

On pourra la comparer à celle des elasticæ 4 donnée à la section 3.4 du chapitre 8 :

$$\dot{\kappa}^2 + \frac{1}{4}\kappa^4 = \frac{b}{2}\kappa^2 + c .$$

6. Les équations d’Euler-Lagrange (II)

Pour s’exercer, on peut reformuler les calculs précédents de façon un peu différente en exprimant la 1-forme $Ldx = L_0 ds$ en termes de longueur d’arc $ds = \sqrt{dx^2 + dy^2}$ dans le plan de base \mathbb{R}^2 et de l’angle de la tangente θ . Comme $f' =$

$\tan(\theta)$ et $f'' = \frac{\theta'}{\cos^2(\theta)}$, $\kappa = \dot{\theta} = \frac{d\theta}{ds}$, on vérifie immédiatement que

$$L_0 = \left(1 + \dot{\theta}^2 \sec^4(\theta)\right)^{\frac{1}{2}}.$$

En effet,

$$\begin{aligned} L_0^2 ds^2 &= \left(1 + \dot{\theta}^2 \sec^4(\theta)\right) ds^2 = \left(1 + (f')^2\right) dx^2 + \left(\frac{d\theta}{ds}\right)^2 \sec^4(\theta) ds^2 \\ &= \left(1 + (f')^2\right) dx^2 + (\theta')^2 \sec^4(\theta) dx^2 = \left(1 + (f')^2 + (f'')^2\right) dx^2 = L^2 dx^2. \end{aligned}$$

Un calcul trivial donne alors :

$$\frac{\partial L_0}{\partial \theta} - \frac{d}{ds} \left(\frac{\partial L_0}{\partial \dot{\theta}} \right) = - \frac{\sec^4(\theta)}{\left(1 + \dot{\theta}^2 \sec^4(\theta)\right)^{\frac{3}{2}}} \left(\ddot{\theta} + 2\dot{\theta}^2 \tan(\theta) \right). \quad (9)$$

On veut minimiser la longueur $\int_{s_1=0}^{s_2} L_0 ds$ avec les conditions aux limites

$$\theta(0) = \theta_1, \theta(s_2) = \theta_2$$

et la contrainte intégrale

$$x_2 = \int_0^{s_2} \cos(\theta(s)) ds = \text{cste}$$

exprimant que les extrémités x_1 et x_2 sont fixes. On est par suite conduit à introduire un multiplicateur de Lagrange λ et à considérer l'équation d'E-L pour le lagrangien $L_1 = L_0 + \lambda \Sigma$ avec $\Sigma = \cos(\theta)$. On trouve :

$$\frac{\partial L_0}{\partial \theta} - \frac{d}{ds} \left(\frac{\partial L_0}{\partial \dot{\theta}} \right) = \lambda \sin(\theta) \quad (10)$$

avec la contrainte au bord $\frac{\partial L_1}{\partial \dot{\theta}}(s_2) = \frac{\partial L_0}{\partial \dot{\theta}}(s_2) = 0$ venant du fait que la valeur en l'extrémité s_2 n'est pas fixée.

Cette équation d'E-L admet une formulation intégrale qui se calcule de la même façon qu'à la section précédente. On part de la dérivée totale :

$$\frac{dL_0}{ds} = \frac{\partial L_0}{\partial s} + \dot{\theta} \frac{\partial L_0}{\partial \theta} + \ddot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} = \dot{\theta} \frac{\partial L_0}{\partial \theta} + \ddot{\theta} \frac{\partial L_0}{\partial \dot{\theta}}$$

et de la dérivée du produit :

$$\frac{d}{ds} \left(\dot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} \right) = \dot{\theta} \frac{d}{ds} \left(\frac{\partial L_0}{\partial \dot{\theta}} \right) + \ddot{\theta} \frac{\partial L_0}{\partial \dot{\theta}}.$$

Mais d'après l'équation (10) on a :

$$\dot{\theta} \frac{d}{ds} \left(\frac{\partial L_0}{\partial \dot{\theta}} \right) = \dot{\theta} \left(\frac{\partial L_0}{\partial \theta} - \lambda \sin(\theta) \right)$$

et donc :

$$\frac{d}{ds} \left(\dot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} \right) = \dot{\theta} \left(\frac{\partial L_0}{\partial \theta} - \lambda \sin(\theta) \right) + \ddot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} .$$

Comme $\dot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} + \ddot{\theta} \frac{\partial L_0}{\partial \ddot{\theta}} = \frac{dL_0}{ds}$, on obtient en définitive :

$$\frac{d}{ds} \left(L_0 - \dot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} \right) = \lambda \dot{\theta} \sin(\theta) = -\lambda \frac{d}{ds} (\cos(\theta))$$

d'où, à une constante près,

$$L_0 - \dot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} = -\lambda \cos(\theta) . \quad (11)$$

Les équations (10) et (11) impliquent l'équation différentielle :

$$\left[\frac{\partial L_0}{\partial \theta} - \frac{d}{ds} \left(\frac{\partial L_0}{\partial \dot{\theta}} \right) \right]^2 + \left[L_0 - \dot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} \right]^2 = \lambda^2$$

soit :

$$(1 + \kappa^2 \sec^4(\theta))^2 + \sec^8(\theta) (\dot{\kappa} + 2\kappa^2 \tan(\theta))^2 = \lambda^2 (1 + \kappa^2 \sec^4(\theta))^3 . \quad (12)$$

7. Le modèle du fibré en cercles

On peut aussi appliquer les calculs précédents au cas, plus symétrique, de la fibration de fibre \mathbb{S}^1 dont le fibré tangent a pour coordonnées locales $(x, y, \theta; \xi, \eta, \varphi)$ et dont la métrique est donnée par :

$$ds^2 = dx^2 + dy^2 + d\theta^2 .$$

On doit alors partir du lagrangien :

$$L(x) = \sqrt{\xi^2 + \eta^2 + \varphi^2} = \sqrt{1 + f'(x)^2 + \frac{f''(x)^2}{(1 + f'(x)^2)^2}} .$$

En effet $\theta = \text{Arctan}(f')$ et donc

$$\varphi = \theta' = \dot{\theta} \frac{ds}{dx} = \frac{f''}{(1 + (f')^2)^{3/2}} (1 + (f')^2)^{1/2} = \frac{f''(x)}{1 + f'(x)^2}$$

La contrainte cinématique s'écrit maintenant $\Sigma = 0$, avec $\Sigma = \theta - \text{Arctan}(\eta)$, et les équations d'E-L donnent :

$$\begin{cases} \frac{d}{dx} \left[\frac{\partial L}{\partial \eta} + \lambda \frac{\partial \Sigma}{\partial \eta} \right] = 0 \\ \lambda \frac{\partial \Sigma}{\partial \theta} - \frac{d}{dx} \frac{\partial L}{\partial \varphi} = 0 \end{cases}$$

avec $\frac{\partial \Sigma}{\partial \theta} = 1$ et $\frac{\partial \Sigma}{\partial \eta} = -\frac{1}{1+\eta^2}$. On obtient donc :

$$\frac{\partial L}{\partial f'} = A - \lambda \frac{\partial \Sigma}{\partial \eta} = A - \left(\frac{\partial \Sigma}{\partial \theta} \right)^{-1} \frac{d}{dx} \left(\frac{\partial L}{\partial \varphi} \right) \frac{\partial \Sigma}{\partial \eta}.$$

On développe alors comme plus haut la dérivée totale :

$$\frac{dL}{dx} = \frac{\partial L}{\partial x} + \frac{\partial L}{\partial f} f' + \frac{\partial L}{\partial f'} f'' + \frac{\partial L}{\partial \varphi} \varphi' = \frac{\partial L}{\partial f'} f'' + \frac{\partial L}{\partial \varphi} \varphi',$$

on écrit :

$$\frac{d}{dx} \left(\varphi \frac{\partial L}{\partial \varphi} \right) = \varphi' \frac{\partial L}{\partial \varphi} + \varphi \frac{d}{dx} \left(\frac{\partial L}{\partial \varphi} \right) = \varphi' \frac{\partial L}{\partial \varphi} + \varphi \left(A - \frac{\partial L}{\partial f'} \right) \left(\frac{\partial \Sigma}{\partial \eta} \right)^{-1},$$

d'où :

$$\frac{dL}{dx} = \frac{d}{dx} \left(\varphi \frac{\partial L}{\partial \varphi} \right) + \frac{\partial L}{\partial f'} f'' - \varphi \left(A - \frac{\partial L}{\partial f'} \right) \left(\frac{\partial \Sigma}{\partial \eta} \right)^{-1}.$$

Mais comme $\frac{\partial \Sigma}{\partial \eta} = -\frac{1}{1+\eta^2}$, $\left(\frac{\partial \Sigma}{\partial \eta} \right)^{-1} = -(1+\eta^2)$, on a $-\varphi \left(\frac{\partial \Sigma}{\partial \eta} \right)^{-1} = f''$ et donc :

$$\frac{d}{dx} \left(L - \varphi \frac{\partial L}{\partial \varphi} \right) = A f'',$$

équation à comparer à l'équation (2). L'intégration donne :

$$L - \varphi \frac{\partial L}{\partial \varphi} = A f' + B.$$

Un calcul analogue à celui effectué précédemment donne alors (avec $g = f'$) :

$$(g')^2 = \frac{(1+g^2)^4 - (1+g^2)^3 (Ag+B)^2}{(Ag+B)^2},$$

équation à comparer à l'équation précédente (3). Dans le cas symétrique $A = 0$, on obtient (avec $k = 1/B$) :

$$(g')^2 = (1+g^2)^3 [k^2 (1+g^2) - 1] \quad (13)$$

à comparer à l'équation (4).

Les figures 11 et 12 comparent les solutions de cette équation (sur l'intervalle $(-0.56, 0.56)$ car des singularités apparaissent sur des intervalles plus grands) avec celles de l'équation (4). Quant à la figure 13 elle compare les solutions des deux équations pour les mêmes conditions aux bord $p_1 = -p_2 = 0.6495$. Le module k change d'une courbe à l'autre : il est de 1.4 pour la première (la plus haute) et de 1.32745 pour la seconde (la plus basse).

On voit que pour des conditions aux limites acceptables (i.e. des pentes pas trop accentuées) les solutions des deux équations sont voisines. Mais elles se séparent

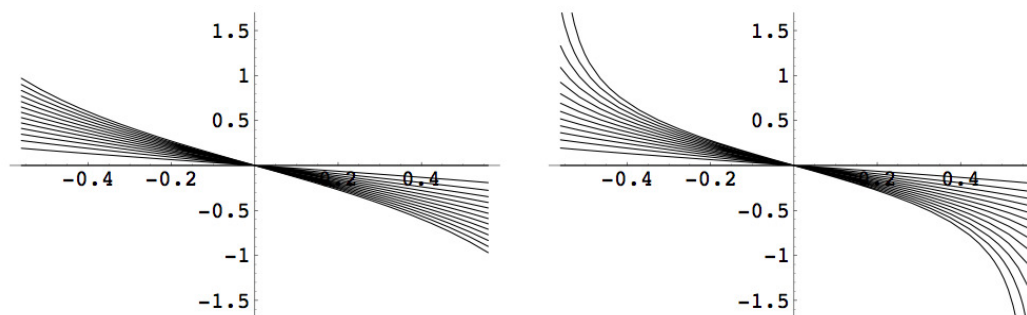
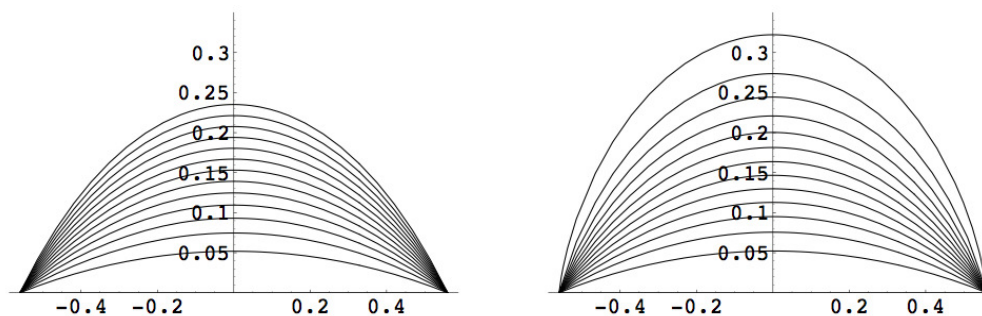
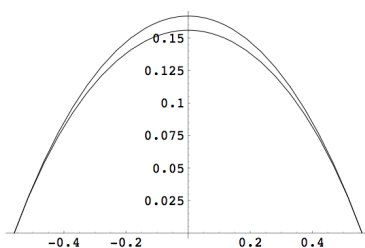
FIGURE 11. Comparaison des familles des solutions (en $g = f'$) des équations (4) et (13).

FIGURE 12. Comparaison des courbes intégrales des équations (4) et (13).

FIGURE 13. Comparaison des courbes intégrales des équations (4) et (13) pour les mêmes conditions aux limites $p_1 = -p_2 = 0.6495$.

nettement lorsque les pentes aux limites s'accroissent. Par exemple pour un module de $k = 2$ dans l'équation (13) on a $p_1 = -p_2 = 6.594$, ce qui correspond à un module de $k = 3.3423$ dans l'équation (4). La figure 14 montre la comparaison des deux solutions. On voit qu'elles divergent notablement et que celle de la première équation monte beaucoup plus haut que celle de la seconde.

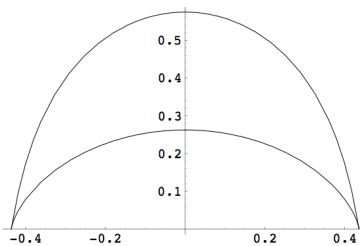


FIGURE 14. Comparaison des courbes intégrales des équations (4) et (13) pour des orientations aux limites suffisamment “verticales”. On voit qu’elles se séparent nettement.

Évidemment, un départage empirique peut paraître sans espoir puisque ces conditions sont précisément celles pour lesquelles les contours subjectifs disparaissent. Mais il existe d’autres possibilités d’expérience. La *motricité* et le mouvement peuvent prendre le relais de la perception et l’on peut demander à des sujets de tracer manuellement des extrapolations.⁸ Il semble que dans ce cas le premier modèle soit nettement plus adéquat.

En ce qui concerne l’équation différentielle sur la courbure, elle est beaucoup plus simple dans le cas du fibré en cercles. On obtient en effet :

$$\begin{cases} \kappa^2 = \frac{(g')^2}{(1+g^2)^3} = \frac{(1+g^2)^3 [k^2(1+g^2)-1]}{(1+g^2)^3} = k^2(1+g^2) - 1 \\ \dot{\kappa} = k^2 g(1+g^2) \end{cases}$$

et l’élimination de g donne l’équation relativement simple :

$$\dot{\kappa}^2 k^2 = (\kappa^2 + 1 - k^2) (1 + \kappa^2)^2, \quad (14)$$

à comparer à l’équation (7).

Cette équation peut être intégrée explicitement. On obtient (C est une constante d’intégration) :

$$\kappa(s) = \pm \left[\frac{(e^{2s} - e^{2C})^2 (k^2 - 1)}{(e^{2s} - e^{2C})^2 - k^2 (e^{2s} + e^{2C})^2} \right]^{\frac{1}{2}}.$$

8. Les courbes de moindre énergie

L’intégration de l’équation différentielle des géodésiques legendriennes est rendue difficile par la présence de racines carrées. Cet obstacle disparaît lorsque l’on considère une autre fonctionnelle dont le lagrangien, couramment appelée “énergie”, est le carré de celui des géodésiques. Les courbes qui minimisent cette énergie n’ont pas d’interprétation géométrique aussi simple que les géodésiques legendriennes, mais elles en sont généralement proches et on peut les expliciter.

8. Nous dirons un mot sur cette motricité à la fin du volume, chapitre 18 section 3.3.

Le nouveau lagrangien s'écrit :

$$L(x) = \xi^2 + \eta^2 + \pi^2 = 1 + f'(x)^2 + f''(x)^2 .$$

Le problème variationnel sous contrainte admet le même système d'équations que précédemment, ce qui donne :

$$\begin{cases} \frac{d}{dx} [2f'(x) - \lambda(x)] = 0 \\ \lambda(x) - \frac{d}{dx} [2f''(x)] = 0 \end{cases}$$

En intégrant la première équation et en éliminant $\lambda(x)$, il vient pour une certaine constante d'intégration C :

$$f'(x) = \frac{1}{2}\lambda(x) + C = f'''(x) + C .$$

Cette équation différentielle du second ordre en f' s'intègre facilement. f est de la forme :

$$f(x) = Ae^x + Be^{-x} + Cx + D$$

où A, B, C, D sont des constantes d'intégration fixées par les conditions aux limites (puisque celles-ci imposent quatre équations indépendantes : valeurs de f et de f' en deux points). La figure 32 en a donné un exemple.

Des courbes analogues sont évoquées par David Mumford dans son étude sur les *elasticæ* où il utilise en dimension 3 le lagrangien :

$$L(x) = \gamma + \beta (1 + z f'(x)^2) + \alpha f''(x)^2 .$$

9. Comparaison avec les *elasticæ*

Nous avons vu que les géodésiques legendriennes satisfont une équation différentielle en la courbure κ qui est relativement complexe, en tout cas plus complexe que celle des *elasticæ*. Mais, réciproquement, les *elasticæ* sont solutions d'un problème variationnel qui, exprimé dans le fibré de contact $C\mathbb{R}^2$ ou le fibré des 1-jets $J^1\mathbb{R}^2$, est à son tour nettement plus complexe que celui des géodésiques legendriennes.

La courbure locale d'une courbe d'équation $y = f(x)$ est donnée, nous l'avons vu plusieurs fois, par $\kappa = \frac{f''(x)}{(1+f'(x)^2)^{3/2}}$. Si l'on fait $\beta = 1$ dans l'équation (1), les *elasticæ* minimisent la fonctionnelle :

$$\begin{aligned} \int_{s_A}^{s_B} (1 + \alpha \kappa^2) ds &= \int_{x_A}^{x_B} \left(1 + \alpha \frac{f''(x)^2}{(1 + f'(x)^2)^3} \right) \sqrt{1 + f'(x)^2} dx \\ &= \int_{x_A}^{x_B} \left(\sqrt{1 + f'(x)^2} + \alpha \frac{f''(x)^2}{(1 + f'(x)^2)^{5/2}} \right) dx . \end{aligned}$$

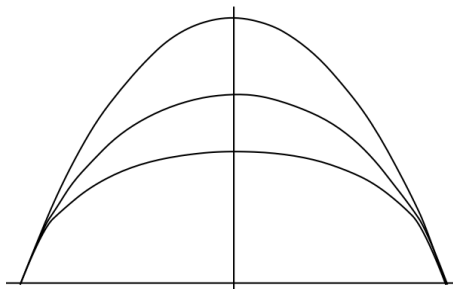


FIGURE 15. Comparaison de trois solutions variationnelles au problème d'interpolation. De haut en bas : (1) Elastica, calculé numériquement à l'aide de l'algorithme proposé par Nitzberg, Mumford et Shiota [390]. La courbe est relativement aplatie car on a plus fortement pénalisé la longueur que la courbure ($\alpha = 1/10$). (2) La projetée de la géodésique legendrienne dans l'espace des jets. (3) La projetée de la courbe d'énergie minimale dans l'espace des jets avec les mêmes conditions aux limites.

La relevée d'un elastica dans le fibré des 1-jets minimise ainsi une fonctionnelle d'action dont le lagrangien présente des analogies avec les lagrangiens de longueur et d'énergie étudiés ci-dessus. La figure 15 compare, dans un cas symétrique, les trois courbes obtenues en appliquant les trois modèles variationnels : elastica, projection d'une géodésique dans l'espace des jets et projection d'une courbe d'énergie minimale dans l'espace des jets.

Passage à $SE(2)$ et au fibré principal

1. Justification du passage

Dans le chapitre précédent nous avons développé des modèles géodésiques situés dans le modèle nilpotent \mathbb{V}_J de $V1$, celui du fibré des 1-jets des courbes planes. Mais nous avons vu au chapitre 5 que le groupe $SE(2)$ des isométries directes du plan opère naturellement sur le fibré \mathbb{V}_J et qu'il est naturel de passer au fibré principal $SE(2) = \mathbb{R}^2 \rtimes SO(2) = \mathbb{R}^2 \rtimes \mathbb{S}^1$ et de se situer dans le second modèle \mathbb{V}_S muni de sa structure de contact ω_S , modèle résoluble non nilpotent dont \mathbb{V}_J est la nilpotentisation. Il est donc naturel de s'interroger sur les possibles modèles géodésiques de contours illusoire modaux dans \mathbb{V}_S .

Nous avons esquissé très brièvement la structure de groupe de Lie de $SE(2)$ dans la section 7 du chapitre 5. Nous allons y revenir maintenant en détail afin de pouvoir lui appliquer ensuite un traitement des problèmes variationnels sur les groupes de Lie dû à Robert Bryant et Phillip Griffiths.

En fait, il s'agit de formuler les modèles variationnels comme ceux des *elasticae* ou des géodésiques legendriennes de façon géométriquement plus profonde en utilisant la méthode du *repère mobile* d'Élie Cartan exprimée par le groupe $G = SE(2)$. Cela est pertinent neurophysiologiquement à un double titre :

- d'abord on peut penser que c'est le groupe G qui se trouve neuralemement implémenté si l'on tient compte *à la fois* des aires $V1$ et $V2$. En effet, nous avons vu que dans ce cas, lorsqu'un élément de contour d'orientation donnée est activé, il en va de même de sa direction orthogonale ;
- ensuite la géométrie de G est universelle pour les problèmes de contours. Elle idéalise géométriquement une architecture fonctionnelle qui, comme l'avaient déjà remarqué profondément Poincaré et Husserl au début de ce siècle, couple la perception avec le repérage spatial et avec le sens *kinesthésique* du regard (mouvements de l'œil et de la tête).¹

1. Sur le "sens du mouvement", cf. l'ouvrage fondamental d'Alain Berthoz [53]. En ce qui concerne la modélisation géométrique de l'eidétique phénoménologique husserlienne des relations entre perception visuelle et mouvements kinesthésiques, cf. notre étude [429].

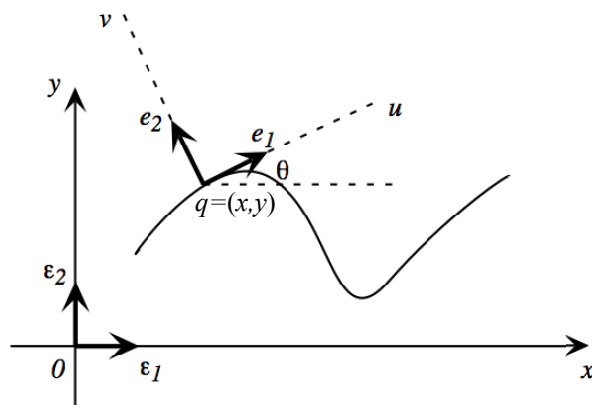


FIGURE 1. Le concept de repère mobile.

2. Les repères mobiles et le fibré principal $SE(2) = \mathbb{R}^2 \rtimes SO(2)$

Soit $(0, \varepsilon_1, \varepsilon_2)$ un repère orthonormé fixe R_0 du plan affine \mathbb{R}^2 de coordonnées (x, y) . Soit γ une courbe dans \mathbb{R}^2 . En chaque point $q = (x, y)$ de \mathbb{R}^2 ,² on peut considérer un repère orthonormé mobile (dit aussi repère de Frenet)

$$R = (q, e_1, e_2) = (q, r_\theta)$$

(noté R_q quand la référence à q est nécessaire) centré en q dont les vecteurs de base (e_1, e_2) sont les vecteurs unitaires respectivement tangent et normal à γ en q (on notera (u, v) les coordonnées par rapport à R_q) (cf. figure 1).³ r_θ est la rotation d'angle θ amenant R_0 sur R .

En termes des coordonnées (x, y, θ) on a donc :

$$\begin{cases} q &= (x, y) \\ e_1 &= (\cos(\theta), \sin(\theta)) \\ e_2 &= (-\sin(\theta), \cos(\theta)) \end{cases}$$

Le repère mobile $R = (q, r_\theta)$ équivaut en fait à un déplacement du plan affine \mathbb{R}^2 . C'est une application affine de \mathbb{R}^2 dont l'action sur un point a de \mathbb{R}^2 est donnée par

$$R(a) = (q, r_\theta)(a) = q + r_\theta(a).$$

Si $R = (q, r_\theta)$ et $S = (s, r_\varphi)$ sont deux tels déplacements, leur composition est donnée, nous l'avons vu aux sections 2.4 du chapitre 3 et 7 du chapitre 5, par le

-
2. Nous aurons parfois besoin de notations distinctes $(q, r$ ou $a)$ pour désigner des points de \mathbb{R}^2 ayant des fonctions différentes dans des contextes différents.
 3. On suppose que tous les repères ont une orientation positive.

produit semi-direct :

$$S \circ R = (s, r_\varphi) \circ (q, r_\theta) = (s + r_\varphi(q), r_{\theta+\varphi}).$$

Cette loi de groupe non commutative peut facilement s'exprimer comme une multiplication matricielle, ce qui simplifie considérablement les calculs. On utilise les nombres complexes $q = x + iy$, $r_\theta =$ (multiplication complexe par $e^{i\theta}$)⁴ et l'on introduit l'espace vectoriel $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R}^2 = \mathbb{R} \times \mathbb{C}$. Il est alors trivial de vérifier que le déplacement $R = (q, r_\theta)$ de \mathbb{R}^2 s'identifie à la restriction aux vecteurs de la forme $\begin{pmatrix} 1 \\ a \end{pmatrix}$ de l'application linéaire de $\mathbb{R} \times \mathbb{C}$ dont la matrice est $g = \begin{pmatrix} 1 & 0 \\ q & e^{i\theta} \end{pmatrix}$, la loi de produit semi-direct des déplacements s'identifiant tout simplement à la multiplication des matrices correspondantes. Il est facile de vérifier que l'inverse g^{-1} de g est donné par la formule :

$$g^{-1} = \begin{pmatrix} 1 & 0 \\ -e^{-i\theta}.q & e^{-i\theta} \end{pmatrix}.$$

où $e^{i\psi}.v$ symbolise l'action de la matrice de rotation d'angle ψ sur le vecteur colonne $v = (v_1, v_2)$.

Remarque. Il faut noter que les matrices que nous utiliserons sont des matrices 3×3 ayant une structure $\begin{pmatrix} [1 \times 1] & [1 \times 2] \\ [2 \times 1] & [2 \times 2] \end{pmatrix}$, $[m \times n]$ symbolisant une matrice à m lignes et n colonnes. Dans ce qui suit, nous essayerons de manifester cette structure autant que faire se peut en adoptant pour les notations une voie moyenne entre, d'un côté, les calculs détaillés avec des matrices et des vecteurs sans structure interne réduits à leurs composantes et, d'un autre côté, la structure complète (explicitant les espaces vectoriels et les bases mis en jeu) des applications linéaires représentées par les matrices. Les notations de cette voie en quelque sorte "semi-intrinsèque" ne seront pas toujours très satisfaisantes mais permettront néanmoins de suivre de façon économique la structure interne des calculs. □

Nous allons utiliser cette représentation très commode pour préciser ce que nous n'avons qu'esquissé à la section 7 du chapitre 5. Elle permet de bien expliciter la structure interne du groupe des déplacements du plan $SE(2)$ qui, bien qu'étant un groupe élémentaire devant être connu de tous, est plus subtil qu'il n'en a l'air. $SE(2)$ est identifiable au groupe de Lie tridimensionnel G des matrices g .

4. Rappelons que la multiplication complexe par $e^{i\theta}$ dans \mathbb{C} correspond dans \mathbb{R}^2 à la matrice de rotation $\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. Le lien avec la formule de De Moivre $e^{i\theta} = \cos(\theta) + i \sin(\theta)$ se fait, nous l'avons déjà vu, à travers l'identification standard de i avec la matrice $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ correspondant à la rotation de $\frac{\pi}{2}$.

Reprenons donc avec ces notations ce que nous savons déjà. Le stabilisateur de 0 (c'est-à-dire l'ensemble des g laissant l'origine invariante) est le sous-groupe $H = SO_2(\mathbb{R})$ des g tels que $q = 0$ (rotations planes pures sans translation) et donc des g de la forme $g = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}$. Le quotient G/H s'identifie à \mathbb{R}^2 (c'est-à-dire au plan de base) et G est le produit semi-direct $(G/H) \rtimes H$. Comme $H = SO_2(\mathbb{R})$, on trouve ainsi une fibration

$$\Pi : G \rightarrow \mathbb{R}^2 = G/H$$

ayant pour fibre un groupe opérant sur le cercle unité \mathbb{S}^1 des orientations du plan. C'est la structure de fibré principal que nous connaissons bien. Au-dessus de chaque point q de \mathbb{R}^2 , il y a un exemplaire du groupe des rotations $SO_2(\mathbb{R})$. La fibration Π opère sur la fibration $\pi : \mathbb{V} \rightarrow \mathbb{R}^2$ ayant pour fibre les orientations du plan (que celles-ci soient codées par \mathbb{P}^1 , \mathbb{S}^1 ou \mathbb{R} à travers $\tan(\theta)$) : si $q \in \mathbb{R}^2$, l'exemplaire de $SO_2(\mathbb{R})$ au-dessus de q opère sur la fibre $\mathbb{V}_q = \pi^{-1}(q)$ en faisant tourner la direction. Le fibré π est associé au fibré principal Π .

Nous sommes ainsi passés naturellement du fibré $\mathbb{V}_J = J^1\mathbb{R}^2$ au fibré principal $SE(2) = \mathbb{R}^2 \rtimes \mathbb{S}^1$ auquel il est associé.

3. Les représentations adjointe et co-adjointe de $G = SE(2)$

Dans le chapitre 5 nous avons exposé la structure de \mathbb{V}_J et de \mathcal{V}_J en termes de groupes et d'algèbres de Lie et nous avons présenté leurs représentations matricielles naturelles, les représentations adjointe et co-adjointe et la forme de Maurer-Cartan. Dans la section 7 du chapitre 5, nous avons commencé à présenter le passage à $G = SE(2)$. Dans la section 3.3 du chapitre 7 nous avons expliqué la classification des algèbres de Lie réelles de dimension 3 et nous avons vu comment $\mathcal{G} = \mathfrak{se}(2)$ s'y positionnait en tant que CLASSE DE BIANCHI VII₀.

Dans cette section, nous allons approfondir les structures de $G = SE(2)$ et $\mathcal{G} = \mathfrak{se}(2)$ en utilisant les représentations matricielles ci-dessus.

3.1. L'algèbre de Lie de $G = SE(2)$

Calculons la structure d'algèbre de Lie sur l'espace tangent $\mathcal{G} = T_e G$ de G en l'origine e .⁵ Cette origine e est l'identité de \mathbb{R}^3 , $e = Id_{\mathbb{R}^3} = \begin{pmatrix} 1 & 0 \\ 0 & I_2 \end{pmatrix}$ (où I_2 est la matrice identité 2×2) et $T_e G$ est l'espace vectoriel tridimensionnel des

5. $e = I_3$ est la matrice identité 3×3 , à ne pas confondre évidemment avec le e de l'exponentielle.

matrices “infinitésimales” $\delta h = \begin{pmatrix} 0 & 0 \\ \delta s & i\delta\varphi \end{pmatrix}$.⁶ En effet, puisque $e^{i\delta\varphi} \simeq 1 + i\delta\varphi$ au premier ordre (avec $i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$), l’élément h correspondant à δs et $\delta\varphi$ s’écrit

$h = \begin{pmatrix} 1 & 0 \\ \delta s & e^{i\delta\varphi} \end{pmatrix} = e + \delta h$ au premier ordre. On remarquera que, conformément à la structure de produit semi-direct de $G = (G/H) \rtimes H$, l’espace vectoriel $\mathcal{G} = \{\delta h\}$ se décompose en somme directe $\mathcal{G} = \mathcal{H} \oplus \mathcal{S}$ avec $\mathcal{H} = \{\delta h | \delta s = 0\}$ (rotations infinitésimales $i.\delta\varphi$) et $\mathcal{S} = \mathcal{G}/\mathcal{H} = \{\delta h | \delta\varphi = 0\}$ (translations infinitésimales δs).

La structure d’algèbre de Lie de \mathcal{G} est donnée par les commutateurs des matrices δh . Ainsi on représente à la fois G et \mathcal{G} dans l’algèbre $M_3(\mathbb{R})$ des matrices 3×3 opérant sur $\mathbb{R}^3 = \mathbb{R} \times \mathbb{C}$. G ne fait intervenir que la structure multiplicative de $M_3(\mathbb{R})$. La structure additive n’intervient pas car $1 + 1 = 2 \neq 1$ et $e^{i\theta} + e^{i\varphi}$ n’est pas en général de la forme $e^{i\psi}$. De même, \mathcal{G} ne fait intervenir que la structure d’algèbre de Lie de $M_3(\mathbb{R})$ car $i\delta\varphi\delta s'$ n’est pas de la forme δr et $(i\delta\varphi).(i\delta\varphi') = -\delta\varphi\delta\varphi'$ n’est pas de la forme $i\delta\psi$.

Soient $\varkappa_1 = \begin{pmatrix} 0 & 0 \\ 1_1 & 0 \end{pmatrix}$, $\varkappa_2 = \begin{pmatrix} 0 & 0 \\ 1_2 & 0 \end{pmatrix}$ (où 1_1 est le vecteur $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ et 1_2 est le vecteur $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$) et $\tau = \begin{pmatrix} 0 & 0 \\ 0 & i \end{pmatrix}$. Notons \varkappa le vecteur ligne de matrices $\varkappa = (\varkappa_1, \varkappa_2) \in \mathcal{G} \times \mathcal{G}$.⁷ On a

$$\delta h = \langle \varkappa, \delta s \rangle + \delta\varphi.\tau$$

(où $\langle \varkappa, \delta s \rangle$ fonctionne comme un “produit scalaire” $\delta s_1.\varkappa_1 + \delta s_2.\varkappa_2$)⁸ et (\varkappa, τ) constitue une base de \mathcal{G} . Si l’on considère les commutateurs $[\delta k, \delta h] = \delta k.\delta h - \delta h.\delta k$, il est trivial de vérifier les relations de commutation :

$$\begin{cases} [\varkappa_i, \varkappa_j] = [\varkappa_i, \varkappa_j] = [\tau, \tau] = 0 \\ [\varkappa, \tau] = -\varkappa.i \end{cases}$$

6. Pour ne pas trop alourdir la présentation et les notations, nous ferons comme certains physiciens et traiterons de façon intuitive les vecteurs tangents comme des perturbations infinitésimales du premier ordre (en utilisant la notation δ). Quand cela sera nécessaire nous reviendrons à leur traitement rigoureux comme vecteurs.

7. Il s’agit donc d’un vecteur ligne 1×2 dont les deux composantes sont des matrices 3×3 .

8. Pour désambigüiser certaines notations, nous utiliserons aussi parfois des points pour indiquer la multiplication par des scalaires : $\delta s_i.\xi_i$ est le produit de la matrice ξ_i par le scalaire δs_i . Les notations ne sont pas très satisfaisantes mais une formulation intrinsèque serait bien plus lourde.

car $\varkappa.\tau = 0$ et $\tau.\varkappa = \varkappa.i$ ($[\varkappa, \tau]$ signifiant ($[\varkappa_1, \tau], [\varkappa_2, \tau]$) et $\varkappa.i$ signifiant $(\varkappa_2, -\varkappa_1)$).⁹ On vérifie également trivialement la relation caractéristique des crochets de Lie :

$$[\delta h, [\delta k, \delta m]] + [\delta k, [\delta m, \delta h]] + [\delta m, [\delta h, \delta k]] = 0 .$$

On vérifie que, comme nous l'avons vu à la section 3.3 du chapitre 4.4, l'algèbre de Lie \mathcal{G} n'est pas nilpotente car, quel que soit le crochet d'ordre n , ses composantes non nulles seront proportionnelles à $\varkappa.i^n$ qui est non nul.

3.2. Changements de repère et automorphismes intérieurs

Nous allons maintenant développer pour $SE(2)$ l'équivalent de ce que nous avons vu à la section 6 du chapitre 5 à propos des représentations adjointes et coadjointes.

Si $g, h \in G$, $g = (q, r_\theta)$ et $h = (s, r_\varphi)$, la translation à gauche $L_g : h \mapsto gh$ s'écrit :

$$L_g(h) = (q + r_\theta(s), r_{\theta+\varphi}) = \begin{pmatrix} 1 & 0 \\ q + e^{i\theta}.s & e^{i(\theta+\varphi)} \end{pmatrix} .$$

Remarque. Les translations à gauche dans G sont pour nous essentielles car elles peuvent être interprétées *kinesthésiquement* (en particulier par les mouvements de l'œil et de la tête). Le contrôle moteur de la vision permet en effet de changer de repère mobile (cf. Berthoz [53]). □

L'application linéaire tangente $T_e L_g$ de L_g en e permet de transporter canoniquement la structure de $\mathcal{G} = T_e G$ sur $T_g G$. On a $T_g G = \{g + \delta g\}$ avec des δg de la forme $\delta g = \begin{pmatrix} 0 & 0 \\ \delta q & e^{i\theta}.i\delta\varphi \end{pmatrix}$ au premier ordre.¹⁰ Pour calculer $T_e L_g$, on écrit

$$L_g(e + \delta h) = g(e + \delta h) = g + g.\delta h = g + \delta g ,$$

ce qui donne

$$\delta g = T_e L_g(\delta h) = g.\delta h .$$

D'où $T_e L_g = g$ (au sens de la multiplication des matrices).¹¹ Si $\delta h = \begin{pmatrix} 0 & 0 \\ \delta s & i\delta\varphi \end{pmatrix}$,

$$\delta g = g.\delta h = \begin{pmatrix} 0 & 0 \\ e^{i\theta}.\delta s & e^{i\theta}.i\delta\varphi \end{pmatrix}$$

et donc le δg est $e^{i\theta}.\delta s$ où, répétons-le, $e^{i\theta}.\delta s$ est l'action de la matrice 2×2 de rotation $e^{i\theta}$ sur le vecteur $\delta s = \begin{pmatrix} \delta s_1 \\ \delta s_2 \end{pmatrix}$.

9. Comme ξ est un vecteur ligne, son produit avec la matrice i est $\xi.i$ et pas $i.\xi$.
 10. On notera la présence du coefficient $e^{i\theta}$ qui vient du fait que δg est un vecteur tangent en g . Lorsque $g = e$ (i.e. $\theta = 0$) ce coefficient disparaît.
 11. Évidemment, $T_e L_g \neq g$ puisque ce sont des éléments d'ensembles différents. Mais $T_e L_g$ étant une application linéaire elle est décrite par une matrice et c'est celle-ci qui est égale à g . Cela est possible parce que G est un groupe de Lie de matrices.

On peut utiliser $T_e L_g$ pour transporter la base (\varkappa, τ) de \mathcal{G} en une base (\varkappa_g, τ_g) de $T_g G$. On obtient

$$\varkappa_g = \varkappa \cdot e^{i\theta} = (\varkappa_1, \varkappa_2) \cdot e^{i\theta} = (\varkappa_{1,g}, \varkappa_{2,g})$$

et

$$\tau_g = \begin{pmatrix} 0 & 0 \\ 0 & e^{i\theta} \cdot i \end{pmatrix}$$

où $\varkappa_{1,g} = \begin{pmatrix} 0 & 0 \\ (e^{i\theta})_1 & 0 \end{pmatrix}$ avec $(e^{i\theta})_1 = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$ la première colonne de la matrice de rotation $e^{i\theta}$, et $\varkappa_{2,g} = \begin{pmatrix} 0 & 0 \\ (e^{i\theta})_2 & 0 \end{pmatrix}$ avec $(e^{i\theta})_2$ la seconde colonne.

On remarquera que dans cette base le vecteur $\delta g = \begin{pmatrix} 0 & 0 \\ e^{i\theta} \cdot \delta s & e^{i\theta} \cdot i \delta \varphi \end{pmatrix}$ de $T_g L_g$ devient le vecteur de composantes $\delta g = (\delta s, \delta \varphi)$.

Si $\delta h = (\delta s, \delta \varphi) \in \mathcal{G}$, son transport par les $T_e L_g$ définit ainsi un champ de vecteurs sur G

$$\delta g = g \cdot \delta h = \begin{pmatrix} 0 & 0 \\ e^{i\theta} \cdot \delta s & e^{i\theta} \cdot i \delta \varphi \end{pmatrix}$$

qui est G -invariant par translations à gauche. Dans les bases G -invariantes (\varkappa_g, τ_g) ce champ est de composantes *constantes* $(\delta s, \delta \varphi)$.

On peut alors revenir de g en e au moyen de la translation à droite $R_{g^{-1}}$. Les automorphismes intérieurs

$$A_g = R_{g^{-1}} \circ L_g = L_g \circ R_{g^{-1}} : h \mapsto ghg^{-1}$$

définissent une représentation A de G dans son propre groupe d'automorphismes $\text{Aut } G$:

$$\begin{aligned} A : G &\rightarrow \text{Aut } G \\ g \in G &\mapsto A_g \in \text{Aut } G \end{aligned}$$

qui s'écrit

$$\begin{aligned} A_g(h) &= ghg^{-1} = \begin{pmatrix} 1 & 0 \\ q & e^{i\theta} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ s & e^{i\varphi} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -e^{-i\theta} \cdot q & e^{-i\theta} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ q + e^{i\theta} \cdot s - e^{i(\theta+\varphi)} e^{-i\theta} \cdot q & e^{i(\theta+\varphi)} e^{-i\theta} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ q + e^{i\theta} \cdot s - e^{i\varphi} \cdot q & e^{i\varphi} \end{pmatrix}. \end{aligned}$$

3.3. La représentation adjointe

Conformément à la section 6 du chapitre 5, la représentation adjointe Ad de G dans $\text{Aut } \mathcal{G}$ est la version infinitésimale de la représentation A :

$$\begin{aligned} Ad : G &\rightarrow \text{Aut } \mathcal{G} \\ g \in G &\mapsto Ad_g = T_e A_g \in \text{Aut } \mathcal{G} \end{aligned}$$

Elle s'obtient, nous l'avons vu, en calculant l'application linéaire tangente

$$T_e A_g = Ad_g$$

de A_g en e , qui est un automorphisme de l'algèbre de Lie \mathcal{G} . Si $h = e + \delta h$, on trouve au premier ordre

$$A_g(h) = A_g(e + \delta h) = e + A_g(\delta h)$$

et on pose $Ad_g(\delta h) = A_g(\delta h)$:

$$\begin{aligned} A_g(\delta h) &= \begin{pmatrix} 0 & 0 \\ e^{i\theta} \cdot \delta s & e^{i\theta} \cdot i\delta\varphi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -e^{-i\theta} \cdot q & e^{-i\theta} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ e^{i\theta} \cdot \delta s - e^{i\theta} \cdot i\delta\varphi (e^{-i\theta} \cdot q) - e^{i\delta\varphi} \cdot q & e^{i\theta} \cdot i\delta\varphi (e^{-i\theta}) \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ e^{i\theta} \cdot \delta s - iq\delta\varphi & i\delta\varphi \end{pmatrix} \end{aligned}$$

car $e^{i\theta} \cdot i \cdot e^{-i\theta} = i$ et donc :

$$Ad_g(\delta h) = \begin{pmatrix} 0 & 0 \\ e^{i\theta} \cdot \delta s - iq\delta\varphi & i\delta\varphi \end{pmatrix} .$$

On vérifie

(i) que $Ad_g(\boldsymbol{\varkappa}) = \boldsymbol{\varkappa} \cdot e^{i\theta}$ au sens où $Ad_g(\boldsymbol{\varkappa})$ est la matrice ligne 1×2 ($Ad_g(\boldsymbol{\varkappa}_1), Ad_g(\boldsymbol{\varkappa}_2)$) de matrices 3×3 qui est le produit de la matrice ligne $\boldsymbol{\varkappa} = (\boldsymbol{\varkappa}_1, \boldsymbol{\varkappa}_2)$ par la matrice de rotation $e^{i\theta}$ et

(ii) que $Ad_g(\tau) = \boldsymbol{\varkappa} \cdot (-iq) + \tau$.

La matrice de Ad_g s'écrit donc dans la base $(\boldsymbol{\varkappa}, \tau)$ de \mathcal{G} de la façon suivante :

$$Ad_g = \begin{pmatrix} e^{i\theta} & -iq \\ 0 & 1 \end{pmatrix} .$$

Il s'agit d'une matrice de type $\begin{pmatrix} [2 \times 2] & [2 \times 1] \\ [1 \times 2] & [1 \times 1] \end{pmatrix}$ agissant sur le vecteur colonne 3×1 des composantes par rapport à la base $(\boldsymbol{\varkappa}, \tau)$ traitée comme un vecteur ligne 1×3 de matrices 3×3 . Autrement dit, si $\gamma = \begin{pmatrix} \sigma \\ \phi \end{pmatrix}$, alias $\begin{pmatrix} \delta s \\ \delta\varphi \end{pmatrix}$, est un vecteur de \mathcal{G} exprimé dans la base $(\boldsymbol{\varkappa}, \tau)$, son image par Ad_g est

$$\begin{aligned} Ad_g \begin{pmatrix} \sigma \\ \phi \end{pmatrix} &= \begin{pmatrix} e^{i\theta} & -iq \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma \\ \phi \end{pmatrix} \\ &= \begin{pmatrix} \sigma' \\ \phi' \end{pmatrix} = \begin{pmatrix} e^{i\theta} \cdot \sigma - iq\phi \\ \phi \end{pmatrix}. \end{aligned}$$

On vérifie facilement que si l'on prend maintenant l'application linéaire tangente de Ad en e , on réobtient bien les relations de commutation de \mathcal{G} :

$$\begin{aligned} T_e Ad &= ad : \mathcal{G} \rightarrow \text{End } \mathcal{G} \\ \gamma &\mapsto ad_\gamma : \varsigma \mapsto ad_\gamma(\varsigma) = [\gamma, \varsigma]. \end{aligned}$$

3.4. Les orbites de la représentation adjointe

Il est facile de calculer les orbites de la représentation adjointe, c'est-à-dire les sous-ensembles de \mathcal{G} de la forme $Ad_g(\delta h)$ pour $\delta h = (\delta s, \delta\varphi)$ fixé et g variable. On a $Ad_g(\delta h) = \delta h'$ avec $\delta\varphi' = \delta\varphi$ et $\delta s' = e^{i\theta} \cdot \delta s - iq\delta\varphi$. On remarque d'abord que Ad_g préserve la composante $\delta\varphi$ de δh et que donc les orbites sont nécessairement dans les plans de \mathcal{G} d'équation $\delta\varphi = cste$. Trois cas sont alors possibles (cf. figure 2).

- Si $\delta\varphi \neq 0$, l'équation $e^{i\theta} \cdot \delta s - iq\delta\varphi = \delta s'$ est toujours résoluble en q pour tous les θ : on prend un θ quelconque et

$$q = -\frac{1}{i\delta\varphi} (\delta s' - e^{i\theta} \delta s).$$

L'orbite de δh est alors tout le *plan* $\delta\varphi = cste$.

- Si $\delta\varphi = 0$ mais $\delta s \neq 0$, alors l'orbite est l'ensemble des $e^{i\theta} \cdot \delta s$ pour θ variable, autrement dit un *cercle*.
- Si $\delta\varphi = 0$ et $\delta s = 0$ (i.e. $\delta h = 0$) alors l'orbite est *ponctuelle* et réduite à l'origine.¹²

3.5. La représentation co-adjointe

En dualisant la construction précédente, on construit, nous l'avons vu, la représentation co-adjointe de G . On part de \mathcal{G}^* l'espace vectoriel dual de \mathcal{G} . Les éléments ς^* de \mathcal{G}^* sont donc par définition les formes linéaires sur \mathcal{G} .¹³ Si (\varkappa^*, τ^*) la base de \mathcal{G}^* duale de la base (\varkappa, τ) de \mathcal{G} (par dualité, il faut la traiter comme un vecteur colonne puisque (\varkappa, τ) est traitée comme un vecteur ligne), ς^* est de la forme $\varsigma^* = \lambda^* \varkappa^* + \mu^* \tau^*$ (les composantes (λ^*, μ^*) étant traitées comme un vecteur ligne). Si

12. Rappelons que si N est une sous-variété de dimension n d'une variété M de dimension m , la codimension de N dans M est la différence $m - n$. Si c est le nombre de conditions indépendantes définissant une orbite, celle-ci est de codimension c dans \mathcal{G} .

13. Si l'on traite \mathcal{G} comme \mathbb{R}^3 alors les formes sont \mathbb{R} -linéaires. Si on traite \mathcal{G} comme $\mathbb{C} \times \mathbb{R}$ alors les formes sont \mathbb{C} -linéaires par rapport à la première composante.

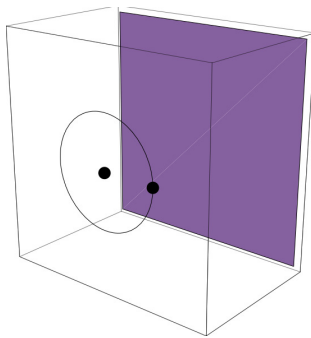


FIGURE 2. Les trois types d'orbites de la représentation adjointe : plan, cercle, point.

$\delta h = \langle \varkappa, \delta s \rangle + \delta \varphi \cdot \tau$, alors, dans les bases duales (\varkappa, τ) et (\varkappa^*, τ^*) , $\varsigma^*(\delta h)$ s'écrit comme le produit intérieur :

$$\langle \varsigma^*, \delta h \rangle = \lambda^* \cdot \delta s + \mu^* \delta \varphi .$$

La représentation co-adjointe est :

$$\begin{aligned} Ad^* : G &\rightarrow Aut \mathcal{G}^* \\ g \in G &\mapsto Ad_g^* \in Aut \mathcal{G}^* \\ \varsigma^* \in \mathcal{G}^* &\mapsto Ad_g^*(\varsigma^*) \in \mathcal{G}^* \\ Ad_g^*(\varsigma^*) : \gamma \in \mathcal{G} &\mapsto \begin{aligned} Ad_g^*(\varsigma^*)(\gamma) &= \langle Ad_g^*(\varsigma^*), \gamma \rangle \\ &= \langle \varsigma^*, Ad_{g^{-1}}(\gamma) \rangle . \end{aligned} \end{aligned}$$

Il s'agit simplement d'utiliser la dualité fournie par le produit intérieur $\langle \bullet, \bullet \rangle$ entre \mathcal{G} et \mathcal{G}^* pour réécrire par adjonction $\langle \varsigma^*, Ad_{g^{-1}}(\gamma) \rangle$ sous la forme $\langle Ad_g^*(\varsigma^*), \gamma \rangle$. On constate que Ad^* possède le type d'une représentation de G dans $Aut \mathcal{G}^*$ et l'on vérifie que c'est bien le cas.

Il est facile de calculer $Ad_g^*(\varsigma^*)$ en utilisant

$$Ad_{g^{-1}} = \begin{pmatrix} e^{-i\theta} & i \cdot e^{-i\theta} \cdot q \\ 0 & 1 \end{pmatrix} .$$

Soient $\varsigma^* = \lambda^* \cdot \varkappa^* + \mu^* \tau^* \in \mathcal{G}^*$ et $Ad_g^*(\varsigma^*) = \lambda'^* \cdot \varkappa^* + \mu'^* \tau^* \in \mathcal{G}^*$. On obtient $\lambda'^* = \lambda^* \cdot e^{-i\theta}$ et $\mu'^* = \lambda^* \cdot i e^{-i\theta} q + \mu^*$. En traitant les éléments de \mathcal{G}^* comme des covecteurs, c'est-à-dire en écrivant $(\lambda^*, \mu^*) \cdot Ad_g^* = (\lambda'^*, \mu'^*)$ (vecteurs lignes), on trouve pour la matrice de Ad_g^* dans la base (\varkappa^*, τ^*) la même matrice que $Ad_{g^{-1}}$ dans la base (\varkappa, τ) mais cette fois elle opère, par dualité, à droite sur des vecteurs lignes et non plus à gauche sur des vecteurs colonnes :

$$Ad_g^* = \begin{pmatrix} e^{-i\theta} & i \cdot e^{-i\theta} \cdot q \\ 0 & 1 \end{pmatrix} .$$

En prenant l'application linéaire tangente à Ad^* en e , on construit une représentation de l'algèbre de Lie \mathcal{G} dans l'algèbre des endomorphismes de \mathcal{G}^* , représentation qui permet de récupérer l'action de \mathcal{G}^* sur les commutateurs de \mathcal{G} :

$$\begin{aligned} ad^* : \mathcal{G} = T_e G &\rightarrow \text{End } \mathcal{G}^* = T_e(\text{Aut } \mathcal{G}^*) \\ \delta h = \gamma \in \mathcal{G} &\mapsto ad_\gamma^* \in \text{End } \mathcal{G}^* \\ \zeta^* \in \mathcal{G}^* &\mapsto ad_\gamma^*(\zeta^*) \in \mathcal{G}^* \\ ad_\gamma^*(\zeta^*) : \eta \in \mathcal{G} &\mapsto ad_\gamma^*(\zeta^*)(\eta) = \langle ad_\gamma^*(\zeta^*), \eta \rangle = \\ &\langle \zeta^*, ad_{-\gamma}(\eta) \rangle = \langle \zeta^*, -[\gamma, \eta] \rangle . \end{aligned}$$

Là encore on s'est borné à transformer par adjonction le produit intérieur $\langle \zeta^*, ad_{-\gamma}(\eta) \rangle$ en $\langle ad_\gamma^*(\zeta^*), \eta \rangle$, à constater que ad^* possède le type d'une telle représentation de \mathcal{G} dans $\text{End } \mathcal{G}^*$ et à vérifier que c'est bien le cas.

3.6. Les orbites de la représentation co-adjointe

Les orbites de la représentation co-adjointe sont les sous-ensembles de \mathcal{G}^* de la forme $Ad_g^*(\zeta_0^*)$ avec $\zeta_0^* = (\lambda_0^*, \mu_0^*)$ fixe et g variable. Comme

$$Ad_g^*(\zeta_0^*) = \lambda'^* \cdot \mathcal{Z}^* + \mu'^* \tau^*$$

avec $\lambda'^* = \lambda_0^* \cdot e^{-i\theta}$ et $\mu'^* = \lambda_0^* \cdot (i \cdot e^{-i\theta} \cdot q) + \mu_0^*$, on voit que lorsque θ varie λ'^* parcourt le cercle $C_{\lambda_0^*}$ du plan engendré par \mathcal{Z}^* centré sur l'origine et passant par $(\lambda_0^*, 0)$. Il y a deux cas (cf. figure 3) :

- Si $\lambda_0^* \neq 0$, le cercle $C_{\lambda_0^*}$ n'est pas dégénéré. L'équation en q et θ :

$$\mu'^* = \lambda_0^* \cdot (i \cdot e^{-i\theta} \cdot q) + \mu_0^*$$

est toujours résoluble en μ'^* car, si par exemple $\lambda_{0,1}^* \neq 0$, il suffit d'utiliser θ pour ramener par rotation $i \cdot e^{-i\theta} \cdot q$ à un vecteur $(r_1, 0)$ et de résoudre $\mu'^* = \lambda_{0,1}^* r_1 + \mu_0^*$ par $r_1 = \frac{\mu'^* - \mu_0^*}{\lambda_{0,1}^*}$. L'orbite de ζ_0^* est donc le cylindre de base le cercle $C_{\lambda_0^*}$.

- Si en revanche $\lambda_0^* = 0$, alors le cercle $C_{\lambda_0^*}$ dégénère et se réduit à l'origine, $\lambda'^* = 0$ et $\mu'^* = \mu_0^*$. ζ_0^* est donc un point fixe (orbite dégénérée) et tous les points de l'axe des μ^* sont des points fixes de Ad^* .

Nous vérifions bien le théorème de Kirillov-Kostant (cf. la section 6.3 du chapitre 5) disant que, pour les groupes résolubles, les orbites de la représentation co-adjointe sont des variétés symplectiques. Les cylindres $C_{\lambda_0^*} \times \mathbb{R}_{\tau^*}$ sont des espaces de phases identifiables aux fibrés cotangents $T^*C_{\lambda_0^*}$ des cercles $C_{\lambda_0^*}$ du plan \mathcal{Z}^* . L'espace de configuration est le cercle $C_{\lambda_0^*}$ des positions décrites par la rotation de λ_0^* (i.e. par les $\lambda^* = \lambda_0^* \cdot e^{-i\theta}$ de module constant $|\lambda_0^*|$) et les $\mu^* \in \mathbb{R}_{\tau^*}$ correspondent aux moments conjugués de l'angle des λ^* (on pourrait noter μ^* p comme un moment). La mesure est $d\theta d\mu^*$ associée à la 2-forme symplectique $d\theta \wedge d\mu^*$.

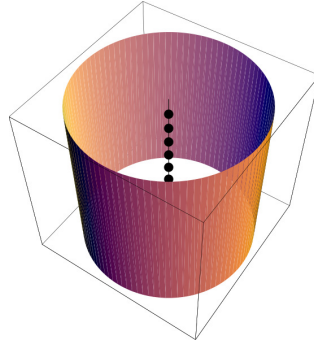


FIGURE 3. Les deux types d'orbites de la représentation co-adjointe : cylindre et droite de points fixes.

Il est également intéressant de relier la géométrie des orbites de la représentation co-adjointe à la structure de $G = SE(2) = \mathbb{R}^2 \rtimes SO(2)$ comme produit semi-direct du groupe des translations \mathbb{R}^2 et du groupe des rotations $SO(2)$. La structure de produit se retrouve au niveau de l'algèbre de Lie \mathcal{G} et de l'espace dual \mathcal{G}^* . Comme le \mathbb{R}^2 de G , que nous noterons $\mathbb{R}_{(q)}^2$ pour éviter les ambiguïtés, est un espace vectoriel, il s'identifie à son dual $\mathbb{R}_{(q^*)}^2$. Comme groupe additif, il s'identifie aussi à son algèbre de Lie (triviale) $\mathbb{R}_{(\mathcal{z})}^2$, algèbre de Lie qui s'identifie elle-même à son dual $(\mathbb{R}_{(\mathcal{z})}^2)^*$. À travers la structure de produit semi-direct, le groupe des rotations $SO(2)$ agit sur $\mathbb{R}_{(q^*)}^2$ par

$$\langle R(\theta) q^*, q \rangle = \langle R(\theta) q^*, R(-\theta) q \rangle .$$

Il agit aussi sur $\mathbb{R}_{(\mathcal{z}^*)}^2$, autrement dit sur le plan \mathcal{z}^* de coordonnées λ^* , par l'action de la représentation co-adjointe que nous avons plus haut $\lambda'^* = \lambda^* \cdot e^{-i\theta}$. Soit alors, comme plus haut, $\lambda_0^* \in \mathcal{z}^*$ et considérons le stabilisateur H_0 de $(\lambda_0^*, 0) \in \mathcal{G}^*$ pour l'action co-adjointe, i.e. le sous-groupe des (q, θ) laissant $(\lambda_0^*, 0)$ fixe. D'après les formules

$$\lambda'^* = \lambda_0^* \cdot e^{-i\theta} = \lambda_0^*, \quad \mu'^* = \lambda_0^* \cdot (i \cdot e^{-i\theta} \cdot q) + 0 = 0,$$

il faut d'abord $e^{-i\theta} = 1$ et donc $\theta = 0$. Il faut ensuite $\lambda_0^* \cdot i \cdot q = 0$. Or $i \cdot q$ est q tourné de $\frac{\pi}{2}$ et le \cdot symbolise le produit scalaire. Il faut par conséquent que $i \cdot q$ soit orthogonal à λ_0^* et par suite que q soit parallèle à λ_0^* . Donc H_0 est le sous-groupe $\mathbb{R}\lambda_0^* \times \{0\}$ de G . On considère alors le quotient $X = G/H_0$ qui est le produit $(\mathbb{R}_{(q)}^2/\mathbb{R}\lambda_0^*) \times SO(2)$. Ce quotient produit d'un cercle ($SO(2)$ est le cercle des θ) par une droite ($\mathbb{R}_{(q)}^2/\mathbb{R}\lambda_0^*$ quotient d'un plan par une droite est une droite) est un cylindre qui s'identifie naturellement à l'orbite co-adjointe. D'abord $SO(2)$ s'identifie naturellement au cercle $C_{\lambda_0^*}$ en le faisant tourner sur lui-même à partir de λ_0^* . Ensuite, on remarquera que, interprétée dans le plan \mathcal{z}^* , la droite $\mathbb{R}\lambda_0^*$ est celle du diamètre passant par λ_0^* du

cercle $C_{\lambda_0^*}$. Le quotient $\mathbb{R}_{(q)}^2/\mathbb{R}\lambda_0^*$ interprété dans \varkappa^* peut ainsi s'identifier à la tangente à $C_{\lambda_0^*}$ en λ_0^* , autrement dit, en termes de mécanique, à la "vitesse" en λ_0^* d'un mobile parcourant le cercle $C_{\lambda_0^*}$. Cette "vitesse" est représentable par un moment et, en ce sens, le cylindre $(\mathbb{R}_{(q)}^2/\mathbb{R}\lambda_0^*) \times SO(2)$ s'identifie à l'espace des phases d'un tel mobile.

4. Le formalisme d'Élie Cartan

4.1. La forme de Maurer-Cartan

Calculons maintenant pour $G = SE(2)$ la forme de Maurer-Cartan que nous avons calculée pour le groupe nilpotent $J^1\mathbb{R}^2$ à la section 10 du chapitre 5. Partons de l'expression générale $g = \begin{pmatrix} 1 & 0 \\ q & e^{i\theta} \end{pmatrix}$ d'un élément g de G . On peut considérer la différentielle de g : $dg = \begin{pmatrix} 0 & 0 \\ dq & i.e^{i\theta}d\theta \end{pmatrix}$ et l'interpréter comme une 1-forme sur G à valeurs dans \mathcal{G} . Cela signifie que les composantes de dg sont des 1-formes sur G , mais que dg possède le type d'un élément de \mathcal{G} . Autrement dit, si $\Omega^1(G)$ est l'espace vectoriel des 1-formes sur G , dg est un élément du produit tensoriel $\Omega^1(G) \otimes \mathcal{G}$ (\otimes étant le produit tensoriel sur \mathbb{R} de \mathbb{R} -espaces vectoriels). Pour exprimer $dg(g)$ comme somme de 3 éléments de $\Omega^1(G) \otimes \mathcal{G}$ dans la base (\varkappa, τ) on pourrait écrire

$$dg(g) = dq(g) \otimes \varkappa + d\theta(g) \otimes \tau.e^{i\theta} .$$

Mais \varkappa étant un vecteur ligne et $dq(g)$ un vecteur colonne il est plus commode pour les calculs d'avoir des produits tensoriels \otimes qui fonctionnent comme des produits de vecteurs lignes à gauche par des vecteurs colonnes à droite. Nous nous placerons donc plutôt dans $\mathcal{G} \otimes \Omega^1(G)$ et écrirons

$$dg(g) = \varkappa \otimes dq(g) + \tau.e^{i\theta} \otimes d\theta(g) .$$

Dans la base (\varkappa_g, τ_g) où $\varkappa_g = \varkappa.e^{i\theta}$ (donc $\varkappa = \varkappa_g.e^{-i\theta}$) et $\tau_g = \tau.e^{i\theta}$, on a donc

$$dg(g) = \varkappa_g.e^{-i\theta} \otimes dq(g) + \tau_g \otimes d\theta(g) = \varkappa_g \otimes e^{-i\theta}.dq(g) + \tau_g \otimes d\theta(g) .$$

Si $g = e$, on obtient $dg(e) = \varkappa \otimes dq(e) + \tau \otimes d\theta(e)$ et donc, en tant qu'élément du produit tensoriel $\mathcal{G}^* \otimes \mathcal{G}$ (en fait $\mathcal{G} \otimes \mathcal{G}^*$ à cause de nos conventions de calcul), $dg(e)$ est l'identité de \mathcal{G} . Cela signifie que $dg(e)$ est une 1-forme sur \mathcal{G} et s'applique donc à des vecteurs $\varsigma \in \mathcal{G}$. Mais comme elle est à valeurs dans \mathcal{G} , on a $dg(e)(\varsigma) \in \mathcal{G}$. L'identité signifie que $dg(e)(\varsigma) = \varsigma$. Plus généralement, dans cette façon de voir les choses, g symbolise un élément *générique* de G , et par conséquent l'application identique de G , ce qui fait que dg peut s'interpréter comme l'application identique du fibré tangent TG .

Le problème (celui de la non-holonomie) est, comme nous l'avons vu à propos du groupe $\mathbb{V}_J = J^1\mathbb{R}^2$ dans la section 10 du chapitre 5, que dg n'est pas invariante par les translations à gauche L_g . C'est bien l'application identique de TG mais pas celle de TG globalement trivialisé par les L_g . En effet, une 1-forme sur G à valeurs dans \mathcal{G} qui est G -invariante par translations à gauche doit avoir des composantes constantes dans les bases des T_g^*G duales des bases (G -invariantes) (\varkappa_g, τ_g) des espaces tangents T_gG . Or ce n'est pas le cas pour dg à cause du coefficient $e^{-i\theta}$ dans $\varkappa_g \cdot e^{-i\theta}$.

L'idée de Cartan est, nous l'avons vu, de translater $dg(e)$ de façon à obtenir une 1-forme sur G à valeurs dans \mathcal{G} qui soit *par construction* invariante par les translations L_g . Soit $\Lambda_G(g) : T_gG \rightarrow \mathcal{G}$ cette 1-forme de Maurer-Cartan. Elle s'interprète très simplement géométriquement. En effet, on a par définition $\Lambda_G(g) = (T_gL_{g^{-1}})^* dg(e)$.¹⁴ Si $\varsigma \in T_gG$ est un vecteur tangent à G en g , $\Lambda_G(g)(\varsigma_g) = T_gL_{g^{-1}}(\varsigma_g)$, autrement dit, $\Lambda_G(g)$ transporte ς_g dans \mathcal{G} au moyen de la trivialisatation globale fournie par les translations à gauche L_g .

On vérifie que l'on a :

$$\Lambda_G(g) = \begin{pmatrix} 0 & 0 \\ e^{-i\theta} \cdot dq & id\theta \end{pmatrix}.$$

Évidemment, $\Lambda_G(e) = dg(e)$ et donc, sur des champs de vecteurs ς invariants à gauche sur G , on a $\Lambda_G(\varsigma) = \varsigma$. Traditionnellement, on écrit Λ_G sous la forme compacte :

$$\Lambda_G = g^{-1}dg$$

où g^{-1} symbolise $(T_gL_{g^{-1}})^*$. Cela donne bien le même résultat puisque, au niveau des matrices,

$$g^{-1}dg = \begin{pmatrix} 1 & 0 \\ i \cdot e^{-i\theta} \cdot q & e^{-i\theta} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ dq & i \cdot e^{i\theta} d\theta \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ e^{-i\theta} \cdot dq & id\theta \end{pmatrix}.$$

Dans la suite nous noterons ω et ρ les 1-formes (à valeurs scalaires respectivement dans \mathbb{C} et \mathbb{R}) $e^{-i\theta} \cdot dq$ (vecteur colonne) et $d\theta$. Avec ces notations, la 1-forme Λ_G (à valeurs vectorielles dans \mathcal{G}) s'écrit :

$$\Lambda_G = \begin{pmatrix} 0 & 0 \\ \omega & \rho \end{pmatrix} = \varkappa \otimes \omega + \tau \otimes \rho$$

où, rappelons-le, (\varkappa, τ) est la base naturelle de \mathcal{G} .

14. On a $T_gL_{g^{-1}} = (T_eL_g)^{-1} : T_gG \rightarrow T_eG$ et donc par dualité $(T_gL_{g^{-1}})^* : T_e^*G = \mathcal{G}^* \rightarrow T_g^*G$.

4.2. Les équations de Maurer-Cartan

Rappelons (cf. encore la section 10 du chapitre 5) que la structure d'algèbre de Lie de \mathcal{G} peut se lire élégamment sur la forme de Maurer-Cartan Λ_G et qu'elle est alors donnée par la formule *universelle*

$$d\Lambda_G = -\frac{1}{2} [\Lambda_G \wedge \Lambda_G] ,$$

cette universalité faisant tout l'intérêt du formalisme. L'idée est de calculer la dérivée extérieure de Λ_G qui est une 2-forme vectorielle G -invariante à valeurs dans \mathcal{G} . Pour ce faire, on calcule d'abord les dérivées extérieures des composantes ω et ρ . On obtient :¹⁵

$$\begin{cases} d\omega &= -i.e^{-i\theta}.(d\theta \wedge dq) + e^{-i\theta}d^2q = i.e^{-i\theta}.dq \wedge d\theta = i.\omega \wedge \rho \\ d\rho &= d^2\theta = 0 \end{cases}$$

et donc

$$d\Lambda_G = \varkappa \otimes d\omega + \tau \otimes d\rho = \varkappa \otimes (i.\omega \wedge \rho) .$$

Mais l'espace $\Omega^1(G) \otimes \mathcal{G}$ (traité comme $\mathcal{G} \otimes \Omega^1(G)$ dans nos calculs) des 1-formes sur G à valeurs dans \mathcal{G} est muni d'un *produit extérieur* qui fait interagir le produit extérieur des 1-formes et le crochet de Lie de \mathcal{G} . En effet, soient $\lambda = \varkappa \otimes \mu + \tau \otimes \nu$ et $\lambda' = \varkappa \otimes \mu' + \nu \otimes \tau'$ deux éléments quelconques de $\mathcal{G} \otimes \Omega^1(G)$. On a par définition :

$$[\lambda \wedge \lambda'] = [\varkappa, \varkappa] \otimes (\mu \wedge \mu') + [\varkappa, \tau] \otimes (\mu \wedge \nu') + [\tau, \varkappa] \otimes (\nu \wedge \mu') + [\tau, \tau] \otimes (\nu \wedge \nu')$$

qui est une 2-forme vectorielle à valeurs dans \mathcal{G} , c'est-à-dire un élément de $\Omega^2(G) \otimes \mathcal{G}$ ($\mathcal{G} \otimes \Omega^2(G)$ dans nos calculs).¹⁶ Comme $[\varkappa, \varkappa] = [\tau, \tau] = 0$, $[\tau, \varkappa] = -[\varkappa, \tau]$ et $[\varkappa, \tau] = -\varkappa.i$, on obtient en définitive :

$$[\lambda \wedge \lambda'] = [\varkappa, \tau] \otimes ((\mu \wedge \nu') - (\nu \wedge \mu')) = -\varkappa.i \otimes ((\mu \wedge \nu') - (\nu \wedge \mu')) .$$

Lorsque $\lambda = \lambda'$, on obtient donc, contrairement à ce qui se passe pour une 1-forme scalaire,

$$[\lambda \wedge \lambda] = -2\varkappa.i \otimes (\mu \wedge \nu) .$$

En particulier

$$[\Lambda_G \wedge \Lambda_G] = -2\varkappa.i \otimes (\omega \wedge \rho) .$$

Mais

$$-2\varkappa.i \otimes (\omega \wedge \rho) = -2\varkappa \otimes (i.\omega \wedge \rho) .$$

En comparant les expressions donnant $d\Lambda_G$ et $[\Lambda_G \wedge \Lambda_G]$ on obtient les équations universelles de Maurer-Cartan qui encodent la géométrie de tout groupe de Lie G :

$$d\Lambda_G = -\frac{1}{2} [\Lambda_G \wedge \Lambda_G] . \tag{1}$$

15. Rappelons que pour toute forme différentielle σ à valeurs *scalaires*, on a $d^2\sigma = 0$.

16. $\Omega^k(G)$ dénote l'espace des k -formes différentielles sur G .

5. Les problèmes variationnels d'après Bryant et Griffiths

Maintenant que nous avons précisé la structure du groupe des déplacements du plan $G = SE(2)$, avec sa structure de fibré principal, ses représentations adjointe et co-adjointe et sa forme de Maurer-Cartan satisfaisant l'équation caractéristique (1), nous pouvons chercher à reformuler nos modèles variationnels de contours subjectifs modaux de façon plus intrinsèque dans ce cadre géométrique plus profond. Nous allons nous appuyer pour cela sur un article de référence "Reduction for constrained variational problems and $\int \frac{\kappa^2}{2} ds$ " [82] où Robert Bryant et Phillip Griffiths ont développé la théorie des problèmes variationnels sur les groupes de Lie et l'ont appliquée à la théorie des *elasticae*. Nous allons résumer leur approche (un peu technique) et l'appliquer ensuite à notre modèle de "géodésiques legendriennes".

5.1. Le système de Pfaff des relevées dans le fibré principal G

Revenons aux courbes (régulières) γ dans le plan rétinien \mathbb{R}^2 et à notre remarque à la fin de la section 7 du chapitre 5. En suivant le repère mobile $R = R_q$ lorsque le point $q = (x, y)$ parcourt γ , on obtient une courbe Γ dans G qui relève γ et que l'on appelle son *relèvement de Frenet*. Dans la mesure où G est un fibré principal sur \mathbb{R}^2 de fibre $H = SO_2(\mathbb{R}) \simeq \mathbb{S}^1$, on peut réinterpréter dans ce nouveau contexte les relèvements legendriens dans $J^1\mathbb{R}^2$ étudiés dans la section 2 du chapitre 3. Si γ est paramétrée par t ($\gamma : I = [0, 1] \rightarrow \mathbb{R}^2, t \mapsto q(t)$), Γ est aussi paramétrée par t . L'élément de longueur d'arc sur γ (et non pas sur Γ) est donné par $ds = \sigma(t)dt$ avec $\sigma(t) = \|q'(t)\|$. Par ailleurs, le repère de Frenet R_q est donné par le vecteur tangent unitaire

$$e_1(t) = \frac{q'(t)}{\sigma(t)} = (\cos(\theta), \sin(\theta))$$

et le vecteur normal unitaire

$$e_2(t) = e_1(t)^\perp = (-\sin(\theta), \cos(\theta)) .$$

Les différentielles de e_1 et e_2 sont par conséquent les 1-formes vectorielles sur \mathbb{R}^2 à valeurs dans \mathbb{R}^2 :

$$\begin{cases} de_1 &= (-\sin(\theta)d\theta, \cos(\theta)d\theta) = d\theta \otimes e_2 = \rho \otimes e_2 \\ de_2 &= -(\cos(\theta)d\theta, \sin(\theta)d\theta) = -d\theta \otimes e_1 = -\rho \otimes e_1 \end{cases}$$

Pour réinterpréter dg dans le repère mobile R de \mathbb{R}^2 , on remarque que, comme $\omega = e^{-i\theta}dq$ par définition (et donc $dq = e^{i\theta}\omega$ dans le repère *fixe*), dq calculé dans le repère *mobile* R s'identifie à ω , ce qui peut s'écrire $dq = \omega_u \otimes e_1 + \omega_v \otimes e_2 =$ (où ω_u et ω_v sont les composantes de ω dans le repère mobile). D'où la réinterprétation de dg dans le repère *mobile* :

$$\begin{cases} dq &= \omega_u \otimes e_1 + \omega_v \otimes e_2 \\ de_1 &= \rho \otimes e_2 \\ de_2 &= -\rho \otimes e_1 \end{cases}$$

Ce système différentiel est lié à G et est donc *indépendant* de toute courbe γ particulière.

Mais par ailleurs, sur une courbe γ particulière, on a, par définition de ds , que dq calculé dans le repère *mobile* est tout simplement $(ds, 0)$ (vecteur colonne) et par conséquent $\omega_u = ds$ et $\omega_v = 0$. Rappelons alors que la courbure κ est donnée par $\kappa = d\theta/ds$ et que donc $\rho = d\theta = \kappa ds = \kappa \sigma dt$. D'où l'expression suivante pour la variation infinitésimale dR du repère mobile R le long de γ :

$$\begin{cases} dq &= ds \otimes e_1 = \sigma(t)dt \otimes e_1 \\ de_1 &= ds\kappa(t) \otimes e_2 = \kappa(t)\sigma(t)dt \otimes e_2 \\ de_2 &= -ds\kappa(t) \otimes e_1 = -\kappa(t)\sigma(t)dt \otimes e_1 \end{cases}$$

En comparant les deux expressions on voit que *le long des courbes* $\gamma(t)$ le système de Pfaff suivant est vérifié :

$$\Pi = \begin{cases} \omega_u - \sigma(t)dt &= 0 \\ \omega_v &= 0 \\ \rho - \kappa(t)\sigma(t)dt &= 0 \end{cases} \quad (\text{Pf})$$

De même que les relevées legendriennes dans $J^1\mathbb{R}^2$ étaient les courbes intégrales de la structure de contact, de même les relevées de Frenet dans G sont les courbes intégrales de Π . On remarquera que pour ces relevées on a, contrairement au cas général, $d\omega = i.\omega \wedge \rho = 0$ puisque, si $\omega_v = 0$ et $\rho = \kappa(t)\sigma(t)dt$

$$\begin{cases} d\omega_u = -\omega_v \wedge \rho = 0 \\ d\omega_v = \omega_u \wedge \rho = \sigma(t)dt \wedge \sigma(t)\kappa(t)dt = \sigma^2(t)\kappa(t)dt \wedge dt = 0 \end{cases}$$

Le système de Pfaff Π est défini sur G mais dépend de t ainsi que des fonctions $\sigma(t)$ et $\kappa(t)$. Nous pouvons alors appliquer exactement la même stratégie que celle que nous avons appliquée au chapitre 3 pour le fibré de contact et le fibré des 1-jets $J^1\mathbb{R}^2$, à savoir introduire des *coordonnées supplémentaires*

$$(\sigma, \kappa, t) \in \mathbb{R}_{(\sigma)}^+ \times \mathbb{R}_{(\kappa)} \times \mathbb{R}_{(t)} = Y$$

et considérer que Π est en fait défini sur la variété produit direct $X = G \times Y$, quitte à introduire ensuite des *conditions supplémentaires* exprimant que σdt est l'élément de longueur et κ la courbure.¹⁷ Sur l'extension X de G , σ et κ ne sont plus des

17. On prolonge trivialement une 1-forme sur G en une 1-forme sur X et on garde le même symbole.

fonctions de t et le système de Pfaff Π prend la forme :

$$\Pi = \begin{cases} \omega_u - \sigma dt & = 0 \\ \omega_v & = 0 \\ \rho - \kappa \sigma dt & = 0 \end{cases}$$

Dans la mesure où $(\omega_u, \omega_v, \rho)$ sont les composantes de $\Lambda_G = \varkappa \otimes \omega + \tau \otimes \rho$, on voudrait pouvoir considérer qu'il consiste tout simplement en l'annulation de la 1-forme :

$$\mu = \Lambda_G - P dt$$

avec $P = (\sigma, 0, \kappa \sigma)$. Comme Λ_G est vue maintenant comme une 1-forme sur

$$X = G \times Y$$

(triviale sur Y) à valeurs dans \mathcal{G} , la cohérence des types impose alors qu'il en aille de même pour $\mu = \Lambda_G - P dt$ et donc que P soit considéré comme un vecteur de \mathcal{G} , nommément le vecteur :

$$P = \begin{pmatrix} 0 & 0 \\ \bar{\sigma} & \kappa \sigma i \end{pmatrix} = \varkappa \otimes \bar{\sigma} + \tau \otimes \kappa \sigma$$

où $\bar{\sigma}$ est le vecteur colonne $\begin{pmatrix} \sigma \\ 0 \end{pmatrix}$. Si l'on note alors \mathcal{A} le sous-espace de \mathcal{G} constitué des P de cette forme, on peut redéfinir Y comme $\mathcal{A} \times \mathbb{R}_{(t)}$ et considérer que Π devient en fait défini sur $X = G \times Y = G \times \mathcal{A} \times \mathbb{R}_{(t)}$.

5.2. La fonctionnelle du problème variationnel

On peut maintenant sélectionner certaines courbes γ de \mathbb{R}^2 en introduisant un principe variationnel pour leurs relevées de Frenet. Pour ce faire, on introduit évidemment un lagrangien L et l'on considère la 1-forme sur X donnée par $\varphi = L dt$. Si $\Gamma : I \rightarrow X$ est une courbe dans X , on lui associe par conséquent "l'énergie" :

$$\Phi(\Gamma) = \int_I \Gamma^* \varphi$$

et l'on cherche les courbes Γ qui minimisent Φ , autrement qui sont des points critiques stables de $\Phi(\Gamma)$.

On analyse ainsi, dans nos applications neurogéométriques, trois niveaux différents de structure reliant la géométrie *universelle* de G à des familles *particulières* de courbes de \mathbb{R}^2 :

1. Le groupe de Lie G et les structures associées \mathcal{G} , \mathcal{G}^* , et Λ_G , ainsi que les représentations adjointe et co-adjointe. Il s'agit là, répétons-le encore une fois, d'un cadre géométrique *universel* implémenté au moyen d'architectures fonctionnelles neuronales.

2. Les trajectoires dans G qui sont les relevées de Frenet Γ des courbes γ de la base \mathbb{R}^2 . Il s'agit du codage de stimuli particuliers dans ce cadre universel.
3. Enfin, pour les contours subjectifs modaux, celles parmi ces courbes qui sont solutions d'un problème variationnel. Il s'agit de l'interprétation variationnelle du problème de la complétion modale.

Dans notre cas, où l'on choisit la métrique euclidienne sur \mathbb{V}_J , la 1-forme φ correspondant au lagrangien de la section 2 du chapitre 10 est donnée, en fonction de dt , ou ds , ou dx , par :

$$\begin{aligned}\varphi &= \left(dq^2 + (\tan^2(\theta))' d\theta^2 \right)^{\frac{1}{2}} \\ &= L dt = \sigma \left(1 + \frac{\kappa^2}{\cos^4(\theta)} \right)^{\frac{1}{2}} dt \\ &= L_0 ds = \left(1 + \frac{\dot{\theta}^2}{\cos^4(\theta)} \right)^{\frac{1}{2}} ds = \left(1 + \dot{\theta}^2 \sec^4(\theta) \right)^{\frac{1}{2}} ds \\ &= L_1 dx = \left(1 + f'^2 + f''^2 \right)^{\frac{1}{2}} dx .\end{aligned}$$

Rappelons (cf. section 6 de ce même chapitre 10) que le lagrangien L_0 montre que l'on peut prendre f' et f'' comme variables même si cela brise la symétrie de rotation. Si l'angle θ subit un déphasage $\theta \rightarrow \theta + \theta_0$, $\dot{\theta}$ reste invariant. On résout donc essentiellement la même équation et l'on revient à la solution initiale en faisant $\theta \rightarrow \theta - \theta_0$. Rappelons aussi que le terme principal de l'équation d'Euler-Lagrange pour L_0 est donné par la formule (9) de la section 6 de ce même chapitre 10 :

$$\frac{\partial L_0}{\partial \theta} - \frac{d}{ds} \left(\frac{\partial L_0}{\partial \dot{\theta}} \right) = - \frac{\sec^4(\theta)}{\left(1 + \dot{\theta}^2 \sec^4(\theta) \right)^{\frac{3}{2}}} \left(\ddot{\theta} + 2\dot{\theta}^2 \tan(\theta) \right) .$$

5.3. Les équations d'Euler-Lagrange (III)

Expliquons maintenant la façon dont Bryant et Griffiths formulent dans ce contexte géométrique général des groupes de Lie les équations d'Euler-Lagrange. Pour calculer celles-ci, on introduit d'abord les multiplicateurs de Lagrange correspondant à la contrainte "cinématique" définie par le système de Pfaff Π (c'est-à-dire $\mu = \Lambda_G - P dt = 0$). On considère par conséquent la 1-forme scalaire sur X :

$$\psi = \varphi + \lambda_1 (\omega_u - \sigma dt) + \lambda_2 \omega_v + \lambda_3 (\rho - \kappa \sigma dt) .$$

Une interprétation géométrique des multiplicateurs de Lagrange consiste à identifier λ à un vecteur de $\mathcal{G}^* \simeq \mathbb{R}^3$ et à écrire les 1-formes ψ sous la forme

$$\psi = \varphi + \langle \lambda, \mu \rangle = \varphi + \langle \lambda, \Lambda_G - P dt \rangle .$$

Dans une telle écriture, $\lambda \in \mathcal{G}^*$ est de type \mathcal{G}^* et $\mu \in \Omega^1(X) \otimes \mathcal{G}$ est une 1-forme vectorielle sur X à valeurs dans \mathcal{G} .¹⁸ La forme μ appliquée à un vecteur tangent ς à X donne un élément $\mu(\varsigma)$ de \mathcal{G} et $\langle \lambda, \mu \rangle = \lambda(\mu(\varsigma))$ est donc de type \mathbb{R} . Comme $\varphi \in \Omega^1(X)$ (1-forme scalaire sur X), le type $[\psi]$ de ψ est bien cohérent, comme on le vérifie facilement :¹⁹

$$[\psi] = [\Omega^1(X) + \langle \mathcal{G}^*, \Omega^1(X) \otimes \mathcal{G} \rangle] = [\Omega^1(X)] + [\Omega^1(X) \otimes \langle \mathcal{G}^*, \mathcal{G} \rangle] = [\Omega^1(X)] .$$

Sur une relevée de Frenet, la contrainte cinématique μ s'annule et l'on a $\psi = \varphi$.

On calcule ensuite la 2-forme scalaire $\Psi = d\psi \in \Omega^2(X)$ en utilisant les équations de Maurer-Cartan (1) de la section 4.2. On obtient (avec $\varphi = Ldt$) :

$$\begin{aligned} \Psi &= dL \wedge dt + \langle d\lambda, \mu \rangle + \langle \lambda, d\mu \rangle \\ &= dL \wedge dt + \langle d\lambda, \mu \rangle + \langle \lambda, d\Lambda_G \rangle - \langle \lambda, dP \wedge dt \rangle \\ &= dL \wedge dt + \langle d\lambda, \mu \rangle - \frac{1}{2} \langle \lambda, [\Lambda_G \wedge \Lambda_G] \rangle - \langle \lambda, dP \wedge dt \rangle . \end{aligned}$$

La vérification des types donne bien $\Omega^2(X)$ pour tous les termes. En effet :

$$\begin{aligned} [dL \wedge dt] &= [\Omega^1(X) \wedge \Omega^1(X)] = [\Omega^2(X)] , \\ [\langle d\lambda, \mu \rangle] &= [\langle \Omega^1(X) \otimes \mathcal{G}^*, \Omega^1(X) \otimes \mathcal{G} \rangle] = [(\Omega^1(X) \wedge \Omega^1(X)) \otimes \langle \mathcal{G}^*, \mathcal{G} \rangle] \\ &= [\Omega^2(X) \otimes \mathbb{R}] = [\Omega^2(X)] , \\ [\langle \lambda, d\mu \rangle] &= [\langle \mathcal{G}^*, \Omega^2(X) \otimes \mathcal{G} \rangle] = [\Omega^2(X) \otimes \langle \mathcal{G}^*, \mathcal{G} \rangle] = [\Omega^2(X) \otimes \mathbb{R}] = [\Omega^2(X)] . \end{aligned}$$

Explicitement, on obtient les formules suivantes :

$$\begin{aligned} \Psi &= \frac{\partial L}{\partial \theta} \rho \wedge dt + \frac{\partial L}{\partial \sigma} d\sigma \wedge dt + \frac{\partial L}{\partial \kappa} d\kappa \wedge dt \\ &\quad + d\lambda_1 \wedge (\omega_u - \sigma dt) - \lambda_1 \omega_v \wedge \rho - \lambda_1 d\sigma \wedge dt \\ &\quad + d\lambda_2 \wedge \omega_v + \lambda_2 \omega_u \wedge \rho \\ &\quad + d\lambda_3 \wedge (\rho - \kappa \sigma dt) - \lambda_3 \sigma d\kappa \wedge dt - \lambda_3 \kappa d\sigma \wedge dt . \end{aligned}$$

Pour obtenir les points critiques de la fonctionnelle $\Phi(\Gamma)$ on exprime alors que les contractions de Ψ par les vecteurs tangents appropriés de $Z = X \times \mathcal{G}^*$ s'annulent. Cela signifie que l'on considère les composantes $\Psi_i = \alpha_i \wedge \beta_i$ ($i = 1, \dots, 8$) de Ψ par

18. Comme dans cette section nous travaillons avec $3 = (1, 1, 1)$ composantes et non plus avec $3 = (2, 1)$ composantes, nous revenons aux formulations standard $\mathcal{G}^* \otimes \mathcal{G}$, $\Omega^1(X) \otimes \mathcal{G}$, etc.

19. Nous notons $[E]$ le type d'une expression E ou d'un ensemble d'expressions analogues. Dans les vérifications de types, les opérations algébriques doivent être interprétées symboliquement et l'on a par exemple $[T + T] = [T]$ si $[T]$ est le type de certains objets munis d'une addition.

rapport aux huit 1-formes $\alpha_i = d\lambda_1, d\lambda_2, d\lambda_3, d\sigma, d\kappa, \rho, \omega_u, \omega_v$ et que l'on annule les 1-formes β_i qui sont leurs coefficients. Comme les Ψ_i sont données par :

$$\left\{ \begin{array}{l} \Psi_1 = d\lambda_1 \wedge (\omega_u - \sigma dt) \\ \Psi_2 = d\lambda_2 \wedge \omega_v \\ \Psi_3 = d\lambda_3 \wedge (\rho - \kappa \sigma dt) \\ \Psi_4 = \frac{\partial L}{\partial \sigma} d\sigma \wedge dt - \lambda_1 d\sigma \wedge dt - \lambda_3 \kappa d\sigma \wedge dt \\ \Psi_5 = \frac{\partial L}{\partial \kappa} d\kappa \wedge dt - \lambda_3 \sigma d\kappa \wedge dt \\ \Psi_6 = -\frac{\partial L}{\partial \theta} dt \wedge \rho - \lambda_1 \omega_v \wedge \rho + \lambda_2 \omega_u \wedge \rho + d\lambda_3 \wedge \rho \\ \Psi_7 = d\lambda_1 \wedge \omega_u - \lambda_2 \rho \wedge \omega_u \\ \Psi_8 = d\lambda_2 \wedge \omega_v + \lambda_1 \rho \wedge \omega_v \end{array} \right.$$

on obtient ainsi le système différentiel de Bryant-Griffiths :

$$\left\{ \begin{array}{l} (1) \quad \omega_u - \sigma dt = 0 \\ (2) \quad \omega_v = 0 \\ (3) \quad \rho - \kappa \sigma dt = 0 \\ (4) \quad \frac{\partial L}{\partial \sigma} dt - \lambda_1 dt - \lambda_3 \kappa dt = 0 \\ (5) \quad \frac{\partial L}{\partial \kappa} dt - \lambda_3 \sigma dt = 0 \\ (6) \quad -\frac{\partial L}{\partial \theta} dt - \lambda_1 \omega_v + \lambda_2 \omega_u + d\lambda_3 = 0 \\ (7) \quad d\lambda_1 - \lambda_2 \rho = 0 \\ (8) \quad d\lambda_2 + \lambda_1 \rho = 0 \end{array} \right. \quad (2)$$

- Les trois premières équations (1), (2) et (3) ne font que redonner le système de Pfaff II de la section 5.1. Ce sont les équations de structure le long d'une relevée de Frenet.
- Les équations (4) et (5) permettent de calculer les multiplicateurs de Lagrange λ_1 et λ_3 . On trouve

$$\lambda_3 = \frac{1}{\sigma} \frac{\partial L}{\partial \kappa}, \quad \lambda_1 = \frac{\partial L}{\partial \sigma} - \frac{\kappa}{\sigma} \frac{\partial L}{\partial \kappa}.$$

- Quant aux deux dernières équations (7) et (8), elles impliquent

$$\lambda_1 d\lambda_1 + \lambda_2 d\lambda_2 = 0,$$

autrement dit la loi de conservation $\lambda_1^2 + \lambda_2^2 = c^2$ ou

$$(\lambda_1, \lambda_2) = \tilde{\lambda} e^{i\varphi} = \left(\tilde{\lambda} \cos(\varphi), \tilde{\lambda} \sin(\varphi) \right)$$

pour une certaine constante $\tilde{\lambda}$. Cela signifie que, en ce qui concerne la composante $\lambda \in \mathcal{G}^*$, les solutions sont incluses dans les orbites de la représentation co-adjointe.

- C'est l'équation (6) $-\frac{\partial L}{\partial \theta} dt - \lambda_1 \omega_v + \lambda_2 \omega_u + d\lambda_3 = 0$ qui est donc la plus importante. Elle permet de calculer λ_2 et représente la véritable équation d'E-L du système. C'est "l'équation d'Euler".

Compte tenu des équations (1) et (2), l'équation d'Euler s'écrit :

$$-\frac{\partial L}{\partial \theta} dt + \lambda_2 \sigma dt + d\lambda_3 = 0 .$$

Mais comme $\lambda_3 = \frac{1}{\sigma} \frac{\partial L}{\partial \kappa}$, elle devient :

$$d\lambda_3 = -\frac{1}{\sigma^2} \frac{\partial L}{\partial \kappa} d\sigma + \frac{1}{\sigma} \left[\frac{\partial^2 L}{\partial \kappa \partial \theta} d\theta + \frac{\partial^2 L}{\partial \kappa \partial \sigma} d\sigma + \frac{\partial^2 L}{\partial^2 \kappa} d\kappa \right] .$$

D'où en définitive, puisque $d\theta = \rho = \kappa \sigma dt$,

$$\lambda_2 \sigma dt - \frac{\partial L}{\partial \theta} dt - \frac{1}{\sigma^2} \frac{\partial L}{\partial \kappa} d\sigma + \kappa \frac{\partial^2 L}{\partial \kappa \partial \theta} dt + \frac{1}{\sigma} \frac{\partial^2 L}{\partial \kappa \partial \sigma} d\sigma + \frac{1}{\sigma} \frac{\partial^2 L}{\partial^2 \kappa} d\kappa = 0 .$$

Dans notre cas, $L = \sigma (1 + \kappa^2 \sec^4 \theta)^{\frac{1}{2}}$ et le calcul donne :²⁰

$$\begin{cases} \lambda_1 &= (1 + \kappa^2 \sec^4(\theta))^{-\frac{1}{2}} \\ \lambda_3 &= \kappa \sec^4(\theta) (1 + \kappa^2 \sec^4(\theta))^{-\frac{1}{2}} \\ \lambda_2 &= -\sec^4(\theta) (1 + \kappa^2 \sec^4(\theta))^{-\frac{3}{2}} (\dot{\kappa} + 2\kappa^2 \tan(\theta)) \end{cases}$$

En termes de L_0 ($\varphi = Ldt = L_0 ds$ et $L = L_0 \sigma$), les multiplicateurs de Lagrange s'expriment simplement par :

$$\begin{cases} \lambda_1 &= L_0 - \kappa \frac{\partial L_0}{\partial \kappa} \\ \lambda_2 &= \frac{\partial L_0}{\partial \theta} - \frac{d}{ds} \left(\frac{\partial L_0}{\partial \kappa} \right) \\ \lambda_3 &= \frac{\partial L_0}{\partial \kappa} \end{cases}$$

On remarquera que $\lambda_2 = \tilde{\lambda} \sin \varphi$ redonne les équations (9) de la section 6 du chapitre 10 et (??) de la section précédente avec $\varphi = \theta$.

La loi de conservation $\lambda_1^2 + \lambda_2^2 = \tilde{\lambda}^2$ fournit alors l'équation différentielle :

$$(1 + \kappa^2 \sec^4(\theta))^2 + \sec^8(\theta) (\dot{\kappa} + 2\kappa^2 \tan(\theta))^2 = \tilde{\lambda}^2 (1 + \kappa^2 \sec^4(\theta))^3$$

qui n'est rien d'autre que l'équation (12) obtenue précédemment à la section 6 du chapitre 10 (avec $\lambda = \tilde{\lambda}$).

20. On remarquera que, étant donnée la forme très particulière de L_0 , on a

$$\lambda_1 = L_0 - \dot{\theta} \frac{\partial L_0}{\partial \dot{\theta}} = \frac{1}{L_0} .$$

5.4. L'interprétation géométrique des équations d'Euler-Lagrange

Bryant et Griffiths donnent l'interprétation géométrique suivante de leur système d'équations. De façon générale, si α et β sont deux 1-formes sur une variété X et si ς est un vecteur tangent, la contraction (produit intérieur) $i_\varsigma(\alpha \wedge \beta)$ de la 2-forme $\alpha \wedge \beta$ par ς est la 1-forme définie par :

$$i_\varsigma(\alpha \wedge \beta) = i_\varsigma(\alpha) \wedge \beta - \alpha \wedge i_\varsigma(\beta) = i_\varsigma(\alpha)\beta - i_\varsigma(\beta)\alpha . \quad (3)$$

Il s'agit bien d'une 1-forme car $i_\varsigma(\alpha) = \alpha(\varsigma)$ et $i_\varsigma(\beta) = \beta(\varsigma)$ sont des scalaires (c'est-à-dire des 0-formes). On prolonge linéairement i_ς aux 2-formes quelconques.

On écrit alors que $i_\varsigma(\Psi) = 0$ pour tous les vecteurs tangents $\varsigma \in TX$ avec

$$\Psi = dL \wedge dt + \langle d\lambda, \mu \rangle - \frac{1}{2} \langle \lambda, [\Lambda_G \wedge \Lambda_G] \rangle - \langle \lambda, dP \wedge dt \rangle .$$

1. Considérons d'abord des vecteurs ς tangents à

$$X = G \times Y = G \times \mathcal{A} \times \mathbb{R}_{(t)}$$

le long du facteur \mathcal{A} , autrement dit (puisque \mathcal{A} est un espace vectoriel et est donc partout identique à son espace tangent) des vecteurs $\varsigma \in \mathcal{A}$, et calculons la 1-forme $i_\varsigma(\Psi)$. D'après la formule (3),

$$i_\varsigma(dL \wedge dt) = dL(\varsigma)dt - dt(\varsigma)dL = dL(\varsigma)dt$$

car $dt(\varsigma) = 0$ puisque $\varsigma \in \mathcal{A}$ n'a pas de composante le long du facteur $\mathbb{R}_{(t)}$ de $X = G \times \mathcal{A} \times \mathbb{R}_{(t)}$. Il en va de même pour le terme comprenant $i_\varsigma(dP \wedge dt)$ qui se réduit à $dP(\varsigma)dt$. Et comme $dP(\varsigma) = \varsigma$, on obtient $i_\varsigma(dP \wedge dt) = \varsigma dt$ et donc

$$i_\varsigma(\langle \lambda, dP \wedge dt \rangle) = \lambda(\varsigma)dt .$$

Les deux autres termes $\langle d\lambda, \mu \rangle$ et $\frac{1}{2} \langle \lambda, [\Lambda_G \wedge \Lambda_G] \rangle$ de $i_\varsigma(\Psi)$ sont nuls car ς n'a aucune composante le long des facteurs de $Z = G \times Y \times \mathcal{G}^*$ sur lesquels les 2-formes composantes sont définies. Comme on a toujours $dt \neq 0$ (condition de transversalité), on obtient en définitive

$$dL(\varsigma) - \langle \lambda, \varsigma \rangle = 0 ,$$

pour tout $\varsigma \in \mathcal{A}$.

2. Considérons maintenant des vecteurs ς tangents à X le long du facteur G et calculons à nouveau $i_\varsigma(\Psi)$.

– Développons d'abord $dL \wedge dt$ en

$$\frac{\partial L}{\partial \theta} d\theta \wedge dt + \frac{\partial L}{\partial \sigma} d\sigma \wedge dt + \frac{\partial L}{\partial \kappa} d\kappa \wedge dt$$

Le terme

$$i_\varsigma \left(\frac{\partial L}{\partial \theta} d\theta \wedge dt + \frac{\partial L}{\partial \sigma} d\sigma \wedge dt + \frac{\partial L}{\partial \kappa} d\kappa \wedge dt \right)$$

donne 0 pour son second terme $d\sigma(\varsigma)dt - dt(\varsigma)d\sigma$ car ς n'a pas de composante le long de \mathcal{A} . Il en va de même pour son troisième terme $d\kappa(\varsigma)dt - dt(\varsigma)d\kappa$. Il ne reste donc que le premier terme

$$i_\varsigma \left(\frac{\partial L}{\partial \theta} d\theta \wedge dt \right) = \frac{\partial L}{\partial \theta} (d\theta(\varsigma)dt - dt(\varsigma)d\theta) .$$

Comme $dt(\varsigma) = 0$, il reste en définitive

$$\frac{\partial L}{\partial \theta} d\theta(\varsigma)dt = \frac{\partial L}{\partial \theta} \rho(\varsigma)dt .$$

- Par ailleurs la contribution de $-\langle \lambda, dP \wedge dt \rangle$ est nulle car ς n'a pas de composante le long de \mathcal{A} .
- Pour le terme $\langle d\lambda, \mu \rangle$ on a (puisque $\mu = \Lambda_G - Pdt$)

$$i_\varsigma (\langle d\lambda, \mu \rangle) = \langle d\lambda(\varsigma), \Lambda_G \rangle - \langle d\lambda, \Lambda_G(\varsigma) \rangle - \langle d\lambda(\varsigma), Pdt \rangle + \langle d\lambda, Pdt(\varsigma) \rangle .$$

Comme $dt(\varsigma) = 0$ et comme $d\lambda$ ne concerne que la composante \mathcal{G}^* de Z et que ς est sans composante le long de \mathcal{G}^* , il ne reste que le terme $-\langle d\lambda, \Lambda_G(\varsigma) \rangle$ qui est lui même égal à $-\langle d\lambda, \varsigma \rangle$ car, par construction, $\Lambda_G(\varsigma) = \varsigma$ (cf. section 4.1).

- Enfin le terme $-\frac{1}{2} \langle \lambda, [\Lambda_G \wedge \Lambda_G] \rangle$ donne

$$i_\varsigma \left(-\frac{1}{2} \langle \lambda, [\Lambda_G \wedge \Lambda_G] \rangle \right) = \left\langle \lambda, i_\varsigma \left(-\frac{1}{2} [\Lambda_G \wedge \Lambda_G] \right) \right\rangle .$$

Mais comme, si ς, η sont deux éléments de \mathcal{G} vus comme champs G -invariants par translations à gauche, on a

$$-\frac{1}{2} [\Lambda_G \wedge \Lambda_G] (\varsigma, \eta) = -[\varsigma, \eta] ,$$

on obtient en termes de 1-formes :

$$i_\varsigma \left(-\frac{1}{2} [\Lambda_G \wedge \Lambda_G] \right) = -\frac{1}{2} [\Lambda_G \wedge \Lambda_G] (\varsigma, \bullet) = -[\varsigma, \bullet] = -[\varsigma, \Lambda_G]$$

où, si α est une 1-forme à valeurs dans \mathcal{G} , $[\varsigma, \alpha]$ est la 1-forme à valeurs dans \mathcal{G} définie par $[\varsigma, \alpha](\eta) = [\varsigma, \alpha(\eta)]$.

On obtient ainsi en définitive le système :

$$\begin{cases} \text{(i)} & \mu = \Lambda_G - Pdt = 0 \\ \text{(ii)} & dL(\varsigma) - \langle \lambda, \varsigma \rangle = 0, \text{ pour tout } \varsigma \in \mathcal{A}. \\ \text{(iii)} & -\frac{\partial L}{\partial \theta} \rho(\varsigma)dt + \langle d\lambda, \varsigma \rangle - \langle \lambda, [P, \varsigma] \rangle dt = 0, \text{ pour tout } \varsigma \in \mathcal{G} \end{cases}$$

On remarque alors que les deux premiers termes de (iii) sont de la forme $\langle \varpi, \varsigma \rangle$ et qu'il est donc intéressant d'écrire le troisième terme de la même façon. Pour cela,

il suffit d'écrire $[P, \varsigma] = U \cdot \varsigma$ avec U matrice 3×3 et ς vecteur colonne et de réécrire $-\langle \lambda, [P, \varsigma] \rangle dt = -\langle \lambda \cdot U, \varsigma \rangle dt$. Mais comme en termes de matrices

$$P = \begin{pmatrix} 0 & 0 & 0 \\ \sigma & 0 & -\kappa\sigma \\ 0 & \kappa\sigma & 0 \end{pmatrix}$$

$$\varsigma = \begin{pmatrix} 0 & 0 & 0 \\ \varsigma_1 & 0 & -\varsigma_3 \\ \varsigma_2 & \varsigma_3 & 0 \end{pmatrix}$$

on a

$$[P, \varsigma] = \begin{pmatrix} 0 & 0 & 0 \\ -\kappa\sigma\varsigma_2 & 0 & 0 \\ \kappa\sigma\varsigma_1 - \sigma\varsigma_3 & 0 & 0 \end{pmatrix}$$

en tant que matrice et donc $[P, \varsigma] = (-\kappa\sigma\varsigma_2, \kappa\sigma\varsigma_1 - \sigma\varsigma_3, 0)$ en tant que vecteur colonne. D'où

$$U = \begin{pmatrix} 0 & -\kappa\sigma & 0 \\ \kappa\sigma & 0 & -\sigma \\ 0 & 0 & 0 \end{pmatrix}$$

et donc

$$-\langle \lambda, [P, \varsigma] \rangle = \left\langle \lambda \cdot \begin{pmatrix} 0 & \kappa\sigma & 0 \\ -\kappa\sigma & 0 & \sigma \\ 0 & 0 & 0 \end{pmatrix}, \varsigma \right\rangle$$

En rappelant que λ est identifié au covecteur $(\lambda_1, \lambda_2, \lambda_3)$ de \mathcal{G}^* et que, $\rho = d\theta$ donne la troisième composante des vecteurs de \mathcal{G} , on voit que $-\frac{\partial L}{\partial \theta} \rho(\varsigma) dt = \langle (0, 0, -\frac{\partial L}{\partial \theta} dt), \varsigma \rangle$.²¹ L'équation (iii) peut donc s'écrire sous la forme :

$$(iii)' \quad (d\lambda_1, d\lambda_2, d\lambda_3) + (\lambda_1, \lambda_2, \lambda_3) \begin{pmatrix} 0 & \kappa\sigma & 0 \\ -\kappa\sigma & 0 & \sigma \\ 0 & 0 & 0 \end{pmatrix} dt + (0, 0, -\frac{\partial L}{\partial \theta}) dt = 0.$$

L'équation (i) redonne les trois premières équations (1), (2), (3) du système (2), l'équation (ii) redonne les équations (4) et (5) permettant de calculer les multiplicateurs de Lagrange λ_1 et λ_3 , les deux premières composantes de (iii)' redonnent les équations (7) et (8) de la loi de conservation exprimant que λ reste dans les orbites de la représentation co-adjointe de G , et enfin la troisième composante de (iii)' redonne l'équation d'Euler (6) qui permet de calculer le multiplicateur λ_2 .

21. Rappelons que les vecteurs de \mathcal{G} sont des vecteurs de dimension 3 qui sont représentés sous la forme de matrices 2×2 et que, suivant le contexte, on les note comme des vecteurs colonnes ou comme des matrices. Idem pour les covecteurs de \mathcal{G}^* .

5.5. Retour sur les modèles du fibré en cercle et des elasticæ

Ce formalisme de Bryant et Griffiths peut facilement être appliqué à d'autres lagrangiens.

Pour le modèle du fibré en cercles issu du lagrangien de la section 7 du chapitre 10, on a

$$L = (dx^2 + dy^2 + d\theta^2)^{\frac{1}{2}} = \sigma (1 + \kappa^2)^{\frac{1}{2}}$$

et l'on obtient les équations :

$$\begin{cases} \lambda_1 &= (1 + \kappa^2)^{-\frac{1}{2}} \\ \lambda_3 &= \kappa (1 + \kappa^2)^{-\frac{1}{2}} \\ \lambda_2 &= -\dot{\kappa} (1 + \kappa^2)^{-\frac{3}{2}} \end{cases}$$

La loi de conservation fournit alors l'équation différentielle :

$$(\dot{\kappa})^2 + (1 + \kappa^2)^2 = \tilde{\lambda}^2 (1 + \kappa^2)^3$$

qui n'est rien d'autre que l'équation (14) de la section 7 du chapitre 10 avec $\tilde{\lambda} = \frac{1}{k}$.

Pour les elasticæ, le lagrangien général est

$$L = (\alpha\kappa^2 + \beta) \sigma .$$

On a donc

$$\frac{\partial L}{\partial \sigma} = \alpha\kappa^2 + \beta, \quad \frac{\partial L}{\partial \kappa} = 2\alpha\kappa\sigma, \quad \frac{\partial L}{\partial \theta} = 0 .$$

L'équation (5) $\frac{\partial L}{\partial \kappa} - \lambda_3\sigma = 0$ donne $\lambda_3 = 2\alpha\kappa$, l'équation (4) donne $(\alpha\kappa^2 + \beta) - \lambda_3\kappa - \lambda_1 = 0$, soit $\lambda_1 = -\alpha\kappa^2 + \beta$. Enfin l'équation d'Euler donne $2\alpha d\kappa + \lambda_2 ds = 0$, soit $\lambda_2 = -2\alpha\dot{\kappa}$. D'où en définitive le système :

$$\begin{cases} \lambda_1 &= -\alpha\kappa^2 + \beta \\ \lambda_3 &= \lambda_3 = 2\alpha\kappa \\ \lambda_2 &= -2\alpha\dot{\kappa} \end{cases}$$

La loi de conservation $\lambda_1^2 + \lambda_2^2 = c^2$ donne l'équation pour κ :

$$4\alpha^2\dot{\kappa}^2 + \alpha^2\kappa^4 = 2\alpha\beta\kappa^2 + c^2 - \beta^2$$

qui n'est rien d'autre que l'équation (4) de la section 3.4 du chapitre 8 pour $\alpha = \frac{1}{2}$.

Nous voyons ainsi comment les différents modèles de ce chapitre s'interprètent élégamment dans le cadre général de la théorie de Bryant et Griffiths pour les problèmes variationnels sur $SE(2)$.

Lagrangiens, hamiltoniens et contrôle optimal

1. Une double généralisation

Pour préciser nos modèles nous aurons besoin d'un approfondissement substantiel des équations d'Euler-Lagrange du calcul des variations classique. Nous avons vu plus haut, en particulier dans la section 9 du chapitre 9 que, classiquement, les équations d'E-L

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) = \frac{\partial L}{\partial q_i}$$

concernent la minimisation d'une intégrale $S(\gamma) = \int_{t_1}^{t_2} L(q(t), \dot{q}(t)) dt$ le long d'une trajectoire $\gamma = q(t)$ dans l'intervalle $[t_1, t_2]$, le Lagrangien $L(q, \dot{q})$ (supposé ici indépendant du temps) étant défini sur le fibré tangent TM d'une variété différentiable M représentant l'espace de configurations du système considéré. Nous avons insisté sur le fait que les dérivées partielles $\frac{\partial L}{\partial q_i}$ ne dépendent que de M et de L alors que les dérivées $\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right)$ sont prises le long des trajectoires $q(t)$, L devenant une fonction composée de t , $L(q(t), \dot{q}(t))$, le long des trajectoires.

Cette approche classique a été généralisée d'une double façon :

1. d'abord, par la transformation de la mécanique lagrangienne en mécanique *hamiltonienne* ;
2. ensuite, dans la seconde moitié du XX^e siècle, par la théorie générale du *contrôle optimal* dont elle devint un cas particulier.

Nous allons donner quelques brèves indications sur cette double généralisation.

2. La formulation hamiltonienne

La version hamiltonienne des équations d'E-L consiste à les formuler non plus à partir d'une fonction – le lagrangien – définie sur le fibré tangent TM mais, *dualement*, à partir d'une fonction H – l'hamiltonien – définie sur le fibré cotangent T^*M . Nous en verrons les avantages. L'un des plus notables est l'existence d'une *structure symplectique canonique* sur les fibrés cotangents T^*M .

La littérature à ce sujet est immense puisqu'elle recouvre pratiquement toute la physique et les introductions disponibles sur le Web sont nombreuses et souvent excellentes. Nous conseillerons au lecteur intéressé la consultation de Ralph Abraham-Jerrold Marsden [2], Vladimir Arnold [20], Jean-Pierre Bourguignon [74], Louis Hand-Janet Finch ou Melvin Calkin [87]. Les applications de la mécanique hamiltonienne sont innombrables et d'une très haute complexité. Le lecteur intéressé pourra par exemple consulter la belle présentation par Alain Chenciner [112] du problème à trois corps depuis Poincaré. Le problème à n -corps est celui des trajectoires de n corps s'attirant d'après la loi de la gravitation de Newton ($n = 2$ correspond aux mouvements keplériens). Très vite (depuis Lagrange et Laplace) on s'est rendu compte qu'il était d'une difficulté inouïe et Henri Poincaré a introduit ses "nouvelles méthodes en mécanique céleste" pour essayer de comprendre l'origine de sa complexité. On ne compte plus les travaux qui lui ont été consacrés, avec de grands moments comme celui de la démonstration du théorème KAM de Andreï Kolmogorov, Vladimir Arnold et Jürgen Moser.¹

2.1. La transformation de Legendre

2.1.1. Cas élémentaires.

La dualisation entre version lagrangienne (TM) et version hamiltonienne (T^*M) s'effectue à travers la *transformée de Legendre*. Pour la définir, considérons le cas élémentaire d'une fonction différentiable $L(v) : \mathbb{R} \rightarrow \mathbb{R}$ et supposons-la strictement *convexe*, ce qui signifie que $L''(v) > 0$. L'exemple le plus simple est la parabole convexe $L(v) = v^2$ avec $L''(v) = 2$.

Nous avons déjà noté dans la section 2.1 du chapitre 3 à propos de l'espace des 1-jets et des relevées legendriennes des courbes planes, la dualité projective qui existe entre les points a d'une courbe γ et les tangentes à γ . Si γ est le graphe de $L(v)$, la tangente en $a = (v, L(v))$ a pour pente $p = L'(v)$. Mais si p est une pente, quelles sont les valeurs de v telles que la tangente en $a = (v, L(v))$ soit p ? Ce sont les solutions de $L'(v) = p$ et la transformée de Legendre fournit un algorithme pour les trouver. On introduit la fonction de *deux* variables (i.e. on *double* le nombre de variables)

$$\mathcal{H}(v, p) = pv - L(v) \text{ ,}$$

autrement dit, on considère les droites de pente p fixée $y = pv + c$ et leur intersection avec le graphe γ d'équation $y = L(v)$. Les points d'intersection sont les solutions de $pv + c = L(v)$ et $p = L'(v)$ signifie simplement que la droite de pente p est tangente

1. On pourra aussi consulter la découverte [113] par Alain Chenciner et Richard Montgomery d'une solution particulièrement intéressante du problème à trois corps dans le cas où leurs masses sont égales.

au graphe γ . Autrement dit, on a $\frac{\partial \mathcal{H}}{\partial v} = p - L'(v)$ et on en déduit que

$$L'(v) = p \iff \frac{\partial \mathcal{H}}{\partial v} = 0,$$

les solutions étant des extrema de la fonction $\mathcal{H}_p(v) = \mathcal{H}(v, p)$ à p constant. Comme $L(v)$ est convexe, il n'y a qu'une seule solution $v(p)$ et l'on peut définir ainsi la transformée de Legendre de $L(v)$ comme la fonction

$$H(p) = \mathcal{H}(v(p), p) = pv(p) - L(v(p))$$

On note que

$$\frac{\partial^2 \mathcal{H}}{\partial v^2} = -L''(v) < 0$$

et donc que l'extremum $v(p)$ de $\mathcal{H}_p(v)$ est en fait un *maximum* (strict). On vérifie immédiatement que $-H(p)$ est bien sûr l'ordonnée de la tangente au graphe γ au point $a(p) = (v(p), L(v(p)))$. En effet, cette tangente est d'équation $y = pv + c$ par construction et donc en $a(p)$ on a $y = L(v(p)) = pv(p) + c$, soit

$$c = L(v(p)) - pv(p) = -H(p).$$

Il est important d'avoir l'intuition géométrique de la situation même si cette dernière est triviale. Commençons par le terme pv de $\mathcal{H}(v, p)$. Dans l'espace $\mathbb{R}_{(v,p,y)}^3$, la surface $y = pv$ est une surface réglée \mathcal{P} dite "paraboloïde hyperbolique" engendrée par les deux familles de droites passant par l'origine $y = p_0v$ (p_0 fixé, v variable) et $y = pv_0$ (v_0 fixé, p variable). Le long des diagonales $p = v$ et $p = -v$, \mathcal{P} est une parabole convexe v^2 ou concave $-v^2$ et l'origine est un point col. Les lignes de niveau $pv = z_0$ sont des hyperboles équilatères. La figure 1 montre la structure de \mathcal{P} . À gauche, \mathcal{P} est représentée avec ses deux familles de génératrices et sa projection sur le plan (p, y) . Au milieu, la structure du point col 0 est manifeste. À droite, on voit les lignes de niveau.

Considérons maintenant le cas de la parabole convexe $L(v) = v^2$. On doit considérer l'équation du second degré $v^2 - pv - c = 0$ dont le discriminant est $\Delta = p^2 + 4c$ et les solutions $v = \frac{p \pm \sqrt{\Delta}}{2}$. Pour $\Delta > 0$, i.e. $c > -\frac{p^2}{4}$, les deux racines sont réelles et pour $\Delta < 0$, i.e. $c < -\frac{p^2}{4}$, les deux racines sont complexes conjuguées. La racine double $p = L'(v) = 2v$ correspond à $\Delta = 0$ soit $c = -\frac{p^2}{4}$. La figure 2 montre la transformée de Legendre $H(p) = -c = \frac{p^2}{4}$.

On voit clairement sur la figure 2 que c est bien l'ordonnée *minimale* pour laquelle la droite $y = pv + c$ intersecte le graphe de $L(v)$ et que donc $H(p) = -c$ est bien *maximal*.

La surface \mathcal{H} de $\mathbb{R}_{(v,p,y)}^3$ d'équation $y = pv - v^2$ est engendrée par des paraboles concaves paramétrées par p . Projetées sur le plan (v, y) , celles-ci passent par l'origine

[ptb]

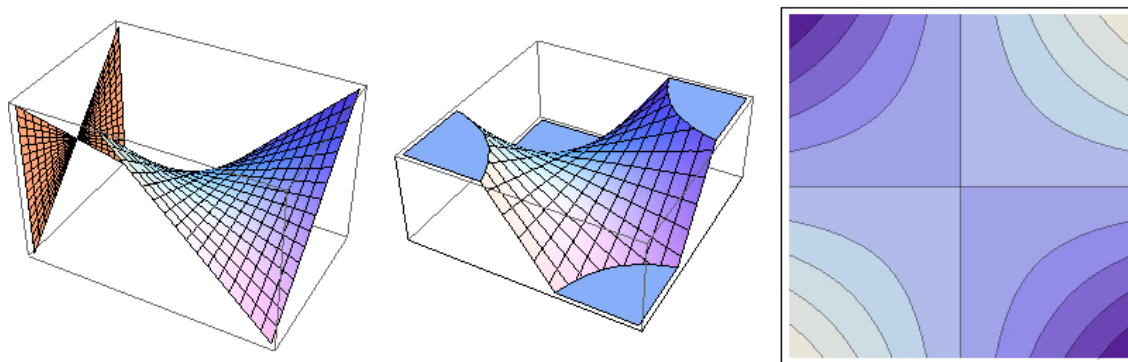


FIGURE 1. La structure du parabolôide hyperbolique \mathcal{P} . À gauche \mathcal{P} avec ses deux familles de génératrices et sa projection sur le plan (p, y) . Au milieu, la structure du point col 0. À droite les lignes de niveau (hyperboles équilatères).

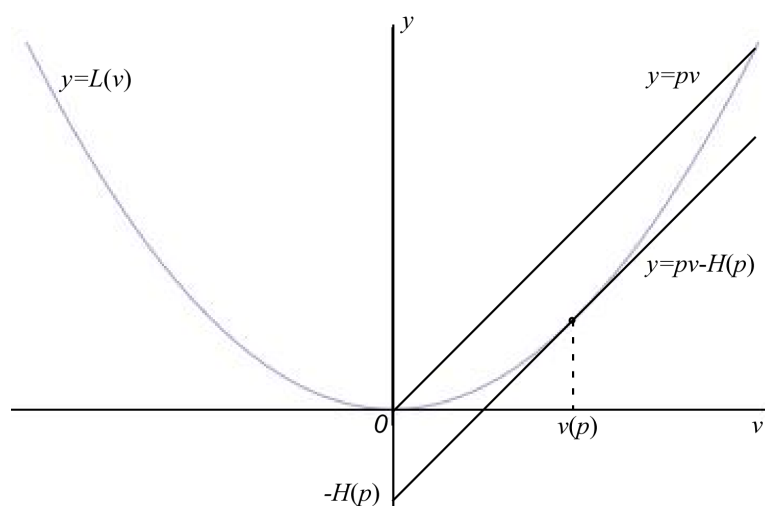


FIGURE 2. La transformée de Legendre du lagrangien quadratique $L(v) = v^2$.

0 et par le point $(p, 0)$ et ont leur sommet au point $\left(\frac{p}{2}, \frac{p^2}{4}\right)$. La figure 3 montre ces projections (à gauche) ainsi que leur déploiement le long de l'axe des p .

Quant à la figure 4, à comparer à la figure 1, elle montre, à gauche, la surface $y = \mathcal{H}(v, p) = pv - v^2$ avec sa projection sur le plan (p, y) et, à droite, ses lignes de niveau. Le contour apparent de \mathcal{H} est la parabole

$$y = \frac{p^2}{4} = H(p) .$$

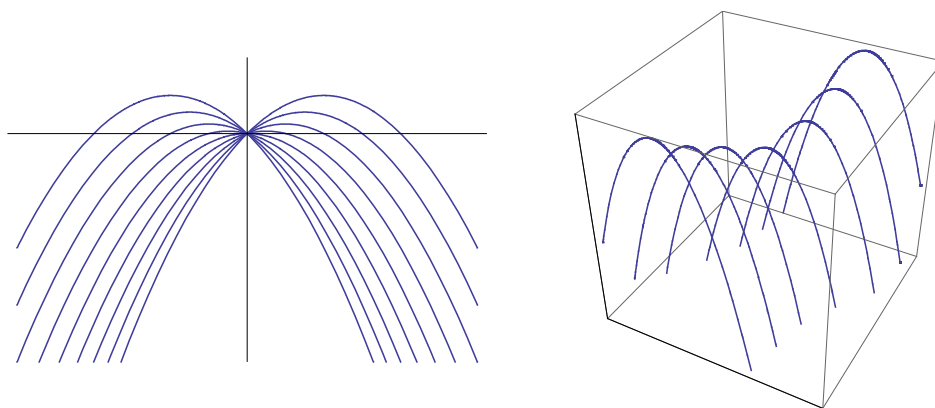


FIGURE 3. *Gauche* : la famille des paraboles concaves $y = pv - v^2$ paramétrées par p . Elles passent par les points $(0, 0)$ et $(p, 0)$ et leur sommet est au point $(\frac{p}{2}, \frac{p^2}{4})$. *Droite* : leur déploiement le long de l'axe des p .

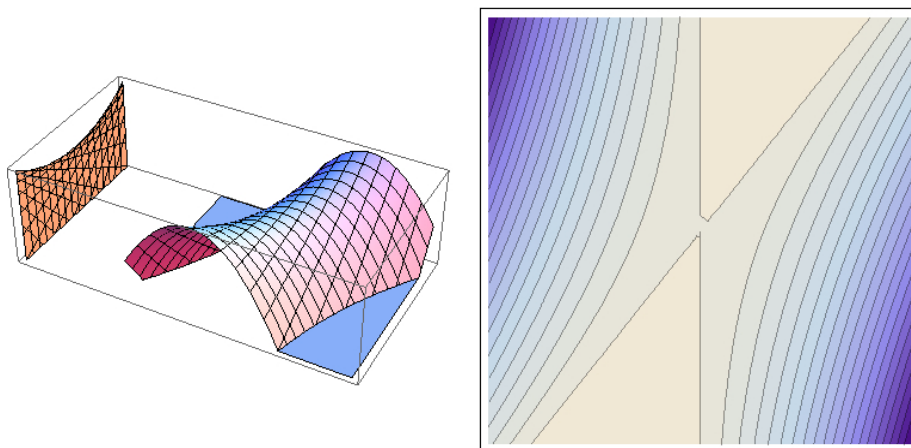


FIGURE 4. *Gauche* : la surface $\mathcal{H}(v, p) = pv - v^2$ avec sa projection sur le plan (p, y) . Son contour apparent est la transformée de Legendre $H(p)$. *Droite* : ses lignes de niveau.

Il faut insister sur cette interprétation géométrique de $H(p)$: la transformée de Legendre $H(p)$ est le *contour apparent* de la surface $\mathcal{H}(v, p)$ sur le plan (p, y) .

Profitons de cette petite illustration du cas le plus trivial pour indiquer un phénomène important qui intervient en cas de non-convexité. Au lieu de considérer un puits de potentiel convexe quadratique comme $L(v) = v^2$, considérons un puits de potentiel *non convexe* avec deux minima séparés par un maximum. Un exemple

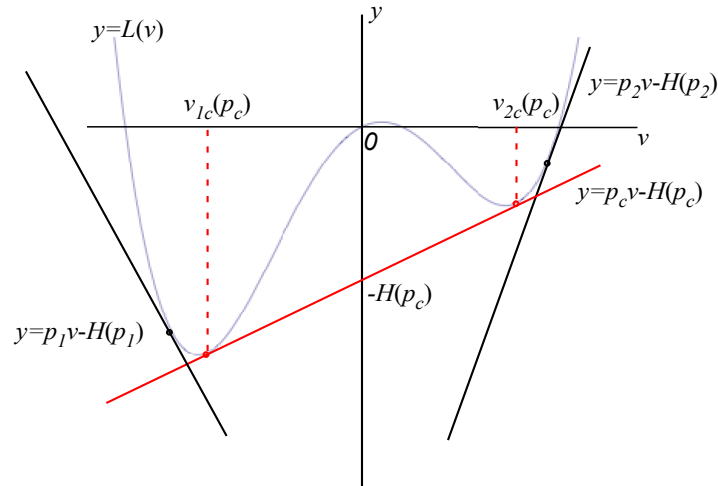


FIGURE 5. Conséquence de la non-convexité de $L(v)$. Pour la tangente critique de pente p_c (en rouge) l'équation $p = L'(v)$ a deux solutions en compétition $v_1(p_c)$ et $v_2(p_c)$ correspondant à une même tangente. Quand on fait tourner p , à la traversée de p_c le point de contact saute brusquement de $v_1(p)$ à $v_2(p)$.

simple est le puits biquadratique

$$L(v) = \frac{v^4}{4} - \frac{v^2}{2} + \frac{v}{8}.$$

Sa transformée de Legendre est représentée à la figure 5 où l'on observe bien la conséquence de la non convexité. Pour les tangentes de pente négative à gauche de la figure, $p_1 < p_{\text{inf } 2}$ = pente du second point d'inflexion, il n'y a qu'un point de contact et il correspond à une racine $v_1(p)$ de l'équation

$$p = L'(v) = v^3 - v + \frac{1}{8}$$

(les deux autres racines étant imaginaires conjuguées). Il en va de même pour les tangentes de pente positive à droite de la figure, $p_2 > p_{\text{inf } 1}$ = pente du premier point d'inflexion, dont le point de contact correspond à une racine $v_2(p)$ (les deux autres racines étant imaginaires conjuguées). Mais pour la tangente critique de pente p_c (en rouge) il y a *deux* points de contact en compétition, autrement dit les deux solutions $v_1(p_c)$ et $v_2(p_c)$ donnent la même tangente. À la traversée de p_c le point de contact *bifurque* donc brusquement de $v_1(p)$ à $v_2(p)$.

Cela se comprend très facilement algébriquement. Une droite quelconque $D_{p,c}$ du plan (v, y) , $y = pv + c$ de pente p et de constante c données, intersecte le graphe γ , $y = L(v)$, de $L(v)$ en au plus 4 points a_i dont les abscisses v_i sont les solutions

réelles de l'équation du 4^e degré $\mathcal{H}(v, p) = pv - L(v) = -c$. Il peut y avoir soit 4 racines réelles, soit 2 racines réelles et deux racines complexes conjuguées, soit 0 racine réelle et deux paires de racines complexes conjuguées. La droite $D_{p,c}$ sera tangente à γ si

$$\frac{\partial \mathcal{H}(v, p)}{\partial v} = p - L'(v) = 0,$$

i.e. $p = v^3 - v + \frac{1}{8}$. On obtient ainsi dans le plan (p, c) paramétrant les droites $D_{p,c}$ une courbe discriminante K séparant les trois domaines où il y a 0, 2, ou 4 racines réelles. Il s'agit de la courbe d'équations paramétriques

$$\begin{aligned} p(v) &= L'(v) = v^3 - v + \frac{1}{8} \\ -c(v) &= pv - L(v) = vL'(v) - L(v) = v \left(v^3 - v + \frac{1}{8} \right) - \left(\frac{v^4}{4} - \frac{v^2}{2} + \frac{v}{8} \right) \\ &= \frac{3v^4}{4} - \frac{v^2}{2} \end{aligned}$$

Elle est représentée dans le plan $(p, y = -c)$ à la figure 6. Il s'agit d'une morphologie K bien connue dite "queue d'aronde". À l'intérieur du triangle curviligne de K défini par le point double et les deux cusps, l'équation $\mathcal{H}(v, p) = -c$ a 4 racines réelles, au-dessus du point double elle a 0 racine réelle et à l'extérieur de ces deux domaines elle a deux racines réelles. Les deux cusps correspondent aux deux points d'inflexion de $L(v)$ car ils sont définis par $p'(v) = 0$ et $y'(v) = 0$, soit $L''(v) = 0$ (ce qui implique automatiquement $y'(v) = L'(v) + vL''(v) - L'(v) = 0$). Le point double $p = \frac{1}{8}$, $y = -c = \frac{1}{4}$ correspond à la tangente double $y = \frac{v}{8} - \frac{1}{4}$ passant par les points $(-1, -\frac{3}{8})$ et $(1, -\frac{1}{8})$. K est le contour apparent de la surface $y = \mathcal{H}(v, p)$ sur le plan (p, y) et le graphe de la transformée de Legendre $H(p)$ qui, à cause de la non-convexité de $L(v)$, n'est plus une fonction univoque.

La figure 7 montre à gauche les graphes de $pv - L(v)$ déployés par rapport à p , autrement dit les sections de la surface $y = \mathcal{H}(v, p)$ pour p constant. À droite, elle montre les lignes de niveau de la surface. La figure 8 montre quant à elle la surface avec sa projection sur le plan $(p, y = -c)$. On y voit bien la queue d'aronde qui est son contour apparent et le graphe de la transformée de Legendre $H(p)$.

2.1.2. Cas général.

Nous avons vu à la section 9 du chapitre 9 que pour le mouvement $x(t)$ d'un point matériel d'énergie cinétique $T = \frac{1}{2}m \|\dot{x}(t)\|^2$ et d'énergie potentielle $V(x(t))$, l'équation de Newton du mouvement équivaut aux équations d'Euler-Lagrange (E-L) pour le lagrangien $L = T - V$. Comme équations minimisant une fonctionnelle, ces équations minimisent l'action $S(\gamma) := \int_{t_0}^{t_1} L(x, \dot{x}) dt$ le long des trajectoires γ allant de $x_0 = x(t_0)$ à $x_1 = x(t_1)$. Cela se généralise aux mouvements de systèmes plus

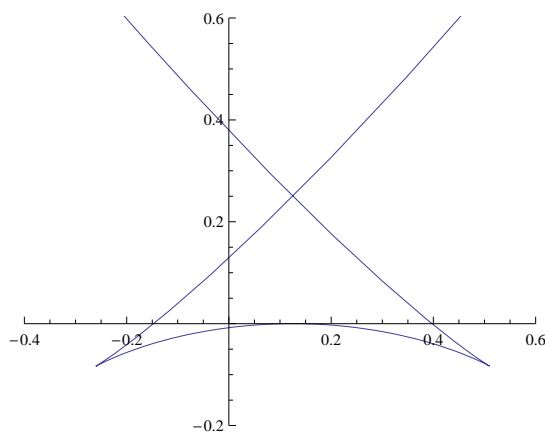


FIGURE 6. La queue d'aronde K discriminant dans le plan $(p, y = -c)$ les droites $D_{p,c}$ ayant 0, 2, ou 4 point d'intersection avec le graphe de $L(v)$.

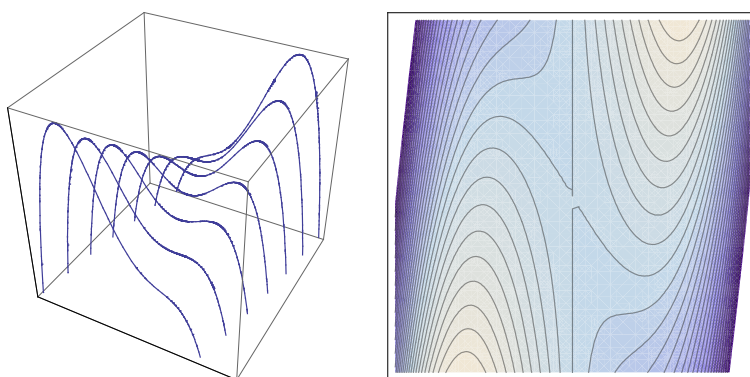


FIGURE 7. *Gauche* : sections de la surface $y = \mathcal{H}(v, p)$ pour p constant. *Droite* : lignes de niveau de la surface.

complexes dont les états instantanés sont descriptibles par les points q d'un espace de configuration M (une variété différentiable) de dimension n . Le lagrangien $L(q, v)$ est alors défini sur le fibré tangent TM de M et, le long d'une trajectoire $q(t)$, devient une fonction $L(q, \dot{q})$.

Remarque sur les notations. Comme nous l'avons déjà mentionné, il est traditionnel en mécanique de noter q les coordonnées généralisées de M et p les "moments conjugués", i.e. les covecteurs éléments du fibré cotangent T^*M . Nous adoptons donc ici ces notations. Dans les chapitres suivants nous changerons parfois cette notation pour des raisons de contexte. \square

L'hamiltonien

$$H(q, p) = \langle p, v \rangle - L(q, v)$$

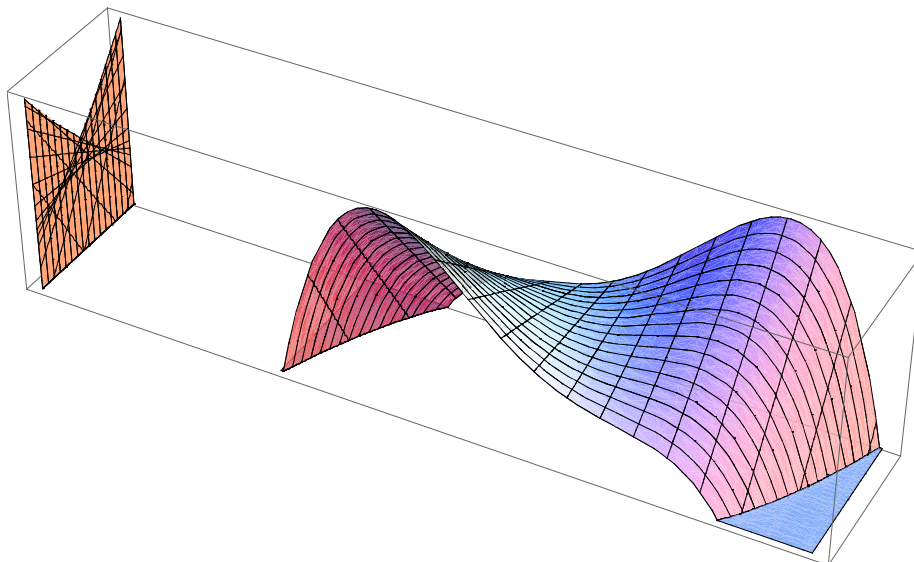


FIGURE 8. Surface $y = \mathcal{H}(v, p)$ avec sa projection sur le plan (p, y) . La queue d'aronde est son contour apparent.

du système est alors la transformée de Legendre du lagrangien $L(q, v)$ par rapport à v . On introduit les moments conjugués p_i ($i = 1, \dots, n$) des coordonnées q_i . Les p appartiennent aux espaces cotangents T_q^*M et $\langle p, v \rangle$ est la valeur du covecteur (de la 1-forme) $p \in T_q^*M$ sur le vecteur tangent $v \in T_qM$.²

En mécanique classique, L dépend souvent de q à travers le potentiel $V(q)$ et de $v = \dot{q}$ à travers l'énergie cinétique qui est une fonction quadratique convexe. La transformée de Legendre par rapport à v est donc bien définie et est paramétrée par q . D'après ce que nous avons vu à la section précédente, elle définit la vitesse $v = \dot{q}$ comme solution de l'équation $p = \frac{\partial L}{\partial v}$, condition qui exprime que $H(q, p)$ maximise la fonction de $3n$ variables

$$\mathcal{H}(q, v, p) = \langle p, v \rangle - L(q, v)$$

puisqu'alors

$$\frac{\partial \mathcal{H}}{\partial v} = p - \frac{\partial L}{\partial v} = 0 .$$

Autrement dit, d'un côté on *ajoute* les n variables supplémentaires p et, d'un autre côté, on *élimine* les n variables v par la condition de maximisation donnant la

2. Dans les présentations élémentaires où $M = \mathbb{R}^n$, on utilise le fait que les espaces tangents $T_q\mathbb{R}^n$ et cotangents $T_q^*\mathbb{R}^n$ sont isomorphes à \mathbb{R}^n pour les identifier entre eux. Un covecteur p devient ainsi un vecteur tangent de même type que v et l'on transforme la dualité $\langle p, v \rangle$ en un produit scalaire $p.v = \sum_{i=1}^n p_i v_i$ qui exprime la structure euclidienne de \mathbb{R}^n .

transformée de Legendre. Un calcul élémentaire montre alors que les équations d'E-L sont équivalentes aux *équations de Hamilton* définies de la façon suivante (dans le cas élémentaire $M = \mathbb{R}^n$).

Écrivons la différentielle totale de H

$$dH = \frac{\partial H}{\partial p} dp + \frac{\partial H}{\partial q} dq$$

et utilisons une notation de type produit scalaire $p.v$ (ce qui signifie qu'on calcule en termes de composantes dans les bases choisies). Cette différentielle dH doit être celle de $p.v - L(q, v)$ avec la condition $p = \frac{\partial L}{\partial v}$. Donc

$$\begin{aligned} d(p.v - L(q, v)) &= v.dp + p.dv - \frac{\partial L}{\partial v} dv - \frac{\partial L}{\partial q} dq \\ &= v.dp - \frac{\partial L}{\partial q} dq \quad (\text{car } p = \frac{\partial L}{\partial v}) \\ &= \dot{q}.dp - \dot{p}dq \quad (\text{car } \frac{\partial L}{\partial q} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) = \dot{p} \text{ par E-L}). \end{aligned}$$

On obtient par conséquent, sous la double condition d'extrémalité

$$\frac{\partial \mathcal{H}}{\partial v} = p - \frac{\partial L}{\partial v} = 0$$

et des équations d'E-L, les équations, dites de Hamilton,

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}.$$

Notons que ces équations impliquent immédiatement le *théorème de Liouville* disant que le flot hamiltonien conserve le volume. En effet, pour un flot $\dot{x} = f(x)$ la conservation du volume s'exprime par l'annulation de la divergence de f . Ici,

$$\operatorname{div}(f) = \frac{\partial}{\partial q} \left(\frac{\partial H}{\partial p} \right) + \frac{\partial}{\partial p} \left(-\frac{\partial H}{\partial q} \right)$$

et donc, trivialement, $\operatorname{div}(f) = 0$.

Toujours en mécanique classique avec $L = T - V$, on a $H = T + V$. En effet (toujours dans le cas élémentaire $M = \mathbb{R}^n$), $T = \frac{1}{2}m \|\dot{q}\|^2$ et $V = V(q)$ et donc $p = \frac{\partial L}{\partial \dot{q}} = m\dot{q}$,

$$H = m\dot{q}.\dot{q} - \frac{1}{2}m \|\dot{q}\|^2 + V = T + V$$

L'hamiltonien H est donc l'*énergie* du système. Dans le cas où, comme ci-dessus, le lagrangien L est indépendant du temps, H est *constant* le long des trajectoires. Si le lagrangien dépend du temps, alors des calculs analogues à ceux que nous venons d'effectuer montrent que $\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}$.

2.2. Les conditions de Legendre et de Weierstrass

Les équations de Hamilton sont des conditions du premier ordre d'extrémalité du coût énergétique qu'est l'intégrale d'action $S(\gamma) = \int_{t_0}^{t_1} L(q(t), v(t)) dt$ le long des trajectoires γ du système. Elles expriment, nous l'avons vu, que la "dérivée première" de la fonctionnelle $S(\gamma)$ s'annule, $\left. \frac{\delta S(\gamma_s)}{\delta s} \right|_{s=0} = 0$, le long des petites perturbations (à extrémités q_0 et q_1 fixes) $\gamma_s = \gamma_* + s\zeta$ de la trajectoire extrémale γ_* . Pour que la trajectoire extrémale soit un minimum – ce que l'on appelle un "minimiseur" –, il faut en plus que la "dérivée seconde" en $s = 0$ soit strictement > 0 , $\left. \frac{\delta^2 S(\gamma_s)}{\delta s^2} \right|_{s=0} > 0$, le long de γ_* . C'est ce que l'on appelle la *condition de Legendre*. Elle est du deuxième ordre.

Il y a en plus une troisième condition – dite *condition de Weierstrass* – qui exprime la convexité globale du lagrangien $L_q(v) = L(q, v)$ comme fonction de v , convexité qui, nous l'avons vu, permet de définir univoquement l'hamiltonien par transformée de Legendre. Si $(v_0, L_q(v_0))$ est un point du graphe de L_q , la convexité dit que le graphe est au-dessus de la tangente en ce point d'équation

$$y = \frac{\partial L(q, v_0)}{\partial v} (v - v_0) + L(q, v_0) .$$

Pour tout v_0 on a donc pour tout v la condition

$$L(q, v) - L(q, v_0) - \frac{\partial L(q, v_0)}{\partial v} (v - v_0) \geq 0 .$$

3. Théorème de Noether et application moment

Récapitulons. Soit Σ un système mécanique, M son espace de configurations et $L : TM \rightarrow \mathbb{R}$ son lagrangien. L'intégrale d'action est

$$S(\gamma) = \int_{t_0}^{t_1} L(q, \dot{q}) dt$$

où γ est un chemin de (q_0, t_0) à (q_1, t_1) . Le principe de moindre action $\delta S = 0$ implique les équations d'Euler-Lagrange :

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) - \frac{\partial L}{\partial q} = 0$$

qui, répétons-le, sont les équations de Newton $\dot{p} = \partial L / \partial q$ pour les moments généralisés $p = \partial L / \partial \dot{q}$ et les forces généralisées $\partial L / \partial q$.

Dans la version hamiltonienne, $H : T^*M \rightarrow \mathbb{R}$ est l'hamiltonien de Σ , défini sur le fibré cotangent T^*M (l'espace des phases de Σ) muni de sa 2-forme *symplectique*

$\omega = \omega_0$ (localement de forme de Darboux $dp \wedge dq$ en coordonnées canoniques).³ Si $F : T^*M \rightarrow \mathbb{R}$ est une observable, on lui associe le champ hamiltonien X_F défini par $i_{X_F}\omega = dF$.

3.1. Le théorème de Noether

L'un des théorèmes fondamentaux des systèmes régis par des principes variationnels de moindre action (lagrangiens et hamiltoniens) est le théorème de Noether établissant une correspondance entre les *symétries* du système Σ et ses *lois de conservation*. Nous l'avons déjà brièvement évoqué dans la section 1.4 du chapitre 4.

3.1.1. *Version lagrangienne.* Soit φ_t un sous-groupe à 1-paramètre de difféomorphismes de M qui sont des symétries du lagrangien L au sens où

$$L(v) = L(D_q\varphi_t(v))$$

pour tout vecteur tangent $v \in T_qM$ ($D_q\varphi_t$ étant l'application linéaire tangente de φ_t en q). Soit X_φ le champ de vecteurs sur M engendrant le flot φ_t . Alors Σ admet la loi de conservation (l'intégrale première) :

$$I : TM \rightarrow \mathbb{R}, (q, \dot{q}) \mapsto I(q, \dot{q}) = \left. \frac{\partial L}{\partial \dot{q}} \cdot \frac{d\varphi_s(q)}{ds} \right|_{s=0} = p \cdot X_\varphi(q) .$$

3.1.2. *Version hamiltonienne.* Supposons que l'hamiltonien H soit invariant sous l'action du flot $\exp(tX_F)$ engendré par le champ hamiltonien X_F . Alors F est une intégrale première de X_H au sens où le crochet de Poisson de F et H s'annule :

$$\{F, H\} = \omega(X_F, X_H) = i_{X_F}i_{X_H}\omega = 0 .$$

Les trois exemples classiques du théorème de Noether sont les groupes de relativité de la Mécanique classique résumés dans le tableau suivant :

<i>Non observabilité d'entités absolues</i>	<i>Groupes de symétrie</i>	<i>Lois de conservation de quantités physiques</i>
Origine du temps	Translations temporelles	Énergie
Origine de l'espace	Translations spatiales	Moment-impulsion
Direction privilégiée	Rotations	Moment angulaire

3. Dans le cas des structures de contact, nous avons noté en général ω la forme de contact de base qui définit la structure et Ω sa différentielle extérieure qui est une forme symplectique sur les hyperplans de contact. Nous pourrions continuer à noter Ω les 2-formes symplectiques, mais dans le cas des variétés symplectiques où il s'agit de la forme essentielle définissant la structure on la note souvent ω . Ce que nous faisons ici.

Cette corrélation entre l'impossibilité d'observer certaines entités absolues et, au contraire, la possibilité d'observer et de mesurer des grandeurs corrélatives conservées est fondamentale. Elle fait partie de ce cadre ("background structure") synthétique a priori qu'imposent les groupes de relativité et dont nous avons déjà traité à propos de la philosophie transcendantale dans la section 1.4 du chapitre 4. Comme le notent Gilles Cohen-Tannoudji et Michel Spiro dans [124],

"la relativité signifie l'impossibilité d'effectuer des mesures *absolues*"

et le théorème de Noether établit

"la correspondance entre *relativité* (non observabilité de certaines entités absolues), *symétries* (invariance par transformation de symétrie) et *lois de conservation* (conservation et donc observabilité de certaines quantités)." ([124], p. 106-107)

3.2. L'application moment

Le caractère synthétique a priori du théorème de Noether peut être considérablement renforcé si l'on se réfère aux travaux mathématiques de Bertram Kostant, Jean-Marie Souriau [509], Vladimir Arnold [20], Alan Weinstein [560], Ralph Abraham & Jerrold Marsden [2] et de nombreux autres spécialistes de géométrie symplectique. Avec le formalisme de *l'application moment*, il devient en effet possible de déduire *directement* des groupes de relativité certaines intégrales premières *préalablement* à la donnée de tout lagrangien ou hamiltonien spécifiques.

L'espace des phases T^*M de Σ est une variété symplectique (P, ω) . Soit G un groupe de Lie agissant *symplectiquement* sur P à travers une action $\Phi : G \times P \rightarrow P$ (pour tout $g \in G$ les $\Phi_g(x)$ sont donc des symplectomorphismes appartenant au groupe symplectique $\text{Sympl}(P, \omega)$). Il y a essentiellement deux façons d'engendrer des champs de vecteurs sur P : soit à partir des symétries, soit à partir des observables.

1. Soit $\mathcal{G} = T_e G$ l'algèbre de Lie de G et soit $\xi \in \mathcal{G}$. À travers l'action Φ , le sous-groupe à 1-paramètre $\exp(t\xi)$ agit sur P comme le flot $\varphi_\xi(t) = \Phi_{\exp(t\xi)}$. Soit

$$\xi_P = \left. \frac{d}{dt} \varphi_\xi(t) \right|_{t=0}$$

le générateur infinitésimal du flot $\varphi_\xi(t)$. On a par construction $\xi_P \in \mathfrak{X}(P)$, où $\mathfrak{X}(P)$ est l'algèbre des champs de vecteurs sur P . L'application $\xi \mapsto \xi_P$ est un morphisme d'algèbres de Lie $\mathcal{G} \rightarrow \mathfrak{X}(P)$, i.e. $[\xi_P, \eta_P] = [\xi, \eta]_P$. Comme G agit symplectiquement, $i_{\xi_P} \omega$ est une 1-forme *fermée* et même exacte si G est connexe. Dans ce cas on a

$$i_{\xi_P} \omega = dF_\xi$$

pour une certaine fonction F_ξ .

2. Soit dans l'autre sens $F : P \rightarrow \mathbb{R}$ une observable. F engendre le champ hamiltonien X_F tel que

$$i_{X_F}\omega = dF .$$

L'idée est alors d'associer à chaque $\xi \in \mathcal{G}$ – de façon naturelle et automatique – une observable F_ξ telle que :

- $X_{F_\xi} = \xi_P$,
- les F_ξ sont des intégrales premières de tout hamiltonien H qui est G -invariant (i.e. tel que $H(x) = H(\Phi_g(x))$, pour tout $g \in G$).

Supposons qu'une telle association existe. Si l'on fixe un point $x \in P$, on peut considérer le système $J(x)$ des valeurs $F_\xi(x)$ (x constant, ξ variable) de l'intégrale première F_ξ . $J(x)$ est une forme linéaire sur \mathcal{G} , i.e. un élément du dual \mathcal{G}^* de \mathcal{G} . On obtient ainsi une application – *l'application moment* :

$$\begin{aligned} J : P &\rightarrow \mathcal{G}^* \\ x \mapsto J(x) : \mathcal{G} &\rightarrow \mathbb{R} \\ \xi \mapsto J(x)(\xi) &= F_\xi(x) . \end{aligned}$$

Théorème de Noether. Si $H : P \rightarrow \mathbb{R}$ est G -invariant, J est un ensemble d'intégrales premières de X_H .

Remarque. Si $(P, \omega) = (T^*M, \omega_0)$ et si G agit sur M , on a alors

$$\begin{aligned} J : T^*M &\rightarrow \mathcal{G}^* \\ \alpha \in T_q^*M &\mapsto J(\alpha) : \xi \rightarrow J(\alpha)(\xi) = F_\xi(\alpha) = \alpha(\xi_M(q)) . \end{aligned}$$

On a donc $F_\xi = P(\xi_M)$ où l'application $P(X)$ est définie pour $X \in \mathfrak{X}(M)$ par

$$\begin{aligned} P(X) : T_q^*M &\rightarrow \mathbb{R} \\ \alpha \in T_q^*M &\mapsto \alpha(X(q)) . \end{aligned}$$

3.3. Symétries et mécanique a priori

Nous venons de voir que, grâce à l'application moment, des intégrales premières peuvent être associées directement (i.e. préalablement à tout hamiltonien) à des groupes de relativité agissant sur des variétés symplectiques. Mais ce caractère "synthétique a priori" des lois de conservation exprimé par le théorème de Noether peut être poussé encore plus loin. Il est en effet possible d'appliquer ces constructions aux groupes de Lie G eux-mêmes en utilisant les résultats fondamentaux d'Alexandre Kirillov sur le fait que, comme nous l'avons expliqué plusieurs fois, les orbites de la représentation co-adjointe de G sont canoniquement munies d'une structure symplectique. Cela permet de dériver des hamiltoniens *directement* des symétries.

Soit donc G un groupe de Lie. G agit sur lui-même par translations à gauche $L_g : h \rightarrow gh$. Rappelons (cf. la section 6 du chapitre 5) que la représentation adjointe

est donnée par

$$\begin{aligned} Ad : G &\rightarrow \text{Aut } \mathcal{G} \\ g &\mapsto Ad_g = T_e(R_{g^{-1}} \circ L_g) \end{aligned}$$

et

$$\begin{aligned} ad : \mathcal{G} &\rightarrow \mathcal{G} \\ \xi &\mapsto ad_\xi : \eta \mapsto ad_\xi(\eta) = [\xi, \eta] \end{aligned}$$

Quant à la représentation co-adjointe, elle est donnée de façon duale par

$$\begin{aligned} Ad^* : G &\rightarrow \text{Aut } \mathcal{G}^* \\ g &\mapsto Ad_g^* : \alpha \mapsto Ad_g^*(\alpha) : \eta \mapsto Ad_g^*(\alpha)(\eta) = \alpha(Ad_{g^{-1}}\eta) \end{aligned}$$

et

$$\begin{aligned} ad^* : \mathcal{G} &\rightarrow \mathcal{G}^* \\ \xi &\mapsto ad_\xi^* : \alpha \mapsto ad_\xi^*(\alpha) : \eta \mapsto ad_\xi^*(\alpha)(\eta) = \alpha(ad_{-\xi}(\eta)) = -\alpha([\xi, \eta]). \end{aligned}$$

Pour aller plus loin on a besoin d'introduire la notion de *structure de Poisson* due à Alan Weinstein. Si (P, ω) est une variété symplectique de coordonnées canoniques (q_i, p_j) , le crochet de Poisson entre deux fonctions $f, g \in \mathcal{C}^\infty(P)$

$$\{f, g\} = \omega(X_f, X_g) = \sum_i \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right)$$

définit une structure de Poisson (SP) sur P , ce qui signifie que $\{f, g\}$ est un crochet de Lie sur l'algèbre $\mathcal{C}^\infty(P)$ qui est relié au produit de $\mathcal{C}^\infty(P)$ par une loi de dérivation

$$\{fg, h\} = f\{g, h\} + \{f, h\}g .$$

Il est possible de reformuler toute la mécanique hamiltonienne en termes de SP. Si $H \in \mathcal{C}^\infty(P)$ est un hamiltonien sur P , le champ hamiltonien X_H (considéré comme une dérivation sur $\mathcal{C}^\infty(P)$) est défini par

$$X_H(f) = \{f, H\} .$$

De même, si $\theta : \mathcal{G} \rightarrow \mathcal{C}^\infty(P)$ est un morphisme d'algèbres de Lie, l'application moment est donnée par

$$\begin{aligned} J : P &\rightarrow \mathcal{G}^* \\ x &\mapsto J(x) : \xi \mapsto J(x)(\xi) = \theta(\xi)(x) . \end{aligned}$$

Considérons alors le diagramme commutatif de fibrations :

$$\begin{array}{ccc} T^*P & \xrightarrow{\Pi} & TP \\ dH \uparrow \downarrow & & \downarrow \uparrow X_H \\ P & \xrightarrow{Id_P} & P \end{array}$$

où Π est le morphisme défini par $X_H = \Pi \circ dH$. Parce qu'elle est issue d'une structure symplectique, la structure de Poisson SP est *non dégénérée* au sens où Π est un *isomorphisme* de fibrations. Ce n'est pas le cas en général mais, néanmoins,

l'image de Π est en général au moins un *feuilletage intégrable* F de P dont les feuilles $F(x)$ sont des variétés symplectiques, l'inclusion canonique $F(x) \hookrightarrow P$ respectant la SP.

Un théorème de Weinstein affirme que toute SP est *localement* le produit direct d'une SP symplectique et d'une SP "transversale" tangente à la SP d'une algèbre de Lie \mathcal{G}^* . Les SP sur les algèbres de Lie \mathcal{G} sont celles qui transforment toutes les applications moment $J : P \rightarrow \mathcal{G}^*$ en applications de Poisson. Elles sont dégénérées. Leurs feuilles symplectiques sont les orbites de la représentation co-adjointe d'un groupe de Lie G d'algèbre de Lie \mathcal{G} . Ces orbites ont une structure symplectique de Kirillov définie de la façon suivante. Soit $\alpha \in \mathcal{G}^*$ et soit $G_\alpha = \{Ad_g^* \alpha\}_{g \in G}$ son orbite. Soit $\theta \in T_\alpha G_\alpha$. On peut supposer que $\theta \in \mathcal{G}^*$ car, \mathcal{G}^* étant un espace vectoriel, on a $T_\alpha \mathcal{G}^* \cong \mathcal{G}^*$. Quand θ est interprété comme un vecteur vitesse en α , il est identifié avec le $\eta \in \mathcal{G}$ tel que $ad_\eta^* \alpha = \theta$. La structure de Kirillov est alors donnée par le crochet de Lie $\sigma(\theta_1, \theta_2) = \alpha([\eta_1, \eta_2])$. En utilisant l'application moment $J : T^*G \rightarrow \mathcal{G}^*$ de l'action de G sur lui-même par translations à gauche, on peut montrer que cette structure symplectique provient, par une opération de "réduction", de la structure symplectique canonique ω_0 de T^*G . Il est par conséquent possible d'engendrer des systèmes hamiltoniens directement à partir de symétries.

4. La théorie du contrôle optimal

4.1. Une vaste généralisation

Une vaste généralisation du calcul variationnel classique dans ses versions lagrangiennes et hamiltoniennes s'est développée avec la théorie du contrôle et, en particulier, du *contrôle optimal*. L'intuition de base reste la même, celle de la trajectoire d'un mobile dans un espace de configurations ou un espace des phases, mais au lieu de penser à des mobiles soumis passivement à des champs de forces externes (comme des points matériels ou des particules en mécanique), on pense plutôt à des mobiles comme des automobiles dont des conducteurs peuvent contrôler les directions avec un volant et les accélérations avec des accélérateurs et des freins. Il s'agit toujours de minimiser un "coût" analogue à une intégrale d'action en mécanique mais désormais il s'agit de calculer le contrôle "optimal" permettant d'effectuer cette minimisation. La version hamiltonienne est bien adaptée à cette situation et les équations d'E-L se trouvent généralisées au moyen d'un théorème fondamental appelé *Principe du Maximum de Pontryagin* (PMP). Comme le dit Andrew Lewis dans son excellente présentation pédagogique (niveau Master) mais néanmoins détaillée de ce théorème [329],

"The Maximum Principle of Pontryagin is (...) a far reaching generalization of the classical subject of the calculus of variation."

La littérature sur le contrôle optimal et le PMP est énorme et très technique car on cherche à obtenir des résultats très généraux. Pour cela, on doit utiliser de nombreux résultats fins et compliqués d'analyse fonctionnelle. Il nous sera donc impossible d'en présenter ici ne serait-ce que des rudiments rigoureux. Nous essayerons néanmoins d'en fournir une introduction simplifiée pas trop inexacte en réservant les précisions aux cas concrets particuliers où les calculs explicites permettront de mieux comprendre certains détails techniques. Nous donnerons plusieurs références bibliographiques au cours de cette esquisse.

On considère toujours un espace de configurations M qui est une variété différentiable de dimension n . Pour simplifier, nous supposons que M est un domaine de \mathbb{R}^n et que l'on dispose donc de coordonnées globales q_i , $i = 1, \dots, n$. On suppose également que l'on s'est donné des contrôles u variant dans un domaine U de \mathbb{R}^m et que l'on dispose donc de coordonnées u_j , $j = 1, \dots, m$. La base de la généralisation est que le système est régi par une équation différentielle

$$\dot{q} = X(q, u) \in T_q M \quad (\text{E})$$

où $X(q, u)$ est un champ de vecteurs sur M dépendant de u , le contrôle $u(t)$ permettant en quelque sorte de "piloter" le mouvement $q(t)$. Les trajectoires du système sont donc à la fois des trajectoires $q(t)$ dans M et des trajectoires $u(t)$ dans U , ces dernières pouvant être choisies. On peut par exemple avoir des systèmes "affines" avec $X(q, u) = X(q) + uY(q)$ ou des systèmes "linéaires" $X(q, u) = A(q) + B(u)$ avec $A \in L(\mathbb{R}^n, \mathbb{R}^n)$ (i.e. A est une matrice $n \times n$) et $B \in L(\mathbb{R}^m, \mathbb{R}^n)$ (i.e. B est une matrice $m \times n$).

À travers $u(t)$, l'équation différentielle (E) dépend du temps et n'est donc pas "autonome". Qui plus est, les contrôles peuvent imposer non seulement des contraintes holonomes sur les positions q mais aussi des contraintes non holonomes sur les contrôles.

On se donne par ailleurs un "coût"

$$J(\gamma) = \int_{t_0}^{t_1} L(q(t), u(t)) dt$$

qu'il s'agit de minimiser, L étant un lagrangien défini sur $M \times U \subseteq \mathbb{R}^n \times \mathbb{R}^m$. On cherchera donc des contrôles optimaux $u_*(t)$ permettant de définir des trajectoires optimales $q_*(t)$ qui à la fois satisfont (E) et minimisent J .⁴ Il y a plusieurs problèmes d'optimisation. Étant données des conditions aux limites q_0 et q_1 et des intervalles temporels $[t_0, t_1]$, on peut d'abord se demander si q_1 est atteignable à partir q_0 avec des contrôles appropriés : c'est le problème de la *contrôlabilité*. Si la réponse est

4. Un exemple intuitif dont tout le monde a l'expérience est celui de garer une voiture dans un créneau étroit le long d'un trottoir. La suite de manœuvres avant/arrière montre le rôle des changements de contrôle. Garer un camion semi-remorque est encore plus compliqué.

positive, on peut se demander, dans le cas $L = 1$ (i.e. $J = t_1 - t_0$), quel est le temps minimal pour cela : l'intervalle $[t_0, t_1]$ est libre et le problème est celui du *temps optimal*. Si L est plus compliqué, on peut se demander quel contrôle optimal permet de minimiser J avec $[t_0, t_1]$ fixé : c'est le problème du *contrôle optimal à intervalle fixé*. On peut aussi supposer qu'on se fixe seulement des sous-variétés S_0 et S_1 de M et l'on cherche la façon optimale de rejoindre un point initial $q_0 \in S_0$ à un point terminal $q_1 \in S_1$. Etc. Tous les problèmes variationnels que nous avons rencontrés jusqu'ici sont des problèmes de contrôle optimal.

Dans le calcul des variations classiques, le contrôle u s'identifie à \dot{q} et le lagrangien $L(q, u)$ est un lagrangien classique $L(q, \dot{q})$; on retombe alors sur les équations d'E-L. Notons toutefois que l'on s'autorise dans ce cas des restrictions non holonomes sur les vitesses car on doit avoir $\dot{q} \in U$. Les lagrangiens $L(q, \dot{q})$ étant des fonctions sur TM , l'optimisation (E-L) donne des EDO sur TM , i.e. des EDO du 2ème ordre sur M (comme l'équation de Newton).

On pourrait croire que tout problème de contrôle optimal peut se ramener à un problème classique d'équations d'E-L. Mais ce n'est pas le cas comme l'explique bien Andrew Lewis dans [329]. Certes, si l'on dispose d'un lagrangien $L(q, \dot{q})$ et d'une équation (E) $\dot{q} = X(q, u)$ on a automatiquement un lagrangien

$$\tilde{L}(q, u) = L(q, X(q, u))$$

et on peut considérer des hamiltoniens associés

$$\mathcal{H}(q, p, u) = \langle p, X(q, u) \rangle - \tilde{L}(q, u) .$$

Mais maximiser \mathcal{H} par rapport à u ne sera en général pas du tout pareil que maximiser \mathcal{H} par rapport à \dot{q} .

4.2. Conditions et contraintes

La définition même d'un problème de contrôle optimal fait immédiatement apparaître tout un ensemble de conditions et de contraintes qui doivent être supposées satisfaites pour arriver à démontrer des théorèmes à la fois suffisamment opérationnels et suffisamment généraux.

Il y a d'abord des contraintes topologiques sur l'espace de configurations $M \subset \mathbb{R}^n$: est-il ouvert, fermé, avec de l'homologie, etc. Il en va de même pour l'ensemble $U \subset \mathbb{R}^m$ des contrôles "admissibles". Si U est fermé, le bord ∂U de U pourra jouer un grand rôle car si, lors de la minimisation du coût J , le contrôle $u(t)$ rejoint ∂U il devra alors soit suivre ∂U soit réentrer dans U , éventuellement en "ricochant" sur ∂U ou en effectuant un saut. De telles possibilités introduisent un grand nombre de difficultés techniques.

D'autres conditions et contraintes sont nécessaires pour que le problème de contrôle ait un sens. D'abord il faut que l'équation de base (E) puisse être intégrée

et donc que la fonction $X(q, u)$ soit intégrable pour tout q pour les contrôles u . Cela impose des contraintes de *mesurabilité* sur les fonctions de contrôle $u(t)$ admissibles. On peut aussi vouloir non seulement l'existence des solutions mais également l'unicité, un minimum de régularité, et une bonne dépendance par rapport aux conditions initiales. De même, il faut que l'intégrale J soit bien définie. Le lagrangien $L(q(t), u(t))$ doit donc être intégrable par rapport à t pour les trajectoires admissibles. C'est une autre contrainte de mesurabilité sur les $u(t)$.

Les contraintes d'intégrabilité et de mesurabilité sont globales et assez faibles. Elles sont à l'opposé des contraintes de dérivabilité qui sont locales et fortes. Elles imposent peu de régularité. C'est pourquoi on considère en général des classes assez larges de fonctions de contrôle $u(t)$. Mais, du coup, on doit surmonter de nombreuses difficultés d'analyse fonctionnelle pour pouvoir démontrer les théorèmes. On cherchera à généraliser des théorèmes classiques comme celui (de type Carathéodory) disant que si $X(q, u)$ est localement intégrable en $u(t)$ pour tout q et de classe C^1 en q pour tout t , alors l'équation (E) possède pour toute condition initiale (q_0, t_0) une solution locale $q(t)$ absolument continue et déterminée de façon unique.

5. Le principe du maximum de Pontryagin (PMP)

Le théorème fondamental du contrôle optimal est le “principe du maximum de Pontryagin” (PMP). Il a été démontré en 1962 par Lev Pontryagin dans son ouvrage de référence [453] *The Mathematical Theory of Optimal Processes* rédigé avec ses étudiants Vladimir Boltyansky, Revaz Gamkrelidze et Evgeni Mischenko. Il a ensuite été développé en particulier par Gamkrelidze dans son traité de 1978 [203] *Principles of Optimal Control Theory*. Pour des précisions historiques, théoriques et techniques, le lecteur pourra consulter entre autres la synthèse d'Andrei Agrachev et Yuri Sachkov de 2002 [10] *Control Theory from the Geometric Viewpoint* paru dans l'*Encyclopedia of Mathematical Sciences*, l'*Introduction to Optimal Control* par Ugo Boscain et Benetto Piccoli [68], les références Berkovitz [48], Gelfand-Fomin [208], Lee-Markus [319], Lewis [329] et Sussmann-Willems [517] ainsi que leurs bibliographies. Pour des précisions sur les applications concrètes en ingénierie et en automatique, le lecteur pourra se référer à l'ouvrage d'Emmanuel Trélat [541].⁵

5.1. Énoncé

Comme l'expliquent Agrachev et Gamkrelidze dans [9] à l'occasion du cinquantième du théorème, le PMP peut se formuler intuitivement de la façon suivante dans le cas où aucun lagrangien L n'intervient. Si X est un champ de vecteurs sur

5. Directeur du Laboratoire Jacques-Louis Lions de la Sorbonne, Emmanuel Trélat est l'un des principaux responsables du groupe de travail “Contrôle” de ce laboratoire.

une variété M , on peut lui associer canoniquement un hamiltonien $h_X : T^*M \rightarrow \mathbb{R}$ au moyen de la formule de dualité $h_X(p) = \langle p, X \rangle$ ou, plus précisément,

$$h_X(q, p_q) = \langle p_q, X_q \rangle$$

pour tout $q \in M$, p_q et X_q étant les valeurs de p et X au point q . Soit \vec{h}_X le champ hamiltonien sur T^*M associé à l'hamiltonien h_X par les équations de Hamilton. Ce champ \vec{h}_X , qui est par définition une section du fibré $T(T^*M)$, peut être considéré comme un "relèvement" de X (qui est un champ sur M) dans T^*M . On notera la succession "champ sur $M \implies$ fonction sur $T^*M \implies$ champ hamiltonien sur T^*M ".

Dans le contexte de la théorie du contrôle, on se donne une classe de champs $X(u)$ paramétrés par des contrôles u_j et on considère les équations différentielles non autonomes sur M données par $\dot{q}(t) = X(u(t))$. Soit alors

$$H = \underset{u}{\text{Sup}} (h_{X(u)})$$

l'hamiltonien pour lequel les contrôles $u_{j,\max}$ satisfont les équations $\frac{\partial H}{\partial u_j} = 0$ permettant de maximiser les $h_{X(u)}$. Les trajectoires de H fournissent par projection sur M des trajectoires de $\dot{q}(t) = X(u_{\max}(t))$ qui sont des trajectoires en quelque sorte "optimales".

En termes de coordonnées, on peut écrire les choses de la façon suivante. Supposons-nous donné un contrôle $u(t)$. Alors $X(q, u) \in T_qM$ définit un champ non autonome $X_u : M \times \mathbb{R} \rightarrow TM$ défini par

$$X_u(q, t) = X(q, u(t)) .$$

Le relèvement \widetilde{X}_u de X_u dans le fibré tangent TM est un champ sur TM , donc une section du fibré $T(TM)$, qui s'écrit (avec des $v_q \in T_qM$ traités comme des vecteurs colonnes et avec les champs identifiés à des opérateurs de dérivation de fonctions $f(q, v_q)$ sur TM)

$$\widetilde{X}_u(v_q, t) = \sum_{i=1}^{i=n} X_{u,i}(q, u(t)) \frac{\partial}{\partial q_i} + \sum_{i,j=1}^{i,j=n} \frac{\partial X_{u,i}}{\partial q_j}(q, u(t)) v_{q,j} \frac{\partial}{\partial v_{q,i}} .$$

La première somme est la composante "horizontale" de \widetilde{X}_u qui n'est rien d'autre que X_u et la seconde somme est la composante "verticale" dans les fibres T_qM . Elle est linéaire en les v_q , les $\frac{\partial X_{u,i}}{\partial q_j}$ étant une matrice $n \times n$ que l'on peut noter $\frac{\partial X_u}{\partial q}$. Les courbes intégrales sont des champs de vecteurs le long des courbes intégrales de X_u , autrement dit des déformations infinitésimales de ces courbes.

De même, le relèvement $\widetilde{X}_u^* = \vec{h}_{X_u}$, qui est une section du fibré $T(T^*M)$, s'écrit (avec des $p_q \in T_q^*M$ traités comme des vecteurs lignes)

$$\vec{h}_{X_u}(p_q, t) = \sum_{i=1}^{i=n} X_{u,i}(q, u(t)) \frac{\partial}{\partial q_i} - \sum_{i,j=1}^{i,j=n} \frac{\partial X_{u,j}}{\partial q_i}(q, u(t)) p_{q,j} \frac{\partial}{\partial p_{q,i}} .$$

C'est le champ hamiltonien de l'hamiltonien

$$h_{X_u}(p_q, t) = \langle p_q, X_u(q, u(t)) \rangle = \sum_{i=1}^{i=n} p_{q,i} X_{u,i}(q, u(t)) ,$$

la matrice $\left\{ \frac{\partial X_{u,j}}{\partial q_i} \right\}_{i,j} = \left(\frac{\partial X_u}{\partial q} \right)^T$ qui agit sur T^*M étant la transposée de la matrice $\frac{\partial X_u}{\partial q}$ qui agit sur TM .

Dans le cas où le problème de contrôle optimal comporte un lagrangien L , les hamiltoniens considérés sont du type

$$\mathcal{H}(q, p, u) = \langle p, X(q, u) \rangle + \alpha L(q, u)$$

et le PMP dit alors la chose suivante :

Théorème (Principe du Maximum de Pontryagin). Supposons l'intervalle temporel $[t_0, t_1]$ fixé. Si $(q_*(t), u_*(t))$ est une solution optimale du problème considéré (avec ses conditions et contraintes), alors *il existe* une trajectoire non triviale $p_*(t)$ et une constante α (un multiplicateur de Lagrange) satisfaisant soit $\alpha = -1$, soit $\alpha = 0$ mais $p_*(t_0) \neq 0$, tels que

(i) $q_*(t)$ et $p_*(t)$ satisfassent les équations de Hamilton de

$$\mathcal{H}(q, p, u) = \langle p, X(q, u) \rangle + \alpha L(q, u)$$

(i.e. $p_*(t)$ est une courbe intégrale du champ hamiltonien $\vec{\mathcal{H}}_{X_u}$ au-dessus de $q_*(t)$),

(ii) $u_*(t)$ maximise \mathcal{H} pour presque tout $t \in [t_0, t_1]$,

autrement dit que

$$\mathcal{H}(q_*, p_*, u_*) = \text{Sup}_{u \in U} (\mathcal{H}(q_*, p_*, u)) = H(q_*, p_*) .$$

Si de plus le contrôle optimal $u_*(t)$ est *borné*, alors $H(q_*, p_*)$ est constant le long de la trajectoire optimale. Si de plus, $(q_*(t), u_*(t))$ est optimale pour joindre une sous-variété source S_0 de M à une sous-variété but S_1 , on peut choisir $p_*(t)$ de façon à ce que $p_*(t_0)$ s'annule sur le sous-espace tangent $T_{q_0}S_0$ de $T_{q_0}M$ et que $p_*(t_1)$ s'annule sur le sous-espace tangent $T_{q_1}S_1$. Cette dernière propriété est une condition de transversalité. \diamond

Remarque. Lorsque le problème considéré suppose que l'intervalle $[t_0, t_1]$ est libre, alors, si $u_*(t)$ est borné, $H(q_*, p_*)$ est non seulement constant mais $= 0$ le long de la trajectoire optimale. \square

L'idée est donc, répétons-le, que l'équation $\mathcal{H}(q_*, p_*, u_*) = H(q_*, p_*)$ permet, par transformée de Legendre, de calculer u_* à partir des équations de Hamilton.

La résolution peut être compliquée, en particulier lorsque le contrôle optimal est au bord de U , ce qui empêche d'utiliser des méthodes classiques comme le théorème des fonctions implicites. Malgré le résultat fondamental qu'est le PMP comme condition nécessaire du premier ordre pour l'extrémalité, le problème de la "synthèse optimale" reste donc difficile. Ainsi que l'expliquent Ugo Boscain et Benetto Piccoli dans leur *Introduction* [68], il faut d'abord considérer des conditions d'ordre supérieur permettant d'éliminer des solutions qui ne sont pas partout au moins localement optimales. Il faut ensuite essayer de réduire toute trajectoire optimale à une concaténation d'arcs de trajectoires contrôlés par des fonctions de contrôle u_i formant une famille *finie*. Quand elle est réalisable, une telle "réduction dimensionnelle" est opérationnellement importante. Il faut également traiter les cas où les solutions ne sont pas forcément uniques, où il peut y avoir compétition entre deux extrémales de mêmes extrémités et de même coût (théorie du cut locus, nous y reviendrons), où il peut exister des extrémales "singulières" ou "abnormales", etc.

Avant de donner un exemple élémentaire et quelques idées intuitives de la preuve, faisons quelques citations du texte d'Agrachev et Gamkrelidze [9] "The Pontryagin Maximum Principle 50 years later". Comme ils le soulignent :

"The maximum condition could be considered here as a generalized elimination algorithm, "dynamically" excluding the parameter u , as it evolves, at almost every instant of time t and thus obtaining the trajectory $p_*(t)$." (p.9)

Ils insistent sur le fait que le PMP est à la fois très général et très concret :

"(The) formulation of the maximum principle, though very simple and universally general, without any exceptions, still works at full swing in amazingly many nontrivial cases, often generating whole families of minimizers (the so called optimal synthesis), and thus yielding complete solutions to otherwise unamenable problems. Such an efficiency could serve as a partial explanation of the fact that, despite its simplicity and generality, the maximum principle is quite hard to prove." (p.2)

Ils explicitent ensuite la signification profonde du PMP comme reformulation d'un problème d'optimisation dans le cadre de la géométrie symplectique de champs hamiltoniens sur des fibrés cotangents :

"As the second remarkable feature of the maximum principle we consider the symplectic invariant form in which it was actually formulated from the very beginning, thus clarifying the real intrinsic meaning of the extremality conditions. The maximum principle canonically assigns to an optimal problem on the configuration manifold a Hamiltonian vector field on the cotangent bundle of the manifold, reducing the solution of the problem to finding trajectories of the Hamiltonian field that satisfy additionally a certain maximum condition. Such procedures that reduce the solution of the initial problem on a

manifold to the study of some canonically constructed infinitesimal objects on its cotangent bundle are often encountered in modern analysis under the name of canonical microlocalization. They readily supply us with invariants of the initial problem, as well as of the manifold itself on which the problem is considered. Historically, the maximum principle was one of the first procedures where this transition was explicitly described.” (p. 2)

5.2. Un exemple élémentaire

Un exemple élémentaire que l'on trouve un peu partout dans les introductions au contrôle optimal est le suivant. Il consiste à prendre un mobile sur \mathbb{R} , à choisir l'accélération $\ddot{x}(t) = u(t)$ comme contrôle et à imposer la contrainte non holonome $u \in [-1, 1] = U$.⁶ Si l'on pose $y = \dot{x}$ et $q = (x, y) \in M = \mathbb{R}^2$, l'équation (E) devient alors

$$\dot{q} = (\dot{x}, \dot{y}) = (y, \ddot{x}) = (y, u) = X(q, u) .$$

Il s'agit donc d'un système affine particulièrement simple $\dot{q} = F(q) + uG(q)$ avec $F(q) = (y, 0)$ et $G(q) = (0, 1)$. Le problème est alors de relier un point (x_0, y_0) , par exemple l'origine $(0, 0)$, à un point (x_1, y_1) en un temps *minimal*. On considère donc le Lagrangien $L(q, u) = 1$.

Pour des covecteurs $p = (\lambda, \mu)$, l'hamiltonien s'écrit

$$\mathcal{H}(q, p, u) = \langle p, X(q, u) \rangle - L = \lambda y + \mu u - 1 .$$

Les équations de Hamilton donnent par conséquent

$$\begin{cases} \dot{x} = \frac{\partial \mathcal{H}}{\partial \lambda} = y, & \dot{y} = \frac{\partial \mathcal{H}}{\partial \mu} = u \\ \dot{\lambda} = -\frac{\partial \mathcal{H}}{\partial x} = 0, & \dot{\mu} = -\frac{\partial \mathcal{H}}{\partial y} = -\lambda \end{cases}$$

On voit que λ est constant, $\lambda = \lambda_0$, et que μ est linéaire en t , $\mu = -\lambda_0 t + \mu_0$. On constate immédiatement que, pour maximiser \mathcal{H} il faut prendre le maximum de μu et donc $u_* = \pm 1$ suivant le *signe* de μ . Le contrôle u a donc toujours une valeur extrême et il saute d'un extrême à l'autre (il subit un “switch”) lorsque μ change de signe. Autrement dit,

$$H(q_*, p_*) = \sup_{u \in U} \mathcal{H}(q, p, u) = \lambda y + |\mu| - 1 .$$

La droite S de switching

$$\mu = -\lambda_0 t + \mu_0 = 0$$

joue ainsi un rôle crucial.

On voit alors comment s'effectue la synthèse optimale. Si $\lambda_0 = 0$, $\dot{\mu} = 0$ et $\mu = \mu_0$ est constant. Si $\mu_0 = 0$, $H = -1$ et le cas est dégénéré. Si $\mu_0 \neq 0$ (toujours avec $\lambda_0 = 0$) alors la droite S n'est jamais atteinte. Les cas intéressants sont ceux

6. C'est le cas d'une voiture. On contrôle l'accélération.

pour lesquels $\lambda_0 \neq 0$. La droite S est alors atteinte pour $t = \frac{\mu_0}{\lambda_0}$ et pour avoir $t > 0$ il faut que λ_0 et μ_0 soient de même signe. Les extrémales satisfont $\dot{x} = y$ et $\dot{y} = u_* = \pm 1$ suivant le signe de μ . On voit donc clairement apparaître une réduction dimensionnelle : les trajectoires optimales sont des concaténations d'arcs de trajectoires paraboliques appartenant aux deux familles $\ddot{x} = \pm 1$

$$\begin{cases} x(t) = x_0 + y_0 t + \frac{1}{2} t^2 & \text{si } \mu > 0, \dot{y} = \ddot{x} = 1 \\ x(t) = x_0 + y_0 t - \frac{1}{2} t^2 & \text{si } \mu < 0, \dot{y} = \ddot{x} = -1 \end{cases}$$

Supposons que l'on démarre de l'origine $(x_0, y_0) = (0, 0)$ pour $t_0 = 0$ (i.e. on part de 0 avec la vitesse nulle, cf. figure 9). Soient P^+ la parabole $x = \frac{1}{2}y^2$, $P^{+,+}$ la demi-parabole P^+ avec $y \geq 0$ (i.e. $x = \frac{1}{2}t^2$, $y = \dot{x} = t \geq 0$), P^- la parabole $x = -\frac{1}{2}y^2$ et $P^{-,-}$ la demi-parabole P^- avec $y \leq 0$ (i.e. $x = -\frac{1}{2}t^2$, $y = \dot{x} = -t \leq 0$). Pour rejoindre un point (x_1, y_1) (i.e. la position x_1 avec la vitesse y_1) qui n'est pas sur $P^{+,+}$ ou $P^{-,-}$, le mobile doit d'abord suivre soit $P^{+,+}$ soit $P^{-,-}$ pendant un certain temps s et, à partir du point (x_s, y_s) , switcher vers la parabole de l'autre famille qui passe par (x_1, y_1) . Supposons par exemple que le mobile démarre par un arc sur $P^{+,+}$ et donc que $\mu_0 > 0$ (et aussi $\lambda_0 > 0$). Après le switch qui se produit au temps $s = \frac{\mu_0}{\lambda_0} > 0$ et au point $\left(\frac{1}{2}\left(\frac{\mu_0}{\lambda_0}\right)^2, \frac{\mu_0}{\lambda_0}\right)$, le mobile continue sur l'arc de parabole

$$\begin{aligned} x(t) &= x_s + y_s(t-s) - \frac{1}{2}(t-s)^2 \\ &= \frac{1}{2}\left(\frac{\mu_0}{\lambda_0}\right)^2 + \frac{\mu_0}{\lambda_0}\left(t - \frac{\mu_0}{\lambda_0}\right) - \frac{1}{2}\left(t - \frac{\mu_0}{\lambda_0}\right)^2 \end{aligned}$$

qui s'écrit aussi $x - x_s = -\frac{1}{2}(y^2 - y_s^2)$. Comme maintenant $\dot{y} = -1$, y décroît à partir de y_s et la trajectoire continue au-dessous de $P^{+,+}$. Et comme la trajectoire doit passer par (x_1, y_1) et que, par construction, $x_s = \frac{1}{2}y_s^2$, on doit donc pouvoir avoir $x_1 - x_s = -\frac{1}{2}(y_1^2 - y_s^2)$, c'est-à-dire $x_1 + \frac{1}{2}y_1^2 = 2x_s$. Mais sur $P^{+,+}$ on a $2x_s > 0$. Si la condition $x_1 + \frac{1}{2}y_1^2 > 0$ n'est pas vérifiée, autrement dit si (x_1, y_1) est à l'intérieur de la parabole P^- ou à l'extérieur de P^- mais au-dessus de $P^{+,+}$, alors le mobile doit partir de l'origine sur la demi-parabole $P^{-,-}$. Bref, $\Sigma = P^{+,+} \cup P^{-,-}$ est une séparatrice. Si (x_1, y_1) est au-dessous de Σ le mobile part sur l'arc $P^{+,+}$ et si au contraire (x_1, y_1) est au-dessus de Σ le mobile part sur l'arc $P^{-,-}$. Pour (x_1, y_1) au-dessous de Σ avec $y_1 \geq 0$, le mobile accélère uniformément jusqu'à x_s puis décélère uniformément jusqu'à x_1 pour y arriver à la vitesse y_1 . Si en revanche $y_1 \leq 0$, le mobile accélère uniformément jusqu'à x_s puis décélère uniformément jusqu'au rebroussement $x_t = x_s + \frac{1}{2}y_s^2 = 2x_s$ où la vitesse s'annule et rebrousse chemin pour arriver à x_1 avec la vitesse $y_1 \leq 0$. La figure 9 montre deux exemples de telles trajectoires.

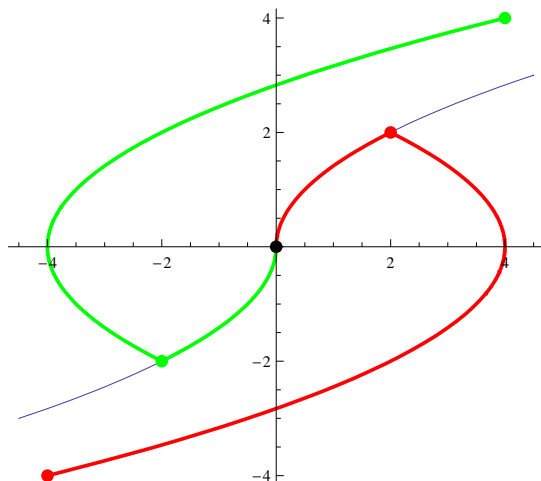


FIGURE 9. Deux trajectoires de la synthèse optimale du système $\dot{q} = (\dot{x}, \dot{y}) = (y, \ddot{x}) = (y, u)$ avec la contrainte non holonome $u \in [-1, 1]$. La trajectoire rouge part de l'origine $(0, 0)$ (point noir) et accélère avec l'accélération constante $+1$ jusqu'au point de switch $(x_s, y_s) = (2, 2)$ (premier point rouge). À partir de là elle décélère avec l'accélération constante -1 jusqu'au point terminal $(x_1, y_1) = (-4, -4)$ (deuxième point rouge). La trajectoire verte est la symétrique.

On voit donc qu'en définitive on repère chaque point (x_1, y_1) au moyen des deux familles de paraboles et qu'avec ce système de coordonnées paraboliques on rejoint (x_1, y_1) à partir de $(0, 0)$ en deux temps, exactement comme en coordonnées cartésiennes on va d'abord de $(0, 0)$ à $(x_1, 0)$ puis de $(x_1, 0)$ à (x_1, y_1) .

5.3. Quelques idées de la preuve (d'après A. Lewis)

Nous nous bornons ici à quelques remarques intuitives et informelles sur la preuve en suivant la présentation pédagogique [329] d'Andrew Lewis; elle est limpide et possède l'avantage d'être détaillée et purement calculatoire, n'utilisant que des coordonnées cartésiennes. Le lecteur remarquera qu'elle est très "eulérienne" au sens du *Methodus* que nous avons commenté à la section 8 du chapitre 9.

5.3.1. Variations et cônes tangents.

Pour démontrer le PMP, il faut d'abord, comme pour démontrer les équations d'E-L, introduire des *variations* des trajectoires. Comme ces dernières sont de type $(q(t), u(t))$, il est naturel de regarder d'abord des variations des $q(t)$ puis ensuite des variations des $u(t)$. En ce qui concerne la première possibilité, on peut, étant donné un contrôle $u(t)$, modifier une trajectoire admissible $q(t) = \gamma(q_0, t_0, t, u(t))$ au moyen d'une déformation $q^s(t)$ de $q(t) = q^0(t)$ dépendant d'un petit paramètre s variant dans un voisinage de 0 et de classe C^1 en s pour tout $t \in [t_0, t_1]$. Cela

définit un champ de vecteurs tangents

$$v(t) = \left. \frac{dq^s(t)}{ds} \right|_{s=0}$$

le long de $q(t)$. Comme les $q^s(t)$ doivent satisfaire l'équation de base (E) $\dot{q}^s(t) = X(q(t), u(t))$, les $v(t)$ doivent satisfaire l'équation

$$\dot{v}(t) = \frac{\partial X(q(t), u(t))}{\partial q} \cdot v(t) .$$

Et, de même que (E) implique que $q^s(t)$ s'obtient à partir de la condition initiale q_0^s , $v(t)$ s'obtient en propageant $v_0 = v(t_0)$ le long de γ au moyen de (E). Cette propagation s'écrit $v(t) = \Phi(t) \cdot v_0$ où $\Phi(t)$ est la matrice

$$\Phi(t) = \frac{\partial \gamma(q_0, t_0, t, u(t))}{\partial q_0}$$

avec $\Phi(t_0) = Id$ (l'identité). En effet, on en tire $\dot{v}(t) = \dot{\Phi}(t) \cdot v_0$. Mais, en écrivant les choses de façon symbolique, on a

$$\begin{aligned} \dot{\Phi}(t) &= \frac{\partial}{\partial t} \left(\frac{\partial \gamma(q_0, t_0, t, u(t))}{\partial q_0} \right) = \frac{\partial}{\partial q_0} \left(\frac{\partial \gamma(q_0, t_0, t, u(t))}{\partial t} \right) \\ &= \frac{\partial}{\partial q_0} (X(q(t), u(t))) = \frac{\partial}{\partial q} (X(q(t), u(t))) \circ \frac{\partial q(t)}{\partial q_0} \\ &= \frac{\partial}{\partial q} (X(q(t), u(t))) \circ \Phi(t) . \end{aligned}$$

On vérifie par conséquent que l'on a bien

$$\begin{aligned} \dot{v}(t) &= \dot{\Phi}(t) \cdot v_0 = \left(\frac{\partial}{\partial q} (X(q(t), u(t))) \circ \Phi(t) \right) \cdot v_0 \\ &= \frac{\partial}{\partial q} (X(q(t), u(t))) \cdot (\Phi(t) \cdot v_0) = \frac{\partial}{\partial q} (X(q(t), u(t))) \cdot v(t) . \end{aligned}$$

Remarque. Les $p(t)$ satisfont l'équation adjointe $\dot{p}(t) = - \left(\frac{\partial X(q(t), u(t))}{\partial q} \right)^T \cdot p(t)$.

□

Mais on peut aussi modifier les contrôles $u(t)$. Une tactique devenue classique, tactique que nous esquissons ici pour le cas où l'intervalle $[t_0, t_1]$ est *fixé*, est de déformer $u(t)$ en $u^s(t)$ au moyen d'une fonction de Dirac. On parle joliment de "needle variations" et nous utiliserons l'expression de N-variations. Au voisinage d'un instant $\tau \in]t_0, t_1]$, plus précisément sur un petit intervalle $[\tau - sl, \tau] \subset [t_0, t_1]$ à gauche de τ , on modifie $u(t)$ en lui donnant une valeur donnée μ . On peut alors

définir une variation v_τ de la trajectoire au moyen de la formule

$$v_\tau = \left. \frac{d}{ds} \right|_{s=0} \gamma(q_0, t_0, \tau, u^s(t)) .$$

Pour que cette formule ait un sens, il faut que $u(t)$ ne soit pas trop pathologique en τ (rappelons que les fonctions de contrôle peuvent être assez sauvages). La condition technique est intuitivement que $u(\tau)$ soit la moyenne de $u(t)$ sur les voisinages infinitésimaux de τ ; on dit alors que τ est un “point de Lebesgue”. Les N-variations à la Dirac forment trivialement des cônes car si $\lambda \in \mathbb{R}^+$, l’intervalle de longueur $s\lambda$ de même τ donne pour le même μ un $v_{\tau,\lambda} = \lambda v_\tau$. On peut par ailleurs considérer plusieurs N-variations en différents points τ_i et des “combinaisons convexes” $\sum_i \lambda_i v_{\tau_i}$

avec $\sum_i \lambda_i = 1$.

Si maintenant on considère un instant $t > \tau$, on peut, comme on l’a fait plus haut pour les variations à $u(t)$ donné, propager v_τ jusqu’en t au moyen des $\Phi(t)$ appropriées (démarrant en τ) en utilisant le contrôle initial $u(t)$ qui n’a pas été modifié pour $t > \tau$. On obtient ainsi des directions tangentes $v_\tau(t)$ de déformations de trajectoires en $t > \tau$. Nous avons vu qu’elles forment des cônes, mais ceux-ci peuvent être très compliqués au sens où une section transverse peut être un ensemble compliqué et pas du tout un bon domaine fermé convexe. Notons alors

$$K_t = K(q_0, t_0, t, u(t)) \subset T_{q(t)}M$$

la *fermeture de l’enveloppe convexe* des $v_\tau(t)$. K_t s’appelle le “cône tangent” en t . Il est la fermeture de l’union des $\Phi(t) \cdot K_\tau$ pour $\tau < t$. Ces cônes tangents sont intimement liés aux ensembles “atteignables” (“reachable sets”) du système considéré. Si q_0, t_0, t_1 sont donnés, l’ensemble atteignable $\mathcal{R}(q_0, t_0, t_1)$ est l’ensemble des cibles $q \in M$ que l’on peut atteindre à partir de q_0 en un temps $[t_0, t_1]$ au moyen de trajectoires admissibles $u(t) : [t_0, t_1] \rightarrow U$ avec $q(t)$ satisfaisant (E).

$$\mathcal{R}(q_0, t_0) = \bigcup_{t_1 \geq t_0} \mathcal{R}(q_0, t_0, t_1)$$

est l’ensemble atteignable à partir de q_0 .

Le problème de la contrôlabilité est celui de l’extension des \mathcal{R} . Il est en général extrêmement difficile. La contrôlabilité *globale* signifie que $\mathcal{R}(q_0, t_0) = M$. La contrôlabilité *locale* en q_0 signifie que pour tout $t > t_0$ $\mathcal{R}(q_0, t_0, t)$ est un voisinage de q_0 , autrement dit qu’à partir de q_0 il existe des trajectoires admissibles allant dans toutes les directions.

5.3.2. Cônes tangents et ensembles atteignables.

Un lemme fondamental reliant les cônes tangents aux ensembles atteignables dit intuitivement (toujours dans le cas d’intervalles $[t_0, t_1]$ fixés) que les points intérieurs

à K sont dans \mathcal{R} au sens suivant : si $q(\bullet)$ est une trajectoire admissible sur $[t_0, t]$ à partir de q_0 , autrement dit si $q(t) \in \mathcal{R}(q_0, t_0, t)$, et si $v_0 \in \overset{\circ}{K}_t$ (l'intérieur de K_t) alors il existe un sous-cône K^\triangleleft de $\overset{\circ}{K}_t$ qui est un voisinage de v_0 (i.e. $v_0 \in \overset{\circ}{K}^\triangleleft$) tel que

$$q(t) + v \in \mathcal{R}(q_0, t_0, t)$$

pour tous les $v \in K^\triangleleft$ de norme assez petite. Ce lemme est loin d'être trivial à démontrer. Comme nous l'avons noté, K_t est la fermeture convexe des $v_\tau(t)$ et on ne sait pas comment les $v_\tau(t)$ y sont effectivement distribués.

Plusieurs intuitions convergent dans ce lemme. D'abord si l'on est arrivé en $q(t) \in \mathcal{R}(q_0, t_0, t)$ et si v_τ est une N-variation pour $\tau < t$, alors cette N-variation pousse $q(t)$ le long du vecteur tangent $v_\tau(t)$. Il s'agit de savoir si l'on peut passer de l'infinitésimal au local, autrement dit si, pour un segment $\lambda v_\tau = v$ de norme assez petite, il existe effectivement une trajectoire admissible $\tilde{q}(\bullet)$ qui arrivera au temps t au point $\tilde{q}(t) = q(t) + v$. Le lemme dit que si v_0 est intérieur à K_t , même si v_0 n'est pas un v_τ , alors cela est vrai pour tous les v d'un voisinage de v_0 dans un sous-cône K^\triangleleft de K_t .

Esquisons maintenant les liens fondamentaux existant entre les cônes tangents et les hamiltoniens. Commençons par les cas sans lagrangiens et notons

$$\mathcal{H}_0(q, p, u) = \langle p, X(q, u) \rangle .$$

Il est trivial que u_* réalise le maximum $H_0(q, p)$ de $\mathcal{H}_0(q, p, u)$ si et seulement si $\langle p, X \rangle \leq 0$ pour tout X de la forme

$$X = X(q, u) - X(q, u_*), u \in U .$$

Cela signifie que si

$$\mathfrak{X}(q) = \{X(q, u) | u \in U\}$$

alors le covecteur p s'annule sur un hyperplan d'appui \mathcal{P} de $\mathfrak{X}(q)$ au sens où $\mathfrak{X}(q)$ est entièrement contenu dans un des demi-espaces séparés par l'hyperplan parallèle à $\ker(p)$ sur lequel p prend la valeur $\langle p, X(q, u_*) \rangle$. On montre alors facilement un lemme disant que si l'on a une réponse adjointe $p(t)$ à $q(t) = \gamma(q_0, t_0, t, u(t))$ telle que pour un $\tau \in [t_0, t_1]$ on ait $\langle p(\tau), v \rangle \leq 0$ pour tout $v \in K_\tau$, alors, pour tout point de Lebesgue $t \leq \tau$, $p(t)$ est une solution extrémale, i.e.

$$H_0(q(t), p(t)) = \sup_{u \in U} \mathcal{H}_0(q(t), p(t), u(t)) .$$

Ce lemme dit donc que le fait qu'une réponse adjointe $p(\tau)$ définisse en τ un plan d'appui du cône tangent K_τ implique que $(q(t), p(t))$ est une extrémale avant τ . L'idée est que l'existence d'un tel plan d'appui signifie que le contrôle est "extrémal" à l'instant τ dans la mesure où les N-variations en τ ne peuvent pointer que dans un demi-espace et que cette extrémalité en τ n'est possible que si le contrôle est tout le temps extrémal avant τ . La preuve est assez simple. On utilise l'évidence

que si $\ker(p)$ est un plan d'appui pour K , i.e. si $\langle p, v \rangle \leq 0$ pour tout $v \in K$, et si $v = \tilde{X} - X$ alors

$$\langle p, \tilde{X} \rangle \leq \langle p, X \rangle .$$

Soit $t \leq \tau$ un point de Lebesgue. On introduit une N-variation v_t en t venant d'un contrôle $\tilde{u}(t)$ modifiant $u(t)$. On propage v_t en τ au moyen du Φ approprié : $v_\tau = \Phi.v_t$. Mais en τ , $\Phi.(X(q, \tilde{u}) - X(q, u)) \in K_\tau$ et donc

$$\langle p(\tau), \Phi.(X(q, \tilde{u})) \rangle - \langle p(\tau), \Phi.(X(q, u)) \rangle \leq 0 ,$$

soit, par dualité,

$$\langle \Phi^T.p(\tau), X(q, \tilde{u}) \rangle - \langle \Phi^T.p(\tau), X(q, u) \rangle \leq 0 .$$

Mais comme $\Phi^T.p(\tau) = p(t)$ on obtient

$$\mathcal{H}_0(q(t), p(t), \tilde{u}(t)) \leq \mathcal{H}_0(q(t), p(t), u(t))$$

pour tout $\tilde{u}(t)$, d'où l'extrémalité.

5.3.3. Un premier théorème.

Un premier théorème dans le cas $L = 0$ et $[t_0, t_1]$ fixé dit que si $u_*(t)$ est un contrôle admissible contrôlant une trajectoire $q_*(t) = \gamma(q_0, t_0, t, u_*(t))$ permettant d'atteindre le bord $\partial\mathcal{R}$ de $\mathcal{R}(q_0, t_0, t_1)$ (i.e. $q_*(t_1) \in \partial\mathcal{R}$), alors il existe une réponse adjointe $p_*(t)$ le long de $q_*(t)$ qui maximise $\mathcal{H}_0(q_*(t), p_*(t), u(t))$ pour presque tout t , i.e.

$$\mathcal{H}_0(q_*(t), p_*(t), u_*(t)) = H_0(q_*(t), p_*(t)) .$$

Qui plus est, si $u_*(t)$ est borné, alors $H_0(q_*(t), p_*(t))$ est constant le long de la trajectoire. Bref, si une trajectoire contrôlée admissible sur $[t_0, t_1]$ permet d'atteindre le bord $\partial\mathcal{R}(q_0, t_0, t_1)$ elle est nécessairement extrême.

Pour le voir, on considère une suite $\{q_j\}_{j \in \mathbb{N}}$ de points q_j extérieurs à \mathcal{R} (i.e. $q_j \notin \overline{\mathcal{R}}$) qui tendent vers $q_*(t_1)$. En normant les $q_j - q_*(t_1)$, on obtient une suite $\{v_j\}$ de vecteurs unitaires dans \mathbb{S}^{n-1} et comme la sphère \mathbb{S}^{n-1} est compacte, il existe une sous-suite convergente et, quitte à prendre une sous-suite des q_j , on peut supposer que c'est la suite $\{v_j\}$ elle-même. La suite converge vers un vecteur unitaire v_0 qui pointe vers l'extérieur du fermé $\overline{\mathcal{R}}$. D'après le lemme fondamental de la sous-section précédente disant que les points intérieurs à K sont dans \mathcal{R} , on ne peut pas avoir $v_0 \in \overset{\circ}{K}_{t_1}$ et donc soit $v_0 \in \partial K_{t_1}$, soit $v_0 \notin K_{t_1}$ (qui, rappelons-le, est par définition un fermé convexe). Cela implique qu'il existe un hyperplan \mathcal{P}_{t_1} séparant K_{t_1} de v_0 . Il s'agit là d'un résultat classique d'analyse convexe. L'intuition est que, comme $q_*(t_1) \in \partial\mathcal{R}$, toute variation de la trajectoire doit garder la trajectoire dans $\overline{\mathcal{R}}$ et donc tout $w \in K_{t_1}$ doit pointer vers l'intérieur. Il existe donc nécessairement un hyperplan séparant les $w \in K_{t_1}$ qui pointent vers l'intérieur de v_0 qui pointe, lui, vers l'extérieur de \mathcal{R} .

Un peu plus précisément, si $v_0 \notin K_{t_1}$ et $w \in K_{t_1}$, soit $r = \|v_0 - w\|$. L'ensemble $C = K_{t_1} \cap B(v_0, r)$ (où $B(v_0, r)$ est la boule fermée de centre v_0 et de rayon r) est l'intersection d'un fermé et d'un compact et est donc un compact. Il s'ensuit que la distance $\|v_0 - w\|$ atteint son minimum w_0 sur C . Posons

$$\begin{aligned}\lambda &= w_0 - v_0 \\ a &= \langle w_0, \lambda \rangle = \|w_0\|^2 - \langle w_0, v_0 \rangle \quad (\text{produit scalaire}).\end{aligned}$$

Comme $v_0 \notin K_{t_1}$ et $w_0 \in K_{t_1}$, on a $\lambda \neq 0$. On montre alors que l'hyperplan $\mathcal{P}_{\lambda, a} = \{v \mid \langle \lambda, v \rangle = a\}$ sépare v_0 et K_{t_1} . D'abord, il est trivial que $\langle \lambda, v_0 \rangle < a$ car, étant donnée la symétrie du produit scalaire, $\langle w_0 - v_0, v_0 \rangle < \langle w_0, w_0 - v_0 \rangle$ n'est qu'une autre façon d'écrire que $\|w_0 - v_0\|^2 > 0$. Ensuite, on montre (c'est plus compliqué) que $\langle \lambda, w \rangle \geq a$ pour tout $w \in K_{t_1}$. On passe alors des λ aux covecteurs p duaux des λ en remplaçant les produits scalaires $\langle \lambda, v \rangle$ par les produits $\langle p, v \rangle$.

Si maintenant $v_0 \in \partial K_{t_1}$ on utilise (comme plus haut pour $q_*(t_1) \in \partial \mathcal{R}$) une suite $\{v_j\}$ de v_j extérieurs à K_{t_1} et convergeant vers v_0 , ainsi que des suites $\{\lambda_j \neq 0\}$ et $\{a_j\}$ telles que $\langle \lambda_j, v_j \rangle < a_j$ et $\langle \lambda_j, w \rangle \geq a$ pour tout $w \in K_{t_1}$. Ces suites existent puisque $v_j \notin K_{t_1}$. On normalise les λ_j en $\alpha_j = \frac{\lambda_j}{\|\lambda_j\|}$ et l'on obtient une suite dans \mathbb{S}^{n-1} compacte qui, quitte à se restreindre à une sous-suite, converge vers un vecteur unitaire α . Soit $\langle \alpha_j, v_j \rangle = c_j$. On a par construction $\langle \alpha_j, K_{t_1} \rangle > c_j$. Soit $c = \langle \alpha, v_0 \rangle$ la limite des c_j (elle existe car $v_j \rightarrow v_0$ et $\alpha_j \rightarrow \alpha$). Pour tout $w \in K_{t_1}$, on a

$$\langle \alpha, w \rangle = \lim_{j \rightarrow \infty} \langle \alpha_j, w \rangle \geq c.$$

D'où la séparation entre v_0 et K_{t_1} .

On peut donc considérer un hyperplan \mathcal{P}_{t_1} séparant K_{t_1} de v_0 et, par conséquent, un p_1 de noyau $\ker(p_1) = \mathcal{P}_{t_1}$ tel que $\langle p_1, v \rangle \leq 0$ pour tout $v \in K_{t_1}$ et $\langle p_1, v_0 \rangle > 0$. Si $p_*(t)$ est la réponse adjointe le long de la trajectoire (q_*, u_*) telle que $p_*(t_1) = p_1$, on peut alors appliquer à $\tau = t_1$ le lemme ci-dessus sur le fait que $\langle p(\tau), v \rangle \leq 0$ pour tout $v \in K_\tau$ est une condition d'extrémalité de t_0 à τ . On montre de plus (la preuve est compliquée) que si $u_*(t)$ est borné, alors $H_0(q_*(t), p_*(t))$ est constant le long de la trajectoire.

Le théorème précédent concerne les trajectoires atteignant $\partial \mathcal{R}(q_0, t_0, t_1)$. Un résultat complémentaire dit que les trajectoires intérieures à \mathcal{R} jusqu'en τ restent à l'intérieur de \mathcal{R} au sens suivant. Si $\tau \in [t_0, t_1]$ et si

$$q(\tau) = \gamma(q_0, t_0, \tau, u) \in \overset{\circ}{\mathcal{R}}(q_0, t_0, \tau),$$

alors $q(t) \in \overset{\circ}{\mathcal{R}}(q_0, t_0, t)$ pour tout $t > \tau$. Cela implique que si une trajectoire de t_0 à t_1 atteint $\partial \mathcal{R}(q_0, t_0, t_1)$ alors, pour tout $t < t_1$, sa restriction de t_0 à t atteint $\overset{\circ}{\mathcal{R}}(q_0, t_0, t)$.

5.3.4. *Le théorème général.*

On peut alors passer de ces résultats préliminaires sur le cas où $L = 0$ au cas où le problème fait intervenir un lagrangien $L(q, u)$. On a donc l'équation (E) $\dot{q} = X(q, u)$ et un "coût" $J(t) = \int_{t_0}^t L(q(t), u(t)) dt$ à minimiser le long des trajectoires de (E), ce qui introduit l'équation différentielle

$$\dot{J}(t) = L(q(t), u(t)) .$$

On peut donc en fait considérer J comme une *coordonnée supplémentaire*. Le nouvel espace de configurations est alors $\widehat{M} = M \times \mathbb{R}$ et l'on décore d'un \bullet les entités \widehat{q} , $\widehat{\mathcal{R}}$, \widehat{K} , etc. intervenant dans ce problème élargi.

Comme l'explique Andrew Lewis [329], un lemme fondamental dit que les trajectoires optimales sont nécessairement *sur le bord* de $\widehat{\mathcal{R}}$ au sens où, si $(q_*(t), u_*(t))$ est optimale, alors

$$\widehat{q}_*(t_1) \in \partial \widehat{\mathcal{R}}(\widehat{q}_*(t_0), t_0, t_1) .$$

En effet, comme $(q_*(t), u_*(t))$ est optimale, $J_*(t_1)$ minimise le coût J le long de toutes les trajectoires $(q(t), u(t))$ joignant $q_*(t_0)$ à $q_*(t_1)$ en temps $[t_0, t_1]$, ce qui s'écrit

$$J_*(t_1) = \inf \left\{ J \in \mathbb{R} \mid (q_*(t_1), J) \in \widehat{\mathcal{R}}(\widehat{q}_*(t_0), t_0, t_1) \right\} .$$

Soit \mathcal{V} un voisinage de $\widehat{q}_*(t_1)$ dans \widehat{M} . Il contient nécessairement des points $(q_*(t_1), J)$ avec $J < J_*(t_1)$ et, à cause de la propriété de minimalité de $J_*(t_1)$, ces points ne peuvent pas appartenir à $\widehat{\mathcal{R}}$. Par conséquent, puisqu'aucun voisinage \mathcal{V} ne peut être dans $\widehat{\mathcal{R}}$, on a $\widehat{q}_*(t_1) \in \partial \widehat{\mathcal{R}}$.

Ce lemme fondamental permet de démontrer la première partie du PMP, à savoir l'existence de réponses adjointes p_* et α_* .

Théorème (première partie du PMP). L'intervalle temporel $[t_0, t_1]$ est supposé fixé. Si $(q_*(t), u_*(t))$ est une trajectoire optimale, il existe une trajectoire (non triviale) $p_*(t)$ absolument continue et une constante $\alpha_* = -1$ ou 0 telles que :

1. soit $\alpha_* = -1$, soit $\alpha_* = 0$ mais $p_*(t_0) \neq 0$;
 2. $p_*(t)$ est une réponse adjointe pour le lagrangien $\alpha_* L$ le long de $(q_*(t), u_*(t))$;
- i.e. $q_*(t)$ et $p_*(t)$ satisfont les équations de Hamilton de l'hamiltonien

$$\mathcal{H}(q, p, u) = \langle p, X(q, u) \rangle + \alpha L(q, u) ;$$

3. $(q_*(t), p_*(t), u_*(t))$ maximise l'hamiltonien pour presque tout $t \in [t_0, t_1]$, i.e.

$$\mathcal{H}(q_*, p_*, u_*) = \sup_{u \in U} (\mathcal{H}(q_*, p_*, u)) = H(q_*, p_*)$$

est vérifié. ◇

La preuve est un simple raffinement de celle montrant que, dans le cas $L = 0$, si une trajectoire admissible $(q_*(t), u_*(t))$ permet d'atteindre $\partial \widehat{\mathcal{R}}(q_0, t_0, t_1)$, alors il existe une réponse adjointe $p_*(t)$ le long de $q_*(t)$ qui maximise $\mathcal{H}_0(q_*(t), p_*(t), u(t))$.

On remarque d'abord que le vecteur $(0, -1) \in \mathbb{R}^n \oplus \mathbb{R}$ n'appartient pas à l'intérieur de $\widehat{K}_{t_1} = \widehat{K}(\widehat{q}_0, t_0, t_1, u_*)$. Si c'était en effet le cas, d'après le lemme disant que les extrémités de trajectoires + des vecteurs v de norme assez petite intérieurs au cône tangent K restent dans l'ensemble atteignable \mathcal{R} , il y aurait des points de $\widehat{\mathcal{R}}$ de coût $< J_*(t_1)$, ce qui est impossible puisque $(q_*(t), u_*(t))$ est optimale. Donc soit $(0, -1) \in \partial \widehat{K}_{t_1}$, soit $(0, -1) \notin \widehat{K}_{t_1}$. Par conséquent (les arguments sont les mêmes que précédemment), il existe un hyperplan $\widehat{\mathcal{P}}_{t_1}$ séparant $(0, -1)$ du convexe fermé \widehat{K}_{t_1} . On prend alors un $\widehat{p}_{*,1} = (p_{*,1}, \alpha_*)$ de noyau $\widehat{\mathcal{P}}_{t_1}$ satisfaisant $\langle \widehat{p}_{*,1}, (0, -1) \rangle \geq 0$ (et donc $-\alpha_* \geq 0$) et $\langle \widehat{p}_{*,1}, \widehat{v} \rangle < 0$ pour tout $\widehat{v} \in \widehat{K}_{t_1}$. Soit alors $\widehat{p}_*(t) = (p_*(t), \alpha_*(t))$ la réponse adjointe avec $\widehat{p}_*(t_1) = \widehat{p}_{*,1}$. Comme le champ \widehat{X} étendant le $X(q, u)$ de l'équation de base (E) ne dépend pas de J , on a $\dot{\alpha}_*(t) = 0$ et donc $\alpha_*(t) = \alpha_* \leq 0$. Si $\alpha_* < 0$, on peut prendre $\alpha_* = -1$ et normaliser \widehat{p}_* en le remplaçant par $-\frac{\widehat{p}_*}{\alpha_*}$. On remarque alors que la linéarité de l'équation adjointe implique $\widehat{p}_*(t) \neq 0$ pour tout $t \in [t_0, t_1]$ et que donc si $\alpha_* = 0$ on doit avoir $p_*(t) \neq 0$ et donc en particulier $p_*(t_0) \neq 0$.

Nous renvoyons à Lewis [329] pour les autres aspects du PMP :

- (i) le fait que si $u_*(t)$ est borné, alors $H(q_*, p_*)$ est constant le long de la trajectoire optimale et que si, de plus, le problème suppose $[t_0, t_1]$ libre, alors $H(q_*, p_*)$ est non seulement constant mais $= 0$;
- (ii) la condition de transversalité disant que si $(q_*(t), u_*(t))$ est optimale seulement pour joindre une sous-variété source S_0 de M à une sous-variété but S_1 alors on peut choisir $p_*(t)$ de façon à ce que $p_*(t_0)$ s'annule sur le sous-espace tangent $T_{q_0}S_0$ de $T_{q_0}M$ et que $p_*(t_1)$ s'annule sur le sous-espace tangent $T_{q_1}S_1$.

6. Géodésiques, caustiques et solutions lagrangiennes

Un problème variationnel central pour notre propos est celui des géodésiques, c'est-à-dire celui des courbes de plus courte distance dans une variété M munie d'une métrique, qu'elle soit riemannienne ou sous-riemannienne. Pour les variétés riemanniennes, ce problème est très ancien. Nous l'avons déjà brièvement évoqué à la section 6.2 du chapitre 6. Mais, même dans les cas classiques les plus simples, il s'est heurté à des difficultés de calcul redoutables tout à fait analogues à celles que nous avons commentées dans le chapitre 9 consacré à l'histoire du calcul des variations. On a là encore un bon exemple de problème physique élémentaire dont le traitement analytique est compliqué.

Rappelons que le problème des géodésiques est double (cf. encore la section 6.2 du chapitre 6).

1. D'abord, il s'agit de trouver leurs équations à partir de la donnée de la métrique. Dans l'espace euclidien, les géodésiques sont des droites (les rayons

lumineux de l'optique géométrique standard). Mais dès que la métrique n'est plus euclidienne les équations deviennent non triviales.

2. Le second problème est que, lorsque la géométrie de M fait que les géodésiques issues d'un point a doivent, après une divergence initiale, reconverger, les géodésiques peuvent avoir une *enveloppe*⁷ que l'on appelle le *lieu conjugué* de a . Un point conjugué de a sur une géodésique γ issue de a est un point où γ devient tangente à une géodésique issue de a qui lui est infiniment voisine (si les deux géodésiques se croisent ce n'est donc pas transversalement au sens habituel). Qui plus est, il peut exister des points où plusieurs géodésiques ont la même longueur et sont en compétition. Ces points constituent le *cut locus* de a et, avec les points conjugués, constituent la *caustique* de a . Ces problèmes sont *globaux*. Les points conjugués sont globaux mais concernent des géodésiques infiniment voisines. Les points du cut locus sont doublement globaux car ils concernent des géodésiques issues du même point et reconvergeant vers un même point après avoir traversé des régions complètement différentes de M . Là encore, les calculs analytiques explicites deviennent vite difficiles et même souvent impraticables et des méthodes qualitatives comme celles introduites par Poincaré ou des méthodes numériques deviennent indispensables.

Pour traiter ces problèmes de caustiques, un bon outil est l'application dite "exponentielle" associée aux géodésiques. En général, étant donné un problème de contrôle optimal de trajectoires $\gamma_u(t)$, on peut considérer les trajectoires partant de $a \in M$ et regarder où elles arrivent en un temps t fixé, par exemple $t = 1$. On définit ainsi une fonctionnelle $E(a, u, t)$ dont il faut étudier la variation lorsque le contrôle u varie. Dire que u est optimal pour la minimisation du coût J signifie d'abord que u est un *point critique* de E (condition du premier ordre) et ensuite que u correspond bien à un minimum de J et pas à un autre type d'extremum (condition du deuxième ordre). Ce que nous avons vu plus haut se généralise et revient à appliquer aux fonctionnelles $J(u)$ la théorie des points critiques des fonctions différentiables $f(x)$ à la Whitney-Thom-Arnold, théorie fondée sur l'analyse des différentielles df et des hessiens $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)$ sur laquelle nous reviendrons à la section 1.4 du chapitre 15.

Dans les problèmes de géométrie sous-riemannienne qui nous intéressent, on regarde les géodésiques γ issues de $a \in M$. Elles dépendent de leur vitesse initiale dans le formalisme lagrangien et de leur moment initial $p \in T_a^*M$ dans le formalisme hamiltonien. En les suivant pendant un temps t , on construit les fronts d'ondes issus de a . On définit une application

$$\text{Exp}(a, p, t) : T^*M \times \mathbb{R} \rightarrow M$$

7. Pour un point de vue général sur les enveloppes dans le contexte inspiré de René Thom qui est le notre, le lecteur pourra se référer à l'article de Thom [528] commenté par Marc Chaperon dans [106].

que l'on peut utiliser sous plusieurs versions en fixant certaines variables à une valeur constante. Une des versions les plus utilisées est $\text{Exp}_a(t) : T_a^*M \rightarrow M$ qui associe à $p \in T_a^*M$ l'extrémité b au temps t de la géodésique $\gamma(t)$ issue de a avec le moment initial p . Au départ, la géodésique $\text{Exp}_{a,p}(t)$ est la seule géodésique reliant a à $b = \text{Exp}_{a,p}(t)$ et le front d'onde issu de a pris au temps t est comme le parallèle sur une sphère obtenu en suivant les méridiens issus d'un point pris comme pôle Nord pendant un temps donné. Mais de même que sur une sphère, après s'être dilatés en un équateur, les fronts d'onde parallèles se recontractent pour aller aboutir au point antipodal \hat{a} de a (le pôle Sud), à la traversée duquel les géodésiques se rencontrent, $\text{Exp}(a, p, t)$ peut avoir des singularités. Il y aura génériquement un domaine $R(a)$ ouvert et dense où tout sera régulier :

(i) les géodésiques seront uniques et réaliseront la distance entre a et $b = \text{Exp}_{a,p}(t)$, pour t fixé,

(ii) $\text{Exp}_a(t) : T_a^*M \rightarrow M$ sera un difféomorphisme local.

Le complémentaire $S(a) = M - R(a)$ sera un lieu *singulier* où l'on trouvera des points très particuliers.

1. Des points conjugués q où, répétons-le, la géodésique perd sa propriété de minimisation par rapport aux autres géodésiques qui lui sont infiniment proches et devient tangente à l'enveloppe des géodésiques issues de a . Cela signifie que le point q est un point critique de l'application exponentielle.
2. Des points de coupure (le cut locus) où la géodésique rentre en conflit avec une autre géodésique issue de a de même longueur mais étant passée par d'autres régions de M (comme lorsque sur une sphère une géodésique issue du pôle Nord le long d'un demi-méridien est supplantée lorsqu'elle travers le pôle Sud par la géodésique suivant le demi-méridien opposé). Si l'on définit plus généralement les points de coupure comme les points où la géodésique perd sa propriété de minimisation globale, alors les premiers points conjugués d'une géodésique sont aussi des points de coupure.

Dans les chapitres suivants, nous allons traiter ces problèmes pour nos modèles de géométrie sous-riemannienne. Dans cette section, nous allons nous borner, à titre de prélude, à rappeler quelques résultats classiques en faisant un *détour pédagogique* par les caustiques en optique et par les géodésiques des ellipsoïdes. Le lecteur intéressé par l'histoire des géodésiques au XIX^e siècle pourra se référer à l'article de Philippe Nabonnand [382].

6.1. Un détour par les caustiques en optique

Présentons d'abord brièvement la théorie des caustiques en optique géométrique. Il s'agit de la théorie des solutions lagrangiennes de l'équation des ondes. Le lecteur intéressé par des approfondissements trouvera une bibliographie détaillée dans notre

mémoire [413] d'*Introduction aux phénomènes critiques*. Citons quelques articles classiques comme par exemple Arnold [19], Berry-Upstill [52], Guckenheimer [234], Thom [532], ainsi que le remarquable séminaire Bourbaki [42] de Daniel Bennequin

6.1.1. *Un exemple.*

Dans l'approximation de l'optique géométrique dans un milieu homogène et isotrope, les rayons lumineux sont, d'après le principe de Fermat⁸, des droites normales aux fronts d'ondes. La nature des géodésiques ne fait donc pas problème. La vraie difficulté est celle des caustiques. Si S_0 est un front d'onde initial il évolue parallèlement à lui-même. Ce qui peut rendre la situation non triviale en introduisant des *singularités* sur les fronts d'ondes est, par conséquent, l'existence d'une *enveloppe* C des normales à S_0 . Si elle existe, celle-ci est dite *caustique* de la propagation⁹ Sur la caustique C il se produit une focalisation de l'énergie lumineuse. L'intensité y devient "infinie" car elle diverge dans l'approximation de l'optique géométrique. C'est pourquoi, dans un médium lumineux où l'on a interposé un écran, seules les caustiques sont observables. Comme le remarque Michael Berry, elles dominent les images optiques et sont phénoménologiquement structurantes.

La figure 10 représente une version simplifiée de la plus intuitive des caustiques, celle "du bol de café au lait lors d'un matin ensoleillé". Les rayons du soleil y sont représentés horizontalement de la gauche vers la droite. Le demi-cercle interne droit qu'ils illuminent constitue le front d'onde S_0 . Les rayons s'y réfléchissent comme sur un miroir conformément à la loi Snell-Descartes pour la réflexion (la normale à S_0 au point d'incidence est la bissectrice de l'angle entre le rayon incident et le rayon réfléchi, ce dernier leur étant coplanaire) et enveloppent la caustique.¹⁰ Celle-ci est la demi-néphroïde d'équations¹¹

$$\begin{cases} x = \frac{1}{4}(3 \cos(t) - \cos(3t)) \\ y = \frac{1}{4}(3 \sin(t) - \sin(3t)) \end{cases}$$

pour $t \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, le cusp se trouvant au point $(\frac{1}{2}, 0)$ correspondant à $t = 0$.

Les caustiques sont tout à fait banales et on peut les rencontrer partout. La figure 11 montre la photo d'une caustique plus complexe que la précédente produite par la réflexion du soleil sur l'assise d'un siège de métro aérien avec le dossier servant d'écran.

8. 1657, le précurseur du principe de moindre action de Maupertuis-Lagrange.

9. La caustique est donc l'évolutive de la famille des fronts d'ondes géométriques et ceux-ci sont différents involutes de celle-là.

10. Dans l'exemple concret, le soleil a une certaine hauteur, les rayons ont un angle d'incidence vertical non nul et les rayons réfléchis intersectent donc la surface du liquide qui fonctionne comme un écran.

11. La néphroïde est une épicycloïde à deux rebroussements qui a été étudiée dès la fin du XVII^e siècle, entre autres par Huygens et Bernouilli.

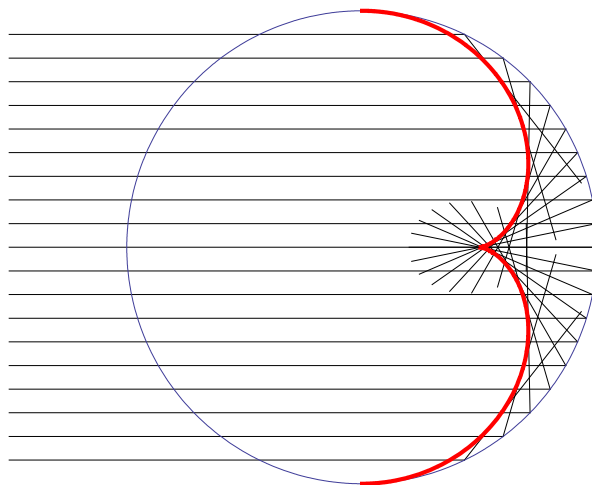


FIGURE 10. Caustique “du bol de café”. Les rayons sont représentés horizontalement de la gauche vers la droite. Le demi-cercle interne droit de centre 0 et de rayon 1 est illuminé par les rayons et constitue le front d’onde initial S_0 . Les rayons s’y réfléchissent (ils sont de longueur 0.8) et enveloppent la caustique. Celle-ci est une demi-néphroïde.

6.1.2. Géométrie des surfaces.

Pour comprendre les singularités que peuvent présenter les caustiques génériquement, on doit faire un peu de géométrie des surfaces. Soit donc S une surface lisse plongée dans \mathbb{R}^3 et choisissons des coordonnées telles qu’elle soit localement en 0 un graphe d’équation $z = z(x, y)$. Considérons les points $q = (u, v, w)$ de \mathbb{R}^3 comme des paramètres “externes” pour la fonction distance entre q et les points a de S . On obtient ainsi une famille de fonctions $V_q(a) = \|q - a\|^2$ à valeurs dans \mathbb{R} de variables “internes” (x, y) et paramétrées par $q \in \mathbb{R}^3$ fonctions dont on étudie les singularités. Si $f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ est une fonction différentiable, un point $a = (x, y) \in U$ est une singularité de f si le gradient de f , $\nabla f = \{\partial f / \partial x, \partial f / \partial y\}$, s’annule en a . Si 0 est singulier et si $f(0) = 0$, alors le développement de Taylor de f en 0 commence au second ordre :

$$f(a) = (\partial^2 f / \partial x^2) x^2 + 2 (\partial^2 f / \partial x \partial y) xy + (\partial^2 f / \partial y^2) y^2 + \dots$$

pour a assez voisin de 0.

Les singularités des fonctions $V_q(a)$ correspondent aux concepts traditionnels : centres de courbure principaux, surfaces focales (*i.e.* caustiques), lignes de courbure, etc.¹² En effet

$$V_q(x, y) = (x - u)^2 + (y - v)^2 + (z(x, y) - w)^2$$

12. Cf. par exemple Porteous [454].



FIGURE 11. Caustique de la réflexion du soleil sur l'assise d'un siège du métro parisien aérien avec le dossier servant d'écran.

et pour q fixé, les points critiques de $V_q(x, y)$ sont donnés par

$$\begin{cases} \frac{\partial V_q}{\partial x} = 2(x - u) + (z - w) \frac{\partial z}{\partial x} = 0 \\ \frac{\partial V_q}{\partial y} = 2(y - v) + (z - w) \frac{\partial z}{\partial y} = 0 \end{cases}$$

Le vecteur $\{x - u, y - v, z - w\}$ est alors $\left\{-(z - w) \frac{\partial z}{\partial x}, -(z - w) \frac{\partial z}{\partial y}, z - w\right\}$ qui est porté par la normale $\left\{-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1\right\}$. Donc “ q est point critique” signifie que q est sur la normale de S en $a = \{x, y, z(x, y)\}$. Prenons pour fixer les idées le cas simple de

$$z(x, y) = \frac{1}{2} (k_1 x^2 + k_2 y^2)$$

et $q = \{0, 0, w\}$. Le (demi)hessien de $V_q(x, y)$ est

$$H(x, y) = \begin{bmatrix} 1 + (z - w) \frac{\partial^2 z}{\partial x^2} + \left(\frac{\partial z}{\partial x}\right)^2 & (z - w) \frac{\partial^2 z}{\partial x \partial y} + \frac{\partial z}{\partial x} \frac{\partial z}{\partial y} \\ (z - w) \frac{\partial^2 z}{\partial x \partial y} + \frac{\partial z}{\partial x} \frac{\partial z}{\partial y} & 1 + (z - w) \frac{\partial^2 z}{\partial y^2} + \left(\frac{\partial z}{\partial y}\right)^2 \end{bmatrix}$$

et donc en 0, $H(0) = \begin{bmatrix} 1 - wk_1 & 0 \\ 0 & 1 - wk_2 \end{bmatrix}$. Le point critique 0 est donc dégénéré (le hessien n'est pas de rang maximal 2) si q est non seulement sur la normale (point critique) mais l'un des centres de courbures $\frac{1}{k_i}$ en a . L'ensemble singulier $K \subset \mathbb{R}^3$ des q pour lesquels $V_q(x, y)$ admet un point critique dégénéré est génériquement une *surface*. Les cas les plus connus où K dégénère sont la sphère (pour laquelle K est réduit à un point) et les cyclides de Dupin (pour lesquelles K est réduit à des courbes).

Dans le cas des ombilics il y a égalité des rayons de courbure principaux et le hessien s'annule. Le point $q = \{0, 0, \frac{1}{k}\}$ est alors une singularité ombilic au sens de Thom-Arnold et K en est génériquement un déploiement universel. Depuis longtemps (Cayley et Darboux) on a étudié en grand détail la structure locale d'une surface au voisinage d'un point ombilic. Dans le cas de stabilité structurelle, on dispose de résultats sur la structure globale des champs de directions principales.¹³

6.1.3. Solutions asymptotiques de l'équation des ondes.

Pour bien comprendre l'optique géométrique il faut tenir compte du fait qu'elle constitue le "squelette" d'un processus *ondulatoire*. Sans entrer dans les détails, disons simplement que l'on analyse le "squelette" géométrique de solutions $v(q, t)$ de l'équation des ondes $Dv = 0$ dans \mathbb{R}^3 où D est l'opérateur différentiel linéaire du 2ème ordre

$$D = \frac{\partial^2}{\partial t^2} - \Delta$$

(Δ étant le Laplacien spatial). La séparation des variables spatiales et temporelle conduit à chercher des solutions stationnaires de fréquence τ qui sont du type

$$v(q, t) = e^{i\tau t} u(q),$$

$u(q)$ étant une amplitude satisfaisant $D_\tau u = 0$, où $D_\tau = \tau^2 + \Delta$, avec pour condition initiale la donnée d'une fonction $u_0(q)$ sur le front d'onde source S_0 . Cela signifie que l'on cherche une propagation spatiale $u(q)$ de $u_0(q)$ sur laquelle vient se greffer un pattern ondulatoire $e^{i\tau t}$.

L'approximation de l'optique géométrique correspond à une fréquence τ *infinie*, *i.e.* à une longueur d'onde = 0. Mais pour $\tau = \infty$, l'opérateur D_τ n'est plus défini. D'où l'idée de chercher des *solutions asymptotiques* relativement à la fréquence τ — c'est-à-dire des solutions u_τ paramétrées par τ , égales à u_0 sur S_0 et définies à une

13. Cf. par exemple Sotomayor-Gutierrez [508]. Il s'agit de résultats classiques.

fonction à décroissance rapide en τ près — de la famille d'équations perturbées $D_\tau u_\tau = \varepsilon_\tau$ où ε_τ est une fonction à décroissance rapide en τ . "À la limite", la fonction u_∞ sera alors solution de l'équation $D_\infty u = 0$.

On cherche donc des solutions spatiales de la forme

$$u_\tau(q) = A_\tau(q)e^{i\theta} ,$$

où $\theta = \tau\varphi(q)$, $\varphi(q)$ étant une phase *spatiale*, et où l'amplitude $A_\tau(q)$ admet un développement asymptotique en τ de la forme :

$$A_\tau(q) \sim \tau^\mu \left[A_0(q) + \frac{1}{\tau} A_1(q) + \dots + \frac{1}{\tau^k} A_k(q) + \dots \right] \text{ avec } A_0 \neq 0 .$$

Lorsque $\tau \rightarrow \infty$, les $\frac{1}{\tau^k} \rightarrow 0$ pour $k \geq 1$. Les points où la solution $v(q, t)$ possède une phase *constante* (les fronts d'onde) satisfont l'équation $\tau(t + \varphi(q)) = \text{cste}$.

Il est facile de calculer $D_\tau u_\tau$ sous la forme d'un développement en puissances décroissantes de τ . Mais comme u_τ doit être solution de $D_\tau u_\tau = \varepsilon_\tau$ où ε_τ est à décroissance rapide (*i.e.* plus rapide que toute puissance négative de τ), les coefficients de ce développement doivent tous être identiquement nuls. Cela donne :

1. L'équation $A_0(1 - \|\nabla\varphi\|^2) \equiv 0$, où $\nabla\varphi$ est le gradient de la phase spatiale, soit, puisque $A_0 \neq 0$,

$$1 - \|\nabla\varphi\|^2 \equiv 0 ,$$

avec la condition initiale $\varphi|_{S_0} = 0$. Cette équation caractéristique de type Hamilton-Jacobi — dite *équation eikonale* — exprime que le gradient de la phase spatiale φ est de module constant et $= 1$. Cela signifie géométriquement que les surfaces de niveau S_t de φ sont des surfaces parallèles. Ce sont les fronts d'ondes de l'approximation géométrique, les lignes de gradient de φ (*i.e.* les courbes intégrales du champ de vecteurs $\nabla\varphi$) étant quant à elles les rayons lumineux.

2. L'équation pour A_0 , dite *équation de transport*,

$$A_0\Delta\varphi - 2\nabla\varphi \cdot \nabla A_0 = 0 ,$$

avec la condition initiale $A_0|_{S_0} = u_0$. Sur un rayon lumineux, comme $q(t)$ satisfait $\frac{dq}{dt} = 2\nabla\varphi$, on a

$$2\nabla\varphi \cdot \nabla A_0 = \nabla A_0 \cdot \frac{dq}{dt} = \frac{dA_0}{dt} ,$$

et l'équation de transport se réduit à une équation différentielle ordinaire linéaire (φ étant supposée connue) :

$$\frac{dA_0}{dt} + A_0\Delta\varphi = 0 .$$

- Il en va de même pour les autres coefficients A_k . Ils sont solutions d'équations linéaires ordinaires sur les rayons lumineux, ces derniers étant les *caractéristiques* de l'équation des ondes.

On peut ainsi, par intégrations successives, obtenir une solution asymptotique *locale*, définie dans un voisinage de la source S_0 . Ce qui fait obstruction à la construction d'une solution *globale* est l'enveloppe des caractéristiques, c'est-à-dire précisément la caustique C , car sur C l'amplitude a_0 *diverge*.

6.1.4. Solutions lagrangiennes.

Pour comprendre ce qui se passe au voisinage du lieu singulier qu'est une caustique, il faut approfondir la description géométrique des géodésiques que sont les rayons et passer au formalisme hamiltonien en introduisant les moments conjugués p des positions $q \in \mathbb{R}^3$. L'hamiltonien sur le fibré cotangent $T^*\mathbb{R}^3$ de \mathbb{R}^3 est

$$H(q, p) = 1 - \|p\|^2 .$$

Il exprime l'équation eikonale et représente ce que l'on appelle techniquement le *symbole principal* de l'opérateur de propagation D_τ .¹⁴ Le lien se fait en prenant pour covecteur p la 1-forme $d\varphi$. En effet, la métrique euclidienne de \mathbb{R}^3 établit un isomorphisme canonique entre $T^*\mathbb{R}^3$ et $T\mathbb{R}^3$ qui permet d'identifier $d\varphi$ au gradient $\nabla\varphi$ et donc l'équation eikonale $1 - \|\nabla\varphi\|^2 = 0$ à

$$1 - \|d\varphi\|^2 = H(q, d\varphi) = 0 .$$

Soit alors Λ_φ le graphe de $d\varphi$ sur l'ouvert U de \mathbb{R}^3 où la phase spatiale φ est définie. Il est facile de vérifier que, pour la structure symplectique canonique de $T^*\mathbb{R}^3$ définie par la 2-forme fondamentale $\Omega = dq \wedge dp$:

- Λ_φ est une sous-variété *lagrangienne* de T^*U , c'est-à-dire une sous-variété Λ de dimension moitié sur laquelle la 2-forme Ω s'annule identiquement. Cela est évident car $\Omega = dq \wedge dp = d\omega$ où ω est la 1-forme canonique pdq sur T^*U . Si $(\xi, \eta) \in T_{(q,p)}(T^*U)$ avec $\xi \in T_qU$, alors $\omega((\xi, \eta)) = p(\xi)$. Sur Λ_φ on a $p = d\varphi$ et donc $\omega((\xi, \eta)) = d\varphi(\xi)$. Par conséquent sur Λ_φ , $\Omega = d\omega = d^2\varphi \equiv 0$.
- Le symbole principal H s'annule sur Λ_φ puisque φ satisfait l'équation eikonale.
- Λ_φ étant un graphe est *transverse* en chacun de ses points aux fibres de la projection canonique $\pi : T^*U \rightarrow U$, ce qui signifie qu'en tout point x de Λ_φ la somme des espaces tangents à Λ_φ et à la fibre de π en x est tout l'espace tangent T_x^*U .

On voit que l'obstruction qu'il y a à construire une solution φ *globale* consiste en un éventuel *défait de transversalité*. D'où l'idée de chercher d'abord des solutions

14. Nous reviendrons plus en détail sur la notion de symbole principal d'un opérateur différentiel dans la section 9 du chapitre 16.

géométriques Λ à l'équation caractéristique et de ne poser que dans un second temps la question de leur représentation fonctionnelle.

On appelle alors “solution lagrangienne” du problème hamiltonien défini par le symbole H une variété lagrangienne Λ sur laquelle H s'annule (conditions (1) et (2)). Lorsque la condition (3) de transversalité est en plus remplie, Λ est localement représentable par le graphe Λ_φ d'une phase φ solution de l'équation caractéristique $H(q, d\varphi) = 0$. Remarquons que le champ $2\nabla\varphi$ (dont les rayons sont les courbes intégrales) est donné par $\frac{\partial H}{\partial p}$. Considérons alors le champ hamiltonien $X_H = \left(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial q}\right)$ défini par H sur l'espace des phases $T^*\mathbb{R}^3$, c'est-à-dire le système d'équations différentielles

$$\begin{cases} \frac{dq}{dt} = \frac{\partial H}{\partial p} = 2\nabla\varphi \\ \frac{dp}{dt} = -\frac{\partial H}{\partial q} = 0 \end{cases}$$

H est constant sur les courbes intégrales de X_H puisque

$$\frac{dH}{dt} = \frac{\partial H}{\partial q} \frac{dq}{dt} + \frac{\partial H}{\partial p} \frac{dp}{dt} = \frac{\partial H}{\partial q} \frac{\partial H}{\partial p} - \frac{\partial H}{\partial p} \frac{\partial H}{\partial q} = 0$$

et comme $H|_\Lambda \equiv 0$ par hypothèse, Λ est réunion de telles courbes, dites courbes *bicaractéristiques*.

Si $\Lambda = \Lambda_\varphi$, $p = d\varphi$, $2\nabla\varphi = \frac{\partial H}{\partial p}(q, d\varphi)$ et donc les rayons (les courbes caractéristiques) sont la *projection* sur \mathbb{R}^3 des courbes bicaractéristiques. La caustique C est par suite le *contour apparent* sur \mathbb{R}^3 de la solution lagrangienne Λ , *i.e.* la projection du lieu critique Σ où la condition (3) de transversalité n'est pas satisfaite.

On voit ainsi clairement à quoi tient l'obstruction due à l'existence de caustiques. La phase spatiale φ s'identifie à une longueur de chemin optique régie par un principe variationnel (principe de Fermat). C'est l'analogue d'une action. Dans \mathbb{R}^3 plusieurs rayons (plusieurs géodésiques) peuvent passer par un même point q . Si α est un paramétrage local de S_0 , φ dépend de α et les origines des rayons passant par q sont données par $\frac{\partial\varphi}{\partial\alpha} = 0$, *i.e.* par les points critiques de $\varphi_q(\alpha)$. Il y a caustique lorsqu'il y a coalescence de deux rayons passant par q , *i.e.* lorsqu'un des points critiques α devient *dégénéré*. L'existence de caustiques est donc lié à la *multiformité* de l'action φ . En passant dans l'espace des phases $T^*\mathbb{R}^3$, l'introduction des coordonnées conjuguées p lève l'obstruction et “désingularise” la situation. Les bicaractéristiques ne se coupent plus et l'on peut construire des solutions lagrangiennes globales. Les caustiques se retrouvent alors par projection de l'espace des phases $T^*\mathbb{R}^3$ sur l'espace de configurations \mathbb{R}^3 .

Nous verrons à la section 8.7.2 du chapitre 16 comment une “chair” ondulatoire se greffe sur ce “squelette” géométrique.

6.2. Un détour par les ellipsoïdes

Dès que l'on quitte l'espace euclidien et que l'on considère des variétés riemanniennes, le problème des géodésiques et de leurs caustiques devient beaucoup plus compliqué. Toujours dans le cadre de ce "prélude", nous allons expliquer un exemple, celui, classique, des ellipsoïdes. Nous n'entrerons pas dans tous les détails, en particulier ceux concernant des valeurs spéciales des paramètres, car il ne s'agit que d'exemples pédagogiques permettant de préciser certains concepts généraux.

6.2.1. La sphère.

Les géodésiques.— On considère donc des surfaces S plongées dans \mathbb{R}^3 dont la métrique est induite par la métrique euclidienne de \mathbb{R}^3 . Pour obtenir une métrique définie sur S il faut paramétrer S par deux coordonnées. C'est Gauss, le grand précurseur de Riemann, qui a introduit les outils de la géométrie des surfaces en tant que telles (indépendamment de leur plongement). Une fois choisie une paramétrisation de S on obtient la métrique par son ds^2 et l'on peut calculer les géodésiques avec les outils considérés dans le chapitre 6 consacré à la géométrie des connexions. Suivant le choix de la paramétrisation, les calculs pourront être plus ou moins praticables.

Par exemple les points $a = (x, y, z)$ de la sphère S^2 d'équation cartésienne

$$x^2 + y^2 + z^2 = 1$$

sont naturellement paramétrés en coordonnées sphériques par

$$\begin{cases} x = \sin(\varphi) \cos(\theta) \\ y = \sin(\varphi) \sin(\theta) \\ z = \cos(\varphi) \end{cases}$$

où $\varphi \in [0, \pi]$ est l'angle du rayon Oa avec l'axe des z (la latitude) et $\theta \in [0, 2\pi]$ est l'angle de la projection du rayon Oa sur le plan (x, y) avec l'axe des x (la longitude). Les pôles sont des points singuliers pour ces coordonnées car la longitude θ n'y est pas définie. Aux pôles, il faut utiliser une autre paramétrisation, par exemple cartésienne.

La métrique est évidente :

$$\begin{aligned}
ds^2 &= dx^2 + dy^2 + dz^2 \\
&= (\cos(\varphi) \cos(\theta) d\varphi - \sin(\varphi) \sin(\theta) d\theta)^2 \\
&\quad + (\cos(\varphi) \sin(\theta) d\varphi + \sin(\varphi) \cos(\theta) d\theta)^2 + (-\sin(\varphi) d\varphi)^2 \\
&= \left(\begin{array}{c} \cos^2(\varphi) \cos^2(\theta) d\varphi^2 - 2 \cos(\varphi) \cos(\theta) \sin(\varphi) \sin(\theta) d\varphi d\theta \\ + \sin^2(\varphi) \sin^2(\theta) d\theta^2 \end{array} \right) \\
&\quad + \left(\begin{array}{c} \cos^2(\varphi) \sin^2(\theta) d\varphi^2 - 2 \cos(\varphi) \sin(\theta) \sin(\varphi) \cos(\theta) d\varphi d\theta \\ + \sin^2(\varphi) \cos^2(\theta) d\theta^2 \end{array} \right) \\
&\quad + \sin^2(\varphi) d\varphi^2 \\
&= d\varphi^2 + \sin^2(\varphi) d\theta^2
\end{aligned}$$

Elle est dégénérée aux pôles où elle se réduit à $ds^2 = d\varphi^2$.

Il y a plusieurs façons de démontrer que les géodésiques sont des arcs de grands cercles, certaines étant très anciennes. Nous avons exposé à la section 6.3 du chapitre 6 sur la “Géométrie des connexions” la démonstration directe à partir des équations générales des géodésiques en géométrie riemannienne.

Mais une façon immédiate consiste à utiliser les symétries de la sphère. Soient A et B deux points. On choisit les axes de façon à ce que A et B soient sur le méridien $\theta = 0$, $\varphi \in [0, \pi]$ et on considère une courbe lisse $\gamma(t) = (x(t), y(t), z(t))$, $t \in [0, 1]$, les joignant. La longueur $\ell(\gamma)$ de γ est l’intégrale

$$\ell(\gamma) = \int_0^1 \sqrt{(\dot{\varphi})^2 + \sin^2(\varphi) (\dot{\theta})^2} dt$$

et évidemment $\ell(\gamma) \geq \int_0^1 \sqrt{(\dot{\varphi})^2} dt = |\varphi(A) - \varphi(B)|$ qui correspond à la longueur de l’arc de méridien AB donné par $\theta = \text{cste}$ et donc $\dot{\theta} = 0$.

Remarque. Il y a donc une infinité continue de géodésiques fermées simples (sans auto-intersection), en particulier, pour des pôles donnés, tous les méridiens et l’unique parallèle qu’est l’équateur, mais tout couple de points antipodaux peuvent être choisis comme pôles. Ce n’est plus du tout le cas pour les ellipsoïdes différents des sphères. \square

Une façon plus sophistiquée, mais générale, de trouver les géodésiques est d’utiliser les équations d’Euler-Lagrange pour le Lagrangien

$$L(\varphi, \theta, \dot{\varphi}, \dot{\theta}) = \sqrt{(\dot{\varphi})^2 + \sin^2(\varphi) (\dot{\theta})^2}.$$

Elles sont

$$\begin{cases} \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\varphi}} \right) = \frac{\partial L}{\partial \varphi} \\ \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}} \right) = \frac{\partial L}{\partial \theta} \end{cases}$$

Comme L ne dépend pas de θ , $\frac{\partial L}{\partial \theta} = 0$ et c'est la seconde équation la plus importante.

$$\frac{\partial L}{\partial \dot{\theta}} = \frac{\dot{\theta} \sin^2(\varphi)}{\sqrt{(\dot{\varphi})^2 + \sin^2(\varphi) (\dot{\theta})^2}} = C$$

$$\frac{d}{dt} \left(\frac{\dot{\varphi}}{\sqrt{(\dot{\varphi})^2 + \sin^2(\varphi) (\dot{\theta})^2}} \right) = \frac{\sin(\varphi) \cos(\varphi) (\dot{\theta})^2}{\sqrt{(\dot{\varphi})^2 + \sin^2(\varphi) (\dot{\theta})^2}}$$

où C est une constante d'intégration. Si $\dot{\varphi} \neq 0$, on peut, en divisant le numérateur et le dénominateur par $\dot{\varphi}$, se ramener à $\frac{\theta' \sin^2(\varphi)}{\sqrt{1 + \sin^2(\varphi) (\theta')^2}} = C$, soit, en élevant au carré,

$$(\theta')^2 \sin^4(\varphi) = C^2 (1 + \sin^2(\varphi) (\theta')^2)$$

$$(\theta')^2 = \frac{C^2}{\sin^2(\varphi) (\sin^2(\varphi) - C^2)}$$

où $\theta' = \frac{d\theta}{d\varphi}$. On notera que cette équation impose $|\sin(\varphi)| > |C|$ et donc $|C| < 1$. Posons $C = \sin(\varphi_0)$, $\varphi_0 \in]-\frac{\pi}{2}, \frac{\pi}{2}[$. Pour $C = 0$ ($\varphi_0 = 0$), si $\sin^2(\varphi) \neq 0$ ($\varphi \neq 0, \pi$, i.e. en dehors des pôles), on a $\theta' = 0$, i.e. $\theta = \theta_0 = \text{cste}$ et la géodésique est un arc de méridien. On suppose donc $C = \sin(\varphi_0) \neq 0$. Pour éviter les problèmes de signes, supposons que $C = \sin(\varphi_0)$ et $\sin(\varphi)$ soient positifs (i.e. $\varphi_0, \varphi \in]0, \frac{\pi}{2}[$) et que l'on ait donc

$$\theta' = \frac{\sin(\varphi_0)}{\sin(\varphi) \sqrt{(\sin^2(\varphi) - \sin^2(\varphi_0))}} .$$

Un changement de variable

$$u = \frac{1}{\tan(\varphi)} = \cot(\varphi), \quad du = -\frac{1}{\sin^2(\varphi)} d\varphi$$

permet alors de se ramener à une simple quadrature :

$$\begin{aligned}
d\theta &= \frac{\sin(\varphi_0)}{\sin(\varphi) \sqrt{(\sin^2(\varphi) - \sin^2(\varphi_0))}} d\varphi = \frac{-\sin(\varphi_0)}{\sin(\varphi) \sqrt{(\sin^2(\varphi) - \sin^2(\varphi_0))}} \sin^2(\varphi) du \\
&= \frac{-\sin(\varphi_0) \sin(\varphi)}{\sqrt{\cos^2(\varphi_0) \sin^2(\varphi) - \sin^2(\varphi_0) \cos^2(\varphi)}} du \\
&= -\frac{du}{\sqrt{\left(\frac{\cos^2(\varphi_0)}{\sin^2(\varphi_0)}\right) - u^2}} \text{ car } \sin(\varphi_0) \neq 0.
\end{aligned}$$

Cette intégrale bien connue donne $\theta = \text{Arccos}\left(\frac{u}{|\cot(\varphi_0)|}\right) + \theta_0$, soit, puisque $u = \cot(\varphi)$,

$$\cos(\theta - \theta_0) |\cot(\varphi_0)| = \cot(\varphi)$$

On note que pour $\varphi = \frac{\pi}{2}$ (i.e. sur l'équateur), $\cot(\varphi) = 0$ et donc $\cos(\theta - \theta_0) = 0$, autrement dit $\theta^\pm = \theta_0 \pm \frac{\pi}{2}$. Ces deux valeurs θ^\pm de θ définissent un diamètre de l'équateur. Considérons maintenant $\theta = \theta_0$. Cela détermine un angle φ modulo π . Comme $\varphi_0, \varphi \in]0, \frac{\pi}{2}[$ par hypothèse simplificatrice, il est égal à φ_0 . Les trois points (φ_0, θ_0) et $(\frac{\pi}{2}, \theta^\pm)$ déterminent un grand cercle de la sphère et il est évident que l'équation de ce grand cercle est l'équation de la géodésique. En effet ; les grands cercles sont les intersections de la sphère avec les plans passant par l'origine. Leurs équations sont donc de la forme $\alpha x + \beta y + \gamma z = 0$ pour certaines constantes (α, β, γ) , soit

$$\alpha \sin(\varphi) \cos(\theta) + \beta \sin(\varphi) \sin(\theta) + \gamma \cos(\varphi) = 0.$$

En divisant par $\sin(\varphi) \sqrt{\alpha^2 + \beta^2}$ si $\sin(\varphi) \neq 0$ (on exclut les pôles) et en posant

$$\frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} = \cos(\theta_0), \quad \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} = \sin(\theta_0),$$

on obtient

$$\cos(\theta_0) \cos(\theta) + \sin(\theta_0) \sin(\theta) = \cos(\theta - \theta_0) = \frac{-\gamma}{\sqrt{\alpha^2 + \beta^2}} \cot(\varphi)$$

qui est de la forme cherchée. On complète ensuite en regardant ce qui se passe pour les autres intervalles de définition, les différents signes et les cas limites. Nous n'entrons pas dans les détails.¹⁵

15. Le problème des géodésiques sur la sphère est élémentaire. Mais il devient techniquement assez sophistiqué lorsqu'il est appliqué à des mesures de géodésie terrestre (même sous l'hypothèse que le globe est sphérique ce qui n'est pas le cas) exigeant des expéditions et de l'astronomie. Cf. par exemple le cours d'"Astronomie sphérique" [64] de Pierre-Ossian Bonnet (1887) à qui l'on doit l'approfondissement des travaux pionniers de Gauss et la formule de Gauss-Bonnet. La technologie des mesures et des cartographies est un univers en soi.

Remarque. Dans ce qui précède, le paramétrage des courbes est quelconque. On utilise souvent le fait que lorsque l'on paramétrise les courbes par la longueur d'arc s , ce qui implique que la vitesse est de module 1, les géodésiques minimisent l'énergie qui est le carré du lagrangien précédent. Notons $\dot{\varphi}^s = \frac{d\varphi}{ds}$ et $\dot{\theta}^s = \frac{d\theta}{ds}$ et considérons donc le lagrangien $E = (\dot{\varphi}^s)^2 + \sin^2(\varphi) (\dot{\theta}^s)^2$. Les équations d'E-L deviennent alors $\dot{\theta}^s \sin^2(\varphi) = C$ et $\ddot{\varphi}^s = \sin(\varphi) \cos(\varphi) (\dot{\theta}^s)^2 = \frac{C^2 \cos(\varphi)}{\sin^3(\varphi)}$. La première équation $\dot{\theta}^s \sin^2(\varphi) = C$ est une intégrale première découverte par Clairaut.¹⁶ Elle est bien la même que $\frac{\partial L}{\partial \theta} = \frac{\dot{\theta} \sin^2(\varphi)}{\sqrt{(\dot{\varphi})^2 + \sin^2(\varphi) (\dot{\theta})^2}} = C$ puisque $\dot{\theta}^s = \frac{d\theta}{ds} = \dot{\theta} \frac{dt}{ds}$. Elle signifie que si α est l'angle de la géodésique avec les méridiens $\theta = \text{cste}$, $\sin(\varphi) \sin(\alpha) = \text{cste}$. \square

Remarque. Nous rencontrons là un bon exemple de ce que sont les progrès mathématiques. On peut démontrer avec des arguments de géométrie euclidienne classique que les grands cercles sont des géodésiques. Mais ces preuves ne se généralisent pas. Avec les équations d'Euler-Lagrange on dispose d'une méthode générale pour trouver les géodésiques. Et, dans le cas de la sphère, la trigonométrie permet de retrouver les grands cercles. \square

Une troisième façon de calculer les géodésiques est de passer du formalisme lagrangien au formalisme hamiltonien. Quand on passe du lagrangien-énergie (paramétrisation par la longueur d'arc) à l'hamiltonien on doit passer de la métrique $g = g_{\mu\nu} dx^\mu dx^\nu$ exprimée par la matrice $(g_{\mu\nu}) = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2(\varphi) \end{pmatrix}$ à la métrique exprimée dualement par la matrice inverse $(g^{\mu\nu}) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sin^2(\varphi)} \end{pmatrix}$. En termes des moments conjugués p_φ et p_θ des variables φ, θ , l'hamiltonien dont les géodésiques sont les projections est

$$H = \frac{1}{2} \left(p_\varphi^2 + \frac{p_\theta^2}{\sin^2(\varphi)} \right)$$

16. Mathématicien prodige, Alexis Claude Clairaut (1713-1765) fut admis à l'Académie des Sciences à 18 ans. Spécialiste de géodésie terrestre, il développa dans son ouvrage de [122] sur la *Théorie de la figure de la terre*, le modèle du globe terrestre comme "sphéroïde", c'est-à-dire comme ellipsoïde de révolution "oblat" (aplati), en corrélant l'aplatissement, la vitesse de rotation de la terre et la pesanteur. En plus des calculs, il participa à des expéditions (en Laponie, il y en eut plusieurs à partir de cette époque) destinées à effectuer des mesures géodésiques. Ce fut aussi un spécialiste de mécanique newtonienne céleste (planètes, lunes, problème à 3 corps, comètes).

et les équations de Hamilton sont par conséquent

$$\begin{aligned}\dot{\varphi} &= \frac{\partial H}{\partial p_\varphi} = p_\varphi, & \dot{\theta} &= \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{\sin^2(\varphi)} \\ \dot{p}_\varphi &= -\frac{\partial H}{\partial \varphi} = \frac{p_\theta^2 \cos(\varphi)}{\sin^3(\varphi)}, & \dot{p}_\theta &= -\frac{\partial H}{\partial \theta} = 0\end{aligned}$$

On voit que $p_\theta = \dot{\theta}^s \sin^2(\varphi)$ est constante. C'est l'intégrale première de Clairaut (1735). Dans le formalisme hamiltonien, elle est une conséquence du théorème de Noether¹⁷, disant que si l'hamiltonien est invariant par un groupe à 1 paramètre de symétries, le moment conjugué est invariant. Ici la symétrie est la rotation θ autour de l'axe des pôles et donc le moment p_θ est constant. L'autre équation de Hamilton donne $\ddot{\varphi}^s = \dot{p}_\varphi = \frac{p_\theta^2 \cos(\varphi)}{\sin^3(\varphi)}$ qui est bien $\ddot{\varphi}^s = \frac{C^2 \cos(\varphi)}{\sin^3(\varphi)}$ avec $C = p_\theta$. On vérifie trivialement que H est constant sur les trajectoires.

L'équation des géodésiques a été intégrée par Legendre (1806) et Bessel (1825). Elle est un cas particulier de celle, nettement plus compliquée, des ellipsoïdes tri-axiaux trouvée par Jacobi et que nous allons expliciter dans la section suivante.

Les caustiques. — Dans le cas de la sphère, les symétries rendent le problème des caustiques trivial. Tout point A peut être pris comme pôle et les géodésiques partant de A sont les méridiens. La caustique de A est son point antipodal \widehat{A} qui est à la fois le seul point conjugué de A , i.e. l'enveloppe des géodésiques partant de A (elles s'y croisent toutes), et le cut locus où toutes les géodésiques entrent en compétition puisqu'elles ont la même longueur.

6.2.2. Les ellipsoïdes de révolution.

Les géodésiques. — Après les sphères, viennent les ellipsoïdes de révolution \mathcal{E}_R (aussi appelés sphéroïdes) pour lesquels la symétrie par rapport à φ est brisée mais où celle par rapport à θ est maintenue. Les pôles sont les deux seuls points *ombilic* qui restent.¹⁸ Pour un \mathcal{E}_R dont l'axe des pôles est l'axe des z , l'équation est

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{c^2} = 1.$$

Traditionnellement, le facteur d'aplatissement est $f = \frac{a-c}{a}$. Si l'on pose $a = 1$ pour simplifier, la paramétrisation en coordonnées sphériques est (on garde $\varphi \in [0, \pi]$ et $\theta \in [0, 2\pi]$)

$$\begin{cases} x = \sin(\varphi) \cos(\theta) \\ y = \sin(\varphi) \sin(\theta) \\ z = c \cos(\varphi) \end{cases}$$

17. Cf. la section 3.1.

18. Rappelons que les points ombilics sont les points où les deux courbures principales de la surface sont égales. En ces points, la surface est infinitésimalement sphérique.

La symétrie de rotation par rapport à φ est brisée et se réduit à la symétrie $\varphi \rightarrow \pi - \varphi$ par rapport à l'équateur. Si $c < 1$, \mathcal{E}_R est aplati ("oblat"), si $c > 1$, \mathcal{E}_R est allongé ("prolat"), et si $c = 1$ on retrouve la sphère. Les pôles sont toujours des points singuliers où la longitude θ n'est pas définie.

La métrique devient

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 \\ &= (\cos^2(\varphi) + c^2 \sin^2(\varphi)) d\varphi^2 + \sin^2(\varphi) d\theta^2, \end{aligned}$$

l'hamiltonien devient

$$H = \frac{1}{2} \left(\frac{p_\varphi^2}{\cos^2(\varphi) + c^2 \sin^2(\varphi)} + \frac{p_\theta^2}{\sin^2(\varphi)} \right)$$

(en dehors des pôles où $\sin(\varphi) = 0$ et où la métrique $ds^2 = d\varphi^2$ devient dégénérée) et les équations de Hamilton sont par conséquent

$$\begin{aligned} \dot{\varphi} &= \frac{\partial H}{\partial p_\varphi} = \frac{p_\varphi}{\cos^2(\varphi) + c^2 \sin^2(\varphi)} \\ \dot{\theta} &= \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{\sin^2(\varphi)} \\ \dot{p}_\varphi &= -\frac{\partial H}{\partial \varphi} = \frac{p_\varphi^2 (c^2 - 1) \cos(\varphi) \sin(\varphi)}{(\cos^2(\varphi) + c^2 \sin^2(\varphi))^2} + \frac{p_\theta^2 \cos(\varphi)}{\sin^3(\varphi)} \\ \dot{p}_\theta &= -\frac{\partial H}{\partial \theta} = 0. \end{aligned}$$

On retrouve l'intégrale première de Clairaut $p_\theta = \text{cste}$, i.e. $\sin(\varphi) \sin(\alpha) = \text{cste}$, α étant l'angle de la géodésique avec les méridiens, et l'équation $\dot{\theta} = \frac{p_\theta}{\sin^2(\varphi)}$. Mais l'équation pour φ est beaucoup plus compliquée.

On peut la simplifier en paramétrant les géodésiques par leur longueur d'arc. Dans ce cas, en effet, H a la valeur constante $\frac{1}{2}$ et par conséquent

$$p_\varphi^2 = (\cos^2(\varphi) + c^2 \sin^2(\varphi)) \left(1 - \frac{p_\theta^2}{\sin^2(\varphi)} \right).^{19}$$

Les méridiens $\theta = \text{cste}$, c'est-à-dire $p_\theta = 0$, sont des géodésiques. Il en va de même pour l'équateur $p_\varphi = 0$, $\varphi = \frac{\pi}{2}$, $\dot{\theta} = p_\theta$. En effet, si $\varphi = \text{cste}$, $\dot{\varphi} = 0$ et donc $p_\varphi = 0$, ce qui implique $\dot{p}_\varphi = 0$ et donc $\frac{p_\theta^2 \cos(\varphi)}{\sin^3(\varphi)} = 0$, d'où $\varphi = \frac{\pi}{2}$ si $p_\theta \neq 0$. Ce sont les seules

19. Pour le fun, on peut vérifier que la valeur de

$$\dot{p}_\varphi = (\cos^2(\varphi) + c^2 \sin^2(\varphi)) \left(1 - \frac{p_\theta^2}{\sin^2(\varphi)} \right) \frac{(c^2 - 1) \cos(\varphi) \sin(\varphi)}{(\cos^2(\varphi) + c^2 \sin^2(\varphi))^2} + \frac{p_\theta^2 \cos(\varphi)}{\sin^3(\varphi)}$$

tirée des équations de Hamilton est bien égale à celle tirée de la dérivation

géodésiques fermées simples. On voit que la rupture de symétrie de rotation par rapport à φ et l'existence d'un seul axe de rotation ne fait subsister que l'équateur comme géodésique parallèle alors que le maintien de la symétrie de rotation par rapport à θ conserve tous les méridiens comme géodésiques.

Indépendamment de cela, on peut, pour mieux calculer (cf. par exemple les travaux de Bernard Bonnard et de ses collaborateurs [63]), faire un changement de variable sur φ en utilisant une intégrale elliptique Φ . On choisit Φ de façon à ce que

$$d\Phi^2 = (\cos^2(\varphi) + c^2 \sin^2(\varphi)) d\varphi^2 ,$$

et donc

$$\Phi(\varphi) = \int_0^\varphi \sqrt{1 - (1 - c^2) \sin^2(t)} dt .$$

Il s'agit de l'intégrale elliptique de deuxième espèce

$$E(\varphi, (1 - c^2))$$

qui rectifie l'arc d'ellipse et que nous avons explicitée à la section 11 du chapitre 9 à partir de Lagrange, Legendre et Gauss. On a $\Phi(-\varphi) = -\Phi(\varphi)$, $\Phi(\frac{\pi}{2}) = E$ (l'intégrale elliptique complète qui fonctionne comme l'équivalent de $\frac{\pi}{2}$ pour φ), $\Phi(\pi - \varphi) = 2E - \Phi(\varphi)$. Le passage de φ à Φ revient à remplacer les arcs de cercles par les arcs d'ellipse correspondants. Dans les coordonnées (Φ, θ) , la métrique se simplifie en

$$ds^2 = d\Phi^2 + \sin^2(\varphi(\Phi)) d\theta^2 = d\Phi^2 + G(\Phi) d\theta^2$$

et la complexité de la géométrie est transférée à la fonction $G(\Phi) = \sin^2(\varphi(\Phi))$.

L'hamiltonien devient

$$H = \frac{1}{2} \left(p_\Phi^2 + \frac{p_\theta^2}{G(\Phi)} \right)$$

$$\begin{aligned} \dot{p}_\varphi &= \frac{d}{d\varphi} \left(\sqrt{(\cos^2(\varphi) + c^2 \sin^2(\varphi)) \left(1 - \frac{p_\theta^2}{\sin^2(\varphi)} \right)} \right) \dot{\varphi} \\ &= \frac{d}{d\varphi} \left(\sqrt{(\cos^2(\varphi) + c^2 \sin^2(\varphi)) \left(1 - \frac{p_\theta^2}{\sin^2(\varphi)} \right)} \right) \frac{\sqrt{(\cos^2(\varphi) + c^2 \sin^2(\varphi)) \left(1 - \frac{p_\theta^2}{\sin^2(\varphi)} \right)}}{\cos^2(\varphi) + c^2 \sin^2(\varphi)} \end{aligned}$$

$$\dot{\varphi}^s = \frac{\sqrt{1 - \frac{p_\theta^2}{\sin^2(\varphi)}}}{\sqrt{\cos^2(\varphi) + c^2 \sin^2(\varphi)}}$$

et les équations de Hamilton sont (apparemment) plus simples :

$$\begin{aligned}\dot{\Phi} &= \frac{\partial H}{\partial p_{\Phi}} = p_{\Phi} \\ \dot{\theta} &= \frac{\partial H}{\partial p_{\theta}} = \frac{p_{\theta}}{G(\Phi)} \\ \dot{p}_{\Phi} &= -\frac{\partial H}{\partial \Phi} = -\frac{1}{2} \frac{p_{\theta}^2 G'(\Phi)}{G(\Phi)^2} \\ \dot{p}_{\theta} &= -\frac{\partial H}{\partial \theta} = 0.\end{aligned}$$

On retrouve l'intégrale première de Clairaut $p_{\theta} = \text{cste}$, et l'équation $\dot{\theta} = \frac{p_{\theta}}{G(\Phi)} = \frac{p_{\theta}}{\sin^2(\varphi)}$. La fonction $G(\Phi)$ est analytiquement plus compliquée que $\sin^2(\Phi)$ mais elle lui reste qualitativement proche.

Remarque. Comme $p_{\Phi} = \dot{\Phi}$, l'hamiltonien H représentant l'énergie s'écrit

$$H = \frac{1}{2} (\dot{\Phi})^2 + \frac{1}{2} \frac{p_{\theta}^2}{G(\Phi)}$$

qui est l'hamiltonien d'un système mécanique d'énergie cinétique $\frac{1}{2} (\dot{\Phi})^2$ et d'énergie potentielle $\frac{1}{2} \frac{p_{\theta}^2}{G(\Phi)}$, p_{θ} étant un paramètre constant. □

On peut alors comprendre qualitativement les géodésiques en remarquant que si on les paramétrise par la longueur d'arc (i.e. en faisant $H = \frac{1}{2}$), elles sont définies par

$$(\dot{\Phi})^2 = (p_{\Phi})^2 = 1 - \frac{p_{\theta}^2}{G(\Phi)}.$$

Pour $p_{\theta} = 0$, on obtient $\theta = \text{cste}$, i.e. un méridien parcouru avec un Φ à vitesse constante. Si l'on veut une géodésique qui soit un parallèle, il faut $\dot{\Phi} = 0$, donc $p_{\Phi} = 0$, donc $\frac{p_{\theta}^2}{G(\Phi)} = 1$ et aussi $\dot{p}_{\Phi} = 0$, donc $G'(\Phi) = 0$, soit $\varphi = \frac{\pi}{2}$. Il s'agit bien de l'équateur avec $\Phi(\frac{\pi}{2}) = E$, $p_{\theta} = G(E) = 1$. En revanche pour les valeurs intermédiaires de p_{θ} , Φ oscille entre deux limites Φ^+ et $\Phi^- = -\Phi^+$ et la géodésique est parcourue à vitesse unité avec la composante $\dot{\theta} = \frac{p_{\theta}}{G(\Phi)}$.

Les caustiques. — Si A est un des pôles (point ombilic), la situation est celle de la sphère et sa caustique est son point antipodal \hat{A} qui est à la fois son lieu conjugué et son cut locus. Mais pour les autres points cette dégénérescence due à la symétrie de révolution disparaît et \hat{A} se déploie en un lieu conjugué générique qui est une *astroïde* à quatre cusps²⁰ et en un cut locus qui est un segment joignant deux cusps opposés, ces cusps étant donc à la fois des points conjugués et des points de coupure (en

20. Comme un carreau de carte à jouer. C'est Jacobi qui introduisit ce terme.

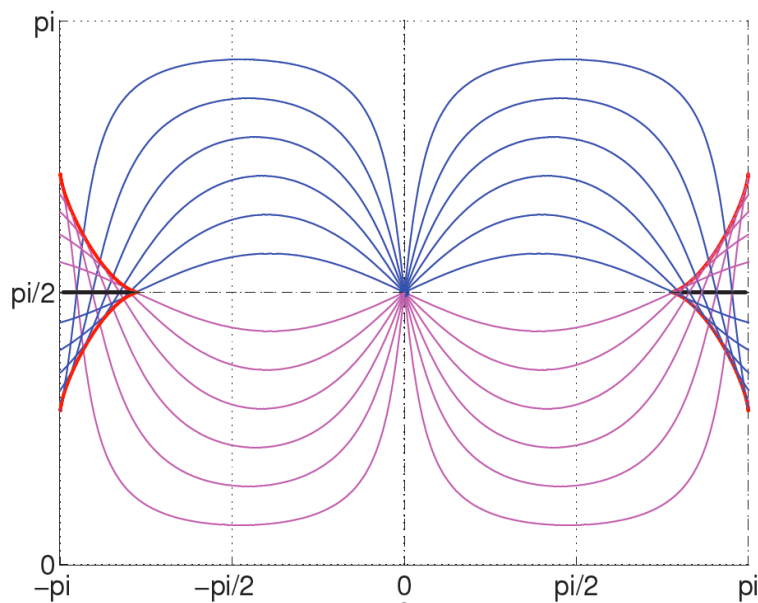


FIGURE 12. Géodésiques issues d'un point de l'équateur d'un ellipsoïde de révolution oblat, ainsi que leur enveloppe (Figure d'Olivier Cots).

dehors de l'astroïde, deux géodésiques passent par chaque point, mais à l'intérieur, il y en a quatre). Ce beau résultat a été démontré par Jacobi dans la sixième de ses *Vorlesungen über Dynamik* [270] de 1842-1843 à Königsberg.²¹ Cette leçon est consacrée au principe de moindre action (“Princip der kleinsten Wirkung”) et on y trouve (p. 47) l'image d'une géodésique tangente à l'astroïde.

Les figures 12 et 13 dues à Bernard Bonnard, Olivier Cots et Lionel Jassionnesse montrent la caustique d'un point de l'équateur dans le cas oblat. La figure 12 montre plusieurs géodésiques oscillantes et leur enveloppe. La figure 13 montre trois géodésiques, l'astroïde de la caustique et le cut locus joignant deux de ses cusps.

Lorsque le point est un point générique (i.e. ni un pôle, ni sur l'équateur), l'astroïde du lieu conjugué a deux cusps sur le parallèle antipodal qui sont reliés par le cut locus et deux cusps sur le méridien antipodal.

En 1842, Jacobi fit la conjecture appelée “la dernière conjecture géométrique de Jacobi” que même pour les ellipsoïdes triaxiaux le lieu conjugué d'un point non ombilical restait une astroïde.

6.2.3. Les ellipsoïdes triaxiaux.

Les géodésiques. — Après les ellipsoïdes de révolution viennent les ellipsoïdes quelconques \mathcal{E} (aussi appelés “triaxiaux”) pour lesquels non seulement la symétrie

21. Éditées en 1866 à titre posthume par Alfred Clebsch.

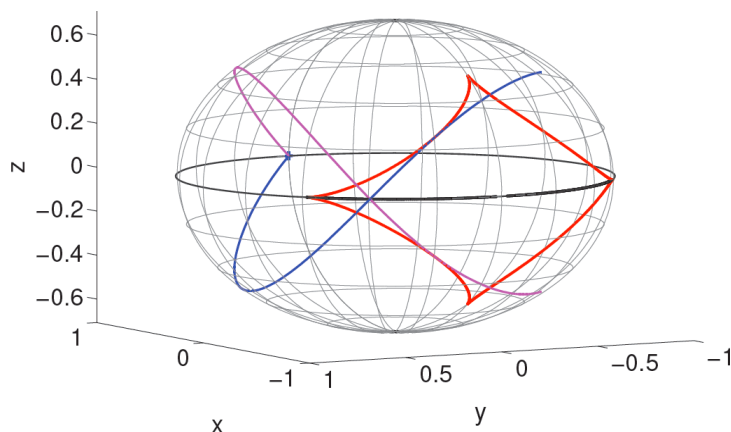


FIGURE 13. Géodésiques d'un point de l'équateur d'un ellipsoïde de révolution oblat : astroïde de la caustique et cut locus joignant deux de ses cusps. (Figure d'Olivier Cots).

par rapport à φ est brisée mais également celle par rapport à θ . L'équation est

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

avec $a > b > c > 0$. Le problème est beaucoup plus complexe (par exemple l'intégrale première de Clairaut n'existe plus) et a posé un véritable défi aux mathématiciens de la première moitié du XIX^e siècle. C'est Jacobi qui en 1839 a utilisé la paramétrisation dite "elliptique" qui lui a permis de ramener le problème des géodésiques à des quadratures. Celles-ci ne sont pas du tout triviales puisqu'elles font intervenir des intégrales abéliennes de type *hyperelliptique*, mais elles ont l'avantage d'exister et d'être explicites.

Les coordonnées elliptiques sont données par les intersections de l'ellipsoïde \mathcal{E} avec les hyperboloïdes confocaux à une ou deux nappes. Elles sont très bien expliquées, avec de belles figures, dans le célèbre traité de 1932 de David Hilbert et Stephan Cohn-Vossen *Anschauliche Geometrie* [251]. Elles sont données par les formules

$$\begin{cases} x = a \cos(\theta) \sqrt{\frac{a^2-b^2}{a^2-c^2} \cos^2(\varphi) + \sin^2(\varphi)} \\ y = b \sin(\theta) \sin(\varphi) \\ z = c \cos(\varphi) \sqrt{\frac{b^2-c^2}{a^2-c^2} \cos^2(\theta) + \sin^2(\theta)} \end{cases}$$

qui redonnent les équations d'un ellipsoïde de révolution oblat lorsque $a = b > c$ et prolata lorsque $a > b = c$ (il faut alors intervertir x et z). On peut continuer à choisir $\theta \in [0, 2\pi]$ et $\varphi \in [0, \pi]$. Les courbes $\theta = \text{cste}$ et $\varphi = \text{cste}$ forment un système de courbes orthogonales qui sont les *lignes de courbure principales*.

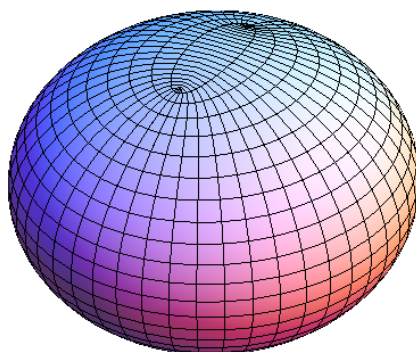


FIGURE 14. Géométrie d'un ellipsoïde triaxial \mathcal{E} pour a voisin de b : $a^2 = 1$, $b^2 = 0.95$, $c^2 = 0.5$. L'axe z est vertical, l'axe y frontal et l'axe x en perspective.

La figure 14 illustre la géométrie d'un \mathcal{E} pour a voisin de b : $a^2 = 1$, $b^2 = 0.95$, $c^2 = 0.5$ (l'axe z est vertical, l'axe y frontal et l'axe x en perspective). Comme $a \sim b$, \mathcal{E} est une petite déformation de \mathcal{E}_R ($a = b$). Dans la "zone Nord" déformant un peu le voisinage du pôle Nord de \mathcal{E}_R , les "parallèles" $\theta = \text{cste}$ sont des cercles déformés en "ellipses" un peu courbées et les "méridiens" $\varphi = \text{cste}$ sont comme dédoublés en branches d'hyperboles un peu tordues.

L'équateur, qui correspond à $\varphi = \frac{\pi}{2}$, est l'ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \{z = 0, x = a \cos(\theta), y = b \sin(\theta)\}$$

du plan (x, y) . Dualement, pour $\theta = \pm \frac{\pi}{2}$ on obtient l'ellipse

$$\frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \{x = 0, y = \pm b \sin(\varphi), z = c \cos(\varphi)\}$$

du plan (z, y) . Par ailleurs, on obtient la section $y = 0$ en faisant $\varphi = 0, \pi$ ou $\theta = 0, \pi$. Comme $y = 0$ puisque $\sin(\varphi) = 0$ ou $\sin(\theta) = 0$, on obtient l'ellipse

$$\frac{x^2}{a^2} + \frac{z^2}{c^2} = 1$$

du plan (x, z) . Mais celle-ci est décomposée en 4 arcs.

Pour $\varphi = 0, \pi$ on obtient une sorte de déploiement des pôles de l' \mathcal{E}_R ($a = b$). Il s'agit de deux petits arcs d'ellipse qui pour $a = b$ se contractent sur les pôles $(0, 0, \pm c)$

$$\begin{cases} x = a \cos(\theta) \sqrt{\frac{a^2-b^2}{a^2-c^2}} \\ y = 0 \\ z = \pm c \sqrt{\frac{b^2-c^2}{a^2-c^2} \cos^2(\theta) + \sin^2(\theta)} \end{cases}$$

Ces arcs se terminent, pour $\theta = 0, \pi$, aux 4 points $(\pm a \sqrt{\frac{a^2-b^2}{a^2-c^2}}, 0, \pm c \sqrt{\frac{b^2-c^2}{a^2-c^2}})$.

Pour $\theta = 0, \pi$ on obtient deux grands arcs d'ellipse qui pour $a = b$ se dilatent en deux méridiens.

$$\begin{cases} x = \pm a \sqrt{\frac{a^2-b^2}{a^2-c^2} \cos^2(\varphi) + \sin^2(\varphi)} \\ y = 0 \\ z = c \cos(\varphi) \sqrt{\frac{b^2-c^2}{a^2-c^2}} \end{cases}$$

Pour $\varphi = 0, \pi$, ces arcs se terminent aux mêmes 4 points $(\pm a \sqrt{\frac{a^2-b^2}{a^2-c^2}}, 0, \pm c \sqrt{\frac{b^2-c^2}{a^2-c^2}})$. Ces 4 points sont 4 ombilics (les seuls) qui dédoublent les deux seuls points ombilicaux de l' \mathcal{E}_R ($a = b$) que sont les pôles. La figure 15 explicite cette géométrie.

Lorsque l'écart entre a et b grandit cette géométrie se déforme. Pour b au voisinage de $\frac{a+c}{2}$ elle est la plus asymétrique, puis lorsque b se rapproche de c elle se rapproche de la géométrie de l' \mathcal{E}_R prolat ($b = c$) de pôles $(\pm a, 0, 0)$. La figure 16 montre cette évolution

Même pour une surface aussi simple qu'un \mathcal{E} triaxial, la métrique est déjà notablement compliquée. Les dérivées partielles des coordonnées sont

$$\begin{cases} \frac{\partial x}{\partial \theta} = -a \sin(\theta) \sqrt{\frac{a^2-b^2}{a^2-c^2} \cos^2(\varphi) + \sin^2(\varphi)} \\ \frac{\partial x}{\partial \varphi} = \frac{a(b^2-c^2) \cos(\theta) \cos(\varphi) \sin(\varphi)}{(a^2-c^2) \sqrt{\frac{a^2-b^2}{a^2-c^2} \cos^2(\varphi) + \sin^2(\varphi)}} \\ \frac{\partial y}{\partial \theta} = b \cos(\theta) \sin(\varphi) \\ \frac{\partial y}{\partial \varphi} = b \cos(\varphi) \sin(\theta) \\ \frac{\partial z}{\partial \theta} = \frac{(a^2-b^2)c \cos(\theta) \cos(\varphi) \sin(\theta)}{(a^2-c^2) \sqrt{\frac{b^2-c^2}{a^2-c^2} \cos^2(\theta) + \sin^2(\theta)}} \\ \frac{\partial z}{\partial \varphi} = -c \sin(\varphi) \sqrt{\frac{b^2-c^2}{a^2-c^2} \cos^2(\theta) + \sin^2(\theta)} \end{cases}$$

Quand on calcule le ds^2 , on vérifie que le terme en $d\theta d\varphi$ s'annule alors que les termes en $d\theta^2$ et $d\varphi^2$ donnent les coefficients $g_{\theta\theta}$ et $g_{\varphi\varphi}$

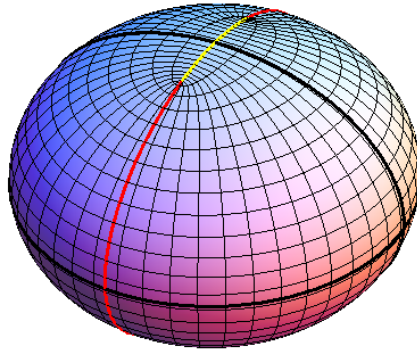


FIGURE 15. L'équateur $z = 0$ est représenté en noir, ainsi que la section $x = 0$. Les petits arcs d'ellipse joignant entre eux les deux ombilics Nord ou Sud sont en jaune, seul celui au Nord étant visible. Les grands arcs d'ellipse joignant un ombilic Nord à l'ombilic Sud colatéral sont en rouge.

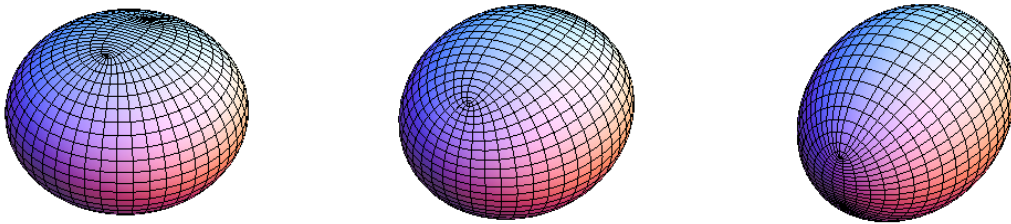


FIGURE 16. La déformation de la géométrie d'un ellipsoïde triaxial entre un cas voisin d'un ellipsoïde de révolution oblat ($a^2 = 1, b^2 = 0.95, c^2 = 0.5$) et un cas voisin d'un ellipsoïde de révolution prolata ($a^2 = 1, b^2 = 0.55, c^2 = 0.5$). Le cas intermédiaire est de paramètres ($a^2 = 1, b^2 = 0.75, c^2 = 0.5$).

$$g_{\theta\theta} = \frac{(a^2 + b^2 + (b^2 - a^2) \cos(2\theta))(a^2 - c^2 + (b^2 - a^2) \cos(2\theta) + (c^2 - b^2) \cos(2\varphi))}{2(a^2 + b^2 - 2c^2 + (b^2 - a^2) \cos(2\theta))}$$

$$g_{\varphi\varphi} = \frac{(b^2 + c^2 + (b^2 - c^2) \cos(2\varphi))(-a^2 + c^2 + (a^2 - b^2) \cos(2\theta) + (b^2 - c^2) \cos(2\varphi))}{2(-2a^2 + b^2 + c^2 + (b^2 - c^2) \cos(2\varphi))}$$

On notera qu'aux ombilics où $\cos(2\theta) = \cos(2\varphi) = 1$, $g_{\theta\theta}$ et $g_{\varphi\varphi}$ sont nuls et la métrique y est totalement dégénérée par rapport aux coordonnées elliptiques. On notera aussi que pour $a = b = 1$ on retrouve les $g_{\theta\theta} = \frac{1}{2}(1 - \cos(2\varphi)) = \cos^2(\varphi)$ et $g_{\varphi\varphi} = \cos^2(\varphi) + c^2 \sin^2(\varphi)$ de \mathcal{E}_R .

L'usage virtuose des coordonnées elliptiques a permis à Jacobi de découvrir qu'elles permettent de *séparer les variables* θ et φ dans l'équation des géodésiques.²² De façon très laconique, il explique dans une annonce à l'Académie de Berlin du 18 avril 1838 ([269], 1839) et une annonce à l'Académie des Sciences de Paris du 28 décembre 1838, qu'il s'agit d'intégrales abéliennes (en fait hyperelliptiques), l'une en θ , l'autre en φ et il trouve une intégrale première qui généralise celle de Clairaut. Après avoir rappelé que le problème

“semble présenter les plus grandes difficultés analytiques. En effet, l'équation différentielle du second ordre dont ce problème dépend, quand on choisit les variables dont on fait ordinairement usage, paraît d'une forme si compliquée, qu'on est facilement rebuté de le traiter.” ([269])

il explique que “après plusieurs tentatives inutiles” la séparation “astucieuse” des variables au moyen des coordonnées elliptiques ramène le problème à des quadratures qui sont pour lui “les plus simples formules du monde”.

$$D = \int \frac{\sqrt{b^2 \cos^2(\varphi) + c^2 \sin^2(\varphi)}}{\sqrt{a^2 - b^2 \cos^2(\varphi) - c^2 \sin^2(\varphi)} \sqrt{(b^2 - c^2) \sin^2(\varphi) - C}} d\varphi - \int \frac{\sqrt{a^2 \sin^2(\theta) + b^2 \cos^2(\theta)}}{\sqrt{a^2 \sin^2(\theta) + b^2 \cos^2(\theta) - c^2} \sqrt{(a^2 - b^2) \sin^2(\theta) + C}} d\theta$$

C étant la constante

$$C = (b^2 - c^2) \sin^2(\varphi) \sin^2(\alpha) - (a^2 - b^2) \sin^2(\theta) \cos^2(\alpha)$$

qui pour $a = b$ redonne l'intégrale première de Clairaut $\sin(\varphi) \sin(\alpha) = \text{cste}$. Pour $a = b$, la seconde intégrale abélienne se trivialisait et devient

$$\int \frac{a}{(a^2 - c^2) \sin(\varphi) \sin(\alpha)} d\theta = \frac{a}{(a^2 - c^2) \sin(\varphi) \sin(\alpha)} \theta + E .$$

Quant à la première intégrale, elle devient l'intégrale elliptique

$$\int \frac{\sqrt{a^2 - (a^2 - c^2) \sin^2(\varphi)}}{(a^2 - c^2) \sin^2(\varphi) \cos(\alpha)} d\varphi .$$

Les trois ellipses sections de \mathcal{E} par les plans de coordonnées sont les seules géodésiques fermées. Les géodésiques qui ne sont pas fermées sont des courbes de

22. Les coordonnées elliptiques et les formules d'intégration des géodésiques forment les leçons 26 à 28 des *Vorlesungen über Dynamik* [270]. Cf. plus haut.

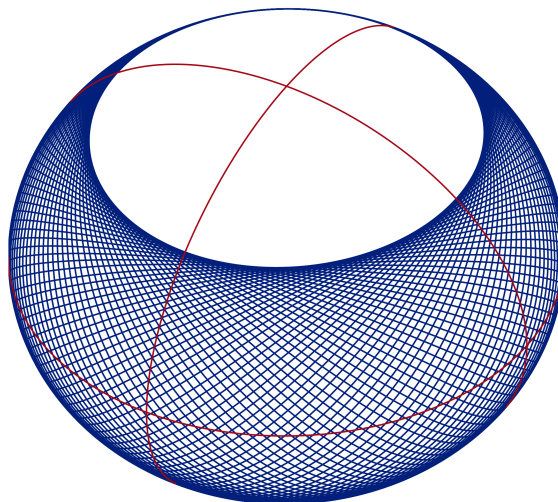


FIGURE 17. Une géodésique de l'ellipsoïde triaxial ($a = 1.01, b = 1, c = 0.8$) spirant indéfiniment de façon dense entre les limites $\varphi^+ = 44.9^\circ$ et $\varphi^- = 134.1^\circ$ (D'après Karney [283])

longueur infinies sur la surface compacte \mathcal{E} et doivent donc s'y “enrouler”. Supposons par exemple $c < 1$ (cas oblat) et considérons une géodésique γ qui ne soit ni un méridien ni l'équateur (géodésiques fermées), et qui part de l'équateur avec un angle $\alpha_0 \neq 0, \frac{\pi}{2}$. Elle monte jusqu'à une latitude φ^+ , hauteur à laquelle elle devient tangente au parallèle φ^+ puis redescend, traverse l'équateur, rejoint la latitude $\varphi^- = \pi - \varphi^+$ à laquelle elle redevient tangente au parallèle, et oscille indéfiniment entre ces deux valeurs en spirant de façon dense sans jamais se fermer. La figure 17 tirée de l'excellente page de Wikipedia “Geodesics on an ellipsoid” créée par Charles Karney montre, pour $a = 1.01, b = 1, c = 0.8$, la géodésique oscillant entre $\varphi^+ = 44.9^\circ$ et $\varphi^- = 134.1^\circ$.

Les caustiques. — Comme nous l'avons noté, “la dernière conjecture géométrique de Jacobi” (1842) était que la caustique d'un point générique d'un ellipsoïde triaxial restait une astroïde. Elle fut à l'origine de nombreux travaux. Elle fut vérifiée numériquement dès la fin du XIX^e siècle (un célèbre dessin de 1882 dû au géomètre Anton von Braunmühl (1853-1908), cf. [382], fut reproduit dans de nombreux ouvrages), mais elle résista à tous les efforts et ne fut rigoureusement démontrée qu'en 2004 (donc plus d'un siècle et demi après) par Jin-ichi Itoh et Kazuyoshi Kiyohara dans leur article [265] “The Cut Loci and the Conjugate Loci on Ellipsoids”. La figure 18 (issue du programme *Mathematica*) de Thomas Waters “The Conjugate Locus on the Triaxial Ellipsoid” [558] montre le lieu conjugué d'un point générique

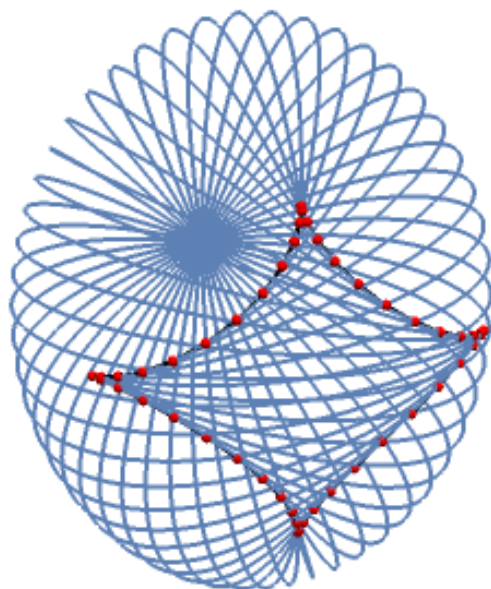


FIGURE 18. Le lieu conjugué d'un point générique de l'ellipsoïde triaxial de demi-axes $a = 1, b = 1.2, c = 1.5$. Plusieurs géodésiques sont représentées ainsi que leurs points de tangence avec leur enveloppe (points rouges). D'après Waters [558].

de l'ellipsoïde de demi-axes $a = 1, b = 1.2, c = 1.5$. Plusieurs géodésiques sont représentées ainsi que leurs points de tangence avec leur enveloppe (points rouges)

Une différence importante avec les ellipsoïdes de révolution (qui sont soit oblates soit prolates) est que sur un ellipsoïde triaxial, il y a des géodésiques à la fois de type oblat et de type prolat et le cut locus joignant deux des cusps de l'astroïde se décompose en un segment central de type prolat encadré par deux segments de type oblat aboutissant aux cusps.

6.2.4. Généralisations.

On peut généraliser cette investigation des géodésiques sur les ellipsoïdes et de leurs caustiques en considérant des surfaces convexes S définies par des métriques analytiques sur la sphère. C'est ce qu'a fait par exemple Henri Poincaré dans un article éblouissant de 1905 [451] en relation avec ses approches qualitatives révolutionnaires du problème des trois corps dont il rappelle "l'extrême complexité". Poincaré s'intéressait particulièrement "aux géodésiques fermées qui jouent ici le rôle des solutions périodiques du problème des trois corps". Nous avons vu que sur un ellipsoïde triaxial il n'existe que trois géodésiques fermées (les trois ellipses sections de l'ellipsoïde par les plans de coordonnées). Les géodésiques fermées sont donc *rare*.

Soit a un point de S pris comme origine. Poincaré considère des coordonnées géodésiques qui sont comme des coordonnées polaires mais basées sur les géodésiques issues de a . On prend une géodésique γ_0 de référence issue de a et si x est sur la géodésique γ issue de a , on le repère (i) par la distance r entre a et x et (ii) par l'angle θ que fait γ avec γ_0 . La métrique est donc définie par un ds^2 de la forme

$$ds^2 = dr^2 + \lambda(r, \theta)^2 d\theta^2 .$$

Si deux géodésiques issues de a , différentes mais infiniment voisines, se rejoignent en un nouveau point c (il peut y avoir une infinité de tels points), le ds^2 est 0 en ce point et donc $dr = 0$ et $\lambda(r, \theta) = 0$. On obtient ainsi l'équation du lieu conjugué $\lambda(r, \theta) = 0$. Poincaré appelle les points conjugués des “foyers”, les cusps des foyers “en pointe” ou “en talon” (métaphore ferroviaire) et le cut locus la “ligne de partage”. Il rappelle alors la théorie des sphéroïdes comme petites déformations de la sphère euclidienne, analyse les géodésiques fermées et leurs géodésiques voisines. Il montre que le nombre total des géodésiques fermées est impair. Il étudie les points doubles des géodésiques fermées (ce sont des croisements normaux car une géodésique ne peut avoir ni cusps ni points où elle est tangente à elle-même) et montre que

“sur une surface convexe quelconque, il y a toujours au moins une géodésique fermée sans point double et il y en a toujours un nombre impair” ([451], p. 254)

(il y en a trois pour les ellipsoïdes triaxiaux). Il développe ensuite la théorie de la stabilité et de l'instabilité des géodésiques fermées et montre que

“si sur une surface convexe quelconque on envisage toutes les géodésiques fermées sans point double l'excès du nombre de celles qui sont stables sur le nombre de celles qui sont instables est constant ; il est donc le même que pour l'ellipsoïde, il est donc égal à 1. Sur une surface convexe, il y a donc toujours au moins une géodésique fermée stable sans point double.” ([451], p. 259)

Par exemple, sur un ellipsoïde triaxial, les deux géodésiques fermées $x = 0, z = 0$ sont stables mais la géodésique intermédiaire $y = 0$ est instable²³

Le cut locus est un arbre fini, dont les points de branchement sont les points où il y a plus de deux géodésiques de même longueur qui sont en compétition. Les extrémités des branches sont quant à elles des cusps du lieu conjugué.

23. Dans l'analogie avec les mouvements astronomiques, les géodésiques que sont les grands cercles de la sphère correspondent aux mouvements keplériens d'une planète autour d'une étoile, et les géodésiques perturbées sur un sphéroïde correspondent aux mouvements d'un problème à trois corps, étoile, planète, satellite.

Retour sur les elasticæ : strates de Maxwell et points conjugués (d'après Y. Sachkov)

Maintenant que nous disposons du cadre général de la théorie du contrôle optimal et des notions de géodésique et de caustique en géométrie riemannienne, il est intéressant de revenir dans ce contexte sur la théorie des elasticæ dont nous avons résumé la théorie classique (eulérienne) dans la section 10 du chapitre 9. Cela nous servira d'introduction à nos modèles de géodésiques sous-riemanniennes qui feront l'objet du chapitre suivant.

L'expression des elasticæ en termes d'intégrales elliptiques (IE) et de fonctions elliptiques (FE) de Jacobi a été reprise en grand détail à la fin des années 2000 par Yuri Sachkov dans deux articles de fond [476] et [477] où il a étudié la façon dont ces courbes extrémales régulières (des minimiseurs qui minimisent localement partout et éventuellement globalement la fonctionnelle d'énergie élastique) peuvent perdre leur extrémalité locale ou globale. Rappelons que (cf. chapitre 12, section 6)

1. on appelle point *conjugué* un point où une courbe extrémale perd son optimalité locale (elle devient instable) ;
2. on appelle point du *cut locus* (cut point) un point où une courbe extrémale perd son optimalité globale.
3. On appelle aussi point de *Maxwell* un point où différentes extrémales donnant la même valeur de la fonctionnelle se rencontrent.

Nous allons exposer les idées générales de ces travaux en renvoyant le lecteur intéressé aux articles originaux pour les détails techniques. Nous allons voir comment la théorie du contrôle optimal permet de considérablement approfondir ce que nous déjà évoqué.

1. Les équations de Hamilton

L'espace des configurations est toujours $\mathbb{V} = SE(2) = \mathbb{R}^2 \times \mathbb{S}^1$ de coordonnées $q = (x, y, \theta)$. On cherche à optimiser certaines courbes Γ (régulières) dans \mathbb{V} . On utilise comme variable "temporelle" t la longueur d'arc s de la *projection* $\gamma(x, y)$

dans le plan de base \mathbb{R}^2 de la courbe cherchée Γ .¹ Les “vitesses” $\dot{\gamma}$ le long de γ sont donc de longueur 1. La courbe Γ est paramétrée de $t = t_0 = 0$ à $t = t_1$ et la contrainte d’isopérimétrie (longueur de γ constante) s’exprime par $t_1 = \ell = \text{cste}$. On prend la courbure $\kappa = \frac{d\theta}{ds}$ comme contrôle. On a donc

$$\begin{cases} \dot{x} = \cos(\theta) \\ \dot{y} = \sin(\theta) \\ \dot{\theta} = u \text{ (contrôle)} = \kappa \text{ (courbure)} \end{cases}$$

avec les conditions aux limites

$$q_0 = q(0) = (x_0, y_0, \theta_0), \quad q_1 = q(t_1) = (x_1, y_1, \theta_1) .$$

L’identité $\cos^2(\theta) + \sin^2(\theta) = 1$ équivaut à $\|\dot{\gamma}\| = 1$. Il s’agit de minimiser la fonctionnelle de Bernoulli du carré de la courbure

$$\mathcal{E}(\Gamma) = \frac{1}{2} \int_0^{t_1} u(t)^2 dt$$

avec des propriétés de régularité convenable sur $u(t)$ et $q(t)$.

Comme dans la section 7 du chapitre 5, on introduit la base $SE(2)$ -invariante de $T\mathbb{V}$ à savoir

$$X_1 = \cos(\theta) \partial_x + \sin(\theta) \partial_y, \quad X_2 = \partial_\theta$$

avec

$$\begin{aligned} [X_1, X_2] &= \sin(\theta) \partial_x - \cos(\theta) \partial_y = X_3, \\ [X_2, X_3] &= X_1, \\ [X_1, X_3] &= 0. \end{aligned}$$

On a

$$\dot{q} = (\dot{x}, \dot{y}, \dot{\theta}) = (\cos(\theta), \sin(\theta), u) = X_1 + uX_2 .$$

Ainsi, on se situe dans l’espace de configurations \mathbb{V} et on y paramétrise les courbes Γ par la longueur d’arc de leur projection γ sur le plan de base. Par construction, les équations impliquent que toute solution Γ est la relevée legendrienne de γ et est donc une intégrale de la distribution de contact $\ker(\omega)$ ($\omega = \cos(\theta) dy - \sin(\theta) dx$ étant la forme de contact). On prend alors la courbure κ de cette projection γ comme contrôle u et on cherche les extrémales dans \mathbb{V} qui satisfont la contrainte appropriée en les pensant comme les trajectoires d’un champ de vecteur $\dot{q} = X(q, u)$ sur \mathbb{V} . En variant u on peut obtenir énormément de courbes, par exemple en enchaînant des intervalles de t où $u = 0$ (segments de droites) avec des intervalles de t où u est constante $\neq 0$ (arcs de cercle). Mais la fonctionnelle \mathcal{E} dépend de t à travers u et sa

1. Ce point est important : s n’est pas l’élément de longueur d’arc de Γ . C’est pourquoi nous le notons t quand $t = s$ intervient comme paramétrage de Γ et pas seulement de γ .

minimisation va définir un contrôle $u(t)$ qui ensuite, par intégrations successives, va permettre de définir d'abord $\theta(t)$ puis $\{x(t), y(t)\}$.

Dans [476], Yuri Sachkov réanalyse en termes de théorie du contrôle optimal l'espace des *elasticae*. Étant donnée une famille de contrôles "admissibles" $u(t)$, on commence par évaluer les points $q_1 = q(t_1) = (x_1, y_1, \theta_1)$ atteignables à partir d'un point initial $q_0 = q(0) = (x_0, y_0, \theta_0)$ en un "temps" t_1 . Comme le système $\{X_1, X_2\}$ est "bracket generating" (condition de Hörmander), nous savons que les trajectoires peuvent aller dans toutes les directions de \mathbb{V} . En fait, on montre que tous les points $q = (x, y, \theta)$ tels que (x, y) soit intérieur au cercle Δ de centre (x_0, y_0) et de rayon t_1 avec θ quelconque sont atteignables, ce qui donne l'intérieur du tore plein $\Delta \times \mathbb{S}^1$ et que, en plus, évidemment, l'extrémité

$$\{x_0 + t_1 \cos(\theta_0), y_0 + t_1 \sin(\theta_0), \theta_0\}$$

du segment de droite de pente θ_0 est atteignable (*elastica* "droit" sans force exercée).

Ce dernier cas est singulier (c'est une extrémale "abnormale" au sens de la théorie du contrôle) et, comme il est parfaitement bien défini, peut être considéré comme résolu. Nous nous restreindrons donc aux cas des "vrais" *elasticae* où une force est exercée sur la lame élastique. On peut montrer que pour tous les points atteignables il existe un contrôle optimal eulérien régulier et même analytique que l'on peut obtenir à partir du Principe du Maximum de Pontryagin (PMP) exposé dans la section 5 du chapitre précédent (12). Sachkov passe donc de la version "lagrangienne" du problème des *elasticae* remontant à Euler à la version "hamiltonienne".

Remarque sur les notations. En accord avec une convention de notation assez habituelle en théorie du contrôle optimal, Sachkov note λ plutôt que p les covecteurs de $T^*\mathbb{V}$ et nous adopterons la même notation dans cette section pour que le lecteur puisse se référer facilement aux textes originaux. \square

On passe de l'espace de configurations \mathbb{V} à l'espace des phases $T^*\mathbb{V}$ avec la projection canonique $\pi : T^*\mathbb{V} \rightarrow \mathbb{V}$ et on considère des hamiltoniens $h : T^*\mathbb{V} \rightarrow \mathbb{R}$. Parmi ceux-ci, il y a ceux h_i associés aux champs X_i et définis par $h_i(\lambda, q) = \langle \lambda, X_i \rangle$, $\lambda \in T_q^*\mathbb{V}$. Si $\lambda = \{\lambda_x, \lambda_y, \lambda_\theta\}$ dans la base holonome $(dx, dy, d\theta)$ duale de $(\partial_x, \partial_y, \partial_\theta)$, autrement dit si

$$\lambda = \lambda_x dx + \lambda_y dy + \lambda_\theta d\theta,$$

on a

$$\begin{cases} h_1(\lambda, q) = \lambda_x \cos(\theta) + \lambda_y \sin(\theta) \\ h_2(\lambda, q) = \lambda_\theta \\ h_3(\lambda, q) = \lambda_x \sin(\theta) - \lambda_y \cos(\theta) \end{cases}$$

Les h_i sont les composantes de λ dans la base

$$\begin{cases} \omega_1 = \cos(\theta) dx + \sin(\theta) dy \\ \omega_2 = d\theta \\ \omega_3 = \sin(\theta) dx - \cos(\theta) dy \end{cases}$$

duale de (X_1, X_2, X_3) .²

De façon générale, comme nous l'avons déjà vu pour les géodésiques, si X est un champ de vecteurs sur \mathbb{V} , on lui associe canoniquement l'hamiltonien $h_X : T^*\mathbb{V} \rightarrow \mathbb{R}$ par

$$h_X(\lambda, q) = \langle \lambda, X \rangle, \quad \lambda \in T_q^*\mathbb{V}.$$

Si \vec{h}_X est le champ hamiltonien sur $T^*\mathbb{V}$ associé à h_X , il relève X (qui est un champ sur \mathbb{V}) dans le fibré cotangent $T^*\mathbb{V}$.

Les hamiltoniens parmi lesquels sélectionner l'hamiltonien optimal pour minimiser la fonctionnelle \mathcal{E} sont de la forme

$$h_u(\lambda, q) = \langle \lambda, \dot{q} \rangle + \mu u^2, \quad \lambda \in T_q^*\mathbb{V}$$

où μ permet de tenir compte de l'intégrande $\frac{1}{2}u^2$ de la fonctionnelle \mathcal{E} . Le cas $\mu = 0$ correspond aux extrémales "abnormales" et lorsque $\mu \neq 0$, on peut normaliser avec $\mu = -\frac{1}{2}$. D'où les hamiltoniens

$$h_u(\lambda, q) = \langle \lambda, \dot{q} \rangle - \frac{1}{2}u^2 = h_1(\lambda, q) + u h_2(\lambda, q) - \frac{1}{2}u^2$$

qui sont les transformées de Legendre de l'intégrande $\frac{1}{2}u^2$ de \mathcal{E} prise comme lagrangien. Conformément au PMP il s'agit d'optimiser le contrôle u au moyen de la condition

$$\frac{\partial h_u}{\partial u} = 0.$$

Pour un contrôle u donné, il est naturel d'écrire $h(\lambda, q, u)$ sous la forme $h_u(\lambda, q)$ avec u comme paramètre. Mais on peut aussi considérer $h(\lambda, q, u)$ comme paramétrée par (λ, q) et l'écrire $h_{(\lambda, q)}(u)$. C'est en fait ce que recouvre la condition $\frac{\partial h_u}{\partial u} = 0$ qui signifie

$$\frac{\partial h_{(\lambda, q)}(u)}{\partial u} = 0$$

pour tout λ . Cela signifie que $h(\lambda, q, u)$ peut être considérée comme une fonction définie sur le fibré cotangent $T^*\mathbb{V}$ et qu'au-dessus de chaque point $(\lambda, q) \in T^*\mathbb{V}$, on place les points critiques de $h_{(\lambda, q)}(u)$. Nous y reviendrons.

Comme $\frac{\partial h_u}{\partial u} = h_2 - u$, la condition de criticité $\frac{\partial h_u}{\partial u} = 0$ s'écrit $u = h_2$, i.e.

$$u(\lambda, q) = h_2(\lambda, q) = \lambda_\theta.$$

Il y a une seule solution \tilde{u} , un maximum unique, qui définit ainsi \tilde{u} comme fonction de (λ, q) . L'hamiltonien $H = \text{Sup}_u(h_u)$ engendrant les elasticæ par projection de ses

2. On peut aussi les noter λ_i .

trajectoires est par conséquent

$$\begin{aligned} H(\lambda, q) &= h_1(\lambda, q) + \frac{1}{2}h_2(\lambda, q)^2 \\ &= \lambda_x \cos(\theta) + \lambda_y \sin(\theta) + \frac{1}{2}\lambda_\theta^2. \end{aligned}$$

Remarque. La dérivée seconde $\frac{\partial^2 h_u}{\partial u^2} = -1$ est < 0 et donc $\tilde{u}(\lambda, q)$ est toujours un maximum non dégénéré de $h_{(\lambda, q)}(u)$. Cela sera utile plus bas. \square

On en tire immédiatement les équations de Hamilton (le champ hamiltonien \vec{H})

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial \lambda_x} = \cos(\theta) \\ \dot{y} = \frac{\partial H}{\partial \lambda_y} = \sin(\theta) \\ \dot{\theta} = \frac{\partial H}{\partial \lambda_\theta} = \lambda_\theta = h_2 = \text{courbure} \\ \dot{\lambda}_x = -\frac{\partial H}{\partial x} = 0 \\ \dot{\lambda}_y = -\frac{\partial H}{\partial y} = 0 \\ \dot{\lambda}_\theta = \dot{h}_2 = -\frac{\partial H}{\partial \theta} = \lambda_x \sin(\theta) - \lambda_y \cos(\theta) = h_3 \end{cases}$$

Par rapport à la structure de fibré $\pi : T^*\mathbb{V} \rightarrow \mathbb{V}$, les trois premières équations constituent un système “horizontal” et les trois dernières un système “vertical”. Les deux premières composantes “verticales” λ_x et λ_y sont constantes et on a le système différentiel des hamiltoniens h_i :

$$\dot{h}_1 = -h_2 h_3, \quad \dot{h}_2 = h_3, \quad \dot{h}_3 = h_1 h_2.$$

Ce système est analytique et ses extrémales et contrôles optimaux sont analytiques.

L’intégration du système, et donc tous les elasticæ de longueur t_1 , sont contenus dans la formule de l’application exponentielle

$$\text{Exp}_{t_1} : T_{q_0}^* \mathbb{V} \rightarrow \mathbb{V}, \quad \lambda_0 \mapsto \text{Exp}_{t_1}(\lambda_0) = \pi \circ e^{t_1 \vec{H}}(\lambda_0) = q(t_1)$$

qui consiste à partir de q_0 avec le covecteur λ_0 , suivre le champ hamiltonien \vec{H} pendant le “temps” t_1 puis à projeter la trajectoire d’abord sur \mathbb{V} puis ensuite sur le plan de base \mathbb{R}^2 . Exp_{t_1} est la section pour le temps-longueur t_1 de l’application exponentielle complète

$$\text{Exp} : T_{q_0}^* \mathbb{V} \times \mathbb{R}^+ \rightarrow \mathbb{V}, \quad \text{Exp}(\lambda_0, t_1) = \underset{t_1}{\text{Exp}}(\lambda_0).$$

La structure des h_i permet facilement de comprendre la classification des elasticæ. En effet

$$\dot{h}_1 h_1 = -h_2 h_3 h_1, \quad \dot{h}_3 h_3 = h_1 h_2 h_3$$

et donc

$$\dot{h}_1 h_1 + \dot{h}_3 h_3 = 0,$$

ce qui signifie que

$$h_1^2 + h_3^2 = \text{cste} = r^2 \geq 0$$

est une intégrale première. Si l'on ramène la trajectoire $\lambda(t)$ dans $T_{q_0}^*\mathbb{V}$ (ce qui est possible grâce à la trivialité du fibré $T^*\mathbb{V}$) elle est donc un cercle, autrement dit, elle s'enroule sur le cylindre constitué des cercles de rayon r dans les espaces cotangents $T_{q(t)}^*\mathbb{V}$.

2. Lien avec le pendule

On retrouve facilement le lien avec la mécanique d'un pendule que nous avons commenté à la section 10 du chapitre 9, qui remonte à Kirchoff, a été approfondi par Max Born et a été souligné dans notre contexte par Andrei Agrachev.³ Il suffit d'introduire des coordonnées polaires dans les plans (h_1, h_3) des $T_q^*\mathbb{V}$ et de poser

$$h_1 = r \cos(\alpha), \quad h_3 = r \sin(\alpha) .$$

Il est trivial de vérifier que l'on a alors, en plus évidemment de $r = \text{cste}$ et $\dot{h}_2 = h_3 = r \sin(\alpha)$, la relation $\dot{\alpha} = h_2$. On en tire immédiatement une équation qui est presque celle du pendule

$$\ddot{\alpha} = \dot{h}_2 = r \sin(\alpha) .$$

On pourrait donc utiliser les coordonnées *cylindriques* (r, α, h_2) dans les fibres $T_q^*\mathbb{V}$ de $T^*\mathbb{V}$, le système dynamique "vertical" devenant

$$\left\{ \dot{r} = 0, \quad \dot{\alpha} = h_2, \quad \dot{h}_2 = \ddot{\alpha} = r \sin(\alpha) \right\} .$$

Mais il est plus commode de se ramener à une vraie équation du pendule. Pour cela on pose

$$\beta = \alpha + \pi$$

(et donc $h_1 = -r \cos(\beta)$ et $h_3 = -r \sin(\beta)$) et on renomme $h_2 = c$. Dans les coordonnées $v = (r, \beta, c)$, le système "vertical" devient

$$\left\{ \dot{r} = 0, \quad \dot{\beta} = c, \quad \dot{c} = \ddot{\beta} = -r \sin(\beta) \right\}$$

qui est le système d'un pendule avec $r = \frac{g}{l}$.⁴ L'énergie du pendule est

$$E = \frac{1}{2}c^2 - r \cos(\beta)$$

3. Rappelons que pour les elasticæ paramétrés par la longueur d'arc la vitesse est unitaire (la loi du mouvement est simple) mais la tangente et la courbure suivent des lois compliquées, alors que pour le pendule la géométrie est simple (cercle) mais la loi du mouvement est compliquée.

4. Le cas $r = 0$ est donc celui d'un pendule immobile ou tournant indéfiniment en absence de gravité.

(énergie cinétique + énergie potentielle) c'est-à-dire la valeur de l'hamiltonien

$$H = h_1 + \frac{1}{2}h_2^2$$

puisque $c = h_2$ et $h_1 = -r \cos(\beta)$.

Remarque. Par rapport à ce que nous avons présenté à la section 10 du chapitre 9, il y a ici quelques changements de notation. Mettons un tilde aux anciennes notations. Alors β correspond à $\tilde{\theta}$, $\tilde{E} = 2k^2 = E + 1$, \tilde{u} correspond à φ (temps de parcours) et donc $\tilde{\varphi} = \text{am}(\varphi)$ (amplitude). \square

Dans les fibres, on a le cylindre de génératrices parallèles à h_2 au-dessus du cercle $h_1^2 + h_3^2 = r^2$ du plan (h_1, h_3) et le cylindre $h_1 + \frac{1}{2}h_2^2 = H$ de génératrices parallèles à h_3 au-dessus de la parabole $h_1 + \frac{1}{2}h_2^2$ du plan (h_1, h_2) . Les relations entre les deux intégrales premières H et r déterminent les types de trajectoire.

1. D'abord, vu sa formule, $H = \frac{1}{2}c^2 - r \cos(\beta) \geq -r$.
2. Pour $H = -r$, on a $c = 0$ et $\cos(\beta) = 1$, i.e. $\beta = 0$ et donc $\alpha = \pi$. C'est la position stable du pendule au repos.
3. Pour $H \in]-r, r[$, le pendule oscille entre les angles $\pm\beta_{\max}$ avec une vitesse $c \in \left[-\sqrt{2(E+r)}, \sqrt{2(E+r)}\right]$.
4. Pour $H = r$, $c = 0$ et $\cos(\beta) = -1$, i.e. $\beta = \pi$ et donc $\alpha = 0$, le pendule est dans sa position d'équilibre instable.
5. Pour $H > r$, le pendule accomplit des tours complets dont le sens dépend du signe de h_2 .

3. Coordonnées de Jacobi

Pour pouvoir analyser en détail les différentes classes d'elasticæ, il est pertinent d'utiliser leur représentation en termes de fonctions elliptiques, ce que fait Yuri Sachkov. On suppose $r = 1$ pour simplifier, donc

$$E = \frac{1}{2}c^2 - \cos(\beta) ,$$

et on considère le portrait de phase (les trajectoires) du pendule dans le cylindre $\mathbb{S}_{(\beta)}^1 \times \mathbb{R}_{(c)}$ des (β, c) . Les trajectoires (β_t, c_t) sont déterminées par leur intersection $(0, c_0)$ avec l'axe des c pour $\beta = 0$. Soit Σ la séparatrice entre les trajectoires oscillantes (domaine ouvert noté C_1 par Sachkov) et les trajectoires tournantes (domaine ouvert noté C_2^+ au-dessus de Σ et C_2^- au-dessous). Comme l'illustre la figure 1, le portrait de phase est *stratifié*.

- Il y a une première strate ouverte de dimension 2 (C_1) qui est l'intérieur de Σ moins l'origine $C_4 = (0, 0)$.

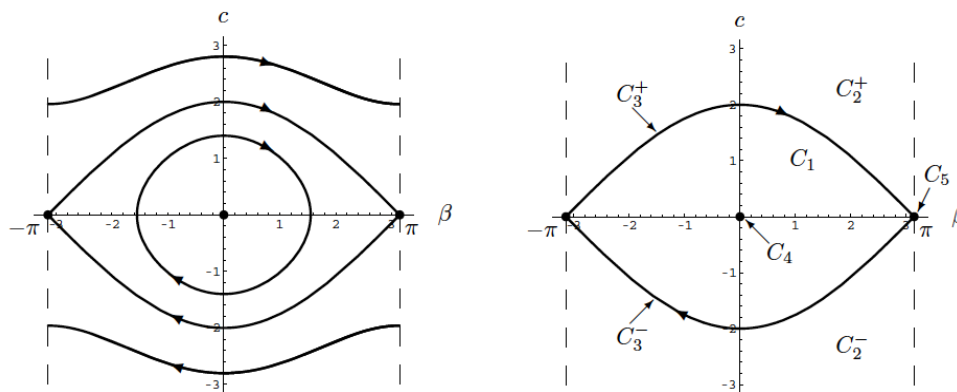


FIGURE 1. À gauche, le portrait de phase C du pendule dans le plan des (β, c) . À droite, sa stratification. (D'après [476], version ArXiv, p. 28).

- Il y a ensuite les 2 strates ouvertes de dimension 2 (C_2^+ et C_2^-) à l'extérieur de Σ .
- Il y a Σ elle-même moins son point singulier $C_5 = (0, \pi)$ qui donne deux strates de dimension 1 (C_3^+ et C_3^-)
- Il y a enfin les deux strates de dimension 0 $C_4 = (0, 0)$ et $C_5 = (0, \pi)$.⁵

Les symétries sont évidentes :

$$\begin{cases} \sigma^1 : (\beta, c) \mapsto (\beta, -c) \\ \sigma^2 : (\beta, c) \mapsto (-\beta, c) \\ \sigma^3 = \sigma^1 \circ \sigma^2 = \sigma^2 \circ \sigma^1 : (\beta, c) \mapsto (-\beta, -c). \end{cases}$$

Les deux premières, σ^1 et σ^2 , renversent le sens du temps alors que la troisième le conserve : si (β_s, c_s) est un arc de trajectoire avec $s \in [0, t]$,

$$\begin{cases} \sigma^1((\beta_s, c_s)) = (\beta_{t-s}, -c_{t-s}) \\ \sigma^2((\beta_s, c_s)) = (-\beta_{t-s}, c_{t-s}) \\ \sigma^3((\beta_s, c_s)) = (-\beta_s, -c_s) \end{cases}$$

On connaît les périodes T du pendule en termes de l'intégrale elliptique complète de première espèce

$$K(k) = \int_0^{\frac{\pi}{2}} \frac{dt}{\sqrt{1 - k^2 \sin^2 t}}$$

où k est le module (cf. la section 7 du chapitre 9).

5. Nous reviendrons plus en détail sur ce portrait de phase dans la section 4.3 du prochain chapitre (14).

– Si $(\beta, c) \in C_1$ (pendule oscillant), alors $c \in (0, 2)$,⁶ $E \in (-1, 1)$,

$$k_1 = \sqrt{\frac{E+1}{2}} = \sqrt{\frac{c^2}{4} + \sin^2\left(\frac{\beta}{2}\right)} \in (0, 1)$$

($k_1 = \frac{c}{2}$ pour $\beta = 0$) et $T = 4K(k_1)$.

– Si $(\beta, c) \in C_2^+$ (pendule tournant), alors $c \in (2, \infty)$, $E > 1$,

$$k_2 = \sqrt{\frac{2}{E+1}} = \frac{1}{\sqrt{\frac{c^2}{4} + \sin^2\left(\frac{\beta}{2}\right)}} \in (0, 1)$$

($k_2 = \frac{2}{c}$ pour $\beta = 0$) et $T = 2K(k_2)k_2$.

– À la limite $E = 1$, $T = \infty$.

Sachkov note $\widehat{C} = C_1 \cup C_2^+ \cup C_3^+$. L'application envoyant le point $(c_0, t) \in \mathbb{R}_{(c)}^+ \times \mathbb{R}_{(t)}$ sur le point (β_t, c_t) de la trajectoire déterminée par $(\beta_0 = 0, c_0)$ est une application analytique. Si on la quotiente par les périodes, elle devient un *difféomorphisme analytique* Φ dont on peut considérer l'inverse $F = \Phi^{-1}$.⁷ Par définition,

$$F(\beta, c) = (c_0, \varphi)$$

qui est la durée rétrograde du pendule nécessaire pour que (β, c) retourne le long de sa trajectoire en $(\beta_0 = 0, c_0)$. $\varphi(\beta, c)$ est analytique et définie modulo les périodes. Il faut insister sur le fait que ce *temps* “pendulaire” correspond à la *longueur* de l'elastica. Comme nous l'avons vu dans la section 7 du chapitre 9, un elastica déploie en quelque sorte *spatialement* la *chronométrie* du pendule, l'angle β devenant l'angle θ de la tangente. Les périodes d'oscillations ou de tours complets du pendule correspondent donc à des “périodes” géométriques de l'elastica, c'est-à-dire à la *reproduction d'un même motif*.

Cela permet à Sachkov d'utiliser les coordonnées spéciales (φ, k) que sont les coordonnées elliptiques de Jacobi. Il le justifie ainsi joliment dans [476] (version ArXiv, p. 29) :

“Elliptic coordinates lift the veil of complexity over the problems governed by the pendulum equation and open their solution to our eyes.”

6. Si $r = 1$ et $\beta = 0$ (intersection de la trajectoire avec l'axe des c), $H = \frac{1}{2}c^2 - 1$ et comme C_3^+ correspond à $H = r = 1$, la valeur limite de c est $\frac{1}{2}c^2 = 2$, i.e. $c = 2$ (car $c > 0$ sur C_3^+).

7. Démontrer que Φ est un difféomorphisme est facile. On considère son Jacobien J et le déterminant D de J . Par définition, $J = \begin{pmatrix} \frac{\partial \beta_t}{\partial c} & \frac{\partial \beta_t}{\partial t} \\ \frac{\partial c_t}{\partial c} & \frac{\partial c_t}{\partial t} \end{pmatrix}$. On sait que $\frac{\partial \beta_t}{\partial t} = \dot{\beta}_t = c_t$ et $\frac{\partial c_t}{\partial t} = \dot{c}_t = -\sin(\beta_t)$. Pour $\frac{\partial \beta_t}{\partial c} = \zeta_t$ et $\frac{\partial c_t}{\partial c} = \eta_t$, on dérive par rapport à t et on commute les dérivées. D'où $\dot{\zeta}_t = \eta_t$ et $\dot{\eta}_t = -\cos(\beta_t)$. Cela permet de montrer immédiatement que, si $D_t = -\zeta_t \sin(\beta_t) - c_t \eta_t$ est le déterminant de J , on a $\dot{D} \equiv 0$ et donc $D_t = D_0 = -c_0$. Or $c_0 \neq 0$ dans le domaine de définition de Φ .

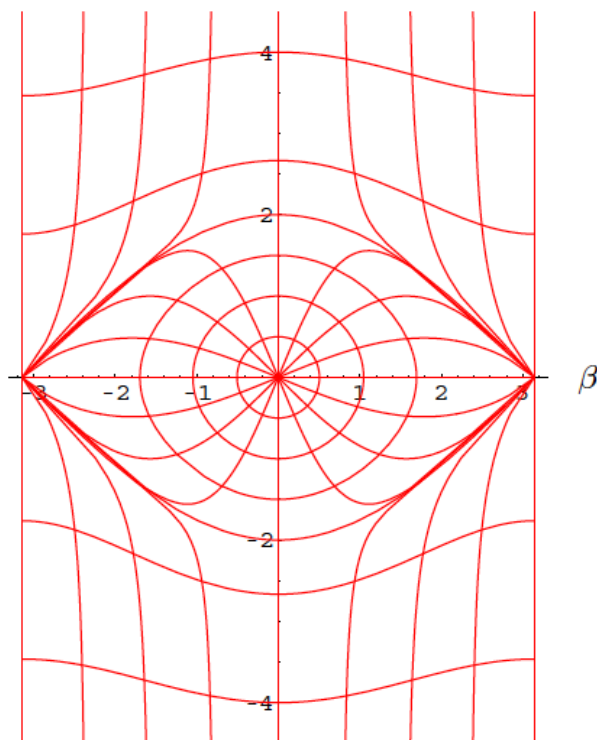


FIGURE 2. Les coordonnées elliptiques (φ, k) introduites par Sachkov. D'après [476] (version ArXiv, p. 29).

Dans C_1 , on a alors, par définition des fonctions elliptiques de Jacobi,

$$\sin\left(\frac{\beta}{2}\right) = k_1 \operatorname{sn}(\varphi, k_1), \quad \frac{c}{2} = k_1 \operatorname{cn}(\varphi, k_1), \quad \cos\left(\frac{\beta}{2}\right) = \operatorname{dn}(\varphi, k_1),$$

$$k_1 = \sqrt{\frac{E+1}{2}} = \sqrt{\frac{c^2}{4} + \sin^2\left(\frac{\beta}{2}\right)}$$

$$T(\varphi) = 4K(k_1)$$

Pour $(\beta_0 = 0, c_0)$, $\varphi_0 = 0$ et, comme $\operatorname{cn}(0) = 1$, $\operatorname{sn}(0) = 0$, $\operatorname{dn}(0) = 1$, on vérifie que les égalités ci-dessus sont bien satisfaites avec $\frac{c_0}{2} = k_1$.

Pour le cas C_2^\pm , on peut prolonger analytiquement les formules du cas C_1 . Comme k_1 devient > 1 , on passe dans le domaine complexe. Pour revenir dans le domaine

réel on utilise $k_2 = \frac{1}{k_1}$ et les formules

$$\begin{aligned} \operatorname{sn} \left(u, \frac{1}{k} \right) &= k \operatorname{sn} \left(\frac{u}{k}, k \right), \\ \operatorname{cn} \left(u, \frac{1}{k} \right) &= \operatorname{dn} \left(\frac{u}{k}, k \right), \\ \operatorname{dn} \left(u, \frac{1}{k} \right) &= \operatorname{cn} \left(\frac{u}{k}, k \right), \\ E \left(u, \frac{1}{k} \right) &= \frac{1}{k} E \left(\frac{u}{k}, k \right) - \frac{1-k^2}{k^2} u. \end{aligned}$$

D'où

$$\begin{aligned} \sin \left(\frac{\beta}{2} \right) &= \pm \operatorname{sn} \left(\frac{\varphi}{k_2}, k_2 \right), \quad \frac{c}{2} = \pm \frac{1}{k_2} \operatorname{dn} \left(\frac{\varphi}{k_2}, k_2 \right), \quad \cos \left(\frac{\beta}{2} \right) = \operatorname{cn} \left(\frac{\varphi}{k_2}, k_2 \right), \\ k_2 &= \sqrt{\frac{2}{E+1}} = \frac{1}{\sqrt{\frac{c^2}{4} + \sin^2 \left(\frac{\beta}{2} \right)}} \end{aligned}$$

$$T(\varphi) = 2K(k_2)k_2$$

Pour le cas limite $k = 1$ des strates C_3^\pm , nous retrouvons les formules déjà rencontrées à la section 10 du chapitre 9,

$$\begin{aligned} \sin \left(\frac{\beta}{2} \right) &= \pm \tanh(\varphi), \\ \frac{c}{2} &= \pm \frac{1}{\cosh(\varphi)}, \\ \cos \left(\frac{\beta}{2} \right) &= \frac{1}{\cosh(\varphi)}. \end{aligned}$$

4. Stratification des espaces cotangents

La stratification C_i du portrait de phase du pendule se transfère en une stratification N_i de l'espace cotangent $N = T_{q_0}^* \mathbb{V}$.

- N_1 correspond à C_1 , soit $r \neq 0$ et $E \in]-r, r[$,
- N_2^\pm correspondent à C_2^\pm , soit $r \neq 0$ et $E \in]r, \infty[$,
- N_3^\pm correspondent à C_3^\pm , soit $r \neq 0$ et $E = r$, $\beta \neq \pi$,
- N_4 correspond à C_4 , soit $r \neq 0$, $E = -r$, $\beta_t \equiv 0$, $c_t \equiv 0$,
- N_5 correspond à C_5 , soit $r \neq 0$, $E = r$, $\beta = \pi$, $c_t \equiv 0$.

Toutes ces strates peuvent se ramener aux cas $r = 1$ par dilatation. On peut ajouter pour être complet les deux cas du pendule sans pesanteur :

- N_6^\pm correspond à $r = 0$, $c \neq 0$, $\beta_t = ct + \beta$, $c_t \equiv c$ (pendule sans pesanteur tournant uniformément)
- N_7 correspond à $r = c = 0$ (pendule sans pesanteur immobile).

5. Intégration des équations pour $k \in [0, 1]$ (pendule oscillant)

Dans les coordonnées elliptiques (k, φ) , le système “vertical” devient trivial à intégrer. Il se réduit en effet à $\dot{k} = 0$ (k est constant puisqu’il reformule l’énergie E) et $\dot{\varphi} = 1$ (puisque φ est un temps de parcours). La formule de base est donc

$$\varphi_t = \varphi_0 + t .$$

À partir de là, on utilise les formules du système “horizontal”, en supposant par exemple $q_0 = 0$. En termes de $\frac{\theta}{2}$, elles s’écrivent :

$$\begin{cases} \dot{x} = \cos(\theta) = 2 \cos^2\left(\frac{\theta}{2}\right) - 1, & x_0 = 0 \\ \dot{y} = \sin(\theta) = 2 \sin\left(\frac{\theta}{2}\right) \cos\left(\frac{\theta}{2}\right), & y_0 = 0 \\ \dot{\theta} = c = \dot{\beta}, & \theta_0 = 0 \end{cases}$$

De la dernière équation on tire

$$\theta_t = \beta_t - \beta,$$

ce qui permet d’intégrer les équations dans les différentes strates en utilisant les formules donnant $\sin\left(\frac{\theta}{2}\right)$ et $\cos\left(\frac{\theta}{2}\right)$. Pour $t = 0$, puisque $\theta_0 = 0$, la condition initiale est $\theta_0 = \beta_0 - \beta = 0$ et donc $\beta_0 = \beta$. Dans $T_{q_0}\mathbb{V}$, le mouvement du pendule commence avec l’angle β et un début de temps de parcours φ .

Par exemple dans la strate N_1 , on a pour $r = 1$, nous l’avons vu,

$$\sin\left(\frac{\beta}{2}\right) = k \operatorname{sn}(\varphi), \quad \cos\left(\frac{\beta}{2}\right) = \operatorname{dn}(\varphi), \quad \frac{c}{2} = k \operatorname{cn}(\varphi) .$$

Par conséquent,

$$\begin{aligned} \sin\left(\frac{\theta_t}{2}\right) &= \sin\left(\frac{\beta_t}{2} - \frac{\beta}{2}\right) = \sin\left(\frac{\beta_t}{2}\right) \cos\left(\frac{\beta}{2}\right) - \cos\left(\frac{\beta_t}{2}\right) \sin\left(\frac{\beta}{2}\right) \\ &= k \operatorname{sn}(\varphi_t) \operatorname{dn}(\varphi) - k \operatorname{sn}(\varphi) \operatorname{dn}(\varphi_t) \\ \cos\left(\frac{\theta_t}{2}\right) &= \operatorname{dn}(\varphi_t) \operatorname{dn}(\varphi) + k^2 \operatorname{sn}(\varphi_t) \operatorname{sn}(\varphi) . \end{aligned}$$

On vérifie la cohérence des formules pour $t = 0$ avec $\varphi_0 = \varphi$: c’est trivial pour le sin et pour le cos on a bien $\operatorname{dn}^2(\varphi) + k^2 \operatorname{sn}^2(\varphi) = 1$. On en tire les formules pour \dot{x} et \dot{y} et donc pour x et y en intégrant des carrés et des produits de fonctions elliptiques de Jacobi. Les calculs deviennent vite un peu compliqués mais ont l’immense avantage

de donner des formules *explicités exactes* des elasticæ. Par exemple, pour x dans la strate N_1 ($r = 1$), on a

$$\dot{x} = 2 \cos^2 \left(\frac{\theta}{2} \right) - 1 = 2 \left(\operatorname{dn}(\varphi_t) \operatorname{dn}(\varphi) + k^2 \operatorname{sn}(\varphi_t) \operatorname{sn}(\varphi) \right)^2 - 1 .$$

L'intégration fera donc intervenir, φ étant constante, φ_t étant variable avec $d\varphi = dt$ les intégrales $\int_{\varphi}^{\varphi_t} \operatorname{dn}^2(u) du$, $\int_{\varphi}^{\varphi_t} \operatorname{sn}^2(u) du$ et $\int_{\varphi}^{\varphi_t} \operatorname{dn}(u) \operatorname{sn}(u) du$ avec des coefficients. Mais on connaît ces intégrales. En effet, si l'on note $\mathbb{E}(u) = E(\operatorname{am}(u), k)$, on a

$$\begin{aligned} \int_{\varphi}^{\varphi_t} \operatorname{dn}^2(u) du &= \mathbb{E}(\varphi_t) - \mathbb{E}(\varphi) \\ \int_{\varphi}^{\varphi_t} \operatorname{sn}^2(u) du &= \frac{1}{k^2} (\varphi_t - \varphi - (\mathbb{E}(\varphi_t) - \mathbb{E}(\varphi))) = \frac{1}{k^2} (t - (\mathbb{E}(\varphi_t) - \mathbb{E}(\varphi))) \\ \int_{\varphi}^{\varphi_t} \operatorname{dn}(u) \operatorname{sn}(u) du &= (1 - \operatorname{cn}(\varphi_t)) - (1 - \operatorname{cn}(\varphi)) = \operatorname{cn}(\varphi) - \operatorname{cn}(\varphi_t) . \end{aligned}$$

D'où les belles formules retrouvées par Sachkov

$$\begin{aligned} x(t) &= 2 \operatorname{dn}^2(\varphi) (\mathbb{E}(\varphi_t) - \mathbb{E}(\varphi)) + 2k^2 \operatorname{sn}^2(\varphi) (t - (\mathbb{E}(\varphi_t) - \mathbb{E}(\varphi))) \\ &\quad + 4k^2 \operatorname{dn}(\varphi) \operatorname{sn}(\varphi) (\operatorname{cn}(\varphi) - \operatorname{cn}(\varphi_t)) - t , \\ y(t) &= 2k (2 \operatorname{dn}^2(\varphi) - 1) (\operatorname{cn}(\varphi) - \operatorname{cn}(\varphi_t)) \\ &\quad - 2k \operatorname{dn}(\varphi) \operatorname{sn}(\varphi) (2(\mathbb{E}(\varphi_t) - \mathbb{E}(\varphi)) - t) . \end{aligned}$$

On vérifie que pour $t = 0$, on obtient bien $x_0 = y_0 = 0$.

On retrouve alors facilement la classification d'Euler exposée à la section 10 du chapitre 9.

1. La courbure $\dot{\theta} = \kappa$ s'annule lorsque $\dot{\beta} = c = 0$, ce qui n'est possible que pour le pendule oscillant (strate C_1 et N_1). Lorsque $\dot{\beta} = 0$, $\beta = \pm\beta_{\max}$. On obtient dans ce cas des elasticæ avec points d'inflexions (dits elasticæ inflexionnels).
 - Lorsque $E \in (-r, 0)$, i.e. $k \in \left(0, \frac{1}{\sqrt{2}}\right)$, les elasticæ n'ont pas de tangentes "verticales" (i.e. orthogonales à l'axe le long duquel se déploie l'elastica).
 - Pour $E = 0$, i.e. $k = \frac{1}{\sqrt{2}}$, les inflexions deviennent verticales : c'est le cas de l'elastica rectangulaire.
 - Pour $E \in (0, E_0)$, i.e. $k \in \left(\frac{1}{\sqrt{2}}, k_0\right)$ où k_0 est la valeur critique solution de $2E(k) - K(k) = 0$, $k_0 \sim 0.908909$, on retrouve les elasticæ "en Ω ".
 - Pour $E = E_0$, $k = k_0$ on retrouve la lemniscate.
 - Pour $k \in (k_0, 1)$, $E \in (E_0, r)$ on retrouve les elasticæ avec des boucles de sens de rotation alternés à cause des inflexions.

- Lorsque $E = r$ et $k = 1$ avec $\theta - \beta \neq \pi$, le pendule approche indéfiniment de son équilibre instable $\theta - \beta \neq \pi$ et l'on obtient l'elastica à une boucle avec une asymptote. Si $\theta - \beta = \pi$, le pendule est sur son équilibre instable et l'elastica est une droite.
2. Ensuite, pour $E > r$ le pendule est tournant et l'on obtient les elasticæ à boucles sans inflexions et donc à sens de rotation constant (le signe de la courbure est constant).

On a comme cas spéciaux l'extrémale abnormale de l'elastica droit : $\theta_t \equiv 0$, $x_t = ct$, $y_t \equiv 0$ et de l'elastica circulaire (pendule tournant uniformément sans pesanteur) : $\theta_t = ct$, $x_t = \frac{\sin(ct)}{c}$, $y_t = \frac{1 - \cos(ct)}{c}$ (cercle de centre $(\frac{1}{c}, 0)$ et de rayon $\frac{1}{c}$).

6. Quelques calculs dans le cas $\varphi_0 = 0$ ($k \in [0, 1]$)

6.1. Géométrie générale

Dans les calculs précédents, les conditions initiales sont $q_0 = 0$, $\beta_0 = \beta$ et $\varphi_0 = \varphi$. Comme on veut pouvoir considérer des arcs d'elastica quelconques, il faut utiliser un système d'axes tel que $\theta_0 = 0$ à l'origine, l'elastica se déployant le long d'un axe défini par β . La complexité des formules générales vient de là. Pour étudier plus en détail la géométrie des elasticæ, il est commode de simplifier ces formules et de prendre $\beta_0 = \varphi_0 = 0$. L'elastica se déploie alors le long de l'axe des x avec $\varphi_t = t$. Cela revient à partir de la position basse du pendule $\beta_0 = 0$, $c_0 = \beta_{t,\max} = 2k$.

Les équations simplifiées sont alors effectivement assez simples :

$$\begin{aligned} \sin\left(\frac{\theta_t}{2}\right) &= k \operatorname{sn}(\varphi_t), \quad \cos\left(\frac{\theta_t}{2}\right) = \operatorname{dn}(\varphi_t) \\ \dot{x}(t) &= 2 \operatorname{dn}^2(\varphi_t) - 1, \quad \dot{y}(t) = 2k \operatorname{sn}(\varphi_t) \operatorname{dn}(\varphi_t) \\ x(t) &= 2\mathbb{E}(\varphi_t) - t, \quad y(t) = 2k(1 - \operatorname{cn}(\varphi_t)) \end{aligned}$$

1. On voit que, puisque $\operatorname{cn}(\varphi_t) = \operatorname{cn}(t)$, $y(t)$ est périodique de période $4K$. C'est une courbe oscillante autour de la droite $y = 2k$. $y(t)$ est minimal et s'annule pour $\varphi_t = 0 \pmod{4K}$, il est maximal de valeur $4k$ pour $\varphi_t = 2K \pmod{4K}$ et vaut $2k$ pour $\varphi_t = K \pmod{2K}$.
2. L'elastica a des tangentes horizontales $\dot{y}(t) = 0$ aux maxima et minima de $y(t)$ où $\operatorname{sn}(\varphi_t) = 0$ ($\operatorname{dn}(\varphi_t)$ est toujours $\neq 0$). Il démarre avec une tangente horizontale et $x(t), y(t) \geq 0$.
3. Il a des points d'inflexions pour $\dot{\theta}_t = \dot{\beta}_t = 2k \operatorname{cn}(\varphi_t) = 0$, i.e. pour $\varphi_t = K \pmod{2K}$.

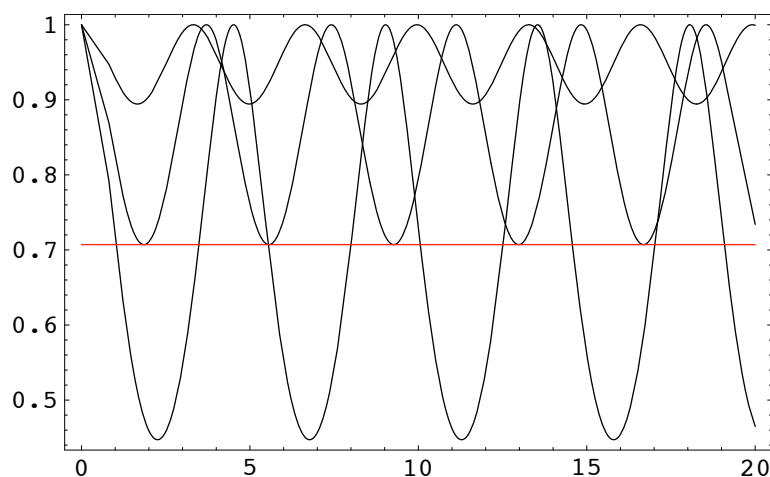


FIGURE 3. Graphes de $\text{dn}(u, k)$ pour $k^2 = 0.2, \frac{1}{2}$ et 0.8 . La droite $\text{dn}^2(u) = \frac{1}{2}$ est en rouge.

4. Il peut avoir des tangentes verticales $\dot{x}(t) = 0$ si l'équation $\text{dn}^2(\varphi_t) = \frac{1}{2}$ possède des solutions. Comme $\text{dn}^2(\varphi_t) = 1 - k^2 \text{sn}^2(\varphi_t)$, cela n'est possible que si $k^2 \geq \frac{1}{2}$ et pour $k^2 = \frac{1}{2}$ les solutions sont alors $\text{sn}^2(\varphi_t) = 1$, soit $\varphi_t = K \bmod(2K)$: ce sont les points d'inflexions (où $y(t) = 2k$) qui ont une tangente verticale. Cela correspond au cas de l'elastica "rectangulaire". La figure 3⁸ montre les graphes de $\text{dn}(u, k)$ pour $k^2 = 0.2, \frac{1}{2}$ et 0.8 ainsi que la droite $\text{dn}^2(u) = \frac{1}{2}$.

Pour illustrer la géométrie des elastica il faut comprendre l'allure du graphe de $x(t)$.

1. $x(t)$ démarre en 0 avec la pente $\dot{x}(t) = 2 \text{dn}^2(0) - 1 = 1$.
2. Pour k assez petit, $\dot{x}(t) > 0$ et la courbe est croissante.
3. Pour $k^2 = \frac{1}{2}$, i.e. $k \sim 0.707107$, $x(t)$ a des points d'inflexions pour $t = \varphi_t = K \bmod(2K)$. Cela donne les points à inflexions verticales de l'elastica rectangulaire.
4. Pour $k^2 > \frac{1}{2}$, les inflexions de $x(t)$ se dédoublent en un maximum et un minimum et l'équation $\text{dn}^2(\varphi_t) = \frac{1}{2}$ a une infinité de solutions qui correspondent à des paires de tangentes verticales encadrant les inflexions de l'elastica. Soit M_1 la première solution, i.e. $M_1 = \text{dn}^{-1}\left(\frac{1}{2}\right) \in [0, K]$. C'est le premier maximum de $x(t)$. La seconde racine $m_1 = \text{dn}^{-1}\left(\frac{1}{2}\right) \in [K, 2K]$ a pour valeur $2K - M_1$ et correspond au premier minimum de $x(t)$. Ensuite les maxima et minima s'enchaînent avec la période $2K$ de $\text{dn}(t)$. Les domaines $[M_n, m_n]$ où $x(t)$ est décroissant correspondent à des arcs rétrogrades de l'elastica.

8. Cf. la figure 8 de la section 7 du chapitre 9.

Le graphe de $x(t)$ est une courbe oscillante autour de la droite de pente

$$\frac{2E(k)}{K(k)} - 1 .$$

Pour le voir, rappelons que $\mathbb{E}(u, k) = E(\operatorname{am}(u), k)$ et que donc, puisque⁹ $K(k) = u\left(\frac{\pi}{2}\right)$,

$$\mathbb{E}(K, k) = E(\operatorname{am}(K), k) = E\left(\operatorname{am}\left(u\left(\frac{\pi}{2}\right)\right), k\right) = E\left(\frac{\pi}{2}, k\right) = E(k) .$$

On a donc, pour un k donné, $x(K) = 2E(k) - K(k)$. Le calcul de $x(nK)$ est une bonne occasion d'utiliser la formule générale d'addition pour \mathbb{E} :

$$\mathbb{E}(u + v) = \mathbb{E}(u) + \mathbb{E}(v) - k^2 \operatorname{sn}(u) \operatorname{sn}(v) \operatorname{sn}(u + v) .$$

On en tire immédiatement

$$\mathbb{E}(nK) = n\mathbb{E}(K) = nE(k)$$

car $\operatorname{sn}(2rK) = 0$ et $\mathbb{E}(K) = E(k)$, puis

$$\mathbb{E}(u + 2nK) = \mathbb{E}(u) + \mathbb{E}(2nK) = \mathbb{E}(u) + 2n\mathbb{E}(K) = \mathbb{E}(u) + 2nE(k) .$$

Pour $\mathbb{E}(u + K)$ et $\mathbb{E}(u + (2n + 1)K)$, la formule est un peu plus compliquée. Comme

$$\operatorname{sn}(u + K) = \frac{\operatorname{cn}(u) \operatorname{dn}(u)}{1 - k^2 \operatorname{sn}^2(u)} = \frac{\operatorname{cn}(u) \operatorname{dn}(u)}{\operatorname{dn}^2(u)} = \frac{\operatorname{cn}(u)}{\operatorname{dn}(u)} ,$$

on a¹⁰

$$\begin{aligned} \mathbb{E}(u + K) &= \mathbb{E}(u) + \mathbb{E}(K) - k^2 \frac{\operatorname{sn}(u) \operatorname{cn}(u)}{\operatorname{dn}(u)} \\ \mathbb{E}(u + (2n + 1)K) &= \mathbb{E}(u) + (2n + 1)\mathbb{E}(K) - k^2 \frac{\operatorname{sn}(u) \operatorname{cn}(u)}{\operatorname{dn}(u)} \\ &= \mathbb{E}(u) + (2n + 1)E(k) - k^2 \frac{\operatorname{sn}(u) \operatorname{cn}(u)}{\operatorname{dn}(u)} . \end{aligned}$$

On voit ainsi que, pour chaque demi-période $2K(k)$, $\mathbb{E}(\varphi_t)$ varie de $2E(k)$ et donc $x(t)$ varie de $2(2E(k) - K(k))$. C'est donc autour de la droite de pente $\frac{2E(k)}{K(k)} - 1$ qu'oscille $x(t)$.

Le signe de $2E(k) - K(k)$ détermine par conséquent le caractère en moyenne ascendant ou descendant du graphe de $x(t)$. La figure 4 représente le graphe de

9. Nous explicitons le paramètre k car les intégrales elliptiques complètes $E(k)$ et $K(k)$ dépendent de k .

10. Pour s'amuser avec les fonctions elliptiques, on peut vérifier qu'en itérant cette formule on réobtient bien $\mathbb{E}(u + 2K)$. On remarquera aussi que, comme $\operatorname{dn}(u)$, le produit $\operatorname{sn}(u) \operatorname{cn}(u)$ est de période $2K$ (et pas seulement $4K$) puisque $\operatorname{sn}(u)$ et $\operatorname{cn}(u)$ changent tous les deux de signe pour $u + 2K$.

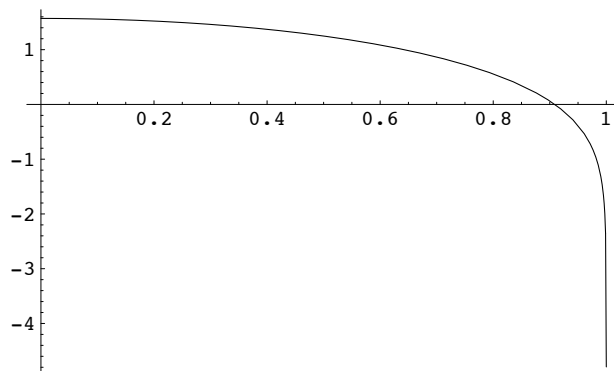


FIGURE 4. Graphe de la fonction $2E(k) - K(k)$ pour $k \in [0, 1]$. Elle s'annule pour $k_0 \sim 0.908909$ et admet $k = 1$ comme asymptote.

cette fonction pour $k \in [0, 1]$. On voit qu'elle part de $\frac{\pi}{2}$ en étant assez faiblement décroissante puis s'annule pour $k_0 \sim 0.908909$ et se met à décroître de plus en plus vite avec $k = 1$ comme asymptote. Ce comportement est dominé par $K(k)$. En effet $E(k)$ varie de $\frac{\pi}{2}$ à 1 et c'est $K(k)$ qui diverge vers $+\infty$ lorsque $k \rightarrow 1^-$.

Pour $k < k_0$, $2E(k) - K(k) > 0$ et, à chaque demi-période $2K$, l'elastica progresse vers les x croissants même s'il y a des arcs rétrogrades.

Pour $k = k_0$, $x(nK) = n(2E(k) - K(k)) = 0$ et l'elastica repasse par le point $(0, 2k_0)$ à chaque quart de période K . C'est le cas de la lemniscate.

Pour $k > k_0$, $2E(k) - K(k) < 0$ et l'elastica régresse vers les x décroissants à chaque demi-période $2K$.

6.2. Catastrophes de conflits

A partir de $k = \frac{1}{\sqrt{2}}$, et tant que $k < k_0$, $2E(k) - K(k)$ reste > 0 , il peut donc y avoir des valeurs critiques de k pour lesquelles l'équation $x(M_1) = x(m_n)$ possède une solution pour un certain entier $n \geq 2$. Au fur et à mesure que k augmente, n augmente. On obtient ainsi une suite d'elasticae "critiques" avec des propriétés particulières. Celles-ci sont structurellement instables car elles n'existent que pour les valeurs critiques de k . En théories des singularités, ces situations s'appellent des "catastrophes de conflits" au sens de Thom.¹¹

6.3. Exemple $k = 0.5$

Pour $k = 0$, l'elastica est l'elastica droit, i.e. l'axe des x . Pour k assez petit, l'elastica commence à onduler avec $y \in [0, 4k]$. Il a des points d'inflexions, des minima et des maxima mais pas de tangentes verticales. Les figures 5 et 6 montrent

11. Pour une introduction à la théorie des singularités cf. [412].

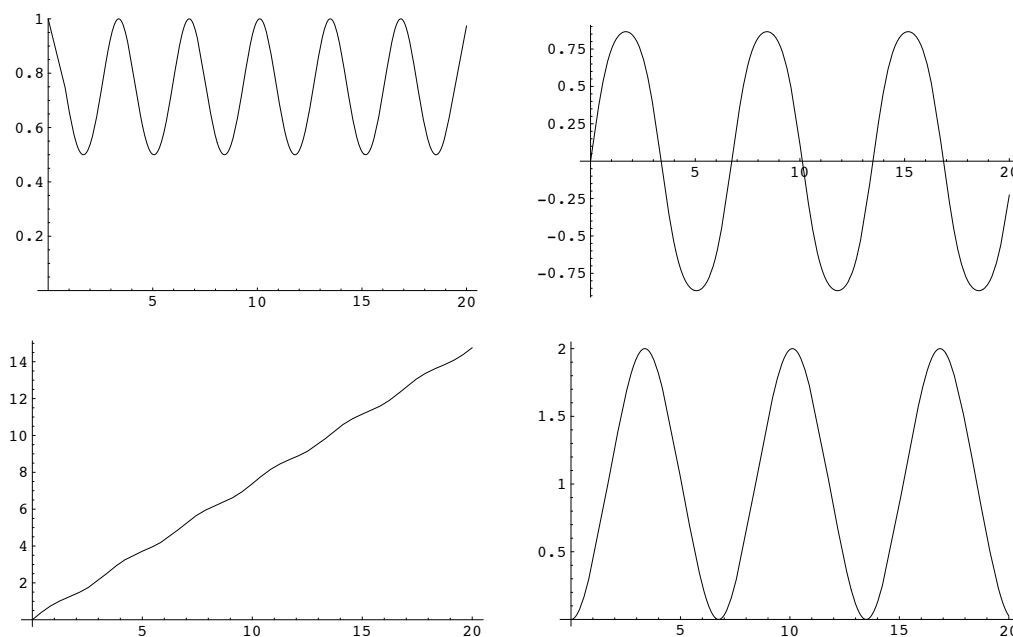


FIGURE 5. Les coordonnées $x(t)$ et $y(t)$ de l'elastica pour $k = 0.5$. En haut à gauche $x'(t)$. En bas à gauche $x(t)$. En haut à droite $y'(t)$. En bas à droite $y(t)$.

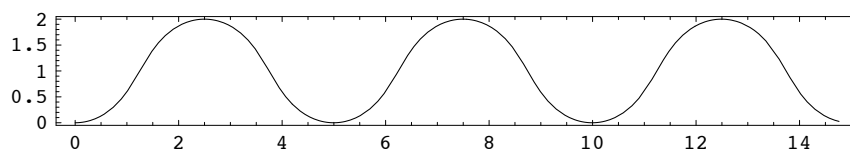


FIGURE 6. L'elastica pour $k = 0.5$.

l'exemple $k = 0.5$. On y voit bien les oscillations de $y(t)$ et la croissance monotone de $x(t)$ ($x'(t)$ ne s'annule pas).

6.4. Exemple $k \sim 0.707107$ (elastica rectangulaire)

Comme nous l'avons vu, pour $k \sim 0.707107$ (elastica rectangulaire), les inflexions deviennent verticales. On voit les inflexions à tangentes horizontales de $x(t)$ apparaître là où $x'(t) = 0$ et les extrema de $y'(t)$ (i.e. $y''(t) = 0$) s'aplatir, ce qui signifie qu'ils deviennent *dégénérés* et que $y(t)$ devient plus rectiligne au voisinage de ses points d'inflexions : la courbe est moins sinusoïdale et plus "en dent de scie". Cf. les figures 7 et 8.

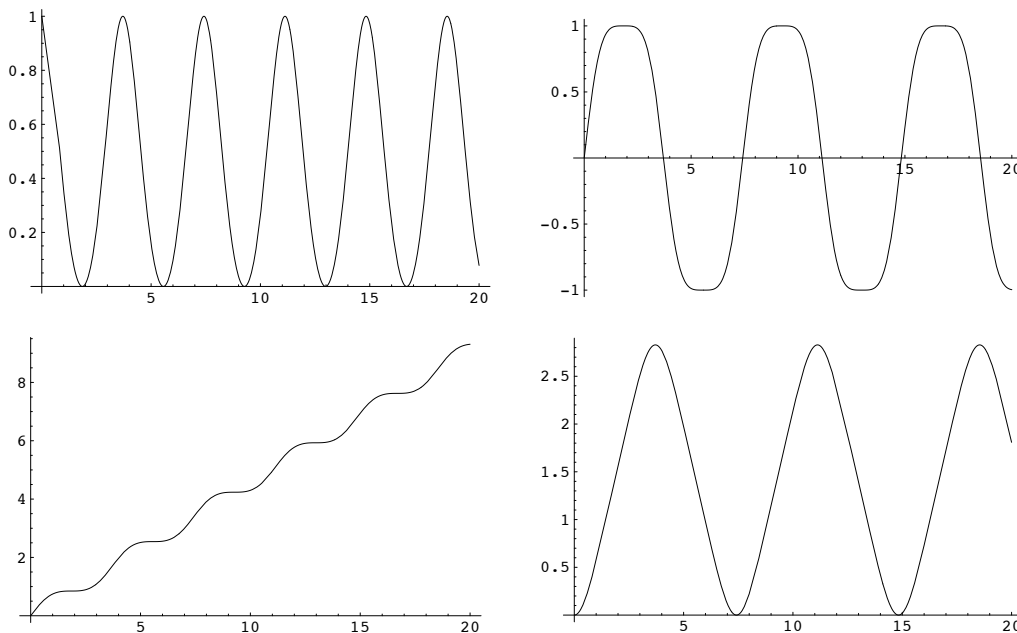


FIGURE 7. Les coordonnées $x(t)$ et $y(t)$ de l'elastica pour $k = 0.707$. En haut à gauche $x'(t)$. En bas à gauche $x(t)$. En haut à droite $y'(t)$. En bas à droite $y(t)$.

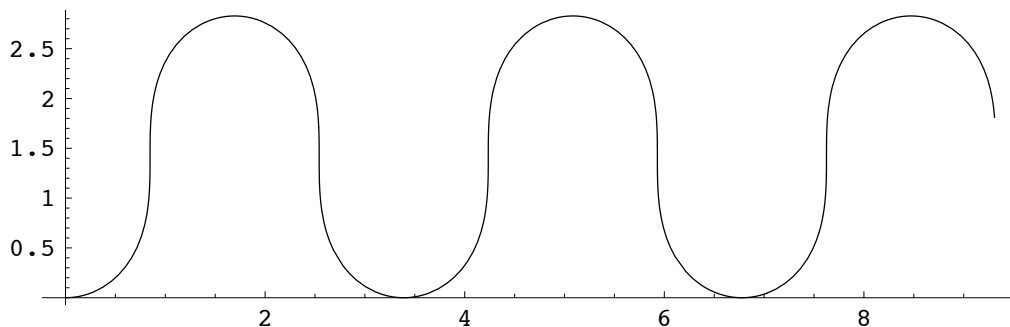


FIGURE 8. L'elastica pour $k = 0.707$.

6.5. Exemple $k = 0.8$

Pour k^2 un peu $> \frac{1}{2}$, $x(t)$ commence à posséder des maxima et des minima où l'elastica a des tangentes verticales encadrant les points d'inflexion. Les figures 9 et 10 traitent le cas $k = 0.8$. On voit les extrema aplatis de $y'(t)$ se creuser légèrement, les maxima dégénérés bifurquant en deux maxima encadrant un minimum et les minima dégénérés en deux minima encadrant un maximum. Soit $M_1 = \text{dn}^{-1}(\frac{1}{2}) \sim 1.23058$ la valeur de t du premier maximum de $x(t)$. Les points de $t = M_1 + 2nK$ à tangentes

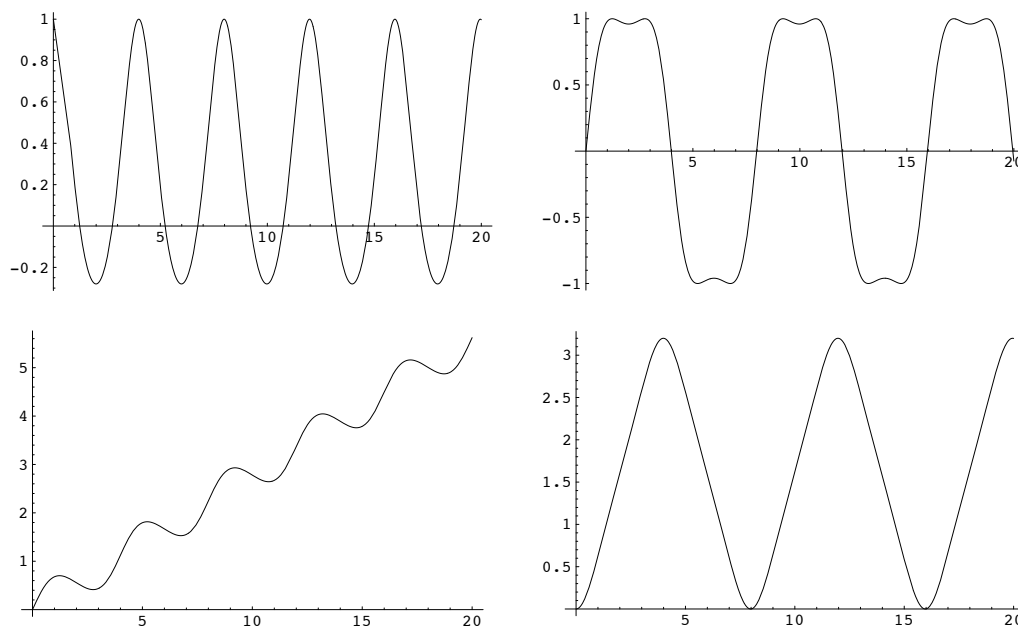


FIGURE 9. Les coordonnées $x(t)$ et $y(t)$ de l'elastica pour $k = 0.8$. En haut à gauche $x'(t)$. En bas à gauche $x(t)$. En haut à droite $y'(t)$. En bas à droite $y(t)$.

verticales ($K \sim 1.9953$) où $x(t)$ possède un maximum local sont marqués d'une flèche rouge et ceux de $t = (2n + 1)K - M_1$ où $x(t)$ possède un minimum local par une flèche bleue.

6.6. Exemple $k \sim 0.8551$

Lorsque k continue à augmenter les arcs où $x(t)$ rétrograde deviennent plus larges et les lacets de l'elastica se rapprochent jusqu'à devenir tangents. Cela vient du fait que comme $2E(k) - K(k)$ décroît, la pente moyenne autour de laquelle $x(t)$ oscille décroît et il arrive un moment où le premier maximum $x(M_1)$ devient égal au deuxième minimum $x(4K - M_1) = x(m_2)$. La résolution numérique de l'équation donne $k \sim 0.8551$. On a alors $M_1 \sim 1.09974$ et $K \sim 2.12409$ et $x(M_1) \sim 0.642955$.

6.7. Exemple $k \sim 0.8858$

Ensuite, au fur et à mesure que $2E(k) - K(k)$ décroît en restant > 0 (i.e. $k < k_0$), les lacets de l'elastica se croisent de plus en plus. La figure représente l'exemple $k = 0.8746$.

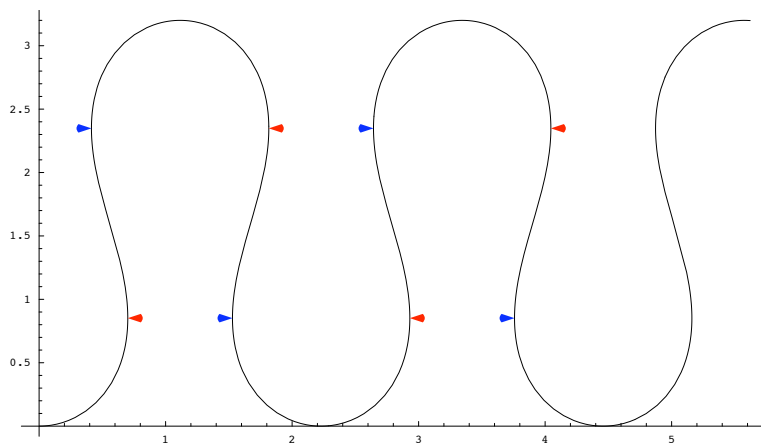


FIGURE 10. L'elastica pour $k = 0.8$. Si $M_1 = \text{dn}^{-1}(\frac{1}{2})$ est la valeur de t du premier maximum de $x(t)$, les points de $t = M_1 + 2nK$ à tangentes verticales où $x(t)$ possède un maximum local sont marqués d'une flèche rouge et ceux de $t = (2n + 1)K - M_1$ où $x(t)$ possède un minimum local par une flèche bleue.

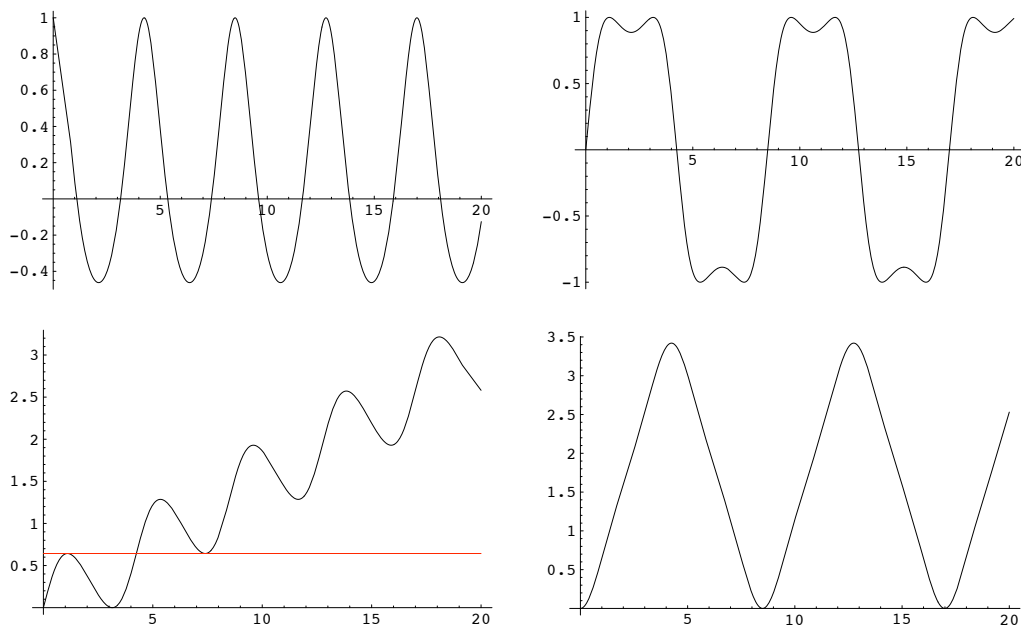


FIGURE 11. Les coordonnées $x(t)$ et $y(t)$ de l'elastica pour $k = 0.8551$. En haut à gauche $x'(t)$. En bas à gauche $x(t)$, la droite d'ordonnée $x(M_1) = x(4K - M_1) \sim 0.642955$ est en rouge. En haut à droite $y'(t)$. En bas à droite $y(t)$.

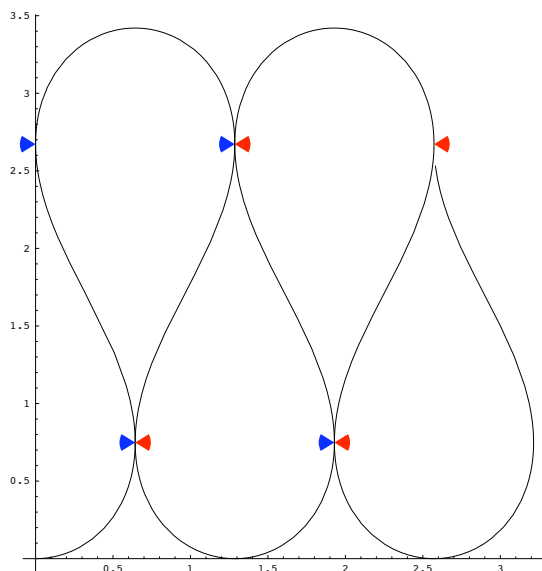


FIGURE 12. L’elastica pour $k = 0.8551$. On note que les points à tangentes verticales de $t = M_1 + 2nK$ (flèches rouges) et de $t = 2(2 + n)K - M_1$ (flèches bleues) coïncident.

Progressivement, les équations $x(M_1) = x(m_n)$ admettent des solutions pour les n successifs. On obtient des elasticæ “critiques” avec des propriétés particulières structurellement instables. La figure 14 représente l’exemple $x(M_1) = x(m_4)$.

6.8. Exemple $k = k_0 \sim 0.908909$ ($2E(k) - K(k) = 0$, lemniscate)

Pour $k \rightarrow k_0^-$ la solution de $x(M_1) = x(m_n)$ n’existe que pour n de plus en plus grand, les lacets de l’elastica se rapprochent de plus en plus les uns des autres et à la limite $k = k_0 \sim 0.908909$, ils se superposent, ce qui donne la lemniscate.

6.9. Exemple $k \sim 0.9624$

Ensuite, pour $k > k_0$, les lacets superposés sur la lemniscate deviennent des boucles de sens alterné qui se séparent et s’éloignent les unes des autres dans le sens des $x < 0$. La pente moyenne de $x(t)$ est désormais > 0 et décroît. Les valeurs critiques de k correspondent maintenant à l’équation $x(m_1) = x(M_n)$ avec $n \geq 3$. Après la dernière valeur critique $k \sim 0.9624$ pour $n = 3$, les boucles deviennent complètement disjointes. La figure 16 montre les cas $k = 0.93$, $k \sim 0.9414$ ($x(m_1) = x(M_4)$) et $k \sim 0.9624$ ($x(m_1) = x(M_3)$). Dans la troisième figure, la droite verticale $x(m_1) = x(M_3) \sim -1.6714$ est en pointillés verts. Quant à la figure 17, elle montre les coordonnées $x(t)$ et $y(t)$ de l’elastica pour la valeur critique $k \sim 0.9624$. On y voit bien l’évolution des profils de $x'(t)$ et $y'(t)$.

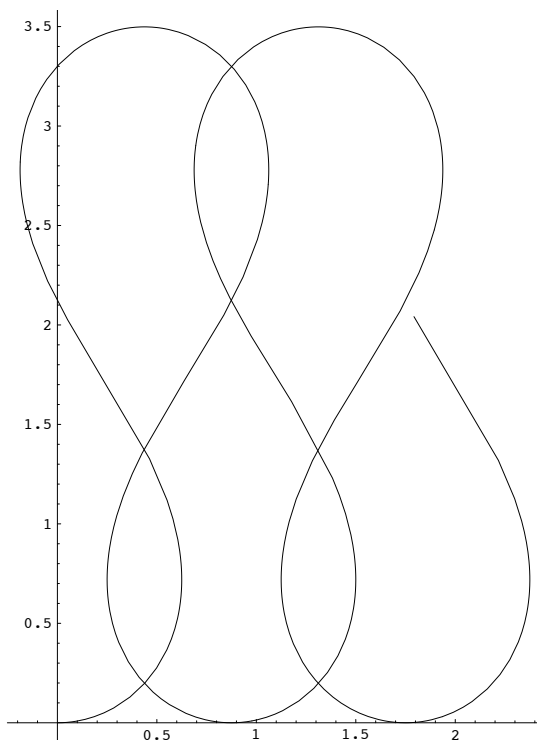


FIGURE 13. L’elastica pour $k = 0.8746$. Les lacets de l’elastica se croisent de plus en plus et $x(t)$ rétrograde suffisamment pour que son premier minimum $x(m_1)$ devienne < 0 .

6.10. Exemple $k : 0.99999 \rightarrow 1$

Ensuite, quand $k \rightarrow 1^-$ et que $2E(k) - K(k)$ plonge asymptotiquement vers $-\infty$, les boucles s’éloignent de plus en plus et, pour $k = 1$, il ne reste plus que la boucle centrale. La figure 18 montre l’exemple de $k = 0.99999$. L’évolution spectaculaire du profil de ces courbes suit celle des fonctions elliptiques illustrée dans la section 7 du chapitre 9. Quant à la figure 19, elle montre l’espacement des boucles (pour voir la boucle centrale en entier on a introduit des $t < 0$). Alors que k est déjà très proche de 1, l’espacement des boucles tendant vers l’infini n’est encore qu’environ 10.

La figure 20 montre la limite $k = 1$ où l’elastica s’exprime en termes de trigonométrie hyperbolique :

$$x(t) = 2 \tanh(t) - t, \quad y(t) = 2 \left(1 - \frac{1}{\cosh(t)} \right).$$

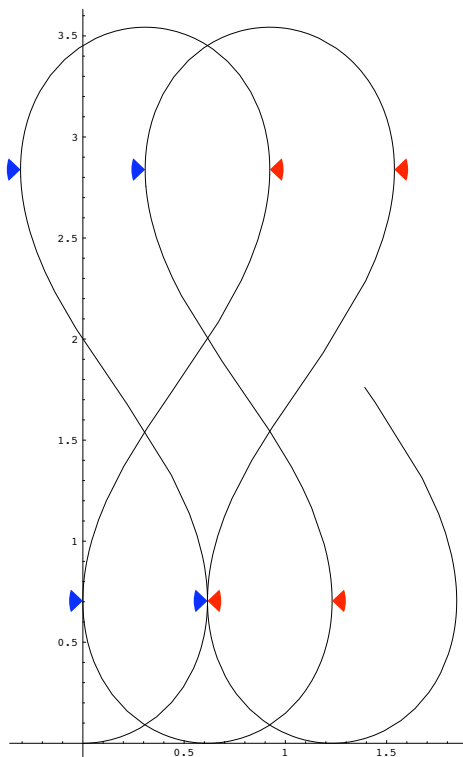


FIGURE 14. L’elastica pour $k = 0.8858$. $x(t)$ rétrograde suffisamment pour que son premier maximum $x(M_1)$ devienne égal à son quatrième minimum $x(m_4)$.

7. Intégration des équations pour $k > 1$ (pendule tournant)

Dans le cas $k > 1$ (pendule tournant) on utilise la transformation $k \rightarrow \frac{1}{k}$. Toutes les fonctions et intégrales elliptiques sont alors de module $\frac{1}{k}$ (i.e. $\operatorname{sn}(u) = \operatorname{sn}(u, \frac{1}{k})$, etc.). On a (pour $r = 1$) dans la strate N_2^+ , $\sin(\frac{\beta}{2}) = \operatorname{sn}(\frac{\varphi}{k})$, $\cos(\frac{\beta}{2}) = \operatorname{cn}(\frac{\varphi}{k})$ et $\frac{c}{2} = \frac{1}{k} \operatorname{dn}(\frac{\varphi}{k})$. Des calculs analogues à ceux des sections précédentes donnent alors les formules de Sachkov :

$$\begin{aligned}
 x(t) &= \frac{2}{k} \left(1 - 2 \operatorname{sn}^2\left(\frac{\varphi}{k}\right) \right) \left(\mathbb{E}\left(\frac{\varphi_t}{k}\right) - \mathbb{E}\left(\frac{\varphi}{k}\right) \right) \\
 &\quad + \frac{4}{k} \operatorname{cn}\left(\frac{\varphi}{k}\right) \operatorname{sn}\left(\frac{\varphi}{k}\right) \left(\operatorname{dn}\left(\frac{\varphi}{k}\right) - \operatorname{dn}\left(\frac{\varphi_t}{k}\right) \right) - \frac{2 - k^2}{k^2} t \\
 y(t) &= \frac{2}{k} \left(2 \operatorname{cn}^2\left(\frac{\varphi}{k}\right) - 1 \right) \left(\operatorname{dn}\left(\frac{\varphi}{k}\right) - \operatorname{dn}\left(\frac{\varphi_t}{k}\right) \right) \\
 &\quad - 2 \operatorname{cn}\left(\frac{\varphi}{k}\right) \operatorname{sn}\left(\frac{\varphi}{k}\right) \left(\frac{2}{k} \left(\mathbb{E}\left(\frac{\varphi_t}{k}\right) - \mathbb{E}\left(\frac{\varphi}{k}\right) \right) - \frac{2 - k^2}{k^2} t \right) .
 \end{aligned}$$

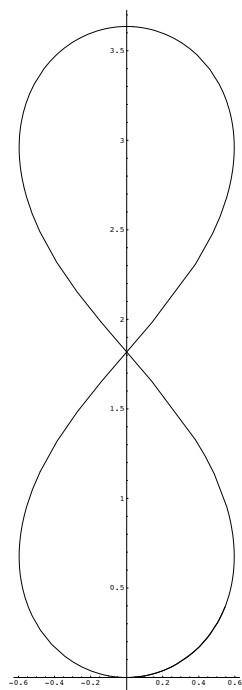


FIGURE 15. L'elastica lemniscate pour $k = k_0 \sim 0.908909$ ($2E(k) - K(k) = 0$).

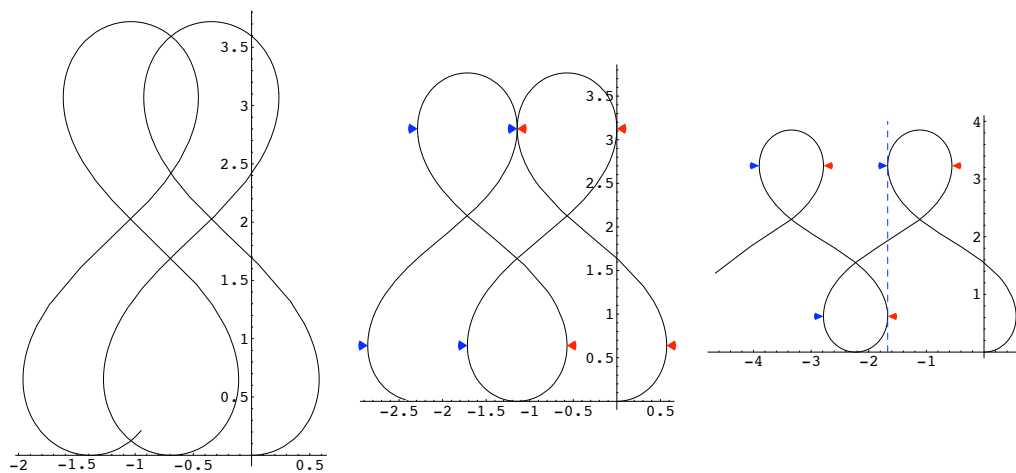


FIGURE 16. De gauche à droite, les elasticæ $k = 0.93 > k_0 \sim 0.908909$ ($2E(k) - K(k) < 0$), $k \sim 0.941403$ ($x(m_1) = x(M_4)$) et $k \sim 0.9624$ ($x(m_1) = x(M_3)$). La droite verticale $x(m_1) = x(M_3) \sim -1.6714$ est en pointillés verts.

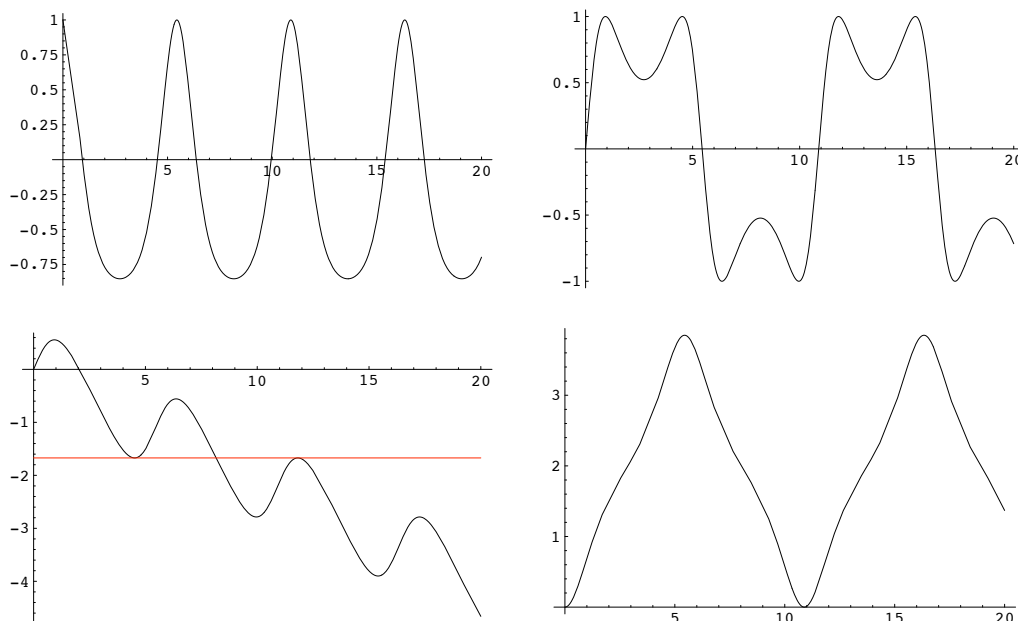


FIGURE 17. Les coordonnées $x(t)$ et $y(t)$ de l'elastica pour $k \sim 0.9624$. En haut à gauche $x'(t)$. En bas à gauche $x(t)$, avec la droite d'ordonnée $x(2K - M_1) = x(M_1 + 4K) \sim -1.6714$ en rouge. En haut à droite $y'(t)$. En bas à droite $y(t)$.

On vérifie que pour $t = 0$, on obtient bien $x_0 = y_0 = 0$.

8. Quelques calculs dans le cas $\varphi_0 = 0$ ($k > 1$)

Les équations simplifiées sont désormais :

$$\begin{aligned} \sin\left(\frac{\theta_t}{2}\right) &= \operatorname{sn}\left(\frac{\varphi_t}{k}\right), \quad \cos\left(\frac{\theta_t}{2}\right) = \operatorname{cn}\left(\frac{\varphi_t}{k}\right) \\ \dot{x}(t) &= 2 \operatorname{cn}^2\left(\frac{\varphi_t}{k}\right) - 1, \quad \dot{y}(t) = 2 \operatorname{sn}\left(\frac{\varphi_t}{k}\right) \operatorname{cn}\left(\frac{\varphi_t}{k}\right) \\ x(t) &= \frac{2}{k} \mathbb{E}\left(\frac{\varphi_t}{k}\right) - \frac{2 - k^2}{k^2} t, \quad y(t) = \frac{2}{k} \left(1 - \operatorname{dn}\left(\frac{\varphi_t}{k}\right)\right). \end{aligned}$$

1. On voit que, puisque dn est de période $2K$ (avec $K = K\left(\frac{1}{k}\right)$), $y(t)$ est maintenant périodique de période $2kK$. C'est une courbe oscillante autour de la droite $y = \frac{1}{k}$. Il est minimal et s'annule pour $\varphi_t = 0 \pmod{2kK}$, il est maximal pour $\varphi_t = kK \pmod{2kK}$.

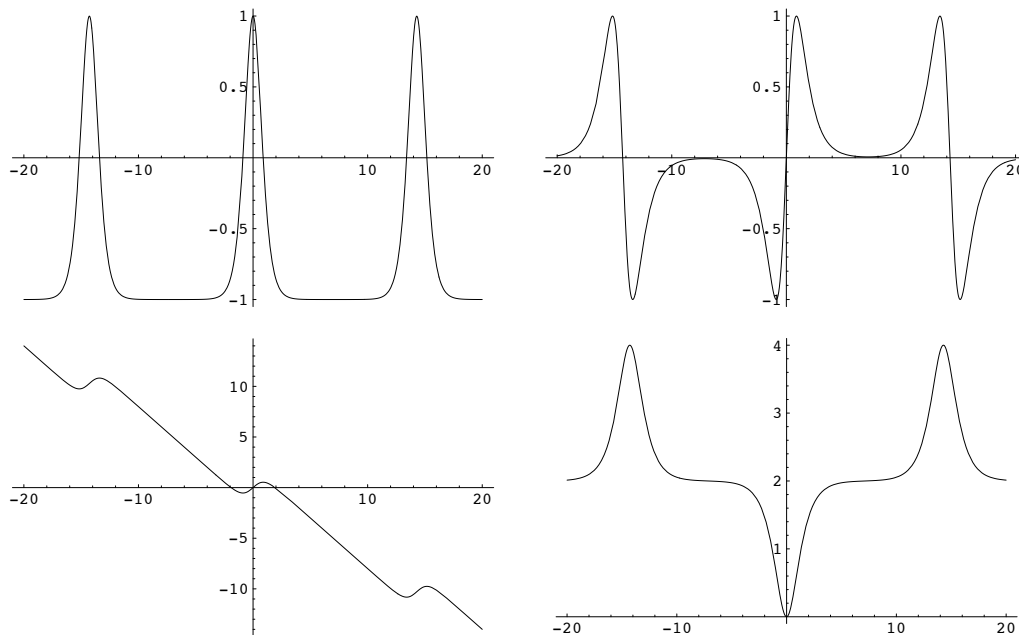


FIGURE 18. Les coordonnées $x(t)$ et $y(t)$ de l'elastica pour $k \sim 0.99999$. En haut à gauche $x'(t)$. En bas à gauche $x(t)$. En haut à droite $y'(t)$. En bas à droite $y(t)$.

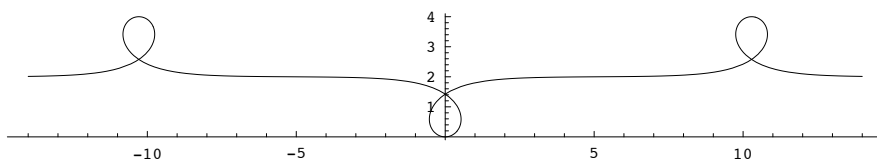


FIGURE 19. L'elastica pour $k \sim 0.99999$. Les boucles s'éloignent de plus en plus les unes des autres (contrairement aux autres figures, $t \in [-20, 20]$ et pas $[0, 20]$).

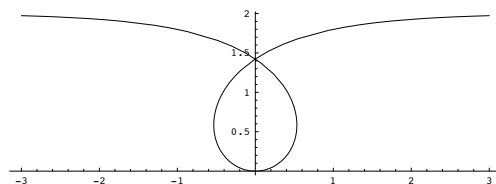
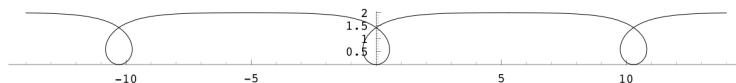


FIGURE 20. L'elastica pour $k = 1$.

2. L'elastica a des tangentes horizontales $\dot{y}(t) = 0$ aux maxima et minima de $y(t)$ où $\text{sn}\left(\frac{\varphi t}{k}\right) = 0$ ou $\text{cn}\left(\frac{\varphi t}{k}\right) = 0$, i.e. pour $\varphi t = 0 \pmod{kK}$. Il démarre avec une tangente horizontale et $x(t), y(t) \geq 0$.

FIGURE 21. L'elastica $k = 1.00001$.

3. L'elastica ne peut pas posséder de points d'inflexions car il faudrait

$$\dot{\theta}_t = \dot{\beta}_t = \frac{2}{k} \operatorname{dn} \left(\frac{\varphi_t}{k} \right) = 0 ,$$

ce qui est impossible.

4. En revanche, l'elastica possède des tangentes verticales $\dot{x}(t) = 0$ pour les racines de l'équation $\operatorname{cn}^2 \left(\frac{\varphi_t}{k} \right) = \frac{1}{2}$ et il en existe toujours.

La géométrie des elasticæ sans inflexions est très simple. Pour k un peu supérieur à 1, la boucle centrale du cas limite $k = 1$ se voit adjoindre des boucles, toutes de même sens de rotation, venant de l'infini. Ces boucles se rapprochent de plus en plus les unes des autres jusqu'à se confondre en un cercle lorsque $k = \infty$.

Les figures 21 et 22 montrent l'exemple $k = \frac{1}{0.99999} \sim 1.00001$. Il est intéressant de les comparer avec les figures 18 et 19 pour voir comment s'effectue l'échange entre $\operatorname{dn}(\varphi_t)$ et $\operatorname{cn} \left(\frac{\varphi_t}{k} \right)$. Quant à la figure 23, elle montre les trois elasticæ $k = 1.075, 1.037, 1.17, 2$. On voit les boucles se croiser de plus en plus.

Le graphe de $x(t)$ est une courbe oscillante autour d'une droite de pente négative tendant vers 0^- lorsque $k \rightarrow \infty$. Cela permet des solutions pour l'équation $x(m_1) = x(M_n)$ avec des n de plus en plus grands. Elles correspondent à une suite infinie de tangences verticales entre boucles au fur et à mesure que celles-ci tendent à se superposer. La première tangence a lieu pour $x(m_1) = x(M_3)$ pour $k \sim 1.088$. A la limite $k = \infty$, les boucles se superposent exactement et l'elastica devient un cercle.

9. Résumé

Résumons les étapes de l'intégration du système

$$\begin{cases} \dot{\beta} = \dot{\alpha} = c = h_2 = \dot{\theta} = \kappa \\ \dot{c} = -r \sin(\beta) \\ \dot{r} = 0, \dot{\theta} = c \\ \dot{x} = \cos(\theta) \\ \dot{y} = \sin(\theta) \end{cases}$$

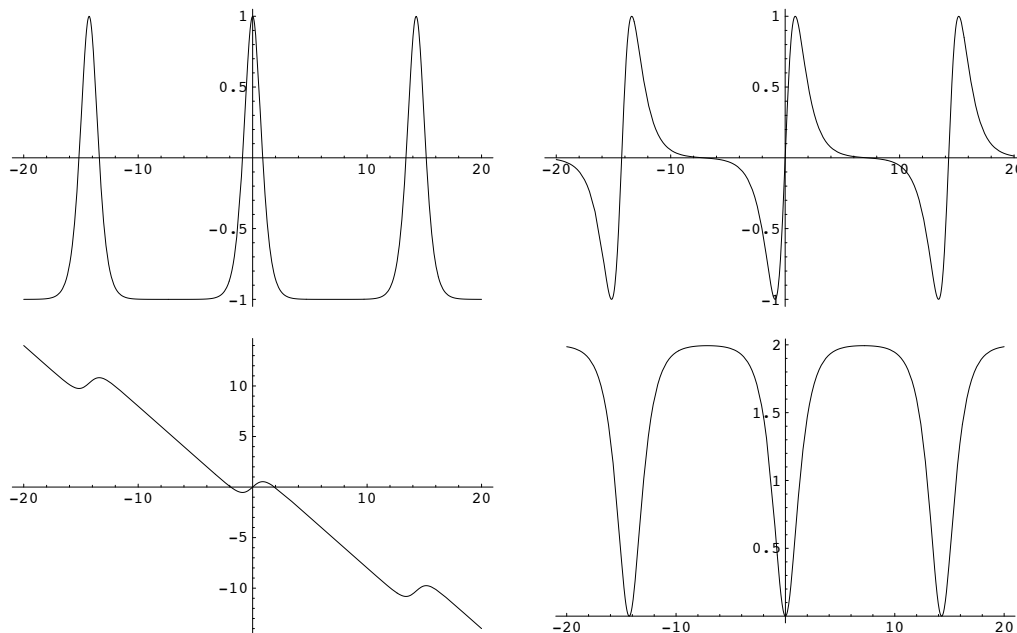


FIGURE 22. La géométrie de l’elastica $k = 1.00001$. En haut à gauche $x'(t)$. En bas à gauche $x(t)$. En haut à droite $y'(t)$. En bas à droite $y(t)$. À comparer avec la figure 18 pour $k = 0.99999$.

1. On veut trouver les elasticæ dans \mathbb{R}^2 . L’espace de configuration est $\mathbb{V} = SE(2) = \mathbb{R}^2 \times \mathbb{S}^1 = \{q = (x, y, \theta)\}$ produit semi-direct du groupe des translations par le groupe des rotations de \mathbb{R}^2 . On paramétrise les relevées lagrangiennes Γ dans \mathbb{V} de courbes γ dans \mathbb{R}^2 par la longueur d’arc $s = t$ de γ . On a donc une vitesse unitaire et $\dot{\theta} = \kappa$ (courbure).
2. En plus des coordonnées classiques (x, y, θ) avec les bases holonomes associées $(\partial_x, \partial_y, \partial_\theta)$ pour les vecteurs tangents de $T\mathbb{V}$ et $(dx, dy, d\theta)$ pour les vecteurs cotangents de $T^*\mathbb{V}$, on a les bases non holonomes $SE(2)$ -invariantes (X_1, X_2, X_3) et $(\omega_1, \omega_2, \omega_3)$. Les composantes des covecteurs $\lambda \in T_q^*\mathbb{V}$ dans la base des ω_i sont les h_i (système “vertical”).
3. Les équations ont pour contrôle $u = \dot{\theta}$ et les trajectoires satisfont

$$\dot{q} = X_1 + uX_2 .$$

Pour minimiser l’intégrale du carré de la courbure (loi de Bernoulli-Euler), on se place dans l’espace des phases $T^*\mathbb{V}$ et, en appliquant le PMP, on trouve $u = h_2$ (courbure), $\dot{q} = X_1 + h_2X_2$ et l’hamiltonien $H = h_1 + \frac{1}{2}h_2^2$.

4. Les équations de Hamilton donnent un système “vertical” qui s’exprime simplement avec les composantes h_i : (h_1, h_3) est sur le cercle de rayon r et

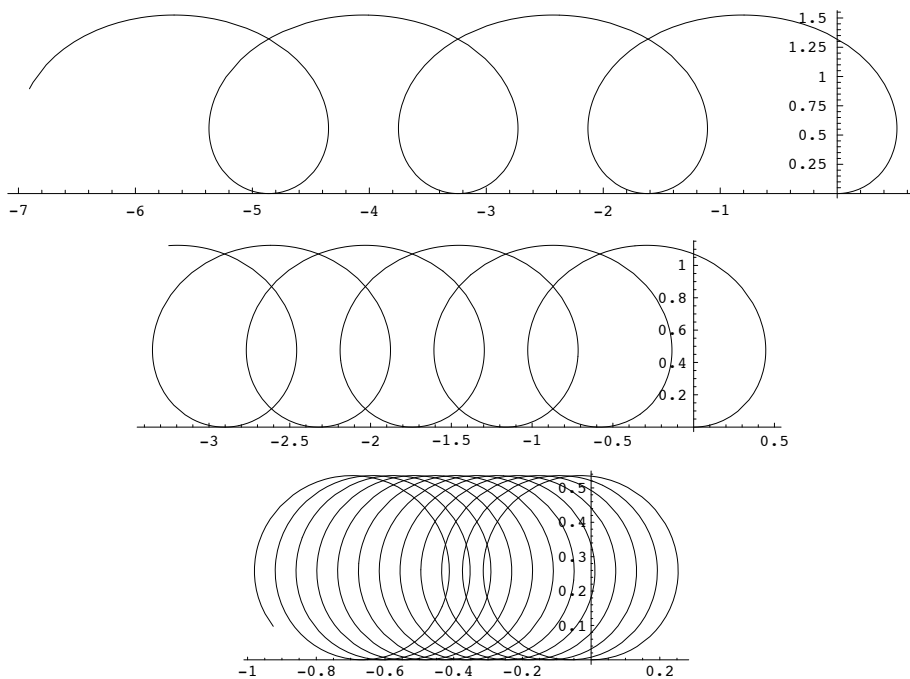


FIGURE 23. Les trois elasticæ $k = 1.075, 1.037, 1.17, 2$. Les boucles se croisent de plus en plus.

$h_1^2 + h_3^2 = r^2$ est une seconde intégrale première. Si on utilise les coordonnées polaires (r, α) dans le plan (h_1, h_3) , r est constant et $\beta = \alpha + \pi$ satisfait l'équation du pendule $\ddot{\beta} = -r \sin(\beta)$ avec $\dot{\beta} = c = h_2 = \dot{\theta} = \kappa$ (courbure). On obtient une stratification naturelle du portrait de phase du pendule sur le cylindre des $(\beta, c = \dot{\beta})$.

5. On introduit alors l'application $(c, t) \mapsto (\beta_t, c_t)$ qui associe à une condition initiale $(\beta_0 = 0, c_0 = c)$ et une durée t la position (β_t, c_t) du pendule au temps t . Modulo les périodes, c'est un difféomorphisme. Son inverse $(\beta_t, c_t) \mapsto (c_0, \varphi)$ donne le temps φ que met la trajectoire du pendule partant de $(\beta_0 = 0, c_0)$ pour arriver en (β_t, c_t) .
6. On reparamétrise l'énergie comme un module k d'IE et on considère les coordonnées "elliptiques" de Jacobi (φ, k) .
7. L'intégration repose
 - sur $\theta_t = \beta_t - \beta$, l'expression de \dot{x} et \dot{y} en termes de fonctions quadratiques de $\sin(\frac{\theta}{2})$ et $\cos(\frac{\theta}{2})$,
 - sur les conversions $\sin(\frac{\beta}{2}) = k \operatorname{sn}(\varphi)$, $\cos(\frac{\beta}{2}) = \operatorname{dn}(\varphi)$,
 - sur les intégrales des carrés et produits de FE.

10. Symétries et points de Maxwell

La partie la plus technique et la plus innovante du travail de Yuri Sachkov consiste à analyser les symétries et les points remarquables (points de Maxwell, points du cut locus, points conjugués) des *elasticæ* en utilisant les formules explicites. Jusqu'ici, nous avons considéré les *elasticæ* individuels, à savoir les trajectoires du champ hamiltonien \vec{H} . Une fois choisi un point de départ q_0 , ce sont les projections dans \mathbb{R}^2 des trajectoires définies par l'exponentielle

$$\text{Exp} : T_{q_0}^* \mathbb{V} \times \mathbb{R}^+ \rightarrow \mathbb{V}, (\lambda, t) \mapsto \pi \circ e^{t\vec{H}}(\lambda) = q(t) .$$

Comme les mouvements du pendule sont temporellement périodiques, ces *elasticæ* sont géométriquement périodiques et, sauf pour la lemniscate et le cercle, des courbes infinies avec reproduction de motifs. Mais lorsque l'on considère des *elasticæ* de longueur finie t , on fixe un temps maximal t et on regarde la section Exp_t de Exp . Il peut alors se produire un ensemble de phénomènes liés à la structure de Exp_t et de la fonctionnelle \mathcal{E}_t : des points critiques, des points critiques dégénérés, etc.¹² Les points de Maxwell en sont un exemple. Avant de les définir, revenons d'abord sur les symétries du système des *elasticæ*.

Les symétries σ^i du portrait de phase C définies en 3 induisent des réflexions¹³ du fibré cotangent $T^*\mathbb{V}$ dont les éléments sont les $\lambda = (\nu, q) \in T^*\mathbb{V}$ avec $q = (x, y, \theta) \in \mathbb{V}$ et $\nu = (\beta, c, r) \in T_q^*\mathbb{V}$. Pour un *elastica* avec $s \in [0, t]$, on a

$$\sigma^1(\theta_s) = \theta_{t-s} - \theta_t, \quad \sigma^1((x_s, y_s)) = R_{-\theta_t}((x_t - x_{t-s}, y_t - y_{t-s}))$$

(où R_α est la rotation d'angle α), à savoir la symétrie par rapport au milieu de la corde $\{(0, 0), (x_t, y_t)\}$,

$$\sigma^2(\theta_s) = \theta_t - \theta_{t-s}, \quad \sigma^2((x_s, y_s)) = R_{\theta_t}((x_t - x_{t-s}, y_{t-s} - y_t)) ,$$

à savoir la symétrie par rapport à la médiatrice de la corde $\{(0, 0), (x_t, y_t)\}$,

$$\sigma^3(\theta_s) = \theta_{-s}, \quad \sigma^3((x_s, y_s)) = (x_s, -y_s) ,$$

à savoir la symétrie par rapport à la corde $\{(0, 0), (x_t, y_t)\}$.

Pour les réflexions des *arcs* d'*elastica*, on prend $\nu = (\beta, c, r)$ comme origine $\nu_0 = (\beta_0, c_0, r)$ pour $s = 0$ (r reste invariant). On a alors (cf. 3)

$$\begin{cases} \sigma^1((\beta, c)) = (\beta_t, -c_t) \\ \sigma^2((\beta, c)) = (-\beta_t, -c_t) \\ \sigma^3((\beta, c)) = (-\beta, -c) \end{cases}$$

12. On peut montrer que les conditions d'optimalité de l'hamiltonien, de régularité des extrémales et d'analyticité du problème font que les points critiques sont nécessairement des *minima* et, qui plus est, des minima *non dégénérés* (mais, évidemment, pas forcément absolus).

13. Les symétries σ^i transforment les *elasticæ* en d'autres *elasticæ* qui en sont les images symétriques. Certains *elastica* peuvent rester invariants et posséder eux-mêmes une symétrie.

En termes de coordonnées elliptiques $(\beta, c, r) \mapsto (k, \varphi, r)$, les symétries σ^i opèrent dans les différentes strates de N en agissant sur φ car r et k (l'énergie) restent invariants. Par exemple dans N_1 , considérons un arc allant de $\varphi = \varphi_0$ à φ_t . σ^1 et σ^3 transforment c en $-c$ et σ^2 laisse c invariant. Par ailleurs, $\sigma^1(\beta) = \beta_t$ et donc, d'après la construction des coordonnées elliptiques,

$$\begin{aligned} \operatorname{sn}(\sigma^1(\varphi)) &= \frac{1}{k} \sin\left(\frac{\sigma^1(\beta)}{2}\right) = \frac{1}{k} \sin\left(\frac{\beta_t}{2}\right) = \operatorname{sn}(\varphi_t) \\ \operatorname{cn}(\sigma^1(\varphi)) &= \frac{1}{k} \frac{\sigma^1(c)}{2} = -\frac{1}{k} \frac{c_t}{2} = -\operatorname{cn}(\varphi_t) . \end{aligned}$$

Cela implique $\sigma^1(\varphi) + \varphi_t = 2K \pmod{4K}$. De même, $\sigma^2(\varphi) + \varphi_t = 0 \pmod{4K}$ et $\sigma^3(\varphi) - \varphi = 2K \pmod{4K}$. On obtient des formules analogues pour les autres strates de N .

Ce qui intéresse Sachkov sont les *points de Maxwell*, c'est-à-dire les points où un elastica perd sa propriété d'optimalité, en particulier à cause de la *compétition* entre arcs symétriques l'un de l'autre. Pour un temps terminal t (i.e. une longueur), les extrémités $q(t) = q_t$ des elasticæ q_s , $s \in [0, t]$ d'origine q_0 sont données par l'application exponentielle du système hamiltonien

$$\operatorname{Exp}_t : T_{q_0}^* \mathbb{V} \rightarrow \mathbb{V}, \quad \lambda \mapsto \pi \circ e^{t\bar{H}}(\lambda) = q(t) .$$

Plus explicitement, les points de Maxwell sont les $\lambda = (\nu, q) \in N$ tels que

- (i) il existe $\tilde{\lambda}$ avec $\tilde{q}_s \neq q_s$ pour $s \in [0, t]$,
- (ii) $\tilde{q}_t = q_t$,
- (iii) la fonctionnelle \mathcal{E}_t est la même pour \tilde{q}_s et q_s .

En utilisant la propriété d'analyticité et l'unicité du prolongement analytique on peut montrer que ces conditions signifient que, après t , \tilde{q}_s minimise mieux la fonctionnelle que q_s .

Il est idoine d'utiliser les symétries. Dans le cas des arcs transformés par symétries, on a $\tilde{q}_s = \sigma^i(q_s)$ et on montre que la condition (iii) d'égalité $\mathcal{E}_t(\tilde{q}_s) = \mathcal{E}_t(q_s)$ est toujours vérifiée. Il faut donc pouvoir éliminer la possibilité que $\sigma^i(q_s)$ soit identique à q_s , autrement dit le cas d'un elastica lui-même symétrique. Il est facile de montrer que, en dehors de cas limites, $\sigma^i(q_s) \equiv q_s$ équivaut à $\sigma^i(\lambda) = \lambda$. Il faut donc $\sigma^i(\lambda) \neq \lambda$ pour satisfaire la condition (i). Reste alors la condition (ii) $\sigma^i(q_t) = q_t$. Considérons par exemple le cas $\sigma^1(q_t) = q_t$. Nous avons vu que $\sigma^1(\theta_t) = -\theta_t$ et $\sigma^1((x_t, y_t)) = R_{-\theta_t}((x_t, y_t))$. Il faut donc $\theta_t = -\theta_t$, i.e. $\theta_t = 0$, et, comme cela implique $\sigma^1((x_t, y_t)) = (x_t, y_t)$, cette condition nécessaire est aussi suffisante : de même que q_0 , q_t doit être un point à tangente horizontale. Pour $i = 3$, il faut $\sigma^3(q_t) = (-\theta_t, x_t, -y_t) = q_t$ et donc $\theta_t = 0 \pmod{\pi}$ et $y_t = 0$. Enfin, pour $i = 2$, la condition est plus compliquée. Il faut $\sigma^2(q_t) = (\theta_t, R_{\theta_t}((x_t, -y_t))) = q_t$ et la condition sur θ_t est automatiquement satisfaite. La contrainte ne porte donc pas

sur θ_t mais sur $R_{\theta_t}((x_t, -y_t)) = (x_t, y_t)$, autrement dit sur la satisfaction des deux équations :

$$\begin{aligned}x_t \cos(\theta_t) + y_t \sin(\theta_t) &= x_t, \\x_t \sin(\theta_t) - y_t \cos(\theta_t) &= y_t.\end{aligned}$$

Si $(x_t, y_t) = (0, 0)$, la condition est évidemment satisfaite. En revanche si $(x_t, y_t) \neq (0, 0)$ et si l'on écrit en coordonnées polaires $x_t = \rho \cos(\chi)$ et $y_t = \rho \sin(\chi)$, elle s'écrit $\theta_t - \chi = \chi$, soit $\chi = \frac{\theta_t}{2}$ ou encore, puisque par construction $x_t \sin(\chi) - y_t \cos(\chi) = 0$,

$$x_t \sin\left(\frac{\theta_t}{2}\right) - y_t \cos\left(\frac{\theta_t}{2}\right) = 0.$$

Sachkov appelle $P = 0$ cette condition.

Ensuite Sachkov étudie les points fixes des symétries σ^i dans N . Soit $\tau = \frac{\varphi_t + \varphi}{2}$ le milieu de l'elastica. Il montre que pour les elastics inflexionnels $\nu \in N_1$, $\sigma^1(\nu) = \nu$ équivaut à $\text{cn}(\tau) = 0$, que $\sigma^2(\nu) = \nu$ équivaut à $\text{sn}(\tau) = 0$, et que $\sigma^3(\nu) = \nu$ est impossible. Cela est facile à comprendre. Si par exemple $\sigma^1(\nu) = \nu$, i.e. $\sigma^1(\varphi) = \varphi$, alors, puisque (cf. plus haut) $\sigma^1(\varphi) + \varphi_t = 2K \pmod{4K}$, on a $\varphi + \varphi_t = 2K \pmod{4K}$ et donc $\tau = K \pmod{2K}$, ce qui équivaut bien à $\text{cn}(\tau) = 0$. Cela signifie que φ et φ_t sont symétriques par rapport à l'axe β , que τ satisfait donc $c = 0$ et correspond à un point d'inflexion (cf. figure 24). De même, si $\sigma^2(\varphi) = \varphi$, alors $\sigma^2(\varphi) + \varphi_t = \varphi + \varphi_t = 0 \pmod{4K}$ et donc $\tau = 0 \pmod{2K}$, ce qui équivaut bien à $\text{sn}(\tau) = 0$. Cela signifie que φ et φ_t sont symétriques par rapport à l'axe c , que τ satisfait donc $\beta = 0$ et correspond à un vertex. Enfin, pour avoir $\sigma^3(\varphi) = \varphi$, il faudrait avoir $\varphi - \varphi = 0 = 2K \pmod{4K}$, ce qui est impossible. Pour les elastics non inflexionnels $\nu \in N_2$, ce sont en revanche $\sigma^1(\nu) = \nu$ et $\sigma^2(\nu) = \nu$ qui sont impossibles et $\sigma^3(\nu) = \nu$ équivaut à $\text{sn}(\tau) \text{cn}(\tau) = 0$. Cela signifie que φ et φ_t sont symétriques par rapport à l'axe c , que τ satisfait donc $\beta = 0$ et correspond à un vertex. Si φ est à gauche de l'axe c et φ_t à droite, le vertex est un bas de boucle, et si φ est à droite et φ_t à gauche, le vertex est un haut de boucle.

On obtient ainsi une description précise des conditions intervenant dans la définition des points de Maxwell déductibles des symétries. Pour les calculer effectivement il faut résoudre les équations $\theta_t = 0 \pmod{\pi}$ et $P = 0$. La condition $\theta_t = 0$ n'est pas trop difficile à expliciter. Posons $\varphi = \tau - p$ et $\varphi_t = \tau + p$, i.e. $p = \frac{\varphi_t - \varphi}{2}$. Par exemple pour les elastics inflexionnels $\nu \in N_1$, $\theta_t = 0$ équivaut à $\frac{\theta_t}{2} = \frac{\beta_0}{2} \pmod{\pi}$, c'est-à-dire à $\text{sn}(\varphi_t) = \text{sn}(\varphi)$ et $\text{dn}(\varphi_t) = \text{dn}(\varphi)$, soit $\varphi_t = \varphi \pmod{4K}$ ou $\varphi_t = 2K - \varphi \pmod{4K}$. Cela donne les conditions $\text{cn}(\tau) = 0$ ou $p = 2Kn$ ($n \in \mathbb{Z}$). On trouve des conditions analogues pour les autres strates de N .

C'est la condition $P = 0$ intervenant dans la symétrie σ^2 qui est la plus technique et se trouve au cœur du problème. Il faut revenir aux expressions donnant $\sin\left(\frac{\theta_t}{2}\right)$ et $\cos\left(\frac{\theta_t}{2}\right)$ en termes de fonctions elliptiques de Jacobi et y remplacer φ_t et φ par τ

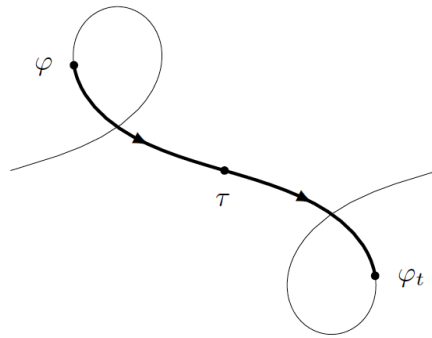


FIGURE 24. Un elastica $\nu \in N_1$ tel que $\sigma^1(\nu) = \nu$. La condition $\text{cn}(\tau) = 0$ avec $\tau = \frac{\varphi_t + \varphi}{2}$ le milieu de l'elastica signifie que φ et φ_t sont symétriques par rapport à l'axe β et que τ satisfait $c = 0$ et correspond à un point d'inflexion.

et p . On obtient une expression compliquée qui, pour $\nu \in N_1$, a la forme

$$P_t = \frac{4k}{\Delta} \text{sn}(\tau) \text{dn}(\tau) f_1(p, k)$$

$$\Delta = 1 - k^2 \text{sn}^2(p) \text{sn}^2(\tau)$$

$$f_1(p, k) = \text{sn}(p) \text{dn}(p) - (2\mathbb{E}(p) - p) \text{cn}(p) .$$

Comme $\text{dn}(\tau) \neq 0$ si $k < 1$, les racines de P_t sont les zéros de $\text{sn}(\tau)$ et les racines de $f_1(p, k)$. Mais pour les points de Maxwell liés à la symétrie σ^2 on doit avoir $\text{sn}(\tau) \neq 0$ et il ne reste que les racines de $f_1(p, k)$. C'est donc l'analyse de la fonction f_1 qui se trouve au cœur du problème. Des calculs compliqués montrent qu'elle possède une suite infinie strictement croissante de racines p_n^1 ($n \in \mathbb{Z}$) avec $p_{-n}^1 = -p_n^1$ (et donc $p_0^1 = 0$) et $p_n^1 \in (-K + 2Kn, K + 2Kn)$. Pour $k = k_0$ (valeur critique), $p_n^1 = 2Kn$. Ces racines ont une interprétation géométrique précise : on prend un point d'inflexion comme origine et on construit les tangentes aux boucles successives de l'elastica ; elles vont par couples (A_n, B_n) à cause des symétries et on montre que $p_n^1 = \frac{A_n B_n}{2}$, $\overline{A_n B_n}$ étant la longueur du segment (A_n, B_n) . La figure 25 montre les deux cas $k < k_0$ et $k > k_0$.

Pour $\nu \in N_2$ (elasticae non inflexionnels), on obtient une formule analogue

$$P_t = \frac{4}{\Delta} \text{sn}(\tau) \text{cn}(\tau) f_2(p, k)$$

mais sans le facteur k , avec $\text{cn}(\tau)$ au lieu de $\text{dn}(\tau)$ et, surtout, une fonction $f_2(p, k)$ sans racines $\neq 0$. Il ne reste donc comme racines de $P_t = 0$ que $p = 0$ (i.e. $\varphi_t = \varphi$), $\text{sn}(\tau) = 0$ et $\text{cn}(\tau) = 0$. Mais pour les points de Maxwell il faut $\text{sn}(\tau) \text{cn}(\tau) \neq 0$ et il ne reste donc que la solution $p = 0$. On peut compléter ces descriptions pour les

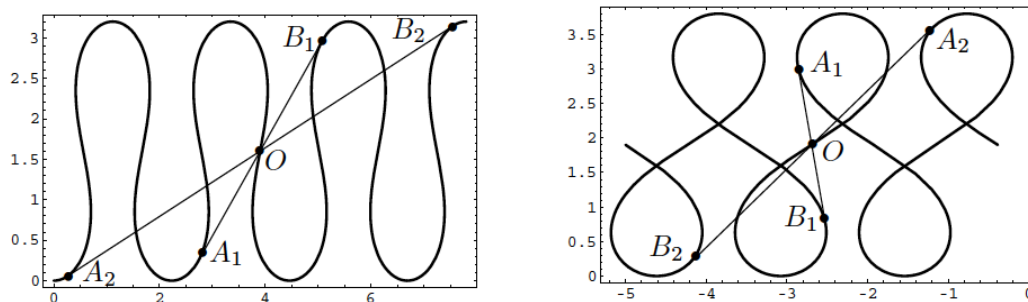


FIGURE 25. Interprétation géométrique des racines de la fonction f_1 intervenant dans la condition $P = 0$ dans le cas des elasticæ inflexionnels. À gauche cas $k < k_0$, à droite cas $k > k_0$ (D'après Sachkov [476]).

autres strates de N . Nous renvoyons le lecteur intéressé aux beaux articles [476] et [477] de Yuri Sachkov où il trouvera les calculs exhaustifs (compliqués).

11. Cut locus et points conjugués

Pour $\lambda \in N = T_{q_0}^* \mathbb{V}$, le point du cut locus t_{cut} est défini, s'il existe, comme le sup des $t_1 > 0$ tels que $q(s)$ soit optimal dans le segment $[0, t_1]$. Autrement dit, c'est un point où une courbe extrémale perd son optimalité globale. Évidemment $t_{\text{cut}}(\lambda) > 0$ car les petits arcs extrémaux sont toujours optimaux. On peut alors montrer que $t_{\text{cut}}(\lambda) \leq T(\lambda)$ où $T(\lambda)$ est la période du pendule.

On appelle point *conjugué* un point où une courbe extrémale perd son optimalité locale (elle devient instable). Un point q_t est un point conjugué s'il est sur l'enveloppe des trajectoires extrémales. Cela signifie que l'application exponentielle :

$$\text{Exp}_t : T_{q_0}^* \mathbb{V} \rightarrow \mathbb{V}, \lambda_0 \mapsto \text{Exp}_t(\lambda_0) = \pi \circ e^{t\bar{H}}(\lambda_0) = q(t)$$

est dégénérée, autrement dit que son Jacobien J , à savoir le déterminant de sa matrice jacobienne \mathfrak{J} , s'annule. En utilisant par exemple les coordonnées de Jacobi (φ, k, r) , il faut calculer

$$\mathfrak{J} = \begin{pmatrix} \frac{\partial x_t}{\partial \varphi} & \frac{\partial x_t}{\partial k} & \frac{\partial x_t}{\partial r} \\ \frac{\partial y_t}{\partial \varphi} & \frac{\partial y_t}{\partial k} & \frac{\partial y_t}{\partial r} \\ \frac{\partial \theta_t}{\partial \varphi} & \frac{\partial \theta_t}{\partial k} & \frac{\partial \theta_t}{\partial r} \end{pmatrix}.$$

On trouvera dans [477] les calculs remarquables de Sachkov donnant la formule de J et la valeur de ses zéros. Ils montrent que pour les elasticæ inflexionnels il existe une infinité de points conjugués $t_{\text{conj}}^n(\lambda)$ le premier, $t_{\text{conj}}^1(\lambda)$ étant situé entre des points de Maxwell dans l'intervalle $[\frac{T}{2}, \frac{3T}{2}]$, entre le premier et le troisième point d'inflexion. Ils montrent également que les elasticæ non inflexionnels n'ont pas de points conjugués.

Contours illusoires modaux et géodésiques sous-riemanniennes

1. Passage à la géométrie sous-riemannienne

Dans les précédents chapitres 8, 10, 11 nous avons présenté plusieurs modèles variationnels de nature géodésique pour les contours illusoires modaux, modèles élaborés au cours des années 1990. Pour ce faire, nous avons utilisé une métrique riemannienne sur \mathbb{V} tout entier en traitant la condition d'intégrabilité comme une contrainte (méthode des multiplicateurs de Lagrange). La structure de contact de \mathbb{V} était donc prise en compte mais sans qu'il existât un lien intrinsèque avec la métrique considérée.

Ce n'est qu'au début des années 2000 que nous nous sommes rendus compte que dans nos premiers modèles nous avons traité "à la main" un problème typique de *géométrie sous-riemannienne* sur le groupe de Lie $\mathbb{V}_J = J^1\mathbb{R}^2$ identifié au groupe de Heisenberg polarisé. Nous avons donc adopté ce point de vue qui consiste à partir de la structure de contact \mathcal{K} sur \mathbb{V}_J pour introduire la métrique *directement sur les plans de contact*.

Comme dans ces précédents chapitres 8, 10, 11, nous pouvions nous placer dans l'espace des 1-jets

$$\mathbb{V}_J = J^1(\mathbb{R}, \mathbb{R}) \text{ (noté } J^1\mathbb{R}^2\text{)}$$

ou dans le groupe

$$\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1 \simeq SE(2) = SO(2) \rtimes \mathbb{R}^2.$$

Le premier cas avait déjà été partiellement traité dans la littérature pour le groupe de Heisenberg non polarisé. En revanche le second, que nous avons analysé avec Alessandro Sarti et Giovanna Citti (cf. notre exposé à la Société Mathématique de France en 2006 [434]), restait ouvert et rencontrait quelques difficultés, sa géométrie sous-riemannienne n'ayant jamais été, aussi bizarre que cela puisse paraître, investiguée en grand détail dans cette optique. Heureusement, notre collègue du CREA Helena Frankowska nous mit en relation avec Andrei Agrachev, l'éminent spécialiste de théorie du contrôle et de géométrie sous-riemannienne que nous avons déjà souvent cité, et, après discussion, Agrachev a résolu le problème. Les sources initiales

que nous utilisâmes furent un exposé qu'il fit à l'Institut Henri Poincaré et un preprint de son élève Igor Moiseev, puis les travaux de son collègue Yuri Sachkov dont nous avons exposé la modélisation des elasticæ dans le chapitre précédent (13).

Dans ce chapitre, nous allons commencer par le premier cas, lui-même double (groupe de Heisenberg polarisé et non polarisé), puis nous en viendrons ensuite au second cas pour l'approfondir.

La géométrie des situations sous-riemanniennes que nous allons expliciter est descriptible de la façon suivante à un niveau général abstrait. On a une variété V de dimension 3 (en général produit d'un plan base (x_1, x_2) par une fibre x_3), un champ de plans non intégrable \mathcal{K} de TV définissant une structure de contact et une métrique g sur \mathcal{K} . \mathcal{K} est le noyau d'une 1-forme ω telle que $\omega \wedge d\omega$ soit une forme volume de V . Le calcul des géodésiques s'effectue à partir d'un lagrangien ou, dans le cadre de la théorie du contrôle optimal, d'un hamiltonien. On retrouve des problèmes analogues à ceux que nous avons rencontrés pour les elasticæ dans le chapitre 13 (cut locus, points conjugués). Mais comme il s'agit de métriques et de géodésiques on rencontre aussi des problèmes nouveaux (sphères, fronts d'onde).

1. Dans la version lagrangienne, les conditions initiales sont données par

$$\begin{aligned} v(0) &= (x_1(0), x_2(0), x_3(0)) \\ \dot{v}(0) &= (\dot{x}_1(0), \dot{x}_2(0), \dot{x}_3(0)) . \end{aligned}$$

Dans la version hamiltonienne elles sont données par $v(0)$ et un vecteur cotangent

$$\varpi(0) = (\xi_1^*(0), \xi_2^*(0), \xi_3^*(0)) .$$

2. L'application exponentielle se définit de la façon suivante. Soit $v = (x_1, x_2, x_3)$ un élément de V et C_v la sous-variété de T_v^*V d'énergie $H(v, \varpi) = 1/2$. L'application exponentielle est l'application $\mathcal{E}_v : \mathbb{R} \times C_v \rightarrow V$ qui associe à (s, ϖ) (s étant la variable temporelle) la projection de l'extrémité au temps s de la H -trajectoire partant de ϖ au temps $s = 0$.

3. La sphère $S(v, r)$ de centre v et de rayon r est l'ensemble

$$S(v, r) = \{v' \mid d_g(v, v') = r\}$$

(d_g étant la distance définie par la métrique g).

4. Le front d'onde $W(v, r)$ est l'ensemble

$$W(v, r) = \{v' \mid \text{il existe une géodésique } \Gamma : v \rightarrow v' \text{ de longueur } r\} .$$

5. Le cut locus de v est l'ensemble

$$\begin{aligned} \{v' \mid v' \text{ est l'extrémité d'une géodésique } \Gamma : v \rightarrow v' \\ \text{qui n'est plus un minimiseur global}\} . \end{aligned}$$

6. Le lieu conjugué de v (ou la caustique) est l'ensemble

$$\Sigma_v = \{\text{lieu singulier de l'application exponentielle } \mathcal{E}_v\}.$$

La difficulté principale vient du fait que, comme l'expliquent Andrei Agrachev et ses collaborateurs dans [8] (p. 378)

“By contrast with the Riemannian case, here the conjugate locus and the cut locus of a point x contain x in their adherence. Consequently, even spheres of arbitrary small radius have singularities.”

2. Les géodésiques sous-riemanniennes de \mathbb{V}_J

Pour approfondir la géométrie sous-riemannienne de V , problème difficile et encore partiellement ouvert, commençons donc par le groupe de Heisenberg polarisé $\mathbb{V}_J = J^1\mathbb{R}^2$. Le problème (en particulier le calcul de la sphère et du front d'onde) peut être entièrement résolu en s'inspirant des travaux de Richard Beals, Bernard Gaveau, Peter Greiner [35] sur le groupe de Heisenberg non polarisé \mathbb{H} . Nous y reviendrons plus bas. Pour une introduction à ces thèmes, on pourra consulter entre autres Strichartz [516] et Vershik-Gershkovich [553].

2.1. Le groupe de Heisenberg polarisé

2.1.1. L'hamiltonien des géodésiques.

Comme nous l'avons souligné à la section 7 du chapitre 5, les métriques sous-riemanniennes sont très différentes des métriques riemanniennes et leur formulation en termes de théorie du contrôle conduit à des problèmes difficiles. Comme le notent Beals, Gaveau et Greiner ([35], p. 634) à propos de \mathbb{H} :

“The results indicate how complicated a control problem can become, even in the simplest situation”.

En fait ce n'est que relativement tardivement que le problème pour \mathbb{H} a été vraiment clarifié. En 1977, dans un texte fondateur, Bernard Gaveau ([206], p. 114) disait encore :

“Le problème variationnel horizontal associé est le problème de minimiser l'énergie d'une courbe de la variété de base sous la condition de Lagrange que son relèvement horizontal est fixé dans le fibré. Ce problème semble non étudié.”

Considéré comme groupe de Lie, l'espace de 1-jets standard $\mathbb{V}_J = J^1\mathbb{R}^2$ de coordonnées $(x, y, p = \tan(\theta))$ a pour produit, rappelons-le (cf. la section 3 du chapitre 5),

$$(x, y, p) \cdot (x', y', p') = (x + x', y + y' + px', p + p') .$$

Nous avons étudié sa structure de contact \mathcal{K} au chapitre 3.

Remarque. Nous gardons ici les notations de ce chapitre 5. □

On choisit comme métrique sous-riemannienne SR sur la structure de contact \mathcal{K} la métrique invariante (à gauche) de produit scalaire $\langle \bullet, \bullet \rangle$ et de norme $\|\bullet\|$ définie par le fait que la base invariante $\{X_1, X_2\}$ de \mathcal{K} soit une base orthonormée. Rappelons que cette métrique n'est pas la métrique euclidienne $\langle \bullet, \bullet \rangle_E, \|\bullet\|_E$. En effet, à cause de la non-holonomie, la métrique euclidienne n'est pas invariante. Même si $\|X_2\|_E = 1$ et $\langle X_1, X_2 \rangle_E = 0$ comme dans le cas euclidien, $\|X_1\|_E = 1 + p^2 \neq 1$ si $p \neq 0$. C'est seulement sur le plan de base (x, y) que les deux métriques coïncident.

Le problème de calculer les géodésiques de cette métrique SR peut facilement se formuler en termes de théorie du contrôle. Si $\Gamma = \{v(s)\}$ est une courbe lisse paramétrée par s dans \mathbb{V}_J , dire qu'elle est une courbe intégrale de \mathcal{K} (i.e. une relevée legendrienne) revient à dire que

$$\dot{v}(s) = u_1 X_1 + u_2 X_2$$

pour des contrôles appropriés u_1 et u_2 ou, en d'autres termes, que

$$\dot{x} = u_1, \dot{y} = pu_1, \dot{p} = u_2,$$

la condition d'intégrabilité $\frac{\dot{y}}{\dot{x}} = p$ étant automatiquement satisfaite.

Pour trouver les géodésiques, il faut minimiser l'intégrale du lagrangien qu'est l'énergie cinétique $\frac{1}{2} \|\dot{v}\|^2$ le long de ces courbes admissibles. La transformée de Legendre conduit alors à l'hamiltonien

$$\begin{aligned} h(v, \varpi) &= \langle \varpi, \dot{v} \rangle - \frac{1}{2} \|\dot{v}\|^2 \\ &= \varpi(u_1 X_1 + u_2 X_2) - \frac{1}{2} \|u_1 X_1 + u_2 X_2\|^2 \end{aligned}$$

pour $\varpi \in T^*\mathbb{V}_J$. Si $\varpi = \xi^* dx + \eta^* dy + \pi^* dp = (\xi^*, \eta^*, \pi^*)$, alors

$$h(v, \varpi) = \xi^* u_1 + \eta^* u_1 p + \pi^* u_2 - \frac{1}{2} (u_1^2 \|X_1\|^2 + 2u_1 u_2 \langle X_1, X_2 \rangle + u_2^2 \|X_2\|^2)$$

et comme $\|X_1\| = \|X_2\| = 1$ et $\langle X_1, X_2 \rangle = 0$, $h(v, \varpi)$ se simplifie en :

$$h(v, \varpi) = \xi^* u_1 + \eta^* u_1 p + \pi^* u_2 - \frac{1}{2} (u_1^2 + u_2^2).$$

On applique alors, comme pour les elasticæ du chapitre précédent (13), le principe du maximum de Pontryagin PMP exposé à la section 5 du chapitre 12. Nous disposons de la famille d'hamiltoniens $h_u : T^*\mathbb{V}_J \rightarrow \mathbb{R}$ définie par

$$h_u(v, \varpi) = \varpi(u_1 X_1 + u_2 X_2) - \frac{1}{2} (u_1^2 + u_2^2)$$

et nous voulons trouver les contrôles u_j qui les optimisent (et, mieux, les maximisent) pour ensuite résoudre les équations de Hamilton. Ces contrôles optimaux $u_{j, \max}$

satisfont la condition du premier ordre $\frac{\partial h}{\partial u_j} = 0$.¹ Soit $H = \text{Sup}_u (h_u)$ l'hamiltonien $h_{u_{\max}}$. Alors, comme nous l'avons vu, les projections sur \mathbb{V}_J des trajectoires de H sont les géodésiques. Ainsi que l'ont souligné Revaz Gamkrelidze et Andrei Agrachev [9] :

“The maximum principle is a first order optimality condition, an elaboration of the classical Lagrange multipliers rule, where $\varpi(t)$ plays the role of the Lagrange multiplier.”

Les conditions d'optimisation sont

$$\begin{cases} \frac{\partial h_u}{\partial u_1} = \varpi(X_1) - u_1 = 0 \\ \frac{\partial h_u}{\partial u_2} = \varpi(X_2) - u_2 = 0 \end{cases}$$

et par conséquent $u_{1,\max} = \varpi(X_1)$ et $u_{2,\max} = \varpi(X_2)$. D'où

$$\begin{aligned} H(v, \varpi) &= u_1 \varpi(X_1) + u_2 \varpi(X_2) - \frac{1}{2} (u_1^2 + u_2^2) = \frac{1}{2} (u_1^2 + u_2^2) \\ &= \frac{1}{2} (\langle \varpi, X_1 \rangle^2 + \langle \varpi, X_2 \rangle^2) \end{aligned}$$

soit, en termes de coordonnées :

$$H(x, y, p, \xi^*, \eta^*, \pi^*) = \frac{1}{2} [(\xi^* + p\eta^*)^2 + \pi^{*2}] .$$

La version lagrangienne de ce problème variationnel s'obtient par transformée de Legendre

$$L = \langle \varpi, \dot{q} \rangle - H = \xi^* \dot{x} + \eta^* \dot{y} + \pi^* \dot{p} - H .$$

Comme $\|\dot{\Gamma}\| = \dot{x}^2 + \dot{p}^2$ pour toute relevée legendrienne, $\pi^* = \dot{p}$ (et donc $\pi^* \dot{p} = \dot{p}^2$), $\dot{x} = \xi^* + p\eta^*$ et donc $\xi^* \dot{x} = \dot{x}^2 - p\eta^* \dot{x}$, $\omega(\dot{\Gamma}) = -p\dot{x} + \dot{y}$, il est trivial de vérifier que

$$\begin{aligned} L &= (\dot{x}^2 + \dot{p}^2) - p\eta^* \dot{x} + \eta^* \dot{y} - \frac{1}{2} (\dot{x}^2 + \dot{p}^2) \\ &= \frac{1}{2} \|\dot{\Gamma}\| + \eta^* \omega(\dot{\Gamma}) . \end{aligned}$$

Le lagrangien L est donc l'énergie cinétique

$$\frac{1}{2} \|\dot{\Gamma}\| = \frac{1}{2} (\dot{x}^2 + \dot{p}^2)$$

1. Rappelons qu'il faut aussi une condition du deuxième ordre pour garantir que la solution extrémale est bien un maximum (théorie des courbes de Jacobi).

de la projection γ des trajectoires Γ sur le plan (x, p) augmentée de la contrainte

$$\eta^* \omega(\dot{\Gamma})$$

où η^* est le multiplicateur de Lagrange. Si la contrainte $\omega(\dot{\Gamma}) = 0$ est satisfaite, alors $L = H$.

Les équations d'Euler-Lagrange s'écrivent alors (la paramétrisation "temporelle" étant notée s)² :

$$\begin{cases} \frac{d}{ds}(\eta^*) = 0, \text{ i.e. } \dot{\eta}^* = 0 \text{ et } \eta^* = \text{cste} \\ \frac{d}{ds}(\dot{x} - p\eta^*) = 0, \text{ i.e. } \ddot{x} - p\dot{\eta}^* - \dot{p}\eta^* = 0, \text{ i.e. } \ddot{x} = \eta^*\dot{p} \\ \frac{d}{ds}(\dot{p}) = -\eta^*\dot{x}, \text{ i.e. } \ddot{p} = -\eta^*\dot{x} \end{cases}$$

On note que

$$(\ddot{x}, \ddot{p}) = \eta^* (\dot{p}, -\dot{x}) ,$$

ce qui signifie que dans le plan (x, p) l'accélération est proportionnelle à la vitesse tournée de $\frac{\pi}{2}$ (trajectoires circulaires uniformes). Cela permet immédiatement d'intégrer cette vitesse.

2.1.2. Calcul des géodésiques.

Les équations de Hamilton sur $T^*\mathbb{V}_J$ dérivées de cet hamiltonien H redonnent évidemment les mêmes résultats. Elles sont les suivantes :

$$\begin{cases} \dot{x}(s) = \frac{\partial H}{\partial \xi^*} = \xi^* + p\eta^* \\ \dot{y}(s) = \frac{\partial H}{\partial \eta^*} = p(\xi^* + p\eta^*) = p\dot{x}(s) \text{ (i.e. } p = \frac{\dot{y}}{\dot{x}} = \frac{dy}{dx}, \text{ condition d'intégrabilité)} \\ \dot{p}(s) = \frac{\partial H}{\partial \pi^*} = \pi^* \\ \dot{\xi}^*(s) = -\frac{\partial H}{\partial x} = 0 \\ \dot{\eta}^*(s) = -\frac{\partial H}{\partial y} = 0 \\ \dot{\pi}^*(s) = -\frac{\partial H}{\partial p} = -\eta^*(\xi^* + p\eta^*) = -\eta^*\dot{x}(s) . \end{cases}$$

Comme H est indépendant de x et y , les dérivées $\dot{\xi}^*(s) = -\frac{\partial H}{\partial x}$ et $\dot{\eta}^*(s) = -\frac{\partial H}{\partial y}$ s'annulent et les moments ξ^* et η^* sont donc *constants*, $\xi^* = \xi_0^*$ et $\eta^* = \eta_0^*$, le long de chaque géodésique. Ce fait simplifie les équations car dès lors

$$\begin{cases} \dot{x}(s) = \xi_0^* + p\eta_0^* \\ \dot{y}(s) = p(\xi_0^* + p\eta_0^*) \\ \dot{\pi}^*(s) = -\eta_0^*(\xi_0^* + p\eta_0^*) \end{cases}$$

On retrouve bien sûr les relations $\ddot{p} = \dot{\pi}^* = -\eta^*\dot{x}$ et $\ddot{x} = \eta^*\dot{p}$, soit

$$(\ddot{x}, \ddot{p}) = \eta_0^* (\dot{p}, -\dot{x}) ,$$

2. Les géodésiques de \mathbb{V}_J sont des courbes paramétrées spatialement que l'on traite intuitivement comme des trajectoires paramétrées temporellement.

qui montrent que dans le plan (x, p) l'accélération est orthogonale à la vitesse et que les géodésiques sont des *cercles* dont le rayon croît quand η_0^* décroît (à la limite $\eta_0^* = 0$ le cercle devient une ligne droite). D'ailleurs,

$$H(x, y, p, \xi^*, \eta^*, \pi^*) = \frac{1}{2} (\dot{x}^2 + \dot{p}^2)$$

et l'hamiltonien H est par conséquent l'énergie cinétique de la projection des trajectoires sur le plan (x, p) .

Pour structurer les calculs, nous introduisons, comme Beals, Gaveau et Greiner l'ont fait dans [35] pour le groupe de Heisenberg non polarisé (cf. la section suivante), la variable complexe $z = (x, p)$ dans le plan de contact K_0 , le moment complexe $\chi^* = (\xi^*, \pi^*)$, et la nouvelle variable

$$\zeta^* = (\xi^* + p\eta^*, \pi^*) = \chi^* + \eta^* \Lambda(z) = (\dot{x}, \dot{p}) ,$$

avec

$$\begin{aligned} \Lambda &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \Lambda^T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ -\Lambda^T + \Lambda &= J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ J^2 &= -\mathbf{1} . \end{aligned}$$

Nous vérifions que $2H$ est la norme de ζ^* :

$$H(z, y, \chi^*, \eta^*) = \frac{1}{2} \langle \zeta^*, \zeta^* \rangle = \frac{1}{2} \|\zeta^*\|^2 = H_0 = \frac{1}{2} \langle \zeta^*(0), \zeta^*(0) \rangle = \frac{1}{2} \|\zeta^*(0)\|^2$$

et que les équations de Hamilton deviennent :

$$\begin{cases} \dot{z}(s) = \frac{\partial H}{\partial \chi^*} = \zeta^*(s) \\ \dot{y}(s) = \frac{\partial H}{\partial \eta^*} = p(\xi^* + p\eta^*) = \langle \zeta^*, \Lambda(z) \rangle \\ \dot{\chi}^*(s) = -\frac{\partial H}{\partial z} = (0, -\eta^*(\xi^* + p\eta^*)) = -\eta^* \Lambda^T(\zeta^*) \\ \dot{\eta}^*(s) = 0, \eta^* = \text{cst} = \eta_0^* \end{cases}$$

L'équation différentielle pour ζ^* est :

$$\begin{aligned} \dot{\zeta}^*(s) &= \dot{\chi}^*(s) + \eta^* \Lambda \dot{z}(s) \text{ (puisque } \eta^* = \text{cste)} \\ &= -\eta^* \Lambda^T \zeta^*(s) + \eta^* \Lambda \zeta^*(s) = \eta^* J \zeta^*(s) \\ (\ddot{x}, \ddot{p}) &= \eta^* J(\dot{x}, \dot{p}) , \end{aligned}$$

J étant la multiplication complexe par $-i$ dans le plan complexe des z . L'exponentielle étant, comme Λ , une matrice 2×2 opérant sur $\zeta^*(0)$, la solution est :

$$\varsigma^*(s) = e^{s\eta^* J} \varsigma^*(0) ,$$

ce qui peut aussi s'écrire $\varsigma^*(s) = e^{-is\eta^*} \varsigma^*(0)$. Cette intégration de ς^* fournit des formules explicites pour les géodésiques.

Pour $\eta^* = 0$, on a $\dot{\varsigma}^* = \dot{\chi}^* = \mathbf{0}$ et par conséquent $\chi^* = \chi_0^* = (\xi_0^*, \pi_0^*)$ et $\varsigma^* = \varsigma_0^*$ sont constants. La variable

$$z = s\chi_0^* + z_0$$

se déplace le long d'une ligne droite dans le plan (x, p) et l'on obtient :

$$\begin{cases} x(s) = s\xi_0^* + x_0 \\ p(s) = s\pi_0^* + p_0 \\ \dot{y}(s) = \langle \varsigma^*, \mathbf{\Lambda}(z) \rangle = \langle \varsigma_0^*, \mathbf{\Lambda}(s\chi_0^* + z_0) \rangle = s\xi_0^* \pi_0^* + \xi_0^* p_0 \\ y(s) = \frac{1}{2} s^2 \xi_0^* \pi_0^* + s\xi_0^* p_0 + y_0 \end{cases}$$

Si $z_0 = 0$, alors $z = s\chi_0^*$ et

$$y(s) = \frac{1}{2} s^2 \xi_0^* \pi_0^* + y_0 .$$

Si de plus $y_0 = 0$, alors $y(s) = \frac{1}{2} s^2 \xi_0^* \pi_0^*$.

Pour $\eta^* \neq 0$ ($\eta^* = \text{cst} = \eta_0^*$), la situation est bien sûr plus compliquée. Supposons que $z_0 = 0$. Alors $\dot{z}(s) = \varsigma^*(s)$ donne

$$\begin{aligned} z(s) &= \int_0^s \varsigma^*(r) dr = \int_0^s e^{-ir\eta_0^*} \varsigma^*(0) dr \\ &= \frac{\varsigma^*(0)}{(-i\eta_0^*)} (e^{-is\eta_0^*} - 1) \\ &= \frac{i}{\eta_0^*} (\varsigma^*(s) - \varsigma^*(0)) . \end{aligned}$$

Cette valeur fournit des formules explicites pour les géodésiques connectant en un temps τ une origine $(x_0 = 0, y_0, p_0 = 0)$ à une extrémité

$$\{x_1 = x(\tau), y_1 = y(\tau), p_1 = p(\tau)\} .$$

En effet,

$$z(\tau) = z_1 = \frac{i}{\eta_0^*} \varsigma^*(0) (e^{-i\tau\eta_0^*} - 1)$$

donne $\varsigma^*(0)$ et puis $\chi_0^* = (\xi_0^*, \pi_0^*)$. Si $e^{-i\tau\eta_0^*} - 1 \neq 1$, i.e. si $\tau\eta_0^* \neq 2k\pi$ (si $\tau\eta_0^* = 2k\pi$ alors $z(\tau) = z_0$), on obtient :

$$\begin{aligned} z(s) &= \frac{e^{-is\eta_0^*} - 1}{e^{-i\tau\eta_0^*} - 1} z_1 = \frac{e^{-i\frac{s}{2}\eta_0^*} (e^{-i\frac{s}{2}\eta_0^*} - e^{i\frac{s}{2}\eta_0^*})}{e^{-i\frac{\tau}{2}\eta_0^*} (e^{-i\frac{\tau}{2}\eta_0^*} - e^{i\frac{\tau}{2}\eta_0^*})} z_1 \\ &= e^{i\frac{(\tau-s)}{2}\eta_0^*} \frac{\sin\left(\frac{s}{2}\eta_0^*\right)}{\sin\left(\frac{\tau}{2}\eta_0^*\right)} z_1 . \end{aligned}$$

En d'autres termes, on obtient pour la partie (x, p) des géodésiques les formules :

$$\begin{cases} x(s) = \frac{\sin\left(\frac{s}{2}\eta_0^*\right)}{\sin\left(\frac{\tau}{2}\eta_0^*\right)} \left(\cos\left(\frac{(\tau-s)}{2}\eta_0^*\right) x_1 - \sin\left(\frac{(\tau-s)}{2}\eta_0^*\right) p_1 \right) \\ p(s) = \frac{\sin\left(\frac{s}{2}\eta_0^*\right)}{\sin\left(\frac{\tau}{2}\eta_0^*\right)} \left(\sin\left(\frac{(\tau-s)}{2}\eta_0^*\right) x_1 + \cos\left(\frac{(\tau-s)}{2}\eta_0^*\right) p_1 \right) \end{cases}$$

qui sont les équations d'un cercle dans le plan des $z = (x, p)$. En effet, on a :

$$x^2 + p^2 - x \left(x_1 + p_1 \cot\left(\frac{\eta_0^*\tau}{2}\right) \right) - p \left(p_1 - x_1 \cot\left(\frac{\eta_0^*\tau}{2}\right) \right) = 0$$

qui est l'équation d'un cercle passant par 0 et z_1 , de centre

$$x_c = \frac{1}{2} \left(x_1 + p_1 \cot\left(\frac{\eta_0^*\tau}{2}\right) \right), p_c = \frac{1}{2} \left(p_1 - x_1 \cot\left(\frac{\eta_0^*\tau}{2}\right) \right)$$

et de rayon

$$r^2 = \frac{1}{4} (x_1^2 + p_1^2) \left(1 + \cot\left(\frac{\eta_0^*\tau}{2}\right) \right) = \frac{1}{4 \sin^2\left(\frac{\eta_0^*\tau}{2}\right)} |z_1|^2 .$$

On notera que l'égalité $H = H_0 = \frac{1}{2} \langle \zeta^*(0), \zeta^*(0) \rangle$ implique que la valeur constante de l'hamiltonien le long d'une trajectoire est :

$$H_0 = \frac{1}{2} \|\zeta^*(0)\|^2 = \frac{\eta_0^{*2}}{8 \sin^2\left(\frac{\eta_0^*\tau}{2}\right)} |z_1|^2 = \frac{\eta_0^{*2}}{2} r^2 .$$

En effet,

$$z_1 = \frac{i}{\eta_0^*} e^{-i\frac{\tau}{2}\eta_0^*} (e^{-i\frac{\tau}{2}\eta_0^*} - e^{i\frac{\tau}{2}\eta_0^*}) \zeta^*(0) = \frac{2}{\eta_0^*} e^{-i\frac{\tau}{2}\eta_0^*} \sin\left(\frac{\tau}{2}\eta_0^*\right) \zeta^*(0) .$$

Pour la composante $y(s)$, les calculs sont plus compliqués. On a

$$\dot{y}(s) = \langle \zeta^*, \mathbf{\Lambda}(z) \rangle = p(\xi^* + p\eta^*)$$

et nous connaissons déjà $\zeta^*(s)$ et $z(s)$. En utilisant les valeurs de x_1, p_1, η_0^* et τ ainsi que les intégrales classiques :

$$\left\{ \begin{array}{l} \int_0^s \cos(\eta^* r) dr = \frac{\sin(\eta^* s)}{\eta^*} \\ \int_0^s \sin(\eta^* r) dr = \frac{1 - \cos(\eta^* s)}{\eta^*} \\ \int_0^s \cos^2(\eta^* r) dr = \frac{2\eta^* s + \sin(2\eta^* s)}{4\eta^*} \\ \int_0^s \sin^2(\eta^* r) dr = \frac{2\eta^* s - \sin(2\eta^* s)}{4\eta^*} \\ \int_0^s \cos(\eta^* r) \sin(\eta^* r) dr = \frac{\sin^2(\eta^* s)}{2\eta^*} \end{array} \right.$$

on obtient

$$\begin{aligned} y(s) - y_0 = & \frac{1}{8(\cos(\eta_0^* \tau) - 1)} [-2\eta_0^* s (x_1^2 + p_1^2) - 4x_1 p_1 \cos(\eta_0^* (s - \tau)) \\ & + 2(x_1^2 - p_1^2) \sin(\eta_0^* (s - \tau)) \\ & + 2x_1 p_1 \cos(\eta_0^* (2s - \tau)) - (x_1^2 - p_1^2) \sin(\eta_0^* (2s - \tau)) \\ & + 2x_1 p_1 \cos(\eta_0^* \tau) + (x_1^2 - p_1^2) \sin(\eta_0^* \tau) \\ & + 2(x_1^2 + p_1^2) \sin(\eta_0^* s)] . \end{aligned}$$

En termes de ξ_0^* , π_0^* , η_0^* et τ , la formule s'écrit :

$$\begin{aligned} y(s) - y_0 = & \xi_0^{*2} \frac{2\eta_0^* s + \sin(2\eta_0^* s)}{4\eta_0^{*2}} - \xi_0^{*2} \frac{\sin(\eta_0^* s)}{\eta_0^{*2}} + \xi_0^* \pi_0^* \frac{\sin^2(\eta_0^* s)}{\eta_0^{*2}} \\ & - \xi_0^* \pi_0^* \frac{1 - \cos(\eta_0^* s)}{\eta_0^{*2}} + \pi_0^{*2} \frac{2\eta_0^* s - \sin(2\eta_0^* s)}{4\eta_0^{*2}} \end{aligned}$$

et l'on note que si l'on développe cette expression à l'ordre 3 on retrouve bien la limite

$$y(s) - y_0 = \frac{1}{2} s^2 \xi_0^* \pi_0^*$$

vue ci-dessus quand $\eta_0^* \rightarrow 0$.

2.1.3. La multiplicité des géodésiques.

Ces équations expliquent l'origine de l'étonnante *multiplicité* des géodésiques sous-riemanniennes connectant deux points. En effet, calculons $y_1 - y_0 = y(\tau) - y_0$. Comme

$$\cos(\eta_0^* \tau) - 1 = -2 \sin^2 \left(\frac{\eta_0^* \tau}{2} \right) ,$$

nous trouvons

$$\begin{aligned}
y_1 - y_0 &= \frac{-1}{16 \sin^2 \left(\frac{\eta_0^* \tau}{2} \right)} [-2\eta_0^* \tau (x_1^2 + p_1^2) - 4x_1 p_1 \\
&\quad + 2x_1 p_1 \cos(\eta_0^* \tau) - (x_1^2 - p_1^2) \sin(\eta_0^* \tau) \\
&\quad + 2x_1 p_1 \cos(\eta_0^* \tau) + (x_1^2 - p_1^2) \sin(\eta_0^* \tau) \\
&\quad + 2(x_1^2 + p_1^2) \sin(\eta_0^* \tau)] \\
&= \frac{1}{8 \sin^2 \left(\frac{\eta_0^* \tau}{2} \right)} [2x_1 p_1 (1 - \cos(\eta_0^* \tau)) + (x_1^2 + p_1^2) (\eta_0^* \tau - \sin(\eta_0^* \tau))] \\
&= \frac{1}{2} x_1 p_1 + \frac{x_1^2 + p_1^2}{4} \left[\frac{\left(\frac{\eta_0^* \tau}{2} \right)}{\sin^2 \left(\frac{\eta_0^* \tau}{2} \right)} - \frac{\cos \left(\frac{\eta_0^* \tau}{2} \right)}{\sin \left(\frac{\eta_0^* \tau}{2} \right)} \right].
\end{aligned}$$

Si nous introduisons la nouvelle variable $\varphi = \frac{\eta_0^* \tau}{2}$, nous voyons que nous devons résoudre l'équation

$$4 \left(y_1 - y_0 - \frac{1}{2} x_1 p_1 \right) = \mu(\varphi) |z_1|^2$$

où $\mu(\varphi)$ est la fonction

$$\mu(\varphi) = \frac{\varphi}{\sin^2(\varphi)} - \cot(\varphi).$$

C'est la fonction $\mu(\varphi)$ qui est la clé de l'étrange comportement des géodésiques sous-riemanniennes. Il s'agit d'une fonction impaire qui diverge pour $\varphi = k\pi$ ($k \neq 0$) (i.e. $\eta_0^* \tau = 2k\pi$, nous avons déjà rencontré cette condition) et présente des points critiques quand $\varphi = \tan(\varphi)$. Mais lorsque $\varphi = \tan(\varphi)$, on a

$$\begin{aligned}
\mu(\varphi) &= \frac{\tan(\varphi)}{\sin^2(\varphi)} - \cot(\varphi) \\
&= \frac{1 - \cos^2(\varphi)}{\cos(\varphi) \sin(\varphi)} = \tan(\varphi) = \varphi
\end{aligned}$$

et les minima de $\mu(\varphi)$ sont sur la diagonale. Le graphe de $\mu(\varphi)$ est représenté à la figure 1.

Pour obtenir les géodésiques connectant une origine $(0, y_0, 0)$ à une même extrémité $\{x_1 = x(\tau), y_1 = y(\tau), p_1 = p(\tau)\}$, nous devons résoudre l'équation

$$4 \left(y_1 - y_0 - \frac{1}{2} x_1 p_1 \right) = \mu(\varphi) \|z_1\|^2$$

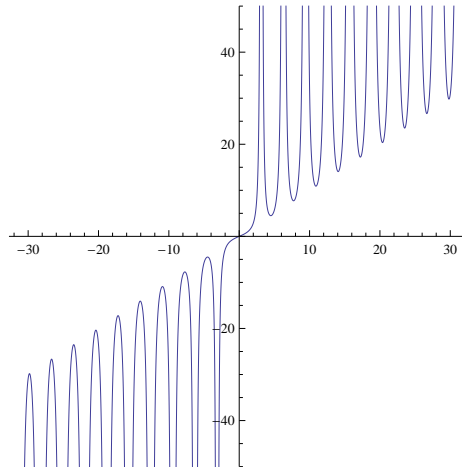


FIGURE 1. La fonction $\mu(\varphi)$ intervenant dans la construction des géodésiques sous-riemanniennes du groupe de Heisenberg polarisé \mathbb{V}_J (les échelles des deux axes ne sont pas les mêmes).

en $\varphi = \frac{\eta_0^* \tau}{2}$ (on peut prendre $\eta_0^* = 2$ et $\tau = \varphi$). Par exemple, pour

$$\{y_0 = 0, x_1 = 2, p_1 = 4, y_1 = 104\} ,$$

nous trouvons $\mu(\varphi) = 20$, ce qui donne 11 solutions :

$$\left\{ \begin{array}{l} \varphi_1 = 2.737, \varphi_2 = 3.552, \varphi_3 = 5.695, \varphi_4 = 6.885, \varphi_5 = 8.680, \varphi_6 = 10.195, \\ \varphi_7 = 11.672, \varphi_8 = 13.506, \varphi_9 = 14.656, \varphi_{10} = 16.845, \varphi_{11} = 17.608 . \end{array} \right\}$$

Nous donnerons d'autres exemples dans la section suivante pour le groupe de Heisenberg non polarisé \mathbb{H} .

2.1.4. La sphère et le front d'onde sous-riemanniens.

Cette structure étonnante des géodésiques implique que la sphère sous-riemannienne S (le lieu des extrémités des géodésiques de longueur sous-riemannienne = 1 partant de 0 et qui sont des minimiseurs *globaux*) et, surtout, le front d'onde W (le lieu des extrémités des géodésiques de longueur sous-riemannienne = 1 partant de 0 et qui ne sont pas forcément des minimiseurs globaux) ont une géométrie assez étrange. Nous allons les calculer explicitement en adaptant les calculs effectués par Richard Beals, Bernard Gaveau, et Peter Greiner dans [35] pour \mathbb{H} , calculs sur lesquels, comme nous l'avons dit, nous reviendrons dans la section suivante. Nous allons voir qu'une propriété fondamentale de cette géométrie, comparée à la géométrie euclidienne, est la complexité du *cut locus* de 0 (c'est-à-dire le lieu des points où les géodésiques partant de 0 s'arrêtent d'être des minimiseurs globaux), ainsi que celle du lieu des points *conjugés* de 0 (c'est-à-dire le lieu des points où les géodésiques partant de 0 perdent leur optimalité locale et deviennent instables,

il s'agit du lieu singulier de l'application exponentielle \mathcal{E} intégrant les géodésiques). Nous avons déjà rencontré ce phénomène au chapitre précédent (13) avec les *elasticæ*.

Pour représenter S et W , il faut d'abord calculer la longueur des géodésiques. Soit donc $\Gamma(s)$ une géodésique partant de 0 au temps $s = 0$ et aboutissant au temps $s = \tau$ au point $(x_1, p_1, y_1) = (z_1, y_1)$. Sa longueur est $L = \int_0^\tau \ell ds$ avec

$$\ell^2 = (\xi^* + p\eta^*)^2 + \pi^{*2}$$

le carré de la norme de $\dot{\Gamma}$ dans le plan de contact muni de sa base orthonormale $\{X_1 = \partial_x + p\partial_y, X_2 = \partial_p\}$. Mais $\ell^2 = 2H = 2H_0$ puisque l'hamiltonien est constant le long des trajectoires, et nous savons que

$$H_0 = \frac{\eta_0^{*2}}{8} \frac{1}{\sin^2\left(\frac{\eta_0^*\tau}{2}\right)} |z_1|^2 .$$

Donc, avec $\frac{\eta_0^*\tau}{2} = \varphi$,

$$L = \left(\frac{\eta_0^*\tau}{2}\right) \frac{1}{\left|\sin\left(\frac{\eta_0^*\tau}{2}\right)\right|} |z_1| = \frac{\varphi}{|\sin(\varphi)|} |z_1| .$$

La sphère S et le front d'onde W sont donc donnés par l'équation fondamentale

$$|z_1| = \frac{|\sin(\varphi)|}{\varphi} .$$

D'où les équations :

$$\begin{cases} x_1 = \frac{|\sin(\varphi)|}{\varphi} \cos(\theta) \\ p_1 = \frac{|\sin(\varphi)|}{\varphi} \sin(\theta) \\ y_1 = \frac{1}{2}x_1p_1 + \frac{\varphi - \sin(\varphi)\cos(\varphi)}{4\varphi^2} \\ = \frac{1}{2} \frac{\sin^2(\varphi)}{\varphi^2} \cos(\theta)\sin(\theta) + \frac{\varphi - \cos(\varphi)\sin(\varphi)}{4\varphi^2} \\ = \frac{\varphi + 2\sin^2(\varphi)\cos(\theta)\sin(\theta) - \cos(\varphi)\sin(\varphi)}{4\varphi^2} \end{cases}$$

Pour $\varphi = 0^+$, $\{x_1, p_1\}$ décrit le cercle unité $x_1 = \cos(\theta)$, $p_1 = \sin(\theta)$ et y_1 décrit au-dessus la courbe

$$y_1 = \frac{1}{2} \cos(\theta)\sin(\theta)$$

comme le montre la figure 2 gauche. C'est l'équateur de la sphère unité S . La même figure 2 à droite montre un quart du profil W_0 du front d'onde W pour $\theta = 0$

3. On utilise les développements limités $\sin(\varphi) \sim \varphi - \frac{\varphi^3}{3!}$, $\cos(\varphi) \sim 1 - \frac{\varphi^2}{2!}$, $\frac{\sin^2(\varphi)}{\varphi^2} \sim 1 - \frac{\varphi^2}{3}$, $\varphi - \cos(\varphi)\sin(\varphi) \sim \frac{\varphi^3}{6}$.

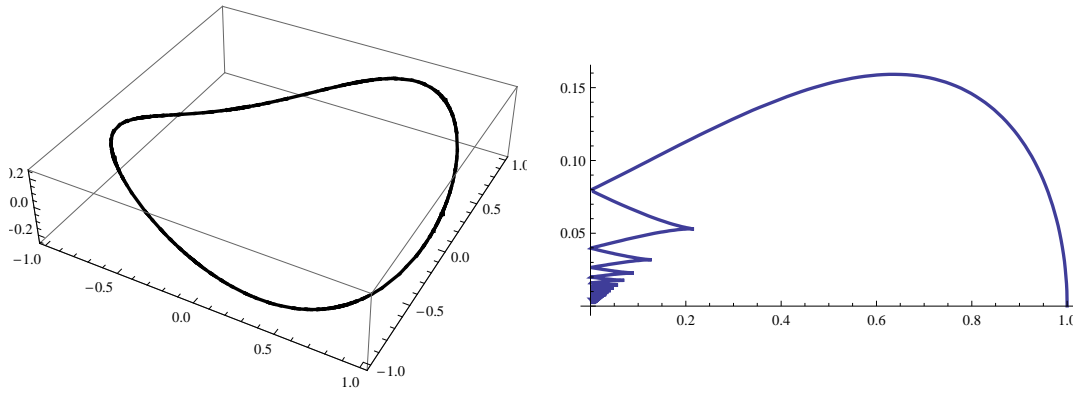


FIGURE 2. À gauche : la courbe $x_1 = \cos(\theta), p_1 = \sin(\theta), y_1 = \frac{1}{2} \cos(\theta) \sin(\theta)$ qui est l'équateur $\varphi = 0$ de la sphère sous-riemannienne de \mathbb{V}_J de rayon 1. À droite : un quart du profil W_0 du front d'onde W pour $\theta = 0$ d'équations $x_1 = \frac{|\sin(\varphi)|}{\varphi}, p_1 = 0, y_1 = \frac{\varphi - \cos(\varphi) \sin(\varphi)}{4\varphi^2}$.

d'équations paramétriques

$$\left\{ x_1 = \frac{|\sin(\varphi)|}{\varphi}, p_1 = 0, y_1 = \frac{\varphi - \cos(\varphi) \sin(\varphi)}{4\varphi^2} \right\} .$$

L'arc supérieur part de $\{1, 0, 0\}$ et arrive sur l'axe vertical des y en $y^1 = \frac{1}{4\pi} \sim 0.785$ pour $\varphi = \pi$ et il y revient aux points $y^k = \frac{1}{4k^2\pi}$ pour $\varphi = k\pi$. Les points cusp du profil $\dot{x} = \dot{y} = 0$ correspondent à la contrainte $\sin(\varphi) - \varphi \cos(\varphi) = 0$ et sont donc sur la courbe C d'équations

$$\left\{ x = \cos(\varphi), y = \frac{1}{4} \cos(\varphi) \sin(\varphi) \right\} .$$

La figure 3 représente C .

Nous montrons aux figures 4 et 5 des images 3D de S et W . La surface externe est la sphère sous-riemannienne S . La partie interne est $W - S$. Elle présente des cercles de plus en plus petits de singularités cusp qui s'accumulent vers 0.⁴ Un comportement aussi complexe est impossible en géométrie riemannienne.

2.1.5. *Singularités des sphères sous-riemanniennes.*

La différence spectaculaire entre la géométrie sous-riemannienne de \mathbb{V}_J et une géométrie riemannienne est que les sphères SR possèdent des singularités et que celles des fronts d'onde s'accumulent même sur le point origine. Le fait que les

4. Un peu comme des disques de discoboles de rayon décroissant dont le bord serait une arête de cusps et le centre serait renflé, les disques étant empilés les uns sur les autres. Nous retrouverons cette structure, mais un peu déformée, dans la section suivante 4.

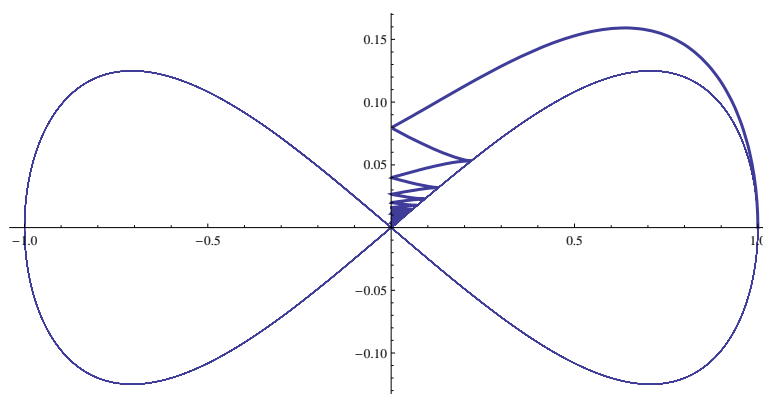


FIGURE 3. La courbe C d'équations $x = \cos(\varphi)$, $y = \frac{1}{4} \cos(\varphi) \sin(\varphi)$ sur laquelle se situent les points cusp du profil W_0 .

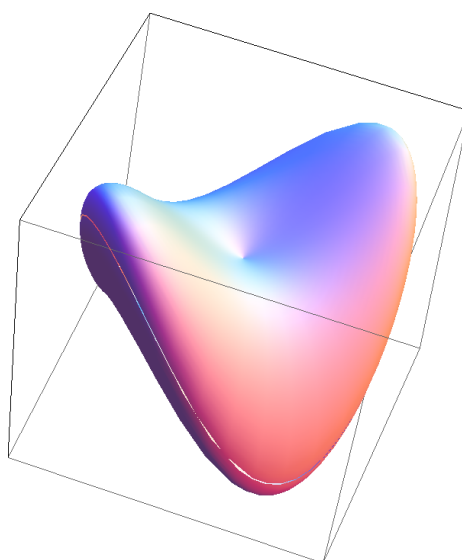


FIGURE 4. La sphère sous-riemannienne S . Elle a une forme de selle avec des singularités aux intersections avec l'axe des y . La ligne blanche correspond à l'équateur $\varphi = 0$.

sphères possèdent des singularités est *général* en géométrie SR et peut être démontré de façon très conceptuelle à partir de tout un ensemble de lemmes techniques. Le lecteur intéressé pourra à ce sujet consulter le chapitre 11 de la “bible” Agrachev-Barilari-Boscain [5]. Les idées directrices sont les suivantes (on utilise les notations générales des chapitres précédents).

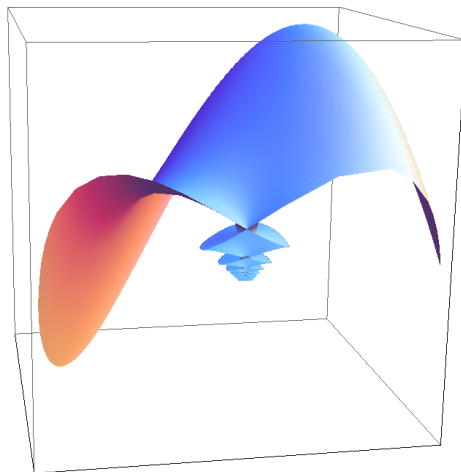


FIGURE 5. Une partie du front d'onde sous-riemannien W . La surface externe est la sphère sous-riemannienne S . La partie interne est $W - S$. Elle présente des cercles de plus en plus petits de singularités cusp qui s'accroissent vers 0.

Le point origine est $q_0 \in M$ et les géodésiques sont les projections des trajectoires du champ hamiltonien \vec{H} définies par l'exponentielle

$$\text{Exp}_t : T_{q_0}^* M \rightarrow M, \lambda_0 \mapsto \pi \circ e^{t\vec{H}}(\lambda_0) = q(t) .$$

On suppose que l'on est dans une "bonne situation" où M est complète et où les extrémales γ sont strictement normales ("strictement" signifie que tous les segments de γ sont normaux). Pour faire plus facilement le lien entre l'hamiltonien et la distance on considère la fonction

$$\Delta(q) = \frac{1}{2}d^2(q_0, q) .$$

Les sphères et les boules fermées sont compactes et les sphères S_r sont les surfaces de niveau de Δ . On regarde alors les propriétés de différentiabilité de Δ . Si Δ est localement C^∞ en tout point de S_r ,⁵ alors S_r est une hypersurface C^∞ de codimension 1, fermée et compacte, sans bord, de M . Si S_r possède des singularités, c'est donc que Δ n'est pas partout C^∞ sur S_r .

On montre que la différentiabilité de Δ en q implique qu'il existe une seule géodésique $\Gamma : q_0 \rightarrow q$ (strictement normale) et que q n'est pas un point conjugué de q_0 le long de Γ . Si Δ est C^∞ au voisinage de q , sa différentielle en q est un élément $d_q\Delta \in T_q^*M$ que l'on peut ramener en arrière dans $T_{q_0}^*M$ par $e^{-t\vec{H}}$ si $q = \Gamma(t)$. Étant donné le choix de Δ , $d_q\Delta = \lambda(t)$ où $\lambda(t)$ est la trajectoire de H dans T^*M

5. La restriction de Δ à S_r est C^∞ puisqu'elle est constante et égale à $\frac{1}{2}r^2$. La condition est que Δ est C^∞ dans un petit voisinage de tout point de S_r .

se projetant sur Γ dans M . Comme $\lambda(t) \neq 0$ par hypothèse de “bonne situation”, $d_q\Delta \neq 0$. On obtient ainsi une application Φ inverse de Exp ,

$$\Phi : S_r \rightarrow T_{q_0}^*M, q \mapsto e^{-t\vec{H}}(d_q\Delta)$$

qui associe à q le λ_0 de l'unique géodésique $\Gamma : q_0 \rightarrow q$ telle que $q = \Gamma(t)$. Par construction $H(\Phi(q)) = \frac{1}{2}r^2$ et donc

$$\Phi : S_r \rightarrow (T_{q_0}^*M) \cap H^{-1}\left(\frac{1}{2}r^2\right),$$

autrement dit

$$\Phi : S_r \rightarrow (T_{q_0}^*M) \cap C_r$$

avec

$$C_r = \left\{ \lambda \in T_{q_0}^*M \mid H(\lambda) = \frac{1}{2}r^2 \right\} = H^{-1}\left(\frac{1}{2}r^2\right).$$

S_r et C_r sont de même dimension $n - 1$ et C_r est C^∞ . On montre que Φ est C^∞ et que, comme $d\Phi$ est injective, elle donc est une *immersion* de S_r dans $T_{q_0}^*M$. L'argumentation repose alors sur les étapes suivante (cf. [5]) :

1. Comme S_r est compacte et Φ est continue, $\Phi(S_r)$ est *compact* et donc fermée dans C_r , ce qui implique que Φ soit un *plongement* de S_r compacte dans une sous-variété C^∞ C_r de même dimension ($n - 1$) dans $T_{q_0}^*M$ (de dimension n).
2. Comme $H = \frac{1}{2} \sum_{j=1}^m \langle \lambda, X_j \rangle^2$, on montre que C_r est diffeomorphe à $\mathbb{S}^{m-1} \times \mathbb{R}^{n-m}$. C'est là le point crucial. Dans le cas sous-riemannien $m < n$ strictement et donc C_r a un facteur \mathbb{R}^{n-m} qui l'empêche d'être compact.
3. La conclusion découle alors de la preuve que $\Phi(S_r)$ (compact) = C_r (non compact). Ce point utilise le fait que comme Φ est un plongement et que les dimensions sont les mêmes (i.e. $n - 1$), son image $\Phi(S_r)$ ne peut pas avoir de bord et est ouverte. Comme elle est également fermée puisque compacte, elle est une composante connexe de C_r et comme C_r est connexe, $\Phi(S_r) = C_r$.

On a par conséquent une bonne régularité de Δ sur l'ensemble \mathfrak{U} des points q atteignables à partir de q_0 par des géodésiques Γ strictement normales, q n'étant pas conjugué de q_0 le long de Γ . On peut montrer que \mathfrak{U} est un *ouvert dense* de M , que Δ est en fait C^∞ sur tout \mathfrak{U} et que donc la “bonne situation” régulière est générique. A priori, le fermé singulier $\mathfrak{S} = M - \mathfrak{U}$ peut être compliqué et contient q_0 car Δ admet des singularités sur toutes les sphères S_r lorsque $r \rightarrow 0$. Mais la structure singulière de Δ n'est quand même pas trop sauvage car on peut montrer que, malgré ses singularités, Δ reste néanmoins *lipschitzienne*, autrement dit qu'il existe une constante K telle que, sur la boule $B(q_0, r)$, on ait la limitation

$$|\Delta(q_1) - \Delta(q_2)| \leq Kd(q_1, q_2) .$$

D'après un célèbre théorème de Hans Rademacher, cela implique que Δ est différentiable *presque partout*, le lieu singulier $\mathfrak{S} = M - \mathfrak{U}$ étant de mesure nulle.⁶

2.2. Le groupe de Heisenberg non polarisé (Beals, Gaveau, Greiner)

2.2.1. Le groupe \mathbb{H} .

Nous avons vu au chapitre 5 que le groupe de Heisenberg polarisé \mathbb{V}_J est isomorphe au groupe de Heisenberg

$$\mathbb{H} = \mathbb{R}^2 \times \mathbb{R} = \mathbb{C} \times \mathbb{R} ,$$

qui a la même algèbre de Lie. Le changement de variable adopté ici est

$$\{x = x_1, y = t + 2x_1x_2, p = 4x_2\}$$

et si $q = (x_1, x_2, t) = (z, t)$ sont des coordonnées, la loi de groupe de \mathbb{H} s'écrit :⁷

$$\begin{aligned} (x_1, x_2, t) \cdot (x'_1, x'_2, t') &= (x_1 + x'_1, x_2 + x'_2, t + t' + 2(x_2x'_1 - x_1x'_2)) \text{ i.e.} \\ (z, t)(z', t') &= (z + z', t + t' + 2\Omega(z, z')) \end{aligned}$$

où $\Omega(z, z')$ est la 2-forme symplectique standard sur le plan des z .

Dans l'espace de base \mathbb{C} , cette loi est la translation $z + z'$, dans la fibre \mathbb{R} elle est la translation augmentée du couplage $2(x_2x'_1 - x_1x'_2)$ qui peut être écrit

$$-2z \wedge z' = -2 \det \begin{pmatrix} x_1 & x'_1 \\ x_2 & x'_2 \end{pmatrix} = \langle \Lambda z, z' \rangle$$

(avec $\Lambda = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$ la multiplication par $-2i$ et $\langle \bullet, \bullet \rangle$ le produit scalaire standard), c'est-à-dire $2 \operatorname{Im}(z\bar{z}')$.

Dans cette section, nous revenons un instant sur les calculs de Beals-Gaveau-Greiner sur \mathbb{H} , calculs que nous avons adaptés à \mathbb{V}_J dans la section précédente. Ce sera d'ailleurs une occasion de dire quelques mots supplémentaires sur \mathbb{H} . Une des différences principales entre le cas \mathbb{H} et le cas $\mathbb{V}_J = \mathbb{H}_{pol}$ est que pour \mathbb{H} la matrice Λ est antisymétrique et que $-\Lambda^T + \Lambda = 2\Lambda$.

La structure de contact de \mathbb{H} translate le plan de contact à l'origine K_0 , c'est-à-dire le plan \mathbb{C} engendré par $X_1 = (1, 0, 0)$, $X_2 = (0, 1, 0)$. Les vecteurs tangents X_1 et X_2 engendrant les plans de contact K_q sont donc :

$$\{X_1 = (1, 0, 2x_2), X_2 = (0, 1, -2x_1)\} .$$

6. C'est un avatar du fameux "lemme de Sard" omniprésent dans la théorie des singularités d'applications différentiables (cf. par exemple [412]).

7. Avec le changement de variables utilisé, le facteur de $(x_2x'_1 - x_1x'_2)$ est 2 au lieu du plus classique $\frac{1}{2}$. Nous avons parlé de cette variante à la section 5.1 du chapitre 5. Cela simplifie un peu les formules plus bas.

Si l'on interprète les champs de vecteurs comme dérivations sur les $f : \mathbb{H} \rightarrow \mathbb{R}$, on a

$$\begin{cases} X_1 = \frac{\partial}{\partial x_1} + 2x_2 \frac{\partial}{\partial t} \\ X_2 = \frac{\partial}{\partial x_2} - 2x_1 \frac{\partial}{\partial t} \end{cases}$$

Si $T = (0, 0, 1)$ est le vecteur unité "vertical", on a la structure d'algèbre de Lie nilpotente de $\mathcal{H} = T_0\mathbb{H}$:

$$[X_1, X_2] = -4T, [X_1, T] = 0, [X_2, T] = 0$$

et tous les commutateurs de niveau 2 sont donc nuls. La forme de contact est

$$\omega = -2x_2 dx_1 + 2x_1 dx_2 + dt,$$

et s'écrit en coordonnées cylindriques $(r \cos(\varphi), r \sin(\varphi), t)$ sous la forme

$$\omega = 2r^2 d\varphi + dt.$$

On a

$$d\omega = 4dx_1 \wedge dx_2 = 4\Omega$$

et

$$\omega \wedge d\omega = 4dt \wedge dx_1 \wedge dx_2$$

est une forme volume, le champ de plans de contact $\mathcal{K} : v \rightarrow K_q$ étant maximale non intégrable.

On notera que la loi de composition de \mathbb{H} vient simplement du fait que le crochet de Lie dans l'algèbre de Lie \mathcal{H} de \mathbb{H} est donné, pour deux vecteurs (ς, τ) et (ς', τ') de \mathcal{H} , par

$$[(\varsigma, \tau), (\varsigma', \tau')] = (0, 4\Omega(\varsigma, \varsigma')) .$$

Si donc $(z, t) = \exp((\varsigma, \tau))$, on a d'après la formule de Baker-Campbell-Hausdorff (car le groupe est nilpotent avec seulement des commutateurs de niveau 1 non nuls, cf. le chapitre 5)

$$\begin{aligned} (z, t)(z', t') &= e^{(\varsigma, \tau)} e^{(\varsigma', \tau')} = e^{(\varsigma, \tau) + (\varsigma', \tau') + \frac{1}{2}[(\varsigma, \tau), (\varsigma', \tau')]} \\ &= e^{(\varsigma + \varsigma', \tau + \tau' + 2\Omega(\varsigma, \varsigma'))} = (z + z', t + t' + 2\Omega(z, z')) . \end{aligned}$$

Si $f : \mathbb{H} \rightarrow \mathbb{R}$ est une fonction différentiable, son gradient relativement à la connexion \mathcal{K} est donné par (cf. chapitre 5, section 2) :

$$\begin{aligned} \nabla_{\mathcal{K}}(f) &= X_1(f)X_1 + X_2(f)X_2 \\ &= \left(\frac{\partial f}{\partial x_1} + 2x_2 \frac{\partial f}{\partial t} \right) (1, 0, 2x_2) + \left(\frac{\partial f}{\partial x_2} - 2x_1 \frac{\partial f}{\partial t} \right) (0, 1, -2x_1) \\ &= \left(\frac{\partial f}{\partial x_1} + 2x_2 \frac{\partial f}{\partial t}, \frac{\partial f}{\partial x_2} - 2x_1 \frac{\partial f}{\partial t}, 2x_2 \frac{\partial f}{\partial x_1} - 2x_1 \frac{\partial f}{\partial x_2} + 4(x_1^2 + x_2^2) \frac{\partial f}{\partial t} \right) . \end{aligned}$$

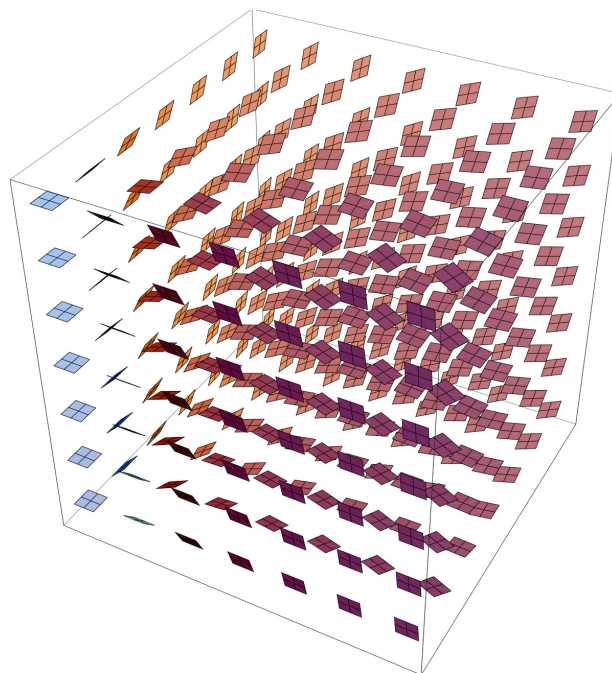


FIGURE 6. La structure de contact du groupe de Heisenberg.

Par ailleurs, si $X = u_1 X_1 + u_2 X_2$ est un champ de vecteurs tangents à \mathcal{K} , sa divergence relativement à \mathcal{K} est donnée par :

$$\operatorname{div}_{\mathcal{K}} X = X_1(u_1) + X_2(u_2) .$$

Si $X = \nabla_{\mathcal{K}}(f)$ alors

$$\operatorname{div}_{\mathcal{K}}(\nabla_{\mathcal{K}}(f)) = X_1(X_1(f)) + X_2(X_2(f)) = 2\Delta_{\mathcal{K}}(f) .$$

La structure de contact de \mathbb{H} est représentée à la figure 6. Le long d'une fibre verticale, z est constant et le plan de contact K_q aussi. Le long de l'axe x_1 , le plan de contact est engendré par

$$\{X_1 = (1, 0, 0), X_2 = (0, 1, -2x_1)\} .$$

Il contient l'axe x_1 et tourne autour de lui avec l'angle $\arctan(-2x_1)$, partant de l'orientation horizontale ($x_1 = 0$) pour arriver à l'orientation verticale ($x_1 = \infty$). Le long de l'axe x_2 , le plan de contact est engendré par

$$\{X_1 = (1, 0, 2x_2), X_2 = (0, 1, 0)\} .$$

Il contient l'axe x_2 et tourne autour de lui avec l'angle $\arctan(2x_2)$, partant de l'orientation horizontale ($x_2 = 0$) pour arriver à l'orientation verticale ($x_2 = \infty$).

2.2.2. *Les géodésiques de \mathbb{H} : hamiltonien et lagrangien.*

Pour calculer les géodésiques, on choisit encore une fois la métrique sur les plans de contact faisant de $\{X_1, X_2\}$ une base orthonormale. Les géodésiques sont alors déductibles de l'hamiltonien H défini sur le fibré cotangent $T^*\mathbb{H}$ par

$$H(x_1, x_2, t, \xi_1^*, \xi_2^*, \theta^*) = \frac{1}{2} [(\xi_1^* + 2x_2\theta^*)^2 + (\xi_2^* - 2x_1\theta^*)^2]$$

(où $(\xi_1^*, \xi_2^*, \theta^*)$ sont les moments conjugués de (x_1, x_2, t)). H est l'opposé du symbole du laplacien sous-riemannien

$$\Delta_H = \frac{1}{2} (X_1^2 + X_2^2) .$$

Les équations de Hamilton sont donc maintenant, avec $\xi^* = (\xi_1^*, \xi_2^*)$

$$\begin{cases} \dot{z}(s) = \frac{\partial H}{\partial \xi^*} = (\xi_1^* + 2x_2\theta^*, \xi_2^* - 2x_1\theta^*) \\ \dot{\xi}^*(s) = -\frac{\partial H}{\partial z} = 2\theta^* ((\xi_2^* - 2x_1\theta^*), -(\xi_1^* + 2x_2\theta^*)) \\ \dot{t}(s) = \frac{\partial H}{\partial \theta^*} = 2x_2(\xi_1^* + 2x_2\theta^*) - 2x_1(\xi_2^* - 2x_1\theta^*) \\ \quad = 2(x_2\xi_1^* - x_1\xi_2^*) + 4(x_1^2 + x_2^2\theta^*) \\ \dot{\theta}^*(s) = -\frac{\partial H}{\partial t} = 0 \end{cases}$$

Comme H est indépendant de t , $\dot{\theta}^*(s) = -\frac{\partial H}{\partial t} = 0$ et θ^* est constant sur les trajectoires.

Nous venons de voir la formulation hamiltonienne du problème. La version lagrangienne (à comparer avec celles des chapitres précédents) s'obtient par transformée de Legendre :

$$L = \langle \xi^*, \dot{z} \rangle + \theta^* \dot{t} - H = \left\langle \xi^*, \frac{\partial H}{\partial \xi^*} \right\rangle + \theta^* \frac{\partial H}{\partial \theta^*} - H .$$

Pour une courbe $\Gamma(s) = (x_1(s), x_2(s), t(s))$, le lagrangien s'écrit donc :

$$\begin{aligned} L(\Gamma, \dot{\Gamma}) &= L(x_1, x_2, t, \dot{x}_1, \dot{x}_2, \dot{t}) = \frac{1}{2} (\dot{x}_1^2 + \dot{x}_2^2) + \theta^* (\dot{t} - 2x_2\dot{x}_1 + 2x_1\dot{x}_2) \\ &= \frac{1}{2} \|\dot{\Gamma}_{\mathcal{K}}\|_{\mathcal{K}}^2 + \theta^* \omega(\dot{\Gamma}) \end{aligned}$$

$\dot{\Gamma}_{\mathcal{K}}$ étant la projection de $\dot{\Gamma}$ sur \mathcal{K} et la norme $\|\bullet\|_{\mathcal{K}}$ étant celle définissant la métrique sous-riemannienne. Il s'agit du lagrangien des géodésiques avec la contrainte

$$\omega(\dot{\Gamma}) = 0$$

qui exprime que $\Gamma(s)$ est une intégrale du champ de plans de contact \mathcal{K} , θ^* étant le multiplicateur de Lagrange. Lorsque la contrainte $\omega(\dot{\Gamma}) = 0$ est satisfaite (i.e.

quand Γ est une courbe intégrale de \mathcal{K} , L est tout simplement *l'énergie cinétique*

$$L(\Gamma, \dot{\Gamma}) = \frac{1}{2} \left\| \dot{\Gamma} \right\|_{\mathcal{K}}^2 .$$

Les équations d'Euler-Lagrange $\frac{d}{ds} \left(\frac{\partial L}{\partial \dot{\Gamma}} \right) = \frac{\partial L}{\partial \Gamma}$ donnent

$$\ddot{x}_1 X_1 + \ddot{x}_2 X_2 = 4\theta^* (\dot{x}_2 X_1 - \dot{x}_1 X_2)$$

soit

$$\ddot{\Gamma}_{\mathcal{K}} = 4\theta^* J'(\dot{\Gamma}_{\mathcal{K}})$$

où J' est la rotation de $-\frac{\pi}{2}$ dans les plans de contact :

$$J'(X_1) = X_2, \quad J'(X_2) = -X_1 .$$

Cette équation peut être clarifiée de la façon suivante (cf. Calin-Mangione [86]). Si T et W sont des champs de vecteurs tangents, la dérivation covariante associée à la connexion \mathcal{K} est définie par :

$$\nabla_T W = (Tg(W, X_1)) X_1 + (Tg(W, X_2)) X_2$$

où le champ T agit comme dérivation et g est le tenseur de la métrique sous-riemannienne (et donc les $g(W, X_i)$ sont des fonctions scalaires). Si $T = t_1 X_1 + t_2 X_2$ et $W = w_1 X_1 + w_2 X_2$, alors

$$\nabla_T W = (t_1 X_1(w_1) + t_2 X_2(w_1)) X_1 + (t_1 X_1(w_2) + t_2 X_2(w_2)) X_2 .$$

En particulier pour une courbe intégrale Γ ,⁸

$$\begin{aligned} \nabla_{\dot{\Gamma}}(\dot{\Gamma}) &= (\dot{x}_1 X_1(\dot{x}_1) + \dot{x}_2 X_2(\dot{x}_1)) X_1 + (\dot{x}_1 X_1(\dot{x}_2) + \dot{x}_2 X_2(\dot{x}_2)) X_2 \\ &= \dot{\Gamma}(\dot{x}_1) X_1 + \dot{\Gamma}(\dot{x}_2) X_2 = \ddot{x}_1 X_1 + \ddot{x}_2 X_2 = \ddot{\Gamma} = 4\theta^* J'(\dot{\Gamma}) . \end{aligned}$$

Par ailleurs, la courbure κ de la connexion \mathcal{K} en $q \in \mathbb{H}$ est définie pour $u \in K_q$ par

$$\kappa(u) = d\omega(u, X_1) X_1 + d\omega(u, X_2) X_2 = d\omega(X_1, X_2) J'(u) = 4J'(u) .$$

Les équations d'Euler-Lagrange s'écrivent donc

$$\nabla_{\dot{\Gamma}}(\dot{\Gamma}) = \theta^* \kappa(\dot{\Gamma}) .$$

Comme le soulignent Calin et Mangione

8. Nous omettons la référence à \mathcal{K} quand elle est évidente.

“In sub-Riemannian geometry the acceleration of the geodesics is equal to the curvature. This keeps the geodesics into the horizontal distribution.” ([86], p. 7)

Le long d’une géodésique $\Gamma(s)$, la norme de la vitesse $\|\dot{\Gamma}\|$ est constante. En effet, appliquons $\dot{\Gamma}$ à $\|\dot{\Gamma}\|^2 = g(\dot{\Gamma}, \dot{\Gamma})$. On obtient, à cause de l’orthogonalité de $\nabla_{\dot{\Gamma}}(\dot{\Gamma})$ et de $\dot{\Gamma}$

$$\dot{\Gamma}g(\dot{\Gamma}, \dot{\Gamma}) = 2g(\nabla_{\dot{\Gamma}}(\dot{\Gamma}), \dot{\Gamma}) = 2\theta^*g(\kappa(\dot{\Gamma}), \dot{\Gamma}) = 0.$$

Par ailleurs, il est facile de calculer l’évolution d’une fonction différentiable $f : \mathbb{H} \rightarrow \mathbb{R}$ le long d’une courbe $\Gamma(s)$. On a en effet

$$\frac{df}{ds} = \frac{\partial f}{\partial s} + X_1(f)\dot{x}_1 + X_2(f)\dot{x}_2 + \frac{\partial f}{\partial t}\omega(\dot{\Gamma}).$$

Si Γ est une courbe intégrale de \mathcal{K} , alors $\omega(\dot{\Gamma}) = 0$ et

$$\frac{df}{ds} = \frac{\partial f}{\partial s} + g(\dot{\Gamma}, \nabla_{\mathcal{K}}(f)).$$

Appliquons cela à l’action $S(x_1, x_2, t, s)$ solution de l’équation de Hamilton-Jacobi

$$\frac{\partial S}{\partial s} + H(x_1, x_2, t, dS) = 0.$$

On a

$$\begin{aligned} H(x_1, x_2, t, dS) &= \frac{1}{2}(dS(X_1)^2 + dS(X_2)^2) = \frac{1}{2}(X_1(S)^2 + X_2(S)^2) \\ &= \frac{1}{2}(X_1(S)^2 + X_2(S)^2) = \frac{1}{2}\|\nabla_{\mathcal{K}}(S)\|^2. \end{aligned}$$

Mais le long d’une géodésique sous-riemannienne $\Gamma(s)$ on a

$$\dot{\Gamma} = \nabla_{\mathcal{K}}(S)$$

et, comme la vitesse est constante = 1, on a $H(x_1, x_2, t, dS) = \frac{1}{2}$. L’équation

$$\|\nabla_{\mathcal{K}}(S)\| = 1$$

est l’équation *eikonale*⁹ de l’“optique géométrique” associée à cette géométrie sous-riemannienne.

9. Cf. chapitre 12, section 6.1.

On notera qu'étant donnée la relation $\dot{\Gamma} = \nabla_{\mathcal{K}}(S)$, la *divergence* par rapport à \mathcal{K} du flot géodésique est, d'après la formule ci-dessus $\operatorname{div}_{\mathcal{K}}(\nabla_{\mathcal{K}}(f)) = 2\Delta_{\mathcal{K}}(f)$, donnée par

$$\operatorname{div}_{\mathcal{K}}(\dot{\Gamma}) = 2\Delta_{\mathcal{K}}(S) .$$

2.2.3. *Les calculs de Beals, Gaveau et Greiner.*

Résumons maintenant les calculs de Beals, Gaveau et Greiner [35] que nous avons adaptés plus haut. Le lecteur pourra les comparer pour en observer les (petites) différences. Si l'on pose $\varsigma^* = \xi^* + \theta^* \Lambda(z)$, $2H$ est la norme de ς^*

$$H(z, t, \xi^*, \theta^*) = \frac{1}{2} \langle \varsigma^*, \varsigma^* \rangle .$$

H étant constant le long des trajectoires,

$$H = H_0 = \frac{1}{2} \langle \varsigma^*(0), \varsigma^*(0) \rangle .$$

Comme H est indépendant de la coordonnée "verticale" t ,

$$\dot{\theta}^*(s) = -\frac{\partial H}{\partial t} = 0$$

et $\theta^* = \text{cste}$ sur les trajectoires. θ^* joue par conséquent le rôle d'un paramètre. Les équations de Hamilton sont données par les formules (on utilise le fait que Λ est antisymétrique, $-\Lambda^T = \Lambda$) :

$$\begin{cases} \dot{z}(s) = \frac{\partial H}{\partial \xi^*} = \varsigma^*(s) \\ \dot{\xi}^*(s) = -\frac{\partial H}{\partial z} = -\theta^* \Lambda^T \varsigma^*(s) = \theta^* \Lambda \varsigma^*(s) \\ \dot{t}(s) = \frac{\partial H}{\partial \theta^*} = \langle \varsigma^*(s), \Lambda z(s) \rangle \\ \dot{\theta}^*(s) = -\frac{\partial H}{\partial t} = 0 \end{cases}$$

On obtient pour ς^* l'équation différentielle

$$\begin{aligned} \dot{\varsigma}^*(s) &= \dot{\xi}^*(s) + \theta^* \Lambda \dot{z}(s) \quad (\text{car } \theta^* = \text{cste}) \\ &= \theta^* \Lambda \varsigma^*(s) + \theta^* \Lambda \varsigma^*(s) = 2\theta^* \Lambda \varsigma^*(s) \end{aligned}$$

dont la solution est :

$$\varsigma^*(s) = e^{2s\theta^* \Lambda} \varsigma^*(0)$$

et comme Λ est la multiplication par $-2i$ dans le plan ς^* , $\varsigma^*(s)$ tourne avec une période de $\frac{\pi}{2\theta^*}$.

On en tire les solutions du système hamiltonien. Si $\theta^* = 0$, alors $\dot{\xi}^* = 0$ et $\xi^* = \varsigma^*$ sont constants $= \xi_0^*$, $z = s\xi_0^*$, $\dot{t} = s \langle \xi_0^*, \Lambda \xi_0^* \rangle = 0$ et donc t est constant $= t_0$: les géodésiques sont des droites horizontales $z = s\xi_0^*$ d'altitude t_0 .

Si en revanche $\theta^* \neq 0$, l'intégration de $\dot{z}(s) = \zeta^*(s)$ donne

$$\begin{aligned} z(s) &= \int_0^s \zeta^*(r) dr \\ &= (2\theta^* \Lambda)^{-1} (e^{2s\theta^* \Lambda} - I) \zeta^*(0) \\ &= (2\theta^* \Lambda)^{-1} (\zeta^*(s) - \zeta^*(0)) . \end{aligned}$$

La projection $z(s)$ de la géodésique part de $z(0) = 0$ et y revient avec la période $\frac{\pi}{2\theta^*}$.

On en tire

$$z(\tau) = (2\theta^* \Lambda)^{-1} (\zeta^*(\tau) - \zeta^*(0)) = (2\theta^* \Lambda)^{-1} (e^{2\tau\theta^* \Lambda} - I) \zeta^*(0)$$

et donc, si $z(\tau) = z \neq 0$, i.e. si $\tau \neq \frac{k\pi}{2\theta^*}$,

$$z(s) = \frac{e^{2s\theta^* \Lambda} - I}{e^{2\tau\theta^* \Lambda} - I} z = e^{(s-\tau)\theta^* \Lambda} \frac{\sinh(s\theta^* \Lambda)}{\sinh(\tau\theta^* \Lambda)} z .$$

Et comme Λ est la multiplication par $-2i$ dans le plan z et comme $\sinh(i\varphi) = i \sin(\varphi)$, on obtient

$$z(s) = e^{(s-\tau)\theta^* \Lambda} \frac{\sin(2s\theta^*)}{\sin(2\tau\theta^*)} z$$

qui est bien défini puisque $z \neq 0$ et $\sin(2\tau\theta^*) \neq 0$ par hypothèse. D'où, sous cette hypothèse, la formule qui relie la valeur initiale $\zeta^*(0) = (\zeta_1^*, \zeta_2^*)$ de ζ^* à la valeur finale $z = z(\tau) = (x_1, x_2)$ de z :

$$H_0 = \frac{1}{2} \langle \zeta^*(0), \zeta^*(0) \rangle = \frac{1}{2} |\zeta^*(0)|^2 = \frac{2\theta^{*2}}{\sin^2(2\tau\theta^*)} |z|^2 ,$$

les deux étant reliés par

$$\begin{aligned} \zeta_1^* &= \frac{2\theta^*}{\sin(2\tau\theta^*)} (x_1 \cos(2\tau\theta^*) - x_2 \sin(2\tau\theta^*)) \\ \zeta_2^* &= \frac{2\theta^*}{\sin(2\tau\theta^*)} (x_1 \sin(2\tau\theta^*) + x_2 \cos(2\tau\theta^*)) . \end{aligned}$$

On trouve en définitive :

$$\begin{cases} x_1(s) = \frac{\sin(2s\theta^*)}{\sin(2\tau\theta^*)} (x_1 \cos(2(\tau-s)\theta^*) - x_2 \sin(2(\tau-s)\theta^*)) \\ x_2(s) = \frac{\sin(2s\theta^*)}{\sin(2\tau\theta^*)} (x_2 \cos(2(\tau-s)\theta^*) + x_1 \sin(2(\tau-s)\theta^*)) \\ t(s) = t_0 + \frac{4s\theta^* - \sin(2s\theta^*)}{2(\sin(2\tau\theta^*))^2} ((x_1)^2 + (x_2)^2) \end{cases}$$

formules exprimées en termes des conditions finales $z = z(\tau)$ que l'on peut également exprimer en termes des conditions initiales $\zeta^*(0) = (\zeta_1^*, \zeta_2^*)$ par

$$\begin{cases} x_1(s) = \frac{\sin(2s\theta^*)}{2\theta^*} (\zeta_1^* \cos(2s\theta^*) + \zeta_2^* \sin(2s\theta^*)) \\ x_2(s) = \frac{\sin(2s\theta^*)}{2\theta^*} (-\zeta_1^* \sin(2s\theta^*) + \zeta_2^* \cos(2s\theta^*)) \\ t(s) = t_0 + \frac{4s\theta^* - \sin[4s\theta^*]}{8\theta^{*2}} ((\zeta_1^*)^2 + (\zeta_2^*)^2) \end{cases}$$

Précisons l'interprétation géométrique de ces formules que nous avons donnée à la section précédente. Pour simplifier, prenons $2\theta^* = 1$. La projection $c(z)$ de la géodésique sur le plan base des z part de 0 au temps $s = 0$ et arrive en $z(\tau) = (x_1, x_2)$ au temps $s = \tau$. Elle revient en 0 au bout d'un temps $s = \pi$. Supposons $\tau < \pi$ (c'est-à-dire que $z(\tau)$ fait moins qu'un tour). La tangente $(x'_1(0), x'_2(0))$ en 0 à $c(z)$ est

$$\left(\frac{\cos(\tau) x_1 - \sin(\tau) x_2}{\sin(\tau)}, \frac{\cos(\tau) x_2 + \sin(\tau) x_1}{\sin(\tau)} \right).$$

Considérons la rotation $r_{\tau-s}$ de $z(\tau)$. La géodésique $c(z)$ se construit en partant de 0 avec la pente $r_\tau(z(\tau))$ et en tournant progressivement le rayon vecteur $r_\tau(z(\tau))$ de r_{-s} tout en le multipliant par $\frac{\sin(s)}{\sin(\tau)}$ de façon à arriver en $z(\tau)$ pour $s = \tau$. Le centre de $c(z)$ est le point de coordonnées (on note la cosécante $\csc(\tau) = \frac{1}{\sin(\tau)}$)

$$\begin{cases} \bar{x}_1 = \frac{1}{2} \csc(\tau)(x_2 \cos(\tau) + x_1 \sin(\tau)) \\ \bar{x}_2 = \frac{1}{2} \csc(\tau)(-x_1 \cos(\tau) + x_2 \sin(\tau)) \end{cases}$$

situé à une distance de $\frac{1}{2} \csc(\tau)$ sur la perpendiculaire en 0 à la tangente $(x'_1(0), x'_2(0))$ de $c(z)$. Quant au rayon R de $c(z)$ il est donné par

$$R^2 = \frac{1}{4} (x_1^2 + x_2^2) \csc^2(\tau).$$

Si l'on applique alors la formule classique disant que l'aire de la partie d'un cercle de rayon R sous-tendue par une corde à distance d du centre est

$$A = \frac{1}{2} \left(\frac{R^2 \operatorname{Arctan} \left(\frac{\sqrt{R^2 - d^2}}{d} \right)}{\sqrt{R^2 - d^2}} - d \right)$$

on constate que $t - t_0$ est proportionnel à l'aire $A(s)$ (où $A(s)$ est considérée comme fonction de s à travers le $d(s)$ de la corde $(0, z(s))$). Il s'agit d'une "loi des aires" schématisée à la figure 7.

Revenons à l'équation donnant t . La valeur finale de t , $t = t(\tau)$, est de la forme

$$t(\tau) = t_0 + \mu(2\tau\theta^*) |z(\tau)|^2$$

avec la fonction $\mu(\varphi)$ que nous connaissons déjà

$$\mu(\varphi) = \frac{\varphi}{\sin^2(\varphi)} - \frac{1 \sin(2\varphi)}{2 \sin^2(\varphi)} = \frac{\varphi}{\sin^2(\varphi)} - \cot(\varphi).$$

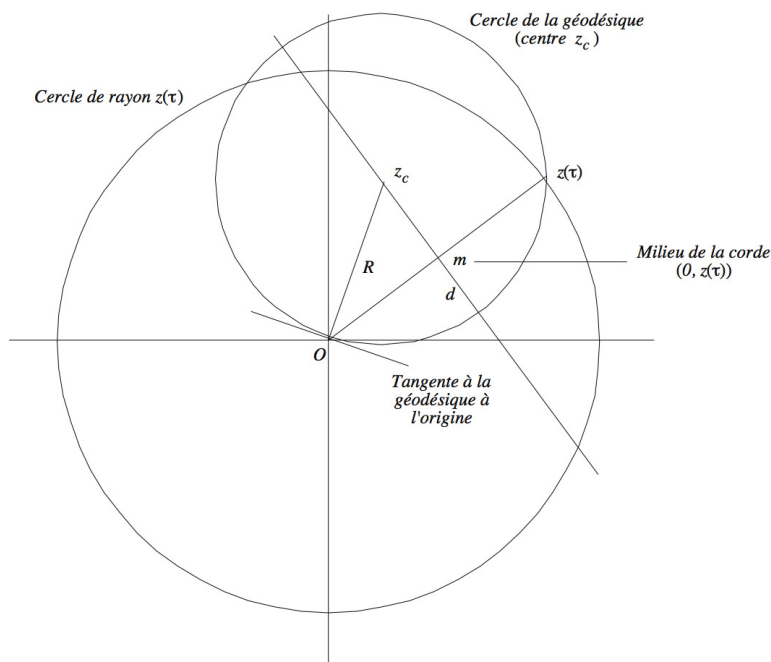


FIGURE 7. La “loi des aires” des géodésiques pour $2\theta^* = 1$. La projection $c(z)$ sur le plan z part de 0 pour $s = 0$ et arrive en $z(\tau)$ pour $s = \tau$. Elle part de 0 avec la pente $r_\tau(z(\tau))$ et se construit en tournant progressivement le rayon vecteur de r_{-s} tout en le multipliant par $\frac{\sin(s)}{\sin(\tau)}$. La composante $t - t_0$ est proportionnelle à l’aire de la partie du cercle géodésique sous-tendue par la corde $(0, z(s))$.

Comme à la section précédente, pour obtenir des géodésiques de même extrémité

$$\{x_1 = x_1(\tau), x_2 = x_2(\tau), t = t(\tau)\}$$

il faut résoudre l’équation

$$t = \mu(2\tau\theta^*) |z(\tau)|^2$$

en $\tau\theta^*$. Si l’on reprend notre exemple de $\mu(\varphi) = 20$ avec $\theta^* = 1$, on obtient 11 solutions :

$$\left\{ \begin{array}{l} \tau_1 = 1.36, \tau_2 = 1.77, \tau_3 = 2.84, \tau_4 = 3.44, \tau_5 = 4.34, \tau_6 = 5.09, \\ \tau_7 = 5.83, \tau_8 = 6.75, \tau_9 = 7.32, \tau_{10} = 8.42, \tau_{11} = 8.80 . \end{array} \right\}$$

Si l’on prend par exemple $x_1 = x_1(\tau) = 2$ et $x_2 = x_2(\tau) = 1$, on obtient $t = 100$ et 11 géodésiques joignant $(0, 0, 0)$ à $(2, 1, 100)$. Nous représentons les deux cas τ_1 et τ_{10} à la figure 8.

Pour la sphère S et le front d’onde W , on obtient la même formule pour la longueur :

$$L = \frac{\varphi}{|\sin(\varphi)|} |z_1|$$

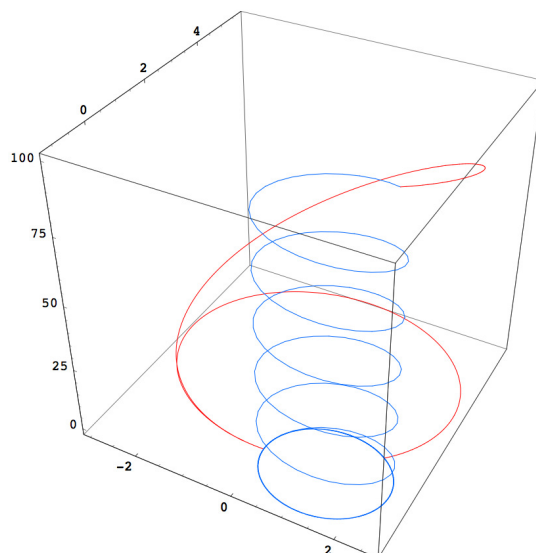


FIGURE 8. Deux géodésiques sous-riemanniennes du groupe de Heisenberg joignant $(0, 0, 0)$ à $(2, 1, 100)$. On représente les géodésiques dans \mathbb{H} et leur projection sur le plan z .

et, pour les géodésiques de longueur 1, les équations

$$\begin{cases} |z_1| = \left| \frac{\sin(\varphi)}{\varphi} \right| \\ |t| = \frac{\varphi - \sin(\varphi) \cos(\varphi)}{\varphi^2} \end{cases}$$

qui sont plus simples que celles de la section précédente pour \mathbb{V}_J . Nous présentons à la figure 9 la section de W dans le premier quadrant du plan (x, t) .

Nous présentons ensuite à la figure 10 la structure de S et de W . Les deux singularités de S sont sur l'axe des t et ressemblent à la singularité “pomme” de l'épicarpe d'une pomme au voisinage du pédoncule.

2.2.4. *Cut locus et points conjugués.*

Comme l'a noté Richard Montgomery dans [375] (p. 12), la loi des aires permet de mettre immédiatement en lumière une propriété remarquable de la géométrie sous-riemannienne de \mathbb{H} qui est impossible en géométrie riemannienne. Considérons un cercle du plan z passant par l'origine 0 et de rayon r (et donc d'aire $A = \pi r^2$), faisons-le tourner autour de 0 et considérons ses relèvements. Ce sont des géodésiques de même origine 0, de même extrémité $(0, 0, A)$ et de même longueur. Le point $(0, 0, A)$ de l'axe des t est donc un point conjugué de 0. Autrement dit, la période $\frac{\pi}{2\theta^*}$ de la projection $c(z)$ est le premier temps conjugué et le temps de coupure de la géodésique.

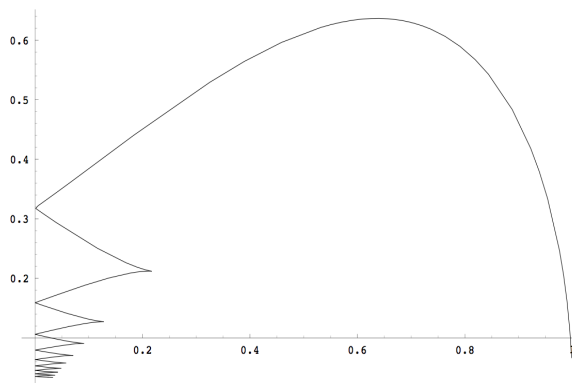


FIGURE 9. Le profil du front d'onde W dans le premier quadrant du plan (x, t) . Le premier arc ($\varphi \in [0, \pi]$) allant de la droite de l'axe des x jusqu'à l'axe t correspond à la sphère (géodésiques globalement minimisantes de longueur 1). Ensuite, chaque intervalle de π de la variation de φ engendre une composante avec un cusp du front W . (D'après Beals et al. [35]).

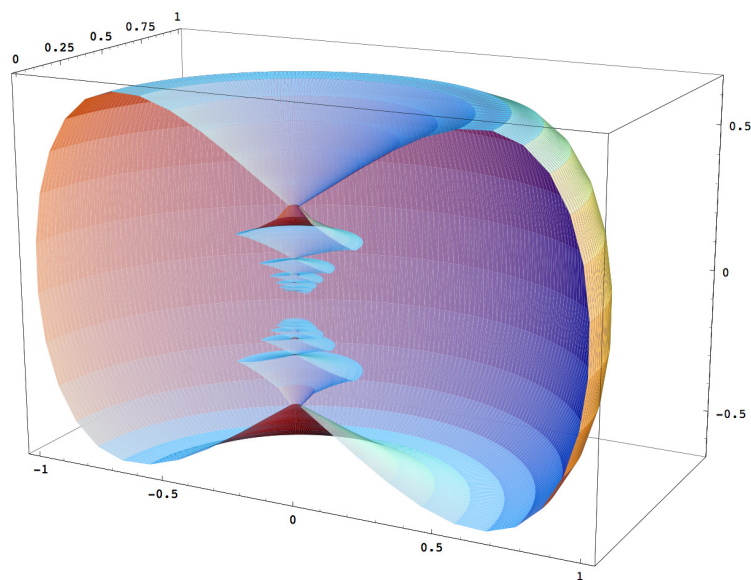


FIGURE 10. La sphère et le front d'onde sous-riemanniens du groupe de Heisenberg. (D'après Beals et al. [35]).

Mais comme A est quelconque, *tous* les points de l'axe des t sont des points conjugués de 0 et le cut locus de 0 est cet axe moins 0. On peut comparer à la géométrie riemannienne de la sphère où un point ne possède essentiellement qu'un seul point conjugué, à savoir son antipode.

3. Les géodésiques des espaces de jets

Avant d'aborder les géodésiques sous-riemanniennes de notre second modèle \mathbb{V}_S , nous allons nous arrêter un instant sur la généralisation du modèle \mathbb{V}_J aux jets d'ordre supérieur $J^k(\mathbb{R}, \mathbb{R})$. Nous pouvons prendre comme exemple la structure d'Engel des jets d'ordre 2 que nous avons explicitée à la section 4 du chapitre 3 et qui correspond à la détection des courbures.

Nous y avons vu qu'on introduit une quatrième variable κ liée à la courbure, qu'on travaille dans l'espace $J^2(\mathbb{R}, \mathbb{R}) = \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}$, et qu'on écrit que l'interprétation naturelle de κ est la dérivée seconde $f''(x)$ pour les courbes d'équation $y = f(x)$. La structure d'Engel est le système de Pfaff

$$\{\omega_J = dy - p dx, \tau_J = dp - \kappa dx\} .$$

La distribution est engendrée par les deux vecteurs

$$\{X_1 = \partial_x + p \partial_y + \kappa \partial_p, X_2 = \partial_\kappa\}$$

avec

$$[X_2, X_1] = \partial_p, [[X_2, X_1], X_1] = \partial_y ,$$

les autres crochets étant nuls.¹⁰ C'est une distribution de Goursat de type (2, 3, 4). Il est facile de généraliser à $k > 2$ avec le système de Pfaff

$$\{\omega_1 = dy - p_1 dx, \dots, \omega_k = dp_{k-1} - p_k dx\}$$

et les générateurs

$$\{X_1 = \partial_x + p_1 \partial_y + \dots + p_k \partial_{p_{k-1}}, X_2 = \partial_{p_k}\}$$

Les géodésiques sont des courbes intégrales de la distribution d'Engel (en général des relevées legendriennes de courbes $y = f(x)$ du plan de base (x, y)) qui sont minimales pour la métrique faisant de $\{X_1, X_2\}$ une base orthonormée et on les projette sur le plan (x, p_k) .

Richard Montgomery et Alejandro Doddoli [146], et également Alfonso Anzaldo-Meneses et Felipe Monroy-Perez, ont analysé les géodésiques des $J^k(\mathbb{R}, \mathbb{R})$ et caractérisé les géodésiques *globalement* minimisantes en étudiant leurs *projections* sur le plan *spécial* (x, p_k) dual du plan (x, y) , projections qui généralisent les cercles et les droites dans le plan (x, p) du cas \mathbb{V}_J . X_1 et X_2 se projettent sur ∂_x et ∂_{p_k} la base standard des plans tangents de ce plan, dx et dp_k est la base standard des plans cotangents et la projection est une *isométrie* entre le plan euclidien (x, p_k) et la métrique des plans de la distribution sous-riemannienne qui en est la relevée.

Un joli résultat, déjà démontré par Andrei Ardentov et Yuri Sachkov [18], est que, pour $k = 2$ (Engel) on obtient les *elasticæ*. Cela est dû à la raison suivante. Soit

10. Dans la section 4 du chapitre 3, X_1 et X_2 étaient notés \mathbf{t}_1^κ et \mathbf{t}_4 .

$\gamma(s) = (x(s), \kappa(s))$ la projection sur le plan (x, κ) d'une géodésique paramétrée par sa longueur d'arc s . Soit $\chi(s)$ sa courbure. On montre que l'on a la relation

$$\chi(s) = a + bx(s)$$

(polynôme de degré $2 - 1 = 1$) qui généralise le cas $\chi(s) = a$ (polynôme de degré $1 - 1 = 0$) de $\mathbb{H}_{pol} = J^1(\mathbb{R}, \mathbb{R})$.¹¹ Or cette relation est caractéristique des elasticæ.

Pour retrouver facilement sur cette base la classification des elasticæ les auteurs ont eu l'ingénieuse idée de comparer une courbe γ donnée par sa paramétrisation $\gamma(s) = (x(s), \kappa(s))$ à sa représentation comme graphe $(x, \kappa(x))$ valide en dehors des points à tangente verticale $\kappa'(x) = \infty$. Comme la métrique euclidienne est $ds^2 = dx^2 + d\kappa^2$, écrivons

$$\begin{aligned} d\kappa &= F(x) ds \\ dx &= \pm \sqrt{1 - F(x)^2} ds \end{aligned}$$

avec $|F(x)| \leq 1$. On sait (cf. la section 4 du chapitre 3) que la courbure $\chi(x)$ de γ est donnée par la formule de Newton $\chi(x) = \frac{\kappa''(x)}{(1 + \kappa'(x)^2)^{3/2}}$. Mais

$$\begin{aligned} \kappa'(x) &= \frac{d\kappa}{dx} = \pm \frac{F(x) ds}{\sqrt{1 - F(x)^2} ds} = \pm \frac{F(x)}{\sqrt{1 - F(x)^2}} \\ 1 + \kappa'(x)^2 &= 1 + \frac{F(x)^2}{1 - F(x)^2} = \frac{1}{1 - F(x)^2} \\ \kappa''(x) &= \pm \frac{F'(x)}{\sqrt{1 - F(x)^2}} \mp \frac{1}{2} \frac{F(x) (-2F(x) F'(x))}{(1 - F(x)^2)^{3/2}} = \pm \frac{F'(x)}{(1 - F(x)^2)^{3/2}} \\ \chi(x) &= \frac{\kappa''(x)}{(1 + \kappa'(x)^2)^{3/2}} = \pm \frac{F'(x)}{(1 - F(x)^2)^{3/2}} (1 - F(x)^2)^{3/2} \\ &= \pm F'(x) . \end{aligned}$$

Les points de γ à tangentes verticales correspondent à $F(x) = \pm 1$.

Mais $\chi(x) = a + bx$ et donc

$$F(x) = c + ax(s) + b \frac{x(s)^2}{2} .$$

En faisant varier F , on retrouve la classification des elasticæ.

11. On montre plus généralement que $\chi(s) = P(x(s))$ où P est un polynôme de degré $k - 1$.

4. Les géodésiques sous-riemanniennes de \mathbb{V}_S (Agrachev, Sachkov)

4.1. Le système hamiltonien sur \mathbb{V}_S

Passons maintenant au calcul des géodésiques du modèle du groupe Euclidien

$$\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1 = \mathbb{R}^2 \rtimes SO(2) = SE(2)$$

muni de sa structure contact naturelle et de la métrique sous-riemannienne associée (cf. la section 7 du chapitre 5 ainsi que le chapitre 11). Comme nous l'avons rappelé dans la section introductive 1 de ce chapitre, après des discussions avec nous, Giovanna Citti et Alessandro Sarti, Andrei Agrachev et certains chercheurs du groupe de la SISSA de Trieste (Yuri Sachkov, Ugo Boschain, Igor Moiseev et aussi Jean-Paul Gauthier de Toulon) résolurent le problème dans les années 2000 (voir en particulier [373]).

Encore une fois, on ramène le problème des géodésiques à un problème de théorie du contrôle optimal. $\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1$ est muni de sa 1-forme de contact

$$\omega_S = \omega = -\sin(\theta) dx + \cos(\theta) dy$$

et la distribution $\mathcal{K}_S = \ker(\omega)$ de plans de contacts est définie par

$$K_q = \text{Span}\{X_1, X_2\} \subset T_q\mathbb{V}_S (q \in \mathbb{V}_S)$$

avec les deux vecteurs tangents de base

$$\{X_1 = \cos(\theta) \partial_x + \sin(\theta) \partial_y, X_2 = \partial_\theta\}$$

dont le crochet de Lie est

$$[X_1, X_2] = \sin(\theta) \partial_x - \cos(\theta) \partial_y = X_3.$$

On a

$$[X_2, X_3] = X_1, [X_1, X_3] = 0,$$

$-X_3$ étant le vecteur caractéristique orthogonal à K_q définissant une échelle par

$$\omega(-X_3) = (-\sin(\theta) dx + \cos(\theta) dy)(-X_3) = \sin^2(\theta) + \cos^2(\theta) = 1.$$

Les relevées legendriennes (intégrales de \mathcal{K}) sont solutions du système différentiel

$$\begin{cases} \dot{x} = u_1 \cos(\theta) \\ \dot{y} = u_1 \sin(\theta) \\ \dot{\theta} = u_2 \end{cases} \quad (1)$$

En effet, dans la mesure où

$$\frac{dy}{dx} = \frac{\dot{y}}{\dot{x}} = \frac{u_1 \sin(\theta)}{u_1 \cos(\theta)} = \tan(\theta),$$

la condition d'intégrabilité est automatiquement satisfaite.

Remarque. On notera que si x et y ne sont pas constants (i.e. si la solution Γ n'est pas dans la fibre) et si le contrôle u_1 s'annule au temps t_c , alors

$$\dot{x}(t_c) = \dot{y}(t_c) = 0$$

et la projection γ de la solution Γ possède une *singularité* (un cusp). Si l'on impose la contrainte $u_1 \neq 0$, alors il ne peut y avoir de singularités. Nous traiterons de ce problème à la section 5. \square

Remarque sur les notations. Dans cette section, pour distinguer le cas de \mathbb{V}_S du cas de \mathbb{V}_J , nous changeons un peu les notations et nous revenons à celles qui sont conventionnelles dans les hamiltoniens en théorie du contrôle : q pour les positions et λ pour les moments conjugués. \square

La géométrie sous-riemannienne sur \mathbb{V}_S est définie en posant que les $\{X_1, X_2\}$ constituent une base orthonormale des K_q . Pour trouver les hamiltoniens qu'il s'agit de maximiser en appliquant le PMP, on part de l'énergie cinétique (définie sur le fibré tangent $T\mathbb{V}_S$)

$$\dot{q}^2 = (u_1 \cos(\theta))^2 + (u_1 \sin(\theta))^2 + u_2^2 = u_1^2 + u_2^2$$

et on considère sa transformée de Legendre (définie sur le fibré cotangent $T^*\mathbb{V}_S$)

$$h(\lambda, q) = \langle \lambda, \dot{q} \rangle - \frac{1}{2} \dot{q}^2$$

λ étant un covecteur de composantes $\lambda = (\lambda_x, \lambda_y, \lambda_\theta)$ dans la base $dx, dy, d\theta$ de $T^*\mathbb{V}_S$. On vérifie immédiatement que

$$h(\lambda, q) = \langle \lambda, u_1 X_1(q) + u_2 X_2(q) \rangle - \frac{1}{2} (u_1^2 + u_2^2).$$

D'après le PMP, pour obtenir l'hamiltonien des géodésiques, on maximise $h(\lambda, q)$ par rapport aux contrôles u_1 et u_2 . Cela donne $\frac{\partial h}{\partial u_1} = \frac{\partial h}{\partial u_2} = 0$, soit

$$\begin{cases} u_1(\lambda, q) = \langle \lambda, X_1(q) \rangle = \lambda_x \cos(\theta) + \lambda_y \sin(\theta) \\ u_2(\lambda, q) = \langle \lambda, X_2(q) \rangle = \lambda_\theta \end{cases}$$

On obtient ainsi l'hamiltonien

$$\begin{aligned} H(\lambda, q) &= \frac{1}{2} (u_1^2 + u_2^2) = \frac{1}{2} (\langle \lambda, X_1(q) \rangle^2 + \langle \lambda, X_2(q) \rangle^2) \\ &= \frac{1}{2} (\lambda_1^2 + \lambda_2^2) \\ &= \frac{1}{2} ((\lambda_x \cos(\theta) + \lambda_y \sin(\theta))^2 + \lambda_\theta^2) \end{aligned}$$

où

$$\{\lambda_1 = \lambda_x \cos(\theta) + \lambda_y \sin(\theta), \lambda_2 = \lambda_\theta, \lambda_3 = \lambda_x \sin(\theta) - \lambda_y \cos(\theta)\}$$

sont les composantes du covecteur λ dans la base duale $(\omega_1, \omega_2, \omega_3)$ de (X_1, X_2, X_3) .¹² On comparera à l'hamiltonien

$$H_{\text{el}}(\lambda, q) = \lambda_1 + \frac{1}{2}\lambda_2^2$$

des elasticæ.

Les équations de Hamilton $\dot{q} = \frac{\partial H}{\partial \lambda}$ et $\dot{\lambda} = -\frac{\partial H}{\partial q}$ donnent pour les λ_i en posant $\lambda'_3 = -\lambda_3$:

$$\begin{cases} \dot{\lambda}_1 = \{\lambda_1, H\} = \lambda_2 \lambda'_3 \\ \dot{\lambda}_2 = \{\lambda_2, H\} = -\lambda_1 \lambda'_3 \\ \dot{\lambda}'_3 = \{\lambda'_3, H\} = -\lambda_2 \lambda_1 \end{cases}$$

Or celui-ci est caractéristique des fonctions de Jacobi cn , sn et dn ce qui explique pourquoi les fonctions elliptiques interviennent si naturellement dans le calcul des géodésiques¹³ En effet

$$\begin{cases} \dot{\text{sn}} = \text{cn} \cdot \text{dn} \\ \dot{\text{cn}} = -\text{sn} \cdot \text{dn} \\ \dot{\text{dn}} = -k \text{cn} \cdot \text{sn} \end{cases}$$

où k est le module de l'intégrale elliptique de première espèce $F(\varphi, k)$ correspondante.

Les équations de Hamilton complètes sont

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial \lambda_x} = \lambda_x \cos^2(\theta) + \lambda_y \cos(\theta) \sin(\theta) \\ \dot{y} = \frac{\partial H}{\partial \lambda_y} = \lambda_y \sin^2(\theta) + \lambda_x \cos(\theta) \sin(\theta) \\ \dot{\theta} = \frac{\partial H}{\partial \lambda_\theta} = \lambda_\theta \\ \dot{\lambda}_x = -\frac{\partial H}{\partial x} = 0 \\ \dot{\lambda}_y = -\frac{\partial H}{\partial y} = 0 \\ \dot{\lambda}_\theta = -\frac{\partial H}{\partial \theta} = -(\lambda_x \cos(\theta) + \lambda_y \sin(\theta))(-\lambda_x \sin(\theta) + \lambda_y \cos(\theta)) \\ = -\lambda_x \lambda_y \cos(2\theta) + \frac{1}{2}(\lambda_x^2 - \lambda_y^2) \sin(2\theta) \end{cases}$$

et définissent des trajectoires dans $T^*\mathbb{V}_S$ dont les projections sur \mathbb{V}_S sont les géodésiques sous-riemanniennes cherchées.

Comme l'énergie $H = \frac{1}{2}(\lambda_1^2 + \lambda_2^2)$ est constante sur les trajectoires¹⁴, cela signifie que, le long des trajectoires, $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ évolue sur le cylindre ayant pour base

-
12. Autrement dit, $\lambda_i = h_i(\lambda, q) = \langle \lambda, X_i \rangle$.
 13. Je remercie André Bellaïche pour cette remarque.
 14. Cela se retrouve immédiatement à partir des EDO des λ_i . En effet

$$\frac{1}{2}(\lambda_1^2 + \lambda_2^2) \cdot = \lambda_1 \dot{\lambda}_1 + \lambda_2 \dot{\lambda}_2 = -\lambda_1 \lambda_2 \lambda_3 + \lambda_2 \lambda_1 \lambda_3 = 0.$$

On a de même $\lambda_1^2 + \lambda_3^2 = \text{cste}$, mais pas $\lambda_2^2 + \lambda_3^2 = \text{cste}$.

le cercle de rayon $c = \sqrt{2H}$ du plan (λ_1, λ_2) . Cela signifie aussi que le moment (λ_x, λ_y) dans le plan base \mathbb{R}^2 est constant. Si on l'écrit comme un nombre complexe $(\lambda_x, \lambda_y) = \rho e^{i\beta}$ ($\rho \geq 0$), ρ et β constants, le changement de base pour λ donne

$$\begin{cases} \lambda_1 = \lambda_x \cos(\theta) + \lambda_y \sin(\theta) = \rho \cos(\beta) \cos(\theta) + \rho \sin(\beta) \sin(\theta) \\ \quad = \rho \cos(\theta - \beta) \\ \lambda_2 = \lambda_\theta \\ \lambda_3 = \lambda_x \sin(\theta) - \lambda_y \cos(\theta) = \rho \cos(\beta) \sin(\theta) - \rho \sin(\beta) \cos(\theta) \\ \quad = \rho \sin(\theta - \beta) \end{cases}$$

et les équations pour x, y, λ_θ deviennent

$$\begin{cases} \dot{x} = \lambda_1 \cos(\theta) = \rho \cos(\theta - \beta) \cos(\theta) \\ \dot{y} = \lambda_1 \sin(\theta) = \rho \cos(\theta - \beta) \sin(\theta) \\ \dot{\lambda}_\theta = \dot{\lambda}_2 = \lambda_1 \lambda_3 = \rho \cos(\theta - \beta) \rho \sin(\theta - \beta) \\ \quad = \frac{1}{2} \rho^2 \sin(2(\theta - \beta)) \end{cases}$$

et l'on note aussitôt

(i) que lorsque $\cos(\theta - \beta) = 0$ on a $\dot{x} = \dot{y} = 0$ et que par conséquent la projection de la géodésique sur le plan (x, y) possède un cusp et

(ii) que lorsque la courbure $\dot{\theta} = \lambda_\theta$ est $= 0$, la projection de la géodésique sur le plan (x, y) possède un point d'inflexion.

H s'écrit

$$H = \frac{1}{2} (\rho^2 \cos^2(\theta - \beta) + \lambda_\theta^2)$$

ce qui donne l'intégrale première d'"énergie"

$$\rho^2 \cos^2(\theta - \beta) + \lambda_\theta^2 = c^2 = 2H$$

(avec $c \geq 0$ et $c = 1$ si $H = \frac{1}{2}$) et l'équation pour $\dot{\theta}$

$$\dot{\theta}^2 = \lambda_\theta^2 = c^2 - \rho^2 \cos^2(\theta - \beta) .$$

On voit que le rapport $\frac{\rho}{c}$ fonctionne comme un module du problème car, si $c > \rho$, alors

$$c^2 - \rho^2 \cos^2(\theta - \beta) > 0$$

et il n'y a pas de contrainte sur cette dernière équation alors que si $c < \rho$ il faut au contraire satisfaire la contrainte

$$\cos^2(\theta - \beta) \leq \frac{c^2}{\rho^2} .$$

On remarque alors que, comme

$$\ddot{\theta} = \dot{\lambda}_\theta = \frac{1}{2} \rho^2 \sin(2(\theta - \beta)) ,$$

si l'on pose

$$\Theta = 2(\theta - \beta) - \pi,$$

on obtient l'équation

$$\ddot{\Theta} = -\rho^2 \sin(\Theta)$$

qui est l'équation d'un pendule de masse 1 et de longueur 1 soumis à la gravité ρ^2 . Étant donnée l'invariance du problème par rotation on peut prendre $\beta = 0$, ce qui donne le système

$$\begin{cases} \dot{x} = \rho \cos^2(\theta) \\ \dot{y} = \rho \cos(\theta) \sin(\theta) = \frac{1}{2}\rho \sin(2\theta) \\ \dot{\theta} = \lambda_\theta \\ \ddot{\theta} = \dot{\lambda}_\theta = \frac{1}{2}\rho^2 \sin(2\theta) \end{cases}$$

avec $\dot{\rho} = 0$ et les conditions initiales

$$\left\{ x(0) = y(0) = 0, \theta(0) = \theta_0, \dot{\theta}^2(0) = c^2 - \rho^2 \cos^2(\theta_0) \right\}.$$

4.2. Intégration dans le cas d'une impulsion de module 1

L'équation du second ordre satisfaite par θ étant l'équation d'un pendule, nous rencontrons, comme pour les elasticæ du chapitre 13, des courbes dont l'orientation de la tangente est "pendulaire". Pour faire les calculs, nous pouvons soit prendre $H = \frac{1}{2}$ (donc $c = 1$ et $t = s$ la longueur d'arc) et faire varier ρ , soit prendre $\rho = 1$, donc $H = \frac{1}{2}c^2$ et faire varier c en en faisant le paramètre définissant le module. Nous nous plaçons d'abord, à titre d'exercice, dans ce second cas, mais nous utiliserons ensuite la variante qu'est le premier cas.

Dans le second cas, nous pouvons poser $2\theta = \pi - \mu$ puis $\mu = 2\varphi$, ce qui conduit pour μ à l'équation standard du pendule

$$\ddot{\mu} = -\sin(\mu)$$

avec l'intégrale première

$$\dot{\varphi}^2 + \sin^2(\varphi) = c^2$$

d'où

$$dt = \pm \frac{1}{c} \frac{d\varphi}{\sqrt{1 - \frac{1}{c^2} \sin^2(\varphi)}}$$

(nous choisissons le signe +). L'intégration en termes d'intégrales elliptiques de première espèce F de module $\frac{1}{c}$ donne, comme nous l'avons longuement expliqué aux sections 7 et 11 du chapitre 9,

$$t = \int_0^t dt = \frac{1}{c} \int_{\varphi(0)}^{\varphi(t)} \frac{d\varphi}{\sqrt{1 - \frac{1}{c^2} \sin^2(\varphi)}} = \frac{1}{c} \left(F\left(\varphi(t), \frac{1}{c}\right) - F\left(\varphi(0), \frac{1}{c}\right) \right).$$

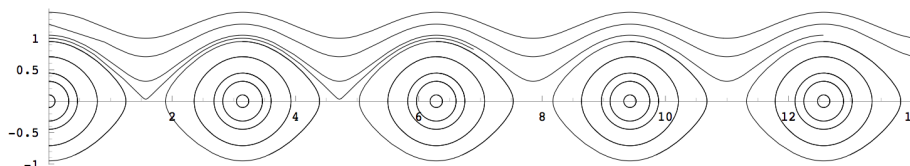


FIGURE 11. Trajectoires du pendule gouvernant les géodésiques.

Le plus simple est de prendre la condition initiale $\varphi(0) = 0$ (c'est-à-dire $\theta(0) = \theta_0 = \frac{\pi}{2}$) ce qui donne, puisque $F\left(0, \frac{1}{c}\right) = 0$,

$$\theta(t) = \frac{\pi}{2} - \operatorname{am}\left(tc, \frac{1}{c}\right) + k\pi$$

où am est l'amplitude de Jacobi. On en déduit (cf. les formules de dérivation de la section 7.4 du chapitre 9)

$$\dot{\theta}(t) = -c \operatorname{dn}\left(tc, \frac{1}{c}\right)$$

et donc, puisque $\operatorname{dn}\left(0, \frac{1}{c}\right) = 1$, $\dot{\theta}(0) = -c$. La figure 11 montre les trajectoires $\varphi(t) = \frac{\pi}{2} - \theta(t)$ dans le plan $(\varphi(t), \dot{\varphi}(t))$ lorsque c varie. Pour $c > 1$ le pendule accomplit des tours complets, la vitesse angulaire étant de module minimal $|\dot{\varphi}(t)| = \sqrt{c^2 - 1} > 0$. Pour $c < 1$, le pendule oscille au contraire entre $-\operatorname{Arcsin}(c)$ et $\operatorname{Arcsin}(c)$, points où $\dot{\varphi}(t)$ s'annule, la vitesse angulaire extrême, $\dot{\varphi}(0) = \pm c$, étant réalisée en 0.

Dans le cas $c > 1$ (module $\frac{1}{c}$ des fonctions elliptiques < 1),

$$\dot{x} = \cos^2(\theta) = \sin^2(\varphi)$$

donne

$$dx = \frac{1}{c} \frac{\sin^2(\varphi) d\varphi}{\sqrt{1 - \frac{1}{c^2} \sin^2(\varphi)}}$$

soit

$$x(t) = \int_0^x dx = \frac{1}{c} \int_{\varphi(0)=0}^{\varphi(t)} \frac{\sin^2(\varphi) d\varphi}{\sqrt{1 - \frac{1}{c^2} \sin^2(\varphi)}} = c \left(F\left(\varphi(t), \frac{1}{c}\right) - E\left(\varphi(t), \frac{1}{c}\right) \right)$$

où E est l'intégrale elliptique de seconde espèce. Mais comme $t = \frac{1}{c} \left(F\left(\varphi(t), \frac{1}{c}\right) \right)$, on a donc

$$x(t) = c^2 t - c E\left(\varphi(t), \frac{1}{c}\right)$$

avec $\varphi(t) = \operatorname{am}\left(tc, \frac{1}{c}\right) + k\pi$.

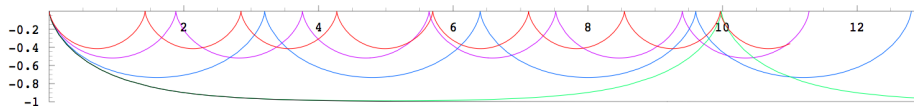


FIGURE 12. Différentes géodésiques sous-riemanniennes pour c^2 décroissant de 2 à 1 : $c^2 = 2$ (rouge), $c^2 = 1.5$ (violet), $c^2 = 1.1$ (bleu), $c = 1.001$ (vert), $c^2 = 1$ (noir). (D'après les calculs d'A. Agrachev).

Pour $y(t)$ l'intégration est triviale puisque $\dot{y} = -\cos(\varphi) \sin(\varphi)$ donne

$$dy = cd \left(\sqrt{1 - \frac{1}{c^2} \sin^2(\varphi)} \right)$$

et donc

$$y(t) = c \left[\sqrt{1 - \frac{1}{c^2} \sin^2(\varphi)} \right]_0^{\varphi(t)} = c \left(\sqrt{1 - \frac{1}{c^2} \sin^2(\varphi)} - 1 \right).$$

Mais comme $tc = F(\varphi(t), \frac{1}{c})$, on a $\sqrt{1 - \frac{1}{c^2} \sin^2(\varphi)} = \operatorname{dn}(tc, \frac{1}{c})$ et donc

$$y(t) = c \left(\operatorname{dn}\left(tc, \frac{1}{c}\right) - 1 \right).$$

La figure 12 montre différentes géodésiques pour c^2 décroissant de 2 à 1.

Pour $c < 1$ (module $\frac{1}{c}$ des fonctions elliptiques > 1), il faut faire le changement de variable

$$c \sin(\psi) = \sin(\varphi) = \cos(\theta)$$

pour changer le module des fonctions elliptiques de $\frac{1}{c}$ en c . φ variant entre $-\operatorname{Arcsin}(c)$ et $\operatorname{Arcsin}(c)$, ψ varie entre 0 et π . On trouve

$$\begin{cases} dt = \pm \frac{d\psi}{\sqrt{1 - c^2 \sin^2(\psi)}} \\ t = F(\psi(t), c) \\ \psi(t) = \operatorname{am}(t, c) \\ \theta(t) = \operatorname{Arccos}(c \operatorname{am}(t, c)) \end{cases}$$

L'intégration de $x(t)$ et $y(t)$ donne alors

$$\begin{cases} x(t) = t - E(\operatorname{am}(t, c), c) \\ y(t) = c(\operatorname{cn}(t, c) - 1) \end{cases}$$

La figure 13 montre différentes géodésiques pour c^2 croissant de 0.1 à 1. On remarquera leurs singularités cusp.

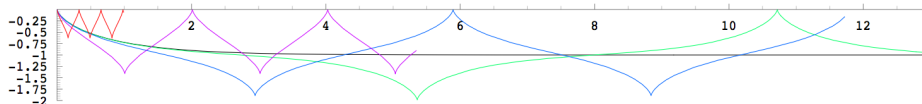


FIGURE 13. Différentes géodésiques sous-riemanniennes pour c^2 croissant de 0.1 à 1 : $c^2 = 0.1$ (rouge), $c^2 = 0.5$ (violet), $c^2 = 0.9$ (bleu), $c^2 = 0.99$ (vert), $c^2 = 1$ (noir). (D'après les calculs d'A. Agrachev).

4.3. Du portrait de phase aux géodésiques

Les sections précédentes ont exposé le développement des modèles \mathbb{V}_J et \mathbb{V}_S au début des années 2000. À la fin des années 2000, Yuri Sachkov et Igor Moiseev ont approfondi leur structure dans plusieurs articles [373], [478] et [479].

Étant donné que les géodésiques sont des courbes à orientation “pendulaire” de la tangente on peut utiliser, comme pour les elasticæ, le portrait de phase du pendule comme fil directeur pour comprendre la structure et les symétries des géodésiques. C'est qu'ont fait Sachkov et Moiseev en utilisant une variante des équations ci-dessus.¹⁵ Comme $H = \frac{1}{2}(\lambda_1^2 + \lambda_2^2)$ est constant, on peut prendre

$$\{\lambda_1 = r \cos(\alpha), \lambda_2 = r \sin(\alpha)\}$$

(r est le c de la section précédente). Ici r est constant et α variable alors que ci-dessus nous avons pris

$$\{\lambda_x = \rho \cos(\beta), \lambda_y = \rho \sin(\beta)\}$$

avec ρ et β tous deux constants.¹⁶

Avec $r = 1$, $H = \frac{1}{2}$ (paramétrisation par la longueur d'arc). On a alors $\lambda_1 = \cos(\alpha) = \rho \cos(\theta - \beta)$ et $\lambda_2 = \sin(\alpha)$. Pour retrouver exactement l'équation du pendule, Sachkov fait, comme ci-dessus, un changement d'angle

$$\gamma^S = 2\alpha + \pi$$

(γ^S est donc de période 4π) et obtient

$$\left\{ \lambda_1 = \sin\left(\frac{\gamma^S}{2}\right), \lambda_2 = -\cos\left(\frac{\gamma^S}{2}\right) \right\}.$$

En notant $\dot{\gamma}^S = c^S$ ($\dot{\gamma}^S = 2\dot{\alpha} = 2\lambda_3$), les équations de Hamilton dans les fibres s'écrivent

$$\begin{cases} \dot{\gamma}^S = c^S = 2\lambda_3 \\ \dot{c}^S = -\sin(\gamma^S) \end{cases}$$

15. Pour que le lecteur intéressé puisse se référer facilement aux textes originaux nous gardons les notations de Sachkov. Pour éviter les confusions nous les décorerons d'un S lorsqu'elles seront déjà utilisées plus haut dans un autre sens.

16. C'est volontairement que nous utilisons ces deux variantes afin que le lecteur puisse se référer facilement aux articles originaux de Sachkov et Moiseev.

(équation du pendule $\ddot{\gamma}^S = -\sin(\gamma^S)$ pour γ^S) et

$$\begin{cases} \dot{x} = \sin\left(\frac{\gamma^S}{2}\right) \cos(\theta) = \lambda_1 \cos(\theta) = \rho \cos(\theta - \beta) \cos(\theta) \\ \dot{y} = \sin\left(\frac{\gamma^S}{2}\right) \sin(\theta) = \lambda_1 \sin(\theta) = \rho \cos(\theta - \beta) \sin(\theta) \\ \dot{\theta} = -\cos\left(\frac{\gamma^S}{2}\right) = \lambda_\theta = \lambda_2 \\ \ddot{\theta} = \frac{1}{2} \sin\left(\frac{\gamma^S}{2}\right) \dot{\gamma}^S = \dot{\lambda}_\theta = \dot{\lambda}_2 = \frac{1}{2} \rho^2 \sin(2(\theta - \beta)) \end{cases}$$

le tout étant bien cohérent car

$$\dot{c}^S = 2\dot{\lambda}_3 = 2\dot{\theta}\lambda_1 = 2\lambda_1\lambda_2 = \sin(2\alpha) = \sin(\gamma^S - \pi) = -\sin(\gamma^S) .$$

On notera que ces équations s'écrivent de façon particulièrement élégante sous la forme

$$\dot{q} = \sin\left(\frac{\gamma^S}{2}\right) X_1 - \cos\left(\frac{\gamma^S}{2}\right) X_2$$

qui montre comment \dot{q} , qui est bien sûr dans le plan de contact $\text{Span}\{X_1, X_2\}$, évolue en étant contrôlé par le mouvement pendulaire.

Les cusps et les points d'inflexion des *projections* des géodésiques sur le plan (x, y) correspondent respectivement à

$$- \{\dot{x}, \dot{y}\} = 0, \text{ par suite } \lambda_1 = \rho \cos(\theta - \beta) = 0, \text{ i.e. } \sin\left(\frac{\gamma^S}{2}\right) = 0 \text{ et donc}$$

$$\dot{q} = -\cos\left(\frac{\gamma^S}{2}\right) X_2$$

porté par $X_2 = \partial_\theta$,

$$- \dot{\theta} = \lambda_2 = 0, \text{ i.e. } \cos\left(\frac{\gamma^S}{2}\right) = 0 \text{ et donc}$$

$$\dot{q} = \sin\left(\frac{\gamma^S}{2}\right) X_1$$

porté par X_1 .

Sachkov et Moiseev reprennent alors la stratification du portrait de phase \mathcal{C} du pendule, stratification qui organise la classification des géodésiques. Le fait que γ^S soit de période 4π est très important car il dédouble la stratification. \mathcal{C} est le cylindre de base le cercle $\mathbb{S}_{4\pi}^1$ et de génératrice l'axe des c^S . L'énergie du pendule dans la description $\lambda^S = (\gamma^S, c^S)$ est

$$E = \frac{1}{2} (\dot{\gamma}^S)^2 - \cos(\gamma^S) = \frac{1}{2} (c^S)^2 - \cos(\gamma^S) .$$

Son minimum est -1 . Comme nous l'avons déjà vu dans le cas des elasticæ aux sections 2 et 3 du chapitre précédent (13), \mathcal{C} est stratifié et se décompose en strates qui sont des sous-variétés de dimensions respectives 2, 1, 0 (de codimension 0, 1,

2) sans bord mais dont le bord est composé de strates de dimension inférieure. Les strates (pour $\gamma^S \in [-\pi, 3\pi]$, γ^S étant de période 4π) sont les suivantes :

1. C_1 : la strate ouverte de dimension 2 des trajectoires oscillantes. Elle est limitée par la séparatrice Σ d'équation $E = 1$ et contient aussi dans son adhérence l'origine $C_4^0 = (0, 0)$ et le point $C_4^{2\pi} = (2\pi, 0)$, ces deux points représentant le pendule au repos. On a $E \in]-1, 1[$, le pendule oscille entre les angles $\pm\gamma_{\max}^S$ avec une vitesse $c^S \in \left[\sqrt{2(E-1)}, \sqrt{2(E+1)} \right]$. C_1 a deux composantes connexes C_1^0 et $C_1^{2\pi}$ car γ^S est de période 4π .
2. C_2^\pm : les 2 strates ouvertes de dimension 2 des trajectoires tournantes. $E > 1$, le pendule accomplit des tours complets dans un sens ou dans l'autre.
3. C_3^\pm : les deux strates de dimension 1 constituant la séparatrice Σ d'équation $E = 1$ moins ses points singuliers $C_5^\pi = (\pi, 0)$ et $C_5^{3\pi} = (-\pi, 0) = (3\pi, 0)$. Elles ont également chacune deux composantes connexes $C_3^{0,\pm}$ et $C_3^{2\pi,\pm}$ car γ^S est de période 4π .
4. $C_4^0 = (0, 0)$ et $C_4^{2\pi} = (2\pi, 0)$: strates ponctuelles de dimension 0. $E = -1$, $c = 0$, $\gamma^S = 2\pi k$, i.e. $\cos(\gamma^S) = 1$. C'est la position stable du pendule au repos.
5. $C_5^\pi = (\pi, 0)$ et $C_5^{3\pi} = (3\pi, 0)$: strates ponctuelles de dimension 0. $E = 1$, $c = 0$ et $\gamma^S = \pi + 2\pi k$, i.e. $\cos(\gamma^S) = -1$. C'est la position d'équilibre instable du pendule.

Les figures 14 et 15 représentent la fonction énergie et le portrait de phase du pendule.

Dans cette variante, les formules des géodésiques que nous avons obtenues plus haut en termes de fonctions elliptiques s'expriment de la façon suivante au moyen des coordonnées de Jacobi (φ, k) introduites à la section 3 du chapitre 13 pour les *elasticæ*.¹⁷

- k est le module de l'intégrale elliptique utilisée : il encode l'énergie E ;
- φ est le temps-longueur "pendulaire" : il encode la longueur de la géodésique.

Dans les coordonnées de Jacobi, le système "vertical" dans les fibres devient trivial car $\dot{k} = 0$ et $\dot{\varphi} = 1$, i.e.

$$\varphi_t = \varphi + t$$

avec $\varphi = \varphi_0$, ce qui signifie que ces coordonnées de Jacobi "*rectifient*" la dynamique du pendule $\ddot{\gamma}^S = -\sin(\gamma^S)$. Pour intégrer le système "horizontal", il faut

17. On les comparera d'ailleurs à celles de ces *elasticæ*.

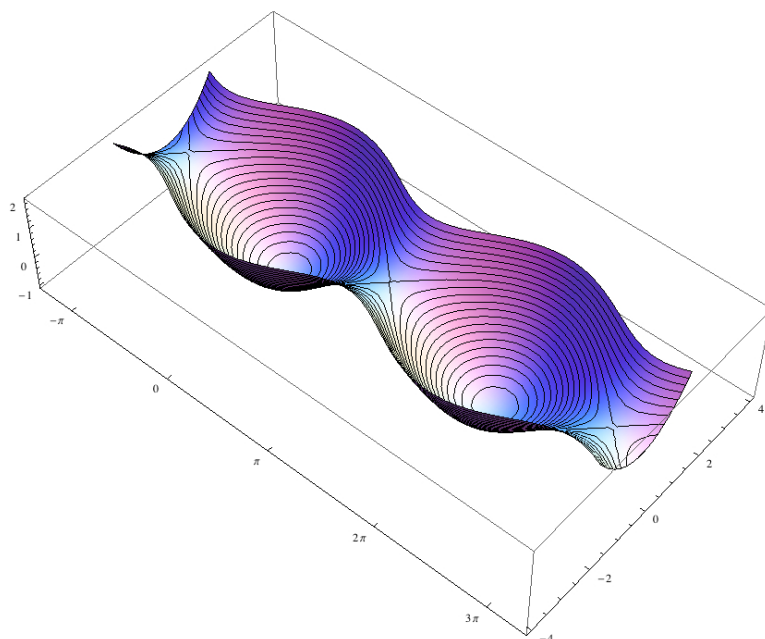


FIGURE 14. La fonction énergie du pendule $E = \frac{1}{2} (c^S)^2 - \cos(\gamma^S)$ pour $\gamma^S \in [-\pi, 3\pi]$ (période de 4π) et $c \in [-4, 4]$. Quelques lignes de niveau sont représentées.

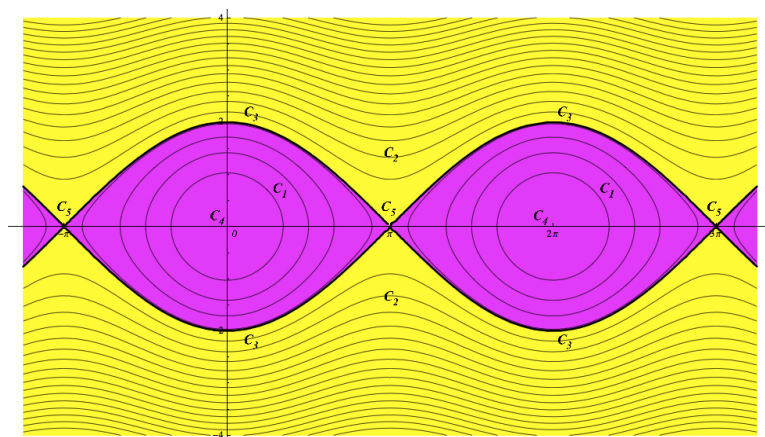


FIGURE 15. Le portrait de phase du pendule et sa stratification pour $\gamma^S \in [-\pi, 3\pi]$ (période de 4π) et $c \in [-4, 4]$. La strate ouverte C_1 (deux composantes connexes) est en magenta. La strate ouverte C_2 (deux composantes connexes) est en jaune. La strate C_3 de dimension 1 (quatre composantes connexes) est en traits épais. Les strates ponctuelles sont C_4 (deux points) et C_5 (deux points mod (4π) , trois représentés).

d'abord calculer c^S , $\sin\left(\frac{\gamma^S}{2}\right)$, $\cos\left(\frac{\gamma^S}{2}\right)$, $\cos(\theta)$, $\sin(\theta)$ pour reformuler \dot{x} , \dot{y} , $\dot{\theta}$ et ensuite intégrer en termes de fonction elliptiques, ce qui est assez facile car on connaît les intégrales des fonctions et des produits de fonctions de Jacobi.¹⁸

Dans les strates C_1 , $k = \sqrt{\frac{E+1}{2}} \in]0, 1[$ et φ est de période $4K(k)$ où $K(k)$ est l'intégrale elliptique complète qui, rappelons-le, est une fonction croissante partant de $K(k) = \frac{\pi}{2}$ et divergeant pour $k \rightarrow 0^-$. Les deux strates C_1^0 et $C_1^{2\pi}$ sont distinguées par le signe sg_1 de $\cos\left(\frac{\gamma^S}{2}\right)$: $sg_1 = +$ dans C_1^0 où $\frac{\gamma^S}{2} \in]-\frac{\pi}{2}, 0, \frac{\pi}{2}[$ et $sg_1 = -$ dans $C_1^{2\pi}$ où $\frac{\gamma^S}{2} \in]\frac{\pi}{2}, \pi, \frac{3\pi}{2}[$. Sachkov et Moiseev trouvent (avec $x_t = x(t)$, etc. et $\mathbb{E}(\varphi) = E(\text{am}(\varphi), k) = \int_0^\varphi \text{dn}^2(r, k) dr$),¹⁹

$$\left\{ \begin{array}{l} c^S = 2k \text{cn}(\varphi, k) \\ \sin\left(\frac{\gamma^S}{2}\right) = sg_1 k \text{sn}(\varphi, k) \\ \cos\left(\frac{\gamma^S}{2}\right) = sg_1 \text{dn}(\varphi, k) \quad (sg_1 = \text{signe de } \cos\left(\frac{\gamma^S}{2}\right)) \\ \cos(\theta_t) = \text{cn}(\varphi, k) \text{cn}(\varphi_t, k) + \text{sn}(\varphi, k) \text{sn}(\varphi_t, k) \\ \sin(\theta_t) = sg_1 (\text{sn}(\varphi, k) \text{cn}(\varphi_t, k) - \text{cn}(\varphi, k) \text{sn}(\varphi_t, k)) \\ x_t = sg_1 \frac{1}{k} (\text{cn}(\varphi, k) (\text{dn}(\varphi, k) - \text{dn}(\varphi_t, k)) + \text{sn}(\varphi, k) (t + \mathbb{E}(\varphi) - \mathbb{E}(\varphi_t))) \\ y_t = \frac{1}{k} (\text{sn}(\varphi, k) (\text{dn}(\varphi, k) - \text{dn}(\varphi_t, k)) - \text{cn}(\varphi, k) (t + \mathbb{E}(\varphi) - \mathbb{E}(\varphi_t))) \\ \theta_t = sg_1 (\text{am}(\varphi) - \text{am}(\varphi_t)) \pmod{2\pi} . \end{array} \right.$$

La dernière formule pour θ_t dérive directement des formules

$$\cos(\text{am}(\varphi)) = \text{cn}(\varphi, k), \quad \sin(\text{am}(\varphi)) = \text{sn}(\varphi, k)$$

18. Dans le calcul numérique de θ à partir de $\cos(\theta)$ et $\sin(\theta)$ en utilisant des arccos et des arcsin, il faut tenir compte du fait que les logiciels comme *Mathematica* utilisent les déterminations $\arccos \in [0, \pi]$ et $\arcsin \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Il faut donc utiliser une instruction du type `If[arcsin > 0, Then arccos, Else - arccos]`.

19. Rappelons que E est l'intégrale elliptique de deuxième espèce et que

$$\begin{aligned} \int_\varphi^{\varphi_t} \text{dn}^2(u) du &= \mathbb{E}(\varphi_t) - \mathbb{E}(\varphi) , \\ \int_\varphi^{\varphi_t} \text{sn}^2(u) du &= \frac{1}{k^2} (t - (\mathbb{E}(\varphi_t) - \mathbb{E}(\varphi))) , \\ \int_\varphi^{\varphi_t} \text{dn}(u) \text{sn}(u) du &= \text{cn}(\varphi) - \text{cn}(\varphi_t) . \end{aligned}$$

de la trigonométrie “chronométrique” du pendule.²⁰ On vérifie que la dérivée de y_t en $t = 0$ étant nulle,²¹ les projections sur le plan (x, y) des géodésiques partant de 0 ont bien une tangente horizontale $\theta = 0$.

Dans les strates C_2 , $k = \sqrt{\frac{2}{E+1}} \in]0, 1[$ et φ est de période $4kK(k)$, les deux strates C_2^\pm étant distinguées par le signe sg_2 de c^S .²² On trouve

$$\left\{ \begin{array}{l} c^S = sg_2 \frac{2}{k} \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \quad (sg_2 = \text{signe de } c) \\ \sin \left(\frac{\gamma^S}{2} \right) = sg_2 \operatorname{sn} \left(\frac{\varphi}{k}, k \right) \\ \cos \left(\frac{\gamma^S}{2} \right) = \operatorname{cn} \left(\frac{\varphi}{k}, k \right) \\ \cos(\theta_t) = k^2 \operatorname{sn} \left(\frac{\varphi}{k}, k \right) \operatorname{sn} \left(\left(\frac{\varphi}{k} \right)_t, k \right) + \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \operatorname{dn} \left(\left(\frac{\varphi}{k} \right)_t, k \right) \\ \sin(\theta_t) = k \left(\operatorname{sn} \left(\frac{\varphi}{k}, k \right) \operatorname{dn} \left(\left(\frac{\varphi}{k} \right)_t, k \right) - \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \operatorname{sn} \left(\left(\frac{\varphi}{k} \right)_t, k \right) \right) \\ x_t = sg_2 k \left(\operatorname{dn} \left(\frac{\varphi}{k}, k \right) \left(\operatorname{cn} \left(\frac{\varphi}{k}, k \right) - \operatorname{cn} \left(\left(\frac{\varphi}{k} \right)_t, k \right) \right) + \operatorname{sn} \left(\frac{\varphi}{k}, k \right) \left(\frac{t}{k} + \mathbb{E} \left(\frac{\varphi}{k} \right) - \mathbb{E} \left(\left(\frac{\varphi}{k} \right)_t \right) \right) \right) \\ y_t = sg_2 \left(k^2 \operatorname{sn} \left(\frac{\varphi}{k}, k \right) \left(\operatorname{cn} \left(\frac{\varphi}{k}, k \right) - \operatorname{cn} \left(\left(\frac{\varphi}{k} \right)_t, k \right) \right) - \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \left(\frac{t}{k} + \mathbb{E} \left(\frac{\varphi}{k} \right) - \mathbb{E} \left(\left(\frac{\varphi}{k} \right)_t \right) \right) \right) \end{array} \right.$$

On vérifie que la dérivée de y_t en $t = 0$ étant nulle,²³ les projections sur le plan (x, y) des géodésiques partant de 0 ont bien une tangente horizontale $\theta = 0$.

Pour la strate limite C_3 d'équation $E = 1$, on a $k = 1$, $\varphi \in \mathbb{R}$ (la période est devenue infinie). Les quatre strates $C_3^{0,\pm}$ et $C_3^{2\pi,\pm}$ sont distinguées par sg_1 et sg_2 . Les limites pour $k = 1$ vues à la section 7 du chapitre 9 donnent

20. C'est un exercice amusant sur les fonctions elliptiques de vérifier *directement* à partir de ces formules un peu compliquées ce que l'on sait depuis le début à savoir que, puisque la paramétrisation t est celle par la longueur d'arc et que la vitesse est donc de module constant $= 1$, on a bien $(x'_t)^2 + (y'_t)^2 + (\theta'_t)^2 = 1$.

21. Elle vaut $\frac{1}{k} \operatorname{cn} \left(\frac{\varphi}{k}, k \right) \left(-1 + k^2 \operatorname{sn}^2 \left(\frac{\varphi}{k}, k \right) + \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \sqrt{1 - k^2 \operatorname{sn}^2 \left(\frac{\varphi}{k}, k \right)} \right) = \frac{1}{k} \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \left(-\operatorname{dn}^2 \left(\frac{\varphi}{k}, k \right) + \operatorname{dn}^2 \left(\frac{\varphi}{k}, k \right) \right) = 0$.

22. On voit que la limite $k = 1$ est bien la même pour C_1 et C_2 et correspond à la valeur $E = 1$ pour laquelle $\frac{2}{E+1} = \frac{E+1}{2}$. Elle correspond à la strate séparatrice C_3 .

23. Elle vaut $\frac{1}{k} \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \left(-1 + k^2 \operatorname{sn}^2 \left(\frac{\varphi}{k}, k \right) + \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \sqrt{1 - k^2 \operatorname{sn}^2 \left(\frac{\varphi}{k}, k \right)} \right) = \frac{1}{k} \operatorname{dn} \left(\frac{\varphi}{k}, k \right) \left(-\operatorname{dn}^2 \left(\frac{\varphi}{k}, k \right) + \operatorname{dn}^2 \left(\frac{\varphi}{k}, k \right) \right) = 0$.

$$\left\{ \begin{array}{l} c^S = sg_2 \frac{2}{\cosh(\varphi)} \\ \sin\left(\frac{\gamma^S}{2}\right) = sg_1 sg_2 \tanh(\varphi) \\ \cos\left(\frac{\gamma^S}{2}\right) = sg_1 \frac{1}{\cosh(\varphi)} \\ \cos(\theta_t) = \frac{1}{\cosh(\varphi) \cosh(\varphi_t)} + \tanh(\varphi) \tanh(\varphi_t) \\ \sin(\theta_t) = sg_1 \left(\frac{\tanh(\varphi)}{\cosh(\varphi_t)} - \frac{\tanh(\varphi_t)}{\cosh(\varphi)} \right) \\ x_t = sg_1 sg_2 \left(\left(\frac{1}{\cosh(\varphi)} \right) \left(\frac{1}{\cosh(\varphi)} - \frac{1}{\cosh(\varphi_t)} \right) + \tanh(\varphi) (t + \tanh(\varphi) - \tanh(\varphi_t)) \right) \\ y_t = sg_2 \left(\tanh(\varphi) \left(\frac{1}{\cosh(\varphi)} - \frac{1}{\cosh(\varphi_t)} \right) - \frac{1}{\cosh(\varphi)} (t + \tanh(\varphi) - \tanh(\varphi_t)) \right) \end{array} \right.$$

En C_4 , $\gamma^S = \frac{\gamma^S}{2} = 0$ et $c^S = 0$, ce qui correspond à la limite $k \rightarrow 0$ (trigonométrie circulaire) dans les formules de C_1 et donc à $\text{cn} \equiv \cos$, $\text{sn} \equiv \sin$, $\text{dn} \equiv 1$,

$$\mathbb{E}(\varphi) = \int_0^\varphi \text{dn}^2(r, k) dr = \varphi.$$

On a $c^S = 0$ et $\gamma^S = 0, 2\pi$ (et donc $\frac{\gamma^S}{2} = 0, \pi$), ce qui donne, puisque

$$\text{dn}(\varphi, k) \sim 1 - \frac{1}{2}k^2 \sin^2(\varphi)$$

pour $k \rightarrow 0$ et que $\varphi_t = \varphi + t$,

$$\left\{ \begin{array}{l} c^S = 2k \cos(\varphi) = 0 \\ \sin\left(\frac{\gamma^S}{2}\right) = sg_1 k \sin(\varphi) = 0 \\ \cos\left(\frac{\gamma^S}{2}\right) = sg_1 \text{dn}(\varphi, 0) = sg_1 \\ \cos(\theta_t) = \cos(\varphi) \cos(\varphi_t) + \sin(\varphi) \sin(\varphi_t) = \cos(\varphi - \varphi_t) = \cos(-t) \\ \sin(\theta_t) = sg_1 (\sin(\varphi) \cos(\varphi_t) - \cos(\varphi) \sin(\varphi_t)) = sg_1 \sin(-t) \\ \theta_t = -sg_1 t \pmod{2\pi} \\ x_t = sg_1 \frac{1}{k} (\cos(\varphi) (\text{dn}(\varphi, k) - \text{dn}(\varphi_t, k)) + \sin(\varphi) (t + \varphi - \varphi_t)) = 0 \\ y_t = \frac{1}{k} (\sin(\varphi, k) (\text{dn}(\varphi, k) - \text{dn}(\varphi_t, k)) - \cos(\varphi) (t + \varphi - \varphi_t)) = 0 \end{array} \right.$$

soit le mouvement uniforme sur le cercle fibre \mathbb{S}^1 au-dessus de l'origine ($x = 0, y = 0$) de \mathbb{R}^2 . Pour $k \sim 0$, on a au premier ordre l'approximation

$$\left\{ \begin{array}{l} \theta_t \sim -sg_1 t \pmod{2\pi} \\ x_t \sim sg_1 \frac{k}{4} (\cos(\varphi) - \cos(\varphi + 2t) + 2t \sin(\varphi)) \\ y_t \sim \frac{k}{2} (\cos(\varphi + t) \sin(t) - t \cos(\varphi)) \end{array} \right.$$

En C_5 , on a $c^S = 0$ et $\gamma^S = \pm\pi$, ce qui correspond à la limite $\varphi = \pm\infty$ dans les formules de C_5 et implique $\theta_t = 0$, $y_t = 0$ et $x_t = sg_1 sg_2 t$ où $sg_1 sg_2$ ($sg_1 sg_2 = 1$ si $\gamma^S = \pi$ et $= -1$ si $\gamma^S = -\pi$) est le signe de $\sin\left(\frac{\gamma^S}{2}\right)$ (la géodésique est la droite des

x). En effet, comme $c^S = 0$, on a, aux signes près, $\frac{1}{\cosh(\varphi)} = 0$, donc $\varphi = \infty$ et par suite $\tanh(\varphi) = 1$ et $\varphi_t = \varphi + t = \varphi = \infty$. D'où $x_t = t$ en C_5^π et $x_t = -t$ en $C_5^{3\pi}$.

À partir de ces formules explicites, qui peuvent sembler un peu compliquées mais sont en fait très simplement structurées, on peut étudier de façon systématique les comportements des géodésiques de $\mathbb{V}_S = SE(2)$, les points conjugués et le cut locus.

Remarque. Il est étonnant et remarquable que cette structure sous-riemannienne, celle de l'un des groupes les plus fondamentaux et apparemment les plus simples de la géométrie élémentaire, n'ait été élucidée qu'en 2006 à partir de motivations théoriques venant en partie des neurosciences de la vision. C'est un exemple très intéressant de feedback sur une théorie préexistante à partir d'une application modélisatrice. Il y en a plusieurs autres en histoire des sciences comme par exemple le feed back de la relativité générale sur la géométrie riemannienne, ou celui de la mécanique quantique à la Heisenberg sur la théorie des opérateurs linéaires, ou encore celui des catastrophes de l'embryogenèse sur la théorie des singularités et de leurs déploiements universels, ou encore celui des réseaux de neurones sur la physique statistique des verres de spins. □

4.4. Images de géodésiques

On dispose donc des formules pour les géodésiques

$$q(s) = \{x(s), y(s), \theta(s)\}$$

issues de l'origine²⁴ en fonction des conditions initiales du contrôle pendulaire

$$\lambda^S = (\gamma^S, c^S)$$

et du temps-longueur t . Elles réalisent l'application exponentielle $\text{Exp}_s(\gamma^S, c^S)$. Lorsque t varie de 0 à $+\infty$, ces trajectoires remplissent $SE(2)$ puisque (théorème de Chow) 0 peut être relié par une géodésique à n'importe quel point q de $SE(2)$ et l'orbite $\widetilde{\lambda^S} = \widetilde{(\gamma^S, c^S)}$ du pendule de contrôle pour une certaine valeur de l'énergie $E = \frac{1}{2}(c^S)^2 - \cos(\gamma^S)$ (et donc une certaine valeur de k puisque k code E) engendre une famille de géodésiques à 1 paramètre qui compose une surface plus ou moins compliquée $Q_{\widetilde{\lambda^S}}$.²⁵

4.4.1. *Considérations préliminaires.* Les images des deux strates C_1 sont séparées de celles des deux strates C_2 par celles des quatre strates C_3 . Pour C_1 et C_2 interviennent des fonctions elliptiques de φ qui possèdent une périodicité. Pour C_3 il n'y a plus de périodicité et $\varphi \in \mathbb{R}$ mais il y a en revanche une limite qui est atteinte

24. Géodésiques aussi notées $q_s = (x_s, y_s, \theta_s)$.
 25. Nous noterons suivant les cas $q(s)$ ou $\Gamma(s)$ (voire q_s ou Γ_s) une géodésique particulière paramétrée par s . Nous noterons Q_I des familles de géodésiques indexées par des paramètres I .

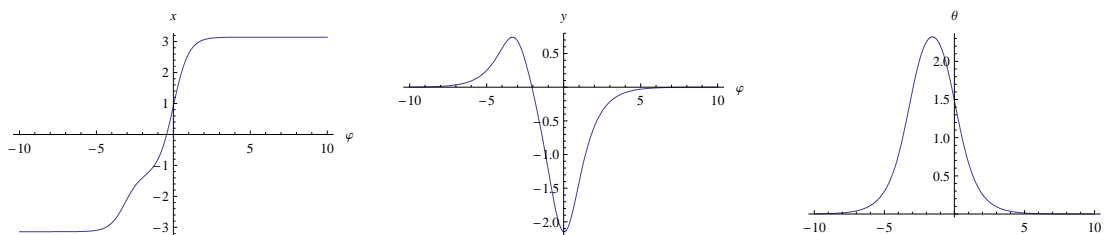


FIGURE 16. Les graphes de x_t , y_t et θ_t pour la moitié de la strate C_3 dans le cas $t = \pi$. Les limites pour $\varphi \rightarrow \infty$, $x_t = \pm\pi$, $y_t = 0$ et $\theta_t = 0$ sont atteintes exponentiellement vite. Avec $\varphi \in [-10, 10]$ on a déjà une approximation très inférieure à la résolution de l'image.

exponentiellement vite, ce qui fait que l'on peut restreindre φ à un intervalle avec une excellente approximation. Quand on considère les équations pour C_3 , on voit que la limite pour $\varphi \rightarrow \pm\infty$ est $\theta_t = 0$ (modulo 2π) ($\cos(\theta_t) = 1$ et $\sin(\theta_t) = 0$), $x_t = \pm sg_1 sg_2 t$ et $y_t = 0$. Ces limites sont atteintes extrêmement vite. La figure 16 montre le cas $t = \pi$ pour la moitié de la strate C_3 . On voit qu'avec $\varphi \in (-10, 10)$ on a déjà une approximation très inférieure à la résolution de l'image.

Mais la convergence est quand même assez subtile. En effet les graphes x_t , y_t et θ_t sont périodiques pour C_1 mais leur période tend lentement vers l'infini à cause de la divergence très lente de $K(k)$. Donc la convergence des graphes de C_1 vers C_3 est très rapide (exponentielle) pour la partie centrale à cause de la convergence des fonctions elliptiques vers des fonctions hyperboliques mais très lente en ce qui concerne le rejet des translatées de la partie centrale vers l'infini. La figure 17 explicite cette convergence. Toujours pour $t = \pi$, on a pris $k = 0.9999999$ déjà très proche de 1 et pourtant le très petit intervalle $\varphi \in [-20, 20]$ est encore suffisant pour manifester la périodicité.

Une deuxième remarque est qu'il existe une différence fondamentale entre les strates ouvertes C_1 et C_2 en ce qui concerne la *période* des fonctions elliptiques de Jacobi qui y interviennent. Pour C_1 , la période $T_1(k) = 4K(k)$ est toujours $> 2\pi$ puisque $K(0) = \frac{\pi}{2}$ et que $K(k)$ est une fonction croissante. Pour C_2 la période $T_2(k) = 4kK(k)$ est > 0 mais aussi petite que l'on veut. C'est l'origine de la structure très singulière du front d'onde. Nous y reviendrons en détail à la section 4.7. La figure 18 montre $T_1(k)$ et $T_2(k)$ pour les valeurs permises de k , $k \in (0, 1)$.

4.4.2. *De C_1 vers C_3 .* Pour C_1^0 , la surface géodésique $Q_{C_1^0, k}$ est une sorte de "cône" de sommet 0 mais

- (i) avec un "sommet" anisotrope allongé le long de la direction x et
- (ii) de forme plus ou moins spiralée et vrillée comme celle des coquilles de cérithes.

La figure 19 montre $Q_{C_1^0, k}$ pour $k = 0.1$ qui stabilise le cas dégénéré $\widetilde{\lambda^S} = \{C_4^0\}$: $x_t = 0$, $y_t = 0$, $\theta_t = -t \pmod{2\pi}$. Pour des raisons de lisibilité du graphique, nous

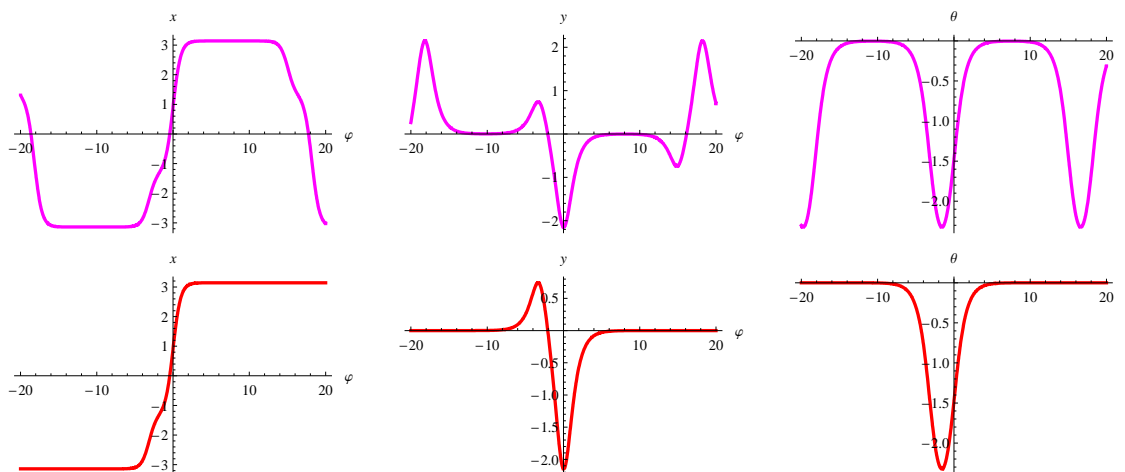


FIGURE 17. La convergence des graphes de C_1 vers C_3 . Elle est très rapide (exponentielle) pour la partie centrale mais très lente en ce qui concerne le rejet des translatées de la partie centrale vers l’infini.

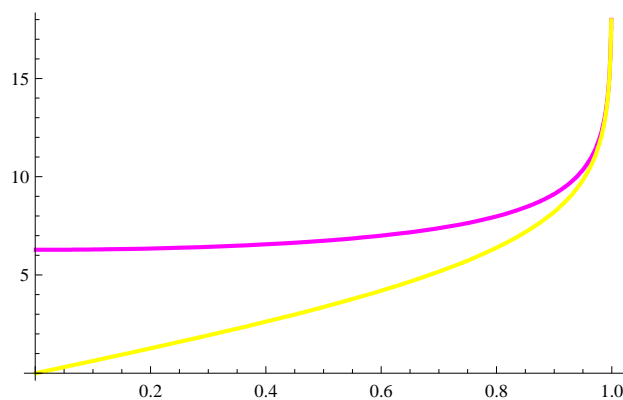


FIGURE 18. Les périodes $T_1(k) = 4K(k)$ pour C_1 et $T_2(k) = 4kK(k)$ pour C_2 pour les valeurs permises de k , $k \in (0, 1)$. $T_1(k)$ (magenta) commence en 2π puisque $K(0) = \frac{\pi}{2}$ et $T_2(k)$ (jaune) commence en 0.

orientons la fibre θ horizontalement et nous représentons θ_t et non pas $\theta_t \bmod (2\pi)$. La représentation $\bmod (2\pi)$ transformerait $Q_{C_4^0,k}$ en un “ouroboros” (cf. plus bas section 4.4.5). Les figures 20 et 21 montrent à l’autre extrême $Q_{\tilde{\lambda}^S} = Q_{C_1^0,k}$ pour $k = 0.9$. Le cône s’élargit et s’aplatit et l’anisotropie à l’origine devient de plus en plus importante. Elles montrent aussi la projection de la géométrie de $Q_{C_1^0,k}$ sur le plan (x, y) au voisinage de 0 (pour $k = 0.9$) ce qui permet de vérifier que toutes les projections des géodésiques démarrent bien avec $\theta = 0$ à l’origine $(0, 0)$. La figure 22 montre $Q_{C_1^0,k}$ pour $k = 0.999$, la limite $k = 1$ correspondant à la strate C_3 du

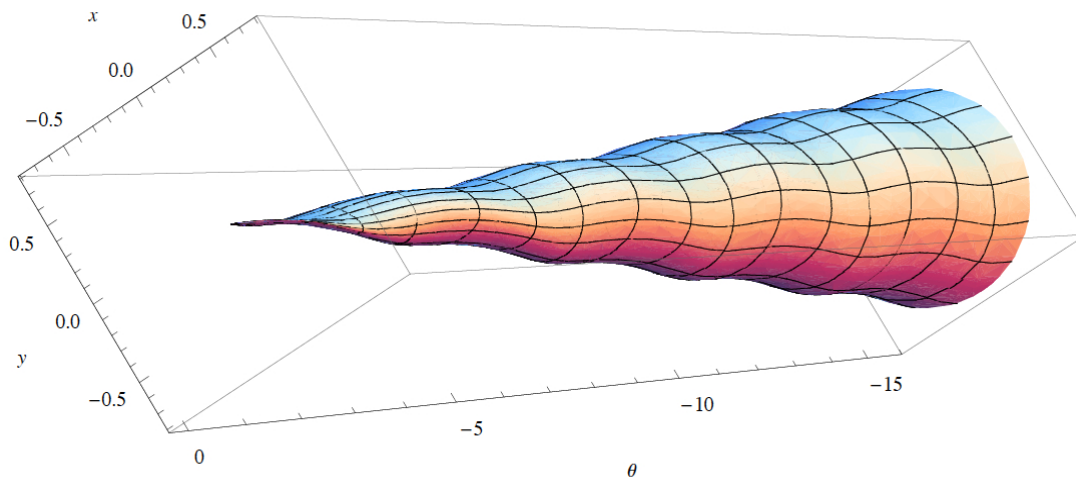


FIGURE 19. Le “cône” $Q_{C_1^0, k}$ des géodésiques issues de 0 pour le contrôle $k = 0.1$ ($E = -0.98$) qui stabilise le cas dégénéré $\lambda^S = \{C_4^0\}$ ($E = -1$ minimale) : $x_t = 0, y_t = 0, \theta_t = -t$. Il est proche du cône dégénéré contracté sur la fibre $\mathbb{S}_{(0,0)}^1$. Le temps-longueur est $t \in [0, 5\pi]$. Pour des raisons de lisibilité, le graphique représente θ_t et non pas $\theta_t \bmod (2\pi)$.

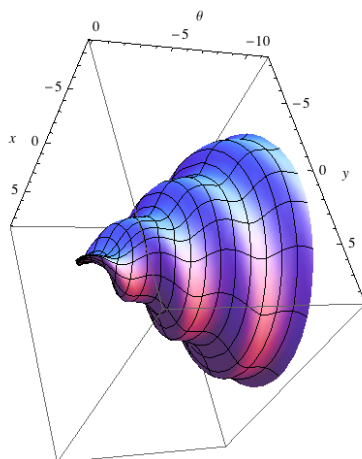


FIGURE 20. Le “cône” $Q_{\lambda^S} = Q_{C_1^0, k}$ des géodésiques issues de 0 pour le contrôle $k = 0.9$ ($E = 0.62$). Le temps-longueur est $t \in [0, 5\pi]$. Pour des raisons de lisibilité, le graphique représente θ_t et non pas $\theta_t \bmod (2\pi)$.

contrôle pendulaire. On voit que le “cône” s’évase de plus en plus jusqu’à devenir asymptotique au plan $\theta = -2\pi = 0 \bmod (2\pi)$ et que l’anisotropie de l’origine devient massive.

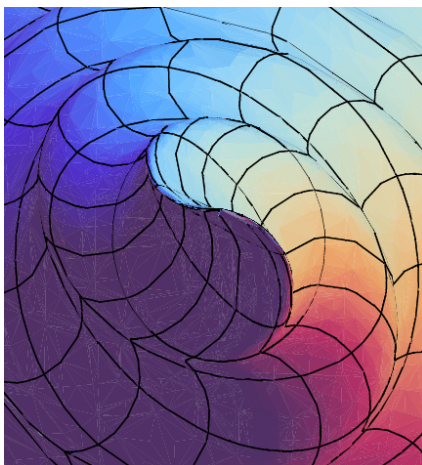


FIGURE 21. La projection de la géométrie de $Q_{C_1^0, k}$ sur le plan (x, y) au voisinage de 0. On vérifie que toutes les projections des géodésiques démarrent bien horizontalement à l'origine $(0, 0)$ avec $\theta = 0$.

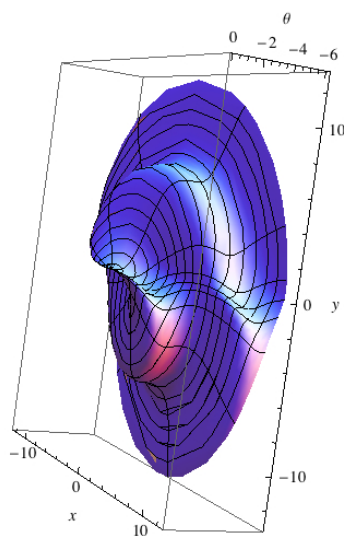


FIGURE 22. $Q_{\tilde{\lambda}^s} = Q_{C_1^0, k}$ pour $k = 0.999$, la limite $k = 1$ correspondant à la strate C_3 du contrôle pendulaire.

Pour la strate $C_1^{2\pi}$ on obtient les mêmes figures avec les changements de signe de y et θ .

Quand on arrive à la strate limite C_3 , on obtient une surface limite séparatrice Q_{C_3} composée de quatre composantes $Q_{C_3^{0,+}}$, $Q_{C_3^{0,-}}$, $Q_{C_3^{2\pi,+}}$, $Q_{C_3^{2\pi,-}}$. La composante $Q_{C_1^0, k \rightarrow 1^-}$ a pour limite la réunion $Q_{C_3^{0,+}} \cup Q_{C_3^{0,-}}$ qui est représentée à droite de la

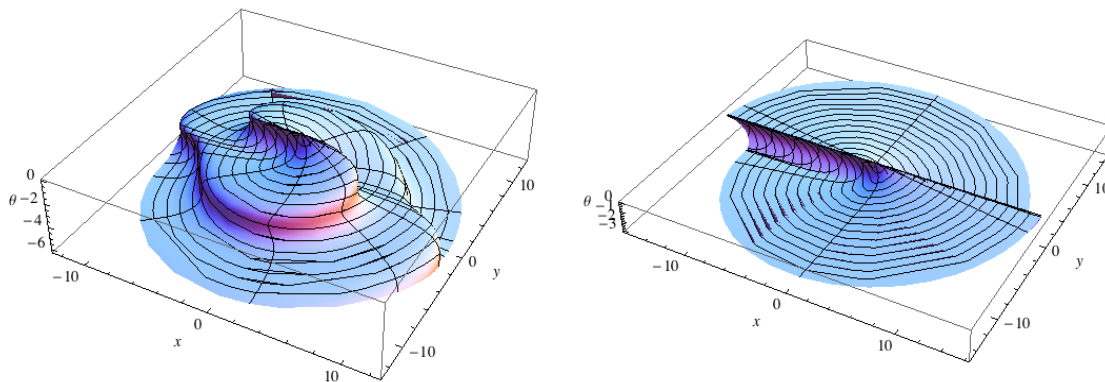


FIGURE 23. À droite, la limite $Q_{C_3^{0,+}} \cup Q_{C_3^{0,-}}$ de $Q_{C_1^{0,k \rightarrow 1^-}}$. À gauche, $Q_{C_1^{0,k=0.999}}$.

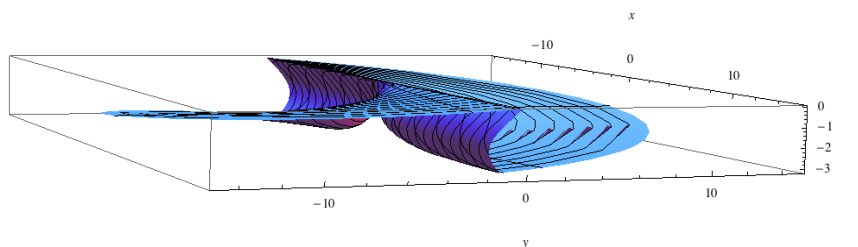


FIGURE 24. Une autre perspective sur la surface $Q_{C_3^{0,+}} \cup Q_{C_3^{0,-}}$.

figure 23. La divergence de la période $4K(k)$ de φ lorsque $k \rightarrow 1^-$, n'est notable que pour k extrêmement proche de 1. Pour $k = 0.999$ on a encore $K(0.999) \sim 4.5$ qui est très faible et $Q_{C_1^{0,k=0.999}}$ (à gauche) est encore assez loin de $Q_{C_3^{0,+}} \cup Q_{C_3^{0,-}}$. Mais on voit que l'anisotropie à l'origine devient l'axe des x qui correspond à $\varphi = \infty$ et que les lignes de niveau $\varphi = \text{cste}$ varient de plus en plus irrégulièrement. Elles stationnent longtemps au voisinage de l'axe des x pour les grandes valeurs de φ puis évoluent rapidement lorsque φ passe par 0 et repartent rapidement vers $\varphi = \infty$ (cf. plus bas la figure 27 des coordonnées de Q_{C_3}). La figure 24 montre une autre perspective sur $Q_{C_3^{0,+}} \cup Q_{C_3^{0,-}}$.

4.4.3. *Arrêt sur C_3 .* Profitons-en pour représenter la surface Q_{C_3} dans son ensemble avec ses quatre composantes. La figure 25 le fait, toujours pour $t \in [0, 5\pi]$.

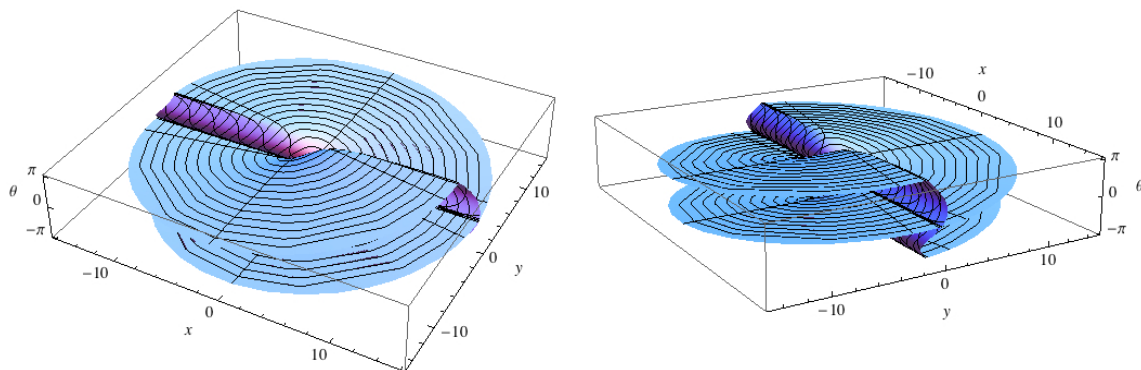


FIGURE 25. La surface $Q_{C_3} = \text{Exp}(C_3)$ de $SE(2)$ décrivant l'évolution de l'image de la strate C_3 lorsque t varie ($t \in [0, 5\pi]$ et $\varphi \in [-20, 20]$). On a choisi de positionner la fibre θ verticalement.

Q_{C_3} résulte de l'intrication de deux surfaces ressemblant à des surfaces “fronces”, chacune possédant deux nappes (l'une supérieure, l'autre inférieure) plutôt “horizontale” (θ variant lentement) se transformant progressivement l'une en l'autre à travers une nappe “neutre” en contournant la singularité à l'origine (le “centre organisateur” de la fronce dirait René Thom) et séparées par une nappe intermédiaire “verticale” en falaise (θ variant rapidement). La première est $Q_{C_3^{0,+} \cup C_3^{2\pi,-}}$ et la seconde $Q_{C_3^{0,-} \cup C_3^{2\pi,+}}$. La figure 26 représente ces deux surfaces. Si l'on considère celle de droite, $Q_{C_3^{0,-} \cup C_3^{2\pi,+}}$, la ligne médiane de la nappe intermédiaire “neutre” en falaise (dans la direction des x) correspond à $\varphi = \infty$ et se prolonge au-delà du “centre organisateur”. La ligne dans la direction des y correspond à $\varphi = 0$. On voit qu'au fur et à mesure que t croît, la hauteur maximale des nappes (supérieure et inférieure) tend exponentiellement vers $\theta = \pm\pi$, la nappe “neutre” passant par le méridien $\varphi = 0$. Nous y reviendrons plus précisément à la section 4.7 (cf. la figure 71).

La figure 27 montre les coordonnées (x, y, θ) de Q_{C_3} en fonction de φ pour $t = 5\pi$. On voit bien qu'elles restent très proches de leurs valeurs asymptotiques sauf dans des domaines restreints (en particulier au voisinage de $\varphi = 0$) où elles varient rapidement. Cela est caractéristique des formules à base d'exponentielles.

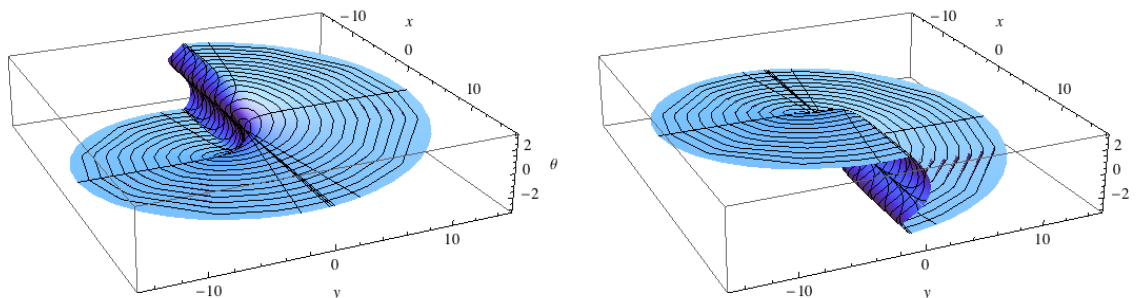


FIGURE 26. Les deux surfaces “fronce” composant la surface $Q_{C_3} = \text{Exp}(C_3)$ de la figure 25. À gauche $Q_{C_3^{0,+} \cup C_3^{2\pi,-}}$. À droite $Q_{C_3^{0,-} \cup C_3^{2\pi,+}}$. Cf. la figure 71 pour les lignes $\varphi = \pm\infty$ et $\varphi = 0$.

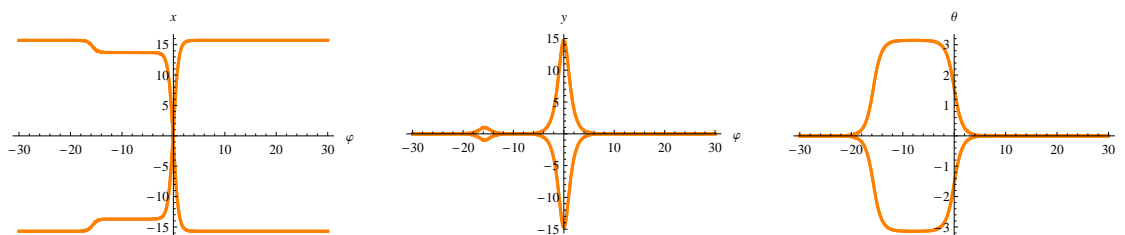


FIGURE 27. Les coordonnées (x, y, θ) de Q_{C_3} en fonction de φ pour $t = 5\pi$.

4.4.4. *De C_3 à C_2 .* Pour C_2 , la figure 28 montre $Q_{C_2^+, k=0.99}$ pour $k = 0.99$. Il s’agit d’une déformation de $Q_{C_3^{0,+} \cup C_3^{2\pi,+}}$ (et non plus de $Q_{C_3^{0,+}} \cup Q_{C_3^{0,-}}$ comme dans le cas de C_1). Cette déformation est représentée à la figure 29.

Mais dès que k décroît, la surface géodésique $Q_{C_2, k}$ devient très singulière et très compliquée. Des approximations jusqu’à l’ordre 3 permettent de comprendre ce qui se passe :

1. à l’ordre 1, x_t vaut $(\cos(\frac{\varphi}{k}) - \cos(\frac{\varphi}{k} + \frac{t}{k}))k$;
2. à l’ordre 3, x_t vaut

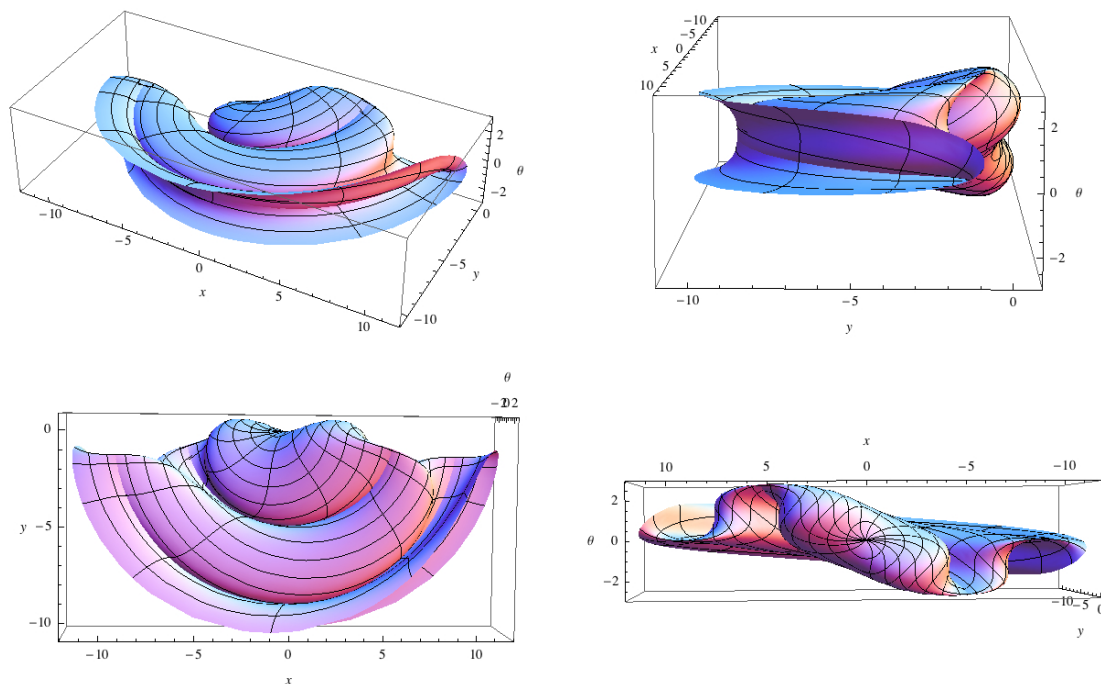


FIGURE 28. Plusieurs points de vue sur la surface géodésique $Q_{C_2^+, k=0.99}$ pour $k = 0.99$.

$$\begin{aligned}
 & \left(\cos\left(\frac{\varphi}{k}\right) - \cos\left(\frac{\varphi}{k} + \frac{t}{k}\right) \right) k + \\
 & \frac{1}{8} \left(\begin{array}{c} 4\left(\cos\left(\frac{\varphi}{k} + \frac{t}{k}\right) - \cos\left(\frac{\varphi}{k}\right)\right) \sin^2\left(\frac{\varphi}{k}\right) + \\ \left(2\frac{\varphi}{k} + 4\frac{t}{k} + \sin\left(2\frac{\varphi}{k}\right) - 2\sin\left(2\left(\frac{\varphi}{k} + \frac{t}{k}\right)\right)\right) \sin\left(\frac{\varphi}{k}\right) + \\ \sin\left(\frac{\varphi}{k} + \frac{t}{k}\right) \left(\sin\left(2\left(\frac{\varphi}{k} + \frac{t}{k}\right)\right) - 2\left(\frac{\varphi}{k} + \frac{t}{k}\right)\right) \end{array} \right) k^3 ;
 \end{aligned}$$

3. à l'ordre 2, y_t vaut

$$\begin{aligned}
 & \frac{1}{4} \left(-2\frac{t}{k} + 4\cos\left(\frac{\varphi}{k}\right) \sin\left(\frac{\varphi}{k}\right) - 4\cos\left(\frac{\varphi}{k} + \frac{t}{k}\right) \sin\left(\frac{\varphi}{k}\right) \right. \\
 & \quad \left. - \sin\left(2\frac{\varphi}{k}\right) + \sin\left(2\left(\frac{\varphi}{k} + \frac{t}{k}\right)\right) \right) k^2 ;
 \end{aligned}$$

4. à l'ordre 1, $\cos(\theta_t)$ vaut 1 ;

5. à l'ordre 2, $\cos(\theta_t)$ vaut

$$1 + \left(-\frac{1}{2} \sin^2\left(\frac{\varphi}{k}\right) + \sin\left(\frac{\varphi}{k} + \frac{t}{k}\right) \sin\left(\frac{\varphi}{k}\right) - \frac{1}{2} \sin^2\left(\frac{\varphi}{k} + \frac{t}{k}\right) \right) k^2 ;$$

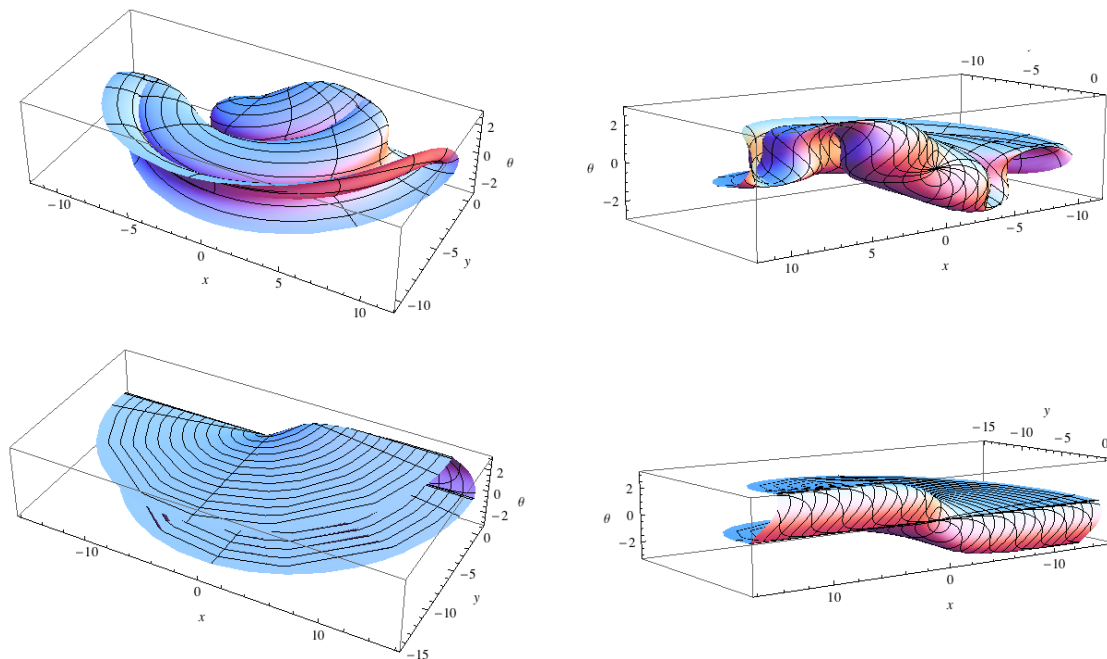


FIGURE 29. Plusieurs points de vue sur la surface géodésique $Q_{C_2^+, k=0.99}$ de la figure 28 (en haut) comme déformation de la surface limite géodésique limite $Q_{C_3^{0,+} \cup C_3^{2\pi,+}}$ (en bas).

- 6. à l'ordre 1, $\sin(\theta_t)$ vaut $(\sin(\frac{\varphi}{k}) - \sin(\frac{\varphi}{k} + \frac{t}{k}))k$;
- 7. à l'ordre 3, $\sin(\theta_t)$ vaut

$$\begin{aligned}
 & (\sin(\frac{\varphi}{k}) - \sin(\frac{\varphi}{k} + \frac{t}{k}))k + \\
 & \frac{1}{8} \left(\begin{array}{l} \cos(\frac{\varphi}{k})(\sin(2\frac{\varphi}{k}) - 2\frac{\varphi}{k}) + \\ \cos(\frac{\varphi}{k} + \frac{t}{k})(2(\frac{\varphi}{k} + \frac{t}{k}) - \sin(2(\frac{\varphi}{k} + \frac{t}{k}))) + \\ 4(\sin(\frac{\varphi}{k}) - \sin(\frac{\varphi}{k} + \frac{t}{k})) \sin(\frac{\varphi}{k}) \sin(\frac{\varphi}{k} + \frac{t}{k}) \end{array} \right) k^3 ;
 \end{aligned}$$

Notons Γ_n l'approximation à l'ordre n de la géodésique Γ . À l'ordre 1 on a donc pour Γ_1 les équations

$$\begin{cases} x_t \sim (\cos(\frac{\varphi}{k}) - \cos(\frac{\varphi}{k} + \frac{t}{k}))k \\ y_t \sim 0 \\ \theta_t \sim \text{Arcsin}((\sin(\frac{\varphi}{k}) - \sin(\frac{\varphi}{k} + \frac{t}{k}))k) \sim (\sin(\frac{\varphi}{k}) - \sin(\frac{\varphi}{k} + \frac{t}{k}))k \end{cases}$$

et $\Gamma_{1,t}$ est un cercle $C_{k,\varphi}$ de centre $\{k \cos(\frac{\varphi}{k}), k \sin(\frac{\varphi}{k})\}$ et de rayon k dans un voisinage $\{x, \theta\}$ de $\{x = 0, \theta = 0\}$ au-dessus de $y = 0$. Il démarre donc bien en 0 pour $t = 0$. Lorsque k est petit fixé et que $\varphi \in [0, 4kK(k)]$ est également fixé, $\Gamma_{1,t}$ décrit

le cercle $C_{k,\varphi}$ avec la période $\frac{t}{k} = 2\pi$, i.e. $t = 2\pi k$. Quand c'est k et t qui sont fixés et φ qui varie, la trajectoire de l'extrémité de $\Gamma_{1,t}$ est le cercle Σ_r de centre l'origine et de rayon $r^2 = 2k^2 \left(1 - \cos\left(\frac{t}{k}\right)\right)$.²⁶

Considérons l'exemple $k = 0.1$. La période de t est 0.2π et

$$\varphi \in [0, 4kK(k) \sim 0.6299] .$$

La figure 30 représente le bouquet de cercles $C_{k=0.1,\varphi}$ pour les 7 valeurs φ allant de 0 à 0.6299 par pas égaux. Elle représente aussi en traits épais oranges, les arcs de ces cercles pour t variant dans le quart de période 0.05π ainsi que le cercle des extrémités Σ_r pour

$$r = \left(2k^2 \left(1 - \cos\left(\frac{t}{k}\right)\right)\right)^{\frac{1}{2}} \sim 0.1414 .$$

Lorsque t varie à k constant et que les géodésiques $\Gamma_{1,t}$ s'enroulent sur les cercles $C_{k,\varphi}$, le rayon r de Σ_r varie périodiquement entre 0 et $r_k = k\sqrt{2}$ comme le montre la figure 31 à gauche. Lorsque c'est t qui est constant et que c'est k qui varie en décroissant vers 0, le rayon r de Σ_r varie aussi périodiquement mais en étant modulé par la droite $r_k = k\sqrt{2}$. Il oscille donc de plus en plus vite lorsque $k \rightarrow 0^+$ comme le montre la figure 31 à droite.

Cette structure de base au premier ordre se complexifie lorsque l'on développe $\{x_t, y_t, \theta_t\}$ aux ordres supérieurs. À l'ordre 2, x_t et θ_t restent les mêmes qu'à l'ordre 1 mais y_t ne s'annule plus et les cercles $C_{k,\varphi}$ deviennent des hélices serrées comme le montre la figure 32. À l'ordre 3, cette hélice déployée parallèlement à l'axe des y se désaxe et s'incline comme le montre également la figure 32.

La structure en bouquet de la figure 30 va ainsi se trouver déployée en hélices désaxées comme l'illustre la figure 33

4.4.5. *Représentations toriques.* Nous avons dit plus haut à propos des surfaces géodésiques $Q_{C_1^0,k}$ où θ_t était représenté comme variant dans \mathbb{R} et non pas mod (2π) que représenter $\theta_t \bmod (2\pi)$ transformerait les "cônes" en "ouroboros".

Pour le voir on peut utiliser des représentations toriques en traitant $SE(2)$ comme la fibration $\mathbb{R}^2 \times \mathbb{S}^1 \rightarrow \mathbb{S}^1$ de fibre \mathbb{R}^2 au-dessus de \mathbb{S}^1 , et non plus, comme nous l'avons fait jusqu'ici, comme la fibration $\mathbb{R}^2 \times \mathbb{S}^1 \rightarrow \mathbb{R}^2$ de fibre \mathbb{S}^1 au-dessus de \mathbb{R}^2 . Si l'on veut facilement visualiser cette fibration, il suffit alors d'utiliser par exemple

26. En effet

$$\begin{aligned} x_t &\sim k \left(\cos\left(\frac{\varphi}{k}\right) \left(1 - \cos\left(\frac{t}{k}\right)\right) + \sin\left(\frac{\varphi}{k}\right) \sin\left(\frac{t}{k}\right) \right) \\ \theta_t &\sim k \left(\sin\left(\frac{\varphi}{k}\right) \left(1 - \cos\left(\frac{t}{k}\right)\right) - \cos\left(\frac{\varphi}{k}\right) \sin\left(\frac{t}{k}\right) \right) \end{aligned}$$

et donc $x_t^2 + \theta_t^2 \sim 2k^2 \left(1 - \cos\left(\frac{t}{k}\right)\right)$.

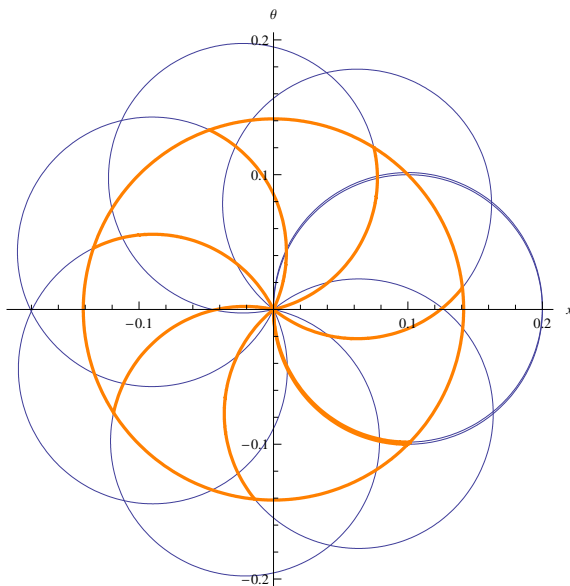


FIGURE 30. La structure des géodésiques au premier ordre pour k petit. L'exemple de $k = 0.1$. La période de t est 0.2π et $\varphi \in [0, 4kK(k) \sim 0.6299]$. Le bouquet de cercles $C_{k=0.1, \varphi}$ pour les 7 valeurs φ allant de 0 à 0.6299 par pas égaux est représenté en traits noirs normaux. En traits épais oranges sont représentés les arcs des $C_{k=0.1, \varphi}$ pour $t \in [0, 0.05\pi]$ ainsi que le cercle des extrémités Σ_r avec $r = (2k^2 (1 - \cos(\frac{t}{k})))^{\frac{1}{2}} \sim 0.1414$.

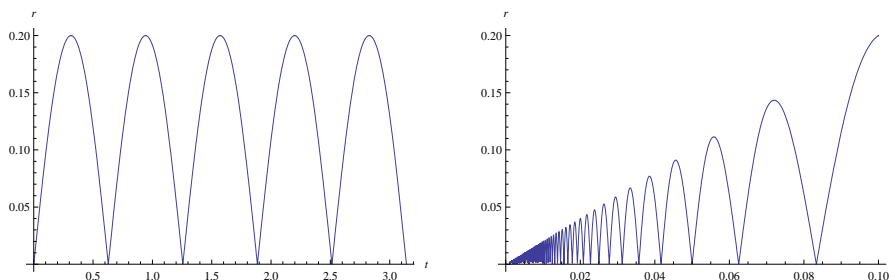


FIGURE 31. Le rayon r du cercle Σ_r des extrémités des géodésiques $\Gamma_{1,t}$ au premier ordre. À gauche : lorsque t varie à k constant et que les $\Gamma_{1,t}$ s'enroulent sur les cercles $C_{k, \varphi}$, r varie périodiquement entre 0 et $r_k = k\sqrt{2}$. À droite : lorsque c'est t qui est constant et que c'est k qui varie en décroissant vers 0, r varie également de façon périodique mais en étant modulé par la droite $r_k = k\sqrt{2}$. Il oscille donc de plus en plus vite lorsque $k \rightarrow 0^+$.

un difféomorphisme simple Δ entre $\mathbb{R}^2 \times \mathbb{S}^1$ et l'intérieur du tore $\mathring{\mathbb{D}}_\Lambda \times \mathbb{S}_{2\Lambda}^1 = \mathbb{T}_\Lambda$ où $\mathbb{S}_{2\Lambda}^1$ est le cercle de rayon 2Λ (l'“âme” du tore, en quelque sorte son “cercle déférent”) et où $\mathring{\mathbb{D}}_\Lambda$ est l'intérieur du disque de \mathbb{R}^2 de centre 0 et de rayon Λ . Il suffit

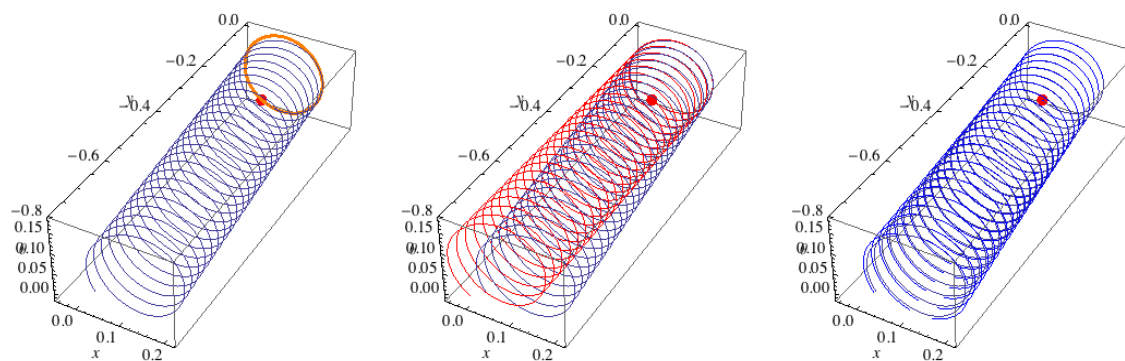


FIGURE 32. Les approximations d'ordre $n = 1, 2, 3$ $\Gamma_{n,t}$ d'une géodésique Γ_t pour k petit ($k = 0.1$), $\varphi = 0.1$ (et donc $\frac{\varphi}{k} = 1$) et t variant de 0 à 5π . La géodésique Γ_t est l'hélice en traits noirs commune aux trois figures et le point rouge est l'origine $\{0, 0, 0\}$ de Γ_t . À gauche, en orange, l'approximation au premier ordre qu'est le cercle $\Gamma_{1,t} = C_{k=0.1, \varphi=0.1}$ ($y_t = 0$). Au milieu, en rouge, l'approximation à l'ordre deux $\Gamma_{2,t}$, qui déploie $\Gamma_{1,t} = C_{k=0.1, \varphi=0.1}$ en hélice le long de l'axe des y . À droite, en bleu, l'approximation à l'ordre trois $\Gamma_{3,t}$ qui désaxe l'hélice $\Gamma_{2,t}$ par rapport à l'axe des y . $\Gamma_{3,t}$ est déjà une assez bonne approximation de Γ_t .

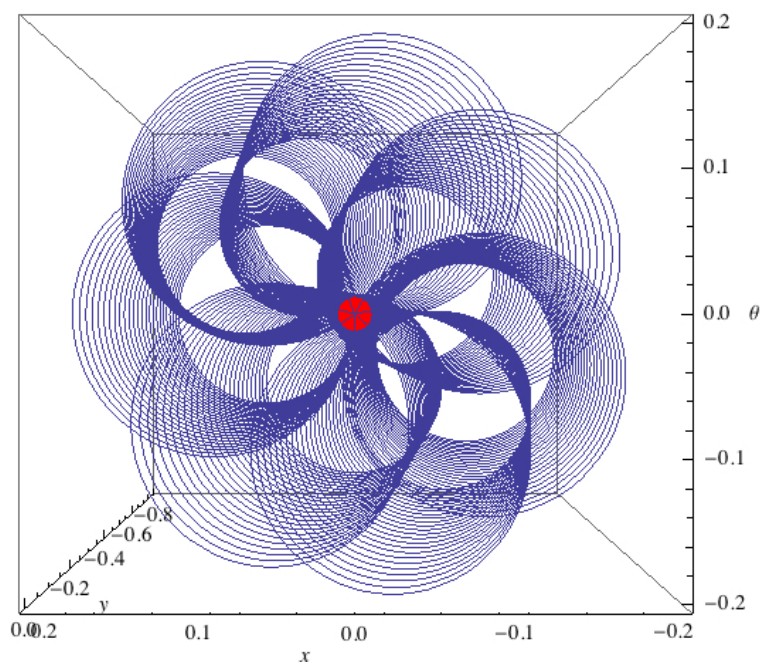


FIGURE 33. Les géodésiques $Q_{C_2,k}$ pour sept valeurs équiréparties de $\varphi \in [0, kK(k)]$. Elles forment un bouquet d'hélices désaxées par rapport à l'axe des y . Le point de vue est l'origine dans la direction des $y < 0$.

de prendre une fonction C^∞ bien régulièrement croissante $d : [0, \Lambda[\rightarrow [0, \infty[$ telle que $d(0) = 0$ et $d(\lambda) \xrightarrow{\lambda \rightarrow \Lambda^-} \infty$ et de poser

$$\Delta(\lambda \cos \psi, \lambda \sin \psi) = (d(\lambda) \cos \psi, d(\lambda) \sin \psi)$$

pour $\lambda \in [0, \Lambda[$ et $\psi \in [0, 2\pi]$. Si l'on veut représenter une partie de $SE(2)$ uniformément borné par B dans les fibres $\mathbb{R}_{(\theta)}^2$, il suffit même de prendre simplement $\Lambda > B$ et $\Delta = Id$. Dans ce cas, si l'on considère le tore $\mathring{\mathbb{D}}_\Lambda \times \mathbb{S}^1 = \mathbb{T}_\Lambda$ dans \mathbb{R}^3 de coordonnées $\{X_{tor}, Y_{tor}, Z_{tor}\}$, on représente $\{x, y, \theta\}$ par

$$X_{tor} = (2\Lambda + x) \cos(\theta), Y_{tor} = (2\Lambda + x) \sin(\theta), Z_{tor} = y.$$

Considérons par exemple le $Q_{C_1^0, k=0.1}$ de la figure 19. On peut prendre $\Lambda = 1$ et, pour $t \in [0, 5\pi]$, θ fait à peu près deux tours et demi. Pour le $Q_{C_1^0, k=0.9}$ de la figure 20, on peut prendre $\Lambda = 10$ et, pour $t \in [0, 5\pi]$, θ fait un peu moins de deux tours. Enfin, pour le $Q_{C_1^0, k=0.999}$ de la figure 22, on peut prendre $\Lambda = 15$ et, toujours pour $t \in [0, 5\pi]$, θ fait à peu près un tour. La figure 34 présente ces trois cas dans des tores de rayon adapté. La figure 35 les représente dans un tore unique de rayon $\Lambda = 15$ de façon à illustrer la façon dont ils s'emboîtent en gigogne. Plus k augmente plus le "cône" s'évase et moins il fait de tours. On notera la morphologie "en fourche" du bord $t = 5\pi$ qui devient de plus en plus prononcée au fur et à mesure que $k \rightarrow 1^-$. Elle est due aux rapides variations de θ_t bien visibles dans la figure 22 et qui, à la limite $k = 1$ (strate C_3), donnent les sauts en falaise des figures 25 et 26. La figure 36 illustre l'origine de cette morphologie. Elle représente 10 lignes de niveau de θ du $Q_{C_1^0, k=0.9}$ de la figure 20. On voit que pour $\theta = -10$ la ligne de niveau est encore un cycle. Mais le bord $\partial Q_{C_1^0, k=0.9}$ pour $t = 5\pi$ étant ondulé, lorsque θ continue à croître, ce cycle se brise en deux arcs qui se contractent progressivement en deux points correspondant aux minima de θ (aux maxima de $|\theta|$).

Considérons maintenant la représentation torique complète de la surface géodésique Q_{C_3} image de la strate limite C_3 bordant les strates ouvertes C_1 et C_2 . La figure 37 montre deux perspectives du début de Q_{C_3} , $t \in [0, \frac{\pi}{2}]$, dans le tore adapté $\mathbb{T}_{\Lambda=2}$. On y voit les quatre feuilletts correspondant aux quatre composantes de C_3 et la façon dont ils s'organisent en deux nappes qui se croisent conformément à la figure 25. La figure 38 montre l'évolution de Q_{C_3} pour $t \in [0, \pi]$ dans le tore adapté $\mathbb{T}_{\Lambda=4}$, puis $t \in [0, 2\pi]$ dans le tore adapté $\mathbb{T}_{\Lambda=8}$ et enfin $t \in [0, 5\pi]$ dans le tore adapté $\mathbb{T}_{\Lambda=20}$ (avec deux perspectives). Quant à la figure 39, elle montre $Q_{C_3, t=5\pi}$ positionné dans le tore adapté $\mathbb{T}_{\Lambda=20}$ avec, en rouge, le cercle de rayon 40 qui est l'âme du tore (son "cercle déferent"). Pour t assez grand $Q_{C_3, t}$ s'approche exponentiellement d'un tour complet.

Nous rencontrerons à nouveau ces représentations toriques à la figure 43 de la prochaine section 4.5.

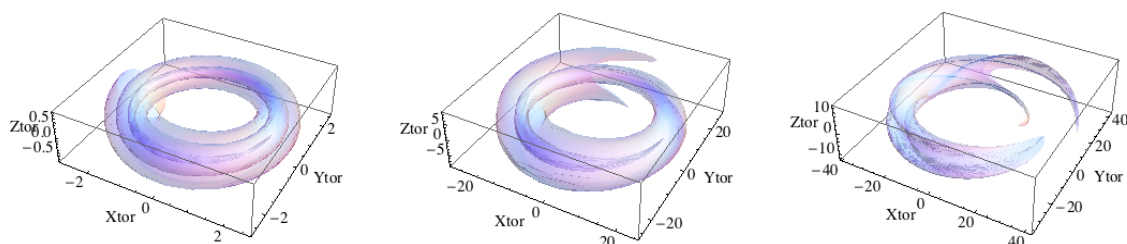


FIGURE 34. Représentations toriques des surfaces géodésiques $Q_{C_1^0, k}$ pour $t \in [0, 5\pi]$ dans des tores \mathbb{T}_Λ de rayon approprié. À gauche, $Q_{C_1^0, k=0.1}$ dans $\mathbb{T}_{\Lambda=1}$: θ fait à peu près deux tours et demi. Au milieu, $Q_{C_1^0, k=0.9}$ dans $\mathbb{T}_{\Lambda=10}$: θ fait un peu moins de deux tours. À droite, $Q_{C_1^0, k=0.999}$ dans $\mathbb{T}_{\Lambda=15}$: θ fait à peu près un tour. On note la morphologie “en fourche” du bord $t = 5\pi$ qui devient de plus en plus prononcée au fur et à mesure que $k \rightarrow 1^-$.

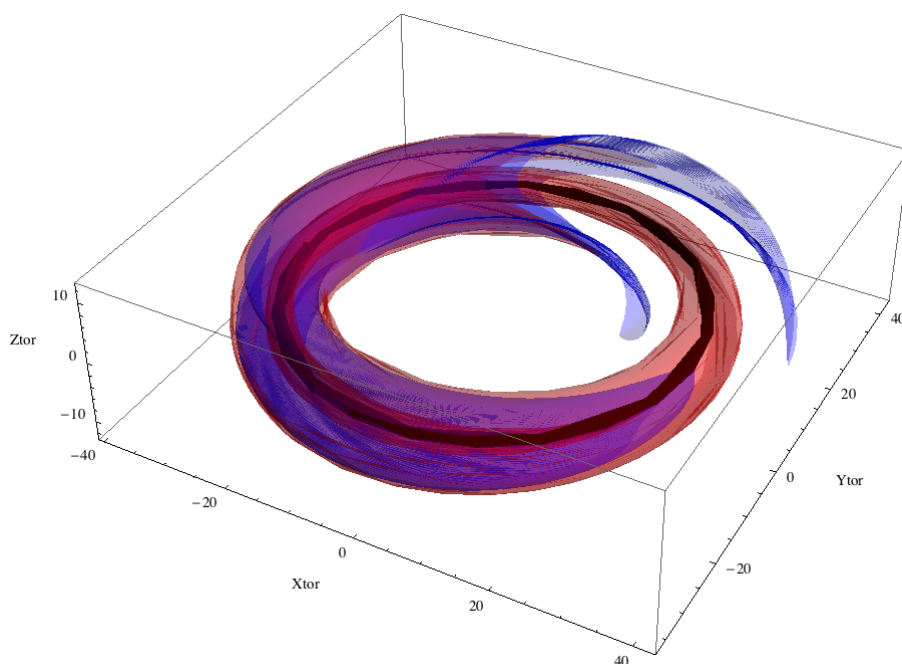


FIGURE 35. Les représentations toriques de $Q_{C_1^0, k=0.1}$ (gris), $Q_{C_1^0, k=0.9}$ (rouge) et $Q_{C_1^0, k=0.999}$ (bleu) dans le même tore $\mathbb{T}_{\Lambda=15}$. La transparence permet d’entrevoir comment ces surfaces géodésiques s’emboîtent en gigogne.

Pour Q_{C_2} les représentations toriques ne sont pas très lisibles. Nous reviendrons sur la géométrie (compliquée) de Q_{C_2} dans la section 4.7 avec l’analyse des sphères et des fronts d’ondes sous-riemanniens.

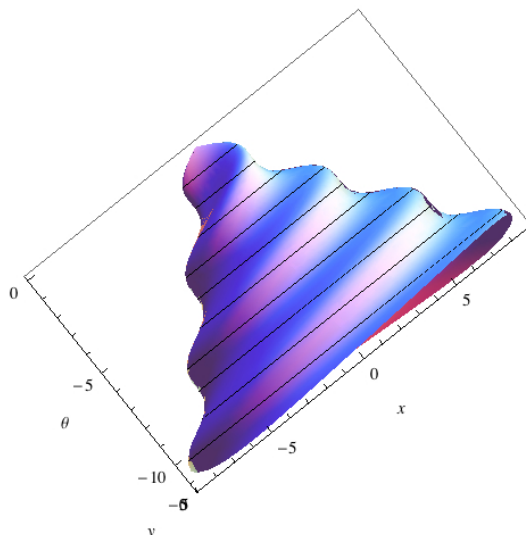


FIGURE 36. Origine de la morphologie “en fourche” du bord des représentations toriques des surfaces géodésiques $Q_{C_1^0, k}$ de la figure 34. On représente 10 lignes de niveau de θ du $Q_{C_1^0, k=0.9}$ de la figure 20. On voit que pour $\theta = -10$ la ligne de niveau est encore un cycle. Mais le bord $\partial Q_{C_1^0, k=0.9}$ pour $t = 5\pi$ étant ondulé, lorsque θ continue à croître, ce cycle se brise en deux arcs qui se contractent progressivement en deux points correspondant aux minima de θ (aux maxima de $|\theta|$).

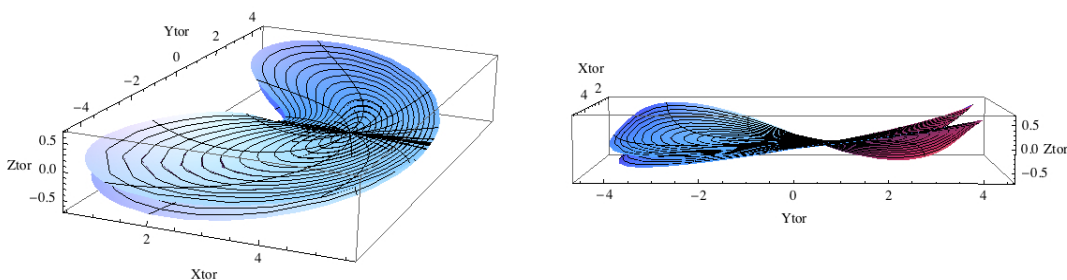


FIGURE 37. Deux perspectives du début de Q_{C_3} , $t \in [0, \frac{\pi}{2}]$, dans le tore adapté $\mathbb{T}_{\Lambda=2}$. Les quatre feuillettes correspondant aux quatre composantes de C_3 s’organisent en deux nappes qui se croisent (cf. figure 25).

4.4.6. *Un feuilletage géodésique compliqué.* La structure de base au premier ordre avec les bouquets de cercles $C_{k, \varphi}$ et les cercles Σ_r de rayon r oscillant va ainsi se déployer le long de l’axe des y mais, à cause des désaxages, manifester des oscillations avec des contractions et des dilatations complexes (un peu comme des “noeuds” et des “ventres” de paquets d’ondes). Les surfaces géodésiques $Q_{\tilde{\lambda}_S}$ “feuillettent”

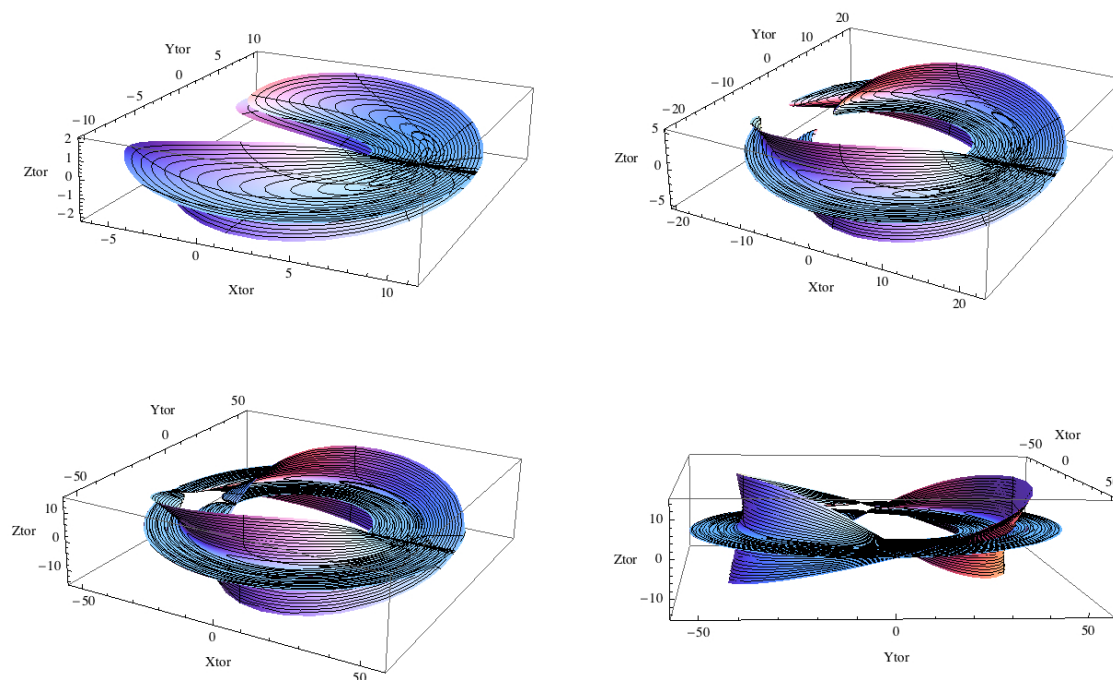


FIGURE 38. L'évolution des surfaces géodésiques Q_{C_3} . En haut à gauche, $t \in [0, \pi]$ dans le tore adapté $\mathbb{T}_{\Lambda=4}$. En haut à droite, $t \in [0, 2\pi]$ dans le tore adapté $\mathbb{T}_{\Lambda=8}$. En bas deux perspectives pour $t \in [0, 5\pi]$ dans le tore adapté $\mathbb{T}_{\Lambda=20}$.

$SE(2)$ mais de façon irrégulière. Les feuilles peuvent être dégénérées, posséder des singularités, des enveloppes, et s'intersecter entre elles. On obtient ainsi une géométrie très compliquée des feuilles géodésiques du groupe très élémentaire $SE(2)$. Comme nous allons le voir, on peut l'analyser en détail.

4.5. Exponentielle, cut locus, points conjugués

Les délicats calculs donnant le cut locus, les points conjugués, la sphère et le front d'onde des géodésiques de $SE(2)$ ainsi que quelques graphiques associés (nous allons en présenter d'autres) se trouvent d'abord dans *Geometry of strokes* [372] d'Igor Moiseev (2006) puis dans [373] déjà cité (2008) et dans les deux papiers [478] et [479] (2009) de Yuri Sachkov. Pour étudier les points où une géodésique $q(s)$ entre en compétition au temps-longueur t avec une autre géodésique $\tilde{q}(s)$ différente, autrement dit, telle que $\tilde{q}(s) \neq q(s)$ pour $s < t$ mais $\tilde{q}(t) = q(t)$, les auteurs ont eu l'idée d'utiliser les *symétries* du *pendule* qui contrôle les géodésiques. En effet, ces symétries transforment des géodésiques en géodésiques et cela permet de définir assez facilement des conditions pour que deux géodésiques symétriques puissent

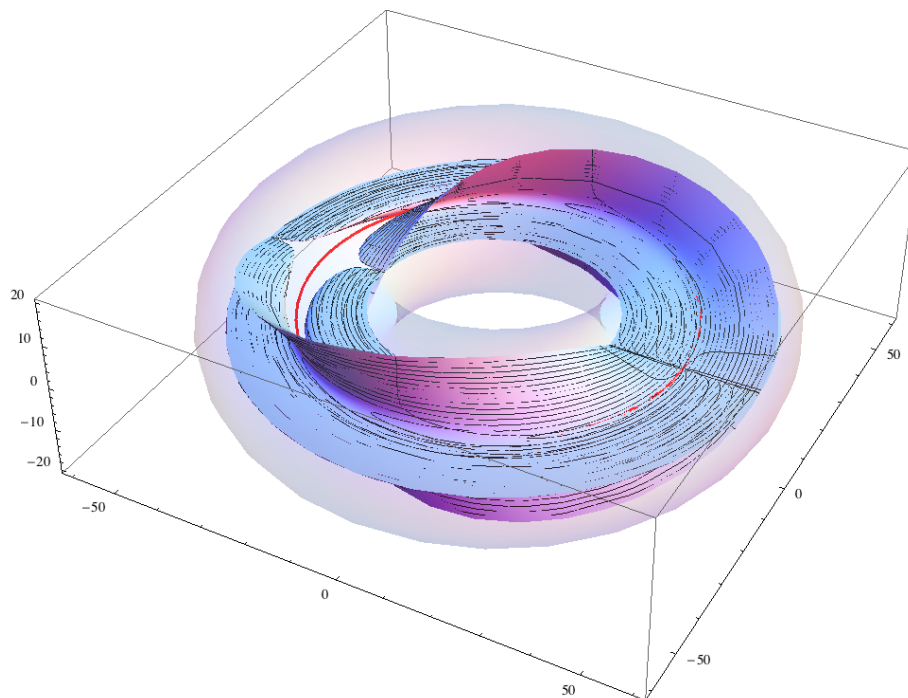


FIGURE 39. La surface géodésique $Q_{C_3, t=5\pi}$ positionnée dans le tore $\mathbb{T}_{\Lambda=20}$ avec, en rouge, le cercle de rayon 40 qui est l'âme du tore (son "cercle déférent"). Pour t assez grand $Q_{C_3, t}$ s'approche exponentiellement d'un tour complet.

entrer en compétition. Ces contraintes, appelées par les auteurs les *strates* et *points de Maxwell*, contraignent très fortement le cut locus et les points conjugués. Les étapes de l'analyse sont les suivantes.

1. les symétries du portrait de phase du pendule ;
2. l'action de ces symétries sur les géodésiques ;
3. les symétries de l'application exponentielle ;
4. les conditions nécessaires imposées par les symétries sur les points et les strates de Maxwell ;
5. les points et les strates de Maxwell effectifs ;
6. les points conjugués et le cut locus.

Les symétries du portrait de phase du pendule sont évidentes. Nous les avons rencontrées dans la section 10 du chapitre 13 à propos des *elasticæ*. Il faut tenir compte du fait que γ^S est ici de période 4π . Il y a donc, en plus de l'identité ε^0 ,

$$\begin{aligned} \varepsilon^1 &: (\gamma^S, c^S) \mapsto (\gamma^S, -c^S), \\ \varepsilon^2 &: (\gamma^S, c^S) \mapsto (-\gamma^S, c^S), \end{aligned}$$

$$\begin{aligned}
\varepsilon^3 &= \varepsilon^1 \circ \varepsilon^2 = \varepsilon^2 \circ \varepsilon^1 : (\gamma^S, c^S) \mapsto (-\gamma^S, -c^S), \\
\varepsilon^4 &: (\gamma^S, c^S) \mapsto (\gamma^S + 2\pi, c^S), \\
\varepsilon^5 &= \varepsilon^1 \circ \varepsilon^4 = \varepsilon^4 \circ \varepsilon^1 : (\gamma^S, c^S) \mapsto (\gamma^S + 2\pi, -c^S), \\
\varepsilon^6 &= \varepsilon^2 \circ \varepsilon^4 = \varepsilon^4 \circ \varepsilon^2 : (\gamma^S, c^S) \mapsto (-\gamma^S + 2\pi, c^S), \\
\varepsilon^7 &= \varepsilon^3 \circ \varepsilon^4 = \varepsilon^4 \circ \varepsilon^3 : (\gamma^S, c^S) \mapsto (-\gamma^S + 2\pi, -c^S).
\end{aligned}$$

Ces symétries ε^i forment le groupe à 8 éléments $(\mathbb{Z}/2\mathbb{Z})^3$. Elles transforment de façon évidente les trajectoires du pendule. Si l'on note $(\gamma^S, c^S)_u = (\gamma_u^S, c_u^S)$, on a, pour $s \in [t_0, t]$,

$$\begin{aligned}
\varepsilon^i((\gamma^S, c^S)_s) &= (\varepsilon^i(\gamma^S, c^S))_{t-s} \text{ si } i = 1, 2, 5, 6 \\
\varepsilon^j((\gamma^S, c^S)_s) &= (\varepsilon^i(\gamma^S, c^S))_s \text{ si } j = 3, 4, 7.
\end{aligned}$$

La figure 40 illustre de façon schématique la façon dont $(\gamma^S, c^S)_0$ est transformé par ε^i en $(\gamma^{S,i}, c^{S,i})_0$:

$$\begin{aligned}
(\gamma^{S,1}, c^{S,1})_0 &= (\gamma^S, -c^S)_t, (\gamma^{S,1}, c^{S,1})_t = (\gamma^S, -c^S)_0, \\
(\gamma^{S,2}, c^{S,2})_0 &= (-\gamma^S, c^S)_t, (\gamma^{S,2}, c^{S,2})_t = (-\gamma^S, c^S)_0, \\
(\gamma^{S,3}, c^{S,3})_0 &= (-\gamma^S, -c^S)_0, (\gamma^{S,3}, c^{S,3})_t = (-\gamma^S, -c^S)_t, \\
(\gamma^{S,4}, c^{S,4})_0 &= (\gamma^S + 2\pi, c^S)_0, (\gamma^{S,4}, c^{S,4})_t = (\gamma^S + 2\pi, c^S)_t, \\
(\gamma^{S,5}, c^{S,5})_0 &= (\gamma^S + 2\pi, -c^S)_t, (\gamma^{S,5}, c^{S,5})_t = (\gamma^S + 2\pi, -c^S)_0, \\
(\gamma^{S,6}, c^{S,6})_0 &= (-\gamma^S + 2\pi, c^S)_t, (\gamma^{S,6}, c^{S,6})_t = (-\gamma^S + 2\pi, c^S)_0, \\
(\gamma^{S,7}, c^{S,7})_0 &= (-\gamma^S + 2\pi, -c^S)_0, (\gamma^{S,7}, c^{S,7})_t = (-\gamma^S + 2\pi, -c^S)_t.
\end{aligned}$$

Les symétries transforment aussi naturellement les géodésiques de la façon suivante : $\varepsilon^i(x_s, y_s, \theta_s) = (x_s^i, y_s^i, \theta_s^i)$ avec

$$\begin{aligned}
(x_s^1, y_s^1, \theta_s^1) &= (\cos(\theta_t)(x_t - x_{t-s}) + \sin(\theta_t)(y_t - y_{t-s}), \\
&\quad \sin(\theta_t)(x_t - x_{t-s}) - \cos(\theta_t)(y_t - y_{t-s}), \theta_t - \theta_{t-s}) \\
(x_s^2, y_s^2, \theta_s^2) &= (-\cos(\theta_t)(x_t - x_{t-s}) - \sin(\theta_t)(y_t - y_{t-s}), \\
&\quad -\sin(\theta_t)(x_t - x_{t-s}) + \cos(\theta_t)(y_t - y_{t-s}), \theta_t - \theta_{t-s}) \\
(x_s^3, y_s^3, \theta_s^3) &= (-x_s, -y_s, \theta_s), (x_s^4, y_s^4, \theta_s^4) = (-x_s, y_s, -\theta_s) \\
(x_s^5, y_s^5, \theta_s^5) &= (\cos(\theta_t)(x_{t-s} - x_t) + \sin(\theta_t)(y_{t-s} - y_t), \\
&\quad -\sin(\theta_t)(x_{t-s} - x_t) + \cos(\theta_t)(y_{t-s} - y_t), \theta_{t-s} - \theta_t) \\
(x_s^6, y_s^6, \theta_s^6) &= (\cos(\theta_t)(x_t - x_{t-s}) + \sin(\theta_t)(y_t - y_{t-s}), \\
&\quad -\sin(\theta_t)(x_t - x_{t-s}) + \cos(\theta_t)(y_t - y_{t-s}), \theta_{t-s} - \theta_t) \\
(x_s^7, y_s^7, \theta_s^7) &= (x_s, -y_s, -\theta_s)
\end{aligned}$$

(en différentiant x_s^i et y_s^i par rapport à s à t constant, on vérifie trivialement que $\frac{y_s^i}{x_s^i}$ est bien égal à $\tan(\theta_s^i)$).

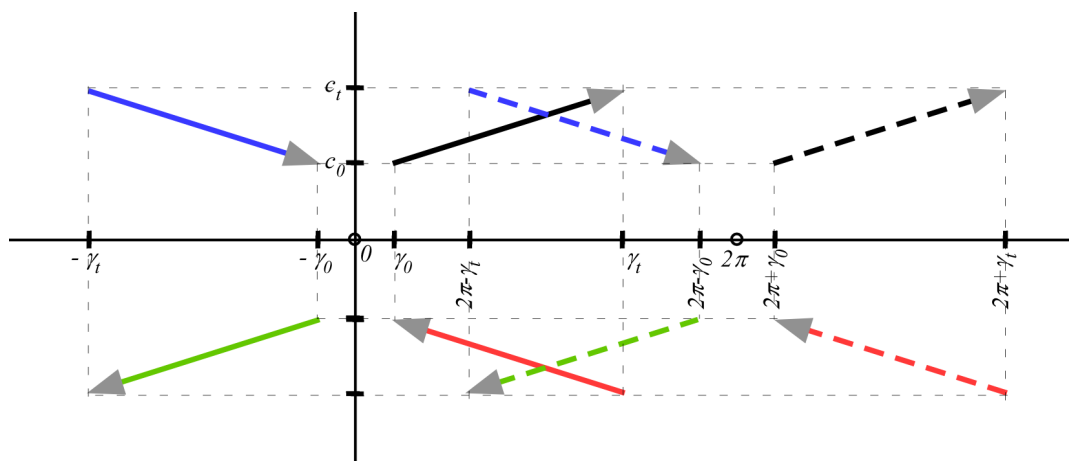


FIGURE 40. Schéma des transformations $(\gamma^{S,i}, c^{S,i})_0 \rightarrow (\gamma^{S,i}, c^{S,i})_t$ d'un arc $(\gamma^S, c^S)_0 \rightarrow (\gamma^S, c^S)_t$ du pendule par les symétries ε^i . Les flèches ne sont pas les trajectoires mais les vecteurs "origine ($s = 0$) \rightarrow extrémité ($s = t$)". $i = 0$ (arc initial) : noir continu, $i = 1$: rouge continu, $i = 2$: bleu continu, $i = 3$: vert continu, $i = 4$: noir pointillé, $i = 5$: rouge pointillé, $i = 6$: bleu pointillé, $i = 7$: vert pointillé.

Si l'on note $(\gamma^{S,i}, c^{S,i})$ l'état initial ($t = 0$) des transformées de la géodésique de référence d'origine (γ^S, c^S) et d'extrémité (γ_t^S, c_t^S) , on a par conséquent

$$\begin{aligned} (\gamma^{S,1}, c^{S,1}) &= (\gamma_t^S, -c_t^S), \\ (\gamma^{S,2}, c^{S,2}) &= (-\gamma_t^S, c_t^S), \\ (\gamma^{S,3}, c^{S,3}) &= (-\gamma^S, -c^S), \\ (\gamma^{S,4}, c^{S,4}) &= (\gamma^S + 2\pi, c^S), \\ (\gamma^{S,5}, c^{S,5}) &= (\gamma_t^S + 2\pi, -c_t^S), \\ (\gamma^{S,6}, c^{S,6}) &= (-\gamma_t^S + 2\pi, c_t^S), \\ (\gamma^{S,7}, c^{S,7}) &= (-\gamma^S + 2\pi, -c^S). \end{aligned}$$

Ces symétries sont géométriquement très simples dans le plan (x, y) . Si l'on tourne $\varepsilon^i(q)$ de façon à faire coïncider (x_t^i, y_t^i) avec (x_t, y_t) et si ℓ est le segment reliant $(x_0, y_0) = (0, 0)$ à (x_t, y_t) dans le plan (x, y) , et ℓ^\perp la droite orthogonale au milieu m_ℓ de ℓ , alors $\varepsilon^5, \varepsilon^6$ deviennent après rotation la réflexion par rapport à m_ℓ , $\varepsilon^1, \varepsilon^2$ deviennent la réflexion par rapport à ℓ^\perp et $\varepsilon^4, \varepsilon^7$ deviennent la réflexion par rapport à ℓ , ε^3 devenant quant à elle l'identité après une rotation de π . Notons que $\varepsilon^5, \varepsilon^6, \varepsilon^7$ s'obtiennent en appliquant ε^4 (c'est-à-dire la symétrie par rapport à l'axe des y orthogonal au plan de contact à l'origine) à $\varepsilon^1, \varepsilon^2, \varepsilon^3$. En faisant $s = t$ dans les formules et en tenant compte du fait que $(x_0, y_0, \theta_0) = (0, 0, 0)$ on obtient l'action des symétries ε^i sur l'extrémité au temps-longueur t des géodésiques issues de l'origine.

Pour illustrer ces symétries, la figure 41 les représente pour un arc de courbe simple. Il s'agit de l'arc A^0 de la cubique $x(s) = s, y(s) = s^3$ pour $s \in [0, t = 1.5]$.

A^0 est tracé en noir et en trait épais. Les différentes symétries sont codées par des couleurs et des pointillés : $A^1 = \varepsilon^1(A^0)$ en rouge continu, $A^2 = \varepsilon^2(A^0)$ en bleu continu, $A^3 = \varepsilon^3(A^0)$ en vert continu, $A^4 = \varepsilon^4(A^0)$ en noir pointillé, $A^5 = \varepsilon^5(A^0)$ en rouge pointillé, $A^6 = \varepsilon^6(A^0)$ en bleu pointillé, $A^7 = \varepsilon^7(A^0)$ en vert pointillé. Pour mieux visualiser les symétries, nous avons également représenté par un point épais la position

(i) au temps $s = 0.5$ pour A^0, A^3, A^4, A^7 qui démarrent en $s = 0$ et sont paramétrées par s et

(ii) au temps $s = t - 0.5 = 1$ pour A^1, A^2, A^5, A^6 qui démarrent en $s = t$ et sont paramétrées par $t - s$.

Nous avons aussi représenté en trait fin la rotation de l'arc initial A^0 amenant son extrémité pour $t = 0.5$ sur celle de $A^i = \varepsilon^i(A^0)$. Ces rotations sont faciles à calculer. Posons

$$x(t) = X = 1.5, \quad y(t) = Y = 1.5^3 = 3.375, \quad \theta = \text{Arctan}(\dot{y}(t)) = 1.424.$$

Soit ϕ^i l'angle de rotation en radians. On obtient :

$\cos(\phi^1) = \frac{(X^2 - Y^2) \cos(\theta) + 2XY \sin(\theta)}{X^2 + Y^2}$ $= 0.636$ $\phi^1 = -0.88143 \text{ rad} = -50.5^\circ$	$\sin(\phi^1) = \frac{(X^2 - Y^2) \sin(\theta) - 2XY \cos(\theta)}{X^2 + Y^2}$ $= -0.772$
$\cos(\phi^2) = \frac{-(X^2 - Y^2) \cos(\theta) - 2XY \sin(\theta)}{X^2 + Y^2}$ $= -0.636$ $\phi^2 = 2.26 \text{ rad} = 129.5^\circ$	$\sin(\phi^2) = \frac{-(X^2 - Y^2) \sin(\theta) + 2XY \cos(\theta)}{X^2 + Y^2}$ $= 0.772$
$\cos(\phi^3) = -1$ $\phi^3 = \pi \text{ radian} = 180^\circ$	$\sin(\phi^3) = 0$
$\cos(\phi^4) = \frac{-X^2 + Y^2}{X^2 + Y^2} = 0.67$ $\phi^4 = 0.8364 \text{ rad} = 47.925^\circ$	$\sin(\phi^4) = \frac{2XY}{X^2 + Y^2} = 0.742$
$\cos(\phi^5) = -\cos(\theta) = -0.1465$ $\phi^5 = 1.7179 \text{ rad} = 98.427^\circ$	$\sin(\phi^5) = \sin(\theta) = 0.9892$
$\cos(\phi^6) = \cos(\theta) = 0.1465$ $\phi^6 = -1.4237 \text{ rad} = -81.573^\circ$	$\sin(\phi^6) = -\sin(\theta) = -0.9892$
$\cos(\phi^7) = \frac{X^2 - Y^2}{X^2 + Y^2} = -0.67$ $\phi^7 = -2.3 \text{ rad} = -132.075^\circ$	$\sin(\phi^7) = \frac{-2XY}{X^2 + Y^2} = -0.742$

On voit très bien sur la figure les trois symétries après rotation : $\varepsilon^5, \varepsilon^6 \rightsquigarrow$ réflexion par rapport à m_ℓ , $\varepsilon^1, \varepsilon^2 \rightsquigarrow$ réflexion par rapport à ℓ^\perp , $\varepsilon^4, \varepsilon^7 \rightsquigarrow$ réflexion par rapport à ℓ , $\varepsilon^3 \rightsquigarrow$ identité.

À partir de ce système bien structuré de symétries, Sachkov et Moiseev vont alors calculer les points et les strates de Maxwell qu'elles peuvent engendrer. Si un moment initial $\lambda^{S,i} = (\gamma^{S,i}, c^{S,i})$ est différent de $\lambda^S = (\gamma^S, c^S)$, les deux géodésiques

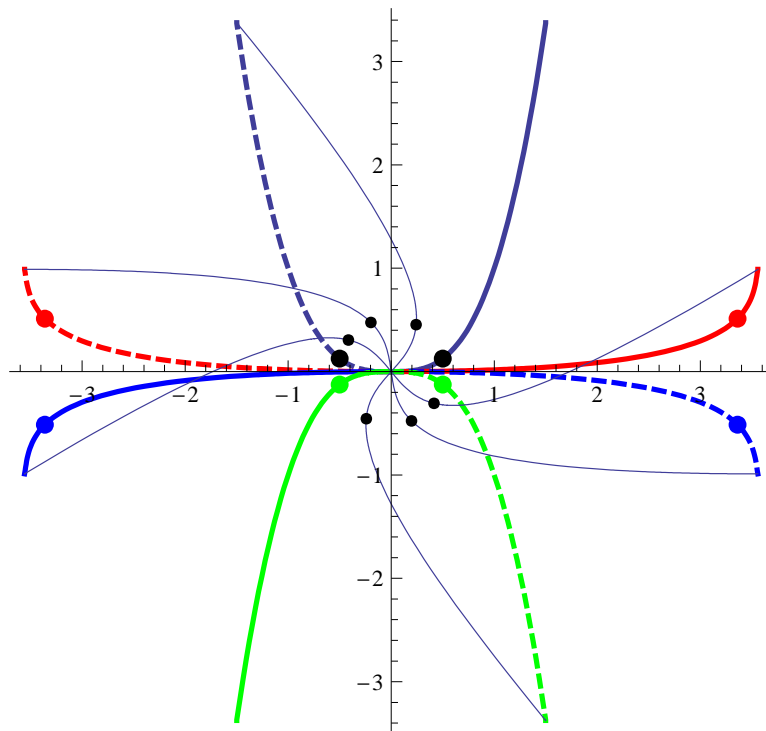


FIGURE 41. Les symétries ε^i pour l'arc simple A^0 de la cubique $x(s) = s$, $y(s) = s^3$ pour $s \in [0, t = 1.5]$. A^0 : noir continu, $A^1 = \varepsilon^1(A^0)$: rouge continu, $A^2 = \varepsilon^2(A^0)$: bleu continu, $A^3 = \varepsilon^3(A^0)$: vert continu, $A^4 = \varepsilon^4(A^0)$: noir pointillé, $A^5 = \varepsilon^5(A^0)$: rouge pointillé, $A^6 = \varepsilon^6(A^0)$: bleu pointillé, $A^7 = \varepsilon^7(A^0)$: vert pointillé. Le point épais est la position au temps $s = 0.5$ pour $\varepsilon^0, \varepsilon^3, \varepsilon^4, \varepsilon^7$ et au temps $s = t - 0.5 = 1$ pour $\varepsilon^1, \varepsilon^2, \varepsilon^5, \varepsilon^6$. On a représenté en traits fins la rotation de A^0 amenant son extrémité sur celle de $A^i = \varepsilon^i(A^0)$ pour $t = 0.5$.

$q(s)$ et $q^i(s)$ sont initialement différentes et on peut calculer à quelle condition elles peuvent se rencontrer au temps t .

Les calculs deviennent vite assez techniques. Si $\lambda^S = (\gamma^S, c^S)$ est un état du pendule et $\nu^S = (\lambda^S, t)$ un état au temps-longueur t , les ν^S décrivent l'espace $N = \mathcal{C} \times \mathbb{R}_t^+$ où \mathcal{C} est le portrait de phase du pendule et la stratification C_i de \mathcal{C} induit évidemment une stratification N^i de N . Les symétries ε^i agissent sur N simplement par $\varepsilon^i(\nu^S) = \nu^{S,i} = (\varepsilon^i(\lambda^S), t)$. Autrement dit, on garde le même paramétrage temporel mais on démarre en $\varepsilon^i(\lambda^S)$ qui, comme nous l'avons vu plus

haut, est donné par

$$\begin{aligned}
 (\gamma^{S,1}, c^{S,1}) &= (\gamma_t^S, -c_t^S) , \\
 (\gamma^{S,2}, c^{S,2}) &= (-\gamma_t^S, c_t^S) , \\
 (\gamma^{S,3}, c^{S,3}) &= (-\gamma^S, -c^S) , \\
 (\gamma^{S,4}, c^{S,4}) &= (\gamma^S + 2\pi, c^S) , \\
 (\gamma^{S,5}, c^{S,5}) &= (\gamma_t^S + 2\pi, -c_t^S) , \\
 (\gamma^{S,6}, c^{S,6}) &= (-\gamma_t^S + 2\pi, c_t^S) , \\
 (\gamma^{S,7}, c^{S,7}) &= (-\gamma^S + 2\pi, -c^S) .
 \end{aligned}$$

L'application exponentielle $\text{Exp} : N \rightarrow M$ s'écrit

$$\text{Exp}(\nu^S) = \text{Exp}(\lambda^S, s) = \text{Exp}(\gamma^S, c^S, s) = \pi \circ e^{s\vec{H}}(\lambda^S) = \pi(\lambda_s^S) = q(s)$$

et on cherche les conditions sur t pour que

- (i) $\lambda^{S,i} \neq \lambda^S$, autrement dit le moment initial de $q^i(s)$ est différent de celui de $q(s)$, ce qui implique $\text{Exp}(\lambda^{S,i}, s) \neq \text{Exp}(\lambda^S, s)$ pour $s < t$,
- (ii) $\text{Exp}(\lambda^{S,i}, t) = \text{Exp}(\lambda^S, t)$.

Dans ce cas on dit que $q^i(t)$ est un *point de Maxwell* de la strate Max^i : au temps-longueur t , la géodésique $q(s)$ entre en compétition avec sa symétrique $q^i(s)$.

Mais nous avons calculé plus haut les transformations amenant $(x^i(t), y^i(t))$ sur $(x(t), y(t))$. Une première condition nécessaire pour que $q^i(t) = q(t)$ est que cette transformation soit l'identité, une seconde condition étant que $\theta^i(t) = \theta(t)$. Notons en coordonnées polaires

$$x(t) = \varrho \cos(\psi), y(t) = \varrho \sin(\psi) .$$

On obtient ainsi des relations entre l'angle ψ du rayon vecteur $q(0) \rightarrow q(t)$ et l'angle θ de la tangente à l'extrémité $q(t)$ de la géodésique.²⁷ Cela donne

- (i) **Symétrie** ε^1 . $2\psi - \theta = 0 \pmod{2\pi}$ ou $\varrho = 0$.
- (ii) **Symétrie** ε^2 . $2\psi - \theta = \pi \pmod{2\pi}$ ou $\varrho = 0$.
- (iii) **Symétrie** ε^3 . Impossible sauf si $\varrho = 0$, et alors θ peut être quelconque.
- (iv) **Symétrie** ε^4 . $2\psi = \pi \pmod{2\pi}$ (i.e. $x = 0$) et $\theta = 0 \pmod{\pi}$.
- (v) **Symétrie** ε^5 . $\theta = \pi \pmod{2\pi}$ ou $\varrho = 0$ et $\theta = 0$ (car il faut $\theta = -\theta$).
- (vi) **Symétrie** ε^6 . $\theta = 0 \pmod{2\pi}$.
- (vii) **Symétrie** ε^7 . $2\psi = 0 \pmod{2\pi}$ (i.e. $y = 0$) et $\theta = 0 \pmod{\pi}$.

La figure 42 montre les différentes relations possibles entre ψ et θ .

27. Donc dans ce qui suit ϱ, ψ, θ sont en fait des $\varrho_t, \psi_t, \theta_t$.

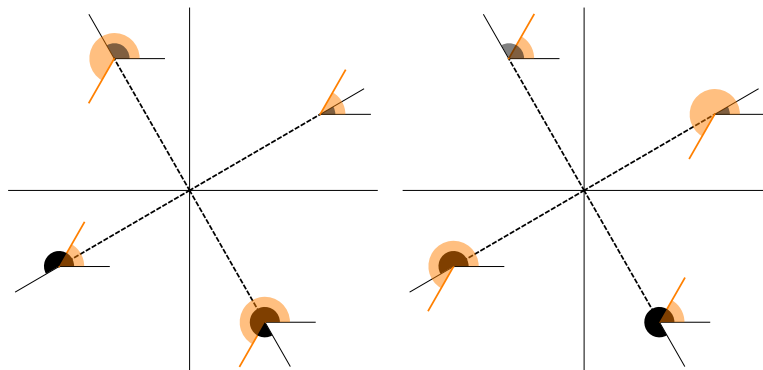


FIGURE 42. Les relations $2\psi - \theta = 0 \pmod{2\pi}$ (à droite) et $2\psi - \theta = \pi \pmod{2\pi}$ (à gauche). En pointillés épais sont représentés les rayons vecteurs $(\cos(\psi), \sin(\psi))$. Les secteurs en noir représentent ψ pour les valeurs $\frac{\pi}{6}, \frac{\pi}{6} + \frac{\pi}{2}, \frac{\pi}{6} + \pi, \frac{\pi}{6} + \frac{3\pi}{2}$. Les secteurs en orange représentent les valeurs de θ correspondantes.

Mais en plus des relations entre ψ et θ il faut vérifier que la condition

$$\lambda^{S,i} = (\gamma^{S,i}, c^{S,i}) \neq \lambda^S = (\gamma^S, c^S)$$

est vérifiée. Cela dépend des strates C_i considérées : on écrit la condition et on utilise les formules. On peut se restreindre aux strates C_1, C_2, C_3 car les strates ponctuelles C_4, C_5 correspondent à des équilibres.

On trouve ainsi pour les 7 symétries ε^i dans les différentes strates $C_j, j = 1, 2, 3$:

1. Strate C_1 .

(i) **Symétrie** ε^1 . $2\psi - \theta = 0 \pmod{2\pi}$ ou $\varrho = 0$. Dans C_1 , la contrainte est

$$(\gamma^{S,1}, c^{S,1}) = (\gamma_t^S, -c_t^S) \neq (\gamma^S, c^S)$$

et elle n'est pas satisfaite si $(\gamma_t^S, -c_t^S) = (\gamma^S, c^S)$. Comme en coordonnées de Jacobi $c^S = 2k \operatorname{cn}(\varphi, k)$ et $\sin\left(\frac{\gamma^S}{2}\right) = sg_1 k \operatorname{sn}(\varphi, k)$ et $\cos\left(\frac{\gamma^S}{2}\right) = sg_1 k \operatorname{dn}(\varphi, k)$,²⁸ il faut²⁹ $\operatorname{cn}(\varphi_t) = -\operatorname{cn}(\varphi)$ et $\operatorname{sn}(\varphi_t) = \operatorname{sn}(\varphi)$ (car sg_1 reste le même puisque $\cos\left(\frac{\gamma^S}{2}\right)$ reste le même puisque $\gamma_t^S = \gamma^S$) et $\operatorname{dn}(\varphi_t) = \operatorname{dn}(\varphi)$. Les formules d'addition des fonctions elliptiques de Jacobi³⁰ montrent alors que cette triple condition équivaut

28. Rappelons que sg_1 est le signe de $\cos\left(\frac{\gamma^S}{2}\right)$.

29. Le module k étant évident dans le contexte, nous ne le précisons plus.

30. Cf. la section 7 du chapitre 9. Rappelons que $2K$ est la demi-période de cn et sn et la période de dn .

à $\varphi_t = 2K - \varphi \pmod{4K}$)³¹ et donc à

$$\operatorname{cn}\left(\frac{\varphi + \varphi_t}{2}\right) = 0.$$

(ii) **Symétrie** ε^2 . $2\psi - \theta = \pi \pmod{2\pi}$ ou $\varrho = 0$. Dans C_1 , la contrainte est

$$(\gamma^{S,2}, c^{S,2}) = (-\gamma_t^S, c_t^S) \neq (\gamma^S, c^S)$$

et elle n'est pas satisfaite si $(-\gamma_t^S, c_t^S) = (\gamma^S, c^S)$. Il faut donc $\operatorname{cn}(\varphi_t) = \operatorname{cn}(\varphi)$, $\operatorname{sn}(\varphi_t) = -\operatorname{sn}(\varphi)$ (car sg_1 reste le même puisque $\cos\left(\frac{\gamma^S}{2}\right)$ reste le même si $\gamma_t^S = -\gamma^S$) et $\operatorname{dn}(\varphi_t, k) = \operatorname{dn}(\varphi, k)$. Cette triple condition équivaut à $\varphi_t = -\varphi \pmod{4K}$ et donc à

$$\operatorname{sn}\left(\frac{\varphi + \varphi_t}{2}\right) = 0.$$

(iii) **Symétrie** ε^3 . $\varrho = 0$, θ quelconque. La contrainte est

$$(\gamma^{S,3}, c^{S,3}) = (-\gamma^S, -c^S) \neq (\gamma^S, c^S)$$

et elle n'est pas satisfaite pour $(\gamma^S, c^S) = (0, 0)$.

(iv) **Symétrie** ε^4 . $2\psi = \pi \pmod{2\pi}$ (i.e. $x = 0$) et $\theta = 0 \pmod{\pi}$ (car $\theta = \pi = -\theta \pmod{\pi}$). La contrainte est

$$(\gamma^{S,4}, c^{S,4}) = (\gamma^S + 2\pi, c^S) \neq (\gamma^S, c^S)$$

et elle est toujours réalisée.

(v) **Symétrie** ε^5 . $\theta = \pi \pmod{2\pi}$ (car $\theta = \pi = -\theta \pmod{\pi}$) ou $\varrho = 0$, $\theta = 0$ (car il faut $\theta = -\theta$). La contrainte est

$$(\gamma^{S,5}, c^{S,5}) = (\gamma_t^S + 2\pi, -c_t^S) \neq (\gamma^S, c^S).$$

Elle est toujours réalisée car la strate C_1 est composée de deux composantes connexes et les trajectoires sont incluses dans une seule de ces composantes. Comme $(\gamma_t^S + 2\pi, -c_t^S)$ et (γ^S, c^S) n'appartiennent pas à la même composante, on ne peut pas avoir $(\gamma_t^S + 2\pi, -c_t^S) = (\gamma^S, c^S)$.

(vi) **Symétrie** ε^6 . $\theta = 0 \pmod{2\pi}$. La contrainte est

$$(\gamma^{S,6}, c^{S,6}) = (-\gamma_t^S + 2\pi, c_t^S) \neq (\gamma^S, c^S).$$

31. En effet, $\operatorname{cn}(u + 2K) = -\operatorname{cn}(u)$, $\operatorname{sn}(u + 2K) = -\operatorname{sn}(u)$, $\operatorname{dn}(u + 2K) = \operatorname{dn}(u)$ et par ailleurs $\operatorname{cn}(u)$ et $\operatorname{dn}(u)$ sont paires et $\operatorname{sn}(u)$ est impaire. Donc $\operatorname{cn}(-u + 2K) = -\operatorname{cn}(-u) = -\operatorname{cn}(u)$, $\operatorname{sn}(-u + 2K) = -\operatorname{sn}(-u) = \operatorname{sn}(u)$, $\operatorname{dn}(-u + 2K) = \operatorname{dn}(-u) = \operatorname{dn}(u)$.

Elle est toujours réalisée car la strate C_1 est ouverte, ainsi que ses deux composantes connexes. Si $\gamma^S \in]-\pi, \pi[$, alors $-\gamma_t^S + 2\pi \in]\pi, 3\pi[$ appartient à l'autre composante, ce qui est impossible puisqu'une trajectoire ne quitte pas sa composante.

- (vii) **Symétrie** ε^7 . $2\psi = 0 \pmod{2\pi}$ (i.e. $y = 0$) et $\theta = 0 \pmod{\pi}$ (car $\theta = \pi = -\theta \pmod{\pi}$). La contrainte est

$$(\gamma^{S,7}, c^{S,7}) = (-\gamma^S + 2\pi, -c^S) \neq (\gamma^S, c^S) .$$

Elle est toujours réalisée pour la même raison que pour la symétrie ε^6 , d'autant plus que $-c^S = c^S$ implique $c^S = 0$ et que les points de $c^S = 0$ n'appartiennent pas à C_1 .

2. Strate C_2 .

- (i) **Symétrie** ε^1 . $2\psi - \theta = 0 \pmod{2\pi}$ ou $\varrho = 0$. Pour la strate C_2 , la contrainte est toujours satisfaite. En effet, $c_s^S \neq 0$ et le signe de c_s^S reste constant sur une trajectoire et on a donc toujours

$$(\gamma^{S,1}, c^{S,1}) = (\gamma_t^S, -c_t^S) \neq (\gamma^S, c^S) .$$

- (ii) **Symétrie** ε^2 . $2\psi - \theta = \pi \pmod{2\pi}$ ou $\varrho = 0$. On a $c^S = sg_2 \frac{2}{k} \operatorname{dn} \left(\frac{\varphi}{k} \right)$,³² $\sin \left(\frac{\gamma^S}{2} \right) = sg_2 \operatorname{sn} \left(\frac{\varphi}{k} \right)$ et $\cos \left(\frac{\gamma^S}{2} \right) = \operatorname{cn} \left(\frac{\varphi}{k} \right)$. La contrainte est

$$(\gamma^{S,2}, c^{S,2}) = (-\gamma_t^S, c_t^S) \neq (\gamma^S, c^S)$$

et elle n'est pas satisfaite si $(-\gamma_t^S, c_t^S) = (\gamma^S, c^S)$. Avec $sg_{2,t} = sg_2$ puisque $c_t^S = c^S$, il faut donc $\operatorname{cn} \left(\frac{\varphi_t}{k} \right) = \cos \left(\frac{\gamma_t^S}{2} \right) = \cos \left(\frac{-\gamma^S}{2} \right) = \cos \left(\frac{\gamma^S}{2} \right) = \operatorname{cn} \left(\frac{\varphi}{k} \right)$, $\operatorname{sn} \left(\frac{\varphi_t}{k} \right) = -\operatorname{sn} \left(\frac{\varphi}{k} \right)$ (puisque $\sin \left(\frac{\gamma_t^S}{2} \right) = \sin \left(\frac{-\gamma^S}{2} \right) = -\sin \left(\frac{\gamma^S}{2} \right)$) et $\operatorname{dn} \left(\frac{\varphi_t}{k} \right) = \operatorname{dn} \left(\frac{\varphi}{k} \right)$. Cela équivaut à $\frac{\varphi_t}{k} = -\frac{\varphi}{k} \pmod{4K}$, soit

$$\operatorname{sn} \left(\frac{\varphi + \varphi_t}{2k} \right) = 0 \pmod{4kK} .$$

- (iii) **Symétrie** ε^3 . $\varrho = 0$, θ quelconque. La contrainte est toujours satisfaite. Elle est en effet

$$(\gamma^{S,3}, c^{S,3}) = (-\gamma^S, -c^S) \neq (\gamma^S, c^S)$$

et elle n'est pas satisfaite que pour $(\gamma^S, c^S) = (0, 0)$ qui n'est pas dans C_2 .

32. sg_2 est le signe de c .

- (iv) **Symétrie** ε^4 . $2\psi = \pi \bmod(2\pi)$, i.e. $x = 0$ et $\theta = 0 \bmod(\pi)$ (car $\theta = \pi = -\theta \bmod(\pi)$). La contrainte est

$$(\gamma^{S,4}, c^{S,4}) = (\gamma^S + 2\pi, c^S) \neq (\gamma^S, c^S)$$

et elle est toujours réalisée.

- (v) **Symétrie** ε^5 . $\theta = \pi \bmod(2\pi)$ (car $\theta = \pi = -\theta \bmod(\pi)$) ou $\varrho = 0$, $\theta = 0$ (car il faut $\theta = -\theta$). La contrainte est

$$(\gamma^{S,5}, c^{S,5}) = (\gamma_t^S + 2\pi, -c_t^S) \neq (\gamma^S, c^S)$$

Elle est toujours réalisée car dans la strate C_2 c_s^S est $\neq 0$ et de signe constant et donc $-c_t^S = c^S$ est impossible.

- (vi) **Symétrie** ε^6 . $\theta = 0 \bmod(2\pi)$. La contrainte est

$$(\gamma^{S,6}, c^{S,6}) = (-\gamma_t^S + 2\pi, c_t^S) \neq (\gamma^S, c^S)$$

et elle n'est pas satisfaite si $(-\gamma_t^S + 2\pi, c_t^S) = (\gamma^S, c^S)$. Avec $sg_{2,t} = sg_2$ puisque $c_t^S = c^S$, il faut donc

$$\operatorname{cn}\left(\frac{\varphi_t}{k}\right) = \cos\left(\frac{\gamma_t^S}{2}\right) = \cos\left(\frac{\pi - \gamma^S}{2}\right) = -\cos\left(\frac{\gamma^S}{2}\right) = -\operatorname{cn}\left(\frac{\varphi}{k}\right)$$

$$\operatorname{sn}\left(\frac{\varphi_t}{k}\right) = \operatorname{sn}\left(\frac{\varphi}{k}\right) \text{ (puisque } \sin\left(\frac{\gamma_t^S}{2}\right) = \sin\left(\frac{\pi - \gamma^S}{2}\right) = \sin\left(\frac{\gamma^S}{2}\right)\text{)}$$

$$\operatorname{dn}\left(\frac{\varphi_t}{k}\right) = \operatorname{dn}\left(\frac{\varphi}{k}\right) .$$

Cela équivaut à $\frac{\varphi_t}{k} = 2K - \frac{\varphi}{k} \bmod(4K)$, soit

$$\operatorname{cn}\left(\frac{\varphi + \varphi_t}{2k}\right) = 0 \bmod(4kK) .$$

- (vii) **Symétrie** ε^7 . $2\psi = 0 \bmod(2\pi)$, i.e. $y = 0$ et $\theta = 0 \bmod(\pi)$ (car $\theta = \pi = -\theta \bmod(\pi)$). La contrainte est

$$(\gamma^{S,7}, c^{S,7}) = (-\gamma^S + 2\pi, -c^S) \neq (\gamma^S, c^S)$$

Elle est toujours réalisée car $-c^S = c^S$ implique $c^S = 0$ alors que les points de $c^S = 0$ n'appartiennent pas à C_2 .

3. Strate C_3 .

- (i) **Symétrie** ε^1 . $2\psi - \theta = 0 \bmod(2\pi)$ ou $\varrho = 0$. Pour la strate C_3 , il en va de même que pour la strate C_2 : $c_s^S \neq 0$ et le signe de c_s^S reste constant sur une trajectoire et on a donc toujours

$$(\gamma^{S,1}, c^{S,1}) = (\gamma_t^S, -c_t^S) \neq (\gamma^S, c^S) .$$

La contrainte est donc toujours satisfaite.

- (ii) **Symétrie** ε^2 . $2\psi - \theta = \pi \bmod(2\pi)$ ou $\varrho = 0$. Dans C_3 , on a $c^S = \frac{2sg_2}{\cosh(\varphi)}$, $\sin\left(\frac{\gamma^S}{2}\right) = sg_1sg_2 \tanh(\varphi)$ et $\cos\left(\frac{\gamma^S}{2}\right) = \frac{sg_1}{\cosh(\varphi)}$ avec $sg_{1,s}$ le signe de $\cos\left(\frac{\gamma_s^S}{2}\right)$ et $sg_{2,s}$ le signe de c_s^S . Comme les sg_i restent les mêmes, la contrainte est

$$(\gamma^{S,2}, c^{S,2}) = (-\gamma_t^S, c_t^S) \neq (\gamma^S, c^S)$$

et elle n'est pas satisfaite si $\tanh(\varphi_t) = -\tanh(\varphi)$ et $\cosh(\varphi_t) = \cosh(\varphi)$, soit $\varphi_t = -\varphi$.

- (iii) **Symétrie** ε^3 . $\varrho = 0$, θ quelconque. La contrainte est

$$(\gamma^{S,3}, c^{S,3}) = (-\gamma^S, -c^S) \neq (\gamma^S, c^S)$$

et elle n'est pas satisfaite pour $(\gamma^S, c^S) = (0, 0)$. Mais comme $(0, 0) \notin C_3$, elle est toujours satisfaite dans C_3 .

- (iv) **Symétrie** ε^4 . $2\psi = \pi \bmod(2\pi)$, i.e. $x = 0$ et $\theta = 0 \bmod(\pi)$ (car $\theta = \pi = -\theta \bmod(\pi)$). La contrainte est

$$(\gamma^{S,4}, c^{S,4}) = (\gamma^S + 2\pi, c^S) \neq (\gamma^S, c^S)$$

et elle est toujours réalisée.

- (v) **Symétrie** ε^5 . $\theta = \pi \bmod(2\pi)$ (car $\theta = \pi = -\theta \bmod(\pi)$) ou $\varrho = 0$, $\theta = 0$ (car il faut $\theta = -\theta$). La contrainte est

$$(\gamma^{S,5}, c^{S,5}) = (\gamma_t^S + 2\pi, -c_t^S) \neq (\gamma^S, c^S) .$$

Elle est toujours réalisée car la strate C_3 est composée de quatre composantes connexes et les trajectoires restent incluses dans une seule de ces composantes. Or dans chaque composante c_s^S est $\neq 0$ et de signe constant. On ne peut donc pas avoir $-c_t^S = c^S$.

- (vi) **Symétrie** ε^6 . $\theta = 0 \bmod(2\pi)$. La contrainte est

$$(\gamma^{S,6}, c^{S,6}) = (-\gamma_t^S + 2\pi, c_t^S) \neq (\gamma^S, c^S) .$$

Elle est toujours réalisée car la strate C_3 ne comprend pas les points C_3 et C_5 et ses quatre composantes connexes ont des $\gamma_s^S \in]-\pi, \pi[\cup]\pi, 3\pi[$ (union disjointe d'intervalles ouverts. Si par exemple $\gamma^S \in]-\pi, \pi[$, alors $-\gamma_t^S + 2\pi \in]\pi, 3\pi[$ appartient à une autre composante, ce qui est impossible puisque les trajectoires ne quittent pas leur composante.

- (vii) **Symétrie** ε^7 . $2\psi = 0 \bmod(2\pi)$, i.e. $y = 0$ et $\theta = 0 \bmod(\pi)$ (car $\theta = \pi = -\theta \bmod(\pi)$). La contrainte est

$$(\gamma^{S,7}, c^{S,7}) = (-\gamma^S + 2\pi, -c^S) \neq (\gamma^S, c^S) .$$

Elle est toujours réalisée pour la même raison que pour la symétrie ε^5 car $-c^S = c^S$ implique $c^S = 0$ et les points de $c^S = 0$ n'appartiennent pas à C_3 .

Si l'on considère $SE(2) = \mathbb{R}^2 \times \mathbb{S}^1$ comme une fibration de base \mathbb{S}^1 et de fibre \mathbb{R}^2_θ , la contrainte $\varepsilon^1 : 2\psi - \theta = 0 \pmod{2\pi}$, soit $\psi = \frac{\theta}{2} \pmod{\pi}$, définit un ruban de Moebius "appuyé" sur \mathbb{S}^1 ($\rho = 0$) et partant de la demi-droite $\psi = 0$ de \mathbb{R}^2_0 pour $\theta = 0$, passant par la demi-droite $\psi = \frac{\pi}{2}$ de \mathbb{R}^2_π pour $\theta = \pi$ et revenant enfin sur la demi-droite $\psi = \pi$ de $\mathbb{R}^2_{2\pi=0}$ pour $\theta = 2\pi = 0$. De même, la contrainte $\varepsilon^2 : 2\psi - \theta = \pi \pmod{2\pi}$, soit $\psi = \frac{\theta}{2} + \frac{\pi}{2} \pmod{\pi}$, définit un ruban de Moebius également "appuyé" sur \mathbb{S}^1 ($\rho = 0$) et partant de la demi-droite $\psi = \frac{\pi}{2}$ de \mathbb{R}^2_0 pour $\theta = 0$, passant par la demi-droite $\psi = \pi$ de \mathbb{R}^2_π pour $\theta = \pi$ et revenant enfin sur la demi-droite $\psi = \frac{3\pi}{2}$ de $\mathbb{R}^2_{2\pi=0}$ pour $\theta = 2\pi = 0$. Quant aux contraintes $\theta = 0$ et $\theta = \pi$ intervenant dans les symétries $\varepsilon^4, \varepsilon^5, \varepsilon^6, \varepsilon^7$, elles correspondent aux fibres \mathbb{R}^2_0 et \mathbb{R}^2_π tout entières.

La figure 43 illustre (la moitié de) ces contraintes dans la représentation "torique" de $SE(2)$ comme fibration $\mathbb{R}^2 \times \mathbb{S}^1 \rightarrow \mathbb{S}^1$ qui, nous l'avons vu plus haut à la section 4.4.5, utilise un difféomorphisme simple entre $\mathbb{R}^2 \times \mathbb{S}^1$ et l'intérieur d'un tore $\mathring{\mathbb{D}} \times \mathbb{S}^1$ où $\mathring{\mathbb{D}}$ est l'intérieur d'un disque.

Une fois obtenues ces contraintes associées aux strates et aux symétries, on peut alors utiliser les formules explicites des géodésiques pour voir à quelles conditions elles peuvent être satisfaites et dans ce cas pour quelles valeurs de t . Les calculs sont assez complexes et nous référons le lecteur à [373] et surtout au magistral "Quaderno di Matematica" de Sachkov de l'Université Bicocca de Milan [479]. Ils font intervenir tout un ensemble de fonctions construites à partir de fonctions elliptiques de Jacobi et d'intégrales elliptiques, fonctions dont il s'agit d'étudier certaines propriétés. Mais certaines conditions se révèlent être assez simples.

Par exemple la contrainte $\theta_t = 0$ pour $t > 0$ est équivalente à $t = 4K(k)n$ sur C_1 , $t = 4kK(k)n$ ou $\operatorname{cn}\left(\frac{\varphi + \varphi t}{2k}\right) = 0$ sur C_2 et est impossible sur C_3 . La contrainte $\theta_t = \pi$ pour $t > 0$ est quant à elle équivalente à $t = 2K(k) + 4K(k)n$ sur C_1 et impossible sur C_2 et sur C_3 .

Pour les conditions $2\psi - \theta = 0 \pmod{2\pi}$ ou $\varrho = 0$ (symétrie ε^1), etc. on retrouve les mêmes conditions : $\operatorname{cn}\left(\frac{\varphi + \varphi t}{2k}\right) = 0$ sur C_1 (qui est bien une condition sur t à cause de φ_t), etc. La seule différence est pour la contrainte $2\psi - \theta = \pi \pmod{2\pi}$ ou $\varrho = 0$ (symétrie ε^1) sur la strate C_2 . En plus de $\operatorname{sn}\left(\frac{\varphi + \varphi t}{2k}\right) = 0 \pmod{4kK(k)}$ il y a aussi les valeurs t_n de t , $t_n = 2kp_n(k)$, associées aux zéros $p_n(k)$ de la fonction

$$f(p, k) = \operatorname{cn}(p) (\mathbb{E}(p) - p) - \operatorname{dn}(p) \operatorname{sn}(p) .$$

On montre que

$$p_n(k) \in [-K(k) + 2K(k)n, 2K(k)n]$$

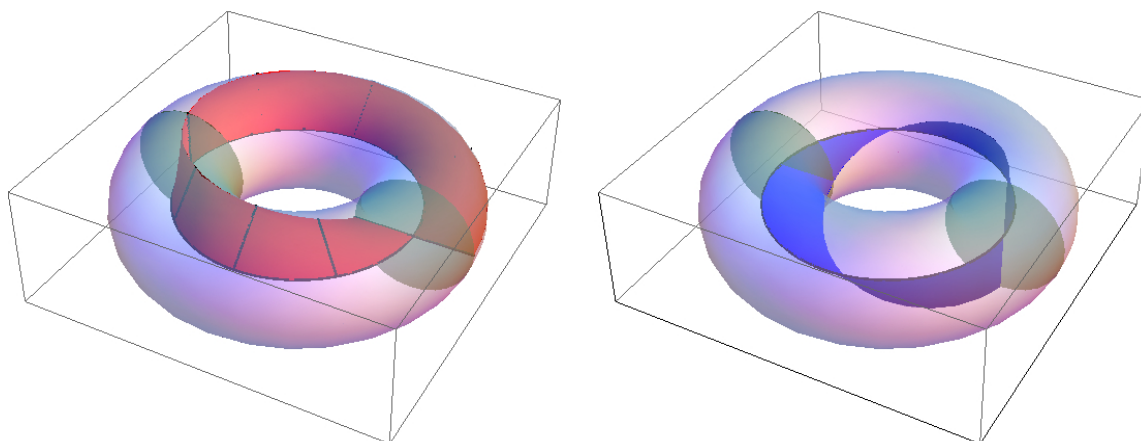


FIGURE 43. Les contraintes $q^i(t) = q(t)$ représentées dans le modèle “torique” $\mathring{\mathbb{D}} \times \mathbb{S}^1 \rightarrow \mathbb{S}^1$ de la fibration $SE(2) = \mathbb{R}^2 \times \mathbb{S}^1 \rightarrow \mathbb{S}^1$ ($\mathring{\mathbb{D}}$ est l’intérieur du disque de \mathbb{R}^2 de centre 0 et de rayon $\frac{1}{2}$). $\mathring{\mathbb{D}} \times \mathbb{S}^1$ est le tore plein ouvert (i.e. sans sa surface). La contrainte $\varepsilon^1 : 2\psi - \theta = 0 \pmod{2\pi}$, soit $\psi = \frac{\theta}{2} \pmod{\pi}$, définit un ruban de Moebius “appuyé” sur \mathbb{S}^1 ($\rho = 0$) dont *une moitié* est représentée en rouge : elle part de la demi-droite $\psi = 0$ de $\mathring{\mathbb{D}}_0$ pour $\theta = 0$, passe par la demi-droite $\psi = \frac{\pi}{2}$ de $\mathring{\mathbb{D}}_\pi$ pour $\theta = \pi$ et revient enfin sur la demi-droite $\psi = \pi$ de $\mathring{\mathbb{D}}_{2\pi=0}$ pour $\theta = 2\pi = 0$. De même, la contrainte $\varepsilon^2 : 2\psi - \theta = \pi \pmod{2\pi}$, soit $\psi = \frac{\theta}{2} + \frac{\pi}{2} \pmod{\pi}$, définit un ruban de Moebius également “appuyé” sur \mathbb{S}^1 ($\rho = 0$) dont *une moitié* est représentée en bleu : elle part de la demi-droite $\psi = \frac{\pi}{2}$ de $\mathring{\mathbb{D}}_0$ pour $\theta = 0$, passe par la demi-droite $\psi = \pi$ de $\mathring{\mathbb{D}}_\pi$ pour $\theta = \pi$ et revient enfin sur la demi-droite $\psi = \frac{3\pi}{2}$ de $\mathring{\mathbb{D}}_{2\pi=0}$ pour $\theta = 2\pi = 0$. Quant aux contraintes $\theta = 0$ et $\theta = \pi$ intervenant dans les symétries $\varepsilon^4, \varepsilon^5, \varepsilon^6, \varepsilon^7$, elles correspondent aux fibres $\mathring{\mathbb{D}}_0$ et $\mathring{\mathbb{D}}_\pi$ tout entières.

et donc

$$t_n \in [-2kK(k) + 4kK(k)n, 4kK(k)n] ,$$

ainsi que $p_{-n}(k) = -p_n(k)$. La figure 44 montre les graphes de ces fonctions pour $k = 0.25, 0.5, 0.75$. On peut y vérifier les encadrements des zéros. Pour $k = 0$ la fonction devient $-\sin(p)$ et, comme $K(0) = \frac{\pi}{2}$ les intervalles contenant les $p_n(0)$ deviennent $[-\frac{\pi}{2} + n\pi, n\pi]$. Le premier est $[\frac{\pi}{2}, \pi]$ et le zéro π est la borne. Pour k petit, les zéros sont proches des bornes. Par exemple, pour $k = 0.25$ le premier zéro non nul est $p \sim 3.09223$ alors que la borne est ~ 3.19248 puisque $K(0.25) \sim 1.59624$. Nous retrouverons cette fonction f à la section 4.7.

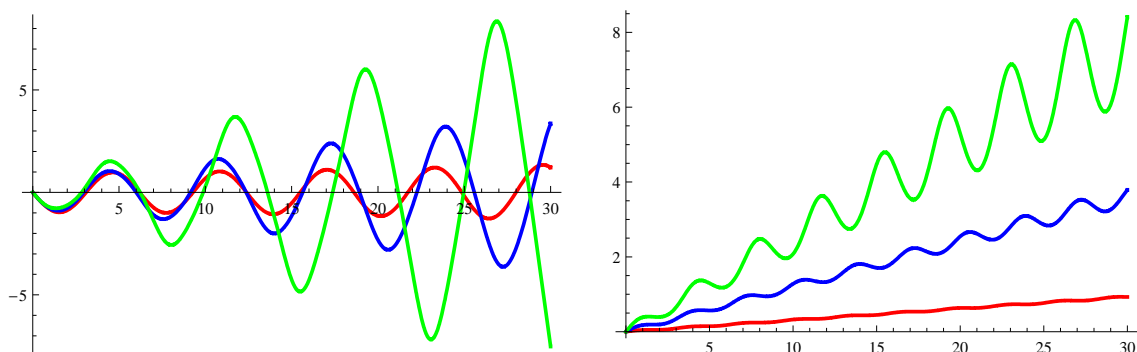


FIGURE 44. Les courbes $f(p, k) = \operatorname{cn}(p)(\mathbb{E}(p) - p) - \operatorname{dn}(p)\operatorname{sn}(p)$ (à gauche) et $k^2 \operatorname{cn}(p)\operatorname{sn}(p) + \operatorname{dn}(p)(p - \mathbb{E}(p))$ (à droite) pour $k = 0.25$ (rouge), 0.5 (bleu), 0.75 (vert). Ce sont les zéros de la première qui sont importants.

Pour la strate C_1 , intervient une fonction analogue

$$k^2 \operatorname{cn}(p)\operatorname{sn}(p) + \operatorname{dn}(p)(p - \mathbb{E}(p))$$

mais elle est toujours > 0 comme le montre aussi la figure 44.

Un théorème dit alors que la géodésique $q(t)$ sur $[0, T]$ est optimale si et seulement si elle ne contient aucun point de Maxwell de ses restrictions à $[0, t]$ ni aucune limite de points de Maxwell. Sachkov en déduit une borne supérieure pour le temps de coupure (cut time) t_c des géodésiques, i.e., répétons-le encore une fois, le temps auquel elles perdent leur propriété d'optimalité. On a toujours $t_c \leq$ premier temps auquel la géodésique rencontre l'ensemble Maxwell. On en dérive après de longs calculs (cf. [479]) le remarquable théorème :

Théorème (Sachkov, 2009) Si $\lambda^S = (\gamma^S, c^S)$, le temps de coupure $t_{cut}(\lambda^S)$ de la géodésique quittant l'origine avec la condition initiale λ^S et le premier temps conjugué sont ³³

1. pour $\lambda^S \in C_1$, $t_{cut} = 2K(k)$, $t_{conj} = +\infty$ (il n'y a pas de points conjugués),
2. pour $\lambda^S \in C_2$, $t_{cut} = 2kp_1(k)$, $t_{cut} = 2kp_1(k) \leq t_{conj} \leq 4K(k)$
3. pour $\lambda^S \in C_3$, $t_{cut} = +\infty$, $t_{conj} = +\infty$,
4. pour $\lambda^S \in C_4$, $t_{cut} = \pi$, $t_{conj} = +\infty$,
5. pour $\lambda^S \in C_5$, $t_{cut} = +\infty$, $t_{conj} = +\infty$.

33. Rappelons que dans les coordonnées de Jacobi (φ, k) le module k encode l'énergie du pendule et φ le temps pendulaire, c'est-à-dire la longueur de la géodésique, et que c^S , $\cos\left(\frac{\gamma^S}{2}\right)$ et $\cos\left(\frac{\gamma^S}{2}\right)$ s'expriment facilement en termes des fonctions elliptiques de Jacobi.

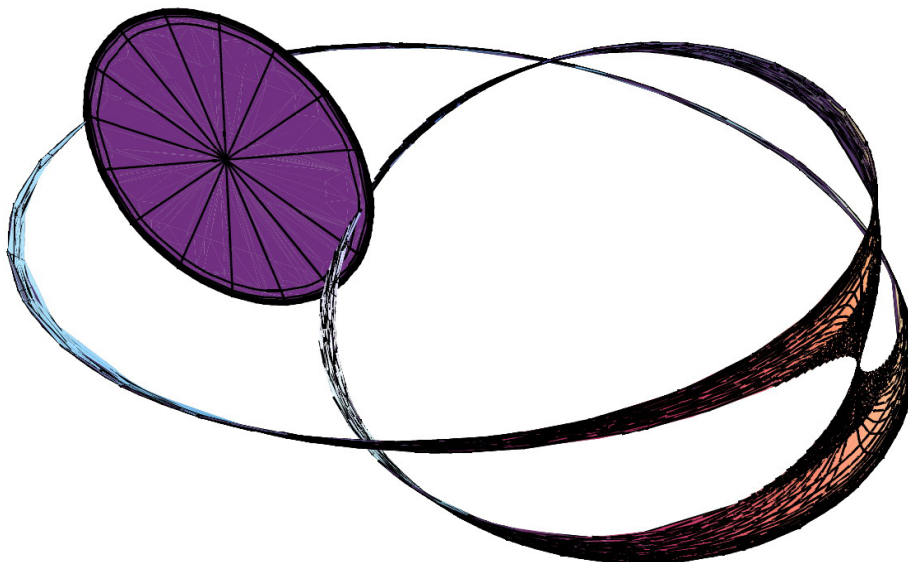


FIGURE 45. Le cut locus de $SE(2)$ dans le modèle toroïdal d'après Yuri Sachkov [479]

La géométrie du cut locus CL de $SE(2)$ est subtile. CL comprend le plan \mathbb{R}_π^2 opposé à l'origine 0 et deux parties partant de 0 (0 non compris) au-dessus de $\theta \in (-\pi, \pi)$ et incluses dans le ruban de Moebius $2\psi - \theta = \pm\pi \pmod{2\pi}$ (dont la moitié est représentée en bleu dans la figure 43). Ces branches sont limitées par une courbe et sont représentées à la figure 45 dans le modèle toroïdal de $SE(2)$.

4.6. Récapitulation

Avant de continuer récapitulons la situation.

1. Nous étudions la géométrie sous-riemannienne du groupe non nilpotent $SE(2)$ et nous essayons de comprendre comment elle déforme et stabilise celle de sa "nilpotentisation" \mathbb{H}_{pol} en en levant les dégénérescences.
2. Les équations de Hamilton donnent les géodésiques partant de 0 avec des conditions initiales λ^S (S pour "Sachkov") et paramétrées par la longueur d'arc t .
3. Les conditions initiales $\lambda^S = (\gamma^S, c^S)$ appartiennent à l'espace des phases \mathcal{C} d'un pendule, celui-ci étant décomposé en strates C_i possédant plusieurs composantes. La dynamique de ces conditions initiales est $\ddot{\gamma}^S = -\sin(\gamma^S)$, $\dot{c}^S = \dot{\gamma}^S$.
4. Les cusps et les points d'inflexion des projections des géodésiques sur le plan (x, y) correspondent respectivement

- (i) à $\sin\left(\frac{\gamma^S}{2}\right) = 0$ et $\dot{q} = -\cos\left(\frac{\gamma^S}{2}\right) X_2$ porté par $X_2 = \partial_\theta$ et
- (ii) à $\cos\left(\frac{\gamma^S}{2}\right) = 0$ et $\dot{q} = \sin\left(\frac{\gamma^S}{2}\right) X_1$ porté par X_1 .
5. Les coordonnées elliptiques de Jacobi (φ, k) “rectifient” la dynamique du pendule car $\dot{k} = 0$ et $\dot{\varphi} = 1$, i.e. $\varphi_t = \varphi + t$ avec $\varphi = \varphi_0$.³⁴ Leurs relations avec les coordonnées $\lambda^S = (\gamma^S, c^S)$ sont explicites et simples :
- pour C_1 ,
- $$\sin\left(\frac{\gamma^S}{2}\right) = sg_1 k \operatorname{sn}(\varphi, k), \cos\left(\frac{\gamma^S}{2}\right) = sg_1 \operatorname{dn}(\varphi, k), c^S = 2k \operatorname{cn}(\varphi, k),$$
- pour C_2 ,
- $$\sin\left(\frac{\gamma^S}{2}\right) = sg_2 \operatorname{sn}\left(\frac{\varphi}{k}, k\right), \cos\left(\frac{\gamma^S}{2}\right) = \operatorname{cn}\left(\frac{\varphi}{k}, k\right), c^S = sg_2 \frac{2}{k} \operatorname{dn}\left(\frac{\varphi}{k}, k\right),$$
- etc.
6. Les équations des géodésiques deviennent donc des fonctions des trois variables (φ, k, t) $\{x(\varphi, k, t) = x_t(\varphi, k), y(\varphi, k, t) = y_t(\varphi, k), \theta(\varphi, k, t) = \theta_t(\varphi, k)\}$ où t est à la fois le temps et la longueur. Elles reposent sur les trigonométries circulaire, elliptique ou hyperbolique suivant les strates C_i de \mathcal{C} . Quand on suit une géodésique, (φ, k) est fixé et t est variable. Il s’agit d’un problème *direct* : on se donne (φ, k) et on calcule les fonctions de t $\{x_t(\varphi, k), y_t(\varphi, k), \theta_t(\varphi, k)\}$, c’est-à-dire la fonction exponentielle $\operatorname{Exp}(\lambda^S, t)$, à partir de son espace *source*. Lorsqu’on fixe t et qu’on fait varier (φ, k) on obtient les surfaces que sont les sphères $S(0, R)$ et les fronts d’onde $W(0, R)$ d’origine 0 et de rayon R .

34. Pour faire les calculs, on peut aussi utiliser des systèmes de coordonnées “normales” qui rectifient les géodésiques elles-mêmes. Andrei Agrachev les définit de la façon suivante. La forme de contact ω étant définie par la translée à gauche de la 1-forme dy en 0, soit Γ l’axe des y constitué des points $(0, u, 0)$. Soit A_Γ le \mathbb{R}^3 de $T^*\mathbb{V}_S$ constitué des duaux des plans de contact le long de l’axe Γ : $A_\Gamma = \{(0, u, 0; \lambda_x, 0, \lambda_\theta)\}$. Sur A_Γ , l’hamiltonien $H = \frac{1}{2}(\lambda_x^2 + \lambda_\theta^2)$ définit une co-métrie et la condition $H = \frac{1}{2}$ signifie que, dans les plans cotangents, le covecteur $(\lambda_x, 0, \lambda_\theta)$ est sur le cercle de rayon 1 d’équation $\lambda_x^2 + \lambda_\theta^2 = 1$. Autrement dit, $A_\Gamma \cap H^{-1}\left(\frac{1}{2}\right)$ est le cylindre C_Γ de rayon 1 et d’axe Γ . Le changement de coordonnées est alors l’application $\Phi : A_\Gamma \rightarrow \mathbb{V}_S$ qui, au point $(0, u, 0; \lambda_x, 0, \lambda_\theta)$ de A_Γ associe la géodésique γ de conditions initiales $x = 0, y = u, \theta = 0, \lambda_x, \lambda_y = 0, \lambda_\theta$, prise au temps $s = 1$. Par Φ^{-1} la géodésique γ devient *la droite* reliant $(0, u, 0)$ à $(\lambda_x, u, \lambda_\theta)$, le point $(s\lambda_x, u, s\lambda_\theta)$ correspondant à $\gamma(s)$. Dans ces coordonnées normales qui “rectifient” les géodésiques on peut alors calculer (c’est très technique) des formes normales pour les cut loci, les lieux conjugués et les fronts d’onde. Ici, Sachkov et Moiseev utilisent les coordonnées de Jacobi rectifiant la dynamique pendulaire des contrôles λ .

7. Mais ces surfaces $S(0, R)$ et $W(0, R)$ possèdent des *singularités*. Il existe dans $\mathbb{V}_S = SE(2)$ un ensemble singulier global \mathfrak{S} de sa géométrie sous-riemannienne (cut locus, lieu conjugué, caustique) qui est le lieu singulier de l'application Exp. Il est très difficile à étudier car il est défini par la dégénérescence du Jacobien

$$J = \frac{\partial(x(\varphi, k, t), y(\varphi, k, t), \theta(\varphi, k, t))}{\partial(\varphi, k, y)}$$

égal au déterminant de la matrice 3×3 des dérivées partielles. Pour ces valeurs singulières, Exp *n'est plus* un difféomorphisme. J est facile à expliciter en appliquant les formules des dérivées des fonctions elliptiques et les formules d'addition pour

$$\begin{aligned} \frac{\varphi + \varphi_t}{2} &= \varphi + \frac{t}{2} \quad (\text{cas } C_1) \\ \frac{\varphi + \varphi_t}{2} &= \frac{\varphi}{k} + \frac{t}{2k} \quad (\text{cas } C_2) \end{aligned}$$

mais est pénible à analyser car la formule finale est compliquée. On la trouve dans le papier [479].

– Pour C_1

$$\begin{aligned} J &= \frac{4}{k^3 (1 - k^2) (1 - k^2 \operatorname{sn}^2(\frac{t}{2}) \operatorname{sn}^2(\frac{\varphi}{2}))} \\ &\quad \times \left(u \operatorname{sn}^2\left(\varphi + \frac{t}{2}\right) + v \operatorname{cn}^2\left(\varphi + \frac{t}{2}\right) \right) \end{aligned}$$

avec u et v définis par

$$\begin{aligned} u &= (1 - k^2) \left(\frac{t}{2} - \mathbb{E}\left(\frac{t}{2}\right) \right) \left(\mathbb{E}\left(\frac{t}{2}\right) - (1 - k^2) \frac{t}{2} \right) \\ v &= \left(\frac{t}{2} - \mathbb{E}\left(\frac{t}{2}\right) \right) \left(\mathbb{E}\left(\frac{t}{2}\right) - (1 - k^2) \frac{t}{2} \right) \\ &\quad + k^2 \operatorname{cn}\left(\frac{t}{2}\right) \operatorname{dn}\left(\frac{t}{2}\right) \operatorname{sn}\left(\frac{t}{2}\right) \left(2\mathbb{E}\left(\frac{t}{2}\right) - (2 - k^2) \frac{t}{2} \right) \\ &\quad + k^2 \left(\left(\mathbb{E}\left(\frac{t}{2}\right) - \frac{t}{2} \right) \left(\mathbb{E}\left(\frac{t}{2}\right) - (1 - k^2) \frac{t}{2} \right) - k^2 \right) \operatorname{sn}^2\left(\frac{t}{2}\right) \\ &\quad + k^4 \operatorname{sn}^4\left(\frac{t}{2}\right). \end{aligned}$$

– Pour C_2

$$J = \frac{4k}{k^3(1-k^2)(1-k^2 \operatorname{sn}^2(\frac{t}{2k}) \operatorname{sn}^2(\frac{\varphi}{k} + \frac{t}{2k}))} \times \left(u \operatorname{sn}^2\left(\frac{\varphi}{k} + \frac{t}{2k}\right) + v \operatorname{cn}^2\left(\frac{\varphi}{k} + \frac{t}{2k}\right) \right)$$

avec

$$u = (1-k^2) \operatorname{sn}\left(\frac{t}{2k}\right) \left(\operatorname{cn}\left(\frac{t}{2k}\right) \operatorname{dn}\left(\frac{t}{2k}\right) \left(\frac{t}{2k} - 2\mathbb{E}\left(\frac{t}{2k}\right)\right) + \operatorname{sn}\left(\frac{t}{2k}\right) \left(\operatorname{dn}^2\left(\frac{t}{2k}\right) + \mathbb{E}\left(\frac{t}{2k}\right)\right) \left(\frac{t}{2k} - \mathbb{E}\left(\frac{t}{2k}\right)\right) \right)$$

$$v = \left(\operatorname{cn}\left(\frac{t}{2k}\right) \left(\mathbb{E}\left(\frac{t}{2k}\right) - \frac{t}{2k}\right) - \operatorname{dn}\left(\frac{t}{2k}\right) \operatorname{sn}\left(\frac{t}{2k}\right) \right) \times \left(\operatorname{cn}\left(\frac{t}{2k}\right) \mathbb{E}\left(\frac{t}{2k}\right) - \operatorname{dn}\left(\frac{t}{2k}\right) \operatorname{sn}\left(\frac{t}{2k}\right) \right) .$$

On reconnaîtra dans la dernière équation la fonction

$$f\left(\frac{t}{2k}, k\right) = \operatorname{cn}\left(\frac{t}{2k}\right) \left(\mathbb{E}\left(\frac{t}{2k}\right) - \frac{t}{2k}\right) - \operatorname{dn}\left(\frac{t}{2k}\right) \operatorname{sn}\left(\frac{t}{2k}\right)$$

introduite plus haut.

8. Une fois ainsi “explicité” le lieu singulier global \mathfrak{S} avec ses strates, il faut alors calculer ses intersections avec les sphères-fronts. Comme les sphères-fronts et \mathfrak{S} sont des surfaces (donc de codimension 1), leurs intersections transverses sont génériquement des lignes de dimension 1 (de codimension 2). On obtient ainsi des lignes singulières qui ont pour extrémités des points singuliers de codimension 3. Le calcul de ces lignes relève d’un problème *inverse* qui, comme en général les problèmes inverses, est beaucoup plus difficile que le problème direct car il exige de trouver des valeurs spéciales de k et φ qui sont des solutions de certaines équations.
9. Pour construire des parties de l’ensemble singulier global \mathfrak{S} on utilise les 7 symétries ε^i de l’espace des phases \mathcal{C} du pendule, symétries qui, avec l’identité, constituent un groupe isomorphe à $(\mathbb{Z}/2\mathbb{Z})^3$. Par exemple

$$\varepsilon^1 : \lambda^S = (\gamma^S, c^S) \mapsto \lambda^{S,1} = (\gamma^{S,1} = \gamma^S, c^{S,1} = -c^S) ,$$

etc.

10. On étend les symétries ε^i aux trajectoires $\lambda_s^S, s \in [0, t]$, dans $\mathcal{C}, \varepsilon^i(\gamma_s^S, c_s^S) = (\gamma_s^{S,i}, c_s^{S,i})$. Par exemple $(\gamma_s^{S,1}, c_s^{S,1}) = (\gamma_{t-s}^S, c_{t-s}^S)$, etc. La *valeur de t* devient essentielle puisque, pour certaines symétries comme ε^1 (en fait pour $i = 1, 2, 5, 6$), la trajectoire $(\gamma_s^{S,1}, c_s^{S,1})$ démarre en $(\gamma_0^{S,1}, c_0^{S,1}) = (\gamma_t^S, c_t^S)$ qui est l’extrémité de la trajectoire $\lambda_s^S, s \in [0, t]$.

11. On étend à partir de là les symétries ε^i aux géodésiques q_s , $s \in [0, t]$, elles-mêmes, $\varepsilon^i(q_s) = q_s^i$. Par exemple,

$$\begin{cases} x_s^1 = \cos(\theta_t)(x_t - x_{t-s}) + \sin(\theta_t)(y_t - y_{t-s}) \\ y_s^1 = \sin(\theta_t)(x_t - x_{t-s}) - \cos(\theta_t)(y_t - y_{t-s}) \\ \theta_s^1 = \theta_t - \theta_{t-s} \end{cases}$$

etc. On vérifie que l'on a bien $q_0^i = (0, 0, 0)$.

12. On peut alors examiner comment une géodésique peut entrer en compétition avec l'une de ses symétriques parce que $q_t^i = q_t$. On obtient pour $s = t$, puisque $(x_0, y_0, \theta_0) = (0, 0, 0)$, des conditions comme par exemple

$$\begin{cases} x_t^1 = \cos(\theta_t)x_t + \sin(\theta_t)y_t = x_t \\ y_t^1 = \sin(\theta_t)x_t - \cos(\theta_t)y_t = y_t \\ \theta_t^1 = \theta_t \end{cases}$$

etc. D'où des équations à résoudre comme

$$\begin{aligned} \cos(\theta)x + \sin(\theta)y &= x \\ \sin(\theta)x - \cos(\theta)y &= y, \end{aligned}$$

etc. Pour les symétries $i = 4, 5, 6, 7$ on trouve aussi la condition $\theta_t^i = -\theta_t = \theta_t$, soit $\theta = 0$.

13. Encore faut-il éliminer les cas triviaux et non pertinents où $q_t^i = q_t$ parce que $q_s^i = q_s$ pour tout $s \in [0, t]$, c'est-à-dire les cas où q^i est tout simplement identique à q . Cela se ramène à éliminer les cas où

$$\lambda^{S,i} = (\gamma^{S,i}, c^{S,i}) = \lambda^S = (\gamma^S, c^S) .$$

Cela dépend des strates. En effet les trajectoires λ_s^S restent dans leur composante connexe et il y aura des *obstructions topologiques* si $\lambda^{S,i}$ appartient à une autre composante connexe que λ^S . Par exemple $\lambda^{S,1} = \lambda^S$ signifie $\gamma_t^S = \gamma^S$ et $c_t^S = -c^S$. Si $\lambda^S \in C_1$, $\lambda^{S,1}$ appartient à la même composante connexe et $\lambda^{S,1} = \lambda^S$ est possible. Il n'y a pas d'obstruction topologique et la condition est donc, si l'on revient aux formules donnant (γ^S, c^S) dans les coordonnées de Jacobi, $\operatorname{sn}(\varphi_t, k) = \operatorname{sn}(\varphi, k)$, $\operatorname{dn}(\varphi_t, k) = \operatorname{dn}(\varphi, k)$, $\operatorname{cn}(\varphi_t, k) = -\operatorname{cn}(\varphi, k)$, ce qui est équivalent à

$$\varphi_t = 2K - \varphi \pmod{4K}.$$

En revanche, si $\lambda^S \in C_2$, $\lambda^{S,1}$ appartient à l'autre composante connexe que λ^S et donc $\lambda^{S,1} = \lambda^S$ est impossible à cause d'une obstruction topologique. Etc.

14. On obtient en définitive les équations devant satisfaire les strates/points de Maxwell où deux géodésiques différentes (i.e. de conditions initiales différentes) entrent en compétition au temps-longueur t . Pour des conditions initiales λ^S

données on définit ainsi des temps-longueur t *singuliers* auxquels la géodésique intersecte le lieu singulier \mathfrak{S} , en l'occurrence des strates de Maxwell.

15. Pour résoudre ces équations, on introduit de nouvelles coordonnées mieux adaptées au problème.

4.7. Les sphères et les fronts d'onde de $\mathbb{V}_S = SE(2)$

Les sphères $S(0, R)$ de $\mathbb{V}_S = SE(2)$, ainsi que leurs extensions que sont les fronts d'onde $W(0, R)$, ressemblent à ceux du groupe de Heisenberg polarisé \mathbb{V}_J lorsque le rayon est assez petit mais leurs singularités sont *nettement* plus compliquées. Pour R assez petit les sphères $S(0, R)$ sont homéomorphes à des sphères \mathbb{S}^2 avec deux lieux singuliers qui ressemblent à des singularités “pomme”³⁵ opposées situées sur les images des strates C_2^\pm . Ces singularités sont des petits segments. À partir de ce segment singulier, la surface $S(0, R)$ se prolonge à l'intérieur par le front $W(0, R)$ constitué d'une “pile de disques”³⁶ de rayon de plus en plus petit s'accumulant sur $(0, 0, 0)$, ayant pour bord un cycle (un cercle déformé) de cusps et séparées par de petits éléments de caustiques déployant des “tetrapaks” (cf. plus bas). Elles ont moins de symétries, mais restent topologiquement identiques. Une différence fondamentale avec \mathbb{V}_J est *qu'il n'y a plus d'invariance par dilatation*.

Nous allons essayer d'illustrer cette structure compliquée.

Remarque. Nous considérons les géodésiques partant de l'origine $(0, 0, 0)$ et donc partant avec une tangente horizontale $\theta = 0$ dans la direction des x . Cela introduit une dissymétrie entre les coordonnées x et y . Si l'on considère les géodésiques partant de $(0, 0, \theta)$ les graphiques ci-dessous tournent de θ . _____ \square

4.7.1. Structure générale.

Dans les formules pour les strates C_1 et C_2 , les fonctions elliptiques de Jacobi ont une période, $T_1(k) = 4K(k)$ pour C_1 et $T_2(k) = 4kK(k)$ pour C_2 (cf. plus haut la figure 18). Comme $\varphi_t = \varphi + t$, si $t = T$, et donc $t = R = T(k)$ puisque t est la longueur, les formules se simplifient. L'équation $t = R = T(k)$ a au plus une

35. Cf. la figure 10 de la section 2.2.3 du chapitre 14. Il s'agit de la singularité de l'épicarpe au voisinage du pédoncule.

36. Un peu comme des disques de discoboles, fins au bord et renflés au centre.

solution $k = k_R$. Pour cette solution,

$$\begin{aligned}
 \mathbb{E}(\varphi_t) &= \mathbb{E}(\varphi + T) = \int_0^{\varphi+T} \operatorname{dn}^2(\psi, k) d\psi \\
 &= \int_0^T \operatorname{dn}^2(\psi, k) d\psi + \int_T^{\varphi+T} \operatorname{dn}^2(\psi, k) d\psi \\
 &= \int_0^T \operatorname{dn}^2(\psi, k) d\psi + \int_0^\varphi \operatorname{dn}^2(\psi' + T, k) d(\psi' + T) \\
 &= \int_0^T \operatorname{dn}^2(\psi, k) d\psi + \int_0^\varphi \operatorname{dn}^2(\psi', k) d\psi' \\
 &= \mathbb{E}(T) + \mathbb{E}(\varphi) \\
 t + \mathbb{E}(\varphi) - \mathbb{E}(\varphi_t) &= T - \mathbb{E}(T) = R - \mathbb{E}(R)
 \end{aligned}$$

Pour C_1 (par exemple C_1^0) il n'y a de solution k_R que si $t \geq 2\pi$ et alors, pour $k = k_R$,

$$\left\{ \begin{array}{l}
 \cos(\theta_R) = \operatorname{cn}(\varphi, k_R) \operatorname{cn}(\varphi, k_R) + \operatorname{sn}(\varphi, k_R) \operatorname{sn}(\varphi, k_R) = 1 \\
 \sin(\theta_R) = sg_1(\operatorname{sn}(\varphi, k_R) \operatorname{cn}(\varphi, k_R) - \operatorname{cn}(\varphi, k_R) \operatorname{sn}(\varphi, k_R)) = 0 \\
 \theta_R = 0 \\
 x_R = sg_1 \frac{1}{k} \left(\begin{array}{l} \operatorname{cn}(\varphi, k_R) (\operatorname{dn}(\varphi, k_R) - \operatorname{dn}(\varphi, k_R)) \\ + \operatorname{sn}(\varphi, k_R) (R + \mathbb{E}(\varphi) - \mathbb{E}(\varphi + R)) \end{array} \right) \\
 \quad = sg_1 \frac{1}{k} \operatorname{sn}(\varphi, k_R) (R - \mathbb{E}(R)) \\
 y_R = \frac{1}{k} \left(\begin{array}{l} \operatorname{sn}(\varphi, k_R) (\operatorname{dn}(\varphi, k_R) - \operatorname{dn}(\varphi, k_R)) \\ - \operatorname{cn}(\varphi, k_R) (R + \mathbb{E}(\varphi) - \mathbb{E}(\varphi + R)) \end{array} \right) \\
 \quad = -\frac{1}{k} \operatorname{cn}(\varphi, k_R) (R - \mathbb{E}(R))
 \end{array} \right.$$

Comme $\operatorname{sn}(\varphi + 2K, k) = -\operatorname{sn}(\varphi, k)$ et $\operatorname{cn}(\varphi + 2K, k) = -\operatorname{cn}(\varphi, k)$, la ligne de niveau exceptionnelle L_{k_R} (de $\theta_R = 0$ constant) est *symétrique* par rapport l'origine.

La figure 46 montre ce qui se passe dans le cas $t = R = 3\pi$ pour les L_k voisines de L_{k_R} . Dans ce cas $4K(k) = 3\pi$ a pour solution $k_{3\pi} \sim 0.915931$.

Pour C_2 (par exemple C_2^+) il y a toujours une solution k_R de $t = R = 4kK(k)$ et alors, pour $k = k_R$,

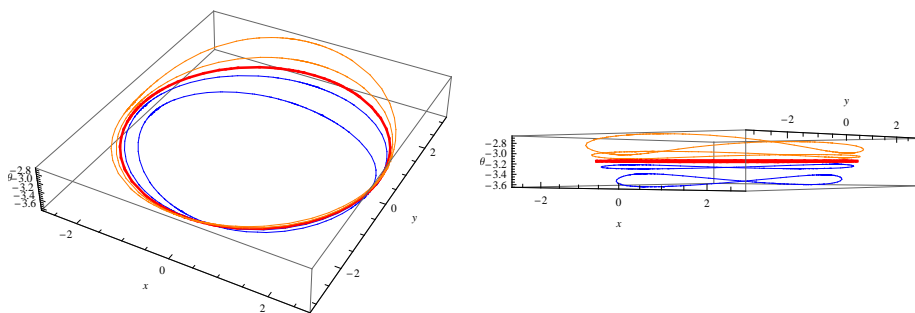


FIGURE 46. La ligne de niveau exceptionnelle L_{k_R} (de $\theta_R = 0$ constant) de l'image de C_1^0 pour le rayon $t = R = 3\pi$ et le module exceptionnel $k_{3\pi} \sim 0.915931$. L_{k_R} est en rouge, les L_k voisines pour $k = 0.85$ et 0.9 sont en bleu et les L_k voisines pour $k = 0.93$ et 0.95 sont en orange.

$$\left\{ \begin{array}{l} \cos(\theta_R) = k_R^2 \operatorname{sn}\left(\frac{\varphi}{k_R}, k_R\right) \operatorname{sn}\left(\frac{\varphi}{k_R}, k_R\right) + \operatorname{dn}\left(\frac{\varphi}{k_R}, k_R\right) \operatorname{dn}\left(\frac{\varphi}{k_R}, k_R\right) = 1 \\ \sin(\theta_R) = k_R \left(\operatorname{sn}\left(\frac{\varphi}{k_R}, k_R\right) \operatorname{dn}\left(\frac{\varphi}{k_R}, k_R\right) - \operatorname{dn}\left(\frac{\varphi}{k_R}, k_R\right) \operatorname{sn}\left(\frac{\varphi}{k_R}, k_R\right) \right) = 0 \\ \theta_R = 0 \\ x_R = k_R \operatorname{sn}\left(\frac{\varphi}{k_R}, k_R\right) \left(\frac{R}{k_R} + \mathbb{E}\left(\frac{\varphi}{k_R}\right) - \mathbb{E}\left(\frac{\varphi}{k_R} + \frac{R}{k_R}\right) \right) \\ \quad = k_R \operatorname{sn}\left(\frac{\varphi}{k_R}, k_R\right) \left(\frac{R}{k_R} - \mathbb{E}\left(\frac{R}{k_R}\right) \right) \\ y_R = -\operatorname{dn}\left(\frac{\varphi}{k_R}, k_R\right) \left(\frac{R}{k_R} + \mathbb{E}\left(\frac{\varphi}{k_R}\right) - \mathbb{E}\left(\frac{\varphi}{k_R} + \frac{R}{k_R}\right) \right) \\ \quad = -\operatorname{dn}\left(\frac{\varphi}{k_R}, k_R\right) \left(\frac{R}{k_R} - \mathbb{E}\left(\frac{R}{k_R}\right) \right) \end{array} \right.$$

Contrairement à ce qui se passe pour C_1 , ici, comme $\operatorname{sn}(\varphi + 2K, k) = -\operatorname{sn}(\varphi, k)$ mais $\operatorname{dn}(\varphi + 2K, k) = \operatorname{dn}(\varphi, k)$, la ligne de niveau exceptionnelle L_{k_R} (de $\theta_R = 0$ constant) est *dégénérée* car y_R étant de période seulement $2K$ prend deux fois la même valeur.

La figure 47 montre ce qui se passe dans le cas $t = R = \frac{\pi}{2}$ pour les L_k voisines de L_{k_R} . Dans ce cas $4kK(k) = \frac{\pi}{2}$ a pour solution $k_{\frac{\pi}{2}} \sim 0.246139$. On notera la très forte variabilité de ces L_k .

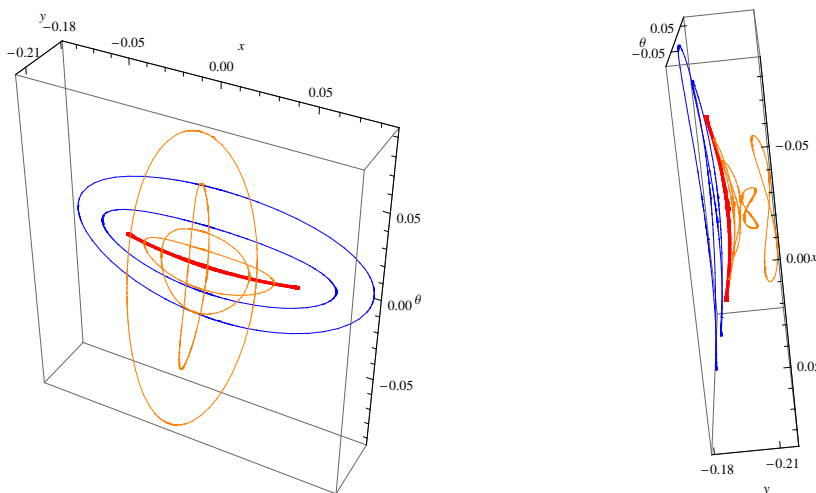


FIGURE 47. La ligne de niveau exceptionnelle dégénérée L_{k_R} (de $\theta_R = 0$ constant) de l'image de C_2^+ pour le rayon $t = R = \frac{\pi}{2}$ et le module exceptionnel $k_{\frac{\pi}{2}} \sim 0.246139$. L_{k_R} est en rouge, les L_k voisines $k = 0.243$ et 0.24 sont en bleu, et les L_k voisines pour $k = 0.248, 0.25, 0.255, 0.26$ sont en orange.

On note alors un phénomène intéressant. Pour $k_{c,R} \sim 0.2541$ voisin de $k_{\frac{\pi}{2}} \sim 0.246139$, il existe une autre ligne de niveau $L_{k_{c,\frac{\pi}{2}}}$ dégénérée d'orientation très transverse à $L_{k_{\frac{\pi}{2}}}$. La géométrie du front $W(0, \frac{\pi}{2})$ entre ces valeurs est intéressante ainsi que le montre la figure 48.

Cette géométrie de transition entre L_{k_R} et $L_{k_{c,R}}$ a la forme d'un "tetrapak"³⁷ dont un modèle simple est la surface

$$\{x = (2 - t) \cos(\theta), y = t \sin(\theta)\}$$

($t \in [0, 2], \theta \in [0, 2\pi]$) représentée à la figure 49.

La figure 50 montre le front $W(0, \frac{\pi}{2})$ pour $k \in [0.24, 0.26]$, avec la distorsion effectuant la transition entre les lignes de niveau dégénérées $k_{\frac{\pi}{2}} \sim 0.246139$ et $k_{c,\frac{\pi}{2}} \sim 0.254126$. C'est une sorte de cuvette oblongue ayant pour fond la ligne singulière $L_{k_{\frac{\pi}{2}}}$, suivie du tetrapak interpolant entre $L_{k_{\frac{\pi}{2}}}$ et $L_{k_{c,\frac{\pi}{2}}}$, suivie d'un calot ayant pour sommet la ligne singulière $L_{k_{c,\frac{\pi}{2}}}$. La figure 51 montre, avec un autre point de vue, les intersections du même front avec les surfaces définissant les lignes singulières.

La figure 52 montre la structure analogue pour le rayon $R = \frac{3\pi}{2}$. La première ligne de niveau dégénérée est pour $k_{\frac{3\pi}{2}} \sim 0.6551$, et la seconde pour $k_{c,\frac{3\pi}{2}} \sim 0.8272$.

Alors que la dégénérescence de la ligne de niveau L_k est évidente pour la valeur k_R elle l'est beaucoup moins pour la valeur $k_{c,R}$. Celle-ci résulte de formules techniques

37. Le tetrapak est l'emballage en forme de berlingot inventé en 1951 par Ruben Rausing.

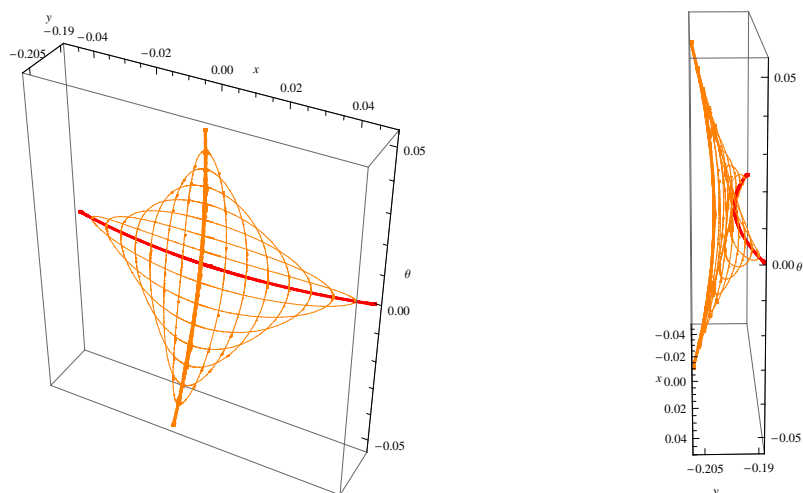


FIGURE 48. Les lignes de niveau L_k entre les lignes dégénérées $k_{\frac{\pi}{2}} \sim 0.246139$ et $k_{c, \frac{\pi}{2}} \sim 0.2541$.

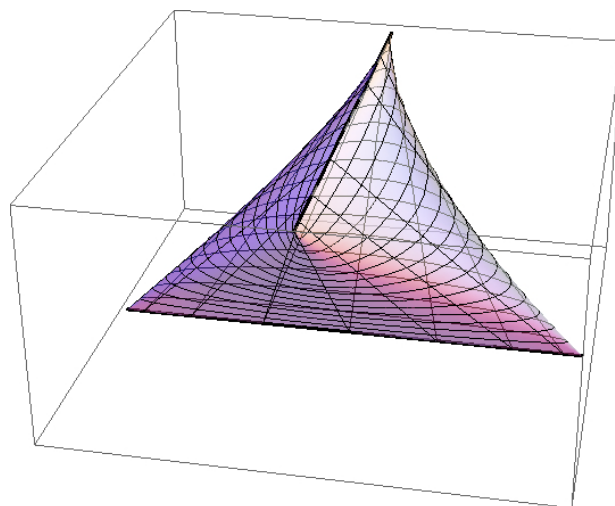


FIGURE 49. Le “tetrapak” d’équations $\{x = (2 - t) \cos(\theta), y = t \sin(\theta), t\}$ ($t \in [0, 2], \theta \in [0, 2\pi]$).

que l’on trouvera dans l’étude de Sachkov et Moiseev [373] déjà souvent citée. $k_{c,R}$ est le *premier zéro* $p_1(k)$ (i.e. le plus grand) de la fonction

$$f(t, k) = \operatorname{cn}\left(\frac{t}{2k}, k\right) \left(\mathbb{E}\left(\frac{t}{2k}\right) - \frac{t}{2k}\right) - \operatorname{dn}\left(\frac{t}{2k}, k\right) \operatorname{sn}\left(\frac{t}{2k}, k\right)$$

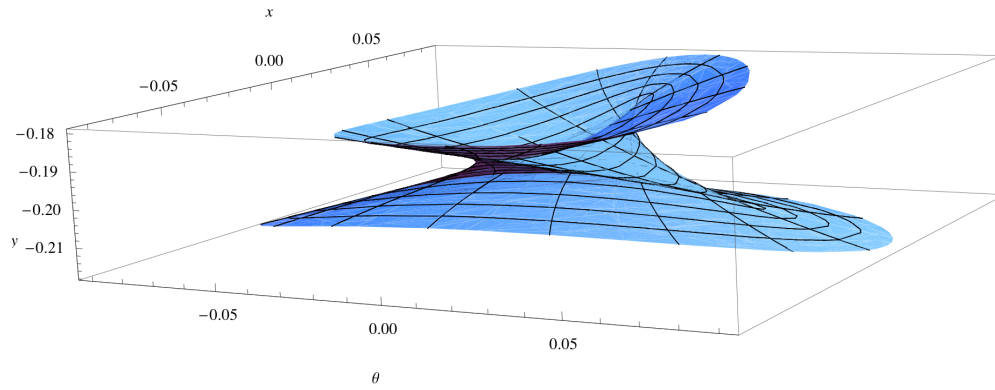


FIGURE 50. Le front $W\left(0, \frac{\pi}{2}\right)$ pour $k \in [0.24, 0.26]$ avec la distorsion effectuant la transition entre les lignes de niveau dégénérées $k_{\frac{\pi}{2}} \sim 0.246139$ et $k_{c, \frac{\pi}{2}} \sim 0.254126$.

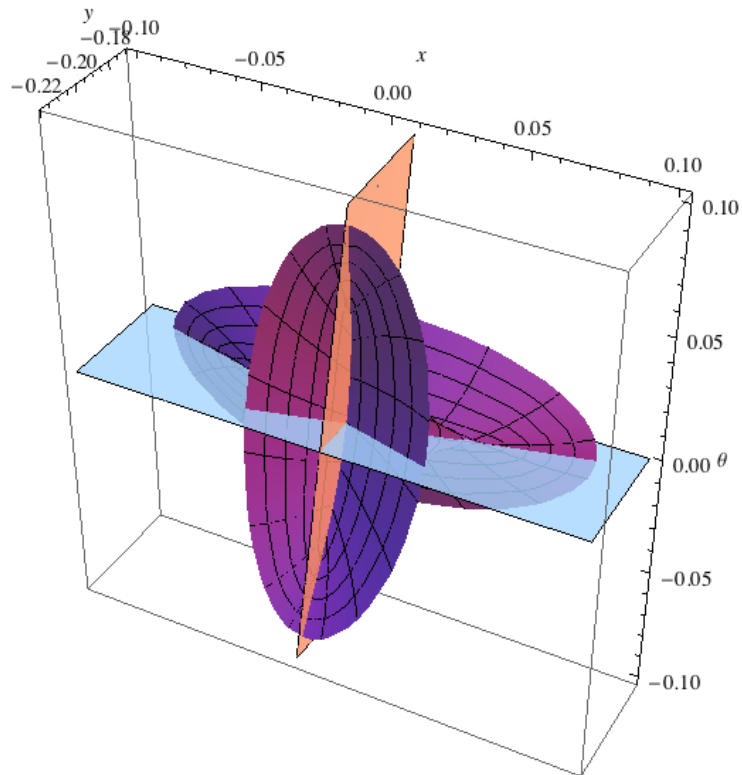


FIGURE 51. Les intersections du front de la figure 50 (sous un autre point de vue) avec les surfaces définissant les lignes singulières.

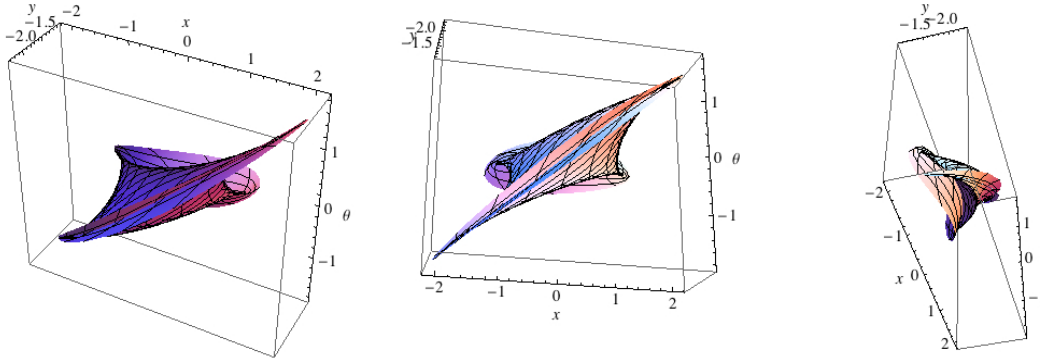


FIGURE 52. Le front $W(0, \frac{3\pi}{2})$ pour $k \in [0.6, 0.85]$ avec la distorsion effectuant la transition entre les lignes de niveau dégénérées $k_{\frac{3\pi}{2}} \sim 0.6551$ et $k_{c, \frac{3\pi}{2}} \sim 0.8272$.

que nous avons rencontrée à la section 4.5. Ce zéro $p_1(k)$ est situé dans l'intervalle $(K(k), 2K(k))$ mais glisse de $2K(k)$ à $K(k)$ lorsque k croît de 0 à 1 et ce comportement est essentiel pour comprendre la structure des points conjugués, du cut locus et de la caustique de $SE(2)$.

Comme nous l'avons alors vu avec le théorème de cette section, cela signifie que l'on est sur une *strate de Maxwell* et que R est le *point de coupure* $t_{cut} = 2kp_1(k)$.

Dans la section 4.5 (figure 44) c'était la condition initiale k qui était donnée et la longueur t de la géodésique qui était variable, ici c'est au contraire la longueur $t = R$ qui est donnée comme rayon de la sphère et c'est k qu'il faut trouver. Il s'agit des points de la sphère $S(0, R)$ et du front $W(0, R)$ qui sont des points de Maxwell et des points de coupure. La figure 53 représente les graphes de $f(\frac{\pi}{2}, k)$ et de $f(\frac{3\pi}{2}, k)$ pour $k \in (0, 1)$.

Évidemment, on aura d'autres lignes de niveau dégénérées $L_{n,R}$ ($L_{1,R} = L_R$) pour $R = 4nkK(k)$. On a toujours $\theta_{n,R} = 0$, mais $x_{n,R} = nx_R$ et $y_{n,R} = ny_R$. La figure 54 montre les $L_{n, \frac{\pi}{2}}$ pour $n = 1, 2, 3$.

Il est intéressant de regarder les extrémités des lignes de niveau dégénérées L_{k_R} lorsque le rayon R varie. Elles parcourent des courbes Γ_n qui correspondent aux extrema de $x_R = nk_R \operatorname{sn}\left(\frac{\varphi}{k_R}, k_R\right) \left(\frac{R}{k_R} - \mathbb{E}\left(\frac{R}{k_R}\right)\right)$. Ces derniers sont réalisés pour $\operatorname{sn}\left(\frac{\varphi}{k_R}, k_R\right) = \pm 1$ et donc $\frac{\varphi}{k_R} = \pm 2K(k_R)$ et comme pour ces valeurs on a $\operatorname{dn}\left(\frac{\varphi}{k_R}, k_R\right) = \sqrt{1 - k^2}$, les équations des Γ_n sont

$$\left\{ \theta = 0, x = \pm nk(4K(k) - \mathbb{E}(4K(k))), y = -n\sqrt{1 - k^2}(4K(k) - \mathbb{E}(4K(k))) \right\} .$$

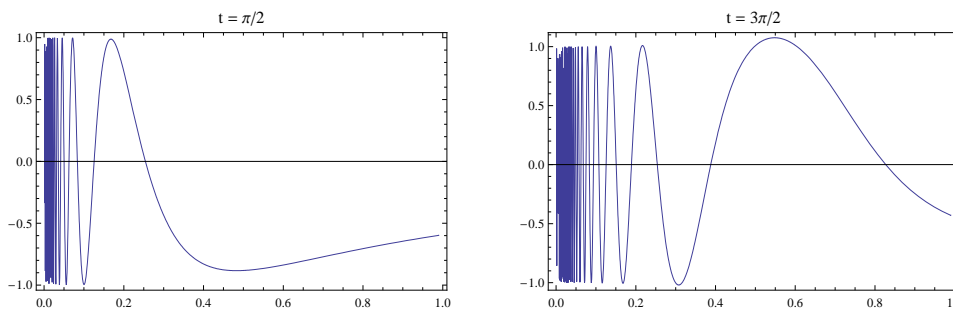


FIGURE 53. Graphes de $f(t, k) = \operatorname{cn}\left(\frac{t}{2k}\right) \left(\mathbb{E}\left(\frac{t}{2k}\right) - \frac{t}{2k}\right) - \operatorname{dn}\left(\frac{t}{2k}\right) \operatorname{sn}\left(\frac{t}{2k}\right)$ pour $k \in (0, 1)$. A gauche le cas $t = R = \frac{\pi}{2}$, à droite le cas $t = R = \frac{3\pi}{2}$. Les zéros $p_n(k)$ sont des points de Maxwell et des points de coupure successifs. On comparera à la figure 44 de la section 4.5 qui donnait les graphes de $f(p, k)$ pour p variable et k fixé.

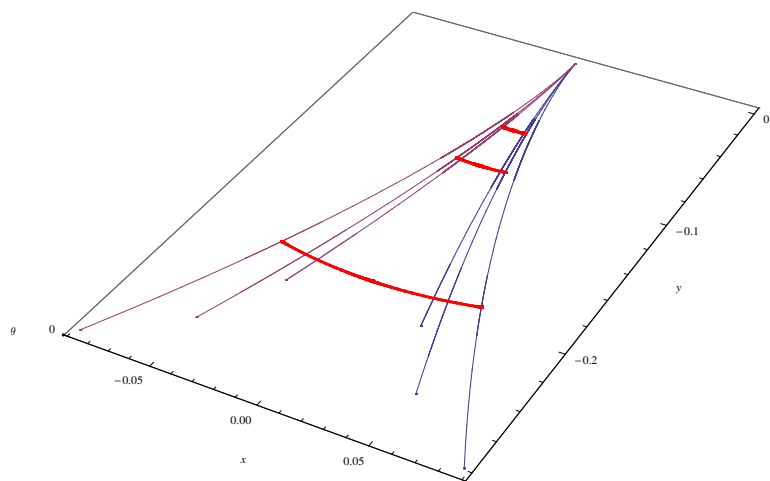


FIGURE 54. En rouge, les trois lignes de niveau dégénérées $L_{n, \frac{\pi}{2}}$ pour $n = 1, 2, 3$. Elles correspondent à $\frac{\pi}{2} = 4nkK(k)$, soit $k = 0.246139$, $k = 0.124513$ et $k = 0.083189$. En noir, les caustiques Γ_n pour $n = 1, 2, 3$. Les extrémités de $L_{n, \frac{\pi}{2}}$ sont sur Γ_n .

La figure 54 montre aussi les courbes Γ_n pour $n = 1, 2, 3$ et l'on voit que les extrémités des $L_{n, \frac{\pi}{2}}$ pour $n = 1, 2, 3$ y sont bien situées. Ces courbes possédant un cusp à l'origine font partie de la géométrie sous-riemannienne globale de $SE(2)$. Dans le cas du modèle nilpotent \mathbb{V}_J (groupe de Heisenberg polarisé), elles se réduisaient tout simplement à l'axe des y , les singularités du front d'onde constituant des cercles de cusps qui sont transverses à cet axe, de plus en plus petits et séparés par des points sur l'axe. Ici, la situation est plus générique et, comme nous allons le voir, il y a encore des cycles de cusps de plus en plus petits séparés par des petites structures en tetrapak plus compliquées qu'un point et appuyées entre autres sur Γ .

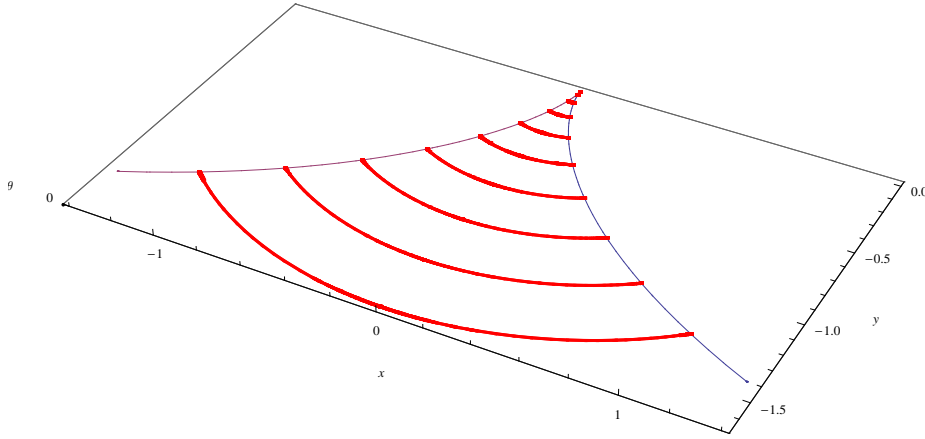


FIGURE 55. La ligne de niveau dégénérée $L_R = L_{1,R}$ pour R variable variant de 0 à $\frac{3\pi}{2}$ par pas de $\frac{\pi}{6}$. Les valeurs de k sont $k_{\frac{\pi}{6}} = 0.083189$, $k_{\frac{2\pi}{6}=\frac{\pi}{3}} = 0.165515$, $k_{\frac{3\pi}{6}=\frac{\pi}{2}} = 0.246139$, $k_{\frac{4\pi}{6}=\frac{2\pi}{3}} = 0.324265$, $k_{\frac{5\pi}{6}} = 0.399162$, $k_{\frac{6\pi}{6}=\pi} = 0.470184$, $k_{\frac{7\pi}{6}} = 0.536784$, $k_{\frac{8\pi}{6}=\frac{4\pi}{3}} = 0.598529$, $k_{\frac{9\pi}{6}=\frac{3\pi}{2}} = 0.655113$. Les extrémités des L_R se déplacent sur la courbe Γ_1 .

Lorsque le rayon R varie, la ligne de niveau dégénérée $L_R = L_{1,R}$ évolue, ses extrémités se déplaçant sur Γ_1 . La figure 55 montre L_R pour R variant de 0 à $\frac{3\pi}{2}$ par pas de $\frac{\pi}{6}$. Les valeurs de k sont

$$\begin{cases} k_{\frac{\pi}{6}} = 0.083189, & k_{\frac{2\pi}{6}=\frac{\pi}{3}} = 0.165515, & k_{\frac{3\pi}{6}=\frac{\pi}{2}} = 0.246139, \\ k_{\frac{4\pi}{6}=\frac{2\pi}{3}} = 0.324265, & k_{\frac{5\pi}{6}} = 0.399162, & k_{\frac{6\pi}{6}=\pi} = 0.470184, \\ k_{\frac{7\pi}{6}} = 0.536784, & k_{\frac{8\pi}{6}=\frac{4\pi}{3}} = 0.598529, & k_{\frac{9\pi}{6}=\frac{3\pi}{2}} = 0.655113. \end{cases}$$

Les lignes de niveau exceptionnelles L_R sont entrelacées avec les autres lignes de niveau exceptionnelles $L_{c,R}$ définies par les zéros de $f\left(\frac{R}{2k}, k\right)$. La figure 56 montre (en orange) les $L_{c,R}$ pour R variant également de 0 à $\frac{3\pi}{2}$ par pas de $\frac{\pi}{6}$. Les valeurs de k sont

$$\begin{cases} k_{c,\frac{\pi}{6}} = 0.083189, & k_{c,\frac{2\pi}{6}=\frac{\pi}{3}} = 0.165515, & k_{c,\frac{3\pi}{6}=\frac{\pi}{2}} = 0.246139, \\ k_{c,\frac{4\pi}{6}=\frac{2\pi}{3}} = 0.324265, & k_{c,\frac{5\pi}{6}} = 0.399162, & k_{c,\frac{6\pi}{6}=\pi} = 0.470184, \\ k_{c,\frac{7\pi}{6}} = 0.536784, & k_{c,\frac{8\pi}{6}=\frac{4\pi}{3}} = 0.598529, & k_{c,\frac{9\pi}{6}=\frac{3\pi}{2}} = 0.655113. \end{cases}$$

Les L_R de la figure 55 sont également représentées (en rouge). On voit la transversalité entre les L_R et les $L_{c,R}$, la courbure des $L_{c,R}$ ainsi que leur rotation. La figure 57 compare les k_R et $k_{c,R}$ et montre comment les $k_{c,R}$ dérivent par rapport aux k_R qui correspondent à des fonctions exactes de périodes. $2K(k)$. Cela est dû au fait que, comme nous l'avons rappelé plus haut, le premier zéro $p_1(k)$ de la fonction $f(t, k)$ est situé dans l'intervalle $(K(k), 2K(k))$ mais glisse de $2K(k)$ à $K(k)$ lorsque k croît de 0 à 1.

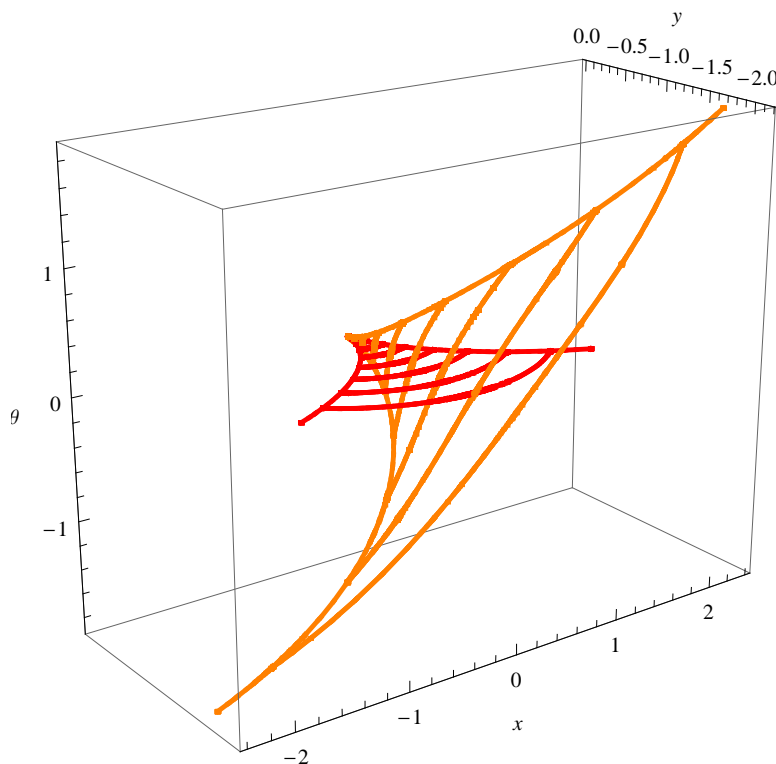


FIGURE 56. En orange, les $L_{c,R}$ pour R variant de 0 à $\frac{3\pi}{2}$ par pas de $\frac{\pi}{6}$. Les valeurs de k sont $k_{c,\frac{\pi}{6}} = 0.083189$, $k_{c,\frac{2\pi}{6}=\frac{\pi}{3}} = 0.165515$, $k_{c,\frac{3\pi}{6}=\frac{\pi}{2}} = 0.246139$, $k_{c,\frac{4\pi}{6}=\frac{2\pi}{3}} = 0.324265$, $k_{c,\frac{5\pi}{6}} = 0.399162$, $k_{c,\frac{6\pi}{6}=\pi} = 0.470184$, $k_{c,\frac{7\pi}{6}} = 0.536784$, $k_{c,\frac{8\pi}{6}=\frac{4\pi}{3}} = 0.598529$, $k_{c,\frac{9\pi}{6}=\frac{3\pi}{2}} = 0.655113$. En rouge, les L_R de la figure 55.



FIGURE 57. Comparaison des k_R et $k_{c,R}$. Les $k_{c,R}$ dérivent par rapport aux k_R qui correspondent à des fonctions exactes de périodes. $2K(k)$.

On voit ainsi que le front $W(0, R)$ de $\mathbb{V}_S = SE(2)$ ressemble à celui de sa nilpotentisation de Heisenberg \mathbb{V}_J mais avec

- (i) les cercles de cusps bordant les “disques” se déformant en des cycles ondulés,
- (ii) l’axe des y constituant la caustique dégénérée de \mathbb{V}_J se démultipliant en les composantes de la caustique bordant les $L_{c,R}$,
- (iii) les points de l’axe des y séparant les “assiettes” se déployant en des petites transitions “en tetrapak” entre les L_R et les $L_{c,R}$.

4.7.2. *Le cas* $R = \frac{\pi}{2}$.

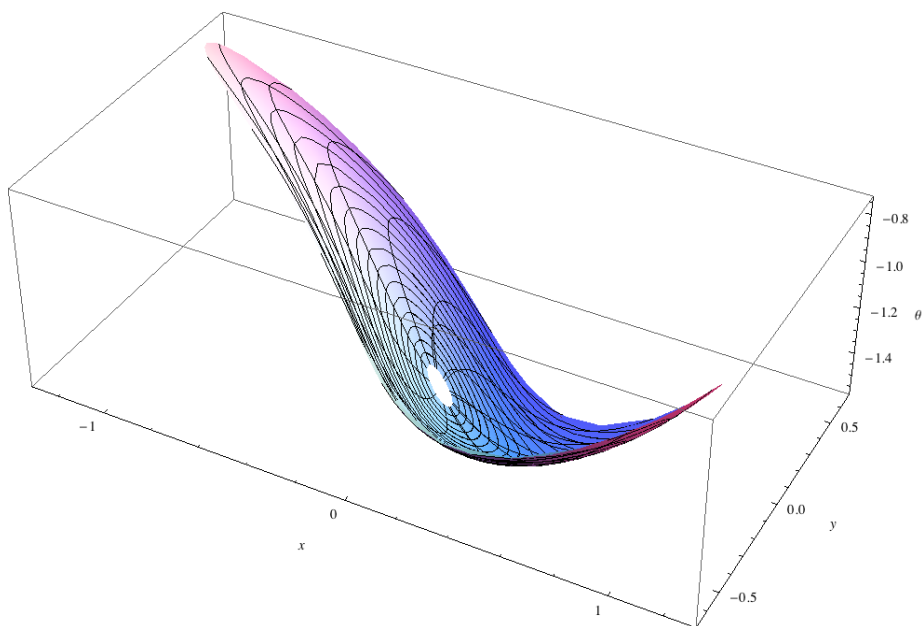


FIGURE 58. L'image $W_{C_1^0, R=\frac{\pi}{2}}$ de la strate C_1^0 pour $k \in (0.1, 0.9)$ dans la sphère-front $W(0, \frac{\pi}{2})$. Pour $k = 0$, les rayons démarrent au point $(0, 0, -\frac{\pi}{2})$.

Les deux strates C_1 s'envoient de façon difféomorphe au temps-longueur t sur deux composantes $W_{C_1, t=R}$ de $W(0, R)$.³⁸ La figure 58 montre l'image $W_{C_1^0, R=\frac{\pi}{2}}$ de la strate C_1^0 ($sg1 = +$) pour $k \in (0.1, 0.9)$. Les courbes parallèles au bord sont les lignes de niveau $k = \text{cste}$ qui se propagent de $(0, 0, -\frac{\pi}{2})$ jusqu'à l'image de la strate C_3 de la figure 60 et les "rayons" sont les lignes de niveau $\varphi = \text{cste}$. On notera qu'à l'approche de $k = 1$ elles deviennent de moins en moins transverses aux lignes de niveau $k = \text{cste}$.

La figure 59 montre l'image $W_{C_3, R=\frac{\pi}{2}}$ des quatre strates C_3 ainsi que sa projection sur le plan (x, y) .

La figure 60 montre les images $W_{C_1^0, R=\frac{\pi}{2}}$ et $W_{C_1^{2\pi}, R=\frac{\pi}{2}}$ des deux strates C_1^0 et $C_1^{2\pi}$ ainsi que l'image $W_{C_3, R=\frac{\pi}{2}}$ des quatre strates C_3 qui les bordent.

Les images des strates C_2^\pm sont beaucoup plus compliquées et beaucoup plus difficiles à bien représenter graphiquement car elles ont une infinité de singularités. La structure générale de $S(0, R)$ est celle d'une "pomme" dont une bande équatoriale (sans singularités) est l'image de $C_1 \cup C_3$ de la figure 60, bande à laquelle on ajoute deux "demi-pommes" avec leur singularité (et pour le front d'onde $W(0, R)$ l'infinité

38. Nous considérons $W(0, R)$ comme une extension de $S(0, R)$, i.e. $S(0, R) \subset W(0, R)$. De façon générale, $W_I(0, R)$ (pour des paramètres I fixés) est le bord $\partial Q_{I, t=R}$ pour $t = R$ de $Q_{I, t=R}$.

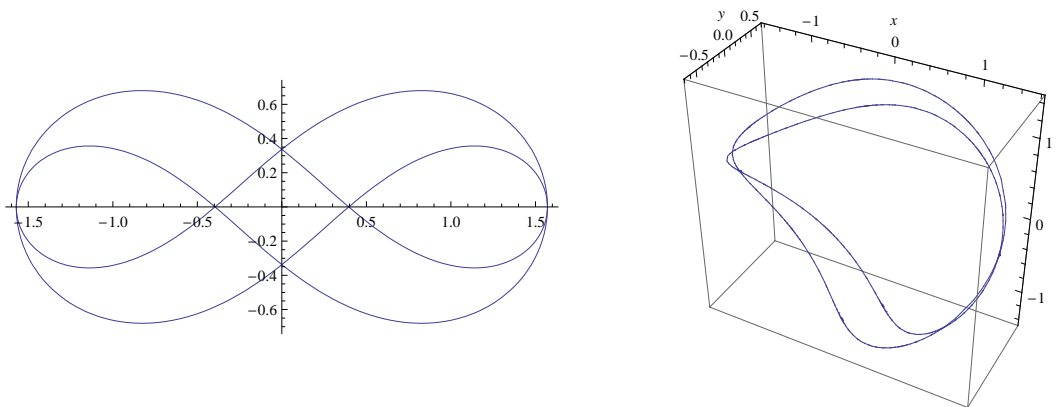


FIGURE 59. Image $W_{C_3, R=\frac{\pi}{2}}$ des quatre strates C_3 . À gauche leur projection sur le plan (x, y) . À droite leur représentation 3D.

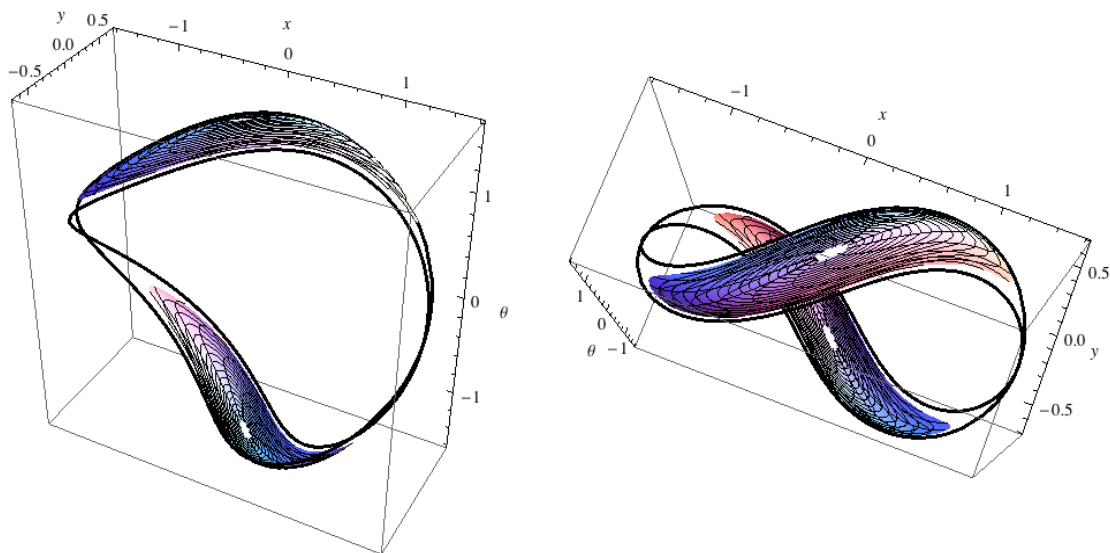


FIGURE 60. Deux points de vue sur les images $W_{C_1^0, R=\frac{\pi}{2}}$ et $W_{C_1^{2\pi}, R=\frac{\pi}{2}}$ des deux strates C_1^0 et $C_1^{2\pi}$ ainsi que sur l'image $W_{C_3, R=\frac{\pi}{2}}$ des quatre strates C_3 qui les bordent.

de cycles de cusps de plus en plus petits convergeant vers l'origine). Nous allons considérer la “demi-pomme” $W_{C_2^+, R=\frac{\pi}{2}}$ image de C_2^+ pour le rayon $R = \frac{\pi}{2}$.

Pour en avoir une idée, on peut commencer par considérer ses lignes de niveau L_k pour $k = \text{cste}$. La figure 62 en montre quelques unes. Ce sont des cercles déformés de “rayon” moyen variable. On y voit clairement deux phénomènes :

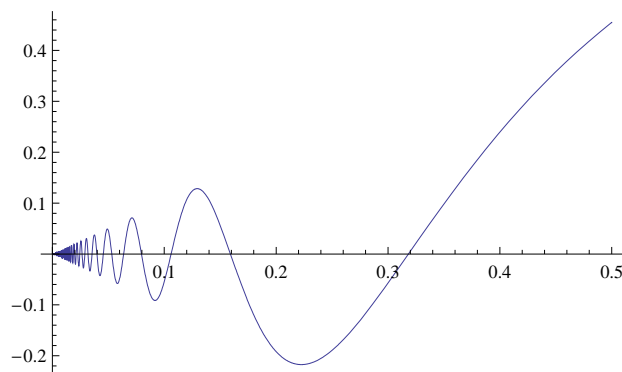


FIGURE 61. Un comportement typique de fonction décroissante tendant vers 0 tout en oscillant de plus en plus fortement. L'exemple de la fonction $x \sin\left(\frac{1}{x}\right)$ au voisinage de 0 ($x \in (0.001, 0.5)$).

- (i) Partant de la limite L_1 qui est l'image de C_3 , L_k se propage qualitativement³⁹ comme un front parallèle.
- (ii) Une fois passée la singularité de $S(0, R)$, les “cercles” L_k ont un “rayon” moyen qui, lorsque $k \rightarrow 0$, décroît tout en oscillant de plus en plus rapidement. Autrement dit, il évolue selon une sorte de sinusoïde de plus en plus rapide enveloppée par une courbe décroissante tendant vers 0, un peu comme la fonction $x \sin\left(\frac{1}{x}\right)$ au voisinage de 0 (cf. figure 61). Ce genre de comportement pose des problèmes d'analyse de convergence qui se reflètent dans des difficultés d'approximation numérique lorsqu'on utilise des logiciels comme *Mathematica*.

Pour les lignes de niveau $\varphi = cste$, qui sont comme des “profils” P_φ , la situation est moins claire car les L_k et les P_φ sont en général fort peu transverses. La figure 63 représente le profil P_0 de $W_{C_2^\pm, R=\frac{\pi}{2}}$ pour $\varphi = 0$ et $k \in [0.0001, 0.9999]$. Elle représente aussi ses projections $\text{Pr}_{0,(y,\theta)}$ et $\text{Pr}_{0,(x,\theta)}$ sur les plans (y, θ) et (x, θ) . La symétrie $(x, y) \rightarrow (-x, -y)$ entre C_2^+ et C_2^- est claire. P_0 part des points symétriques de C_3

$$\left\{ \pm x = 1 - \frac{1}{\cosh\left(\frac{\pi}{2}\right)}, \pm y = -\frac{\pi}{2} + \tanh\left(\frac{\pi}{2}\right), -\arccos\left(\frac{1}{\cosh\left(\frac{\pi}{2}\right)}\right) \right\}$$

$$\sim \{0.6015, -0.6536, -1.1609\}$$

La figure 64 montre comment le profil P_0^+ de $W_{C_2^+, R=\frac{\pi}{2}}$ croise les lignes de niveau L_k . Il s'agit d'un “demi-profil” car on ne considère que la strate C_2^+ .

La figure 65 montre quant à elle les projections d'autres demi-profils P_φ^+ de l'image $W_{C_2^+, R=\frac{\pi}{2}}$ de C_2^+ pour $\varphi = \frac{\pi}{4}, \pi, 2\pi$. On voit que le comportement reste

39. C'est-à-dire avec des distorsions quantitatives mais avec un type qualitatif stable.

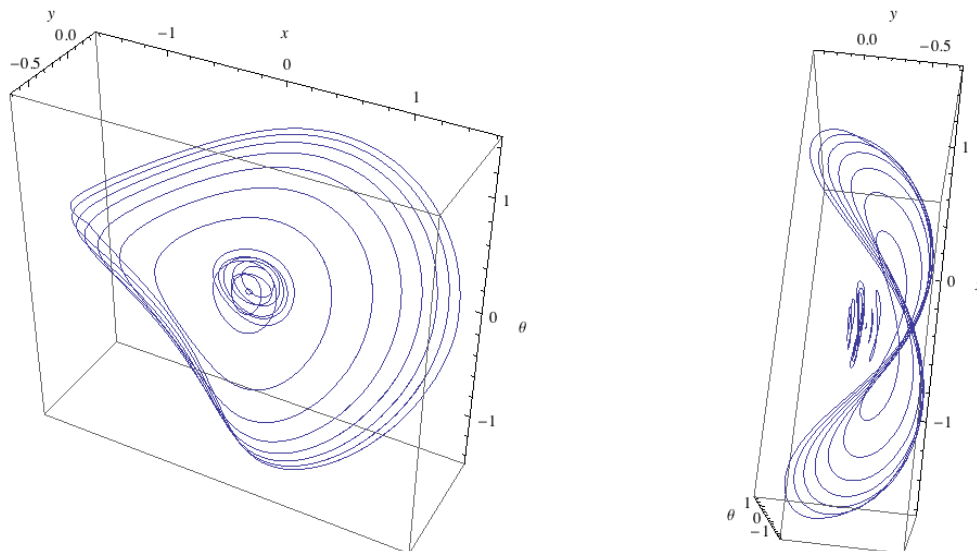


FIGURE 62. Deux points de vue sur les lignes de niveau L_k de l'image $W_{C_2^+, R=\frac{\pi}{2}}$ de C_2^+ dans $W(0, \frac{\pi}{2})$ pour $k = 0.07, 0.1, 0.15, 0.75, 0.2, 0.225, 0.25, 0.275, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.9999$. Partant de $k = 0.9999$, L_k se propage comme une sorte de front parallèle dont la taille décroît *en moyenne* mais en *oscillant* (comme dans la figure 61). Lorsque $k \rightarrow 0^+$, la taille de L_k décroît en moyenne tout en oscillant de plus en plus rapidement. Cela se voit nettement à droite où l'un des L_k est presque réduit à un point.

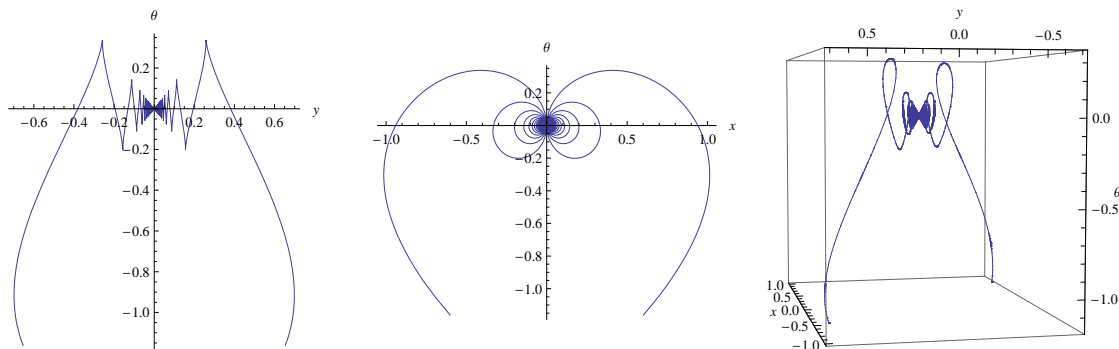


FIGURE 63. Le profil P_0 de l'image $W_{C_2^\pm, R=\frac{\pi}{2}}$ au temps-longueur $R = \frac{\pi}{2}$ de la double strate C_2^\pm pour $\varphi = 0$ et $k \in [0.0001, 0.9999]$. À gauche, la projection $\text{Pr}_{0,(y,\theta)}$ sur le plan (y, θ) . L'axe y est horizontal et l'axe θ vertical. Au milieu, la projection $\text{Pr}_{0,(x,\theta)}$ sur le plan (x, θ) . L'axe x est horizontal et l'axe θ vertical. À droite, une vue 3D de P_0 .

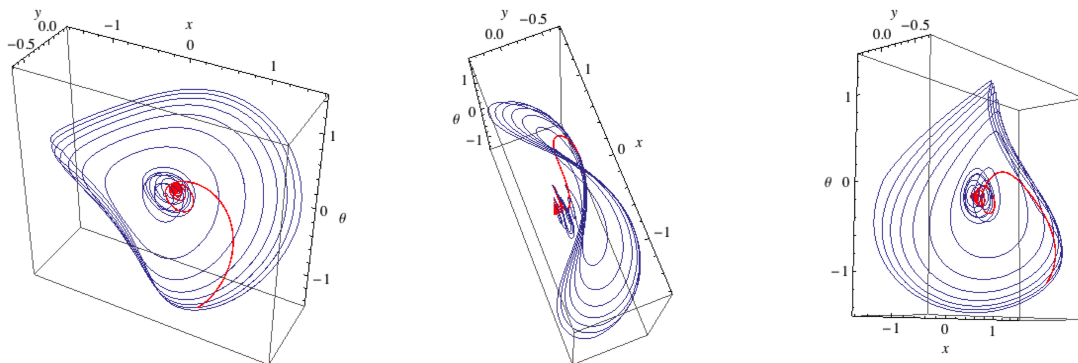


FIGURE 64. En rouge, le “demi-profil” P_0^+ , $\varphi = 0$, de l’image $W_{C_2^+, R=\frac{\pi}{2}}$ de C_2^+ dans $W(0, \frac{\pi}{2})$. Le module k part de $k = 0.9$ et décroît jusqu’à $k = 0.007$. Les trois points de vue montrent comment P_0^+ intersecte les lignes de niveau L_k en spiralant vers 0.

qualitativement le même mais que, au fur et à mesure que φ augmente, P_φ^+ accomplit au départ plus de tours. Les P_φ^+ spiralent vers l’origine mais cette structure générale commune varie beaucoup au niveau du détail.

Pour que le lecteur puisse tester un peu plus quantitativement la variabilité de détail des demi-profils P_φ^+ à l’intérieur de leur structure qualitative commune, la figure 66 montre le tableau de leurs coordonnées x, y pour $\varphi = 0, \frac{\pi}{4}, \frac{\pi}{2}, \pi, 2\pi$.

La figure 67 représente l’image $W_{C_2^+, R=\frac{\pi}{2}}$ de C_2^+ dans $W(0, \frac{\pi}{2})$. Nous avons choisi $k \in (0.07, 0.8)$ pour avoir assez de singularités proches de 0 et une ligne niveau $L_{0.8}$ bien distinguable de l’image de C_3 qui est la limite de L_k pour $k \rightarrow 1^-$. La figure 68 représente un zoom pour $k \in (0.07, 0.26)$.

Enfin la figure 69 représente (sous deux points de vue) la sphère $S(0, \frac{\pi}{2})$ et le front d’onde $W(0, \frac{\pi}{2})$. La borne $k \leq 0.8$, qui laisse un vide autour de l’image de C_3 , ainsi que la transparence des surfaces permettent d’entrevoir la structure interne.

4.7.3. Le cas $R = \frac{3\pi}{2}$.

Lorsque le rayon R de la sphère $S(0, R)$ et du front d’onde $W(0, R)$ augmente, les distorsions par rapport au modèle \mathbb{V}_J du groupe de Heisenberg polarisé deviennent de plus en plus fortes. Avant de présenter la structure du cas $R = \frac{3\pi}{2}$, précisons ce qui se passe en revenant sur l’évolution de l’image $Q_{C_3} = \text{Exp}(C_3)$ de la strate C_3 lorsque t varie que nous avons présentée plus haut à la section 4.4. La figure 70 représente avec un code couleur l’évolution du bord ∂Q_{C_3} de Q_{C_3} au temps-longueur t pour t variant entre $\frac{\pi}{4}$ et 2π . La figure 71 montre les courbes $\varphi = 0$ et les limites $\varphi = \pm\infty$ pour les quatre strates.

Ces très fortes déformations se lisent sur la géométrie de $S(0, R)$ et $W(0, R)$ lorsque le temps-longueur t croît. La figure 72 montre la déformation des images des

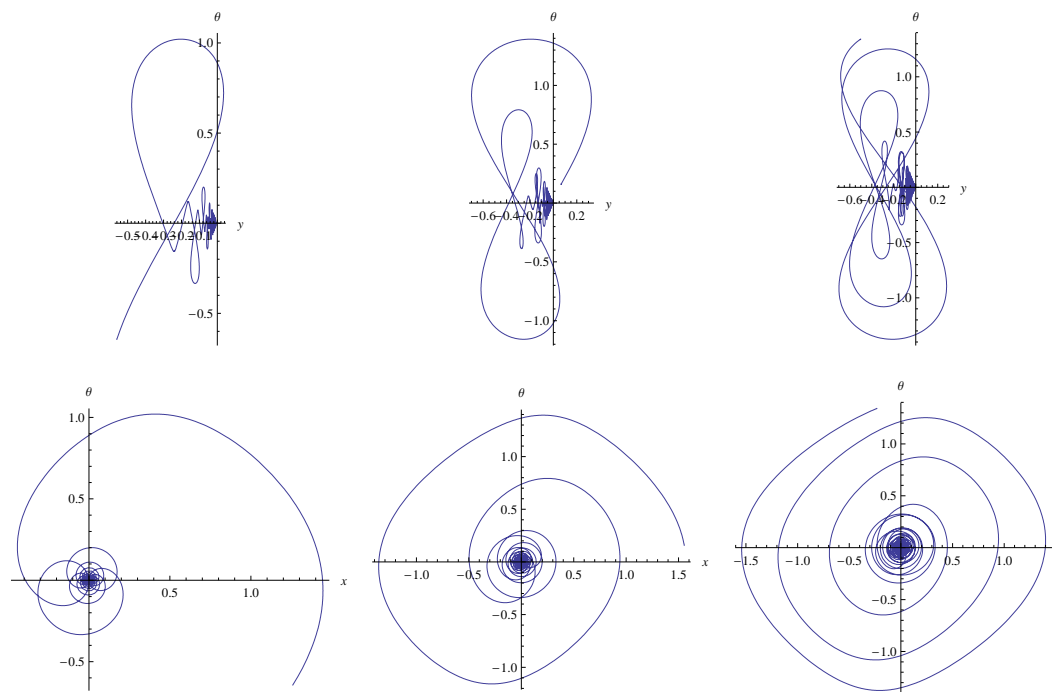


FIGURE 65. De gauche à droite, les demi-profil P_φ^+ de l'image $W_{C_2^+, R=\frac{\pi}{2}}$ de C_2^+ dans $W(0, \frac{\pi}{2})$ pour $\varphi = \frac{\pi}{4}, \pi, 2\pi$. En haut la projection $\text{Pr}_{0,(y,\theta)}$ sur le plan (y, θ) . L'axe y est horizontal et l'axe θ vertical. En bas, la projection $\text{Pr}_{0,(x,\theta)}$ sur le plan (x, θ) . L'axe x est horizontal et l'axe θ vertical.

orbites $k = 0.95, 0.97, 0.99, 0.995, 0.999$ de C_1^0 pour $t = \frac{3\pi}{2}$. Afin d'aider les lecteurs qui ont une faculté limitée de rotation mentale d'images 3D, nous présentons quatre points de vue différents.

La figure 73 montre quatre points de vue sur l'image $W_{C_1^0, R=\frac{3\pi}{2}}$ de la strate C_1^0 au temps-longueur $t = R = \frac{3\pi}{2}$ pour $k \in (0.1, 0.999)$. Les courbes "parallèles" au bord sont les lignes de niveau $k = \text{cste}$ qui se propagent de $(0, 0, -\frac{3\pi}{2})$ jusqu'à l'image des strates $C_3^{0,+} \cup C_3^{0,-}$, et les "rayons" sont les lignes de niveau $\varphi = \text{cste}$.

La figure 74 montre quatre autres points de vue sur l'image plus complète $W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ des strates C_1^0 et $C_1^{2\pi}$ au temps-longueur $t = R = \frac{3\pi}{2}$ pour $k \in [0.1, 0.999]$ ainsi que l'image $W_{C_3, R=\frac{3\pi}{2}}$ des quatre strates C_3 qui les bordent.

Encore plus que dans le cas $R = \frac{\pi}{2}$, les images $W_{C_2^\pm, R=\frac{3\pi}{2}}$ des strates C_2^\pm sont beaucoup plus compliquées avec leur infinité de singularités. La figure 75 montre quelques lignes de niveau L_k pour $k = \text{cste}$ ainsi que la façon dont le demi profil P_0^+ de $W_{C_2^+, R=\frac{3\pi}{2}}$ (en rouge) croise ces lignes de niveau L_k en spiraland vers 0.

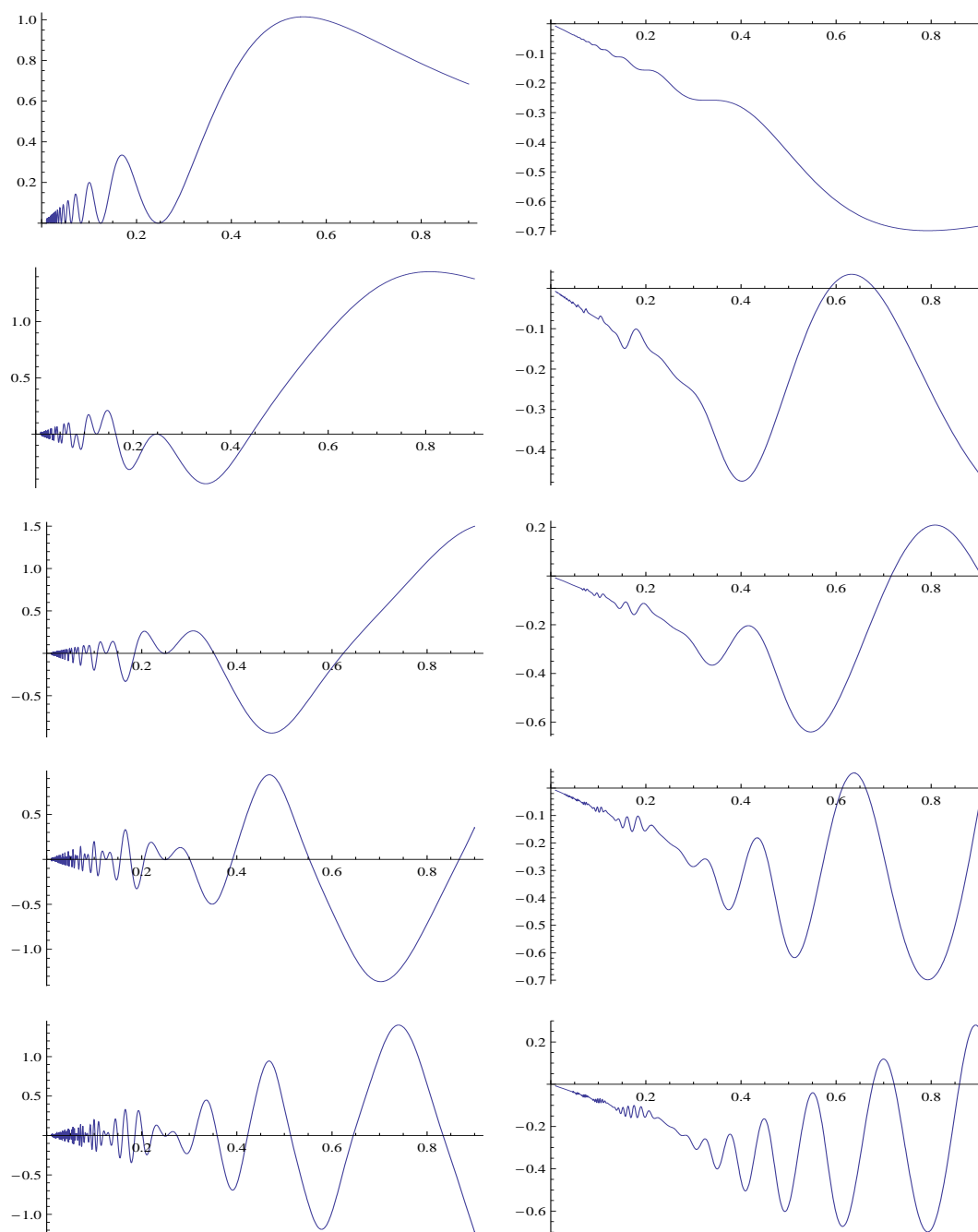


FIGURE 66. Les coordonnées des demi-profil P_{φ}^+ . De gauche à droite x et y . De haut en bas $\varphi = 0, \frac{\pi}{4}, \frac{\pi}{2}, \pi, 2\pi$. On voit bien la variabilité de détail à l'intérieur d'une structure qualitative connue.

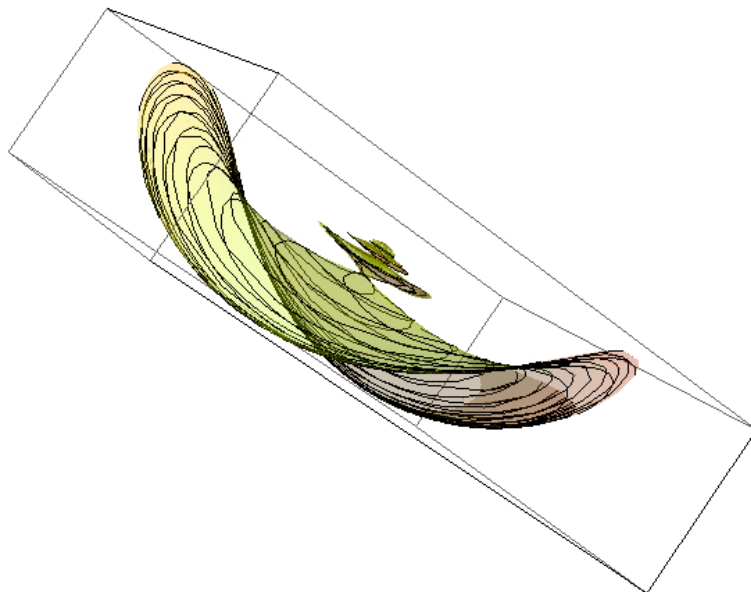


FIGURE 67. L'image $W_{C_2^+, R=\frac{3\pi}{2}}$ de la strate C_2^+ dans $W(0, \frac{\pi}{2})$ pour $k \in (0.07, 0.8)$.

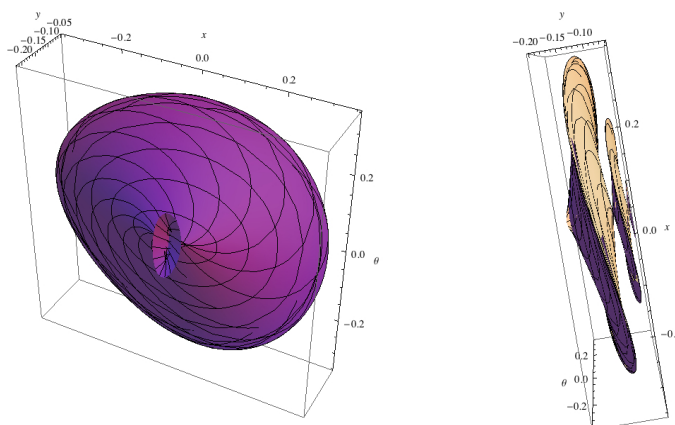


FIGURE 68. Zoom pour $k \in (0.07, 0.26)$ de l'image $W_{C_2^+, R=\frac{3\pi}{2}}$ de la strate C_2^+ dans $W(0, \frac{\pi}{2})$ présentée à la figure 67.

Pour voir comment les demi-profil P_φ^+ de $W_{C_2^+, R=\frac{3\pi}{2}}$ varient avec φ dans le cas $t = R = \frac{3\pi}{2}$, la figure 76 les représente pour $\varphi = 0, \frac{\pi}{4}, \pi, 2\pi$ et $k \in [0.0001, 0.9999]$. Ce sont encore des demi-profil car on ne considère que la strate C_2^+ . On notera

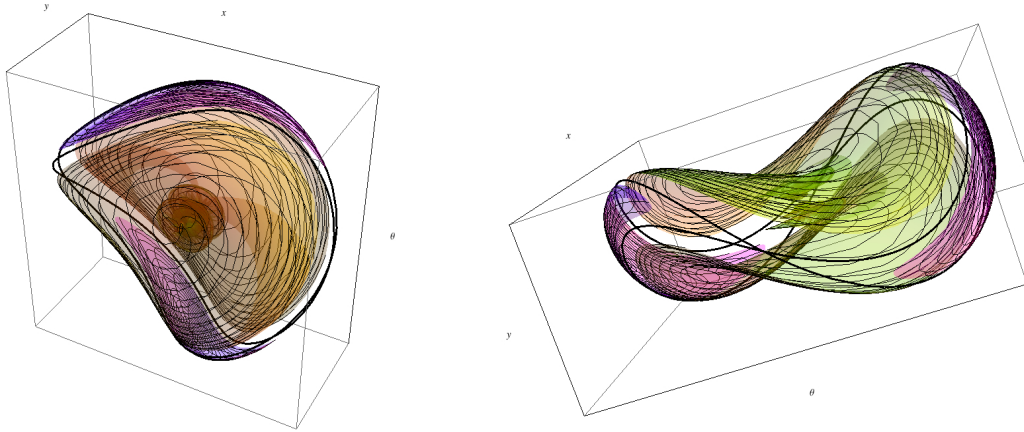


FIGURE 69. Deux points de vue sur la sphère $S(0, \frac{\pi}{2})$ et le front d'onde $W(0, \frac{\pi}{2})$. La borne $k \leq 0.8$ ainsi que la transparence permettent d'entrevoir la structure interne.

les multiples façons dont ils spiralent vers l'origine lorsque $k \rightarrow 0^+$ en suivant la géométrie compliquée de $W_{C_2^+, R=\frac{3\pi}{2}}$.

La figure 77 représente quant à elle les projections $\text{Pr}_{0,(y,\theta)}$ et $\text{Pr}_{0,(x,\theta)}$ sur les plans (y, θ) et (x, θ) du profil P_0 complet de $W_{C_2^\pm, R=\frac{\pi}{2}}$ pour $\varphi = 0$ et $k \in [0.01, 0.99]$. La symétrie $(x, y) \rightarrow (-x, -y)$ entre C_2^+ et C_2^- est claire. P_0 part des points symétriques de C_3

$$\left\{ \pm x = 1 - \frac{1}{\cosh\left(\frac{3\pi}{2}\right)}, \pm y = -\frac{3\pi}{2} + \tanh\left(\frac{3\pi}{2}\right), -\arccos\left(\frac{1}{\cosh\left(\frac{3\pi}{2}\right)}\right) \right\}$$

$$\sim \{0.982, -3.7125, -1.5528\}$$

La figure 78 montre quant à elle les projections d'autres demi-profil P_φ^+ de l'image $W_{C_2^+, R=\frac{3\pi}{2}}$ de C_2^+ pour $\varphi = \frac{\pi}{4}, \pi, 2\pi$. On comparera avec la figure 65.

La figure 79 représente l'image $W_{C_2^+, R=\frac{3\pi}{2}}$ de C_2^+ dans $W(0, \frac{3\pi}{2})$ pour $k \in [0.01, 0.999]$. Comme plus haut pour le rayon $R = \frac{3\pi}{2}$, nous recollons, pour des raisons de routines graphiques deux demi-surfaces correspondant à $k \in [0.01, 0.9]$ et $k \in [0.9, 0.999]$.

La figure 80 montre comment $W_{C_2^+, R=\frac{3\pi}{2}}$ s'emboîte dans le $W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ de la figure 74.

Enfin, les figures 81 et 82 montrent deux perspectives sur le front d'onde complet

$$W\left(0, \frac{3\pi}{2}\right) = W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}} \cup W_{C_2^+, R=\frac{3\pi}{2}} \cup W_{C_2^-, R=\frac{3\pi}{2}} .$$

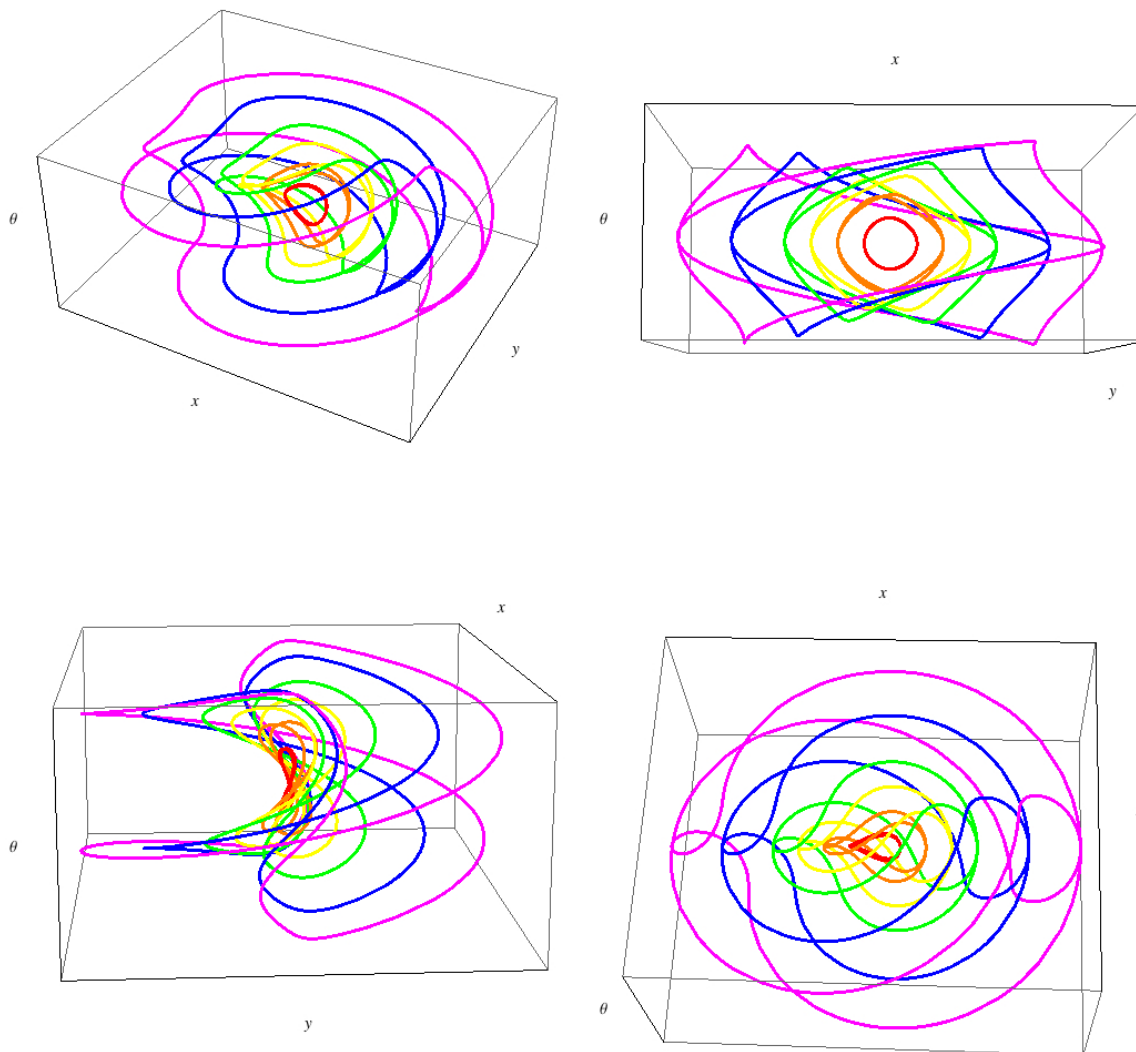


FIGURE 70. L'évolution du bord ∂Q_{C_3} de l'image Q_{C_3} des quatre strates C_3 au temps-longueur t pour t variant entre $\frac{\pi}{4}$ et 2π . Le code couleur est $\frac{\pi}{4}$ (rouge), $\frac{\pi}{2}$ (orange), $\frac{3\pi}{4}$ (jaune), π (vert), $\frac{3\pi}{2}$ (bleu), 2π (magenta). Il y a quatre points de vue.

Dans ces figures, l'angle θ n'est pas représenté modulo 2π . La figure 83 reprend la figure 74 mais représente cette fois $W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ et $W_{C_3, R=\frac{3\pi}{2}}$ modulo 2π . Le cylindre vertical représente le saut de θ de $-\pi$ à π . De même les figures 84, 85 reprennent les figures 81 et 82 du front d'onde complet $W(0, \frac{3\pi}{2})$ mais cette fois modulo 2π .

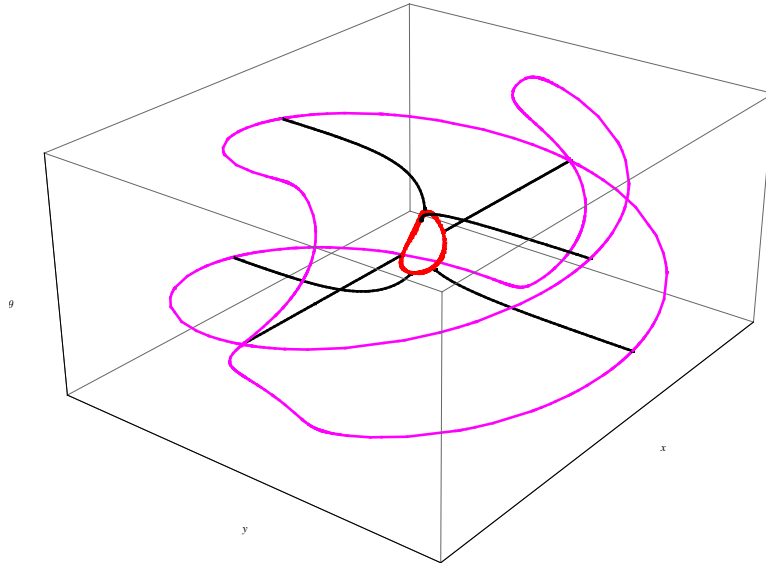


FIGURE 71. La courbe $\varphi = 0$ et les limites $\varphi = \pm\infty$ des quatre strates de Q_{C_3} entre $t = \frac{\pi}{4}$ (rouge) et $t = 2\pi$ (magenta). Les quatre composantes de $\varphi = 0$ sont en trait noir continu et les deux composantes $\varphi = \pm\infty$ sont en trait noir pointillé. On peut facilement suivre les quatre composantes de l'image de C_3 pour $t = 2\pi$ de $\varphi = -\infty$ à $\varphi = 0$ puis de $\varphi = 0$ à $\varphi = \infty$.

4.7.4. *Le cas $R = \frac{3\pi}{2}$: représentations toriques.*

En fait, pour illustrer la forme torique de $W(0, R)$ pour un rayon $R > \pi$, il faut utiliser une représentation torique comme celle introduite à la section 4.4.5 et observer que $W(0, R)$ y accomplit plus d'un tour puisque θ calculé directement débordé l'intervalle $[-\pi, \pi]$. Les nouvelles coordonnées sur l'intérieur du tore $\mathring{\mathbb{D}}_\Lambda \times \mathbb{S}_{2\Lambda}^1 = \mathbb{T}_\Lambda$ ⁴⁰ sont

$$X_{tor} = (2\Lambda + x) \cos(\theta), \quad Y_{tor} = (2\Lambda + x) \sin(\theta), \quad Z_{tor} = y.$$

Pour $W(0, \frac{3\pi}{2})$, $\Lambda = 5$ est idoine.

La figure 86 montre (en magenta) l'image $W_{C_1^0, R=\frac{3\pi}{2}}$ dans le tore $\mathbb{T}_{\Lambda=5}$ de la strate C_1^0 au temps-longueur $t = R = \frac{3\pi}{2}$ pour $k \in [0.01, 0.999]$. La "corne" correspond aux k inférieurs et la "fourche" aux k supérieurs.

La figure 87 montre (en vert) l'image $W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ dans le tore $\mathbb{T}_{\Lambda=5}$ de la strate $C_1^{2\pi}$ au temps-longueur $t = R = \frac{3\pi}{2}$ pour $k \in [0.01, 0.999]$. La "corne" correspond aux k inférieurs et la "fourche" aux k supérieurs.

Quant à la figure 88, elle montre que $W_{C_1^0, R=\frac{3\pi}{2}}$ et $W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ s'intersectent dans le tore $\mathbb{T}_{\Lambda=5}$, ce qui est à l'origine de la forme torique de $W(0, \frac{3\pi}{2})$.

40. Rappelons que $\mathbb{S}_{2\Lambda}^1$ est le cercle de rayon 2Λ (l'"âme" du tore, son "cercle déférent") et que $\mathring{\mathbb{D}}_\Lambda$ est l'intérieur du disque de \mathbb{R}^2 de centre 0 et de rayon Λ .

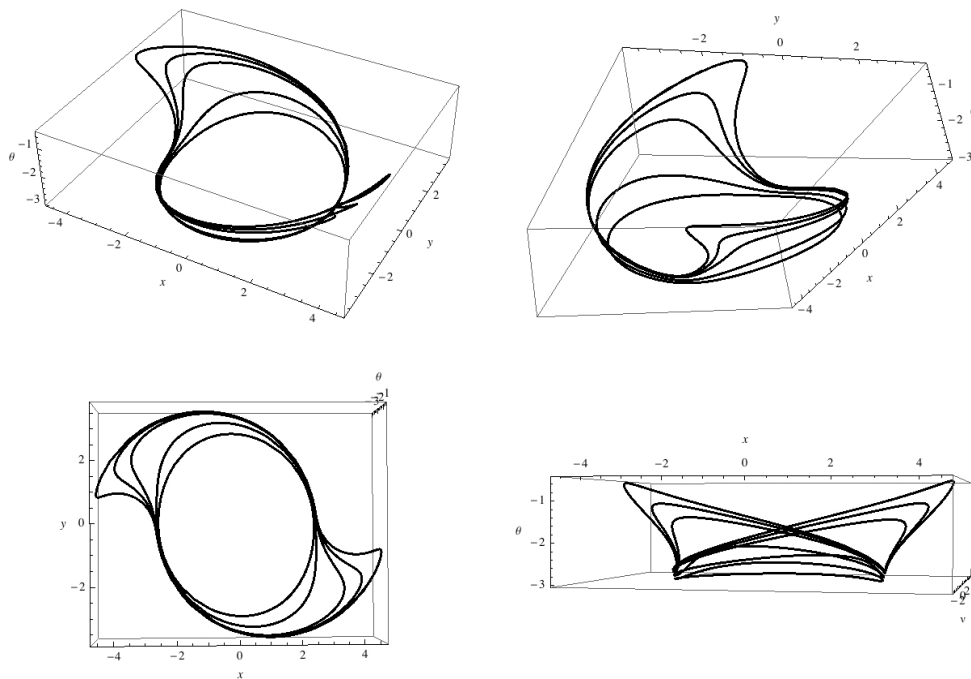


FIGURE 72. Le bord $\partial Q_{C_1^0, k} = W_{C_1^0, k, t = \frac{3\pi}{2}}$ de $Q_{C_1^0, k}$ pour $t = \frac{3\pi}{2}$ et $k = 0.95, 0.97, 0.99, 0.995, 0.999$ (quatre points de vue).

Plus précisément, $W_{C_1^0, R = \frac{3\pi}{2}}$ et $W_{C_1^{2\pi}, R = \frac{3\pi}{2}}$ s'intersectent pour k égal à l'unique solution de $2K(k) = \frac{3\pi}{2}$, $k_0 = 0.91593\dots$ ⁴¹ En effet, pour cette valeur critique on a $\varphi_t = \varphi + 2K$ et donc, puisque $\text{cn}(\varphi + 2K, k) = -\text{cn}(\varphi, k)$, $\text{sn}(\varphi + 2K, k) = -\text{sn}(\varphi, k)$ mais $\text{dn}(\varphi + 2K, k) = \text{dn}(\varphi, k)$, on obtient

41. Comme $K(0) = \frac{\pi}{2}$, le graphe de $2K(k)$ est une courbe croissante commençant à π et c'est donc à partir de $R = \pi$ que $W(0, R)$ commence à avoir une forme torique.

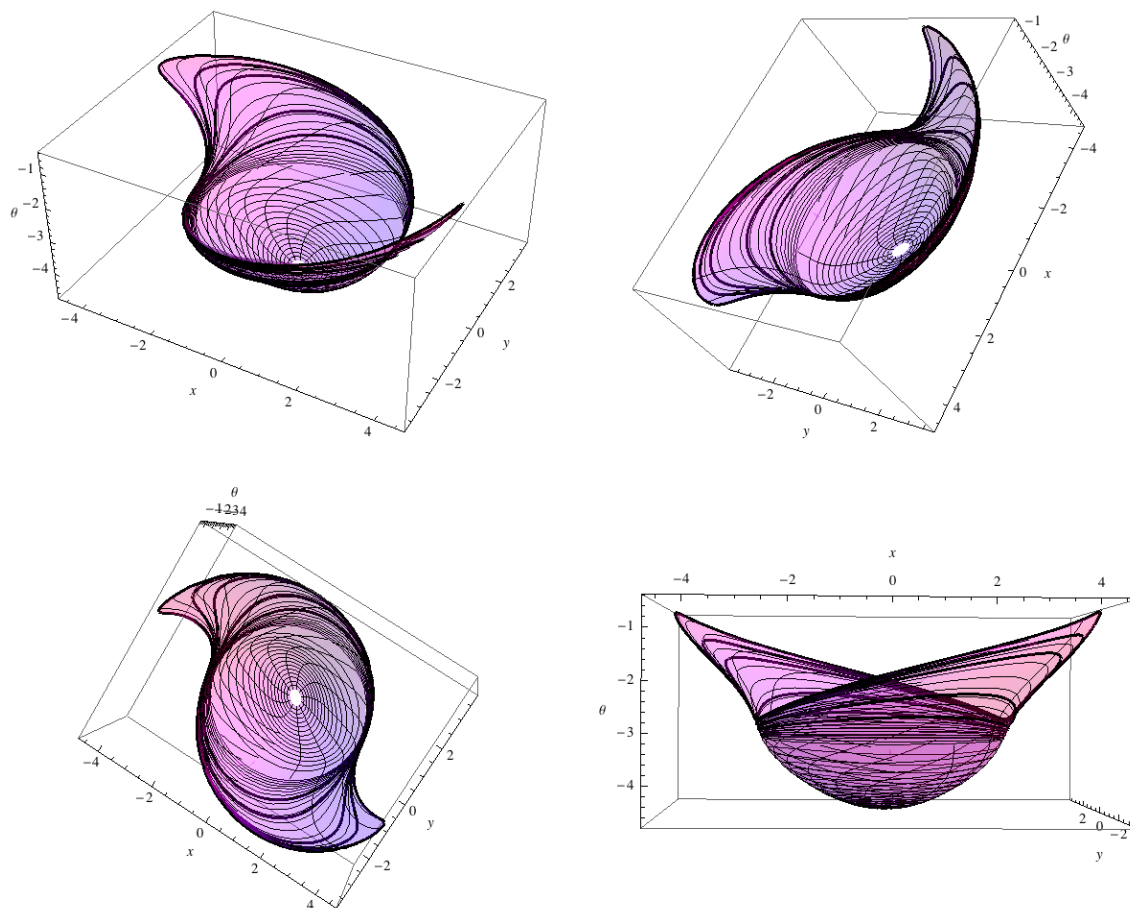


FIGURE 73. Quatre points de vue sur l'image $W_{C_1^0, R=\frac{3\pi}{2}}$ de la strate C_1^0 au temps-longueur $t = R = \frac{3\pi}{2}$ pour $k \in [0.1, 0.999]$. Le fait qu'à partir d'un certain niveau k les lignes de niveau soient plus serrées est lié à la raison suivante. Les niveaux k que *Mathematica* sélectionne par défaut dans l'intervalle $[0.1, 0.999]$ ne sont pas assez serrés pour k proche de 1 et la représentation de la surface devient mauvaise. On pourrait choisir une "MeshFunction" appropriée pour pallier cette dégradation. Mais nous préférons recoller les deux sous-surfaces $k \in [0.1, 0.9]$ et $k \in [0.9, 0.999]$ car cela permet de visualiser directement le niveau intermédiaire 0.9.

(i) pour C_1^0

$$\left\{ \begin{array}{l} \cos\left(\theta_{\frac{3\pi}{2}}\right)^0 = \operatorname{cn}(\varphi, k) \operatorname{cn}(\varphi + 2K, k) + \operatorname{sn}(\varphi, k) \operatorname{sn}(\varphi + 2K, k) \\ \quad = -\operatorname{cn}^2(\varphi, k) - \operatorname{sn}^2(\varphi, k) = -1 \\ \sin\left(\theta_{\frac{3\pi}{2}}\right)^0 = \operatorname{sn}(\varphi, k) \operatorname{cn}(\varphi + 2K, k) - \operatorname{cn}(\varphi, k) \operatorname{sn}(\varphi + 2K, k) \\ \quad = -\operatorname{sn}(\varphi, k) \operatorname{cn}(\varphi, k) + \operatorname{cn}(\varphi, k) \operatorname{sn}(\varphi, k) = 0 \\ x_{\frac{3\pi}{2}}^0 = \frac{1}{k} \left(\begin{array}{l} \operatorname{cn}(\varphi, k) (\operatorname{dn}(\varphi, k) - \operatorname{dn}(\varphi + 2K, k)) \\ + \operatorname{sn}(\varphi, k) \left(\frac{3\pi}{2} + \mathbb{E}(\varphi) - \mathbb{E}(\varphi + 2K)\right) \end{array} \right) \\ \quad = \frac{1}{k} \operatorname{sn}(\varphi, k) \left(\frac{3\pi}{2} - \mathbb{E}(2K)\right) \\ y_{\frac{3\pi}{2}}^0 = \frac{1}{k} \left(\begin{array}{l} \operatorname{sn}(\varphi, k) (\operatorname{dn}(\varphi, k) - \operatorname{dn}(\varphi + 2K, k)) \\ - \operatorname{cn}(\varphi, k) \left(\frac{3\pi}{2} + \mathbb{E}(\varphi) - \mathbb{E}(\varphi + 2K)\right) \end{array} \right) \\ \quad = -\frac{1}{k} \operatorname{cn}(\varphi, k) \left(\frac{3\pi}{2} - \mathbb{E}(2K)\right) \\ X_{tor}^0 = \left(2\Lambda + x_{\frac{3\pi}{2}}^0\right) \cos\left(\theta_{\frac{3\pi}{2}}\right)^0 = -\left(2\Lambda + x_{\frac{3\pi}{2}}^0\right) \\ Y_{tor}^0 = \left(2\Lambda + x_{\frac{3\pi}{2}}^0\right) \sin\left(\theta_{\frac{3\pi}{2}}\right)^0 = 0 \\ Z_{tor}^0 = y_{\frac{3\pi}{2}}^0 \end{array} \right.$$

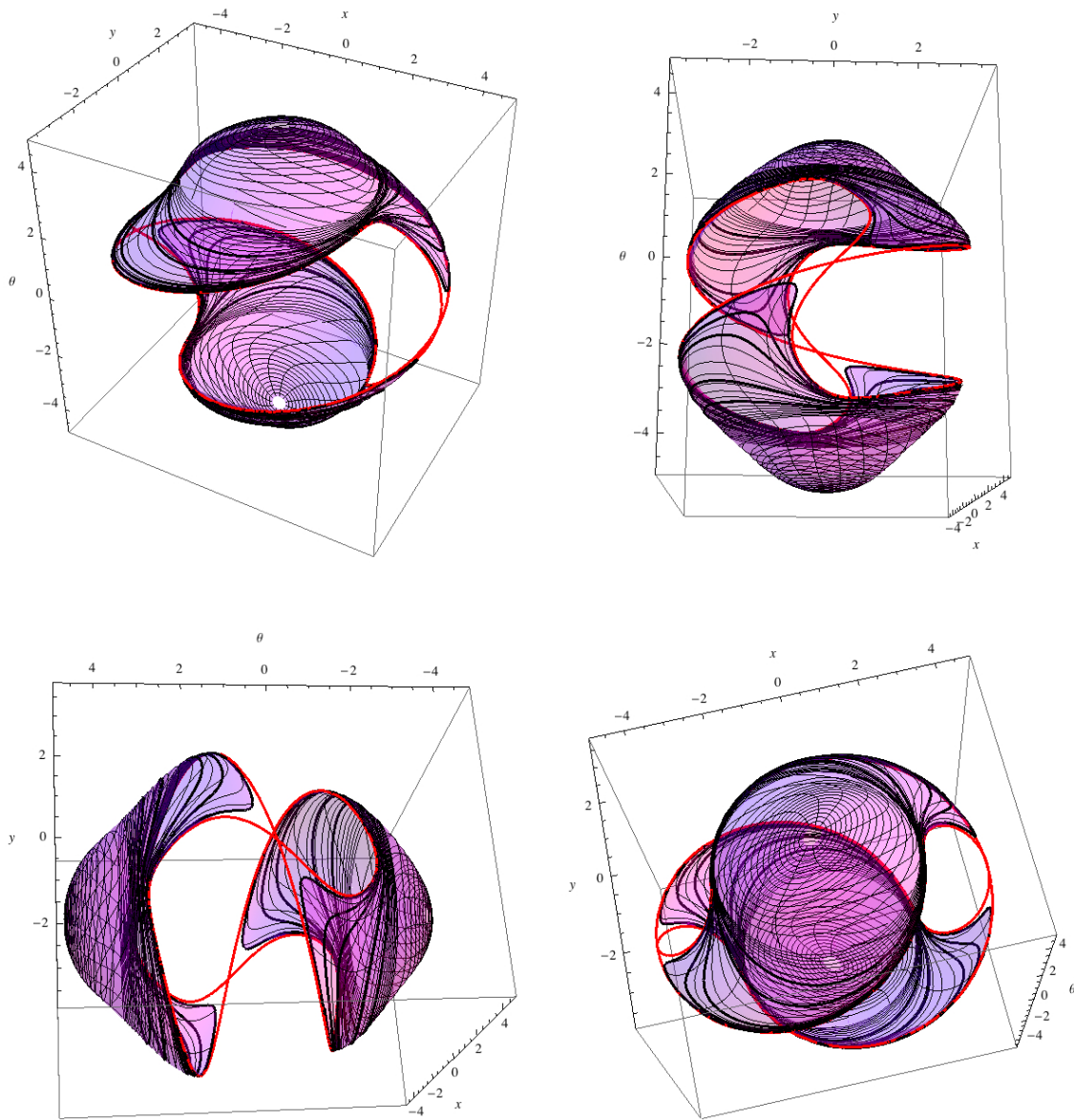


FIGURE 74. Quatre points de vue sur l'image $W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ des strates C_1^0 et $C_1^{2\pi}$ au temps-longueur $t = R = \frac{3\pi}{2}$ pour $k \in [0.1, 0.999]$ ainsi que l'image $W_{C_3, R=\frac{3\pi}{2}}$ des quatre strates C_3 qui les bordent.

(ii) pour $C_1^{2\pi}$ avec la même valeur de φ

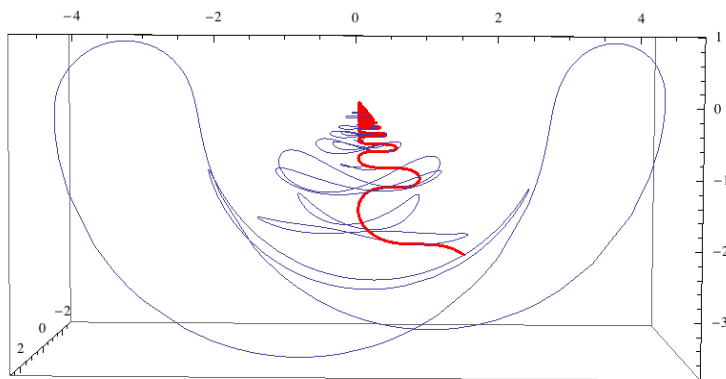


FIGURE 75. Les lignes de niveau L_k de l'image $W_{C_2^+, R=\frac{3\pi}{2}}$ de C_2^+ dans $W(0, \frac{3\pi}{2})$ pour $k = 0.07, 0.1, 0.15, 0.75, 0.2, 0.225, 0.25, 0.275, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.9999$. Pour k assez voisin de 1, L_k se propage comme une sorte de front parallèle. Au contraire, lorsque $k \rightarrow 0^+$, la taille de L_k décroît tout en oscillant de plus en plus rapidement. En rouge, le demi-profil P_0^+ , $\varphi = 0$ qui croise les lignes de niveau L_k en spirant vers 0.

$$\left\{ \begin{array}{l} \cos\left(\theta_{\frac{3\pi}{2}}\right)^{2\pi} = \operatorname{cn}(\varphi, k) \operatorname{cn}(\varphi + 2K, k) + \operatorname{sn}(\varphi, k) \operatorname{sn}(\varphi + 2K, k) = -1 \\ \sin\left(\theta_{\frac{3\pi}{2}}\right)^{2\pi} = -(\operatorname{sn}(\varphi, k) \operatorname{cn}(\varphi + 2K, k) - \operatorname{cn}(\varphi, k) \operatorname{sn}(\varphi + 2K, k)) = 0 \\ x_{\frac{3\pi}{2}}^{2\pi}(\varphi, k) = -\frac{1}{k} \begin{pmatrix} \operatorname{cn}(\varphi, k) (\operatorname{dn}(\varphi, k) - \operatorname{dn}(\varphi + 2K, k)) \\ + \operatorname{sn}(\varphi, k) \left(\frac{3\pi}{2} + \mathbb{E}(\varphi) - \mathbb{E}(\varphi + 2K)\right) \end{pmatrix} \\ \qquad \qquad \qquad = -\frac{1}{k} \operatorname{sn}(\varphi, k) \left(\frac{3\pi}{2} - \mathbb{E}(2K)\right) \\ y_{\frac{3\pi}{2}}^{2\pi}(\varphi, k) = -\frac{1}{k} \operatorname{cn}(\varphi, k) \left(\frac{3\pi}{2} - \mathbb{E}(2K)\right) \\ X_{tor}^{2\pi}(\varphi, k) = \left(2\Lambda + x_{\frac{3\pi}{2}}^{2\pi}(\varphi, k)\right) \cos\left(\theta_{\frac{3\pi}{2}}\right)^{2\pi} \\ Y_{tor}^{2\pi}(\varphi, k) = \left(2\Lambda + x_{\frac{3\pi}{2}}^{2\pi}(\varphi, k)\right) \sin\left(\theta_{\frac{3\pi}{2}}\right)^{2\pi} = 0 \\ Z_{tor}^{2\pi}(\varphi, k) = y_{\frac{3\pi}{2}}^{2\pi}(\varphi, k) \end{array} \right.$$

On voit donc que, sn étant une fonction impaire et cn une fonction paire, si l'on change φ en $-\varphi$ alors

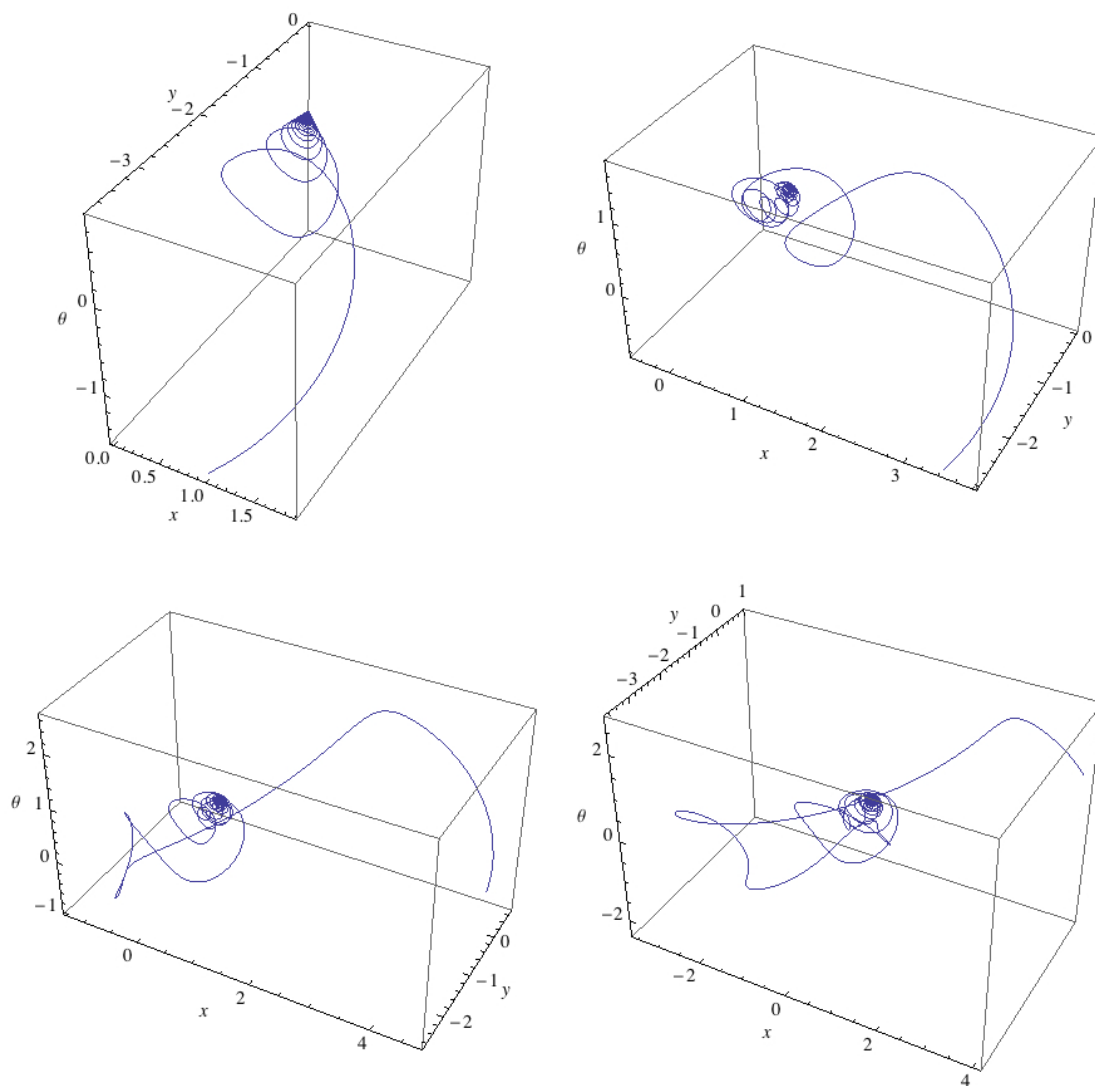


FIGURE 76. Les demi-profils P_φ^+ de $W_{C_2^+, R=\frac{3\pi}{2}}$ pour $\varphi = 0, \frac{\pi}{4}, \pi, 2\pi$ et $k \in [0.0001, 0.9999]$.

$$\left\{ \begin{array}{l} x_{\frac{3\pi}{2}}^{2\pi}(-\varphi, k) = -\frac{1}{k} \operatorname{sn}(-\varphi, k) \left(\frac{3\pi}{2} - \mathbb{E}(2K)\right) = \frac{1}{k} \operatorname{sn}(\varphi, k) \left(\frac{3\pi}{2} - \mathbb{E}(2K)\right) \\ \quad = x_{\frac{3\pi}{2}}^0(\varphi, k) \\ y_{\frac{3\pi}{2}}^{2\pi}(-\varphi, k) = -\frac{1}{k} \operatorname{cn}(-\varphi, k) \left(\frac{3\pi}{2} - \mathbb{E}(2K)\right) = -\frac{1}{k} \operatorname{cn}(\varphi, k) \left(\frac{3\pi}{2} - \mathbb{E}(2K)\right) \\ \quad = y_{\frac{3\pi}{2}}^0(\varphi, k) \\ X_{tor}^{2\pi}(-\varphi, k) = \left(2\Lambda + x_{\frac{3\pi}{2}}^{2\pi}(-\varphi, k)\right) \cos\left(\theta_{\frac{3\pi}{2}}\right)^{2\pi} = -\left(2\Lambda + x_{\frac{3\pi}{2}}^{2\pi}(-\varphi, k)\right) \\ \quad = -\left(2\Lambda + x_{\frac{3\pi}{2}}^0(\varphi, k)\right) = X_{tor}^0(\varphi, k) \\ Y_{tor}^{2\pi}(-\varphi, k) = \left(2\Lambda + x_{\frac{3\pi}{2}}^{2\pi}(-\varphi, k)\right) \sin\left(\theta_{\frac{3\pi}{2}}\right)^{2\pi} = 0 = Y_{tor}^0(\varphi, k) \\ Z_{tor}^{2\pi}(-\varphi, k) = y_{\frac{3\pi}{2}}^{2\pi}(-\varphi, k) = y_{\frac{3\pi}{2}}^0(\varphi, k) = Z_{tor}^0(\varphi, k) \end{array} \right.$$

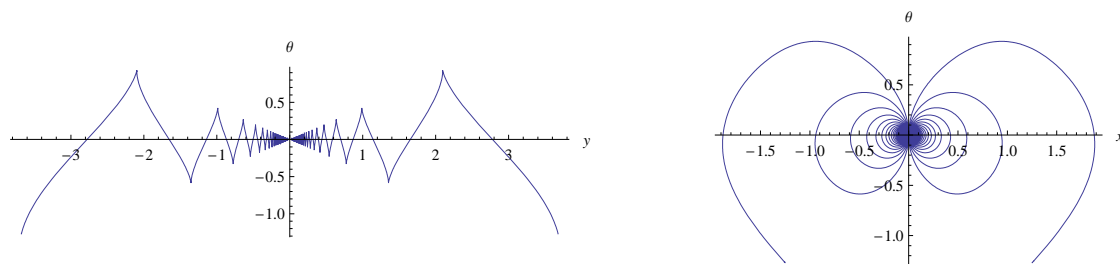


FIGURE 77. Le profil complet P_0 de l'image $W_{C_2^\pm, R=\frac{3\pi}{2}}$ de la double strate C_2^\pm pour $\varphi = 0$ et $k \in [0.01, 0.99]$. Pour $k = 0.99$, P_0 part de $\{\pm x \sim 1.2515, \pm y \sim -3.6777, \theta \sim -1.2691\}$. À gauche, la projection $\text{Pr}_{0,(y,\theta)}$ sur le plan (y, θ) . L'axe y est horizontal et l'axe θ vertical. À droite, la projection $\text{Pr}_{0,(x,\theta)}$ sur le plan (x, θ) .

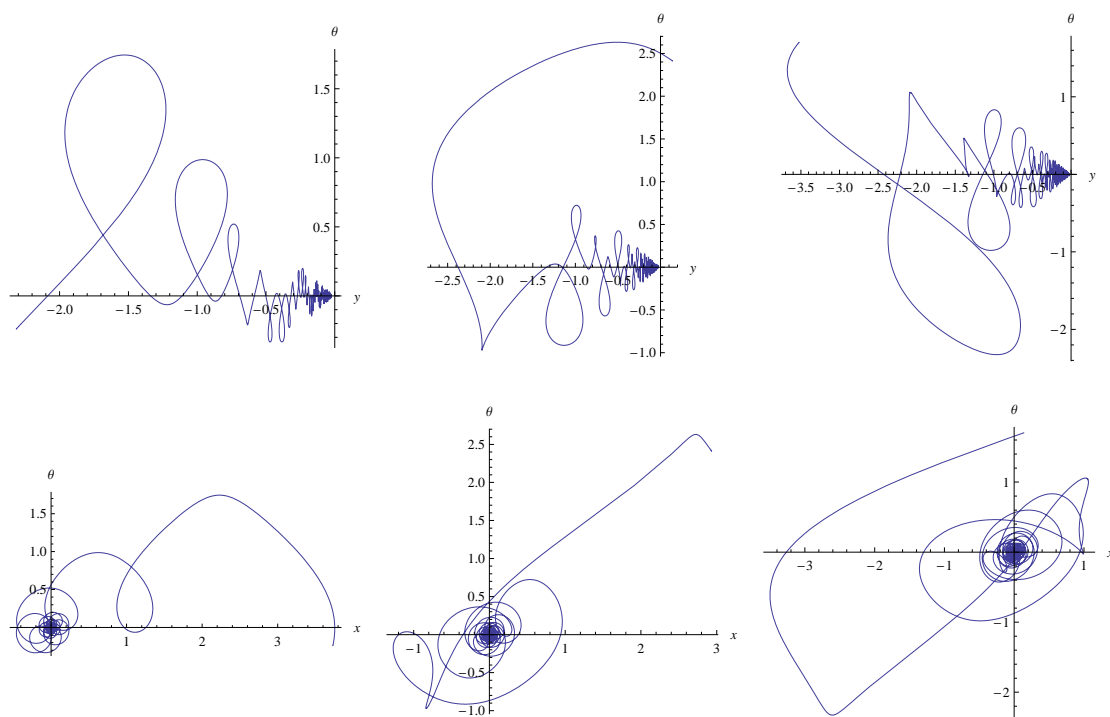


FIGURE 78. De gauche à droite, les profils P_φ^+ de l'image $W_{C_2^+, R=\frac{3\pi}{2}}$ de C_2^+ dans $W(0, \frac{3\pi}{2})$ pour $\varphi = \frac{\pi}{4}, \pi, 2\pi$. En haut la projection $\text{Pr}_{0,(y,\theta)}$ sur le plan (y, θ) . L'axe y est horizontal et l'axe θ vertical. En bas, la projection $\text{Pr}_{0,(x,\theta)}$ sur le plan (x, θ) . L'axe x est horizontal et l'axe θ vertical. À comparer avec la figure 65.

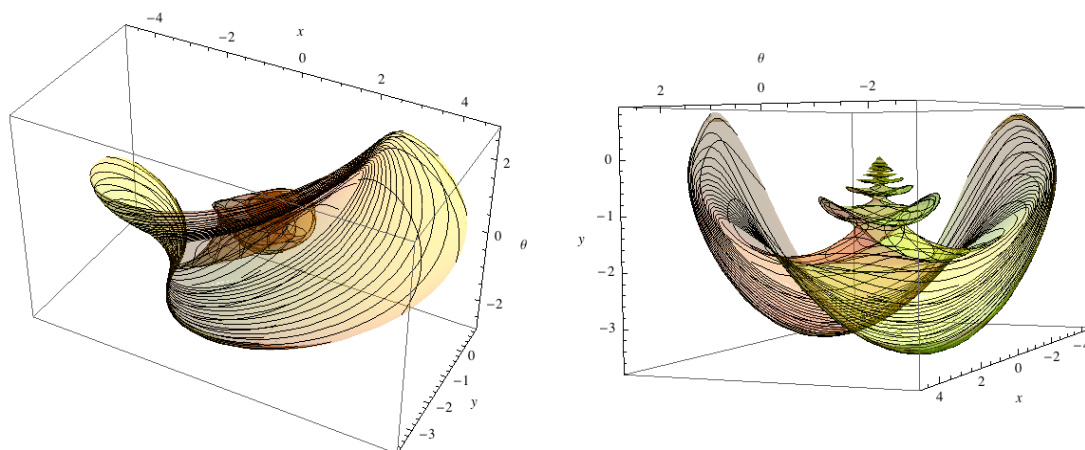


FIGURE 79. Deux points de vue sur l'image $W_{C_2^+, R=\frac{3\pi}{2}}$ de la strate C_2^+ dans $W(0, \frac{3\pi}{2})$ pour $k \in [0.01, 0.999]$.

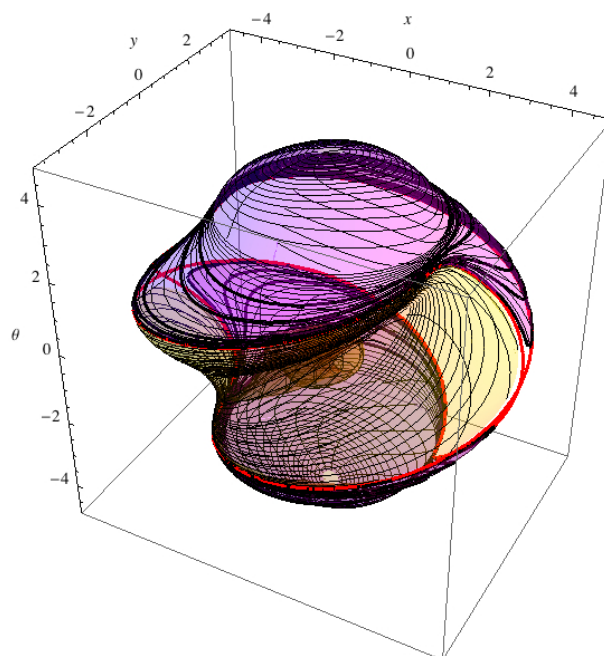


FIGURE 80. L'image $W_{C_2^+, R=\frac{3\pi}{2}}$ de la strate C_2^+ dans $W(0, \frac{3\pi}{2})$ emboîtée dans le $W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ de la figure 74.

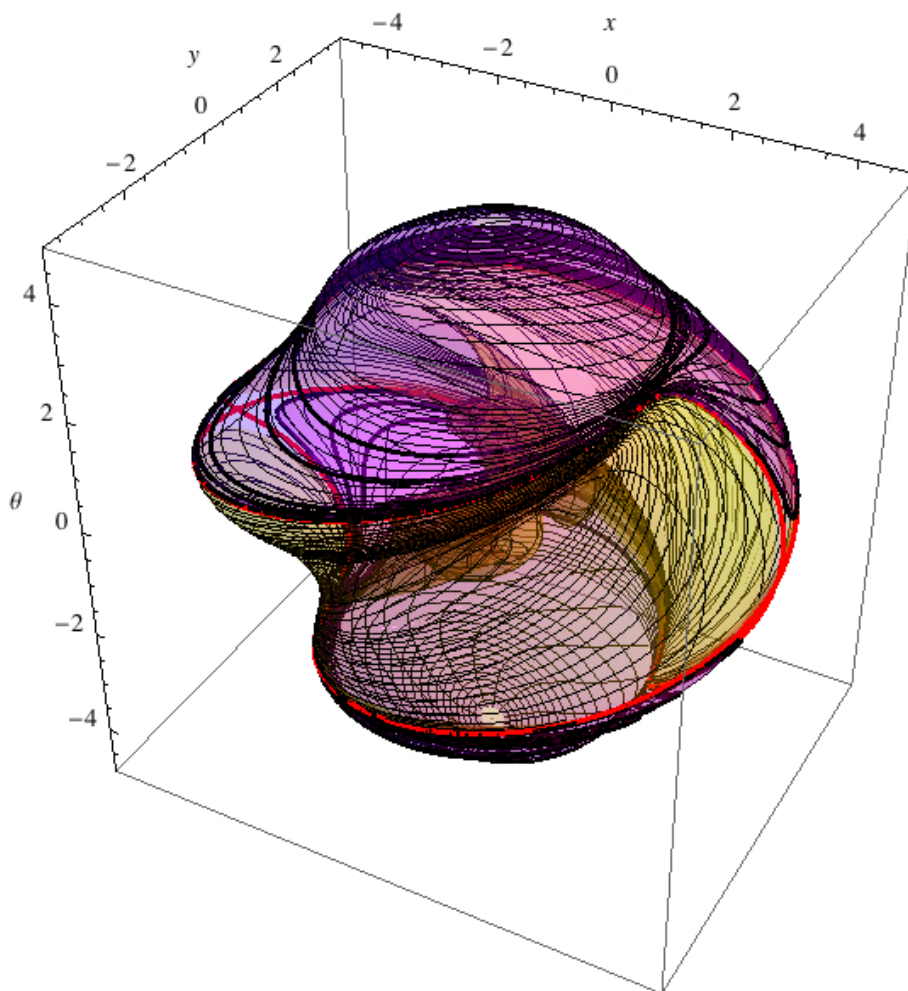


FIGURE 81. Une perspective sur le front d'onde complet $W(0, \frac{3\pi}{2}) = W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}} \cup W_{C_2^+, R=\frac{3\pi}{2}} \cup W_{C_2^-, R=\frac{3\pi}{2}}$.

Les points de l'intersection $W_{C_1^0, R=\frac{3\pi}{2}} \cap W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ sont donc situés, conformément à ce que nous avons vu à la section 4.5, sur *l'ensemble de Maxwell* puisque les géodésiques partant de l'origine avec les deux covecteurs différents $(k_0, \varphi) \in C_1^0$ et $(k_0, -\varphi) \in C_1^{2\pi}$ les atteignent au même temps-longueur $t = \frac{3\pi}{2}$. Nous avons vu dans cette section que le temps de coupure t_{cut} est $2K(k)$, ce qui correspond bien ici à $2K(k_0) = \frac{3\pi}{2}$ par définition de k_0 . La figure 89 montre, dans la représentation torique, les deux géodésiques partant de l'origine 0 avec les covecteurs $(k_0, \varphi = \frac{\pi}{4}) \in C_1^0$ et $(k_0, -\varphi = -\frac{\pi}{4}) \in C_1^{2\pi}$ et arrivant en $t = \frac{3\pi}{2}$ au même point $\{-11.7472, 0, -1.96668\}$ de l'ensemble de Maxwell $W_{C_1^0, R=\frac{3\pi}{2}} \cap W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$.

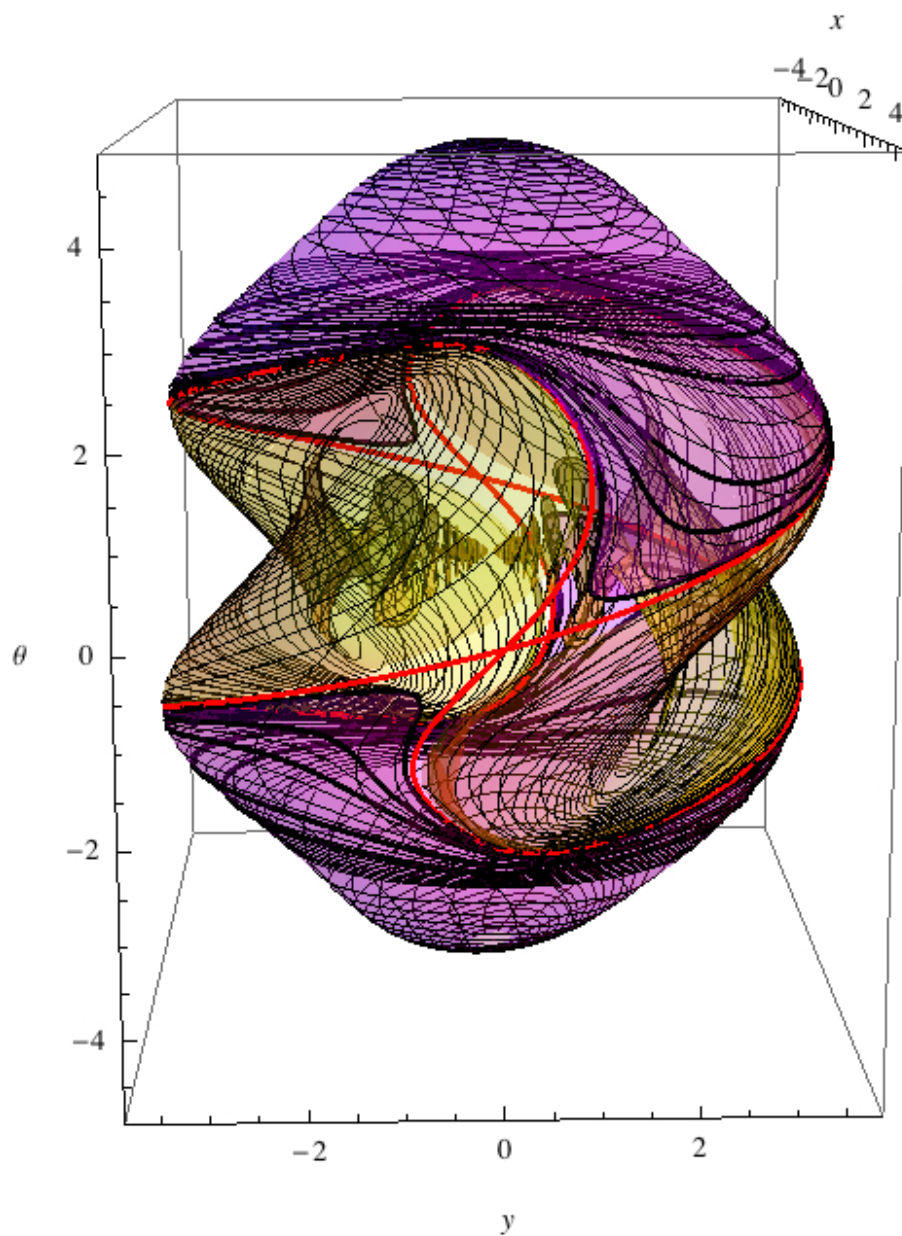


FIGURE 82. Une autre perspective sur le front d'onde complet $W(0, \frac{3\pi}{2}) = W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}} \cup W_{C_2^+, R=\frac{3\pi}{2}} \cup W_{C_2^-, R=\frac{3\pi}{2}}$.

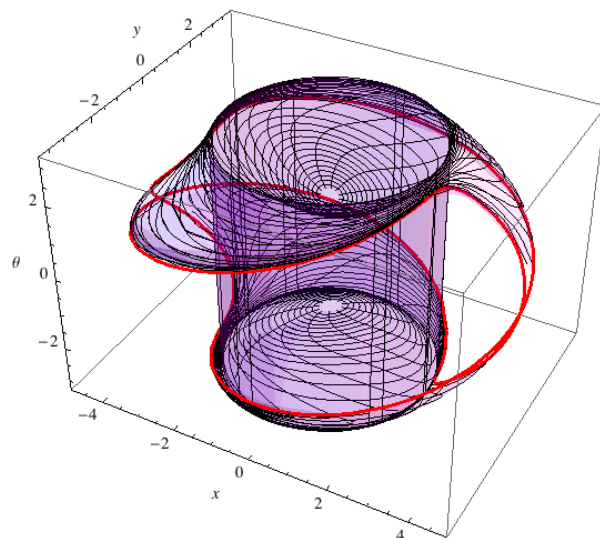


FIGURE 83. Les strates $W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ et $W_{C_3, R=\frac{3\pi}{2}}$ modulo 2π . Le cylindre vertical représente le saut de θ de $-\pi$ à π .

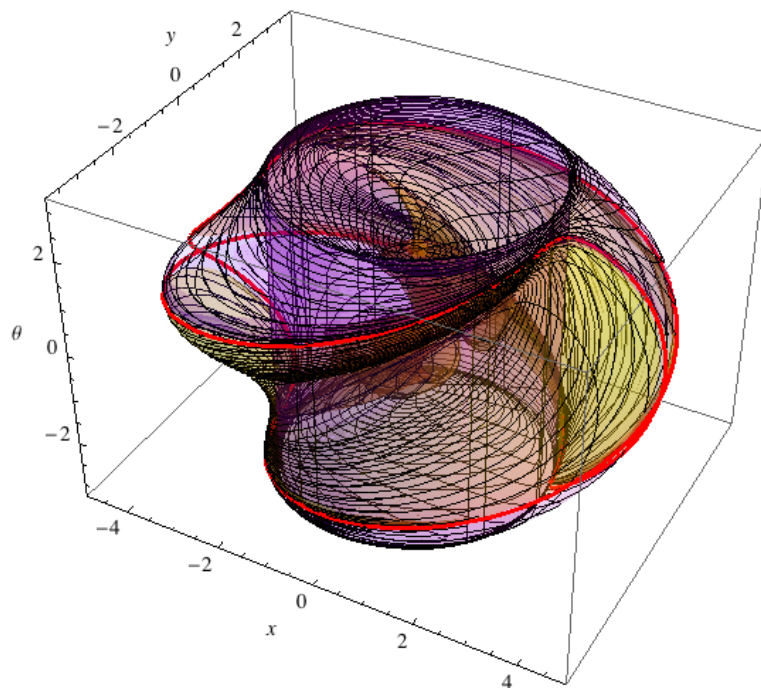


FIGURE 84. Le front d'onde complet $W(0, \frac{3\pi}{2})$ modulo 2π . Le cylindre vertical représente le saut de θ de $-\pi$ à π .

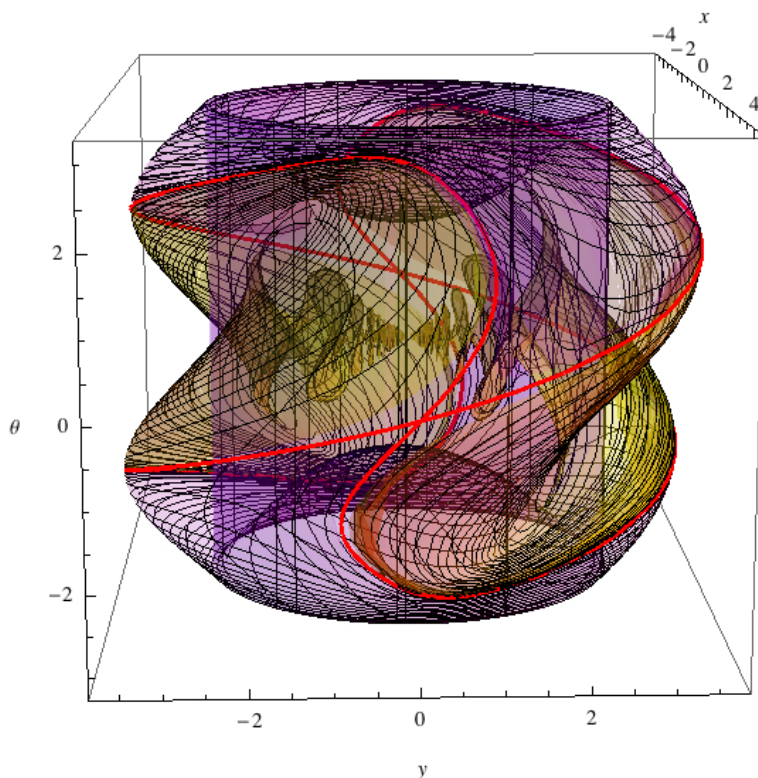


FIGURE 85. De nouveau le front d'onde complet $W(0, \frac{3\pi}{2})$ modulo 2π . Le cylindre vertical représente le saut de θ de $-\pi$ à π .

En ce qui concerne les images dans le tore $\mathbb{T}_{\Lambda=5}$ des strates C_2 , les figures 90 et 91 montrent $W_{C_2^+, R=\frac{3\pi}{2}}$ et $W_{C_2^-, R=\frac{3\pi}{2}}$ tandis que la figure 92 montre leur union $W_{C_2^+, R=\frac{3\pi}{2}} \cup W_{C_2^-, R=\frac{3\pi}{2}}$. Pour alléger ces figures en éliminant la convergence de la "pile de singularités" vers l'origine, nous prenons $k \in [0.5, 0.999]$. Comme pour le rayon $R = \frac{3\pi}{2}$ la première ligne de niveau dégénérée correspond à $k_{\frac{3\pi}{2}} \sim 0.6551$ et la seconde à $k_{c, \frac{3\pi}{2}} \sim 0.8272$, $k \geq 0.5$ suffit pour avoir les premiers "disques" de $W(0, \frac{3\pi}{2}) - S(0, \frac{3\pi}{2})$.

La figure 93 montre le front d'onde complet

$$W\left(0, \frac{3\pi}{2}\right) = W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}} \cup W_{C_2^+, R=\frac{3\pi}{2}} \cup W_{C_2^-, R=\frac{3\pi}{2}}$$

dans le tore $\mathbb{T}_{\Lambda=5}$ et la figure 94 en montre les composantes de profil et de haut.

Enfin, la figure 95 représente dans le tore $\mathbb{T}_{\Lambda=5}$ les strates W_{C_3} et, avec la figure 96, la façon dont elles bordent les strates W_{C_1} et W_{C_2} .

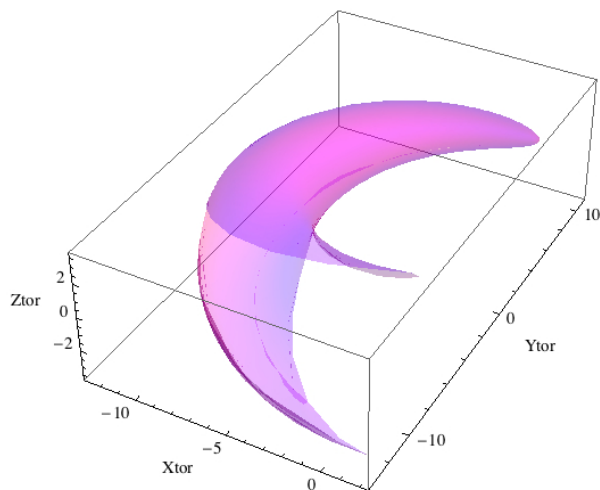


FIGURE 86. Image $W_{C_1^0, R=\frac{3\pi}{2}}$ dans le tore $\mathbb{T}_{\Lambda=5}$ de la strate C_1^0 au temps-longueur $t = R = \frac{3\pi}{2}$ pour $k \in [0.01, 0.999]$. La “corne” correspond aux k inférieurs et la “fourche” aux k supérieurs.

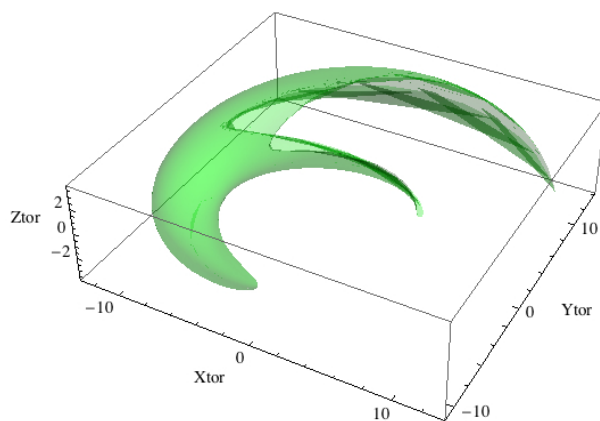


FIGURE 87. Image $W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ dans le tore $\mathbb{T}_{\Lambda=5}$ de la strate $C_1^{2\pi}$ au temps-longueur $t = R = \frac{3\pi}{2}$ pour $k \in [0.01, 0.999]$. La “corne” correspond aux k inférieurs et la “fourche” aux k supérieurs.

4.8. Non commutativité et complexité

Le lecteur peu versé dans les équations différentielles sera peut-être surpris de la complexité des formules régissant la géométrie de $SE(2)$. Elles sont pourtant inhérentes à un problème non seulement très simple mais relevant de la géométrie la plus élémentaire, celle du plan euclidien. Le groupe des déplacements $SE(2)$ est beaucoup plus “simple” (au sens non mathématique) que le groupe des rotations

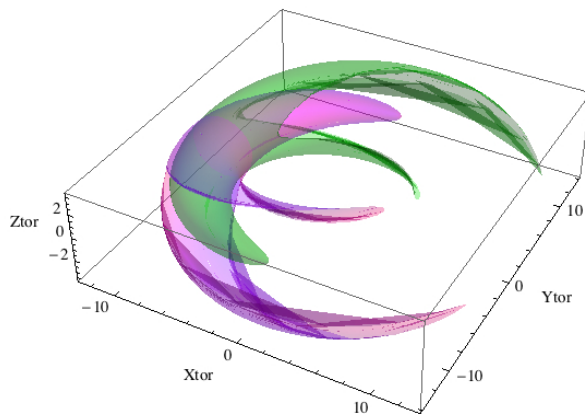


FIGURE 88. $W_{C_1^0, R=\frac{3\pi}{2}}$ et $W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ s'intersectent dans le tore $\mathbb{T}_{\Lambda=5}$.

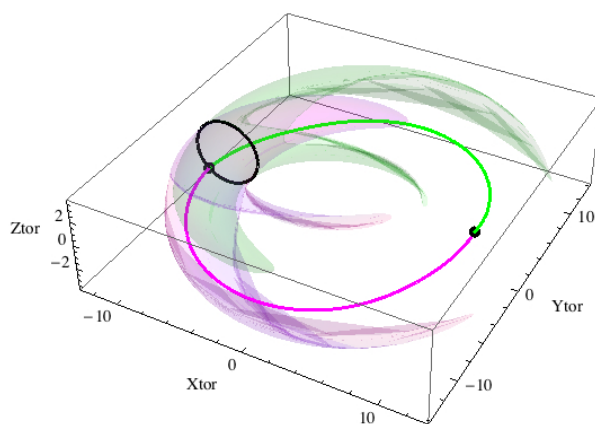


FIGURE 89. Les deux géodésiques partant de l'origine 0 (point noir côté droit) avec les covecteurs $(k_0, \varphi = \frac{\pi}{4}) \in C_1^0$ (magenta) et $(k_0, -\varphi = -\frac{\pi}{4}) \in C_1^{2\pi}$ (vert) et arrivant en $t = \frac{3\pi}{2}$ au même point $\{-11.7472, 0, -1.96668\}$ de l'ensemble de Maxwell $W_{C_1^0, R=\frac{3\pi}{2}} \cap W_{C_1^{2\pi}, R=\frac{3\pi}{2}}$ (point noir du méridien noir côté gauche).

$SO(3)$ de \mathbb{R}^3 . D'où peut provenir une telle complexité immanente à une telle simplicité? Elle est due à la *non-commutativité*.

Revenons sur la remarque du début de la section 4.1 et le système d'équations différentielles pour les λ_i (avec $\lambda'_3 = -\lambda_3$) qui est celui des fonctions elliptiques de Jacobi :

$$\begin{cases} \dot{\lambda}_1 = \{\lambda_1, H\} = \lambda_2 \lambda'_3 \\ \dot{\lambda}_2 = \{\lambda_2, H\} = -\lambda_1 \lambda'_3 \\ \dot{\lambda}'_3 = \{\lambda'_3, H\} = -\lambda_2 \lambda_1 \end{cases}$$

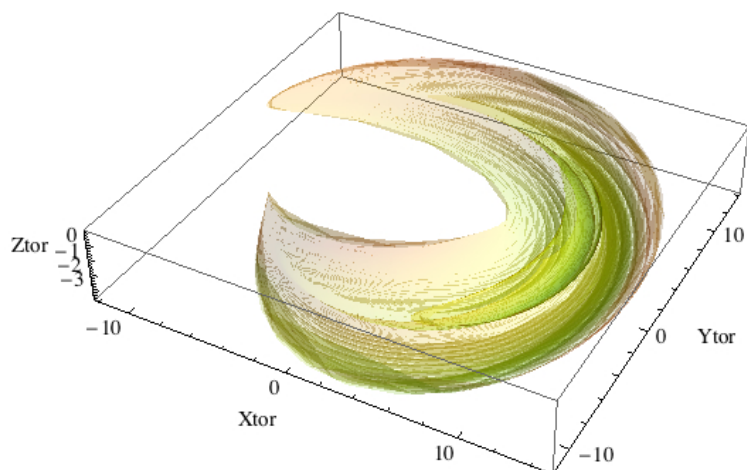


FIGURE 90. $W_{C_2^+, R=\frac{3\pi}{2}}$ dans le tore $\mathbb{T}_{\Lambda=5}$ pour $k \in [0.5, 0.999]$.

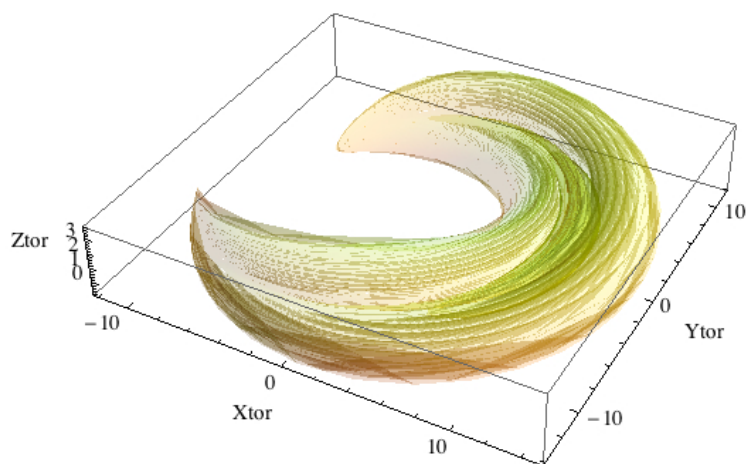


FIGURE 91. $W_{C_2^-, R=\frac{3\pi}{2}}$ dans le tore $\mathbb{T}_{\Lambda=5}$ pour $k \in [0.5, 0.999]$.

À partir de ces équations différentielles très simples, il est trivial de calculer des développements limités pour les $\lambda_i(t)$. On pose

$$\begin{aligned}\lambda_1(t) &= \alpha + \sum_{j \geq 1} a_j t^j, \\ \lambda_2(t) &= \beta + \sum_{j \geq 1} b_j t^j, \\ \lambda_3'(t) &= \gamma + \sum_{j \geq 1} c_j t^j,\end{aligned}$$

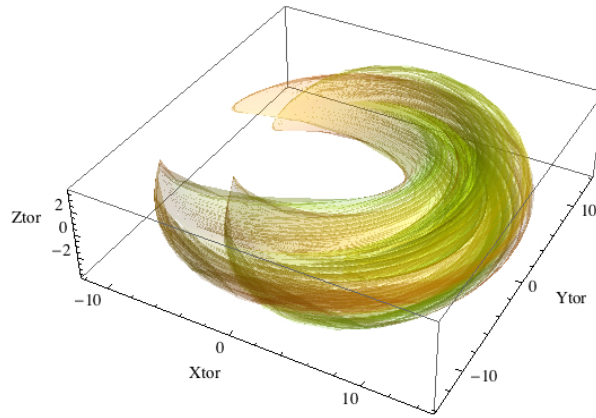


FIGURE 92. L'union $W_{C_2^+, R=\frac{3\pi}{2}} \cup W_{C_2^-, R=\frac{3\pi}{2}}$ dans le tore $\mathbb{T}_{\Lambda=5}$ pour $k \in [0.5, 0.999]$.

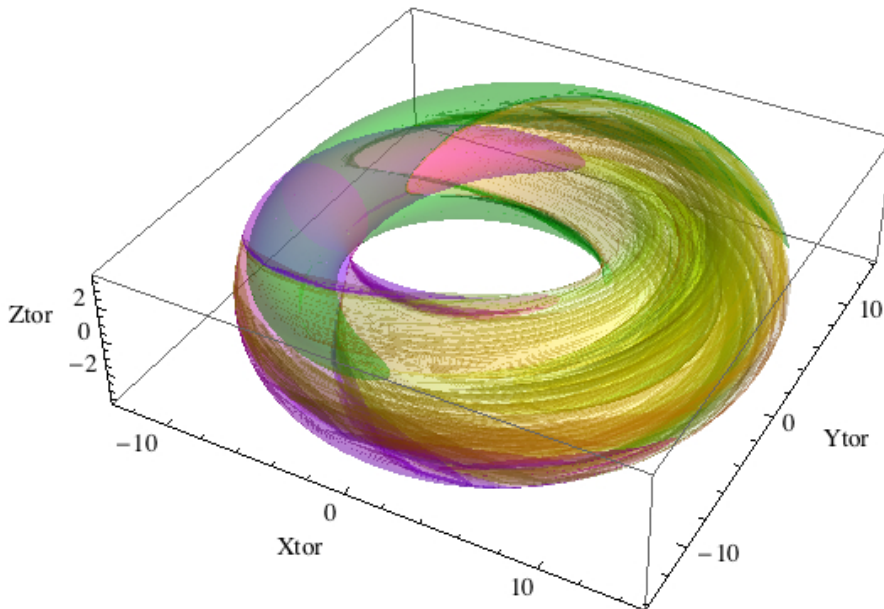


FIGURE 93. Le front d'onde complet $W\left(0, \frac{3\pi}{2}\right) = W_{C_1^0, R=\frac{3\pi}{2}} \cup W_{C_1^{2\pi}, R=\frac{3\pi}{2}} \cup W_{C_2^+, R=\frac{3\pi}{2}} \cup W_{C_2^-, R=\frac{3\pi}{2}}$ dans le tore $\mathbb{T}_{\Lambda=5}$.

on insère ces valeurs dans les équations et on égale les coefficients des t^j . On obtient ainsi itérativement les coefficients. Par exemple à l'ordre 4 on obtient :

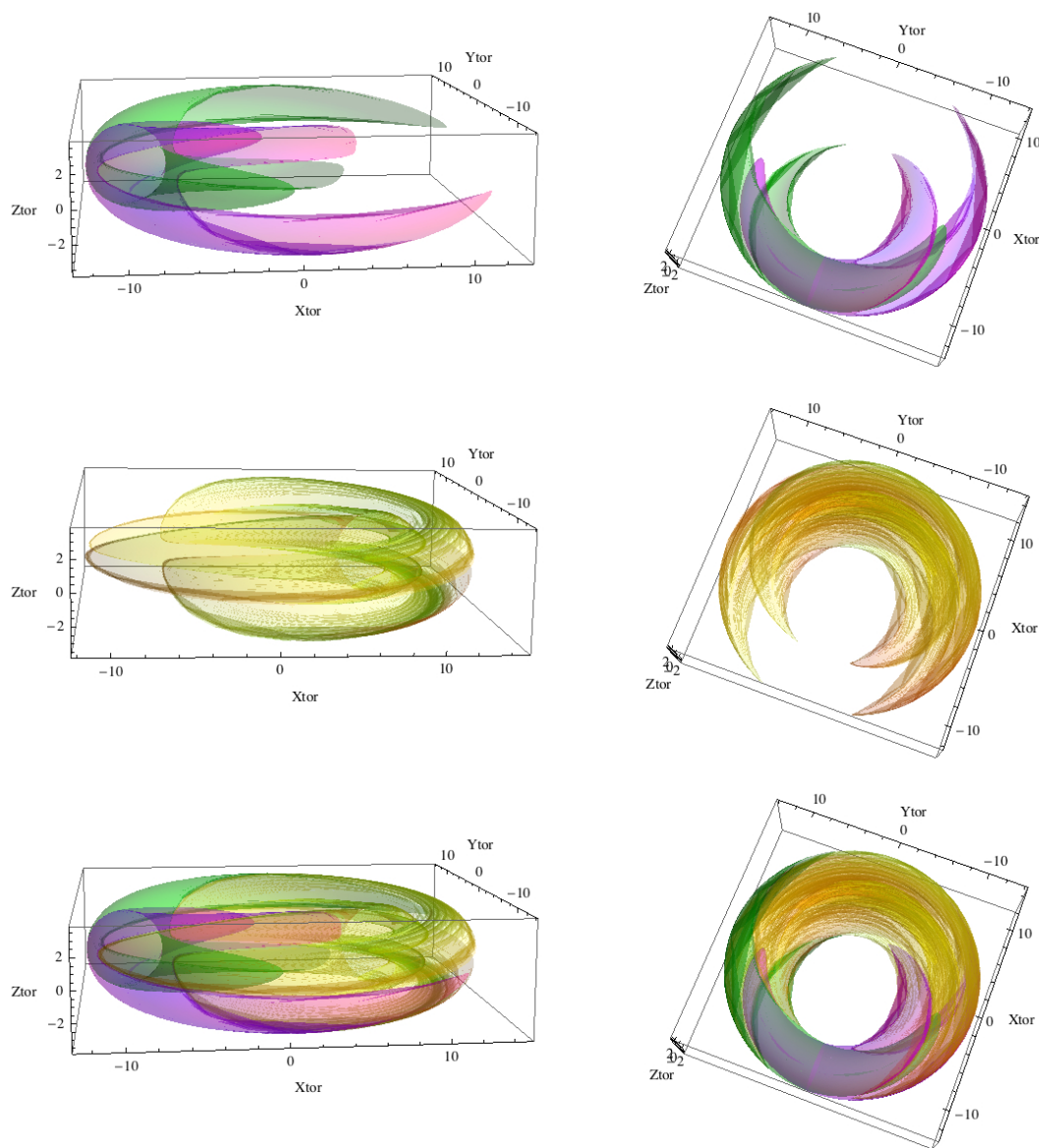


FIGURE 94. Deux perspectives (de profil à gauche et de haut à droite) sur le front d'onde $W(0, \frac{3\pi}{2})$ dans le tore $\mathbb{T}_{\Lambda=5}$. En haut : les deux composantes de $W_{C_1, R=\frac{3\pi}{2}}$. Au milieu : les deux composantes de $W_{C_2, R=\frac{3\pi}{2}}$. En bas : le front d'onde complet $W(0, \frac{3\pi}{2})$.

$$\begin{cases} \lambda_1 = \alpha + \beta\gamma t - \frac{1}{2}\alpha(\beta^2 + \gamma^2)t^2 - \frac{1}{6}\beta\gamma(-4\alpha^2 + \beta^2 + \gamma^2)t^3 \\ \quad - \frac{1}{24}\alpha(-\beta^4 - 14\gamma^2\beta^2 - \gamma^4 + 4\alpha^2(\beta^2 + \gamma^2))t^4 \\ \lambda_2 = \beta - \alpha\gamma t + \frac{1}{2}\beta(\alpha^2 - \gamma^2)t^2 - \frac{1}{6}\alpha\gamma(\alpha^2 - 4\beta^2 - \gamma^2)t^3 \\ \quad + \frac{1}{24}\beta(\alpha^4 - 2(2\beta^2 + 7\gamma^2)\alpha^2 + \gamma^4 + 4\beta^2\gamma^2)t^4 \\ \lambda'_3 = \gamma - \alpha\beta t + \frac{1}{2}(\alpha^2 - \beta^2)\gamma t^2 - \frac{1}{6}\alpha\beta(\alpha^2 - \beta^2 - 4\gamma^2)t^3 \\ \quad + \frac{1}{24}\gamma(\alpha^4 - 2(7\beta^2 + 2\gamma^2)\alpha^2 + \beta^4 + 4\beta^2\gamma^2)t^4. \end{cases}$$

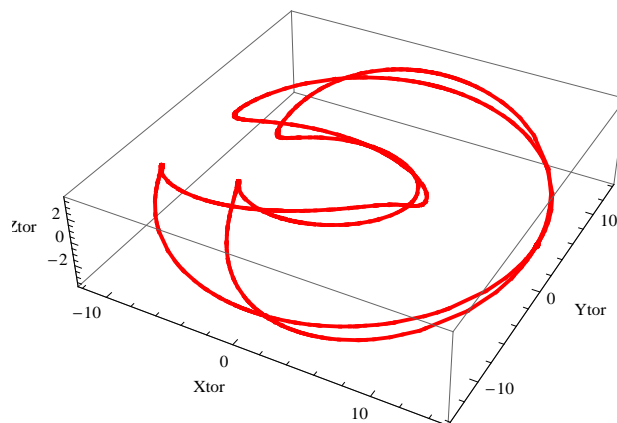


FIGURE 95. Les strates W_{C_3} dans le tore $\mathbb{T}_{\Lambda=5}$.

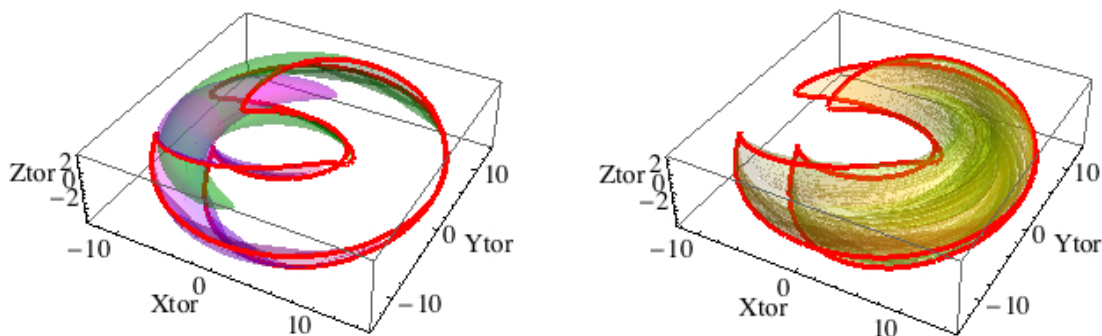


FIGURE 96. Façon dont les strates W_{C_3} bordent les strates W_{C_1} et W_{C_2} dans le tore $\mathbb{T}_{\Lambda=5}$.

On a donc un schéma numérique itératif simple mais où, comme souvent, l'itération engendre de la complexité.

Les intégrales elliptiques et les fonctions de Jacobi possèdent de nombreuses propriétés remarquables reliant par exemple des fonctions à des dérivées d'autres fonctions. Quand on fait des développements limités en tenant compte de ces propriétés, on retombe sur les développements limités construits "à la main".

5. Cusps des projections des géodésiques

Nous avons vu que les projections γ de certaines géodésiques sous-riemanniennes Γ de $\mathbb{V}_S = SE(2)$ peuvent présenter des cusps. Or, si l'on modélise les contours illusoires modaux par des projections de géodésiques, il faut tenir compte du fait qu'on n'observe pas empiriquement de tels cusps. Mais, par ailleurs, il faut aussi tenir compte du fait que, lorsque la courbure est trop forte, on observe des contours illusoires *linéaires par morceaux* possédant des *points anguleux*. Il peut donc exister des contours illusoires avec singularités.

5.1. Le modèle \mathbb{V}_S^\emptyset

Pour éclairer le problème, on peut commencer à analyser les géodésiques *sans cusp* de \mathbb{V}_S . Yuri Sachkov a fait une première analyse dans [477] et l'a ensuite approfondie avec Ugo Boscain, Remco Duits et Francesco Rossi dans [69] et [158] (cf. aussi [159]).

5.1.1. La contrainte d'élimination des cusps.

On peut considérer le problème des géodésiques sous-riemanniennes sans cusp Γ de \mathbb{V}_S comme la relevée du problème de la minimisation de la fonctionnelle

$$J = \int_0^\ell \sqrt{1 + \kappa^2(s)} ds$$

pour les courbes $\gamma : a(0) = a_0 \rightarrow a(\ell) = a_1$ du plan euclidien de base \mathbb{R}^2 paramétrées par leur longueur d'arc s (dans \mathbb{R}^2), $\kappa(s)$ étant la courbure et ℓ la longueur de γ considérée comme libre.⁴² En effet, comme γ est paramétrée par s , les vecteurs tangents $\{\dot{x}(s), \dot{y}(s)\}$ dans la base sont unitaires :

$$\{\dot{x}(s) = \cos(\theta(s)), \dot{y}(s) = \sin(\theta(s))\} .$$

Le carré de l'élément de longueur de la relevée legendrienne Γ (paramétrée par s) de γ est

$$\begin{aligned} \left(\dot{x}(s)^2 + \dot{y}(s)^2 + \dot{\theta}(s)^2 \right) ds^2 &= \left(\cos^2(\theta(s)) + \sin^2(\theta(s)) + \kappa^2(s) \right) ds^2 \\ &= \left(1 + \kappa^2(s) \right) ds^2 \end{aligned}$$

et donc les géodésiques sont bien les minimiseurs de J .

Ce modèle n'est évidemment applicable que si la projection γ est paramétrable par s , autrement dit en dehors des cusps où $\dot{x}(s) = \dot{y}(s) = 0$ et où la tangente à Γ

42. On pourrait introduire un paramètre $\beta > 0$, et considérer la base orthonormée $\{\beta X_1, X_2\}$ ainsi que la fonctionnelle $J_\beta = \int_0^\ell \sqrt{\kappa^2(s) + \beta^2} ds$. Mais on peut rescaler le problème de façon à obtenir $\beta = 1$.

est donc “verticale” (tangente à la fibre \mathbb{S}^1), la s -paramétrisation perdant alors tout sens. Par rapport au système contrôlé général \mathbb{V}_S

$$\left\{ \dot{x} = u_1 \cos(\theta), \dot{y} = u_1 \sin(\theta), \dot{\theta} = u_2 \right\},$$

on doit donc supposer $u_1 \neq 0$ et la paramétrisation par s donne alors $u_1 = 1$,

$$\left\{ \dot{x} = \cos(\theta), \dot{y} = \sin(\theta), \dot{\theta} = u = \kappa \right\}.$$

Nous noterons \mathbb{V}_S^\emptyset (\emptyset pour “sans cusp”) le sous-modèle de \mathbb{V}_S qu’est $SE(2)$ restreint à $u_1 = 1$, $u_2 = u$. \mathbb{V}_S^\emptyset est associé naturellement à la paramétrisation par s des géodésiques mais on peut aussi paramétrer celles-ci par la longueur d’arc t de Γ dans $SE(2)$ en se restreignant aux valeurs de t contraintes par l’apparition de cusps.

On a donc

$$dt = \sqrt{|\dot{x}|^2 + |\dot{y}|^2 + |\dot{\theta}|^2} ds = \sqrt{1 + \kappa^2(s)} ds$$

puisque $|\dot{x}|^2 + |\dot{y}|^2 = \cos^2(\theta) + \sin^2(\theta) = 1$ et $\dot{\theta} = \kappa$. Donc

$$t(s) = \int_0^s \sqrt{1 + \kappa^2(r)} dr$$

et la longueur L de Γ est par conséquent $t(\ell)$. Si l’on note \diamond' la dérivée de \diamond par rapport à t (\diamond restant la dérivée de \diamond par rapport à s), alors la dérivée $\theta' = \frac{d}{dt}\theta(t(s))$ calculée dans la s -paramétrisation est

$$\theta'(s) = \dot{\theta}(s) \frac{ds}{dt} = \frac{\kappa(s)}{\sqrt{1 + \kappa^2(s)}}.$$

Nous la noterons $\varkappa(s)$.⁴³ On a évidemment $0 \leq \varkappa^2(s) < 1$ (la limite $\varkappa^2(s) = 1$ correspondant à la divergence de κ),

$$\begin{aligned} \dot{\varkappa}(s) &= \frac{\dot{\kappa}(s)}{(1 + \kappa^2(s))^{\frac{3}{2}}}, \\ \ddot{\varkappa}(s) &= \frac{\ddot{\kappa}(s)(1 + \kappa^2(s)) - 3\dot{\kappa}^2(s)\kappa(s)}{(1 + \kappa^2(s))^{\frac{5}{2}}}, \\ \varkappa^2(s) &= \frac{\varkappa^2(s)}{1 - \varkappa^2(s)}, \\ \frac{ds}{dt} &= \sqrt{1 - \varkappa^2(s)}. \end{aligned}$$

On a aussi $|x'(t)|^2 + |y'(t)|^2 + |\theta'(t)|^2 = 1$.

43. Le \varkappa est une variante du kappa κ .

5.1.2. *L'hamiltonien de \mathbb{V}_S^\emptyset .*

Calculons l'hamiltonien pour \mathbb{V}_S^\emptyset . Le lagrangien est $\sqrt{1+u^2}$ et les hamiltoniens dépendant du contrôle u obtenus par transformation de Legendre sont

$$h(\lambda, q) = \langle \lambda, \dot{q} \rangle - \sqrt{1+u^2} = \lambda_x \cos(\theta) + \lambda_y \sin(\theta) + \lambda_\theta u - \sqrt{1+u^2} .$$

La maximisation de $h(\lambda, q)$ par rapport au contrôle u conforme au PMP donne $\frac{\partial h}{\partial u} = 0$, soit

$$\lambda_\theta = \frac{u}{\sqrt{1+u^2}} ,$$

ce qui implique

$$\lambda_\theta^2 < 1 \text{ et } \lambda_\theta u - \sqrt{1+u^2} = -\sqrt{1-\lambda_\theta^2} .$$

L'hamiltonien H est par conséquent

$$H(\lambda, q) = \lambda_x \cos(\theta) + \lambda_y \sin(\theta) - \sqrt{1-\lambda_\theta^2}$$

et les équations de Hamilton, devenues beaucoup plus simples, donnent

$$\begin{cases} \dot{x}(s) = \frac{\partial H}{\partial \lambda_x} = \cos(\theta(s)) \\ \dot{y}(s) = \frac{\partial H}{\partial \lambda_y} = \sin(\theta(s)) \\ \dot{\theta}(s) = \frac{\partial H}{\partial \lambda_\theta} = \frac{\lambda_\theta(s)}{\sqrt{1-\lambda_\theta^2(s)}} = \kappa(s) \\ \dot{\lambda}_x(s) = -\frac{\partial H}{\partial x} = 0 \\ \dot{\lambda}_y(s) = -\frac{\partial H}{\partial y} = 0 \\ \dot{\lambda}_\theta(s) = -\frac{\partial H}{\partial \theta} = -(-\lambda_x(s) \sin(\theta(s)) + \lambda_y(s) \cos(\theta(s))) \\ \quad = \lambda_x(s) \sin(\theta(s)) - \lambda_y(s) \cos(\theta(s)) \end{cases}$$

(qui, notons-le, supposent évidemment que les cusps $\dot{x}(s) = \dot{y}(s) = 0$ sont impossibles puisque $\cos(\theta) = \sin(\theta) = 0$ est impossible). On en tire immédiatement

$$\dot{\theta}(s) = \kappa(s) = \frac{\lambda_\theta(s)}{\sqrt{1-\lambda_\theta^2(s)}} ,$$

soit, comme nous l'avons vu puisque $u = \kappa$ sur une extrémale,

$$\lambda_\theta(s) (= \lambda_2(s)) = \frac{\kappa(s)}{\sqrt{1+\kappa^2(s)}}$$

et, par conséquent,

$$\lambda_\theta(s) = \theta'(s) = \varkappa(s) .$$

On en tire également

$$\dot{\lambda}_\theta(s) = \lambda_x(s) \sin(\theta(s)) - \lambda_y(s) \cos(\theta(s)) = \lambda_3(s) .$$

En fait, nous avons vu plus haut que, par rapport à la t -paramétrisation dans le fibré $SE(2)$, les EDO pour la dynamique verticale (avec $\diamond' = \frac{d\circ}{dt}$) sont

$$\begin{cases} \lambda_1'(t) = -\lambda_2(t) \lambda_3(t) \\ \lambda_2'(t) = \lambda_1(t) \lambda_3(t) \\ \lambda_3'(t) = \lambda_2(t) \lambda_1(t) \end{cases}$$

En termes de la s -paramétrisation dans la base (avec $\diamond = \frac{d\circ}{ds}$), elles deviennent simplement $\dot{\lambda}_2(s) = \lambda_3(s)$ et $\dot{\lambda}_3(s) = \lambda_2(s)$ car $\lambda_1 = \frac{1}{\sqrt{1+\kappa^2(s)}} = \frac{ds}{dt}$, relation qui implique en effet

$$\begin{cases} \dot{\lambda}_2(s) = \lambda_2'(t(s)) \frac{dt}{ds} = \lambda_1(t(s)) \lambda_3(t(s)) \frac{1}{\lambda_1(t(s))} = \lambda_3(t(s)) \\ \dot{\lambda}_3(s) = \lambda_3'(t(s)) = \lambda_2(t(s)) \lambda_1(t(s)) \frac{1}{\lambda_1(t(s))} = \lambda_2(t(s)) \end{cases}$$

Et comme $\lambda_2 = \lambda_\theta = \varkappa$, on voit que $\dot{\varkappa} = \lambda_3$ et $\ddot{\varkappa} = \dot{\lambda}_3 = \lambda_2 = \varkappa$.

Remarque. Comme l'ont montré les auteurs de [158], cette équation est dérivable de la condition de stationnarité de $J = \int_0^\ell \sqrt{1 + \kappa^2(s)} ds$ le long de $\gamma(s)$ par la même méthode que celle obtenue par David Mumford pour les elasticæ, méthode que nous avons exposée à la section 3.4 du chapitre 8. On considère une petite variation $\delta\gamma(s)$ de $\gamma(s)$ dans la direction normale en laissant les extrémités $\gamma(0)$ et $\gamma(\ell)$ fixes. Comme $\gamma(s)$ est paramétrée par la longueur d'arc, la tangente unitaire est $T(s) = \dot{\gamma}(s)$ et la normale unitaire est $N(s)$ avec

$$\ddot{\gamma}(s) = \dot{T}(s) = \kappa(s) N(s) .$$

Les calculs montrent que la condition de stationnarité s'écrit

$$\left(\frac{\kappa(s)}{\sqrt{1 + \kappa^2(s)}} \right)'' + \frac{\kappa^3(s)}{\sqrt{1 + \kappa^2(s)}} - \kappa(s) \sqrt{1 + \kappa^2(s)} = 0$$

c'est-à-dire $\left(\frac{\kappa(s)}{\sqrt{1 + \kappa^2(s)}} \right)'' = \frac{\kappa(s)}{\sqrt{1 + \kappa^2(s)}}$, soit $\ddot{\varkappa}(s) = \varkappa(s)$. □

5.1.3. Les portraits de phases.

Nous avons vu que le modèle \mathbb{V}_S avec la paramétrisation t est piloté par l'équation du pendule

$$(\gamma^S)'' = -\sin(\gamma^S)$$

($\gamma^S \in [-\pi, 3\pi]$ de période 4π) avec

$$\sin\left(\frac{\gamma^S}{2}\right) = \lambda_1, \quad \cos\left(\frac{\gamma^S}{2}\right) = -\lambda_2, \quad (\gamma^S)' = c^S = 2\lambda_3 .$$

Nous noterons Π_S le portrait de phase. Le modèle \mathbb{V}_S^\emptyset avec la paramétrisation s est piloté quant à lui par l'équation hyperbolique beaucoup plus simple $\ddot{\varkappa}(s) = \varkappa(s)$ avec $\varkappa(s) \in]-1, 1[$. Nous noterons Π_S^\emptyset le portrait de phase.

Les solutions de l'EDO $\ddot{\varkappa}(s) = \varkappa(s)$ sont bien connues

$$\varkappa(s) = \varkappa(0) \cosh(s) + \dot{\varkappa}(0) \sinh(s)$$

et satisfait la loi de conservation

$$\lambda_1^2 + \lambda_3^2 = 1 - \varkappa^2(s) + \dot{\varkappa}^2(s) = c_\varkappa^2 = \text{cste}$$

(on choisit $c_\varkappa \geq 0$). Elles s'arrêtent lorsqu'un cusp est atteint pour la valeur maximale de s ,

$$s_{\max} = \log \left(\frac{1 + c_\varkappa}{|\varkappa(0)| + |\dot{\varkappa}(0)|} \right),$$

définie par la condition initiale $(\varkappa(0), \dot{\varkappa}(0))$.⁴⁴ Ces cusps des projections $\gamma(s)$ où la relevée legendrienne $\Gamma(s)$ possède une tangente "verticale" (i.e. tangente à la fibre) peuvent aussi être le point de départ de géodésiques. Ils correspondent à $\kappa = \infty$, c'est-à-dire à $\varkappa(0) = 1$ ou $\varkappa(l) = 1$ puisque

$$\varkappa(s) = \frac{\kappa(s)}{\sqrt{1 + \kappa^2(s)}}.$$

L'intégration des trajectoires donne le résultat suivant. Comme le problème est invariant par L -translations dans $SE(2)$, on peut se borner à calculer une famille génératrice de géodésiques sans cusp partant de $q_0 = 0$ et permettant d'obtenir toutes les autres par L -translations. Dans [69], [158] et [159], les auteurs analysent soigneusement le domaine \mathcal{R} (\mathbb{R} pour "reachable") de $SE(2)$ constitué des extrémités q_I "admissibles" atteignables par des géodésiques sans cusp $q_0 \rightarrow q_I$ issues de 0. Ils calculent explicitement \mathcal{R} et son bord $\partial\mathcal{R}$. Un résultat important est que si $q_I \in \mathcal{R}$ alors $q_I + (0, 0, \pi) \notin \mathcal{R}$. Autrement dit le portrait de phase de \mathbb{V}_S défini sur $[-\pi, 3\pi]$ doit être réduit pour \mathbb{V}_S^\emptyset à sa partie centrale $[0, 2\pi]$.

Ils montrent que la famille de trajectoires suivantes $\tilde{\gamma}(s)$ constitue une famille génératrice avec les conditions initiales $\varkappa(0) \in [-1, 1]$, $\dot{\varkappa}(0) \in \mathbb{R}$ et $s \leq s_{\max}$:

44. Si par exemple $\varkappa(s) = 1$ et si l'on pose $e^s = X$, on obtient l'équation du second degré

$$(\varkappa(0) + \dot{\varkappa}(0)) X^2 - 2X + (\varkappa(0) - \dot{\varkappa}(0)) = 0$$

de discriminant

$$\Delta = 4 - 4(\varkappa(0) + \dot{\varkappa}(0))(\varkappa(0) - \dot{\varkappa}(0)) = 4c_\varkappa^2.$$

Les solutions sont donc $s = \frac{1 \pm c_\varkappa}{\varkappa(0) + \dot{\varkappa}(0)}$.

$$\begin{cases} \tilde{x}(s) = \frac{1}{c_{\varkappa}} \varkappa(s) = \frac{1}{c_{\varkappa}} (\varkappa(0) \cosh(s) + \dot{\varkappa}(0) \sinh(s)) \\ \tilde{y}(s) = -\frac{1}{c_{\varkappa}} \int_0^s \sqrt{1 - \varkappa^2(r)} dr \\ \cos(\tilde{\theta}(s)) = \frac{1}{c_{\varkappa}} \dot{\varkappa}(s) \\ \sin(\tilde{\theta}(s)) = -\frac{1}{c_{\varkappa}} \sqrt{1 - \varkappa^2(s)} \end{cases}$$

(i) Dans le cas critique $c_{\varkappa} = 1$ (i.e. $\varkappa^2(0) = \dot{\varkappa}^2(0)$) et $\dot{\varkappa}(0) = -\varkappa(0)$ on a

$$\begin{cases} \tilde{x}(s) = \varkappa(0) e^{-s} \\ \tilde{y}(s) = -\int_0^s \sqrt{1 - \varkappa^2(r)} dr \\ \cos(\tilde{\theta}(s)) = -\varkappa(0) e^{-s} \\ \sin(\tilde{\theta}(s)) = -\sqrt{1 - \varkappa^2(0)} e^{-2s} \end{cases}$$

et donc $s_{\max} = \infty$ (car $1 - \varkappa^2(0) e^{-2s} \geq 0$ pour $s \geq 0$ si $\varkappa(0) \in [-1, 1]$), la trajectoire rejoignant l'axe des y pour $s \rightarrow \infty$.

(ii) En revanche, dans le second cas critique $c_{\varkappa} = 1$ et $\dot{\varkappa}(0) = \varkappa(0)$

$$\begin{cases} \tilde{x}(s) = \varkappa(0) e^s \\ \tilde{y}(s) = -\int_0^s \sqrt{1 - \varkappa^2(r)} dr \\ \cos(\tilde{\theta}(s)) = \varkappa(0) e^s \\ \sin(\tilde{\theta}(s)) = -\sqrt{1 - \varkappa^2(0)} e^{2s} \end{cases}$$

on a $s_{\max} = -\log(|\varkappa(0)|)$ ($1 - \varkappa^2(0) e^{2s_{\max}} = 0$) et pour $s \rightarrow s_{\max}$, la trajectoire rejoint $\tilde{x}(s_{\max}) = \pm 1$.

Ces calculs permettent d'obtenir les valeurs pour l'extrémité des géodésique pour $s = \ell$:

$$\begin{cases} x(\ell) = \frac{1}{c_{\varkappa}^2} \left(\dot{\varkappa}(0) (\varkappa(\ell) - \varkappa(0)) + \sqrt{1 - \varkappa^2(0)} \int_0^{\ell} \sqrt{1 - \varkappa^2(s)} ds \right) \\ y(\ell) = \frac{1}{c_{\varkappa}^2} \left(\sqrt{1 - \varkappa^2(0)} (\varkappa(\ell) - \varkappa(0)) - \dot{\varkappa}(0) \int_0^{\ell} \sqrt{1 - \varkappa^2(s)} ds \right) \\ \cos(\theta(\ell)) = \frac{1}{c_{\varkappa}^2} \left(\dot{\varkappa}(0) \dot{\varkappa}(\ell) + \sqrt{1 - \varkappa^2(0)} \sqrt{1 - \varkappa^2(\ell)} \right) \\ \sin(\theta(\ell)) = \frac{1}{c_{\varkappa}^2} \left(\dot{\varkappa}(\ell) \sqrt{1 - \varkappa^2(0)} - \dot{\varkappa}(0) \sqrt{1 - \varkappa^2(\ell)} \right) \end{cases}$$

Ce système *s'inverse* facilement car on vérifie trivialement que, puisque

$$c_{\varkappa}^2 = 1 - \varkappa^2(0) + \dot{\varkappa}^2(0) ,$$

$$\begin{cases} \dot{\varkappa}(\ell) = \dot{\varkappa}(0) \cos(\theta(\ell)) + \sqrt{1 - \varkappa^2(0)} \sin \theta(\ell) \\ \varkappa(\ell) = \varkappa(0) + x(\ell) \dot{\varkappa}(0) + y(\ell) \sqrt{1 - \varkappa^2(0)} \end{cases}$$

Le lien entre les deux modèles \mathbb{V}_S et \mathbb{V}_S^\emptyset est facile à établir en tenant compte du fait que $t = t(s)$ pour \mathbb{V}_S et $s = s(t)$ pour \mathbb{V}_S^\emptyset et que les dérivées doivent donc être traitées comme des dérivées de fonctions composées. En effet d'un côté $c^S(t) = 2\lambda_3(t) = 2\dot{\varkappa}(s(t))$ et d'un autre $c^S = (\gamma^S)'$ et donc

$$\begin{aligned} (c^S)'(t) &= \frac{dc^S(t)}{dt} = \frac{dc^S(t)}{d\dot{\varkappa}} \frac{d\dot{\varkappa}}{ds} \frac{ds}{dt} = 2\ddot{\varkappa} \frac{ds}{dt} = 2\kappa\sqrt{1-\varkappa^2}, \\ (c^S)' &= (\gamma^S)'' = -\sin(\gamma^S). \end{aligned}$$

L'égalité $-\sin(\gamma^S) = 2\kappa\sqrt{1-\varkappa^2}$ est facilement résoluble par

$$\varkappa(s) = -\cos\left(\frac{\gamma^S(t(s))}{2}\right),$$

ce qui correspond bien à $\cos\left(\frac{\gamma^S}{2}\right) = -\lambda_2$ puisque $\lambda_2 = \varkappa$. On remarquera que cela redonne bien aussi

$$\lambda_1 = \sin\left(\frac{\gamma^S}{2}\right) = \sqrt{1 - \cos^2\left(\frac{\gamma^S}{2}\right)} = \sqrt{1 - \varkappa^2} = \frac{1}{\sqrt{1 + \kappa^2}} = \frac{ds}{dt}.$$

On obtient ainsi un difféomorphisme Φ entre la partie centrale $\gamma^S \in [0, 2\pi]$ du portrait de phase $\Pi_S = (\gamma^S, c^S)$ du pendule pour \mathbb{V}_S et le portrait de phase hyperbolique $\Pi_S^\emptyset = (\varkappa, \dot{\varkappa})$, $\varkappa \in [-1, 1]$ pour \mathbb{V}_S^\emptyset . La stratification de $\Pi_S \upharpoonright_{[0, 2\pi]}$ en strates C_i se transforme par Φ en une stratification C_i^\emptyset , la séparatrice C_3^\emptyset entre C_1^\emptyset et C_2^\emptyset étant constituée des droites $\dot{\varkappa} = \pm\kappa$, i.e. $c_\varkappa = 1$. La figure 97 illustre ce difféomorphisme.

On vérifie qu'à travers Φ les formules des géodésiques de \mathbb{V}_S^\emptyset en termes de s -paramétrisation redonnent bien les formules des géodésiques de \mathbb{V}_S exprimées en termes de fonctions elliptiques de Jacobi dans les coordonnées de Jacobi (φ, k) associées à la t -paramétrisation.

5.1.4. Les cônes accessibles.

Un des grands intérêts du modèle \mathbb{V}_S^\emptyset pour notre propos est qu'il permet de définir des cônes d'accessibilité pour le $\theta(\ell) = \theta_1$ des géodésiques sans cusp. Considérons une projection de géodésique $\gamma(s)$ partant comme convenu de 0 avec une pente $\theta(0) = 0$ et se terminant sans cusp en (x_1, y_1) avec une pente $\theta(\ell) = \theta_1$. On peut alors calculer les projections de géodésiques sans cusp arrivant au même point (x_1, y_1) du plan de base mais avec une pente $\hat{\theta}_1$ pouvant être différente de θ_1 . Les auteurs montrent dans [158] et [159] que les $\hat{\theta}_1$ possibles sont dans un cône $[\theta_{1,\min}, \theta_{1,\max}]$ entourant θ_1 et qui peut être de trois types, $\text{cond}(y_1)$ étant une condition un peu compliquée sur y_1 :

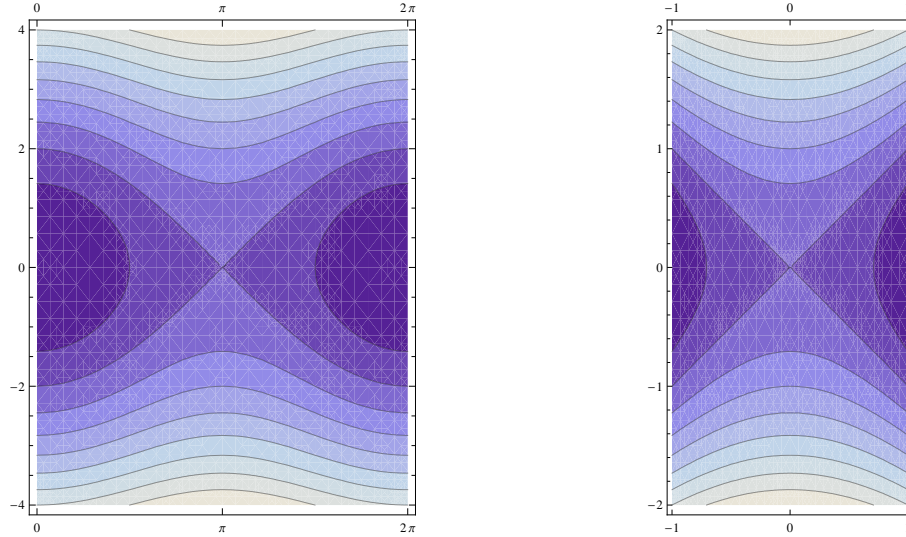


FIGURE 97. Le difféomorphisme Φ entre la partie centrale $\gamma^S \in [0, 2\pi]$ du portrait de phase $\Pi_S = (\gamma^S, c^S)$ du pendule pour \mathbb{V}_S et le portrait de phase hyperbolique $\Pi_S^\infty = (\varkappa, \dot{\varkappa})$, $\varkappa \in [-1, 1]$ pour \mathbb{V}_S^∞ .

- soit $x_1 < 2$ et $\text{cond}(y_1)$ et $y_1 > 0$ et alors $\theta_{1,\min}$ est le $\hat{\theta}_1$ d'une géodésique démarrant par un cusp et finissant en (x_1, y_1) alors que $\theta_{1,\max}$ est le $\hat{\theta}_1$ d'une géodésique finissant par un cusp en (x_1, y_1) ,
- soit $x_1 < 2$ et $\text{cond}(y_1)$ et $y_1 < 0$ et alors $\theta_{1,\min}$ est le $\hat{\theta}_1$ d'une géodésique finissant par un cusp en (x_1, y_1) alors que $\theta_{1,\max}$ est le $\hat{\theta}_1$ d'une géodésique démarrant par un cusp et finissant en (x_1, y_1) ,
- soit $x_1 \geq 2$ ou $\neg\text{cond}(y_1)$ ⁴⁵ et alors $\theta_{1,\min}$ est le $\hat{\theta}_1$ d'une géodésique finissant par un cusp en (x_1, y_1) alors que $\theta_{1,\max}$ est le $\hat{\theta}_1$ d'une géodésique finissant aussi en (x_1, y_1) par un cusp, mais un cusp de tangente verticale d'orientation opposée à celle de $\theta_{1,\min}$.⁴⁶

Comme par hypothèse les géodésiques considérées partent de 0 et donc avec $\theta(0) = 0$, elles partent dans la direction des $x \geq 0$ et l'on montre que, si elles sont sans cusp, elles *restent* dans le demi-plan $x \geq 0$ même si elles peuvent revenir en arrière. À la limite elles peuvent revenir sur $x = 0$ en y_1 mais seulement si leurs extrémités $(0, 0)$ et $(0, y_1)$ sont des points cusp avec $\theta_1 = \pi$ si $y_1 > 0$ et $\theta_1 = -\pi$ si $y_1 < 0$. En effet, les formules ci-dessus impliquent que si $x(\ell) = 0$,

45. Dans $\neg\text{cond}(y_1)$ \neg est le symbole logique de la négation.

46. Dire qu'une géodésique sans cusp de \mathbb{V}_S^∞ commence ou se termine par un point cusp signifie que si on la prolongeait dans \mathbb{V}_S ce point deviendrait un cusp.

alors $y(\ell) = \frac{\varkappa(\ell) - \varkappa(0)}{\sqrt{1 - \varkappa^2(0)}}$, $\dot{x}(\ell) = -\dot{x}(0)$ (i.e. $\cos(\theta(\ell)) = -1$), $\dot{x}(\ell) \sqrt{1 - \varkappa^2(0)} - \dot{x}(0) \sqrt{1 - \varkappa^2(\ell)} = 0$ (i.e. $\sin(\theta(\ell)) = 0$) et donc $\varkappa^2(0) = \varkappa^2(\ell)$.

Les cônes d'accessibilité aux points $(0, y_1)$ de l'axe des y sont donc réduits à une seule direction $\theta_1 = \pm\pi$. Plus $x_1 > 0$ augmente, plus les cônes d'accessibilité s'ouvrent, mais jamais complètement : $\theta_{1,\max} - \theta_{1,\min}$ reste $< \pi$ car si (x_1, y_1, θ_1) est atteignable, alors, comme nous l'avons vu, $(x_1, y_1, \theta_1 + \pi)$ ne l'est pas.

5.1.5. Le fibré en cercles.

Ce modèle s -paramétré est équivalent à notre modèle de fibré en cercles $\mathbb{R}^2 \times \mathbb{S}^1 = SE(2)$, exposé à la section 7 du chapitre 10. Le fibré tangent a pour coordonnées locales $(x, y, \theta; \xi, \eta, \varphi)$, la métrique est $dt^2 = dx^2 + dy^2 + d\theta^2$ et le lagrangien est :

$$L(x) = \sqrt{\xi^2 + \eta^2 + \varphi^2} = \sqrt{1 + f'(x)^2 + \frac{f''(x)^2}{(1 + f'(x)^2)}}$$

(\diamond' indiquant la dérivée de \diamond par rapport à t et \diamond la dérivée par rapport à s) avec $\theta = \text{Arctan}(f')$,

$$\varphi = \theta' = \dot{\theta} \frac{ds}{dx} = \frac{f''(x)}{1 + f'(x)^2}$$

et la contrainte cinématique $\Sigma = \theta - \text{Arctan}(\eta) = 0$. Il correspond bien à la minimisation de $J = \int_0^\ell \sqrt{1 + \kappa^2(s)} ds$.

5.2. Le modèle $\mathbb{V}_P = SE(2) / (\sim \text{mod } \pi)$

En fait, comme nous l'avons expliqué dans la section 1 du chapitre 10, dans la mesure où les orientations ne sont neurophysiologiquement définies que modulo π , il faut travailler dans le quotient $\mathbb{V}_P = M \times \mathbb{P}^1 = SE(2) \text{ (mod } \pi)$. Mais si l'on interdit les cusps, alors \mathbb{V}_P° est équivalent à \mathbb{V}_S° car alors $\gamma(s) + (0, 0, \pi)$ donne simplement la même courbe $\gamma(s)$ mais parcourue en sens inverse. Dans \mathbb{V}_P° , les géodésiques sans cusp $0 = (0, 0, 0) \rightarrow q_1 = (x_1, y_1, \theta_1)$ sont celles pour lesquelles soit $(x_1, y_1, \theta_1) \in \mathcal{R}$, soit $(x_1, y_1, \theta_1 + \pi) \in \mathcal{R}$, soit $(-x_1, y_1, -\theta_1) \in \mathcal{R}$, soit $(-x_1, y_1, -\theta_1 + \pi) \in \mathcal{R}$.

Si l'on admet les cusps, la géométrie du modèle $\mathbb{V}_P = SE(2) / (\sim \text{mod } \pi)$ se déduit de celle de \mathbb{V}_S mais avec un changement de la métrique. En effet si on note $\widehat{\diamond}$ les éléments de \mathbb{V}_P et si $\widehat{q}_1 \in \mathbb{V}_P$, alors la distance $\widehat{d}(\widehat{0}, \widehat{q}_1)$ est la longueur minimale des géodésiques de \mathbb{V}_S $(0, 0, 0) \rightarrow (x_1, y_1, \theta_1)$, ou $(0, 0, \pi) \rightarrow (x_1, y_1, \theta_1)$, ou $(0, 0, 0) \rightarrow (x_1, y_1, \theta_1 + \pi)$, ou enfin $(0, 0, \pi) \rightarrow (x_1, y_1, \theta_1 + \pi)$. Cela fait donc diminuer la distance.

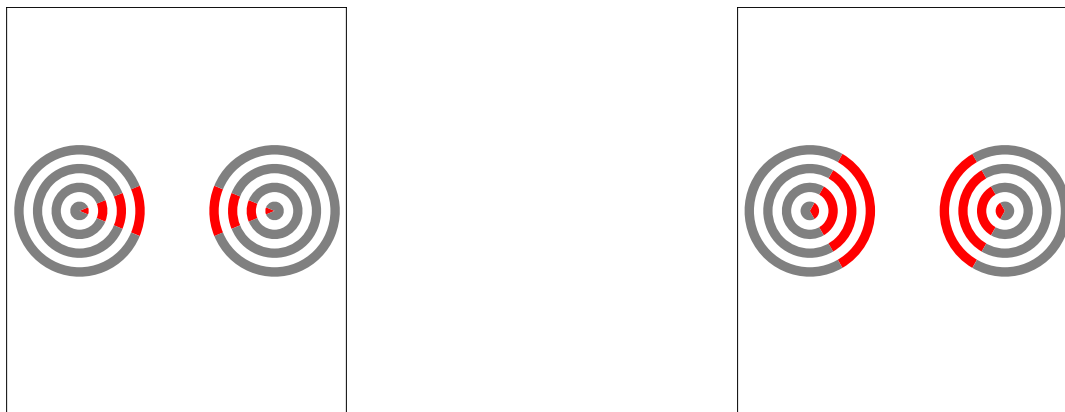


FIGURE 98. Configuration à deux pacmen avec effet néon analogue à la figure 6 à quatre pacmen du chapitre 8, section 2.1. Pour une ouverture faible des pacmen (à gauche) les sujets voient majoritairement des contours illusoires courbes. En revanche, pour une ouverture assez forte des pacmen (à droite) les sujets voient majoritairement des contours illusoires droits avec un point anguleux.

5.3. Cusps ou points anguleux ?

Revenons à la neurophysiologie des contours illusoires modaux. La question expérimentale de leurs singularités est très intéressante. Il semble qu'on n'observe pas de cusps mais qu'on observe en revanche des points anguleux. Considérons par exemple la configuration à deux pacmen avec effet néon de la figure 98 (qui est du même ordre que la figure 6 à quatre pacmen de la section 2.1 du chapitre 8). Pour une ouverture faible des pacmen (à gauche) les sujets voient majoritairement des contours illusoires courbes. En revanche, pour une ouverture assez forte des pacmen (à droite) les sujets voient majoritairement des contours illusoires droits avec un point anguleux. Comme nous l'avons déjà expliqué dans cette section 8–2.1, cela montre que le système visuel peut utiliser différentes stratégies de complétion. Lorsque le sujet observe un contour courbe, il diffuse la courbure le long du contour. En revanche, lorsqu'il observe un point anguleux, il annule partout la courbure sauf au point anguleux où il la concentre. Il y a donc une compétition entre une stratégie de diffusion et une stratégie de concentration-trivialisation.

Cette compétition de stratégies de complétion modale est encore plus nette pour certains CIs du second ordre. Nous en avons également parlé à la section 8–2.1 à propos du fait que (cf. Vol I, section 5.8. du chapitre 5) certains neurones de V2 (“end-stopped cells”) détectent des orientations virtuelles orthogonales à des extrémités de lignes. Reprenons la figure 99 dont la bistabilité est frappante, le sujet observant alternativement soit un cercle, soit un carré : en fixant l'image assez

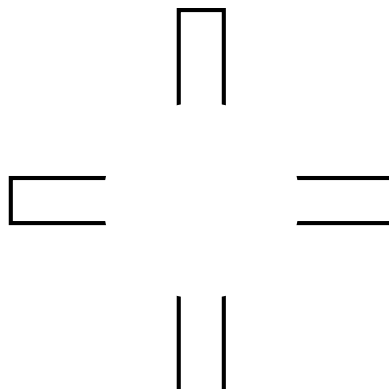


FIGURE 99. Contours illusoires bistables faisant intervenir l'aire V_2 . Il existe une compétition entre la forme "carré" et la forme "cercle" des contours subjectifs (reprise de la figure 17 de la section 8, 2.1).

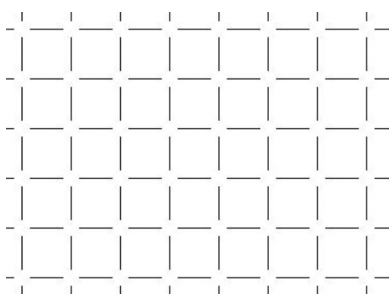


FIGURE 100. De petits cercles illusoires émergent aux extrémités des lignes. On observe une certaine bistabilité, les cercles bifurquant en carrés au bout d'un certain temps.

longtemps le percept bifurque spontanément et périodiquement d'un cas à l'autre. Nous allons en donner une explication possible.

5.4. La bistabilité des contours illusoires

La bistabilité ces CIs modaux est spectaculaire dans l'illusion visuelle d'Ehrenstein (figure 100) même s'il y a une certaine dominance des cercles.

Dans le contexte d'une explication variationnelle des contours illusoires, il faut comprendre comment les deux solutions suivantes peuvent entrer en compétition

- (i) une géodésique,
- (ii) une solution linéaire par morceaux concentrant toute la courbure en des points anguleux.

Dans le plan base euclidien \mathbb{R}^2 la solution linéaire par morceaux est évidemment *toujours* plus longue que l'arc de cercle et il ne peut donc pas y avoir compétition

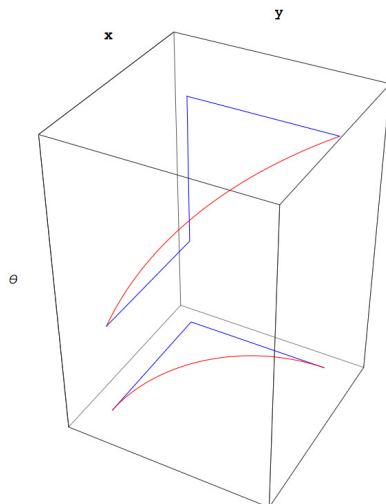


FIGURE 101. Les relevées legendriennes de deux contours illusoires satisfaisant les mêmes conditions aux limites : une solution géodésique et une solution linéaire par morceaux avec une singularité de la courbure.

entre les deux stratégies. En revanche dans le modèle \mathbb{V} de $V1$, il peut y avoir compétition. La figure 101 montre les relevées legendriennes des deux solutions.

Giovanna Citti et Alessandro Sarti, ont fait le petit calcul suivant. Considérons un quart de cercle γ de rayon r dans la base \mathbb{R}^2 en supposant que le poids relatif des connexions horizontales pour la courbure $k = \frac{1}{r}$ soit donné par un paramètre α (avec $\alpha = 1$ lorsque la courbure est nulle). La longueur de sa relevée legendrienne Γ est $L_1 = \int_{\gamma} \sqrt{1 + \alpha^2 k^2} ds$ avec ds la longueur d'arc. Considérons maintenant la relevée de la solution linéaire par morceaux en supposant que le poids relatif du saut dans \mathbb{V} lié au point anguleux dans la base \mathbb{R}^2 soit donné par un paramètre β . Sa longueur est donc $L_2 = 2r + \beta \frac{\pi}{2}$. Il pourra y avoir compétition lorsque $L_1 = L_2$, c'est-à-dire (en élevant au carré) lorsque

$$\frac{\pi^2}{4} (r^2 + \alpha^2) = 4r^2 + 2\pi\beta r + \beta^2 \frac{\pi^2}{4} .$$

Comme $16 > \pi^2$, il y a toujours une racine positive si $\beta < \alpha$

$$r = \frac{-4\pi\beta + \sqrt{\pi^2(16 - \pi^2)\alpha^2 + \pi^4\beta^2}}{16 - \pi^2} .$$

Par exemple dans le cas le plus simple où $\alpha = 1$ (les connexions horizontales ont une force indépendante de la courbure) et $\beta = 0$ (il existe des détecteurs de coins) on trouve comme rayon critique $r_c = \frac{\pi}{\sqrt{16 - \pi^2}} \sim 0.5$. Pour $\alpha = 1$ et $\beta = \frac{1}{2}$, $r_c \sim 0.478$. Pour $r > r_c$ on a $L_1 > L_2$ et c'est la solution linéaire par morceaux

qui l'emporte. Pour $r < r_c$ on a $L_1 < L_2$ et c'est au contraire la solution cercle qui l'emporte. Il serait intéressant de faire des expériences suffisamment précises pour tester et raffiner des modèles de ce genre. L'important ici était simplement de montrer que, dans \mathbb{V} muni de sa géométrie sous-riemannienne, il peut effectivement y avoir bistabilité des contours illusoirs.

6. Quelques autres résultats

6.1. Variations des géodésiques

Dans ce qui précède nous avons plusieurs fois considéré des familles de géodésiques Q_I indexées par des paramètres $I : Q_{\widetilde{\lambda S}}, Q_{C_i, k}, Q_{C_i, t}$, etc. De façon générale, pour une métrique sous-riemannienne sur une variété de contact M de dimension n , on peut étudier les variations d'une géodésique $\Gamma = q(t)$ de condition initiale le covecteur $\lambda \in T_q^*M$ lorsque l'on varie un peu λ .⁴⁷ Pour ce faire, on peut suivre la façon dont la structure symplectique canonique de T^*M évolue le long de la géodésique $q(t)$, puis considérer un vecteur tangent du second ordre $\Xi \in T_\lambda(T^*M)$ et regarder comment il opère sur cette évolution. Cela conduit à la théorie des *courbes de Jacobi*, bien connue en géométrie riemannienne et développée de façon sophistiquée en géométrie sous-riemannienne par Andrei Agrachev et Revaz Gamkrelidze en 1997-1998 et aussi synthétisée dans le chapitre 15 du traité [5] d'Agrachev-Barilari-Boscain. Elle relève de problématiques du second ordre comme celles que nous avons très brièvement évoquées dans la section 2.2 du chapitre 12.

En géométrie riemannienne classique les éléments métriques standard de courbure (courbure de Riemann, courbure sectionnelle, courbure scalaire, courbure de Ricci) permettent de calculer les géodésiques comme nous l'avons vu dans la section 6.2 du chapitre 6 et la section 6 du chapitre 12. Il s'agit de généraliser ces notions au cas sous-riemannien en tenant compte du fait qu'il n'y a plus de connexion de Levi-Civita, qu'il n'y a plus de transport parallèle et que la distance n'est plus une fonction C^∞ . C'est l'œuvre de nombreux spécialistes comme par exemple Luigi Ambrosio, Fabrice Baudoin, Nicola Garofalo, Ludovic Rifford, Cédric Villani, ou Igor Zelenko.

Si $\Xi \in T_\lambda(T^*M)$, on peut le transporter en $\Xi(t)$ le long de l'extrémale $\lambda(t) = e^{t\vec{H}}(\lambda)$ au moyen de la différentielle $D_\lambda(e^{t\vec{H}})$ de $e^{t\vec{H}}$ en λ . Précisons bien les espaces

47. Donc $q(0) = q$ et $q(t)$ est la projection sur M , $\pi(\lambda(t))$, de l'extrémale $\lambda(t) = e^{t\vec{H}}(\lambda) \in T_{q(t)}^*M$ dans T_q^*M satisfaisant $\lambda(0) = \lambda$.

en jeu. L'exponentielle $e^{t\vec{H}}$ du champ hamiltonien \vec{H} est une application

$$\begin{aligned} e^{t\vec{H}} : T^*M &\rightarrow T^*M \\ \lambda = (q \in M, \lambda \in T_q^*M) &\in T^*M \\ \mapsto e^{t\vec{H}}(\lambda) = \lambda(t) &= (q(t) \in M, \lambda(t) \in T_{q(t)}^*M) \in T^*M \end{aligned}$$

(si $\lambda \in T_q^*M$, nous utilisons aussi la notation λ pour $(q, \lambda) \in T^*M$ s'il n'y a pas de confusion). $D_{\lambda(t)}(e^{t\vec{H}})$ est donc une application linéaire $T_\lambda(T^*M) \rightarrow T_{\lambda(t)}(T^*M)$ et $D_\lambda(e^{t\vec{H}})(\Xi) \in T_{\lambda(t)}(T^*M)$. Parmi les $\Xi \in T_\lambda(T^*M)$ il y a les Ξ "verticaux" gardant $q = \pi(\lambda)$ constant et ne variant λ que dans la fibre T_q^*M . Ces Ξ appartiennent à l'espace "vertical" $V_\lambda = T_\lambda(T_q^*M)$ (qui est isomorphe comme espace vectoriel à T_q^*M car ce dernier est lui-même un espace vectoriel) et $D_\lambda(e^{t\vec{H}})$ envoie V_λ dans un sous-espace $V_{\lambda(t)} \subset T_{\lambda(t)}(T^*M)$.

On utilise alors la propriété de V_λ d'être un sous-espace *lagrangien* de $T_\lambda(T^*M)$. Nous avons déjà rencontré cette notion à propos des caustiques en optique dans la section 6.1 du chapitre 12. Cela signifie que V_λ est de dimension n moitié de celle de $T_\lambda(T^*M)$ et que la forme symplectique canonique de T^*M s'annule identiquement sur V_λ . Or, par définition, $e^{t\vec{H}}$ préserve cette structure symplectique ce qui implique que l'image $V_{\lambda(t)}$ de V_λ par $D_\lambda(e^{t\vec{H}})$ dans $T_{\lambda(t)}(T^*M)$ est également un sous-espace lagrangien. On peut alors, en utilisant le champ hamiltonien inverse $e^{-t\vec{H}}$ (qui préserve évidemment lui aussi la structure symplectique de T^*M) ramener tous les $T_{\lambda(t)}(T^*M)$ dans le même $T_\lambda(T^*M)$. Soit donc $J_\lambda(t) = D_{\lambda(t)}(e^{-t\vec{H}})V_{\lambda(t)}$ (on part cette fois de $T_{\lambda(t)}(T^*M)$ pour revenir dans $T_\lambda(T^*M)$). Les $J_\lambda(t)$ constituent *une courbe de sous-espaces lagrangiens* de $T_\lambda(T^*M)$ avec, par construction, la condition initiale $J_\lambda(0) = V_\lambda$ et la propriété $J_\lambda(t+s) = D_{\lambda(t)}(e^{-t\vec{H}})(J_{\lambda(t)}(s))$.⁴⁸ Ces courbes sont les *courbes de Jacobi* et, comme l'expliquent Andrei Agrachev, Davide Barilari et Ugo Boscain ([5], p. 547)

"The Jacobi curve at point $\lambda \in T^*M$ parametrizes all the possible geodesic variations of the geodesic associated with an initial covector λ ."

Elles permettent d'analyser des invariants symplectiques du flot géodésique :

"Heuristically, we would like to extract geometric properties of the sub-Riemannian structure by studying the symplectic invariants of its geodesic

48. Il est important de noter que, bien que $V_\lambda = T_\lambda(T_q^*M)$ soit "vertical", $J_\lambda(t)$ n'est pas en général "vertical" dans $T_\lambda(T^*M)$, i.e. en général $J_\lambda(t) \not\subset T_\lambda(T_q^*M)$.

flow, i.e. the flow of \vec{H} . The simplest idea is to look for invariants of its linearisation.” ([5], p. 542)

Nous renvoyons à [5] pour la théorie des courbes dans l’espace \mathfrak{L}_λ (la grassmannienne) des sous-espaces lagrangiens de l’espace vectoriel symplectique $T_\lambda(T^*M)$. Elle permet de démontrer tout un ensemble de résultats très significatifs. Esquissons-en très brièvement quelques uns.

On considère donc le vectoriel $T_\lambda = T_\lambda(T^*M)$ de dimension $2n$ avec la 2-forme symplectique canonique Ω . Les sous-espaces de dimension k de T_λ forment la grassmannienne $\mathfrak{G}_k(T_\lambda)$ qui est une variété *compacte* de dimension $k(\dim(T_\lambda) - k) = k(2n - k)$.⁴⁹ La grassmannienne $\mathfrak{G}_n(T_\lambda)$ des sous-espaces de dimension moitié n est donc de dimension $n(2n - n) = n^2$. Si l’on tient compte de la structure symplectique Ω , il est naturel de considérer les sous-espaces U de T_λ qui sont *symplectiques* au sens où la 2-forme Ω y est non dégénérée, et, à l’opposé, les sous-espaces U de T_λ qui sont *isotropes* au sens où la 2-forme Ω s’y annule identiquement. Si l’on appelle “orthogonal symplectique” U^\perp de U le sous-espace des $\Xi \in T_\lambda$ tels que $\Omega(\Xi, \Theta) = 0$ pour tout $\Theta \in U$, alors U est symplectique si $U \cap U^\perp = \{0\}$ et au contraire U est isotrope si $U \subseteq U^\perp$. On dit alors que U est un sous-espace *lagrangien* s’il est isotrope de dimension maximale n , autrement dit si $U = U^\perp$. C’est l’opposé d’un sous-espace symplectique. Notons J plutôt que U un sous-espace lagrangien de T_λ (les $J_\lambda(t)$ sont donc des J). Les J forment une grassmannienne lagrangienne \mathfrak{L}_λ qui est une sous-variété compacte de dimension $\frac{n(n+1)}{2}$ de la grassmannienne $\mathfrak{G}_n(T_\lambda)$ qui est quant à elle de dimension n^2 .

Soit alors $J(t)$ une courbe dans \mathfrak{L}_λ avec $J(0) = J$ et soit $\dot{J}(0) \in T_J\mathfrak{L}_\lambda$ son vecteur tangent à l’origine. $\dot{J}(0)$ est identifiable à une *forme quadratique* sur J à travers une identification globale entre $T_J\mathfrak{L}_\lambda$ et l’espace des formes quadratiques sur J . En effet, si $\Xi \in J \subset T_\lambda$, on l’étend en une courbe $\Xi(t) \in J(t)$ et l’on considère $\dot{\Xi} = \dot{\Xi}(0)$ vu également comme élément de T_λ . On peut alors définir \dot{J} sur J par $\dot{J}(\Xi) = \Omega(\Xi, \dot{\Xi})$ et l’on montre que $\dot{J}(\Xi)$ est indépendante de l’extension $\Xi(t)$ choisie. En fait, $T_J\mathfrak{L}_\lambda$ est un sous-espace de l’espace tangent en J , $T_J(\mathfrak{G}_n(T_\lambda))$, de la grassmannienne $\mathfrak{G}_n(T_\lambda)$. Or celui-ci est canoniquement isomorphe aux applications linéaires entre J et le quotient T_λ/J à travers l’association à $\Xi \in J$ du vecteur $\dot{\Xi} \bmod(J) \in T_\lambda/J$. Mais comme J est lagrangienne par hypothèse, Ω établit un isomorphisme entre T_λ/J et le dual J^* . Pour tout $\Phi \in T_\lambda$, cet isomorphisme associe à $\Phi \bmod(J) \in T_\lambda/J$ l’application linéaire $\Omega(\Phi, \Theta)$ sur J et l’on montre que les hypothèses font que cet isomorphisme est bien défini. Donc à $\Xi \in J$ on associe $\Omega(\Xi, \dot{\Xi})$ et l’on montre que,

49. La compacité généralise celle des espaces projectifs : $\mathfrak{G}_1(T)$, l’espace des droites de T , est le projectivisé de T .

toujours parce que J est lagrangienne, cette application linéaire de J dans J^* est auto-adjointe et correspond donc à une forme quadratique.

Dans le cas de J_λ , on montre alors que $\dot{J}_\lambda(0) = -2H|_{T_q^*M}$. En effet, si $\Xi \in J_\lambda$,

$$\dot{J}_\lambda(0)(\Xi) = \Omega(\Xi, \dot{\Xi}) = \Omega\left(\Xi, \frac{d}{dt}\Big|_{t=\dot{\alpha}} e^{-t\vec{H}}(\Xi)\right) = \Omega\left(\Xi, \left[\vec{H}, \Xi\right]\right).$$

Mais on montre que, comme J_λ est “vertical”, alors pour tout $\Phi \in T_\lambda$ on a $\Omega(\Xi, \Phi) = \langle \Xi, \pi_*\Phi \rangle$ ⁵⁰ et il reste donc à calculer $\pi_*\left([\vec{H}, \Xi]\right)$. On trouve $\dot{J}_\lambda(0)(\Xi) = -2H(\Xi)$. Une conséquence immédiate est que le *rang* de $\dot{J}_\lambda(0)$ est le rang de H sur T_q^*M et que si la métrique est vraiment sous-riemannienne $\dot{J}_\lambda(0)$ n’est pas de rang n et donc la courbe $J_\lambda(t)$ dans \mathfrak{L}_λ ne peut pas être régulière en 0.

Ces constructions permettent de mieux calculer les *points conjugués*. On montre en effet que si $\Gamma = \{q(s) = \text{Exp}_q(s\lambda)\}$ est la géodésique issue de q avec le covecteur λ et si Γ est normale, alors $q(t)$ est conjugué de q le long de Γ si et seulement si $J_\lambda(t) \cap J_\lambda(0) \neq \{0\}$ (on dit alors que $J_\lambda(t)$ est conjugué de $J_\lambda(0) = J_\lambda$ le long de $J_\lambda(s)$).

6.2. Modèles stochastiques

Nous avons vu dans la section 3.4 du chapitre 8 que David Mumford a donné une interprétation *stochastique* de son modèle des elasticæ en supposant que la courbure $\kappa(s)$ de γ dans \mathbb{R}^2 (s étant la longueur d’arc) est un bruit blanc et que l’angle $\theta(s)$ est donc un mouvement brownien.⁵¹ En termes de théorie du contrôle cela revient à considérer le processus stochastique

$$\begin{cases} \dot{x} = \cos(\theta) \\ \dot{y} = \sin(\theta) \\ \dot{\theta} \sim N(0, \sigma^2) \end{cases}$$

où $\dot{\theta}$ est maintenant une variable aléatoire normale (gaussienne) de moyenne nulle et de variance σ^2 . Ce processus a été étudié par Gonzalo Sanguinetti dans sa thèse [483] dirigée par Giovanna Citti et Alessandro Sarti et également par Remco Duits

50. Soit $\Omega = \sum_{i=1}^{i=n} d\lambda_i \wedge dq_i$ en coordonnées canoniques. Comme Ξ est vertical, il est de la forme $\Xi = \sum_{i=1}^{i=n} \Xi_{\lambda_i} \partial_{\lambda_i}$ avec les composantes $\Xi_{q_i} = 0$ et par conséquent

$$\Omega(\Xi, \Phi) = \sum_{i=1}^{i=n} (\Xi_{\lambda_i} \Phi_{q_i} - \Xi_{q_i} \Phi_{\lambda_i}) = \sum_{i=1}^{i=n} \Xi_{\lambda_i} \Phi_{q_i} = \langle \Xi, \pi_*\Phi \rangle.$$

51. Répétons que nous reviendrons techniquement sur le mouvement brownien à propos des processus de diffusion dans la section 9.1.5 du chapitre 17.

et Markus van Almsick (cf. [156]). Il s'agit d'un mécanisme d'advection-diffusion régi par une équation de Fokker-Planck.

Les équations de Fokker-Planck, qui généralisent l'équation de diffusion de la chaleur à des processus stochastiques plus compliqués, sont d'un usage ubiquitaire en physique statistique,⁵² par exemple dans le mouvement brownien où l'on étudie les trajectoires stochastiques (continues mais non dérivables) de particules "lourdes" soumises à des chocs aléatoires avec les molécules du milieu. On ne peut pas avoir accès expérimentalement aux mouvements réels mais seulement à la probabilité $P(x, t)$ que la particule soit à la position x au temps t . En 1908, Paul Langevin écrivit une équation qui devint le modèle des équations de Fokker-Planck de la forme

$$\frac{\partial P}{\partial t}(x, t) = - \sum_i \partial_{x_i} (D_1(x) P(x, t)) + \frac{1}{2} \sum_{i,j} \partial_{x_i x_j}^2 (D_2(x) P(x, t))$$

où $D_1(x)$ et $D_2(x)$ sont des coefficients. Le premier terme est un terme d'advection (ou de "drift") et le second un terme de diffusion. La littérature consacrée à cette classe très particulière d'EDP est très vaste. Le lecteur intéressé pourra consulter par exemple les ouvrages de Hannes Risken [465] et Grigorios Pavliotis [402].

Dans notre cas, l'advection se produit le long de la direction X_1 et la diffusion de θ se produit le long de la direction X_2 . La solution fondamentale de l'équation de Fokker-Planck étant trop difficile à expliciter pour le modèle $\mathbb{V}_S = SE(2)$, les auteurs cités de [483] et [156] se sont focalisés sur la nilpotentisation (le cône tangent) de $SE(2)$ qu'est le modèle \mathbb{V}_J du groupe de Heisenberg polarisé.

Soit $v_0 = (x_0, y_0, \theta_0) = (a_0, \theta_0)$ un point dans \mathbb{V}_S et suivons un chemin aléatoire partant de v_0 . S'il n'y a pas de bruit, la trajectoire est déterministe et se réduit à une ligne droite satisfaisant le principe de stricte coaxialité (pas de courbure) :

$$\{\theta = \theta_0, x = x_0 + \cos(\theta_0) t, y = y_0 + \sin(\theta_0) t\}.$$

S'il y a un bruit stochastique, si $v = (x; y, \theta) = (a, \theta)$ est un élément générique de \mathbb{V}_S et si $P(v, t)$ est la probabilité de trouver la marche aléatoire en v au temps t , l'EDP d'évolution de P avec la condition initiale $P_0(v) = P(v, 0)$ est

$$\begin{aligned} \frac{\partial P}{\partial t}(v, t) &= - \left(\cos(\theta) \frac{\partial P}{\partial x}(v, t) + \sin(\theta) \frac{\partial P}{\partial y}(v, t) \right) + \frac{\sigma^2}{2} \frac{\partial^2 P}{\partial \theta^2}(v, t) \\ \frac{\partial P}{\partial t}(v, t) &= -X_1(P(v, t)) + \frac{\sigma^2}{2} (X_2)^2(P(v, t)) . \end{aligned}$$

52. Elles interviennent par exemple, comme nous l'avons expliqué dans plusieurs cours et travaux (cf. [422] et [435], chapitre 8, section 3), dans la théorie de Yoshiki Kuramoto et Hiroaki Daido de la synchronisation des réseaux d'oscillateurs.

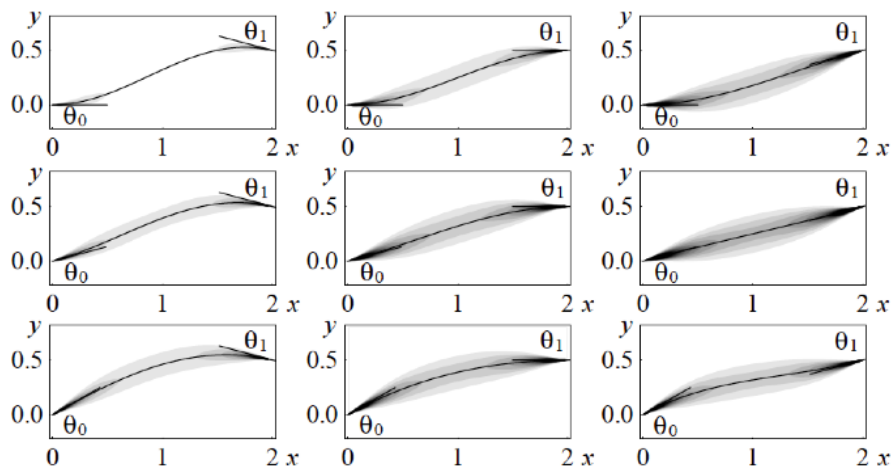


FIGURE 102. Complétion d'un contour avec les conditions au bord $v_0 = (a_0, \theta_0)$ et $v_1 = (a_1, \theta_1)$. $a_0 = (0, 0)$, $a_1 = (2, 0.5)$, $\theta_0 = 0^\circ, 15^\circ, 30^\circ$ (de haut en bas), $\theta_1 = -15^\circ, 0^\circ, 15^\circ$ (de gauche à droite), $\sigma = 1$, $\alpha = 2$. Les lignes sont les courbes (elasticæ de Mumford) optimisant la complétion. (D'après [156]).

Pour l'approximation \mathbb{V}_J où θ est petit, $\theta \sim \tan(\theta) = p$, et l'équation de Fokker-Planck devient

$$\frac{\partial P}{\partial t}(v, t) = - \left(\frac{\partial P}{\partial x}(v, t) + p \frac{\partial P}{\partial y}(v, t) \right) + \frac{\sigma^2}{2} \frac{\partial^2 P}{\partial p^2}(v, t) .$$

On peut calculer explicitement la fonction de Green, c'est-à-dire la solution ayant pour condition initiale la fonction de Dirac $P_0(v) = \delta_0$ à l'origine. Pour calculer un contour illusoire avec des conditions au bord $v_0 = (a_0, \theta_0)$ et $v_1 = (a_1, \theta_1)$, les auteurs considèrent deux processus directionnels, un processus forward partant de v_0 et un processus backward partant de v_1 , et calculent la probabilité de collision de ces deux marches aléatoires. La figure 102, due à Remco Duits et Markus van Almsick, donne un exemple des processus de complétion pour $a_0 = (0, 0)$, $a_1 = (2, 0.5)$, $\theta_0 = 0^\circ, 15^\circ, 30^\circ$ (de haut en bas), $\theta_1 = -15^\circ, 0^\circ, 15^\circ$ (de gauche à droite), $\sigma = 1$, $\alpha = 2$, où α paramétrise la loi de décroissance $\alpha e^{-\alpha t}$ du temps de parcours. Elle représente les distributions marginales, c'est-à-dire l'intégrale sur θ du produit des solutions forward/backward.

6.3. Autres groupes de Lie

Suite aux travaux de Yuri Sachkov et d'Igor Moiseev sur le groupe euclidien $SE(2)$, d'autres collaborateurs d'Andrei Agrachev à la SISSA de Trieste, en particulier Ugo Boscain et Francesco Rossi [67] ont analysé la géométrie sous-riemannienne d'autres groupes de Lie classiques de la géométrie élémentaire, en particulier la

sphère $\mathbb{S}^3 \simeq SU(2)$, le groupe des rotations $SO(3)$ et le groupe $SL(2)$. Comme il y insistent, il s’agissait là de résultats nouveaux car, à l’époque,

“To our knowledge, no global structure of the cut locus is known in sub-Riemannian geometry apart the one of the Heisenberg group.”

Leur principal résultat est le suivant.

Théorème (Boscain et Rossi, 2007).

- (i) Pour $SU(2) \simeq \mathbb{S}^3$ le cut locus (CL) est un grand cercle \mathbb{S}^1 moins l’origine.
- (ii) Pour $SO(3)$ le CL est un ensemble stratifié constitué du recollement en un point d’un plan projectif \mathbb{P}^2 et d’un cercle \mathbb{S}^1 épointé de l’origine.
- (iii) Pour $SL(2)$ le CL est un ensemble stratifié constitué du recollement en un point d’un plan \mathbb{R}^2 et d’un cercle \mathbb{S}^1 épointé de l’origine. ◇

Les auteurs donnent également des formules explicites pour la distance sous-riemannienne et la longueur des géodésiques. Par exemple, pour $SU(2)$ identifié à la sphère réelle \mathbb{S}^3 de \mathbb{C}^2 , la distance de (a, b) à l’origine est

$$\begin{cases} 2\sqrt{\arg(a)(2\pi - \arg(a))} & \text{si } b = 0 \\ \psi(a) & \text{si } b \neq 0 \end{cases}$$

où $\psi(a) = t$ est l’unique solution du système d’équations (c est une constante)

$$\begin{cases} \frac{-ct}{2} + \arctan\left(\frac{c}{\sqrt{1+c^2}} \tan\left(\frac{\sqrt{1+c^2}t}{2}\right)\right) = \arg(a) \\ \frac{\sin\left(\frac{\sqrt{1+c^2}t}{2}\right)}{\sqrt{1+c^2}} = \sqrt{1 - |a|^2} \\ t \in \left(0, \frac{2\pi}{\sqrt{1+c^2}}\right) \end{cases}$$

6.4. Caustiques génériques (Agrachev, Gauthier, Zakalyukin)

Agrachev, Gauthier et Zakalyukin [7] ont réussi à calculer des formes normales (c’est-à-dire des formes maximales simples à équivalence près) des caustiques génériques des structures de contact dans \mathbb{R}^3 . Il existe des systèmes de coordonnées par rapport auxquels l’hamiltonien H peut s’écrire comme une déformation de celui de Heisenberg

$$H(x_1, x_2, t, \xi_1^*, \xi_2^*, \theta^*) = \frac{1}{2} [(\xi_1^* + 2x_2\theta^*)^2 + (\xi_2^* - 2x_1\theta^*)^2]$$

à savoir

$$\begin{aligned} & H(x_1, x_2, t, \xi_1^*, \xi_2^*, \theta^*) \\ &= \frac{1}{2} [(\xi_1^* + \beta(\xi_1^*x_2 - \xi_2^*x_1) + 2\gamma x_2\theta^*)^2 + (\xi_2^* - \beta(\xi_1^*x_2 - \xi_2^*x_1) - 2\gamma x_1\theta^*)^2] \end{aligned}$$

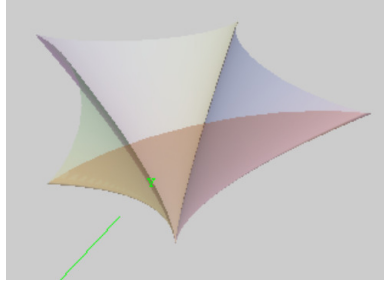


FIGURE 103. Le déploiement de la caustique Σ du groupe de Heisenberg (centre Z) dans le cas générique. Σ bifurque en quatre arêtes cuspidales. (D'après Agrachev et al. [7]).

où $\beta(x_1, x_2, t)$ et $\gamma(x_1, x_2, t)$ satisfont les conditions aux limites

$$\beta(0, 0, t) = 0, \quad \gamma(0, 0, t) = 1, \quad \nabla_{x,y}\gamma(0, 0, t) = 0.$$

Le cas Heisenberg correspond à $\beta = 0$ et $\gamma = 1$. En 0, $\beta = x_1 = x_2 = 0$ et $\gamma = 1$ et $H = \frac{1}{2}(\xi_1^{*2} + \xi_2^{*2})$, donc $H = \frac{1}{2}$ implique $\xi_1^{*2} + \xi_2^{*2} = 1$ et il existe par conséquent un angle φ tel que

$$\xi_1^* = \cos(\varphi), \quad \xi_2^* = \sin(\varphi).$$

Pour le groupe de Heisenberg \mathbb{H} , cas hautement non générique, la caustique Σ est l'axe des t qui est le centre Z de H . Pour le cas générique, dans les nouvelles variables $z = x_1 + ix_2$, $t = h^2$ et φ , elle est donnée au troisième ordre en h par les formules (A est une constante)

$$\begin{cases} z(h, \varphi) = -2Ah^3(3e^{-i\varphi} + e^{3i\varphi}) \\ t(h) = h^2. \end{cases}$$

L'axe des t se sépare en quatre arêtes de cusps. Nous représentons cette forme normale à la figure 103.

L'intersection des arêtes cuspidales avec le plan $t = h^2$ est donnée par $dz/d\varphi = 0$, c'est-à-dire

$$3ih^3(e^{3i\varphi} - e^{-i\varphi}) + 4h^4(e^{i(4\varphi+\varphi_0)} + e^{-i(2\varphi+\varphi_0)}) = 0.$$

Au premier ordre en h , on obtient les 4 cusps pour les valeurs suivantes de φ

$$\begin{cases} \varphi_1 = \frac{2}{3}\cos(\varphi_0)h \\ \varphi_2 = \frac{\pi}{2} - \frac{2}{3}\sin(\varphi_0)h \\ \varphi_3 = \pi - \frac{2}{3}\cos(\varphi_0)h \\ \varphi_4 = \frac{3\pi}{2} + \frac{2}{3}\sin(\varphi_0)h \end{cases}$$

Remarque. Nous avons rencontré plus haut dans la section 4.7 ces arêtes cuspidales pour $SE(2)$ mais avec d'autres coordonnées, les coordonnées de Jacobi utilisées par Yuri Sachkov.

6.5. Embarquement pour l'univers des jets

Après avoir un peu exploré les géodésiques sous-riemanniennes dans le cas particulier de nos modèles, nous allons maintenant, dans un nouveau chapitre, explorer de façon plus générale la géométrie des espaces de jets qui se trouve à la base de toute notre réflexion et constitue selon nous, en basse dimension, l'univers mathématique idoine pour les modèles d'architectures fonctionnelles corticales rétinotopiques.

Géométrie des distributions de jets

Nous avons vu dans les précédents chapitres les exemples de géométrie sous-riemannienne et de géodésiques qui nous intéressent en neurogéométrie. Fidèles à nos options pédagogiques qui consistent à monter du particulier au général plutôt que de descendre du général au particulier comme il en va dans la plupart des traités mathématiques, nous avons préféré traiter d'abord et directement de ces exemples concrets. Mais il est utile de contextualiser un peu ces résultats dans le contexte d'éléments de géométrie plus abstraits et plus généraux. Nous avons commencé à le faire dans la section 7 du chapitre 5 et nous nous proposons de continuer à le faire dans ce chapitre, mais encore très succinctement car le domaine est immense.

Comme toujours, il existe un certain nombre de propriétés structurelles qui découlent assez simplement des hypothèses générales et d'autres propriétés beaucoup plus subtiles et difficiles à démontrer.

1. Généralités sur les distributions d'hyperplans tangents

1.1. Distributions générales

On se donne une variété différentiable M de dimension n et une distribution régulière \mathfrak{D} d'hyperplans tangents de dimension $k < n$, c'est-à-dire en tout point $x \in M$, un k -hyperplan \mathfrak{D}_x de $T_x M$ variant différemment avec x .¹ En relation avec la théorie des connexions² les k -hyperplans \mathfrak{D}_x sont souvent appelés "horizontaux" et les courbes "admissibles", c'est-à-dire tangentes à \mathfrak{D}_x en chacun de leurs points x , "horizontales". Ce n'est pas toujours une qualification très heureuse car, par exemple pour la distribution de contact des 1-jets, les plans de contact dans $\mathbb{R}_{(x,y,p)}$ sont verticaux par rapport au plan de base (x, y) . Mais, comme nous l'avons vu dans la section 2 du chapitre 6, ils sont "horizontaux" par rapport au plan de base (x, p) qui fonctionne comme un espace des phases.

Une première partie de la théorie consiste à étudier de tels champs. C'est déjà un problème immense si l'on songe qu'en dimension 1 il s'agit de champs d'orientations

1. Plus haut, nous avons souvent noté ces distributions \mathcal{K} lorsqu'il s'agissait de distributions de contact.

2. Cf. chapitre 6.

et que si l'on remplace "orientation" par "vecteur tangent" (un vecteur tangent pouvant être nul) il s'agit des champs de vecteurs tangents (sans singularités) autrement dit des équations différentielles ordinaires ! Souvent la distribution \mathfrak{D} est définie explicitement par un système de Pfaff, c'est-à-dire par l'annulation de $(n - k)$ 1-formes explicites ω_j , ce qui permet de faire des calculs. De toute façon, étant donnée une distribution \mathfrak{D} , on peut toujours considérer les formes différentielles qui s'annulent sur \mathfrak{D} et en particulier la distribution \mathfrak{D}^\perp dans le fibré cotangent T^*M constitué des 1-formes différentielles qui s'annulent sur \mathfrak{D} . Il est le dual du quotient TM/\mathfrak{D} .

Une seconde partie de la théorie consiste à se donner une métrique $g_{\mathfrak{D}}$ sur \mathfrak{D} , c'est-à-dire une forme bilinéaire symétrique définie positive $g_{\mathfrak{D}}(x)$ sur les \mathfrak{D}_x qui varie différentiablement avec x . Là aussi c'est un problème immense puisque, pour $k = n$ au lieu de $k < n$, il recouvre toute la géométrie riemannienne !

1.2. Courbure d'une distribution

On peut généraliser aux distributions quelconques ce que nous avons vu à la section 3 du chapitre 6 pour les connexions, leurs 1-formes ω et leurs distributions noyau $\mathfrak{D} = \ker(\omega)$.

On considère l'algèbre de Lie des champs de vecteurs X sur \mathfrak{D} (i.e. "horizontaux" : $X(x) \in \mathfrak{D}_x$ en tout x , notation $X \in \mathfrak{D}$).³ La courbure Ω de \mathfrak{D} est définie comme le crochet de Lie modulo \mathfrak{D} de ces $X \in \mathfrak{D}$. C'est la 2-forme sur $\mathfrak{D} \subset TM$ à valeurs dans le quotient TM/\mathfrak{D} ,

$$\kappa : \Lambda^2 \mathfrak{D} \rightarrow TM/\mathfrak{D} ,$$

qui vaut $\kappa(X, Y) = -[X, Y]$ modulo \mathfrak{D} pour $X, Y \in \mathfrak{D}$.⁴ Elle mesure la *non-intégrabilité* de \mathfrak{D} et permet de reformuler l'opposition entre les conditions de Frobenius et d'Hörmander déjà explicitées à la section 2.8 du chapitre 3. L'intégrabilité signifie en effet que $[X, Y] \in \mathfrak{D}$ pour $X, Y \in \mathfrak{D}$ (condition de Frobenius) et par conséquent \mathfrak{D} est intégrable si et seulement si elle est de courbure nulle.

On notera que, comme

$$[X, Y]_i = \sum_j \left(X_j \frac{\partial Y_i}{\partial x_j} - Y_j \frac{\partial X_i}{\partial x_j} \right)$$

3. Évidemment, dans le cas des groupes de Lie G on peut, comme nous l'avons constamment fait, considérer les champs de vecteurs tangents invariants (à gauche) et ramener leur algèbre de Lie à une structure d'algèbre de Lie sur $T_e G$. Ici, M n'est pas supposée être un groupe et la distribution n'est pas engendrée par translations à partir d'un k -hyperplan d'un $T_x M$ privilégié. La situation est donc plus complexe.

4. Rappelons que $\Lambda^2 \mathfrak{D}$ est l'algèbre antisymétrique des 2-vecteurs.

lorsque l'on identifie les champs de vecteurs X à des opérateurs de dérivation $X = \sum_i X_i \frac{\partial}{\partial x_i}$ des fonctions $C^\infty f$ sur M (cf. section 2 du chapitre 5), on a

$$[fX, Y] = f[X, Y] - Y(f)X$$

et que donc

$$[fX, Y] \equiv f[X, Y] \text{ modulo } \mathfrak{D}$$

si $X \in \mathfrak{D}$ (et idem pour $[X, fY] \equiv f[X, Y]$ modulo \mathfrak{D}), ce qui montre que $\kappa_x(X, Y)$ ne dépend que des valeurs *ponctuelles* $X(x)$ et $Y(x)$ de X et Y et non pas de leur structure locale en x . La courbure est l'un des plus simples invariants algébriques de \mathfrak{D} .

Si l'on passe au dual avec la distribution \mathfrak{D}^\perp de T^*M duale de TM/\mathfrak{D} , la courbure $\kappa : \Lambda^2 \mathfrak{D} \rightarrow TM/\mathfrak{D}$ se dualise en l'application κ^* de \mathfrak{D}^\perp dans le dual $(\Lambda^2 \mathfrak{D})^*$ de $\Lambda^2 \mathfrak{D}$ qui associe à une 1-forme ω s'annulant sur \mathfrak{D} la 2-forme

$$\kappa^*(\omega)(X, Y) = -\omega([X, Y]) \quad (X, Y \in \mathfrak{D}) .$$

En effet, d'après la formule (D) de la section 7.2.7 du chapitre 2

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]) .$$

Donc, si $X, Y \in \mathfrak{D}$, $\omega(X) = \omega(Y) = 0$ et par conséquent $-\omega([X, Y]) = d\omega(X, Y)$. Autrement dit, $\kappa^*(\omega)(X, Y) = -\omega([X, Y])$ est simplement la restriction de $d\omega$ à \mathfrak{D} qui est aussi la restriction de la courbure Ω de ω (au sens de la section 3 du chapitre 6) à \mathfrak{D} .⁵

1.3. Typologie des distributions

Dans la section 1 du chapitre 7 consacrée aux algèbres de Lie \mathcal{G} (en particulier nilpotentes) nous avons vu comment s'introduisent naturellement des "suites centrales descendantes" et des "drapeaux" construits à partir de commutateurs itérés de \mathcal{G} . Il en va de même pour les distributions. Entre les extrêmes constitués par la distribution donnée \mathfrak{D} de dimension k et TM (distribution triviale de dimension maximale n) on peut construire de différentes façons des suites de distributions intermédiaires en itérant des commutateurs bien choisis.

La première suite est le drapeau de Lie des

$$\mathfrak{D}^0 = \mathfrak{D}, \mathfrak{D}^j = \mathfrak{D}^{j-1} + [\mathfrak{D}, \mathfrak{D}^{j-1}] ,$$

qui satisfait

$$[\mathfrak{D}^j, \mathfrak{D}^s] \subset \mathfrak{D}^{j+s} .$$

5. Lorsque \mathfrak{D} est de codimension 1 et le noyau $\ker(\omega)$ d'une 1-forme, la courbure Ω de ω au sens de la section 3 du chapitre 6 est donnée par la formule de Maurer-Cartan $\Omega = d\omega + \frac{1}{2}[\omega \wedge \omega]$ et redonne, pour les vecteurs $X, Y \in \mathfrak{D}$, $\Omega(X, Y) = d\omega(X, Y) = -\omega([X, Y]) = \kappa^*(\omega)(X, Y)$.

Si donc $\{X_i\}$, $i = 1, \dots, k$, est localement une base de \mathfrak{D} au voisinage de $x \in M$, \mathfrak{D}^j est engendré par les crochets $[X_{j_1}, \dots, [X_{j_{r-1}}, X_{j_r}]]$ des r -uples de X_i avec $r \leq j$. Si la distribution \mathfrak{D} est *complètement non holonome*, autrement dit si le drapeau de Lie des \mathfrak{D}^j va de $\mathfrak{D} = \mathfrak{D}^1$ jusqu'à $TM = \mathfrak{D}^r$ pour un certain r , le r minimal s'appelle alors le "degré de non holonomie" ou le *niveau* ("step") de \mathfrak{D} . Si de plus \mathfrak{D} et son drapeau sont *réguliers*, c'est-à-dire que tous les \mathfrak{D}_x^j sont de dimension *constante*, alors la suite des dimensions $v_j = \dim(\mathfrak{D}^j)$ (avec $v_1 = k$ et $v_r = n$) est bien définie et s'appelle alors le (petit) "vecteur de croissance" \mathfrak{gv} ("growth vector") de \mathfrak{D} (nous allons préciser cette notion plus bas). Par exemple le groupe de Heisenberg est de vecteur de croissance $(2, 3)$. Les structures de contact \mathcal{K} sur une variété de contact M de dimension (nécessairement) impaire $2r - 1$, structures sur lesquelles nous allons revenir à la section 2, sont de vecteur de croissance $\mathfrak{gv} = (2(r - 1), 2r - 1)$.

Il est facile de voir qu'il existe de fortes contraintes dimensionnelles générales sur les vecteurs de croissance des distributions régulières (k, n) et l'on comprend intuitivement qu'il est naturel de chercher des *formes normales locales* (cf. plus bas section 1.5) c'est-à-dire des exemples les plus simples possibles auxquels on peut se ramener par un difféomorphisme local (un changement de coordonnées locales). Évidemment, cette structure locale maximale simplifiée ne dit rien de la complexité globale de la distribution \mathfrak{D} . Il suffit de penser que si $k = 1$, \mathfrak{D} est un champ de droites et que de tels champs peuvent être d'une complexité inouïe, comme celle des champs de vecteurs partout non nuls. De même si $k = 2$ et si \mathfrak{D} est intégrable, ses intégrales sont des surfaces régulières dans M et celles-ci peuvent aussi être d'une complexité inouïe. En revanche si M est un groupe de Lie G alors \mathfrak{D} s'obtient par translation à partir d'un sous-espace de l'algèbre de Lie \mathcal{G} et la complexité globale se restreint considérablement.

Donnons quelques exemples.

1. Dans le cas $k = 2$, $n = 3$, $\mathfrak{gv} = (2, 3)$, il faut que $[X_1, X_2] = X_3$ avec $[X_1, X_3]$ et $[X_2, X_3]$ dans $\text{Span}\{X_1, X_2, X_3\} = TM$ avec, d'après Jacobi,

$$[[X_2, X_3], X_1] + [[X_3, X_1], X_2] = 0 .$$

Cela donne localement la structure de Heisenberg.

2. Dans le cas $k = 2$, $n = 4$, $[X_1, X_2] = X_3$ et nécessairement $\mathfrak{gv} = (2, 3, 4)$ avec les crochets entre les trois X_i engendrant un quatrième vecteur indépendant X_4 , tous les crochets successifs étant contraints à rester dans

$$\text{Span}\{X_1, X_2, X_3, X_4\} = TM .$$

Cela donne localement la structure d'Engel.

3. Mais dès que $k = 2$, $n = 5$ on voit apparaître des contraintes plus subtiles. En effet comme la dimension est assez grande, on a la possibilité d'avoir

$[X_1, X_3] = X_4$ et $[X_2, X_3] = X_5$ avec la contrainte de Jacobi

$$[X_5, X_1] - [X_4, X_2] = 0$$

et ce sera même le cas *génériquement*. En contraignant les crochets successifs à rester dans $\text{Span}\{X_1, X_2, X_3, X_4, X_5\} = TM$ on obtient un $\mathfrak{gv} = (2, 3, 5)$. Une forme normale a été calculée par Cartan :

$$\begin{aligned} X_1 &= \partial_{x_1}, \quad X_2 = \partial_{x_2} + x_1\partial_{x_3} + \frac{1}{2}x_1^2\partial_{x_4} + x_1x_2\partial_{x_5}, \\ [X_1, X_2] &= X_3 = \partial_{x_3} + x_1\partial_{x_4} + x_2\partial_{x_5}, \\ [X_1, X_3] &= X_4 = \partial_{x_4}, \quad [X_2, X_3] = X_5 = \partial_{x_5}, \quad [X_1, X_4] = 0, \quad [X_1, X_5] = 0, \\ [X_2, X_4] &= 0, \quad [X_2, X_5] = 0, \quad [X_3, X_4] = 0, \quad [X_3, X_5] = 0, \quad [X_4, X_5] = 0. \end{aligned}$$

La matrice 5×5 de passage des ∂_{x_i} aux X_i étant de déterminant 1, les X_i sont linéairement indépendants et donc $\mathfrak{gv} = (2, 3, 5)$.

4. Un exemple de $k = 2$, $n = 5$ de $\mathfrak{gv} = (2, 3, 4, 5)$ est

$$X_1 = \partial_{x_1}, \quad X_2 = \partial_{x_2} + x_1\partial_{x_3} + \frac{1}{2}x_1^2\partial_{x_4} + \frac{1}{6}x_1^3\partial_{x_5},$$

le changement du coefficient de ∂_{x_5} faisant que les cinq X_i ne sont plus linéairement indépendants et qu'il faut un crochet supplémentaire pour arriver à TM . On trouve

$$\begin{aligned} [X_1, X_2] &= X_3 = \partial_{x_3} + x_1\partial_{x_4} + \frac{1}{2}x_1^2\partial_{x_5}, \quad [X_1, X_3] = X_4 = \partial_{x_4} + x_1\partial_{x_5}, \\ [X_2, X_3] &= X_5 = 0, \quad [X_1, X_4] = X'_5 = \partial_{x_5}, \quad [X_1, X'_5] = 0, \quad [X_2, X_4] = 0, \\ [X_2, X'_5] &= 0, \quad [X_3, X_4] = 0, \quad [X_3, X'_5] = 0, \quad [X_4, X'_5] = 0. \end{aligned}$$

La matrice 5×5 de passage des ∂_{x_i} aux X_1, X_2, X_3, X_4, X'_5 étant encore de déterminant 1, les X_i sont linéairement indépendants et donc $\mathfrak{gv} = (2, 3, 4, 5)$. Une telle structure est en partie obligatoire. En effet comme les cinq vecteurs $\{X_1, X_2, X_3, X_4, X_5\}$ ne sont pas indépendants puisque $\text{Span}\{X_1, \dots, X_5\}$ est de dimension 4 par hypothèse, on peut, par changement linéaire de coordonnées, se ramener à $X_5 = [X_2, X_3] = 0$. On pose alors

$$X'_5 = [X_1, X_4] = [X_1, [X_1, [X_1, X_2]]].$$

Par Jacobi

$$[X_2, [X_1, X_3]] + [X_1, [X_3, X_2]] + [X_3, [X_2, X_1]] = 0$$

et comme $[[X_1, X_2], [X_2, X_1]] = 0$, on a

$$[X_2, X_4] = -[X_1, [X_3, X_2]] = 0.$$

Jacobi implique également

$$[X_2, X_5] = -[X_3, X_4].$$

En jouant sur les degrés de liberté disponibles on peut se ramener à

$$[X_1, X_5] = 0, [X_2, X_5] = 0, [X_3, X_4] = 0, [X_3, X'_5] = 0, [X_4, X_5] = 0 .$$

Les distributions de $\mathfrak{gv} = (2, 3, 4, 5)$ où \mathfrak{gv} est une suite d'entiers *successifs* sont (sous certaines conditions supplémentaires) des distributions dites *de Goursat* (avec $k = 2, n = 5$). Chaque $\mathfrak{D}^j = \mathfrak{D}^{j-1} + [\mathfrak{D}, \mathfrak{D}^{j-1}]$ y rajoute exactement un vecteur supplémentaire linéairement indépendant à \mathfrak{D}^{j-1} . Elles sont *non génériques* dans l'espace des distributions (k, n) et satisfont des contraintes exceptionnelles. Nous allons y revenir dans la section 3.

La suite croissante du drapeau de Lie $\mathfrak{D}^0 = \mathfrak{D} \subset \mathfrak{D}^1 \subset \dots \subset \mathfrak{D}^r$ est en chaque point $x \in M$ ce que l'on appelle une *filtration* de sous-algèbres de Lie de $T_x M$, mais ce n'est pas en tant que telle une *graduation*, c'est-à-dire une somme directe de sous-algèbres \mathfrak{A}_i telles que le crochet de deux éléments de degrés respectifs i et j soit $i + j$. Toutefois, elle engendre en chaque point x une graduation

$$gr(\mathfrak{D}) = \mathfrak{D} \oplus \mathfrak{D}^1/\mathfrak{D} \oplus \dots \oplus \mathfrak{D}^r/\mathfrak{D}^{r-1}$$

sur laquelle nous allons revenir plus bas section 8.3 à propos des *groupes de Carnot*.

Lorsque les $v_j(x)$ ne sont pas localement constants, il peut s'introduire alors des points *singuliers* de la distribution \mathfrak{D} . L'exemple sans doute le plus connu est la *distribution de Martinet* sur \mathbb{R}^3 où \mathfrak{D} est le noyau de la 1-forme

$$\omega = dy - p^2 dx$$

(et non pas $\omega = dy - p dx$ comme pour Heisenberg) avec

$$\omega \wedge d\omega = -2p dx \wedge dy \wedge dp$$

qui est nulle pour $y = 0$. Les plans \mathfrak{D}_x sont engendrés par les vecteurs

$$\{X = \partial_x + p^2 \partial_y, Y = \partial_p\} .$$

On a $[X, Y] = -2p \partial_y$ et donc si $p = 0$, $[X, Y] = 0$ et il faut un autre crochet pour engendrer \mathbb{R}^3 , en l'occurrence

$$[[X, Y], Y] = 2 \partial_y .$$

En dehors du plan $p = 0$ le vecteur de croissance est $(2, 3)$ et ω est de contact, mais sur le plan $p = 0$ il est $(2, 2, 3)$ et ω n'est plus de contact.

Avant de préciser les choses, disons d'abord quelques mots sur la méthode géométrique utilisée pour aboutir à l'existence de *formes normales*. La distribution de Martinet y reviendra à titre d'exemple.

1.4. Le principe de Thom : stratifications et transversalité

Dans beaucoup de travaux sur les formes normales en géométrie différentielle, la méthode est d'abord géométrique. Elle consiste

1. à utiliser d'abord des descripteurs géométriques *intrinsèques* de la structure locale Σ de la situation au point considéré,
2. à trouver, mais seulement dans un second temps, des *coordonnées*
 - (a) adaptées à Σ ,
 - (b) mais par ailleurs arbitraires du moment qu'elles respectent cette contrainte d'adaptation,
3. à effectuer des développements de Taylor avec ces coordonnées,
4. à chercher quels termes de ces développements on peut éliminer par des difféomorphismes de Σ respectant son type de structure.

En général les descripteurs de Σ consisteront à partir d'un certain espace, par exemple celui $\mathcal{E}_{\mathfrak{D}}$ des distributions \mathfrak{D} de type (k, n) sur M , et à introduire des conditions successives réalisées sur des sous-espaces de codimensions de plus en plus grandes. Si \mathfrak{D}_0 est de codimension c , son orbite $\widetilde{\mathfrak{D}}_0$ par les difféomorphismes de M respectant son type de structure admettra, du moins dans les cas simples, une section transverse $\mathfrak{D}_0^{\#}$ de dimension c et, dans $\mathfrak{D}_0^{\#}$, les conditions successives s'exprimeront géométriquement par une *stratification* Σ , c'est-à-dire des sous-espaces de $\mathfrak{D}_0^{\#}$ de dimensions décroissantes (donc de codimensions croissantes dans $\mathfrak{D}_0^{\#}$) avec des propriétés d'incidence, les strates de dimensions j étant dans l'adhérence de strates de dimension $> j$. Cet empilement de lieux singuliers exprime les possibilités de contraindre la structure des \mathfrak{D} .⁶

Si l'on considère maintenant des champs dans TM , on peut analyser leurs relations d'incidence avec les strates de Σ : elles sont essentiellement des conditions de transversalité ou de tangence. On utilise alors un *principe de transversalité*, profond à la fois mathématiquement et philosophiquement, que l'on pourrait appeler "principe de Thom" : les singularités sont la conséquence de défauts de transversalité dans les relations d'incidence avec des strates d'ensembles stratifiés décrivant les structures géométriques considérées. La transversalité induit la régularité et la stabilité structurelle à partir d'un certain niveau alors que la non-transversalité augmente l'instabilité structurelle et le niveau de singularité. Un exemple typique de cette méthode est la classification des singularités d'applications différentiables par Thom, Boardman, Mather, Arnold *et al.* (cf. notre compilation déjà citée [412]).

6. La notion de stratification est assez intuitive mais son traitement mathématique rigoureux est très délicat. Les idées fondatrices sont en grande partie dues à Hassler Whitney et René Thom, en particulier dans le texte princeps de ce dernier "Ensembles et morphismes stratifiés" [529], texte commenté par Bernard Teissier et David Trotman dans [526].

1.5. Formes normales

Un certain nombre de théorèmes fondamentaux expliquent que, sous des conditions de structure assez générales, une distribution \mathfrak{D} peut être mise *localement* sous *forme normale*, au sens où il existe un changement différentiable de coordonnées locales, autrement dit un difféomorphisme local, qui préserve son type et ramène localement ses équations à une forme typique universelle. Ces théorèmes *d'unicité locale* d'une structure d'un certain type sont essentiels pour comprendre la géométrie locale des distributions car les formes normales permettent de faire des calculs valables localement qui sont *explicites, universels* et aussi simples que possibles. Ils sont également philosophiquement profonds car ils montrent que l'usage d'un modèle algébrique simple est empiriquement justifié dans la mesure où celui-ci est structurellement stable et universel.

Il existe en géométrie différentielle énormément de théorèmes de réduction de structures à des formes normales. Ils consistent, répétons-le, à faire des développements locaux en séries de Taylor des descripteurs de la structure puis à voir quels sont les termes de ces développements que l'on peut éliminer par des changements différentiables de coordonnées locales.

Exemple 1 (champs de vecteurs). Par exemple si X est un champ de vecteurs différentiable sur une variété M de dimension n et si $X(x_0) \neq 0$ il existe des coordonnées locales (x_1, \dots, x_n) au voisinage de x_0 dans lesquelles X est le champ constant $X(x) = \partial_x$.

Exemple 2 (fonctions différentiables). Dans la théorie des fonctions différentiables $f : M \rightarrow \mathbb{R}$, il existe une foule de théorèmes de ce genre. Ils se démontrent en deux temps. D'abord on essaye de réduire au maximum le degré des termes des développements de Taylor au moyen de changements de coordonnées. Si l'on peut arriver jusqu'au degré k mais pas jusqu'au degré $k-1$, on dit que, localement en x_0 , f est *déterminé par son jet d'ordre k* . Ensuite on essaye de simplifier au maximum le jet d'ordre k par des changements, cette fois-ci algébriques, de coordonnées. Par exemple si x_0 est un point *régulier* de M , c'est-à-dire un point où le vecteur ∇f des dérivées partielles

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$

n'est pas nul, alors l'image réciproque de $f(x_0)$ est localement en x_0 une hypersurface S_0 et les images réciproques de $f(x_0) + \varepsilon$ où ε varie dans un assez petit voisinage de 0 sont des hypersurfaces "parallèles" S_ε formant localement en x_0 un feuilletage de codimension 1 traversé par une courbe transverse x_ε passant par x_0 . On peut adapter des coordonnées locales à ce feuilletage de façon à ce que f devienne tout simplement la fonction $f(x) = x_1$. Cette "linéarisation" signifie que si l'on prend le

développement de Taylor de f en x_0 ,

$$f(x) = f(x_0) + \sum_i x_i \frac{\partial f}{\partial x_i}(x_0) + \sum_{i,j} x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) + \dots$$

on peut éliminer tous les termes de degré ≥ 2 , ne garder que les termes linéaires, f étant donc déterminée par son jet d'ordre 1 localement en x_0 . À partir d'un système de coordonnées locales où f est égale à son 1-jet

$$f(x) = f(x_0) + \sum_i x_i \frac{\partial f}{\partial x_i}(x_0) ,$$

on peut alors, par un changement simplement linéaire de coordonnées à la source M et une translation de $f(x_0)$ au but \mathbb{R} , poser $f(x_0) = 0$ et $\nabla f(x_0) = (1, 0, \dots, 0)$ et l'on obtient $f(x) = x_1$.

De même, si x est un point *singulier* (on dit aussi point critique) où $\nabla f = 0$ mais où le hessien, à savoir la matrice symétrique $Hf = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)$ des dérivées secondes⁷, n'est pas dégénéré en x_0 (on parle dans ce cas de point critique non dégénéré), alors un célèbre théorème, le *théorème de Morse*, dit qu'il existe un système de coordonnées locales (x_1, \dots, x_n) en x_0 (que l'on peut prendre pour 0) tel que

$$f(x) = f(0) - (x_1^2 + \dots + x_k^2) + x_{k+1}^2 + \dots + x_n^2 .$$

Le nombre k est défini *intrinsèquement* (indépendamment des coordonnées particulières choisies) et s'appelle *l'indice* de x_0 . En raison de cette forme normale, les points critiques non dégénérés sont aussi dits *quadratiques*. Le théorème de Morse implique qu'en un point singulier non dégénéré, une fonction $f : M \rightarrow \mathbb{R}$ est déterminée par son jet d'ordre 2. Il se démontre en montrant d'abord que l'on peut éliminer tous les termes du développement de Taylor de degré > 2 . Dans un tel système de coordonnées, f est égale à son jet d'ordre 2 et donc, puisque $\nabla f = 0$, $f(x) = f(0) + x^t Hf(0) x$ (x est le vecteur colonne des x_i et x^t leur vecteur ligne). Il suffit dès lors, par un changement linéaire de coordonnées, de ramener la forme quadratique non dégénérée $x^t H x$ à ses axes principaux pour réduire f à sa forme normale de Morse.

Les points critiques quadratiques sont structurellement stables. Ce n'est pas le cas des points critiques dégénérés qui sont des singularités pouvant de "déployer" suivant un certain nombre de paramètres de stabilisation. Nous avons rencontré de tels déploiements dans la section 5.11. du chapitre 5 du Vol I dans le modèle de Swindale couplant les cartes d'orientations de $V1$ et leurs pinwheels avec les cartes de direction de mouvement.⁸

7. Nous avons déjà rencontré les hessiens dans la section 6.1 du chapitre 12.

8. Cf. également notre article [444].

La théorie des singularités et de leurs déploiements est un immense sujet. Selon nous, l'une des meilleures introductions se trouve dans les articles d'Alain Chenciner [110] et [111]. Le lecteur intéressé trouvera aussi de nombreuses précisions techniques à son sujet dans notre compilation [412] de la théorie de Thom-Mather et sa bibliographie.

1.6. Classification des germes de 1-formes sur \mathbb{R}^3

1.6.1. Les cas simples.

Avant que d'en venir à des théorèmes de structure un peu généraux, considérons comme autre exemple celui des germes en 0 de 1-formes sur $M = \mathbb{R}^3$. Soit \mathfrak{D} la distribution noyau de ω . Une fois choisies des coordonnées (x, y, z) , ω s'écrit

$$\omega = f(x, y, z) dx + g(x, y, z) dy + h(x, y, z) dz .$$

On peut supposer $\omega(0) \neq 0$ et donc que l'une des fonctions C^∞ f, g, h est $\neq 0$ en 0. Une première adaptation des coordonnées consiste à les choisir de façon à ce que $g(0) \neq 0$ et à se restreindre à un voisinage de 0 où g est partout $\neq 0$. En divisant par g , on peut alors se ramener au cas

$$\omega = dy + f(x, y, z) dx + h(x, y, z) dz .$$

Par changement linéaire de coordonnées, on peut supposer $f(0) = h(0) = 0$. On essaye alors de changer la coordonnée y de façon à éliminer h (mais pas f car sinon $\omega = dy$ est déjà sous forme minimale). Si l'on prend $y = y_1 + \varphi(x, y_1, z)$ avec φ commençant à l'ordre 2 en 0, le déterminant jacobien $(x, y, z) \rightarrow (x_1 = x, y_1, z_1 = z)$ est $1 + \frac{\partial \varphi}{\partial y_1}$ (et donc 1 en 0) et il faut seulement satisfaire la deuxième contrainte $1 + \frac{\partial \varphi}{\partial y_1} \neq 0$, φ pouvant être par ailleurs quelconque. La 1-forme ω s'écrit alors

$$\begin{aligned} \omega &= dy + f(x, y, z) dx + h(x, y, z) dz \\ &= dy_1 + \frac{\partial \varphi}{\partial x} dx + \frac{\partial \varphi}{\partial y_1} dy_1 + \frac{\partial \varphi}{\partial z} dz + f(x, y, z) dx + h(x, y, z) dz \\ &= \left(1 + \frac{\partial \varphi}{\partial y_1}\right) dy_1 + \left(\frac{\partial \varphi}{\partial x} + f\right) dx + \left(\frac{\partial \varphi}{\partial z} + h\right) dz . \end{aligned}$$

On choisit φ (troisième contrainte) de façon à ce que $\frac{\partial \varphi}{\partial z} + h = 0$ et on divise par $1 + \frac{\partial \varphi}{\partial y_1}$. On se ramène ainsi à

$$\omega = dy + f(x, y, z) dx .$$

Un vecteur tangent $X = (\xi, \eta, \pi) \in T_m M$ est dans \mathfrak{D} si et seulement si $\eta + f\xi = 0$, i.e.

$$X = (\xi, \eta = -f\xi, \pi) .$$

C'est à partir de là que l'on doit utiliser la méthodologie des stratifications et considérer la 3-forme ($dV = dx \wedge dy \wedge dz$ étant une forme volume) :

$$\begin{aligned}\omega \wedge d\omega &= (dy + f dx) \wedge (d(f) \wedge dx) = \\ &= (dy + f dx) \wedge ((\partial_x f) dx + (\partial_y f) dy + (\partial_z f) dz) \wedge dx \\ &= (\partial_z f) dy \wedge dz \wedge dx = F dV .\end{aligned}$$

Il est intéressant de calculer le champ de vecteurs tangents spécial Z tel que le produit intérieur $i_Z dV$ de la 3-forme volume dV par X soit la 2-forme $\omega \wedge dF$. Cela signifie que si $Z = (\xi_Z, \eta_Z, \pi_Z)$ on a

$$\begin{aligned}\omega \wedge dF &= (dy + f dx) \wedge ((\partial_x F) dx + (\partial_y F) dy + (\partial_z F) dz) \\ &= (\partial_x F) dy \wedge dx + (\partial_z F) dy \wedge dz + f (\partial_y F) dx \wedge dy + f (\partial_z F) dx \wedge dz \\ &= (f (\partial_y F) - \partial_x F) dx \wedge dy + f (\partial_z F) dx \wedge dz + (\partial_z F) dy \wedge dz \\ i_Z dV &= \pi_Z dx \wedge dy - \eta_Z dx \wedge dz + \xi_Z dy \wedge dz .\end{aligned}$$

Il faut donc

$$\xi_Z = \partial_z F, \eta_Z = -f (\partial_z F), \pi_Z = f (\partial_y F) - \partial_x F .$$

On vérifie que le produit intérieur $i_Z dF = dF(Z)$ est $\equiv 0$ et que donc Z est partout tangent aux surfaces de niveau de $F = \partial_z f$ (si F n'est pas constante car si F est constante $Z \equiv 0$). En effet

$$\begin{aligned}i_Z dF = dF(Z) &= (\partial_x F) \xi_Z + (\partial_y F) \eta_Z + (\partial_z F) \pi_Z \\ &= (\partial_x F) (\partial_z F) - f (\partial_y F) (\partial_z F) + (\partial_z F) (f (\partial_y F) - \partial_x F) \equiv 0 .\end{aligned}$$

La valeur de $\omega = dy + f dx$ sur Z est

$$\omega(Z) = \eta_Z + f \xi_Z = -f (\partial_z F) + f (\partial_z F) = 0$$

et donc $Z \in \mathfrak{D}$. Inclus dans \mathfrak{D} et tangent aux surfaces de niveau de F , Z est le champ caractéristique de \mathfrak{D} et de ω .

- (i) Si $\partial_z f(0) = F(0) \neq 0$, i.e. si f varie au premier ordre le long de l'axe vertical des z , alors $\omega \wedge d\omega(0) \neq 0$ et ω est de contact en 0. C'est une condition *générique* qui signifie que si \mathfrak{D} est engendrée par les deux champs X et Y , les trois vecteurs X , Y et $Z = [X, Y]$ sont linéairement indépendants. En prenant alors $-f$ comme coordonnée z ,⁹ on obtient la forme normale de Darboux $dy - p dx$ pour Heisenberg.¹⁰ Les \mathfrak{D} de Darboux forment un ouvert dense dans l'espace des \mathfrak{D} et sont déterminées par leur jet d'ordre 1.

9. z est renommée p pour l'occasion.

10. On voit ainsi comment notre premier modèle \mathbb{V}_J se situe dans l'univers des possibles.

- (ii) Si en revanche $\partial_z f(0) = F(0) = 0$, alors $\omega \wedge d\omega(0) = 0$. Si $F \equiv 0$, $\omega \wedge d\omega \equiv 0$ et \mathfrak{D} est intégrable (condition d'intégrabilité de Frobenius) et induit un feuilletage de $M = \mathbb{R}^3$ par des surfaces. Cette situation est localement triviale et on peut ramener \mathfrak{D} à un champ de plans parallèles. La situation intéressante est quand le lieu où $\partial_z f = F = 0$ est un lieu singulier S pour ω . Comme F n'est pas $\equiv 0$, S est une surface qui est *régulière* en 0 si $dF(0) \neq 0$, autrement dit si les dérivées secondes $\partial_{z^2}^2 f(0)$, $\partial_{zy}^2 f(0)$ et $\partial_{zx}^2 f(0)$ ne sont pas toutes les trois nulles; les vecteurs tangents $X = (\xi, \eta, \pi)$ à S satisfont $dF(X) = 0$, i.e.

$$\xi \partial_{zx}^2 f + \eta \partial_{zy}^2 f + \pi \partial_{z^2}^2 f = 0 .$$

Si en plus $(\omega \wedge dF)_0 \neq 0$,¹¹ autrement dit si les dérivées secondes $\partial_{z^2}^2 f(0)$, $\partial_{zx}^2 f(0)$ ne sont pas toutes les deux nulles,¹² S est *transverse* à \mathfrak{D}_0 en 0 et donc *partout* transverse dans un voisinage de 0 car la transversalité est une condition *ouverte*. En effet, si $X, Y \in T_0 M$,

$$(\omega \wedge dF)_0(X, Y) = \omega_0(X) dF_0(Y) - \omega_0(Y) dF_0(X) .$$

Donc $(\omega \wedge dF)_0 = 0$ implique que si Y est tangent à S en 0 ($dF_0(Y) = 0$) alors $\omega_0(Y) dF_0(X)$ pour tout X et donc, si dF_0 n'est pas $\equiv 0$, on a $\omega_0(Y) = 0$, autrement dit Y est tangent à \mathfrak{D}_0 et S et \mathfrak{D}_0 sont tangentes en 0. Si ce n'est pas le cas et si par exemple $\partial_{z^2}^2 f(0) \neq 0$, f est équivalente à $-z^2$ et (en renommant z par p) on obtient la forme normale de Martinet $\omega = dy - p^2 dx$. Les \mathfrak{D} de Martinet forment un sous espace de codimension 1 dans l'espace des \mathfrak{D} et sont déterminées par leur jet d'ordre 2.

Dans ce dernier cas, \mathfrak{D} induit sur S ¹³ un *feuilletage* par des courbes intégrales *régulières* γ . On sait que le champ Z est tangent aux surfaces de niveau de F et comme S est la surface de niveau $F = 0$, Z est tangent à S . On sait aussi que Z est dans \mathfrak{D} . Les courbes γ sont donc les intégrales du champ caractéristique Z .¹⁴

11. Dans cette section, pour éviter des confusions de notations, nous notons $m = \{x, y, z\}$ (et non plus x) les points de M et, pour simplifier, nous notons C_m plutôt que $C(m)$. les valeurs des champs C en un point m .

12.

$$\begin{aligned} \omega \wedge dF &= (dy + f dx) \wedge ((\partial_{zx}^2 f) dx + (\partial_{zy}^2 f) dy + (\partial_{z^2}^2 f) dz) \\ &= (\partial_{zx}^2 f) dy \wedge dx + (\partial_{z^2}^2 f) dy \wedge dz + f (\partial_{zy}^2 f) dx \wedge dy + f (\partial_{z^2}^2 f) dx \wedge dz , \\ (\omega \wedge dF)_0 &= (\partial_{zx}^2 f)(0) dy \wedge dx + (\partial_{z^2}^2 f)(0) dy \wedge dz \text{ car } f(0) = 0 . \end{aligned}$$

13. S est une sous-variété fixe alors que \mathfrak{D} est un champ de plans tangents qui varie avec le point.

14. Reprenons le calcul à la main. Les vecteurs tangents $X = (\xi, \eta, \pi)$ qui sont tangents à la fois à S et \mathfrak{D} en $m \in M$ satisfont trois conditions : (a) $\partial_z f = F = 0$ ($m \in S$), (b) $\eta = -f\xi$ ($X \in \mathfrak{D}_m$), (c) $\xi \partial_{zx}^2 f + \eta \partial_{zy}^2 f + \pi \partial_{z^2}^2 f = 0$ ($X \in T_m S$). Si $\partial_{z^2}^2 f \neq 0$, on a $\pi = \frac{\xi}{\partial_{z^2}^2 f} (-\partial_{zx}^2 f + f \partial_{zy}^2 f)$

On montre que ces γ du feuilletage de S par \mathfrak{D} sont des *géodésiques* (“ab-normales”¹⁵) pour toute métrique sous-riemannienne sur \mathfrak{D} . Elles sont également *rigides* au sens où si δ est une autre courbe intégrale de \mathfrak{D} de mêmes extrémités que γ et assez proche de γ , alors δ n’est qu’une reparamétrisation de γ et donc est géométriquement la même courbe. Cela se voit immédiatement sur la forme normale. S est le plan $p = 0$ et, comme \mathfrak{D} est le noyau de $\omega = dy - p^2 dx$, \mathfrak{D} induit sur S le feuilletage par les droites $p = 0, y = \text{cste}$. Prenons par exemple γ correspondant à $y = 0$, d’équation $\gamma(t) = (t, 0, 0)$ entre t_1 et t_2 . Soit $\delta(t) = (x(t), y(t), z(t))$ une courbe voisine de $\gamma(t)$ satisfaisant

$$y'(t) = z^2(t) x'(t)$$

(et donc $\omega(\gamma(t)) = 0$ puisque z est renommée p) et

$$y(t_1) = z(t_1) = y(t_2) = z(t_2) = 0$$

(i.e. $\delta(t_1)$ et $\delta(t_2)$ sont, comme $\gamma(t_1)$ et $\gamma(t_2)$, sur l’axe des x). Comme $\delta(t)$ est assez voisine de $\gamma(t)$ et que pour $\gamma(t)$ on a $x'(t) = 1$, on peut supposer $x'(t) > 0$ partout pour $\delta(t)$ ce qui implique $y'(t) \geq 0$ et donc $y(t)$ croissante. Mais comme $y(t_1) = y(t_2) = 0$, la seule possibilité est $y(t) \equiv 0$ et donc $z^2(t) \equiv 0$ et $z(t) \equiv 0$. Bref, $\delta(t) = (x(t), 0, 0)$ est bien une reparamétrisation de $\gamma(t) = (t, 0, 0)$.

1.6.2. *Les cas plus compliqués (Zhitomirskii).*

Mais on peut pousser plus loin la dégénérescence en supposant que $\partial_{z^2}^2 f(0)$ et $\partial_{zx}^2 f(0)$ sont toutes deux nulles ($(\omega \wedge dF)_0 = 0$) même si S est régulière en 0 (i.e. $dF_0 \neq 0$ et donc $\partial_{zy}^2 f(0) \neq 0$). S et \mathfrak{D}_0 sont alors *tangentes* en 0 et le feuilletage de S par \mathfrak{D} devient *singulier* en 0 à cause de ce défaut de transversalité. Les conditions

- (a) $\partial_z f = F = 0$ ($m \in S$),
- (b) $\eta = -f\xi$ ($X \in \mathfrak{D}_m$),
- (c) $\xi \partial_{zx}^2 f + \eta \partial_{zy}^2 f + \pi \partial_{z^2}^2 f = 0$ ($X \in T_m S$),

donnent en 0 (c) $\eta \partial_{zy}^2 f(0) = 0$ et, si $\partial_{zy}^2 f(0) \neq 0$, $\eta = 0$ qui est la condition (b). Les conditions de tangence à S et \mathfrak{D} deviennent identiques en 0. Michael Zhitomirskii a étudié ce cas nettement plus compliqué (cf. par exemple [573]). La dégénérescence $\partial_{z^2}^2 f(0) = \partial_{zx}^2 f(0) = 0$ signifie que le champ caractéristique

$$Z = (\xi_Z = \partial_z F, \eta_Z = -f(\partial_z F), \pi_Z = f(\partial_y F) - \partial_x F)$$

s’annule en 0 puisque $f, \partial_x F$ et $\partial_z F$ s’annulent. La distribution \mathfrak{D} induit donc sur S un champ Z qui s’annule en 0. Pour analyser un tel champ au voisinage de 0 on doit d’abord le *linéariser* et regarder les valeurs propres λ_1, λ_2 de son linéarisé L_0 .

et ces X sont donc portés par la droite $\text{Span}\left\{1, -f, \frac{-\partial_{zx}^2 f + f \partial_{zy}^2 f}{\partial_{z^2}^2 f}\right\}$. En 0 cela donne, puisque $f(0) = 0, X = \left(\xi, 0, -\xi \frac{\partial_{zx}^2 f}{\partial_{z^2}^2 f}\right)$. Si $\partial_{z^2}^2 f = 0$, alors $\partial_{zx}^2 f(0) \neq 0$ par hypothèse. On change de coordonnées locales et l’on obtient un résultat analogue.

15. Cf. la section 5 du chapitre 12.

On montre que $\lambda_1 + \lambda_2 = \text{tr}(L_0) = 0$. En effet, si on considère Z comme un champ dans \mathbb{R}^3 nul à l'origine, son linéarisé L est donné comme la transformation linéaire de T_0M de matrice ¹⁶

$$\begin{aligned}
 L &= \begin{pmatrix} \partial_x \xi_Z & \partial_y \xi_Z & \partial_z \xi_Z \\ \partial_x \eta_Z & \partial_y \eta_Z & \partial_z \eta_Z \\ \partial_x \pi_Z & \partial_y \pi_Z & \partial_z \pi_Z \end{pmatrix} \\
 &= \begin{pmatrix} \partial_x \partial_z F & \partial_y \partial_z F & \partial_z \partial_z F \\ -\partial_x (f(\partial_z F)) & -\partial_y (f(\partial_z F)) & -\partial_z (f(\partial_z F)) \\ \partial_x (f(\partial_y F) - \partial_x F) & \partial_y (f(\partial_y F) - \partial_x F) & \partial_z (f(\partial_y F) - \partial_x F) \end{pmatrix} \\
 &= \begin{pmatrix} \frac{\partial_{xz}^2 F}{\partial_{yz}^2 F} & \frac{-\partial_x (f(\partial_z F)) - f \partial_{xz}^2 F}{\partial - (\partial_y f)(\partial_z F) - f \partial_{yz}^2 F} & \frac{(\partial_x f)(\partial_y F) + f \partial_{xy}^2 F - \partial_{x^2}^2 F}{\partial (\partial_y f)(\partial_y F) + f \partial_{y^2}^2 F - \partial_{yx}^2 F} \\ \frac{\partial_{yz}^2 F}{\partial_{z^2}^2 F} & \frac{\partial - (\partial_y f)(\partial_z F) - f \partial_{yz}^2 F}{-\partial_z (f(\partial_z F)) - f \partial_{z^2}^2 F} & \frac{(\partial_x f)(\partial_y F) + f \partial_{xy}^2 F - \partial_{x^2}^2 F}{(\partial_z f)(\partial_y F) + f \partial_{zy}^2 F - \partial_{zx}^2 F} \end{pmatrix}
 \end{aligned}$$

En 0, $\mathfrak{D}_0 = T_0S$ est défini par $dy = 0$ i.e. est restreint aux vecteurs tangents avec $\eta = 0$ et l'on doit donc restreindre L à la matrice 2×2 obtenue en enlevant la colonne et la ligne de la composante η . Et puisque $f = 0$ et $dF = 0$, on obtient en définitive

$$L_0 = \begin{pmatrix} \partial_{xz}^2 F(0) & \partial_{z^2}^2 F(0) \\ -\partial_{x^2}^2 F(0) & -\partial_{zx}^2 F(0) \end{pmatrix}$$

qui est bien de trace nulle. Le polynôme caractéristique de L_0 est un polynôme du second degré à coefficients réels. Il a donc soit deux valeurs propres réelles opposées, soit deux valeurs propres complexes conjuguées qui, devant être opposées, sont nécessairement imaginaires purs. Comme ω n'est définie qu'à un facteur multiplicatif près, on peut normaliser λ_1, λ_2 . On se ramène ainsi à 3 cas :

1. $\lambda_1 = 1, \lambda_2 = -1$. Le champ Z est tangent à un foyer. On dit que le point singulier est "hyperbolique". Dans un repère (u, v) où L_0 devient diagonale, si l'on regarde les trajectoires $(u(t), v(t))$ les points (u_0, v_0) suivent une loi $(u(t) = u_0 e^t, v(t) = v_0 e^{-t})$. Les points sur l'axe des u divergent exponentiellement vers l'infini sur cet axe et ceux sur l'axe des v convergent exponentiellement vers 0 sur cet axe. Les autres trajectoires sont des hyperboles équilatères

16. La troisième forme de la matrice étant trop large nous avons décalé les colonnes dans la mise en page. Elles sont séparées par des lignes horizontales.

d'équation algébrique $uv = u_0v_0$. Une forme normale calculée par M. Zhitomirskii est

$$\omega = dy + (xy + x^2z + ax^3) dz$$

(a un nombre réel). Ces points sont déterminés par leurs jets d'ordre 4. Ils sont de codimension 3 et sont donc isolés dans \mathbb{R}^3 . Leur petit vecteur de croissance est $(2, 2, 2, 3)$.

2. $\lambda_1 = i$, $\lambda_2 = -i$. Le champ Z est tangent à une rotation. On dit que le point singulier est "elliptique". Dans un repère (u, v) où L_0 devient diagonale, si l'on regarde les trajectoires $(u(t), v(t))$ les points (u_0, v_0) suivent une loi $(u(t) = u_0 \cos(t), v(t) = v_0 \sin(t))$. Les points sur l'axe des u oscillent entre -1 et $+1$ sur cet axe et il en va de même de ceux sur l'axe des v . Les autres trajectoires sont des ellipses d'équation algébrique $\left(\frac{u}{u_0}\right)^2 + \left(\frac{v}{v_0}\right)^2 = 1$. Une forme normale calculée par M. Zhitomirskii est

$$\omega = dy + \left(xy + \frac{x^3}{3} + xz^2 + ax^3z^2 \right) dz .$$

Ces points sont déterminés par leurs jets d'ordre 5. Ils sont de codimension 3 et sont donc eux aussi isolés dans \mathbb{R}^3 . Leur petit vecteur de croissance est encore $(2, 2, 2, 3)$.

3. $\lambda_1 = \lambda_2 = 0$. $L_0 = 0$ et donc

$$\partial_{x^2}^2 F(0) = \partial_{xz}^2 F(0) = \partial_{z^2}^2 F(0) = 0 .$$

C'est un degré de dégénérescence supplémentaire et il faut pousser l'analyse encore plus loin.

On voit à quel point l'analyse se complique et comment interviennent des degrés de plus en plus élevés de dégénérescence (mesurés par des codimensions de strates) qui peuvent être décrits comme des conditions de non-transversalité au moyen de formes différentielles, de leurs dérivées extérieures, de produits extérieures et de produits intérieurs. C'est cette méthodologie de Cartan-Thom qui pilote la géométrisation du calcul différentiel.

Après avoir donné l'exemple le plus simple de classification de distributions quelconques afin de montrer toute la complexité interne que cache cette simplicité apparente, nous allons revenir maintenant sur plusieurs théorèmes de formes normales que nous avons déjà rencontrés pour certains types bien précis de distributions.

2. Géométrie de contact et géométrie symplectique

Encore une fois en accord avec nos choix pédagogiques d'ouverture de nos modèles vers les théories de mathématique pure qui les contextualisent, nous allons maintenant préciser certaines notions de géométrie de contact et de géométrie symplectique avant que d'en venir à la géométrie générale des espaces de jets.

Nous avons déjà développé, à propos de nos exemples concrets, de nombreuses propriétés des structures de contact et des structures symplectiques, entre autres la condition d'intégrabilité et la symplectisation (chapitre 3, sections 2.2, 2.6, 3.2), la théorie de Kirillov des orbites de la représentation co-adjointe des groupes de Lie (chapitre 5, section 6.3), ainsi que la notion de connexion (chapitre 6, section 2). Et dans les chapitres 12, 13, 14 nous avons montré à quel point les structures symplectiques sont omniprésentes dans les structures hamiltoniennes du contrôle optimal. Dans cette section, nous voudrions brièvement esquisser quelques éléments supplémentaires unifiant leur théorie générale.

2.1. Les variétés de contact et leur symplectisation

Commençons par la notion de variété de contact. Nous suivons le classique traité [20] de Vladimir Arnold, l'une des "bibles" de notre génération. De façon la plus générale, une variété de contact est une variété différentiable M munie d'un champ différentiable d'hyperplans tangents \mathcal{K} complètement non intégrable¹⁷, c'est-à-dire n'admettant aucune hypersurface intégrale. La distribution \mathcal{K} est localement définie comme noyau d'une 1-forme ω définie, nous l'avons vu, à un facteur constant $\lambda \in \mathbb{R}^*$ près¹⁸ en chaque point et donc à la multiplication près par une fonction f partout $\neq 0$. Comme $d(f\omega) = f d\omega + df \wedge \omega$ (formule de Leibniz), on voit que sur \mathcal{K} où $\omega = 0$, $d(f\omega) = f d\omega$ et donc $d\omega \upharpoonright_{\mathcal{K}}$ est également définie à la multiplication près par une fonction partout $\neq 0$. \mathcal{K} serait intégrable si $d\omega \upharpoonright_{\mathcal{K}} = 0$ ce qui est équivalent, nous l'avons vu plusieurs fois, à la condition de Frobenius $\omega \wedge d\omega = 0$. Et à l'autre extrême, \mathcal{K} est de contact si le rang de $d\omega_x$ est partout égal à $n - 1$ (condition de non dégénérescence, $d\omega_x \upharpoonright_{K_x}$ a un noyau trivial) : pour tout $X_x \in K_x$, $X_x \neq 0$, il existe $Y_x \in K_x$ tel que $d\omega_x(X_x, Y_x) \neq 0$. Comme

$$d\omega(X, Y) = -\omega([X, Y])$$

sur \mathcal{K} ¹⁹, on voit que la non-dégénérescence signifie bien la complète non-intégrabilité. Cette condition implique que M soit de dimension impaire, car sur une variété de dimension paire $2m$ les hyperplans tangents sont de dimension impaire $2m - 1$ et sur un espace vectoriel de dimension impaire $2m - 1$ toute forme bilinéaire

17. Nous renotons \mathcal{K} la distribution plutôt que \mathfrak{D} afin de garder nos notations initiales pour les distributions de contact.

18. $\lambda \in \mathbb{R}^*$, le groupe multiplicatif des réels $\neq 0$.

19. Cf. la section précédente (1).

antisymétrique comme $d\omega_x$ est dégénérée et admet un vecteur $X_x \in K_x$, $X_x \neq 0$, tel que $d\omega_x(X_x, Y_x) = 0$ pour *tout* $Y_x \in K_x$.

De même que les fibrés cotangents constituent la source principale des variétés symplectiques, la généralisation de la construction de la variété de contact \mathbb{V}_J des éléments de contact (a, p) du plan \mathbb{R}^2 (qui nous a servi de point de départ dans le chapitre 3) constitue la principale source des variétés de contact. Le formalisme est évidemment plus compliqué mais la construction est essentiellement la même.

Si R est une variété différentiable de dimension r , les éléments de contact sont les couples (a, H_a) d'un point a de R et d'un hyperplan tangent H_a de $T_a R$.²⁰ Ils forment une variété M de dimension impaire $2r - 1 = n$ (r pour a et $r - 1$ pour H_a) qui est le *projectivisé* du fibré cotangent T^*R .²¹ C'est un fibré sur R de fibre M_a isomorphe à l'espace projectif \mathbb{P}^{r-1} . En effet, un H_a est le noyau d'un rayon de 1-formes $\lambda H_a^* \neq 0$ en a .²²

Soit M la variété des éléments de contact de R . On peut directement généraliser la construction de la section 2.2 du chapitre 3 et de sa figure 2. Soient $a \in R$ et H_a un hyperplan de $T_a R$, définissant un point $x = (a, H_a)$ de M . On veut définir un hyperplan K_x de $T_x M$. M étant de dimension impaire $n = 2r - 1$, TM est de dimension $2n = 4r - 2$ et $T_x M$ de dimension $n = 2r - 1$. K_x doit donc être de dimension paire $n - 1 = 2r - 2 = 2(r - 1)$. Pour définir K_x on considère la projection canonique $\pi : M \rightarrow R$, $x = (a, H_a) \simeq (a, \lambda H_a^*) \mapsto a$ et son prolongement π_* aux fibrés tangents, et, si $X_x = (\alpha, \eta^*) \in T_x M$ (où $\alpha \in T_a R$ et $\eta^* \in T_{H_a} \mathbb{P}(T_a^* R)$), on pose que $X_x \in K_x$ si et seulement si $\pi_*(X_x) \in H_{\pi(x)}$. Autrement dit, on impose que lors du déplacement infinitésimal de (a, H_a) défini par X_x le déplacement infinitésimal de a soit dans H_a , aucune contrainte n'étant imposée par ailleurs à la variation infinitésimale de H_a dans l'espace des hyperplans. Donc X_x peut varier le long de $r - 1$ degrés de liberté et comme l'espace des H_a est de dimension $r - 1$, on obtient bien $\dim(K_x) = 2(r - 1)$.

Pour montrer que (M, \mathcal{K}) ainsi définie est bien une variété de contact, autrement dit que le champ d'hyperplans tangents \mathcal{K} est non dégénéré il faut préciser la situation en termes de formes différentielles. Pour cela on "symplectise" (M, \mathcal{K}) en généralisant la construction exposée à la section 3.2 du chapitre 3 à propos de \mathbb{V}_J comme espace des 1-jets. On considère les 1-formes *de contact* $\omega_x \in T_x^* M$ définies (à un facteur $\lambda \in \mathbb{R}^*$ près) comme celles dont les noyaux sont les K_x . Leur ensemble est un sous-ensemble du fibré cotangent $T^* M$ et s'appelle la "symplectisée" de M .

20. Dans le modèle \mathbb{V}_J , R est le plan rétinien, a un point de R , p_a une orientation en a et H_a la droite du plan tangent $T_a R$ d'orientation p_a .

21. Par rapport au chapitre 3 où une variété de contact était notée V en référence à l'aire $V1$ et son modèle \mathbb{V}_J , nous changeons un peu les notations car nous notons M une variété de contact (ou une variété symplectique) générale.

22. La 1-forme $H_a^* \neq 0 \in T_a^* R$ mais, n'étant définie qu'à un facteur constant $\lambda \in \mathbb{R}^*$ près, n'intervient que par la droite λH_a^* qui la supporte, d'où le passage au projectivisé.

C'est le sous-fibré $\pi : \Sigma^*M \rightarrow M$ de $\pi : T^*M \rightarrow M$ ayant comme fibre au-dessus de x les ω_x de contact. Comme les fibres sont de dimension 1, Σ^*M est de dimension paire $n + 1 = 2r$. La condition de non-dégénérescence disant que (M, \mathcal{K}) est une variété de contact exprime alors le fait que Σ^*M est une variété *symplectique*.²³

Plus précisément, il existe une 1-forme *canonique* α sur la symplectisée Σ^*M . Soit $x \in M$ et une forme de contact $\omega_x \in \Sigma_x^*M$ au-dessus de x . Un vecteur tangent τ en ω_x à Σ^*M est le composé d'un vecteur tangent "horizontal" $t_x \in T_xM$ tangent à M en x et d'un vecteur tangent "vertical" $\xi_{\omega_x} \in T_{\omega_x}(\Sigma_x^*M)$ tangent à la fibre Σ_x^*M en ω_x . On définit alors $\alpha(\tau)$ comme la valeur que prend ω_x sur la projection $\pi_*(\tau) = t_x$ et on a donc, comme sur la figure 2 du chapitre 3, section 2.2,

$$\alpha(\tau) = \omega_x(\pi_*(\tau)) = \omega_x(t_x) .$$

On voit immédiatement qu'on aura $\alpha(\tau) = 0$ si et seulement si $\pi_*(\tau) = t_x \in K_x$, ce qui signifie que le noyau de α en ω_x est l'hyperplan "vertical" produit de la fibre Σ_x^*M et de l'hyperplan K_x .

Le résultat clé est que la 1-forme α (qui, étant définie sur un espace de dimension paire, ne peut pas être de contact) possède une dérivée extérieure $d\alpha$ qui est une forme *symplectique* (cf. Arnold [20]). Comme $d\alpha$ est automatiquement fermée puisqu'elle est exacte ($d(d\alpha) = d^2\alpha = 0$), il faut montrer qu'elle est *régulière* (non dégénérée), autrement dit que pour tout $\tau = (t_x; \xi_{\omega_x})$, $\tau \neq 0$, de $T_{\omega_x}(\Sigma^*M)$ ($t_x \in T_xM$ "horizontal" et $\xi_{\omega_x} \in T_{\omega_x}(\Sigma_x^*M)$ "vertical"), il existe $\rho = (r_x; \zeta_{\omega_x})$ de $T_{\omega_x}(\Sigma^*M)$ tel que $d\alpha(\tau, \rho)$ soit $\neq 0$. En effet, \mathcal{K} peut être définie localement par une 1-forme ϖ et les ω_x peuvent être définis comme les $(\lambda\varpi)_x$, où $\lambda \in \mathbb{R}^*$ est la coordonnée "verticale" sur les fibres, et donc $\alpha = \lambda\pi^*(\varpi)$ où $\pi^*(\varpi)$ est le pull-back sur Σ^*M de ϖ (qui est définie sur M) au moyen de la projection $\pi : \Sigma^*M \rightarrow M$. On dérive de cette expression de α l'expression de $d\alpha$,

$$d\alpha = d\lambda \wedge \pi^*(\varpi) + \lambda\pi^*(d\varpi) ,$$

ce qui permet de montrer que $d\alpha$ est régulière et donc symplectique. On utilise le fait que

$$\begin{aligned} d\alpha(\tau, \rho) &= d\lambda(\tau)\pi^*(\varpi)(\rho) - d\lambda(\rho)\pi^*(\varpi)(\tau) + \lambda\pi^*(d\varpi)(\tau, \rho) \\ &= d\lambda(\tau)\varpi(\pi_*(\rho)) - d\lambda(\rho)\varpi(\pi_*(\tau)) + \lambda(d\varpi(\pi_*(\tau), \pi_*(\rho))) \\ &= d\lambda(\tau)\varpi(r) - d\lambda(\rho)\varpi(t) + \lambda(d\varpi(t, r)) . \end{aligned}$$

On distingue trois cas.

23. On peut noter Σ^*M simplement Σ lorsqu'on la traite comme variété symplectique générale en oubliant qu'elle est une sous-variété du fibré cotangent T^*M .

1. Si $\varpi(\pi_*(\tau)) = \varpi(t) \neq 0$, on peut prendre un $\rho \neq 0$ “vertical” (i.e. tel que $\pi_*(\rho) = r = 0$) et alors

$$\begin{aligned} d\alpha(\tau, \rho) &= -d\lambda(\rho) \varpi(\pi_*(\tau)) = -d\lambda(\rho) \varpi(t) \\ &\neq 0 \end{aligned}$$

puisque $\varpi(\pi_*(\tau)) = \varpi(t) \neq 0$ et $d\lambda(\rho) \neq 0$ puisque λ est la coordonnée “verticale” et ρ est “vertical”.

2. Si $\varpi(\pi_*(\tau)) = \varpi(t) = 0$ mais τ n’est pas “vertical”, on peut prendre un ρ satisfaisant également $\varpi(\pi_*(\rho)) = \varpi(r) = 0$. Alors

$$d\alpha(\tau, \rho) = \lambda d\varpi(\pi_*(\tau), \pi_*(\rho)) = \lambda d\varpi(t, r)$$

Comme τ n’est pas “vertical” $\pi_*(\tau) = t \neq 0$ alors que $\varpi(\pi_*(\tau)) = \varpi(t) = 0$. Mais comme (M, \mathcal{K}) est une variété *de contact*, $d\varpi$ est non dégénérée sur \mathcal{K} et il existe donc un ρ tel que $d\varpi(\pi_*(\tau), \pi_*(\rho)) = d\varpi(t, r) \neq 0$.

3. Enfin, si $\tau \neq 0$ est “vertical”, i.e. si $t = \pi_*(\tau) = 0$ (et donc $\varpi(\pi_*(\tau)) = \varpi(t) = 0$) mais $\xi_{\omega_x} \neq 0$, i.e. $d\lambda(\tau) \neq 0$, il suffit de prendre un ρ tel que $\varpi(\pi_*(\rho)) = \varpi(r) \neq 0$.

2.2. Formes normales de Darboux

2.2.1. Les formules.

Localement, les 1-formes de contact et les 2-formes symplectiques (qui sont étroitement intriquées comme nous venons de le voir) possèdent des formes normales universelles d’après le célèbre *théorème de Darboux*. Nous l’avons évoqué à la section 4 du chapitre 3. Soit (M, \mathcal{K}) une variété de contact de dimension $2r - 1$ avec $\mathcal{K} = \ker(\omega)$ pour ω une 1-forme de contact. Le théorème de Darboux dit qu’il existe alors en tout point des coordonnées locales (q_j, p_j, z) , $j = 1, \dots, r - 1$, telles que, dans ces coordonnées, ω prenne la forme normale

$$\omega = dw + \sum_{j=1}^{j=r-1} p_j dq_j .$$

Dans une analogie mécanique hamiltonienne les coordonnées (q_j, p_j) sont des variables conjuguées de position et de moment et la variable w est une “énergie”. Notre modèle \mathbb{V}_J avec $\omega_J = dy - p dx$ est une forme normale pour les coordonnées $\{q_1 = x, p_1 = -p, w = y\}$.

Ce théorème se démontre en prenant la symplectisée $\Sigma = \Sigma^*M$ de M (de dimension $2r$) avec sa forme symplectique $\Omega = d\alpha$ et en appliquant le théorème de Darboux symplectique disant qu’il existe alors en tout point s de Σ des coordonnées locales (p_j, q_j) , $j = 0, \dots, r - 1$, $s = 0$, telles que, dans ces coordonnées, Ω soit sous

la forme normale

$$\Omega = \sum_{j=0}^{j=r-1} dp_j \wedge dq_j .$$

Autrement dit, toute variété symplectique est localement un espace des phases classique \mathbb{R}^{2r} avec des coordonnées conjuguées (p_j, q_j) . Pour construire ces coordonnées, on utilise la géométrie très spéciale et remarquable des structures symplectiques.

2.2.2. Structures symplectiques, riemanniennes et complexes.

Revenons sur la trinité géométrique que nous avons évoquée dans la section 3.4 du chapitre 3. Les propriétés des structures symplectiques ne sont pas intuitives mais le deviennent un peu plus si on les immerge dans des structures complexes où elles interfèrent avec des structures riemanniennes. Pour les illustrer, nous ferons appel à la structure symplectique standard $\Omega = dp \wedge dq$ de \mathbb{R}^2 ou $\Omega = \sum_{j=0}^{j=r-1} dp_j \wedge dq_j$ de \mathbb{R}^{2r} . Dans le cas \mathbb{R}^2 , $\Omega(\tau, \rho)$ est l'aire orientée du parallélogramme $P(\tau, \rho)$ défini par les vecteurs tangents τ et ρ . Dans le cas \mathbb{R}^{2r} , $\Omega(\tau, \rho)$ est la somme aires orientées des projections de $P(\tau, \rho)$ sur les plans (p_j, q_j) .

D'abord, Ω étant une forme bilinéaire antisymétrique en tout point, on a

$$\Omega(\tau, \tau) = 0$$

pour tout vecteur tangent. On a une relation de Ω -orthogonalité naturelle $\tau \perp_{\Omega} \rho$ si $\Omega(\tau, \rho) = 0$. Elle est très différente de la relation d'orthogonalité euclidienne, que nous appellerons E -orthogonalité, puisque tout vecteur tangent τ est orthogonal à lui-même, on dit "isotrope". C'est évident puisque $P(\tau, \tau)$ est un parallélogramme dégénéré d'aire 0. Dans le cas \mathbb{R}^2 , on a $\Omega(\tau, \rho) \neq 0$ si τ et ρ ne sont pas colinéaires, ce qui est caractéristique du fait que p et q sont des coordonnées conjuguées. Si l'on note $H^{\perp_{\Omega}}$ l'orthogonal d'un sous-espace, on a $\mathbb{R}\tau^{\perp_{\Omega}} = \mathbb{R}\tau$ pour tout $\tau \neq 0$.

Notons une complémentarité remarquable dans \mathbb{R}^2 entre la forme symplectique $\Omega(\tau, \rho) = (dp \wedge dq)(\tau, \rho)$ et le produit scalaire euclidien $\langle \tau, \rho \rangle$. En effet, si l'on note τ_p et τ_q les composantes des vecteurs tangents, on a

$$\begin{aligned} \Omega(\tau, \rho) &= (dp \wedge dq)(\tau, \rho) = \tau_p \rho_q - \tau_q \rho_p \\ &= \langle (-\tau_q, \tau_p), (\rho_p, \rho_q) \rangle = \langle J\tau, \rho \rangle \\ &= \langle (\tau_p, \tau_q), (\rho_q, -\rho_p) \rangle = -\langle \tau, J\rho \rangle \end{aligned}$$

avec $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ la matrice de la rotation de $\frac{\pi}{2}$ dans le plan. Autrement dit, l'aire orientée de $P(\tau, \rho)$ est égale à la valeur de la projection orthogonale sur ρ de τ tourné de $\frac{\pi}{2}$. On peut alors identifier \mathbb{R}^2 au plan complexe \mathbb{C} de coordonnée $z = p + iq$. L'opérateur J , qui satisfait $J^2 = -1$, est dans ce cas tout simplement

la multiplication par i . Le produit hermitien dans \mathbb{C} est

$$\begin{aligned}\bar{\tau}\rho &= (\tau_p - i\tau_q)(\rho_p + i\rho_q) = \tau_p\rho_p + \tau_q\rho_q + i(-\tau_q\rho_p + \tau_p\rho_q) \\ &= \langle \tau, \rho \rangle + i\Omega(\tau, \rho) .\end{aligned}$$

Dans le cas de \mathbb{R}^{2r} , $r > 1$, les choses sont un peu plus compliquées. En effet si \mathbf{p}_j et \mathbf{q}_j sont les vecteurs unitaires sur les axes de coordonnées, on a $\mathbf{p}_j \perp_{\Omega} \mathbf{p}_j$ et $\mathbf{q}_j \perp_{\Omega} \mathbf{q}_j$ (isotropie) ainsi que $\Omega(\mathbf{p}_j, \mathbf{q}_j) = 1 \neq 0$ comme dans le cas $r = 1$, mais on a aussi $\mathbf{p}_j \perp_{\Omega} \mathbf{p}_k$, $\mathbf{q}_j \perp_{\Omega} \mathbf{q}_k$ et $\mathbf{p}_j \perp_{\Omega} \mathbf{q}_k$ si $j \neq k$. En effet, si $j \neq k$, \mathbf{p}_j et \mathbf{q}_j sont E -orthogonaux à \mathbf{p}_k et \mathbf{q}_k et toutes les projections du P défini par $(\mathbf{p}_j, \mathbf{p}_k)$, $(\mathbf{q}_j, \mathbf{q}_k)$ ou $(\mathbf{p}_j, \mathbf{q}_k)$ sur les plans $(\mathbf{p}_j, \mathbf{q}_j)$ sont nulles. Donc tout vecteur de la base symplectique $\{\mathbf{p}_j, \mathbf{q}_j\}_j$ est Ω -orthogonal à tous les autres sauf un, qui est son conjugué. On peut dire que \mathbb{R}^{2r} est décomposé en produit de r plans symplectiques à la fois Ω -orthogonaux et E -orthogonaux entre eux. Pour calculer $\Omega(\tau, \rho)$ on regarde ce qui se passe dans les plans (p_j, q_j) , on applique les multiplications par J_j (rotation de $\frac{\pi}{2}$ dans le plan (p_j, q_j)), on prend les produits scalaires et on les somme. La structure euclidienne est celle faisant de la base symplectique une base orthonormée, la structure complexe est celle de \mathbb{C}^r avec les variables complexes $z_j = p_j + iq_j$ et la forme hermitienne est encore $\langle \tau, \rho \rangle + i\Omega(\tau, \rho)$. C'est l'intrication entre structure symplectique, structure euclidienne et structure complexe qui est la clé de la compréhension de la géométrie symplectique.

Dans \mathbb{R}^{2r} il y a des k -plans H isotropes, i.e. tels que $\Omega|_H \equiv 0$ qui sont de dimension > 1 . Il suffit que H et JH (où J consiste à appliquer tous les J_j) soient E -orthogonaux. Comme H et JH sont de dimension k , cela implique $k \leq r$. D'après les règles de Ω -orthogonalité que nous venons de voir, les k -plans de coordonnées (u_m) , $m = 1, \dots, k$, sont isotropes si et seulement si les u_m ne comprennent pas une paire de coordonnées conjuguées (p_j, q_j) . On obtient les r -plans de coordonnées isotropes de dimension maximale r en partitionnant $\{0, \dots, r-1\}$ en deux sous-ensembles A et B et en prenant les coordonnées p_j pour $j \in A$ et les coordonnées q_j pour $j \in B$.

Si Σ est une variété symplectique, on peut toujours construire en tout point s une base symplectique de $T_s\Sigma$ comme pour \mathbb{R}^{2r} . On démarre avec un $\mathbf{p}_0 \neq 0 \in T_s\Sigma$ quelconque. Comme Ω est régulière, il existe un $\mathbf{q}_0 \neq 0 \in T_s\Sigma$ non Ω -orthogonal à \mathbf{p}_0 . On le normalise et l'on obtient la structure symplectique standard sur le plan $(\mathbf{p}_0, \mathbf{q}_0)$. On considère alors le Ω -orthogonal $(\mathbf{p}_0, \mathbf{q}_0)^{\perp_{\Omega}}$ qui est $\mathbf{p}_0^{\perp_{\Omega}} \cap \mathbf{q}_0^{\perp_{\Omega}}$ de dimension $2r-2$. Comme Ω est régulière, Ω restreinte à $(\mathbf{p}_0, \mathbf{q}_0)^{\perp_{\Omega}}$ est symplectique et l'on peut itérer la construction.

Une propriété remarquable de toute forme symplectique Ω est d'établir un isomorphisme entre $T_s\Sigma$ et $T_s^*\Sigma$. Celui-ci associe à tout vecteur tangent $\tau \in T_s\Sigma$ la 1-forme $\tau^* \in T_s^*\Sigma$ définie par $\tau^*(\rho) = \Omega(\rho, \tau)$. C'est bien un isomorphisme parce que Ω est non dégénérée. Si f est une fonction sur Σ et df sa différentielle, le champ

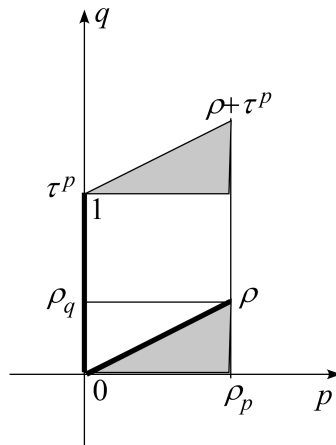


FIGURE 1. Si $\rho = (\rho_p, \rho_q)$ et $\tau^p = (0, 1)$, l'aire du parallélogramme $P(\tau^p, \rho) = 0 \rightarrow \rho \rightarrow \rho + \tau^p \rightarrow \tau^p \rightarrow 0$

est égale à ρ_p . On le montre en déplaçant le triangle grisé du haut sur le triangle grisé du bas pour obtenir un carré de côtés ρ_p et 1 et donc d'aire ρ_p .

de vecteurs tangents associé τ^j est défini par $df(\rho) = \Omega(\rho, \tau^j)$. Dans le cas standard de \mathbb{R}^2 avec $\Omega = dp \wedge dq$, le τ^p associé à dp doit satisfaire $dp(\rho) = \Omega(\rho, \tau^p)$, autrement dit, si l'on continue à noter ρ_p et ρ_q les composantes des vecteurs tangents, $\rho_p = \rho_p \tau_q^p - \rho_q \tau_p^p$ pour tout ρ et donc il faut $\tau_q^p = 1$ et $\tau_p^p = 0$. On note que τ^p est E -orthogonal à l'axe des p et l'on vérifie immédiatement sur la figure 1 que l'aire du parallélogramme $P(\tau^p, \rho)$ est bien égale à ρ_p . Dans le cas standard \mathbb{R}^{2r} le τ^{p_j} associé à p_j a toutes ses composantes nulles sauf celle correspondant à la coordonnée q_j conjuguée de p_j qui est = 1.

On travaille localement en $s = 0$. On prend d'abord une première coordonnée p_0 qui est une courbe C^∞ passant par s . Comme Σ est une variété différentiable, on peut trouver $2r - 1$ autres coordonnées transformant localement Σ en un \mathbb{R}^{2r} : tout point s assez voisin de 0 sera un couple (p_0, s') avec s' un point d'un feuilletage local de Σ en s par des hypersurfaces transverses à la coordonnée et d'équation $dp_0 = 0$. Mais en utilisant Ω , on associe à la différentielle dp_0 un champ de vecteurs tangents $X_0(s)$ tel que pour tout champ local $Y(s)$, $dp_0(Y) = \Omega(Y, X_0)$. Comme $X_0(0) \neq 0$ on peut prendre une hypersurface H de dimension $2r - 1$ transversale à X_0 en $s = 0$. Localement, tout point s d'un voisinage de $s = 0$ peut être atteint à partir d'un point h de H en suivant pendant un certain temps t_s le champ X_0 . On prend alors t_s comme coordonnée q_0 , H devenant l'hypersurface $q_0 = 0$. Par construction, la dérivée de q_0 le long de X_0 est = 1. On peut itérer cette construction car en prenant le sous-espace Σ' d'équations $p_0 = 0$ et $q_0 = 0$ on obtient une hypersurface de H de

dimension $2r - 2$ sur laquelle Ω est encore symplectique. En utilisant ces coordonnées symplectiques on ramène Ω à la forme normale de Darboux $\Omega = \sum_{j=0}^{j=r-1} dp_j \wedge dq_j$.

Cette forme normale disant que, localement, toute variété symplectique Σ de dimension $2r$ est la variété symplectique standard \mathbb{R}^{2r} implique que la géométrie symplectique-euclidienne-complexe de \mathbb{R}^{2r} est *localement universelle* pour les variétés symplectiques.

2.2.3. Sous-variétés lagrangiennes.

On appelle en toute généralité “sous-variété lagrangienne” d’une variété symplectique Σ de dimension $2r$ une sous-variété L

(i) dont tous les plans tangents $T_s L$ sont isotropes (i.e. $\Omega_s(\tau, \rho) = 0$ pour tous $\tau, \rho \in T_s L$) et

(ii) qui est de dimension maximale r .

Si l’on identifie localement Σ à un espace des phases $\pi : T^*\mathbb{R}^r \rightarrow \mathbb{R}^r$ (\mathbb{R}^r de coordonnées q) et si f est une fonction C^∞ sur \mathbb{R}^r , sa différentielle df engendre une sous-variété lagrangienne de Σ transverse aux fibres de π (i.e. $df(q)$ est une section de π). Mais il peut y avoir des sous-variétés lagrangiennes qui ne sont pas transverses aux fibres et dont la projection $\pi(L)$ comporte des lieux singuliers. Cela est crucial en optique et en mécanique hamiltonienne où ces lieux singuliers fonctionnent comme des “caustiques” (cf. notre compilation [413] et sa bibliographie). Nous en avons vu un exemple typique à la section 6 du chapitre 12 avec les caustiques en optique ainsi que dans le chapitre 14 avec les géodésiques des modèles \mathbb{V}_J et \mathbb{V}_S .

2.2.4. Forme de contact de Darboux.

Revenons alors à la forme normale de Darboux des variétés de contact (M, \mathcal{K}, ω) . On considère la symplectisée $\pi : \Sigma^* M \rightarrow M$ ayant pour fibre au-dessus de $x \in M$ les $\mathbb{R}^* \omega_x$ avec ω_x la 1-forme ω en x . L’application $x \rightarrow \omega_x$ qui relève ω en α , est une section d’image \widetilde{M} de la projection canonique π , section qui est un difféomorphisme local de M au voisinage de x sur \widetilde{M} au voisinage de ω_x . Au voisinage de ω_x dans $\Sigma^* M$ on peut choisir des coordonnées symplectiques de Darboux locales (p_j, q_j) , $j = 0, \dots, r - 1$, normalisant la 2-forme symplectique $d\alpha$ et telles que \widetilde{M} soit d’équation $p_0 = 0$ puisque le choix de p_0 est arbitraire dans la construction de la forme normale. Comme $d\alpha = \sum_{j=0}^{j=r-1} dp_j \wedge dq_j$, on a

$$\alpha = \sum_{\underline{j=0}}^{j=r-1} p_j dq_j + dw$$

où w est une fonction nulle au point ω_x . Sur \widetilde{M} d'équation $p_0 = 0$, α devient la 1-forme

$$\alpha = \sum_{j=1}^{j=r-1} p_j dq_j + dw$$

et en transportant α sur ω on obtient la normalisation de Darboux de ω .

Une des conséquences remarquables de la forme normale de Darboux est que toute variété de contact M de dimension $2r - 1$ est *localement* la variété des éléments de contact d'une variété R de dimension r . Autrement dit, il y a toujours une *variété de base* sous-jacente R de dimension r à partir de laquelle M s'obtient par la construction que nous avons explicitée plus haut : éléments de contact (a, H_a) , H_a noyau d'un rayon de 1-formes $\lambda H_a^* \neq 0$ en a , $\lambda \in \mathbb{R}^*$.

On remarquera que $\alpha_1 = \sum_{j=1}^{j=r-1} p_j dq_j$ est elle-même une forme symplectique sur l'espace de dimension $2r - 2$ des (p_j, q_j) , $j = 1, \dots, r - 1$. Notons M_0 ce sous-espace d'équation $p_0 = 0$, $q_0 = 0$ de la symplectisée $\Sigma^* M$ de M . M s'appelle la "contactisée" de M_0 et l'on a donc une chaîne

$$M_0^{2r-2} \xrightarrow{\text{contactisée}} M^{2r-1} \xrightarrow{\text{symplectisée}} (\Sigma^* M)^{2r} .$$

2.2.5. Relevées legendriennes.

Les sous-variétés "legendriennes" L d'une variété de contact (M, \mathcal{K}, ω) de dimension $2r - 1$ sont, par définition, les sous-variétés qui sont des variétés intégrales de \mathcal{K} de dimension maximale $r - 1$. Cela signifie que $\omega \equiv 0$ sur L . On peut alors généraliser la construction de la relevée legendrienne du graphe d'une fonction $y = f(x)$ dans son espace de jets, construction qui a été à l'origine de tout notre développement depuis la section 2 du chapitre 3. Soit S une hypersurface régulière (de dimension $r - 1$) d'une variété R de dimension r . En chaque point $a \in S$ on peut considérer le plan tangent H_a de S . Les $x = (a, H_a)$ sont des éléments de contact, donc des éléments de M , et ils forment une sous-variété L_S de dimension $r - 1$ de M . Mais lorsque a varie infinitésimalement sur S , son vecteur tangent est dans H_a puisque $H_a = T_a S$. Or nous avons vu plus haut que le fait que le déplacement infinitésimal de a soit dans H_a caractérise l'hyperplan de contact K_x de M en x . L_S est donc une sous-variété intégrale de dimension maximale de \mathcal{K} , c'est-à-dire une sous-variété legendrienne. On l'appelle la "relevée legendrienne" de S .

2.2.6. Stabilité structurelle des structures de contact.

L'existence de formes normales locales pour les formes de contact est un résultat de *stabilité structurelle* pour ces distributions qui sont de type $(2(r - 1), 2r - 1)$ où la stabilité signifie que, localement, toute distribution \mathcal{K}' assez voisine de \mathcal{K} est

difféomorphe à \mathcal{K} par un difféomorphisme local préservant la structure. Une condition dimensionnelle nécessaire pour la stabilité a été démontrée par Jakubczyk-Brzytycki en 1984 et Vershik-Gershkovich en 1988. Elle affirme que si une distribution de type (k, n) est structurellement stable, alors nécessairement $k(n - k) \leq n$ (cf. Montgomery [375]). Cela est dû intuitivement au fait suivant. La dimension de la grassmannienne $Gr(k, n)$ des k -plans dans \mathbb{R}^n est $k(n - k)$ et les distributions de type (k, n) sont localement des applications C^∞ de \mathbb{R}^n dans $Gr(k, n)$. Elles dépendent donc de $k(n - k)$ fonctions C^∞ dont on peut regarder les jets successifs G^N . Par ailleurs les difféomorphismes locaux dépendent de n fonctions dont on peut aussi regarder les jets successifs D^N . La stabilité d'une structure (M, \mathcal{K}) signifie que l'action des difféomorphismes sur (M, \mathcal{K}) engendre *toutes* les (M', \mathcal{K}') assez voisines de (M, \mathcal{K}) pour la topologie naturelle (la topologie de Whitney) sur l'espace des distributions de type (k, n) . Mais si $k(n - k) - n > 0$, quand on augmente l'ordre N des jets, la dimension de G^N tend vers l'infini comme $\frac{1}{N!} (k(n - k) - n) N^n$ ce qui est beaucoup plus rapide que la dimension de D^N et il est donc impossible qu'il n'y ait qu'une seule orbite. D'où la condition dimensionnelle $k(n - k) \leq n$.²⁴ Elle autorise les champs de droites $(1, n)$ car $1 \cdot (n - 1) = n - 1 \leq n$ (mais ces champs ne peuvent évidemment pas être complètement non holonomes), les structures de type $(n - 1, n)$ car $(n - 1) \cdot 1 = n - 1 \leq n$ (ces structures sont les structures de contact si $n = 2r - 1$ est impair, en particulier $(2, 3)$ donne Heisenberg), $(2, 4)$ car $2(n - 2) = 2n - 4 \leq n$ impose $n \leq 4$ (c'est la structure d'Engel), mais interdit tous les autres types (c'est facile à vérifier).

2.3. Champs de contact

2.3.1. Champs de contact sur \mathbb{V}_J .

Nous venons de voir que les structures de contact et les structures symplectiques gouvernent des géométries essentiellement de même nature en dimensions respectivement impaires et paires. Nous avons vu par ailleurs longuement dans les chapitres 12, 13 et 14 comment la structure symplectique des espaces de phases permettait d'associer à un hamiltonien un champ de vecteurs hamiltonien défini par les équations de Hamilton et laissant invariante la forme symplectique. Il est donc naturel de dire un mot sur la notion analogue, celle dite de *champ de contact*, en géométrie de contact. Nous nous restreindrons à notre structure de contact de base définie sur $\mathbb{V} = J^1(\mathbb{R}, \mathbb{R})$ (aussi noté $\mathbb{V}_J = J^1(\mathbb{R}^2)$) par $\omega = dy - p dx$.²⁵ Soit X un champ de

24. Ce genre de calcul dimensionnel est standard dans les preuves de stabilité structurelle d'applications différentiables entre variétés. Le lecteur intéressé par leur technique en trouvera de nombreux exemples dans notre compilation [412] et sa bibliographie.

25. Comme $d\omega = dx \wedge dp = d\omega_0$, où $\omega_0 = -p dx$ est la forme symplectique standard sur l'espace de phases $\mathbb{R}_{(x,p)}^2$, on relève la mécanique hamiltonienne de \mathbb{R}^2 .

vecteurs (C^∞) sur \mathbb{V} . C'est l'opérateur de dérivation $X = \sum_{i=1}^n X_i \frac{\partial}{\partial x_i}$ de composantes X_i relativement aux dérivations particulières $\frac{\partial}{\partial x_i}$ associées aux coordonnées (locales) x_i (cf. la section 6.3 du chapitre 2). Il agit non seulement sur les fonctions sur M mais aussi sur les k -formes différentielles ϖ comme dérivée de Lie $\mathcal{L}_X \varpi$ (cf. la section 7.2.12 du même chapitre). Nous avons vu la "formule d'homotopie" de Cartan qui permet de définir $\mathcal{L}_X \varpi$ comme une somme de deux termes :

$$\mathcal{L}_X \varpi = i_X d\varpi + d(i_X \varpi) .$$

Rappelons que dans le cas de \mathbb{V}_J , si $X = \xi \partial_x + \eta \partial_y + \pi \partial_p$, on a

$$\begin{aligned} i_X \omega &= \omega(X) = \eta - p\xi \\ d(i_X \varpi) &= d\eta - pd\xi - \xi dp \\ i_X d\omega &= i_X(dx \wedge dp) = -\pi dx + \xi dp \end{aligned}$$

En effet,

$$(dx \wedge dp)(X, X') = \xi \pi' - \xi' \pi$$

et donc

$$i_X(dx \wedge dp)(X') = \xi dp(X') - \pi dx(X') .$$

Nous l'avons vu plusieurs fois²⁶. Par conséquent,

$$\begin{aligned} \mathcal{L}_X \omega &= -\pi dx + \xi dp + d\eta - pd\xi - \xi dp \\ &= -\pi dx + d\eta - pd\xi \\ &= -\pi dx + (\partial_x \eta) dx + (\partial_y \eta) dy + (\partial_p \eta) dp - p((\partial_x \xi) dx + (\partial_y \xi) dy + (\partial_p \xi) dp) \\ &= (-\pi + \partial_x \eta - p(\partial_x \xi)) dx + (\partial_y \eta - p(\partial_y \xi)) dy + (\partial_p \eta - p(\partial_p \xi)) dp . \end{aligned}$$

Cette expression générale de la façon dont un champ de vecteurs quelconque *déforme* la 1-forme de contact ω permet de définir les *champs de contact* : X est un champ de contact si le flot φ_t qu'il définit *conserve* la distribution de contact $\mathcal{K} = \ker(\omega)$ et comme celle-ci est définie par ω à un facteur $\neq 0$ près, cela signifie

26. À titre d'exercice scolaire, on peut s'amuser à vérifier à la main la formule générale de la section 1.2

$$d\omega(X, X') = X\omega(X') - X'\omega(X) - \omega([X, X']) .$$

On écrit

$$\begin{aligned} X\omega(X') &= X(\eta' - p\xi') = X(\eta') - X(p\xi') = X(\eta') - pX(\xi') - \pi\xi', \\ X'\omega(X) &= X'(\eta - p\xi) = X'(\eta) - X'(p\xi) = X'(\eta) - pX'(\xi) - \pi'\xi, \\ \omega([X, X']) &= (X(\eta') - X'(\eta)) - p(X(\xi') - X'(\xi)) . \end{aligned}$$

que le flot φ_t (qui part de $\varphi_0 = Id$) transforme ω en $\omega_{t,v} = F(t,v)\omega_v$ ($v \in \mathbb{V}$) et puisque l'on part de $F(0,v) \equiv 1$, on peut écrire

$$\omega_{t,v} = e^{\Lambda(t,v)}\omega_v$$

avec $\Lambda(0,v) \equiv 1$. Comme X est le générateur infinitésimal du groupe à un paramètre φ_t , on obtient

$$\mathcal{L}_X\omega = \lambda\omega$$

avec $\lambda = \left. \frac{d\Lambda(t)}{dt} \right|_{t=0}$ une certaine fonction C^∞ sur \mathbb{V} . De façon plus précise, les difféomorphismes φ_t sont solutions de l'équation différentielle

$$\frac{d}{dt}\varphi_t(v) = X(\varphi_t(v)) \in T_{\varphi_t(v)}\mathbb{V}$$

et, si φ_t^* est l'opérateur linéaire exprimant la façon dont φ_t agit sur les formes différentielles, on a

$$\left. \frac{d}{dt}(\varphi_t^*(\omega)) \right|_{t=0} = \varphi_t^*(\mathcal{L}_X\omega) .$$

Par conséquent, si $\varphi_t^*(\omega) = e^{\Lambda(t)}\omega$, on a

$$\varphi_t^*(\mathcal{L}_X\omega) = \left. \frac{d\Lambda(t)}{dt} \right|_{t=0} e^{\Lambda(t)}\omega = \left. \frac{d\Lambda(t)}{dt} \right|_{t=0} \varphi_t^*(\omega)$$

et, en appliquant $(\varphi_t^*)^{-1}$ (qui existe puisque φ_t^* est un isomorphisme linéaire), $\mathcal{L}_X\omega = \lambda\omega$ avec $\lambda = \left. \frac{d\Lambda(t)}{dt} \right|_{t=0}$.

D'où les conditions

$$\begin{aligned} \partial_y\eta - p(\partial_y\xi) &= \lambda \\ -\pi + \partial_x\eta - p(\partial_x\xi) &= -\lambda p \\ \partial_p\eta - p(\partial_p\xi) &= 0 . \end{aligned}$$

Sur le plan de base $p = 0$, elles se réduisent à

$$\partial_p\eta = 0, \partial_y\eta = \lambda, \partial_x\eta = \pi ,$$

conditions qui ne concernent que les dérivées partielles de la composante η . Si $p, \lambda \neq 0$, l'élimination de λ donne

$$\begin{aligned} \partial_p\eta &= p(\partial_p\xi) \\ \pi &= \partial_x\eta - p(\partial_x\xi) + p(\partial_y\eta) - p^2(\partial_y\xi) . \end{aligned}$$

La première condition est automatiquement satisfaite si ξ et η ne dépendent pas de p , ce qui signifie que la projection de X sur le plan de base $\mathbb{R}_{(x,y)}^2$ reste constante lorsqu'on se déplace le long d'une fibre de la projection $\mathbb{V} \rightarrow \mathbb{R}_{(x,y)}^2$ et que sur cette fibre (où x et y restent fixes) π varie comme un polynôme du second degré en p . Nous reviendrons plus bas sur ces champs de contact particuliers.

On peut retrouver ce résultat à la main, ce qui est un bon exercice scolaire bien que non rigoureux. Soit $v = (x, y, p) \in \mathbb{V}$, $X = \xi \partial_x + \eta \partial_y + \pi \partial_p \in T_v \mathbb{V}$. Soit $\varphi(v) = v + \delta v = v + \varepsilon X$ un petit déplacement de v dans la direction X . L'application linéaire tangente $T_v \varphi$ de φ en v est de matrice

$$T_v \varphi = \begin{pmatrix} 1 + \varepsilon \partial_x \xi & \varepsilon \partial_y \xi & \varepsilon \partial_p \xi \\ \varepsilon \partial_x \eta & 1 + \varepsilon \partial_y \eta & \varepsilon \partial_p \eta \\ \varepsilon \partial_x \pi & \varepsilon \partial_y \pi & 1 + \varepsilon \partial_p \pi \end{pmatrix} .$$

Le plan de contact K_v en v est $\text{Span} \{ \partial_x + p \partial_y, \partial_p \} = \text{Span} \{ (1, p, 0), (0, 0, 1) \}$. Sa transformée $T_v \varphi(K_v)$ par $T_v \varphi$ est

$$\text{Span} \left\{ \begin{array}{l} (1 + \varepsilon \partial_x \xi + p \varepsilon \partial_y \xi, \varepsilon \partial_x \eta + p(1 + \varepsilon \partial_y \eta), \varepsilon \partial_x \pi + p \varepsilon \partial_y \pi), \\ (\varepsilon \partial_p \xi, \varepsilon \partial_p \eta, 1 + \varepsilon \partial_p \pi) \end{array} \right\}$$

et l'on veut qu'elle soit égale au plan de contact $K_{\varphi(v)}$ en $\varphi(v)$ qui est

$$\text{Span} \{ \partial_x + (p + \varepsilon \pi) \partial_y, \partial_p \} = \text{Span} \{ (1, p + \varepsilon \pi, 0), (0, 0, 1) \} .$$

On écrit donc que les générateurs de $T_v \varphi(K_v)$ sont dans $K_{\varphi(v)}$, i.e.

$$\begin{aligned} (1 + \varepsilon \partial_x \xi + p \varepsilon \partial_y \xi, \varepsilon \partial_x \eta + p(1 + \varepsilon \partial_y \eta), \varepsilon \partial_x \pi + p \varepsilon \partial_y \pi) &= a(1, p + \varepsilon \pi, 0) + b(0, 0, 1) \\ (\varepsilon \partial_p \xi, \varepsilon \partial_p \eta, 1 + \varepsilon \partial_p \pi) &= c(1, p + \varepsilon \pi, 0) + d(0, 0, 1) . \end{aligned}$$

On en tire d'abord $d = 1 + \varepsilon \partial_p \pi$ (qui fixe d) puis $c = \varepsilon \partial_p \xi$ et $c(p + \varepsilon \pi) = \varepsilon \partial_p \eta$, soit $(\varepsilon \partial_p \xi)(p + \varepsilon \pi) \sim \varepsilon p \partial_p \xi = \varepsilon \partial_p \eta$ (car $\varepsilon \neq 0$ mais $\varepsilon^2 = 0$) et donc

$$p \partial_p \xi = \partial_p \eta ,$$

ce qui est la première condition. On en tire ensuite d'abord $b = \varepsilon \partial_x \pi + p \varepsilon \partial_y \pi$ (qui fixe b) puis $a = 1 + \varepsilon \partial_x \xi + p \varepsilon \partial_y \xi$ et $a(p + \varepsilon \pi) = \varepsilon \partial_x \eta + p(1 + \varepsilon \partial_y \eta)$, soit

$$\begin{aligned} (1 + \varepsilon \partial_x \xi + p \varepsilon \partial_y \xi)(p + \varepsilon \pi) &\sim \varepsilon \pi + p(1 + \varepsilon \partial_x \xi + p \varepsilon \partial_y \xi) = \varepsilon \partial_x \eta + p(1 + \varepsilon \partial_y \eta) \\ \pi + p(\partial_x \xi + p \partial_y \xi) &= \partial_x \eta + p(\partial_y \eta) \\ \pi &= \partial_x \eta - p(\partial_x \xi) + p(\partial_y \eta) - p^2(\partial_y \xi) , \end{aligned}$$

ce qui est la seconde condition.

Évidemment, un X de contact peut être localement ou même globalement nul ce qui correspond à un flot localement ou globalement égal à l'identité. La condition d'être non nul est une condition ouverte à cause des hypothèses de régularité et si U est l'ouvert de \mathbb{V} où $X \neq 0$, X ne peut pas être à la fois de contact et "horizontal" localement sur U , i.e. satisfaire $\omega(X) \equiv 0$. En effet $\omega(X) \equiv 0$ localement signifie que $\eta \equiv p\xi$, ce qui implique $\partial_p \eta \equiv p(\partial_p \xi) + \xi$. Mais si X est de contact $\partial_p \eta \equiv p(\partial_p \xi)$ et donc $\xi \equiv 0$, et donc $\eta \equiv 0$ si X est horizontal, et donc $\pi \equiv 0$ si X est de contact. Cela montre qu'un champ de contact est *presque partout transverse* à la distribution de contact \mathcal{K} là où il n'est pas localement nul. Mais cela n'empêche évidemment pas X d'être dans \mathcal{K} en des points *particuliers*. Pour cela il suffit que la condition $p = \frac{\eta}{\xi}$

soit satisfaite au point considéré. Plus bas, nous verrons par exemple le cas de champ de contacts où ξ et η sont indépendants de p et où il y a donc, pour chaque point (x, y) du plan de base, une (et une seule) valeur de p telle que $\eta(x, y) = p\xi(x, y)$. Pour cette valeur particulière, X est horizontal.

L'exemple prototypique d'un champ de contact est le *champ caractéristique de Reeb* $\chi = \frac{\partial}{\partial y}$ défini à la section 4 du chapitre 3 qui est partout non nul et partout transverse à la distribution \mathcal{K} . Pour χ , $\eta_\chi \equiv 1$ et $\xi_\chi \equiv 0$, $\pi_\chi \equiv 0$ et donc χ satisfait trivialement les conditions. Ce champ de Reeb χ existe pour toute structure de contact et est défini par $i_\chi d\omega \equiv 0$ et $i_\chi \omega = \omega(\chi) \equiv 1$ (normalisation), ce qui implique

$$\mathcal{L}_\chi \omega = i_\chi d\omega + d(i_\chi \omega) \equiv 0 ,$$

autrement dit $\lambda \equiv 0$.²⁷ χ étant partout transverse à \mathcal{K} , le fibré tangent de \mathbb{V} se décompose en somme directe $T\mathbb{V} = \mathcal{K} \oplus \mathbb{R}\chi$ et tout champ X peut s'écrire $X = f\chi + Y$ où f est une certaine fonction C^∞ sur \mathbb{V} et Y un champ horizontal. Les champs $f\chi = (\xi \equiv 0, \eta = f, \pi \equiv 0)$ multiples de χ sont de contact si $\partial_x f = -p\partial_y f$ et $\partial_p f = 0$. On voit que pour respecter la transversalité $X \pitchfork \mathcal{K}$ si $X \neq 0$ il faut que Y s'annule avec f , ce qui définit des domaines où X est nul avec un flot agissant comme identité.

2.3.2. La fonction génératrice.

On montre alors que les X de contacts sont *caractérisés* par la fonction $f = \omega(X)$ et que X est reconstructible à partir de f par la formule

$$\begin{aligned} X_f &= \xi \partial_x + \eta \partial_y + \pi \partial_p \\ &= -(\partial_p f) \partial_x + (f - p\partial_p f) \partial_y + (\partial_x f + p\partial_y f) \partial_p . \end{aligned}$$

Si $\partial_x f = -p\partial_y f$ et $\partial_p f = 0$ on a bien $X = f\partial_y = f\chi$.²⁸ On vérifie trivialement que X_f est bien de contact. En effet,

$$\begin{aligned} \partial_p \eta - p(\partial_p \xi) &= \partial_p f - \partial_p f - p\partial_{p^2}^2 f + p\partial_{p^2}^2 f = 0 \\ \partial_y \eta - p(\partial_y \xi) &= \partial_y f - p\partial_{yp}^2 f + p\partial_{yp}^2 f = \partial_y f = \lambda \\ -\pi + \partial_x \eta - p(\partial_x \xi) &= -(\partial_x f + p\partial_y f) + \partial_x f - p\partial_{xp}^2 f + p\partial_{xp}^2 f \\ &= -p\partial_y f = -\lambda p . \end{aligned}$$

On vérifie aussi que la correspondance $X \rightarrow f_X$ et $f \rightarrow X_f$ est bijective. Pour le voir on utilise les formules $i_{X_f} \omega = \omega(X_f) = f$ et $i_{X_f} d\omega = df(\chi)\omega - df$. La première

27. On vérifie que pour notre ω de \mathbb{V}_J , $\omega(\chi) \equiv 1$ implique $\eta - p\xi \equiv 1$ et $i_\chi d\omega \equiv 0$ implique $\xi dp - \pi dx = 0$, donc $\xi \equiv 0$, $\pi \equiv 0$ et donc $\eta \equiv 1$.

28. Cette formule est pour la géométrie de contact l'équivalent de la correspondance entre hamiltoniens et champs hamiltoniens en géométrie symplectique puisqu'un hamiltonien H sur le plan de phases (x, p) définit le champ hamiltonien $X_H = -(\partial_p H) \partial_x + (\partial_x H) \partial_p$.

formule n'est que la définition de f et la seconde²⁹ se vérifie facilement :

$$d\omega(X_f, X') = \xi_f \pi' - \xi' \pi_f$$

alors que

$$df(\chi)\omega(X') = (\partial_y f)(\eta' - p\xi')$$

et

$$df(X') = (\partial_x f)\xi' + (\partial_y f)\eta' + (\partial_p f)\pi' .$$

Le facteur de ξ' dans $df(\chi)\omega - df$ est $-p(\partial_y f) - \partial_x f = -\pi$, celui de π' est $-\partial_p f = \xi$ et celui de η' est $\partial_y f - \partial_y f = 0$, ce qui donne l'égalité cherchée. On remarquera que celle-ci montre de façon plus intrinsèque que X_f est bien de contact car $\mathcal{L}_{X_f}\omega = i_{X_f}d\omega + d(i_{X_f}\omega)$ devient

$$\mathcal{L}_{X_f}\omega = df(\chi)\omega - df + df = df(\chi)\omega$$

qui est bien de la forme $\lambda\omega$ avec $\lambda = df(\chi)$. On a $f_{X_f} = f$ puisque $\omega(X_f) = f$ et, dans l'autre sens, $X_{f_X} = X$ car, X étant de contact et f_X étant $\eta - p\xi$, on a :

$$\begin{aligned} \xi_{f_X} &= -\partial_p f = -\partial_p \eta + \xi + p\partial_p \xi = \xi \\ \eta_{f_X} &= f - p\partial_p f = \eta - p\xi + p\xi = \eta \\ \pi_{f_X} &= \partial_x f + p\partial_y f = \partial_x \eta - p\partial_x \xi + p(\partial_y \eta - p\partial_y \xi) = \pi . \end{aligned}$$

Cette correspondance entre fonctions f et champs de contact X faisant des f les "fonctions génératrices" des X (comme les hamiltoniens sont les fonctions génératrices des champs hamiltoniens) permet de définir, comme dans le cas symplectique, une structure d'algèbre de Lie sur l'espace $\mathcal{C}^\infty(\mathbb{V})$ des f au moyen du crochet de Lie $\{f, g\} = \omega([X_f, X_g])$, le développement de la formule donnant

$$\{f, g\} = \omega([X_f, X_g]) = \partial_x f \partial_p g - \partial_p f \partial_x g + (f - p\partial_p f) \partial_y g - (g - p\partial_p g) \partial_y f$$

En mécanique hamiltonienne dans $\mathbb{R}_{(x,p)}^2$, on a

$$\{f, g\} = \partial_x f \partial_p g - \partial_p f \partial_x g = d\omega(X_f, X_g) .$$

Il s'agit bien de la même formule car

$$d\omega(X_f, X_g) = X_f \omega(X_g) - X_g \omega(X_f) - \omega([X_f, X_g])$$

donne

$$X_f(g) - X_g(f) - \omega([X_f, X_g])$$

et, comme

$$d\omega(X_f, X_g) = X_f(g) = -X_g(f) ,$$

on a en fait

$$0 = -X_g(f) - \omega([X_f, X_g]) ,$$

29. Elle correspond en mécanique hamiltonienne dans $\mathbb{R}_{(x,p)}^2$ à $i_{X_H}d\omega = -df$.

soit

$$\omega([X_f, X_g]) = -X_g(f) = X_f(g) = d\omega(X_f, X_g) .$$

2.3.3. Relèvement des champs de l'espace de base.

Dans la vaste littérature à ce sujet, le lecteur intéressé à des précisions pourra consulter par exemple le texte de Boris Doubrov et Boris Komrakov [150]. Ces auteurs étudient en particulier les champs de contact qui *relèvent* des champs de vecteurs X_0 dans le plan de base $\mathbb{R}_{(x,y)}^2$. Dans ce cas, $X_0 = \xi_0 \partial_x + \eta_0 \partial_y$ avec ξ_0, η_0 des fonctions de (x, y) ne dépendant pas de p . La fonction caractéristique $f = \eta_0 - p\xi_0$ garde un sens et on peut donc considérer le champ de contact X_f , en gardant à l'esprit que f est dans ce cas *linéaire* en p . Comme

$$X_f = -(\partial_p f) \partial_x + (f - p\partial_p f) \partial_y + (\partial_x f + p\partial_y f) \partial_p$$

et que les fonctions ξ_0, η_0 ne dépendent pas de p , on a

$$\begin{aligned} -\partial_p f &= \xi_0 \\ f - p\partial_p f &= \eta_0 - p\xi_0 + p\xi_0 = \eta_0 \end{aligned}$$

et donc X_f relève bien X_0 . La composante verticale est bien entendu

$$\pi = \partial_x \eta_0 - p\partial_x \xi_0 + p(\partial_y \eta_0 - p\partial_y \xi_0) .$$

Il est intéressant de vérifier que si $\gamma(t) = (x(t), y(t))$ est une intégrale de X_0 dans $\mathbb{R}_{(x,y)}^2$ alors sa relevée legendrienne $\Gamma(t)$ est une intégrale de X_f . Montrons-le naïvement là où $\xi_0 \neq 0$. Par hypothèse, $\dot{x}(t) = \xi_0$ et $\dot{y}(t) = \eta_0$. Quand on relève $\gamma(t)$ on obtient $\Gamma(t) = \left(x(t), y(t), p(t) = \frac{\dot{y}(t)}{\dot{x}(t)} = \frac{\eta_0}{\xi_0}\right)$ dont le vecteur tangent est $(\dot{x}(t), \dot{y}(t), \dot{p}(t))$. Mais $\dot{p}(t) = \frac{\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)}{\dot{x}(t)^2}$ et

$$\begin{aligned} \ddot{x}(t) &= (\partial_x \xi_0) \dot{x}(t) + (\partial_y \xi_0) \dot{y}(t) = \xi_0 (\partial_x \xi_0) + \eta_0 (\partial_y \xi_0) \\ \ddot{y}(t) &= (\partial_x \eta_0) \dot{x}(t) + (\partial_y \eta_0) \dot{y}(t) = \xi_0 (\partial_x \eta_0) + \eta_0 (\partial_y \eta_0) \\ \dot{p}(t) &= \frac{1}{\xi_0^2} (\xi_0 (\xi_0 (\partial_x \eta_0) + \eta_0 (\partial_y \eta_0)) - \eta_0 (\xi_0 (\partial_x \xi_0) + \eta_0 (\partial_y \xi_0))) \\ &= \partial_x \eta_0 + p\partial_y \eta_0 - p\partial_x \xi_0 - p^2 \partial_y \xi_0 = \pi . \end{aligned}$$

En relevant le “portrait de phase” de X_0 (i.e. le feuilletage du plan de base par les trajectoires $\gamma(t)$), on obtient donc une surface Σ de \mathbb{V} (qui n'est évidemment pas une surface intégrale de ω puisque celles-ci n'existent pas) sur laquelle la distribution \mathcal{K} des plans de contact découpe les relevées legendriennes $\Gamma(t)$ des $\gamma(t)$.

Explicitons quelques exemples. Si X_0 est un champ de pente constante p_c , i.e. dont les trajectoires $\gamma(t)$ sont géométriquement des droites paramétrées qui satisfont partout $\eta_0(x, y) = p_c \xi_0(x, y)$, alors $\partial_x \eta_0 = p_c \partial_x \xi_0$ et $\partial_y \eta_0 = p_c \partial_y \xi_0$ et par conséquent

$$\pi = \partial_x \eta_0 + p \partial_y \eta_0 - p \partial_x \xi_0 - p^2 \partial_y \xi_0 = p_c \partial_x \xi_0 + p_c^2 \partial_y \xi_0 - p_c \partial_x \xi_0 - p_c^2 \partial_y \xi_0 = 0.$$

Le champ de contact X_f relevant X_0 est donc tout simplement X_0 dans chaque plan $p = \text{cste}$. Pour la valeur spéciale $p = p_c$ les trajectoires $\Gamma(t)$ de $X_f = X_0$ dans le plan $p = p_c$ sont les relevées legendriennes des trajectoires $\gamma(t)$ de X_0 . Il est normal qu'elles soient des droites de p -composante nulle puisque la p -composante est associée à la courbure des $\gamma(t)$ et que celle-ci est nulle.

Comme autre exemple où les trajectoires $\gamma(t)$ de X_0 sont des droites mais non plus parallèles, considérons dans le plan de base le champ radial X_0 ayant pour trajectoires les rayons (en coordonnées polaires)

$$x(t) = t \cos(\theta), y(t) = t \sin(\theta),$$

$t \geq 0$, θ fixé. On a

$$X_0 = (\xi_0(t) = \cos(\theta), \eta_0(t) = \sin(\theta)).$$

et donc $p(t) = \tan(\theta)$ et $\dot{p}(t) = 0$. Le champ de contact X_f relevant X_0 est encore X_0 dans chaque plan $p = \text{cste}$ et quand on relève les trajectoires $\gamma(t)$ en $\Gamma(t)$, on obtient les rayons, mais ventilés le long de l'axe des p suivant la contrainte $p = \tan(\theta)$.

Pour illustrer les effets de la courbure des $\gamma(t)$ considérons le champ

$$x(t) = r \cos(t), y(t) = r \sin(t), r = \text{cste} \geq 0,$$

i.e.

$$X_0 = (\xi_0 = -r \sin(t), \eta_0 = r \cos(t)),$$

dont les trajectoires sont les cercles concentriques centrés sur l'origine. On a

$$p(t) = -\frac{r \cos(t)}{r \sin(t)} = -\cot(t)$$

et

$$\dot{p}(t) = \frac{1}{\sin^2(t)} \geq 1, \dot{p}(0) = \infty, \dot{p}\left(\frac{\pi}{2}\right) = 1.$$

Pour calculer la composante π de X_f , on utilise les formules standard de passage des coordonnées cartésiennes à des coordonnées polaires. Si $f(x, y) = F(r, \theta)$, on a pour $r \neq 0$ (nous supposons que c'est le cas)

$$\begin{aligned} \partial_x f &= \frac{x}{r} \partial_r F - \frac{y}{r^2} \partial_\theta F \\ \partial_y f &= \frac{y}{r} \partial_r F + \frac{x}{r^2} \partial_\theta F. \end{aligned}$$

Dans notre cas

$$\begin{aligned}\partial_x \xi_0 &= -\frac{x}{r} \sin(\theta) + \frac{y}{r^2} r \cos(\theta) = 0, \\ \partial_y \xi_0 &= -\frac{y}{r} \sin(\theta) - \frac{x}{r^2} r \cos(\theta) = -1, \\ \partial_x \eta_0 &= \frac{x}{r} \cos(\theta) + \frac{y}{r^2} r \sin(\theta) = 1, \\ \partial_y \eta_0 &= \frac{y}{r} \cos(\theta) - \frac{x}{r^2} r \sin(\theta) = 0,\end{aligned}$$

et donc

$$\begin{aligned}\pi &= \partial_x \eta_0 + p \partial_y \eta_0 - p \partial_x \xi_0 - p^2 \partial_y \xi_0 \\ &= 1 + p^2.\end{aligned}$$

On vérifie que lorsque $p = -\cot(t)$ on retrouve bien

$$\pi = 1 + \frac{\cos^2(t)}{\sin^2(t)} = \frac{1}{\sin^2(t)} = \dot{p}(t).$$

La figure 2 illustre cet exemple. Elle représente (en rouge) l'arc de trajectoire $\gamma(t)$ de X_0 pour $r = 1$ et $\theta \in [\frac{\pi}{16}, \frac{3\pi}{16}]$, ainsi que sa relevée legendrienne $\Gamma(t)$ (en bleu). Elle montre aussi la fibre "au dessus" (ou plutôt dans ce cas en-dessous car $p < 0$ pour l'intervalle choisi de θ) du point $\theta = \frac{\pi}{8}$ ($x \sim 0.924$, $y \sim 0.383$). Pour les 3 points v , $p = 0, -1, -\cot(\frac{\pi}{8}) \sim -2.414$, elle représente d'une part le plan de contact K_v et d'autre part le champ de contact X_f et (en transparence) le plan P_f qu'il engendre avec l'axe vertical des p . Pour des raisons de lisibilité, les X_f sont scalés par le facteur 0.2. On voit comment les plans K_v et P_f qui sont bien transverses pour $p = 0$ se rapprochent et fusionnent pour $p = -\cot(\frac{\pi}{8})$, X_f devenant inclus dans K_v .

La figure 3 représente le graphique sous un autre point de vue qui rend ce phénomène plus visible.

2.3.4. Surfaces convexes.

Les champs de contact en dimension 3 permettent de définir une notion importante qui est celle de *surface convexe*. Une surface plongée S dans une variété de contact M de dimension 3 est dite *convexe* s'il existe un champ de contact X qui lui est partout transverse (ce qui exige évidemment que X soit partout $\neq 0$ sur S). Les points $x \in M$ où $X_x \in K_x$ engendrent alors une séparatrice ("dividing set") Γ_S de S dont on montre que la classe d'isotopie est indépendante du X choisi. Considérons par exemple le cas de Heisenberg non polarisé

$$\omega = dz + xdy - ydx = dz + r^2 d\theta$$

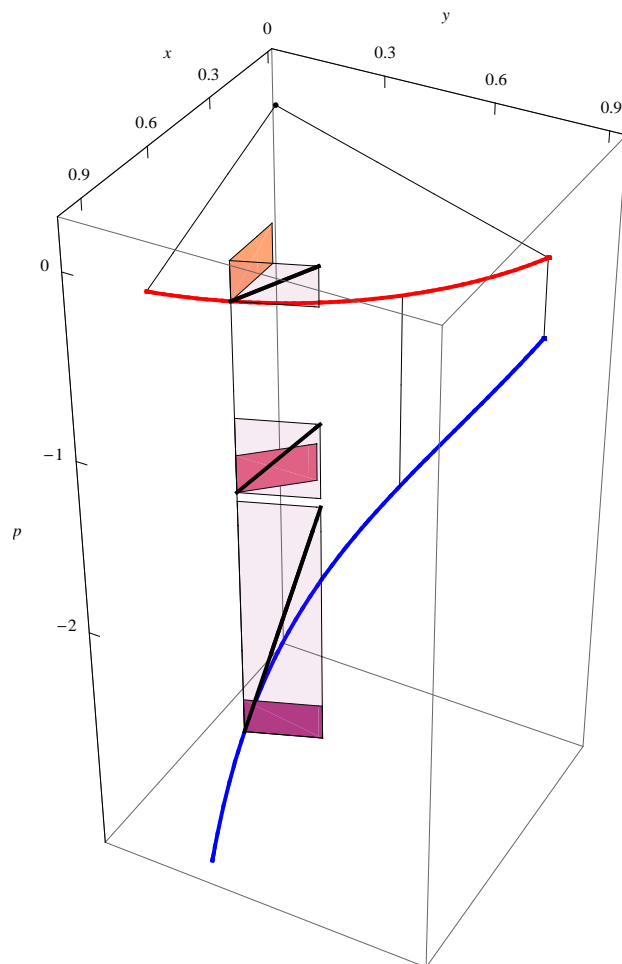


FIGURE 2. En rouge l'arc de trajectoire $\gamma(t)$ de X_0 pour $r = 1$ et $\theta \in [\frac{\pi}{16}, \frac{3\pi}{16}]$, ainsi que (en bleu et restreinte à $\theta \in [\frac{1.5\pi}{16}, \frac{3\pi}{16}]$ pour rendre l'image plus lisible) sa relevée legendrienne $\Gamma(t)$. Les échelles des axes x, y et de l'axe p ne sont pas les mêmes. Pour les 3 points $v, p = 0, -1, -\cot(\frac{\pi}{8}) \sim -2.414$, de la fibre du point $\theta = \frac{\pi}{8}$ ($x \sim 0.924, y \sim 0.383$) on représente (par un rectangle opaque) le plan de contact K_v ainsi que le champ de contact X_f et (par un rectangle transparent) le plan P_f qu'il engendre avec l'axe vertical des p . Pour des raisons de lisibilité, les X_f sont scalés par le facteur 0.2 et comme X_0 est de norme 1, il est de norme 0.2 dans l'image. Les plans K_v et P_f transverses pour $p = 0$ se rapprochent et fusionnent pour $p = -\cot(\frac{\pi}{8})$, X_f devenant inclus dans K_v .

(en coordonnées cylindriques semi-polaires) et considérons le champ

$$X = r\partial_r + 2z\partial_z .$$

Il s'agit bien d'un champ de contact car

$$\mathcal{L}_X\omega = i_X d\omega + d(i_X\omega) = i_X(2rdr \wedge d\theta) + d(2z) = 2r^2d\theta + 2dz = 2\omega$$

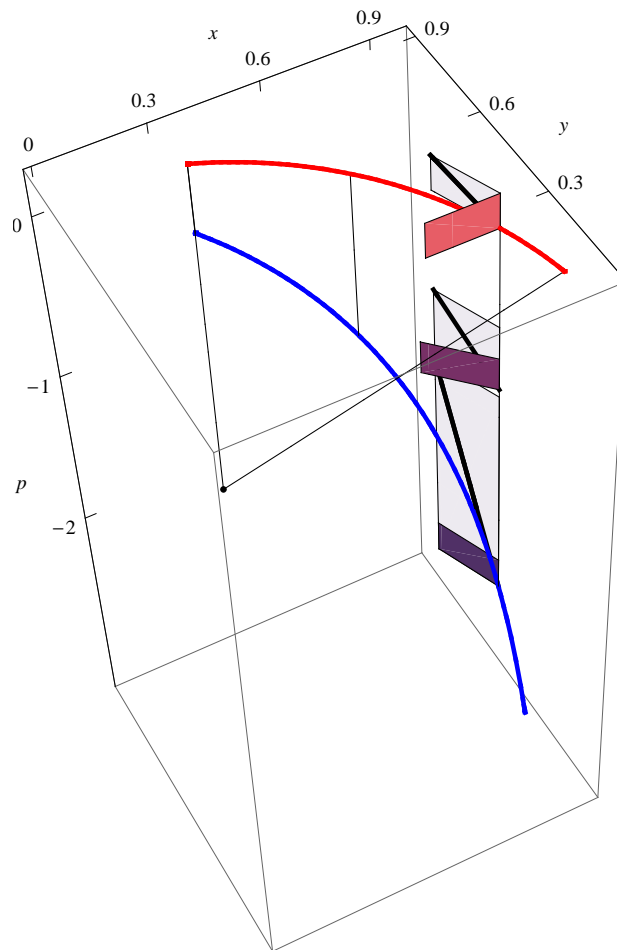


FIGURE 3. Le graphique de la figure 2 sous un autre point de vue.

Par définition, les cylindres S de rayon $r_0 \neq 0$ sont transverses à X et sont donc des surfaces convexes. La séparatrice est le cercle $z = 0$ où $X = r\partial_r \in \mathcal{K}$.

Les surfaces convexes ont été étudiées par de nombreux auteurs dont Ko Honda et Emmanuel Giroux. Nous y reviendrons plus bas dans la section 4.8.

Nous arrêterons ici ces quelques remarques sur les champs de contact et nous renvoyons à Doubrov-Komrakov [150] pour la classification de leurs algèbres de Lie.

3. Géométrie des espaces de jets et distributions de Goursat

3.1. Jets et géométrie des équations différentielles

Nous avons vu à la section 2.10 du chapitre 3³⁰ que la géométrie des structures de contact sur \mathbb{R}^{2n+1} fournit le cadre naturel de la résolution des équations différentielles du premier ordre. Rappelons en effet que lorsque l'on cherche la solution d'une EDO $F(x, y, \dot{y}) = 0$, on considère d'un côté la surface S de $J^1(\mathbb{R}, \mathbb{R})$ d'équation $F(x, y, p) = 0$ et d'un autre côté la structure de contact $dy - p dx$. Les solutions de l'EDO sont les courbes intégrales de \mathcal{K} satisfaisant $dF = 0$ (i.e. tangentes à S). Si S et \mathcal{K} sont transverses, le champ de droites $TS \cap \mathcal{K}$ définit la solution. De même, si l'on cherche à résoudre une EDP de n variables x_i , on considère $J^1(\mathbb{R}^n, \mathbb{R})$ et les variétés legendriennes de $dy - \sum_i p_i dx_i$ tangentes à $F(x_i, y, \partial_{x_i} y) = 0$. La résolution des EDO et EDP d'ordre supérieur conduit à considérer des espaces de jets d'ordre supérieur. Dans tous les cas, il est essentiel de comprendre la position de l'hypersurface particulière $F = 0$ par rapport à la géométrie intrinsèque et universelle de l'espace de jets approprié puisque les solutions satisfont des conditions de tangence entre les deux.

Cette approche géométrique a commencé à être développée par Sophus Lie au début des années 1870 et a ouvert cet immense champ de recherche qui a explosé à partir des années 1950. Nous voudrions y faire une petite incursion.

3.2. Les dérivées “cachées” de Leibniz à Montgomery

Depuis le début, nos modèles neurogéométriques ont consisté d'une façon ou d'une autre à utiliser la géométrie très spéciale des espaces de jets qui géométrise les développements de Taylor et l'extraordinaire conception leibnizienne-eulérienne consistant

- (i) à introduire les dérivées successives comme des nouvelles variables indépendantes,
- (ii) à enrichir le calcul symbolique à la Leibniz avec ces nouveaux symboles,
- (iii) puis à introduire des conditions imposant d'interpréter numériquement ces symboles comme les dérivées.

Pour citer à nouveau l'expression de Richard Montgomery et Michail Zhitomirskii dans [377] à propos de la nouvelle variable p dans la 1-forme de contact $\omega = dy - p dx$, on peut dire que

“ p has the ‘hidden’ meaning of $\frac{dy}{dx}$ ”.

Comme nous l'avons vu à la section 8 du chapitre 9 sur l'histoire du Calcul des Variations, tout se trouve déjà dans le génial *Methodus* d'Euler. Ce qu'ont ajouté les travaux postérieurs (tout aussi géniaux) est l'idée d'une géométrie universelle

30. À propos de l'article de Thom [531] “Sur les équations différentielles multiformes et leurs intégrales singulières” de 1972.

des systèmes de Pfaff qui expriment les conditions *d'implémentation concrète* de ces calculs symboliques.

Partis de courbes dans le plan décrites comme des graphes de fonctions $y = f(x)$, nous avons d'abord introduit la variable p et la 1-forme $\omega = dy - p dx$ dont l'annulation garantit que la variable indépendante p est interprétée comme la dérivée $\frac{dy}{dx}$. Cela nous a conduit à l'espace des 1-jets avec sa structure de contact canonique. Puis nous avons introduit la courbure κ comme nouvelle variable indépendante avec le système de Pfaff imposant son interprétation comme courbure. Cela nous a conduit à l'espace des 2-jets avec sa structure d'Engel canonique. Nous voudrions maintenant dire un mot de la façon dont cette géométrisation des développements de Taylor au moyen des espaces de jets peut se continuer au-delà de l'ordre 2.

On utilise la notion de "prolongement" de Cartan. Soit (M, \mathfrak{D}) une distribution (complètement non holonome) de type (k, n) . Si $0 < \ell < k$ on peut regarder les sous-vecteurs V_x de dimension ℓ de \mathfrak{D}_x , i.e. la grassmannienne $Gr_\ell(\mathfrak{D}_x)$. On obtient ainsi un fibré au-dessus de M de fibre $Gr_\ell(\mathfrak{D}_x)$ qui est un "prolongement" de M si l'on introduit les formes différentielles exprimant que ces fibrés sont comme des fibrés de jets. Les structures de contact et d'Engel que nous avons rencontrées sont bien de ce type. Pour les structures de contact, on part de \mathbb{R}^2 de dimension $n = 2$ et de coordonnées $a = (x, y)$ avec la distribution triviale $(2, 2)$ (i.e. le fibré tangent $T\mathbb{R}^2$), on prend $\ell = n - 1 = 1$ (i.e. les droites des $T_a\mathbb{R}^2$) et l'on obtient $M = \mathbb{P}(T^*\mathbb{R}^2)$ de dimension $2n - 1 = 3$ avec sa nouvelle coordonnée "verticale" p pour les fibres³¹ ainsi que sa structure de contact \mathcal{K} noyau de $\omega = dy - p dx$ (1-jets). L'annulation de ω force à interpréter p comme une tangente $p = \frac{dy}{dx}$.

Ensuite, on prolonge cette distribution en prenant des droites dans les K_x définies par l'annulation de $\tau = dp - \kappa dx$ et l'on obtient la structure d'Engel $E \rightarrow M \rightarrow \mathbb{R}^2$ avec sa nouvelle coordonnée "verticale" κ que la condition $\tau = 0$ force à interpréter comme une courbure $\kappa = \frac{dp}{dx} = \frac{d^2y}{dx^2}$ si p est déjà forcé à être une tangente $p = \frac{dy}{dx}$ par la condition $\omega = 0$. Et l'on itère.

3.3. Distributions et drapeaux de Goursat

De façon générale, si (M, \mathfrak{D}) est une distribution de type (k, n) on peut considérer deux drapeaux. Le premier est le *drapeau de Lie* introduit à la section 1.3 :

$$\mathfrak{D} = \mathfrak{D}^0, \quad \mathfrak{D}^1 = \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}], \quad \mathfrak{D}^j = \mathfrak{D}^{j-1} + [\mathfrak{D}, \mathfrak{D}^{j-1}] = \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}^{j-1}], \quad \dots$$

\mathfrak{D}^j est donc l'espace engendré par les commutateurs $[X, [Y, [Z, \dots]]]$ d'éléments de \mathfrak{D} de l'ordre 0 (un commutateur d'ordre 0 est un $X \in \mathfrak{D}$) à l'ordre j . Même si les k -plans \mathfrak{D}_x des $T_x M$ sont tous de dimension k , cela n'implique évidemment pas que

31. La valeur $p = \infty$ est autorisés si les fibres sont des droites projectives compactifiant les fibres de l'espace des 1-jets : cf. la section 2.1 du chapitre 3.

les \mathfrak{D}_x^j soient de dimension constante. Le “petit” vecteur de croissance de \mathfrak{D} en x est

$$\{\nu_0 = k, \nu_1(x) = \dim(\mathfrak{D}_x^1), \dots\} .$$

Si les $\nu_j(x)$ sont *constants* on a une première notion de régularité, disons L -régularité (L pour “Lie”) pour la distinguer d’autres formes de régularités qui vont intervenir.

Le second drapeau, dit *drapeau dérivé*, est

$$\mathfrak{D}^{(0)} = \mathfrak{D}, \mathfrak{D}^{(1)} = \mathfrak{D}^{(0)} + [\mathfrak{D}^{(0)}, \mathfrak{D}^{(0)}] = \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}], \mathfrak{D}^{(j)} = \mathfrak{D}^{(j-1)} + [\mathfrak{D}^{(j-1)}, \mathfrak{D}^{(j-1)}],$$

et le “grand” vecteur de croissance de \mathfrak{D} en x est

$$\{n_0 = k, n_1(x) = \dim(\mathfrak{D}_x^{(1)}), \dots\} .$$

On a $\mathfrak{D}^j \subset \mathfrak{D}^{(j)}$ car $\mathfrak{D}^{(j)}$ contient les crochets de $[\mathfrak{D}^{(j-1)}, \mathfrak{D}^{(j-1)}]$ et pas seulement ceux de $[\mathfrak{D}, \mathfrak{D}^{(j-1)}]$.³²

Là encore, si les $n_j(x)$ sont *constants*, on a une seconde notion de régularité, disons D -régularité (D pour “drapeau”). On dit aussi que \mathfrak{D} est *totale*ment régulière. Si l’on a les deux régularités et si les $n_j = \nu_j$ (les “petits” et “grands” vecteurs de croissance sont constants et égaux) on peut parler de LD -régularité.

Si \mathfrak{D} est complètement non holonome son drapeau dérivé s’arrête en $\mathfrak{D}^{(s)} = TM$, et, en cas de D -régularité, son “grand” vecteur de croissance est

$$(n_0 = k, \dots, n_s = \dim(\mathfrak{D}^{(s)}) = n) .$$

On dit alors que (M, \mathfrak{D}) de type (k, n) est une *distribution de Goursat* si le grand vecteur de croissance est

$$n_{j+1} = n_j + 1 ,$$

autrement dit si à chaque pas l’adjonction des crochets augmente la dimension exactement de 1. Clairement, les espaces de jets $J^s(\mathbb{R}, \mathbb{R})$ de courbes du plan qui sont des graphes $y = f(x)$ de fonctions $C^\infty f : \mathbb{R} \rightarrow \mathbb{R}$ constituent des distributions de Goursat.

Un *théorème fondamental d’Élie Cartan* démontre en partie la réciproque en affirmant que les espaces de jets $J^s(\mathbb{R}, \mathbb{R})$ sont localement les formes normales³³ locales *génériques* des distributions de Goursat \mathfrak{G} de type $(2, n = 2 + s)$, autrement dit qu’il existe un *ouvert dense* de l’espace de ces \mathfrak{G} admettant cette forme normale (cf. les grands articles, classiques, de Cartan [90], [92], [93]). Les points x d’une \mathfrak{G} sur M qui admettent cette forme normale locale sont dits *réguliers*.³⁴

Les espaces de jets $J^s(\mathbb{R}, \mathbb{R})$ avec leur forme normale de Cartan sont des \mathfrak{G} dont le petit vecteur de croissance est égal au grand vecteur de croissance. Un théorème

32. Nous donnerons plus bas à la section 3.4 un exemple simple de Giaro, Kumpera, Ruiz où $\mathfrak{D}^j \subsetneq \mathfrak{D}^{(j)}$.

33. On utilisait aussi à l’époque l’expression de forme “canonique”.

34. Il y a donc plusieurs notions de régularité qu’il ne faut pas confondre même si elles interfèrent.

de Murray (1994) montre la réciproque : les formes normales de Cartan sont les seules à satisfaire cette propriété structurelle.

Remarque. Si l'on voulait être rigoureux, il faudrait préciser la notion de "localité" utilisée ici. Dans ce contexte, "local" signifie "sur un voisinage aussi petit que l'on veut d'un point x " et donc "sur un voisinage infinitésimal". Pour définir rigoureusement cette notion sans faire appel à l'analyse non standard, il faut considérer le filtre \mathcal{F} des voisinages U de x et prendre la limite sur \mathcal{F} des restrictions aux U de la structure géométrique Σ considérée. On parle alors du "germe" de Σ en x . Dans la suite de cette section "local" signifiera "germe". \square

Qui plus est, un *théorème de réduction de Zhitomirskii* de 1990 démontre que l'on peut ramener localement l'étude de toute distribution de Goursat \mathfrak{G} de type (k, m) à celle d'une \mathfrak{G} de type $(2, m - (k - 2))$. Ce sont donc les \mathfrak{G} de type $(2, n = 2 + s)$ dont il faut comprendre en détail la structure géométrique *locale* en étudiant comment elle peut être *singulière* (non générique) et différer de celle d'un espace de jets.

En approfondissant les choses, on a découvert que la généralisation des structures de Heisenberg et d'Engel au-delà des 2-jets n'était pas du tout aussi simple qu'on avait pu le croire au prime abord et que la contrainte d'être une distribution de Goursat était une contrainte *extrêmement forte*. Quand on quitte les espaces de jets avec leurs formes normales de Cartan, la géométrie des distributions de Goursat \mathfrak{G} devient compliquée et même "monstrueuse" pour reprendre une expression de Richard Montgomery. En dimensions 3 et 4 la propriété pour une distribution \mathfrak{D} d'être de Goursat est *générique*, valide sur un ouvert dense de l'espace des \mathfrak{D} muni de sa topologie naturelle et, qui plus est, les \mathfrak{G} sont partout régulières. Dès la dimension ≥ 5 ce n'est plus du tout le cas, les formes normales se multiplient, ne se réduisent plus à la forme normale de Cartan des espaces de jets et le sous-espace des \mathfrak{G} devient de codimension *infinie* dans l'espace des \mathfrak{D} . Cela est dû à un simple fait dimensionnel. On montre en effet qu'en dimension 5, i.e. pour le type $(2, 5)$, génériquement $\dim(\mathfrak{D}^{(1)}) = 3$ et $\dim(\mathfrak{D}^{(2)}) = 5$ et il y a donc un saut dimensionnel par rapport à la condition pour \mathfrak{D} d'être de Goursat $\dim(\mathfrak{D}^{(1)}) = 3$, $\dim(\mathfrak{D}^{(2)}) = 4$, $\dim(\mathfrak{D}^{(3)}) = 5$. Cela est intuitif car si \mathfrak{D} est engendrée par X_1 et X_2 on peut s'attendre à ce que au niveau de $\mathfrak{D}^{(2)}$ ont ait en général

$$X_1, X_2, [X_1, X_2] = X_3, [X_1, X_3] = X_4, [X_2, X_3] = X_5$$

sans relations particulières entre ces 5 vecteurs puisqu'on est en dimension 5 et qu'aucune contrainte purement dimensionnelle ne les empêche d'être linéairement indépendants.

Remarque. Il faut bien faire attention ici au fait que la notion de *codimension* d'un espace n'est pas intrinsèque mais "contextuelle", i.e. relative à un espace ambiant dont l'espace considéré est un sous-espace. L'espace $\mathcal{E}_{\mathfrak{D}}$ des distributions \mathfrak{D} sur M est un espace fonctionnel de dimension infinie et l'espace $\mathcal{E}_{\mathfrak{G}}$ des \mathfrak{G} est d'une certaine codimension relativement à lui. Mais dans la suite nous considérerons des

classes particulières \mathcal{F} de \mathfrak{G} possédant une certaine propriété et qui auront une certaine codimension, finie et même faible, relativement à $\mathcal{E}_{\mathfrak{G}}$. Ces \mathcal{F} seront a fortiori de codimension infinie dans $\mathcal{E}_{\mathfrak{D}}$ mais c'est leur codimension dans $\mathcal{E}_{\mathfrak{G}}$ qui sera pertinente. \square

On peut voir ce qui se passe en considérant la structure d'*itération* des jets $J^s(\mathbb{R}, \mathbb{R})$. Nous nous bornerons ici à quelques simples calculs illustratifs. Partons des 1-jets (dimension 3, Heisenberg) avec la variable de base x et la 1-forme

$$\omega = \omega_1 = dy - p dx .$$

La distribution \mathfrak{D} est engendrée, nous l'avons vu plusieurs fois, par

$$\{\mathfrak{t}_1 = \partial_x + p\partial_y, \mathfrak{t}_2 = \partial_p\}$$

avec $[\mathfrak{t}_1, \mathfrak{t}_2] = \mathfrak{t}_3 = -\partial_y$, $[\mathfrak{t}_1, \mathfrak{t}_3] = [\mathfrak{t}_2, \mathfrak{t}_3] = 0$.³⁵ On a donc, puisque $\partial_x = \mathfrak{t}_1 + p\mathfrak{t}_3$,

$$\mathfrak{D}^1 = \mathfrak{D}^{(1)} = \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}] = \text{Span}\{\mathfrak{t}_1, \mathfrak{t}_2\} + \text{Span}\{\mathfrak{t}_3\} = \text{Span}\{\partial_x, \partial_y, \partial_p\} = TM$$

Si l'on note $t \pitchfork u$ la relation d'indépendance linéaire entre deux vecteurs, on vérifie immédiatement

- que $\mathfrak{t}_1 \pitchfork \mathfrak{t}_2$ partout puisque ∂_p est \pitchfork au plan $\text{Span}\{\partial_x, \partial_y\}$ si $p \neq 0$ et $\partial_p \pitchfork \partial_x$ si $p = 0$,
- que $\mathfrak{t}_2 \pitchfork \mathfrak{t}_3$ partout et
- que $\mathfrak{t}_1 \pitchfork \mathfrak{t}_3$ si $p = 0$ mais également si $p \neq 0$.

On a donc une *LD*-régularité et les deux vecteurs de croissance sont partout $\{2, 3\}$.

Lorsque l'on passe aux 2-jets (dimension 4, Engel), on introduit en plus de ω_1 la nouvelle 1-forme $\tau = \omega_2 = dp - q dx$ et \mathfrak{D} est alors engendrée³⁶ par

$$\{\mathfrak{t}_1 = \partial_x + p\partial_y + q\partial_p, \mathfrak{t}_4 = \partial_q\}$$

avec les commutateurs (\mathfrak{t}_2 et \mathfrak{t}_3 restant les mêmes)

$$\begin{aligned} [\mathfrak{t}_1, \mathfrak{t}_4] &= -\partial_p = -\mathfrak{t}_2, \quad [\mathfrak{t}_1, \mathfrak{t}_2] = \mathfrak{t}_3 = -\partial_y, \\ [\mathfrak{t}_1, \mathfrak{t}_3] &= [\mathfrak{t}_2, \mathfrak{t}_3] = [\mathfrak{t}_2, \mathfrak{t}_4] = [\mathfrak{t}_3, \mathfrak{t}_4] = 0 . \end{aligned}$$

35. Depuis le début de l'ouvrage nous notons \mathfrak{t}_1 et \mathfrak{t}_2 les générateurs de \mathfrak{D} pour les 1-jets. Mais \mathfrak{t}_2 correspond à la troisième variable p . On aurait donc pu opter pour la notation $\mathfrak{t}_3 = \partial_p$ et $[\mathfrak{t}_1, \mathfrak{t}_3] = \mathfrak{t}_2 = -\partial_y$ qui serait plus facile à généraliser aux jets d'ordre supérieur. Nous préférons toutefois maintenir la notation initiale.

36. Pour rendre ces petits calculs plus lisibles, nous utilisons ici pour les nouvelles variables indépendantes forcées à être des dérivées successives par l'annulation du système de Pfaff définissant les jets, la notation p, q (et non plus κ, r, s, \dots remontant à Euler, plutôt que des variables numérotées x_3, x_4 , etc.

Donc

$$\begin{aligned}\mathfrak{D}^1 &= \mathfrak{D}^{(1)} = \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}] = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_4 \} + \text{Span} \{ \mathbf{t}_2 \} \\ &= \text{Span} \{ \mathbf{t}_1, \mathbf{t}_4, \mathbf{t}_2 \}, \\ \mathfrak{D}^2 &= \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}^1] = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_4 \} + \text{Span} \{ \mathbf{t}_2, \mathbf{t}_3 \} \\ &= \text{Span} \{ \mathbf{t}_1, \mathbf{t}_4, \mathbf{t}_2, \mathbf{t}_3 \} = TM\end{aligned}$$

car $\mathbf{t}_3 = [\mathbf{t}_1, \mathbf{t}_2] \in [\mathfrak{D}, \mathfrak{D}^1]$ et $\partial_x = \mathbf{t}_1 + p\mathbf{t}_3 - q\mathbf{t}_2$. Par ailleurs, pour la même raison,

$$\begin{aligned}\mathfrak{D}^{(2)} &= \mathfrak{D}^{(1)} + [\mathfrak{D}^{(1)}, \mathfrak{D}^{(1)}] = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_4, \mathbf{t}_2 \} + \text{Span} \{ \mathbf{t}_2, \mathbf{t}_3 \} \\ &= \text{Span} \{ \mathbf{t}_1, \mathbf{t}_4, \mathbf{t}_2, \mathbf{t}_3 \} = TM\end{aligned}$$

et, comme tous les vecteurs \mathbf{t}_j sont partout \pitchfork entre eux, il y a LD -régularité et les deux vecteurs de croissance sont partout $(2, 3, 4)$. On voit clairement comment fonctionne l'itération qui conduit aux distributions de Goursat $J^s(\mathbb{R}, \mathbb{R})$ (formes normales de Cartan) définies par le système de Pfaff $\{\omega_1 = 0, \dots, \omega_s = 0\}$.

Lorsque les deux vecteurs de croissance sont égaux et constants, on dit que \mathfrak{D} satisfait la condition *forte* de Goursat et l'on montre que cela implique que le modèle local soit celui des jets.

3.4. Apparition de singularités (Giaco-Kumpera-Ruiz)

Pour les dimensions 3 et 4 $J^1(\mathbb{R}, \mathbb{R})$ et $J^2(\mathbb{R}, \mathbb{R})$ sont les seules formes normales de Goursat mais pour la dimension 5 la situation se complique. Si l'on ajoute à ω_1 et ω_2 la nouvelle 1-forme $\omega_3 = dq - rdx$, \mathfrak{D} devient maintenant engendrée par

$$\{ \mathbf{t}_1 = \partial_x + p\partial_y + q\partial_p + r\partial_q, \mathbf{t}_5 = \partial_r \}$$

avec les relations de commutation

$$\begin{aligned}[\mathbf{t}_1, \mathbf{t}_5] &= -\partial_q = -\mathbf{t}_4, \quad [\mathbf{t}_1, \mathbf{t}_4] = -\partial_p = -\mathbf{t}_2, \quad [\mathbf{t}_1, \mathbf{t}_2] = \mathbf{t}_3 = -\partial_y, \\ [\mathbf{t}_1, \mathbf{t}_3] &= [\mathbf{t}_2, \mathbf{t}_3] = [\mathbf{t}_2, \mathbf{t}_4] = [\mathbf{t}_2, \mathbf{t}_5] = [\mathbf{t}_3, \mathbf{t}_4] = [\mathbf{t}_3, \mathbf{t}_5] = [\mathbf{t}_4, \mathbf{t}_5] = 0.\end{aligned}$$

On a donc

$$\begin{aligned}\mathfrak{D}^1 &= \mathfrak{D}^{(1)} = \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}] = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_5 \} + \text{Span} \{ \mathbf{t}_4 \} = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_5, \mathbf{t}_4 \}, \\ \mathfrak{D}^2 &= \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}^1] = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_5 \} + \text{Span} \{ \mathbf{t}_4, \mathbf{t}_2 \} = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_5, \mathbf{t}_4, \mathbf{t}_2 \}, \\ \mathfrak{D}^3 &= \mathfrak{D} + [\mathfrak{D}, \mathfrak{D}^3] = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_5 \} + \text{Span} \{ \mathbf{t}_4, \mathbf{t}_2, \mathbf{t}_3 \} = \text{Span} \{ \mathbf{t}_1, \mathbf{t}_5, \mathbf{t}_4, \mathbf{t}_2, \mathbf{t}_3 \} \\ &= TM\end{aligned}$$

car $\partial_x = \mathbf{t}_1 + p\mathbf{t}_3 + q\mathbf{t}_2 + r\mathbf{t}_4$ et par ailleurs

$$\begin{aligned}\mathfrak{D}^{(2)} &= \mathfrak{D}^{(1)} + [\mathfrak{D}^{(1)}, \mathfrak{D}^{(1)}] = \text{Span} \{ \mathfrak{t}_1, \mathfrak{t}_5, \mathfrak{t}_4 \} + \text{Span} \{ \mathfrak{t}_4, \mathfrak{t}_2 \} = \text{Span} \{ \mathfrak{t}_1, \mathfrak{t}_5, \mathfrak{t}_4, \mathfrak{t}_2 \} \\ \mathfrak{D}^{(3)} &= \mathfrak{D}^{(2)} + [\mathfrak{D}^{(2)}, \mathfrak{D}^{(2)}] = \text{Span} \{ \mathfrak{t}_1, \mathfrak{t}_5, \mathfrak{t}_4, \mathfrak{t}_2 \} + \text{Span} \{ \mathfrak{t}_4, \mathfrak{t}_2, \mathfrak{t}_3 \} \\ &= \text{Span} \{ \mathfrak{t}_1, \mathfrak{t}_5, \mathfrak{t}_4, \mathfrak{t}_2, \mathfrak{t}_3 \} = TM\end{aligned}$$

et l'on a encore LD -régularité et les deux vecteurs de croissance sont bien partout $(2, 3, 4, 5)$. C'est la forme normale $J^3(\mathbb{R}, \mathbb{R})$ de Cartan des 3-jets. Remarquons que la régularité vient du fait que les vecteurs de base $\partial_x, \partial_y, \partial_p, \partial_q, \partial_r$ interviennent chacun dans l'un des \mathfrak{t}_j avec un coefficient *constant* (en l'occurrence 1), même s'ils interviennent dans d'autres \mathfrak{t}_j avec des coefficients *variables*.

Le problème est que, dès cette dimension 5, il existe des formes normales de Goursat *exceptionnelles* correspondant à une perte de L -régularité même si la D -régularité reste satisfaite. C'est une subtile découverte d'André Giaro, Antonio Kumpera et Ceferino Ruiz qui ont montré en 1978 (cf. [211] et [308]) que, en plus de ω_3 , il existe une autre (et une seule) forme normale de Goursat $\{\omega_1 = 0, \omega_2 = 0, \omega'_3 = 0\}$ avec

$$\omega'_3 = dx - rdq$$

qui échange x et q dans ω_3 . Cette autre distribution \mathfrak{D}' est engendrée par

$$\{ \mathfrak{t}'_5 = \mathfrak{t}_5 = \partial_r, \mathfrak{t}'_1 = \partial_q + r(q\partial_p + p\partial_y + \partial_x) \}$$

comme on le vérifie immédiatement. Les relations de commutations sont maintenant

$$\begin{aligned}[\mathfrak{t}'_1, \mathfrak{t}'_5] &= -(q\partial_p + p\partial_y + \partial_x) = -\mathfrak{t}'_4, \quad [\mathfrak{t}'_1, \mathfrak{t}'_4] = \partial_p = \mathfrak{t}'_2 = \mathfrak{t}_2, \\ [\mathfrak{t}'_5, \mathfrak{t}'_4] &= 0, \quad [\mathfrak{t}'_1, \mathfrak{t}'_2] = -r\partial_y = -\mathfrak{t}'_3, \quad [\mathfrak{t}'_5, \mathfrak{t}'_2] = 0, \quad [\mathfrak{t}'_1, \mathfrak{t}'_3] = 0, \quad [\mathfrak{t}'_2, \mathfrak{t}'_3] = 0, \quad [\mathfrak{t}'_4, \mathfrak{t}'_3] = 0,\end{aligned}$$

mais

$$[\mathfrak{t}'_4, \mathfrak{t}'_2] = \partial_y = \mathfrak{t}''_3, \quad [\mathfrak{t}'_5, \mathfrak{t}'_3] = \partial_y = \mathfrak{t}''_3$$

et l'on note que le vecteur de base ∂_y intervient d'abord dans \mathfrak{t}'_3 avec le coefficient variable r et seulement ensuite dans $[\mathfrak{t}'_4, \mathfrak{t}'_2]$ et $[\mathfrak{t}'_5, \mathfrak{t}'_3]$ avec le coefficient constant 1. C'est cela la source du comportement *singulier* de \mathfrak{D}' . On obtient le drapeau de Lie

$$\begin{aligned}\mathfrak{D}^1 &= \mathfrak{D}'^{(1)} = \mathfrak{D}' + [\mathfrak{D}', \mathfrak{D}'] = \text{Span} \{ \mathfrak{t}'_1, \mathfrak{t}'_5 \} + \text{Span} \{ \mathfrak{t}'_4 \} = \text{Span} \{ \mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4 \}, \\ \mathfrak{D}^2 &= \mathfrak{D}' + [\mathfrak{D}', \mathfrak{D}^1] = \text{Span} \{ \mathfrak{t}'_1, \mathfrak{t}'_5 \} + \text{Span} \{ \mathfrak{t}'_4, \mathfrak{t}'_2 \} = \text{Span} \{ \mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4, \mathfrak{t}'_2 \}, \\ \mathfrak{D}^3 &= \mathfrak{D}' + [\mathfrak{D}', \mathfrak{D}^2] = \text{Span} \{ \mathfrak{t}'_1, \mathfrak{t}'_5 \} + \text{Span} \{ \mathfrak{t}'_4, \mathfrak{t}'_2, \mathfrak{t}'_3 \} = \text{Span} \{ \mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4, \mathfrak{t}'_2, \mathfrak{t}'_3 \}.\end{aligned}$$

Le problème est que les 5 vecteurs \mathfrak{t}'_j

$$\mathfrak{t}'_1 = \partial_q + r(q\partial_p + p\partial_y + \partial_x), \quad \mathfrak{t}'_2 = \partial_p, \quad \mathfrak{t}'_3 = r\partial_y, \quad \mathfrak{t}'_4 = \partial_x + p\partial_y + q\partial_p, \quad \mathfrak{t}'_5 = \partial_r$$

ne sont linéairement indépendants que si $r \neq 0$ puisque le déterminant de la matrice 5×5 les ayant comme vecteurs lignes est r . Sur l'hyperplan H d'équation $r = 0$,

$\mathfrak{D}_H^3 = \mathfrak{D}_H^2$ puisque $\mathfrak{t}'_3 = 0$. H est un lieu L -singulier et ses points sont singuliers au sens où la forme normale de \mathfrak{D}' en leur voisinage n'est pas celle de Cartan.

Il faut donc pousser jusqu'à \mathfrak{D}'^4 car alors le dernier commutateur

$$[\mathfrak{t}'_5, \mathfrak{t}'_3] = \mathfrak{t}''_3 = \partial_y$$

donne ∂_y avec un coefficient constant (= 1).

$$\begin{aligned} \mathfrak{D}'^4 &= \mathfrak{D}' + [\mathfrak{D}', \mathfrak{D}'^3] = \text{Span} \{\mathfrak{t}'_1, \mathfrak{t}'_5\} + \text{Span} \{\mathfrak{t}'_4, \mathfrak{t}'_2, \mathfrak{t}''_3\} \\ &= \text{Span} \{\mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4, \mathfrak{t}'_2, \mathfrak{t}'_3\} = TM \end{aligned}$$

car

$$\begin{cases} \partial_x = \mathfrak{t}'_4 - p\mathfrak{t}''_3 - q\mathfrak{t}'_2 \\ \partial_y = \mathfrak{t}''_3 \\ \partial_p = \mathfrak{t}'_2 \\ \partial_q = \mathfrak{t}'_1 - r\mathfrak{q}\mathfrak{t}'_2 - r\mathfrak{p}\mathfrak{t}''_3 - r(\mathfrak{t}'_4 - p\mathfrak{t}''_3 - q\mathfrak{t}'_2) = \mathfrak{t}'_1 - r\mathfrak{t}'_4 \\ \partial_r = \mathfrak{t}'_5 \end{cases}$$

Le petit vecteur de croissance est $\{2, 3, 4, 5\}$ en dehors de H mais $\{2, 3, 4, 4, 5\}$ sur H .

En revanche pour le drapeau dérivé on obtient

$$\begin{aligned} \mathfrak{D}'^{(1)} &= \mathfrak{D}'^1 = \mathfrak{D}' + [\mathfrak{D}', \mathfrak{D}'] = \text{Span} \{\mathfrak{t}'_1, \mathfrak{t}'_5\} + \text{Span} \{\mathfrak{t}'_4\} = \text{Span} \{\mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4\} \\ \mathfrak{D}'^{(2)} &= \mathfrak{D}'^{(1)} + [\mathfrak{D}'^{(1)}, \mathfrak{D}'^{(1)}] = \text{Span} \{\mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4\} + \text{Span} \{\mathfrak{t}'_4, \mathfrak{t}'_2\} = \text{Span} \{\mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4, \mathfrak{t}'_2\} \\ \mathfrak{D}'^{(3)} &= \mathfrak{D}'^{(2)} + [\mathfrak{D}'^{(2)}, \mathfrak{D}'^{(2)}] = \text{Span} \{\mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4, \mathfrak{t}'_2\} + \text{Span} \{\mathfrak{t}'_4, \mathfrak{t}'_2, \mathfrak{t}'_3, \mathfrak{t}''_3\} \\ &= \text{Span} \{\mathfrak{t}'_1, \mathfrak{t}'_5, \mathfrak{t}'_4, \mathfrak{t}'_2, \mathfrak{t}'_3\} = TM . \end{aligned}$$

On voit que la différence avec les \mathfrak{D}'^j est que pour obtenir $\mathfrak{t}''_3 = \partial_y$ on n'a pas à attendre $\mathfrak{D}'^{(4)}$ avec $[\mathfrak{t}'_5, \mathfrak{t}'_3] = \mathfrak{t}''_3$ mais qu'on l'obtient dès $\mathfrak{D}'^{(3)}$ avec $[\mathfrak{t}'_4, \mathfrak{t}'_2] = \mathfrak{t}''_3$. \mathfrak{D}' est donc D -régulière bien que L -singulière et est une distribution de Goursat de grand vecteur de croissance $\{2, 3, 4, 5\}$ partout.

En plus, dès la dimension 6 (4-jets), il apparaît des constantes inéliminables et même, à partir de la dimension 10, des "modules", c'est-à-dire des *continua* de formes normales.

3.5. La récurrence K-R

On obtient ainsi une récurrence de formes normales pour les singularités des \mathfrak{G} en dimension $n \geq 5$ (et donc $s = n - 2 \geq 3$). Les deux premières 1-formes sont $\omega_1 = dy - p dx = dy - p_1 dx$ et $\omega_2 = dp - q dx = dp_1 - p_2 dx$. Ensuite, étant donné un sous-ensemble K de $J = \{3, \dots, s\}$ de $j \geq 3$ singuliers (i.e. tels que ω_j ne soit pas de

Cartan), la récurrence est $\omega_j = du_j - v_j dw_j$ avec $u_j = v_{j-1}$, $w_j = w_{j-1}$, $v_j = p_j + c_j$ ($c_j \in \mathbb{R}$ une constante), i.e.

$$\omega_j = dv_{j-1} - (p_j + c_j) dw_{j-1} ,$$

si $j \notin K$ est régulier et $u_j = w_{j-1}$, $w_j = v_{j-1}$, $v_j = p_j$, i.e.

$$\omega_j = dw_{j-1} - p_j dv_{j-1}$$

si $j \in K$ est singulier.

Dans le cas $n = 5$, $s = 3$, $J = \{3\}$ et il n'y a que 2 possibilités.

1. Si 3 est régulier,

$$\omega_3 = dp_2 - p_3 dx = dq - r dx$$

(on peut faire $c_3 = 0$) et \mathfrak{G} est engendrée par

$$\{\partial_r, r\partial_q + q\partial_p + p\partial_y + \partial_x\} .$$

2. Si 3 est singulier,

$$\omega'_3 = dx - p_3 dp_2 = dx - r dq$$

et \mathfrak{G} est engendrée par

$$\{\partial_r, \partial_q + r(q\partial_p + p\partial_y + \partial_x)\} .$$

Dans le cas $n = 6$, $s = 4$, $J = \{3, 4\}$ et il existe 4 sous-ensembles de $J : \{\emptyset\}, \{3\}, \{4\}, \{3, 4\}$. Il y a 5 possibilités.

1. Si 3 et 4 sont réguliers ($K = \{\emptyset\}$),

$$\omega_3 = dp_2 - p_3 dx = dq - r dx$$

et

$$\omega_4 = dp_3 - p_4 dx = dr - s dx$$

(Cartan, on peut faire $c_4 = 0$). \mathfrak{G} est engendrée par

$$\{\partial_s, s\partial_r + r\partial_q + q\partial_p + p\partial_y + \partial_x\} .$$

2. Si 3 est régulier et 4 singulier ($K = \{4\}$),

$$\omega_3 = dp_2 - p_3 dx = dq - r dx$$

et

$$\omega'_4 = dx - p_4 dp_3 = dx - s dr$$

(on peut faire $c_4 = 0$). \mathfrak{G} est engendrée par

$$\{\partial_s, \partial_r + s(r\partial_q + q\partial_p + p\partial_y + \partial_x)\} .$$

3. Si 3 est singulier et 4 régulier ($K = \{3\}$),

$$\omega'_3 = dx - p_3 dp_2 = dx - r dq$$

et

$$\omega_4 = dp_3 - (p_4 + c_4) dp_2 = dr - (s + c) dq .$$

\mathfrak{G} est engendrée par

$$\{\partial_s, (s + c) \partial_r + \partial_q + r (q \partial_p + p \partial_y + \partial_x)\} .$$

Ce cas se divise en 2 sous-cas :

(a) $c = 0$,

(b) $c \neq 0$ et on peut normaliser en prenant $c = 1$.

4. Si 3 et 4 sont singuliers ($K = \{3, 4\}$),

$$\omega'_3 = dx - p_3 dp_2 = dx - r dq$$

et

$$\omega'_4 = dp_2 - p_4 dp_3 = dq - s dr$$

(on peut faire $c_4 = 0$). \mathfrak{G} est engendrée par

$$\{\partial_s, \partial_r + s (\partial_q + r (q \partial_p + p \partial_y + \partial_x))\} .$$

La combinatoire se complexifie rapidement et devient vite très calculatoire. Le lecteur intéressé (et courageux) pourra se référer aux calculs de Piotr Mormul, par exemple dans [109], allant jusqu'à l'ordre 8.

3.6. Géométrie des singularités (Montgomery-Zhitomirskii)

Ces phénomènes de singularités découverts par Giaro, Kumpera et Ruiz ont inauguré une intense période de recherche sur la prolifération des formes normales de Goursat possibles en dimension ≥ 5 . On pourra se référer par exemple, outre aux articles déjà cités et parmi tant d'autres (la littérature est vaste), aux travaux de Mohamad Cheaito et Piotr Mormul ([109], 1999, [379], 2000) qui classifient les modèles locaux jusqu'à la dimension 8, de William Pasillas-Lépine et Witold Respondek ([400], 2001) et, surtout, de Richard Montgomery et Michail Zhitomirskii ([376], 2001, [377], 2010) qui ont approfondi la *géométrie* des singularités. Nous allons nous appuyer sur leur article [376] pour esquisser le "hard problem" de classification dont il s'agit.

Remarque. Avant que de commencer, rappelons encore une fois que le fait que les jets permettent de comprendre à l'ordre s la structure *locale* d'objets différentiels au moyen de valeurs *ponctuelles* (celles des jets jusqu'à l'ordre s) est la raison fondamentale de leur applicabilité à la neurogéométrie. Nous y avons insisté d'emblée, dès la section 4 de notre *Introduction* 1, en parlant de la "difficulté principale" de

la possibilité même d'un "calcul différentiel neuronal". Les micro-modules neuro-naux comme les hypercolonnes d'orientation évaluent des valeurs *ponctuelles*, des champs rétinotopiques de tels micro-modules sont des jets et les connexions neuronales cortico-corticales entre micro-modules implémentent les noyaux de formes différentielles définissant les distributions de contact successives. Le plan \mathbb{R}^2 est celui approximant la rétine et les $\mathbb{P}_{\mathfrak{M}}^s$ sont implémentés jusqu'à $s = 2$ ou 3 corticalement de façon rétinotopique par des architectures fonctionnelles. La différence géométrique nette entre les "basic variables" (x, y) de \mathbb{R}^2 et les "engrafted variables" (au sens de David Hubel³⁷) p, q, r, s, \dots des jets successifs (nos variables eulériennes) est donc anatomiquement réalisée dans l'implémentation neuronale, suite à une très longue évolution darwinienne. La géométrie "immanente" ("from within" au sens de Gromov) que nous utilisons pour modéliser les architectures fonctionnelles de la vision plonge ses racines dans la phylogenèse. Il est donc essentiel pour la modélisation géométrique d'utiliser le fait que des structures *locales* à un certain ordre peuvent devenir *ponctuelles* à un ordre plus élevé si l'on pousse assez loin leur développement en jets en rajoutant des dimensions et en compensant par des distributions de Goursat. □

L'approche géométrique est nécessaire car c'est elle qui explique les formes normales exprimées en termes de coordonnées. Comme nous l'avons indiqué dans la section 1.4, la géométrie explicite des stratifications auxquelles les "bonnes" coordonnées doivent s'adapter. La géométrie est donc *première*. Dans [376] Richard Montgomery et Michail Zhitomirskii utilisent la méthode des stratifications et des propriétés de transversalité pour étudier la classification des structures locales des \mathfrak{G} de Goursat sur une variété M sous l'action des difféomorphismes locaux Diff_x préservant leur structure. Cette classification est plus fine que celle des formes normales de Kumpera-Ruiz. Le nombre de types de singularités augmente très vite, en fait exponentiellement, à partir de la dimension $n = 5$. Nous avons vu que pour $n = 3$ (Heisenberg) et $n = 4$ (Engel) il n'y a qu'une orbite (forme normale de Cartan). Pour $n = 5$, il y en a 2, 5 pour $n = 6$, 13 pour $n = 7$, 34 pour $n = 8$, 93 pour $n = 9$ et une *infinité* pour $n \geq 10$ (résultats de Kumpera, Ruiz, Gaspar, Mormul).

La classification montre que les singularités locales des \mathfrak{G} sont *structurellement stables* relativement à l'espace des \mathfrak{G} de type $(2, n)$ muni de la topologie C^{n-1} et sont déterminées à l'ordre $n - 2$.

Remarque. Il y a deux façons de décrire récursivement une distribution de Goursat de type $(2, n)$. Nous avons adopté jusqu'ici une description "ascendante" partant de \mathfrak{D} et arrivant à TM par commutateurs successifs. D'où les suites ascendantes de distributions

$$\mathfrak{D}^{(0)} = \mathfrak{D}, \dots, \mathfrak{D}^{(k)} = \mathfrak{D}^{(k)} + [\mathfrak{D}^{(k)}, \mathfrak{D}^{(k)}], \dots, \mathfrak{D}^{(s)} = TM,$$

37. Cf. Vol I, chapitre 4, section 4.3.2.

$s = n - 2$, de rangs respectifs $2, \dots, k + 2, \dots, n$. Mais on peut aussi adopter une description “descendante” partant de $\mathfrak{D}_{(0)} = TM$ de dimension n pour arriver à $\mathfrak{D}_{(s)}$, $s = n - 2$, de dimension $n - s = 2$, avec des

$$\mathfrak{D}_{(j)} = \mathfrak{D}_{(j+1)} + [\mathfrak{D}_{(j+1)}, \mathfrak{D}_{(j+1)}]$$

de *co*-rangs j , i.e. de dimension $n - j$. $\mathfrak{D}_{(s)} = \mathfrak{D}^{(0)} = \mathfrak{D}$ est de corang $n - 2$, i.e. de rang 2, et l'on a $\mathfrak{D}_{(j)} = \mathfrak{D}^{(k)}$ avec $k = n - (j + 2)$. La description descendante est particulièrement bien adaptée lorsque l'on décrit les $\mathfrak{D}_{(j)}$ comme les noyaux des 1-formes $\omega_1, \dots, \omega_s$: $\mathfrak{D}_{(1)}$ est définie par $\omega_1 = 0$, $\mathfrak{D}_{(2)}$ par $\omega_1 = 0, \omega_2 = 0$, etc. C'est elle qu'utilisent les auteurs. □

L'outil principal est le *feuilletage caractéristique* d'une distribution \mathfrak{D} . Nous avons rencontré une notion analogue, dès notre introduction de la structure de contact de l'espace des 1-jets $J^1(\mathbb{R}, \mathbb{R})$ au chapitre 3, section 4. Nous avons la 1-forme de contact $\omega = dy - pdx$ sur M (notée alors ω_J sur \mathbb{V}_J), les générateurs $\mathfrak{t}_1 = \frac{\partial}{\partial x} + p\frac{\partial}{\partial y}$ et $\mathfrak{t}_2 = \frac{\partial}{\partial p}$ de la distribution de contact \mathfrak{D} (notée alors \mathcal{K}) et le crochet $[\mathfrak{t}_1, \mathfrak{t}_2] = \mathfrak{t}_3 = -\frac{\partial}{\partial y}$ engendrant le champ caractéristique de Reeb $\chi = \frac{\partial}{\partial y}$ partout transverse à \mathfrak{D} , $\mathfrak{D} \oplus \chi$ étant une décomposition en somme directe de TM . La trace de ce champ sur \mathfrak{D} était $\{0\}$ puisque $\mathfrak{D} \cap \chi = \{0\}$. Autrement dit, la trace du champ caractéristique sur \mathfrak{D} était nulle.

Mais avec la distribution d'Engel ($n = 4, s = 2$) de $J^2(\mathbb{R}, \mathbb{R})$, ce n'est plus le cas et la structure s'enrichit. La distribution $\mathfrak{D}_{(1)} = \ker(\omega_1)$, $\omega_1 = dy - pdx$, est engendrée par

$$\{\mathfrak{t}_1 = \partial_x + p\partial_y + q\partial_p, \mathfrak{t}_2 = \partial_p, \mathfrak{t}_4 = \partial_q\}$$

et la distribution

$$\mathfrak{D} = \mathfrak{D}_{(2)} = \ker(\omega_1) \cap \ker(\omega_2) ,$$

$\omega_1 = dy - pdx, \omega_2 = dp - qdx$, est engendrée par

$$\{\mathfrak{t}_1 = \partial_x + p\partial_y + q\partial_p, \mathfrak{t}_4 = \partial_q\}$$

car $\omega_2(\partial_p) = 1 \neq 0$. Les commutateurs sont

$$[\mathfrak{t}_1, \mathfrak{t}_4] = -\partial_p = -\mathfrak{t}_2, [\mathfrak{t}_1, \mathfrak{t}_2] = \mathfrak{t}_3 = -\partial_y,$$

$$[\mathfrak{t}_1, \mathfrak{t}_3] = [\mathfrak{t}_2, \mathfrak{t}_3] = [\mathfrak{t}_2, \mathfrak{t}_4] = [\mathfrak{t}_3, \mathfrak{t}_4] = 0 .$$

Considérons alors la 2-forme $d\omega_1 = dx \wedge dp$ (qui est la forme symplectique canonique de l'espace des phases (x, p)). Si $\mathfrak{t} = (\xi, \eta, \pi, \kappa)$ et $\mathfrak{t}' = (\xi', \eta', \pi', \kappa')$ sont deux vecteurs tangents au même point x de M , $d\omega_1(\mathfrak{t}, \mathfrak{t}') = -\omega_1([\mathfrak{t}, \mathfrak{t}']) = \xi\pi' - \xi'\pi$ sur $\mathfrak{D}_{(1)}$.³⁸ Les $\mathfrak{t} \in \mathfrak{D}_{(1)}$ sont caractérisés par $\mathfrak{t} = (\xi, p\xi, \pi, \kappa)$, mais la contrainte $\eta = p\xi$ n'intervient pas dans $d\omega_1(\mathfrak{t}, \mathfrak{t}')$ qui ne fait intervenir que (ξ, π) . Le noyau de $d\omega_1 \upharpoonright_{\mathfrak{D}_{(1)}}$ est par définition l'ensemble des $\mathfrak{t} \in \mathfrak{D}_{(1)}$ tels que $d\omega_1(\mathfrak{t}, \mathfrak{t}') = 0$ pour tout

38. Quand il est clair que l'on travaille localement en un point x , nous omettons l'indice x : \mathfrak{t} est un $\mathfrak{t} \in T_x M$, ω signifie ω_x , \mathfrak{D} signifie \mathfrak{D}_x , etc. Les structures en x sont linéaires.

$\mathbf{t}' \in \mathfrak{D}_{(1)}$. Pour cela il faut donc que $\mathbf{t} = (\xi, p\xi, \pi, \kappa)$ soit tel que $\xi\pi' - \xi'\pi = 0$ pour tout $\mathbf{t}' = (\xi', p\xi', \pi', \kappa')$, autrement dit

$$\mathbf{t} = (0, 0, 0, \kappa) .$$

En x , le noyau de $d\omega_1 \upharpoonright_{\mathfrak{D}_{(1)}}$ est l'axe des κ donc un sous-espace de dimension 1 (une droite) dans $\mathfrak{D}_{(1)}$ et comme $\mathfrak{D}_{(1)}$ est de dimension 3, il est de codimension 2 dans $\mathfrak{D}_{(1)}$ et donc un sous-espace linéaire de codimension 3 dans les T_xM (qui sont de dimension 4). Notons $\mathfrak{L}_{(1)}$ ce noyau. Lorsque x varie c'est le feuilletage caractéristique de $\mathfrak{D}_{(1)}$. Il est partout l'axe des κ qui appartient à tous les $\mathfrak{D}_{(1),x}$.

On peut ensuite passer à la 2-forme $d\omega_2 = dx \wedge dq$. Si $\mathbf{t} = (\xi, \eta, \pi, \kappa)$ et $\mathbf{t}' = (\xi', \eta', \pi', \kappa')$, on a

$$[\mathbf{t}, \mathbf{t}'] = (0, \xi'\pi - \xi\pi', \xi'\kappa - \xi\kappa', 0)$$

et

$$d\omega_2(\mathbf{t}, \mathbf{t}') = -\omega_2([\mathbf{t}, \mathbf{t}']) = \xi\kappa' - \xi'\kappa$$

sur $\mathfrak{D}_{(2)}$. On considère alors maintenant sa restriction à $\mathfrak{D}_{(2)}$. Comme les $\mathbf{t} \in \mathfrak{D}_{(2)}$ en x sont caractérisés par $\mathbf{t} = (\xi, p\xi, q\xi, \kappa)$, la contrainte $\eta = p\xi, \pi = q\xi$ n'intervient pas dans $d\omega_2(\mathbf{t}, \mathbf{t}')$ qui ne fait intervenir que (ξ, κ) . Le noyau de $d\omega_2 \upharpoonright_{\mathfrak{D}_{(2)}}$ est constitué des $\mathbf{t} \in \mathfrak{D}_{(2)}$ tels que $d\omega_2(\mathbf{t}, \mathbf{t}') = 0$ pour tout $\mathbf{t}' \in \mathfrak{D}_{(2)}$. Pour cela il faut donc que $\mathbf{t} = (\xi, p\xi, q\xi, \kappa)$ soit tel que $\xi\kappa' - \xi'\kappa = 0$ pour tout $\mathbf{t}' = (\xi', p\xi', q\xi', \kappa')$, autrement dit $\mathbf{t} = (0, 0, 0, 0)$. Le noyau de $d\omega_2 \upharpoonright_{\mathfrak{D}_{(2)}}$ est donc nul dans $\mathfrak{D}_{(2)}$ (et donc de codimension 2 dans $\mathfrak{D}_{(2)}$ qui est de dimension 2). Notons $\mathfrak{L}_{(2)}$ ce noyau = $\{0\}$. C'est le feuilletage caractéristique de $\mathfrak{D}_{(2)}$.

À partir de cet exemple initial d'Engel, Richard Montgomery et Michail Zhitomirskii enclenchent une remarquable *induction dimensionnelle* qui illumine l'origine des singularités et permet de classer les structures locales des distributions de Goursat.

1. Étant donnée une distribution de Goursat de type $(2, n)$.

$$\mathfrak{D}_{(0)} = TM \supset \cdots \supset \mathfrak{D}_{(j)} = \mathfrak{D}_{(j+1)} + [\mathfrak{D}_{(j+1)}, \mathfrak{D}_{(j+1)}] \supset \cdots \supset \mathfrak{D}_{(s)} ,$$

$s = n - 2 \geq 2$, avec des $\mathfrak{D}_{(j)}$ de corangs j définis par l'annulation des 1-formes successives $\{\omega_1, \dots, \omega_s\}$, les auteurs montrent d'abord que les restrictions des 2-formes $d\omega_j$ à $\mathfrak{D}_{(j)}$ sont non dégénérées. Par définition ω_j s'annule sur $\mathfrak{D}_{(j)}$ qui est son noyau mais n'est pas identiquement nulle sur $\mathfrak{D}_{(j-1)}$. Son annulation définit même $\mathfrak{D}_{(j)}$ dans $\mathfrak{D}_{(j-1)}$. Sur $\mathfrak{D}_{(j)}$,

$$d\omega_j(\mathbf{t}, \mathbf{t}') = -\omega_j([\mathbf{t}, \mathbf{t}']) .$$

Mais comme par construction $[\mathfrak{D}_{(j+1)}, \mathfrak{D}_{(j+1)}] \subset \mathfrak{D}_{(j)}$, $\omega_j([\mathbf{t}, \mathbf{t}']) = 0$ (on doit ici supposer $j > s$ afin que $\mathfrak{D}_{(j+1)}$ ait un sens) et donc

$$d\omega_j(\mathbf{t}, \mathbf{t}') = 0$$

pour $\mathbf{t}, \mathbf{t}' \in \mathfrak{D}_{(j+1)}$. Dans l'autre sens, $[\mathfrak{D}_{(j)}, \mathfrak{D}_{(j)}] \subset \mathfrak{D}_{(j-1)}$ et donc $d\omega_j$ n'est pas identiquement nulle sur $\mathfrak{D}_{(j)}$. Bref, $d\omega_j \upharpoonright_{\mathfrak{D}_{(j)}}$ est une 2-forme non nulle sur $\mathfrak{D}_{(j)}$ possédant un sous-espace *isotrope* de codimension 1. Cela implique que $d\omega_j \upharpoonright_{\mathfrak{D}_{(j)}}$ soit de rang 2 (i.e. non dégénérée). On montre ensuite qu'il en va de même pour $j = s$.

2. Les auteurs définissent ensuite les feuilletages caractéristiques $\mathfrak{L}_{(j)}$ comme les *noyaux* dans $\mathfrak{D}_{(j)}$ de ces $d\omega_j \upharpoonright_{\mathfrak{D}_{(j)}}$. En utilisant le fait que

$$d\omega_j(\mathbf{t}, \mathbf{t}') = -\omega_j([\mathbf{t}, \mathbf{t}'])$$

sur $\mathfrak{D}_{(j)}$, ils montrent que les $\mathfrak{L}_{(j)}$ sont les espaces des $\mathbf{t} \in \mathfrak{D}_{(j)}$ tels que pour tout $\mathbf{t}' \in \mathfrak{D}_{(j)}$ on ait $[\mathbf{t}, \mathbf{t}'] \in \mathfrak{D}_{(j)}$. Ce sont des sous-algèbres de Lie d'après Jacobi. En effet, si $\mathbf{t}, \mathbf{u} \in \mathfrak{L}_{(j)}$, alors $[\mathbf{t}, \mathbf{u}] \in \mathfrak{D}_{(j)}$ et pour tout $\mathbf{t}' \in \mathfrak{D}_{(j)}$,

$$[[\mathbf{t}, \mathbf{u}], \mathbf{t}'] = -[[\mathbf{u}, \mathbf{t}'], \mathbf{t}] - [[\mathbf{t}', \mathbf{t}], \mathbf{u}]$$

par Jacobi. Or $[\mathbf{u}, \mathbf{t}'] \in \mathfrak{D}_{(j)}$ car $\mathbf{u} \in \mathfrak{L}_{(j)}$ et $\mathbf{t}' \in \mathfrak{D}_{(j)}$ et donc

$$-[[\mathbf{u}, \mathbf{t}'], \mathbf{t}] = [\mathbf{t}, [\mathbf{u}, \mathbf{t}']] \in \mathfrak{D}_{(j)}$$

car $\mathbf{u} \in \mathfrak{L}_{(j)}$. Il en va de même pour le second terme et donc

$$[[\mathbf{t}, \mathbf{u}], \mathbf{t}'] \in \mathfrak{D}_{(j)} .$$

Cela signifie que la distribution $\mathfrak{L}_{(j)}$ est *involutive*.

3. Les $\mathfrak{L}_{(j)}$ sont de codimension 2 dans $\mathfrak{D}_{(j)}$ et forment un *drapeau* démarrant avec $\mathfrak{L}_{(1)}$ de codimension 3 dans TM (puisque $\mathfrak{D}_{(1)}$ est de codimension 1) et décroissant de une dimension à chaque pas jusqu'à $\mathfrak{L}_{(s)}$ ($s = n - 2$) qui se réduit à $\mathfrak{L}_{(s)} = \{0\}$. Le positionnement des $\mathfrak{L}_{(j)}$ par rapport aux $\mathfrak{D}_{(j)}$ est donnée par le "sandwich lemma" disant que, si $j < s$, $\mathfrak{L}_{(j)}$ est en fait une sous-distribution de codimension 1 de $\mathfrak{D}_{(j+1)}$ et comme $\mathfrak{D}_{(j+1)}$ est elle-même de codimension 1 dans $\mathfrak{D}_{(j)}$, $\mathfrak{L}_{(j)}$ est bien de codimension 2 dans $\mathfrak{D}_{(j)}$. Ce sont les relations d'*incidence* entre les $\mathfrak{L}_{(j)}$, $\mathfrak{D}_{(j+1)}$ et $\mathfrak{D}_{(j)}$ qui permettent de comprendre l'origine des singularités.
4. Les singularités peuvent apparaître à partir de $n \geq 5$, c'est-à-dire $s \geq 3$. Comme $\mathfrak{L}_{(j)}$ est de codimension 2 dans $\mathfrak{D}_{(j)}$ et que $\mathfrak{D}_{(j+2)}$ est également de codimension 2 dans $\mathfrak{D}_{(j)}$ puisque \mathfrak{D} est de Goursat, on peut *comparer* les sous-espaces linéaires $\mathfrak{L}_{(j),x}$ et $\mathfrak{D}_{(j+2),x}$ dans $\mathfrak{D}_{(j),x}$. La propriété pour \mathfrak{D} d'être *régulière* (réductible à la forme normale de Cartan) est alors *géométriquement* caractérisable par $\mathfrak{L}_{(j)} \neq \mathfrak{D}_{(j+2)}$ pour $j = 1, \dots, s - 2 = n - 4$ en tout point x . Si $\mathfrak{L}_{(j),x} \neq \mathfrak{D}_{(j+2),x}$ on peut dire que \mathfrak{D} est *j-régulière* en x .
5. \mathfrak{D} sera par conséquent *singulière* en x si $\mathfrak{L}_{(j),x} = \mathfrak{D}_{(j+2),x}$ pour au moins un $j \in J = \{1, \dots, s - 2\}$. Les formes normales de Kumpera-Ruiz correspondent

alors au choix d'un sous-ensemble K de J et de \mathfrak{D} de Goursat qui sont j -régulières pour $j \notin K$ et k -singulières pour $k \in K$. Cela fait 2^{s-2} types possibles et on montre qu'on les obtient *tous*.

6. Pour effectuer la classification des germes des \mathfrak{D} au moyen des orbites de l'action de Diff_x (classification plus fine que celle de Kumpera-Ruiz), Montgomery et Zhitomirskii la ramène de façon remarquable à un problème basique de classification *projective* en explicitant la façon dont on descend d'un cran dans la chaîne descendante des $\mathfrak{D}_{(j)}$ de corangs j .

- (a) Si l'on a une \mathfrak{G} de corang $s \leq n - 3$, on peut descendre d'un cran pour obtenir une \mathfrak{G} de corang $s + 1$. $\mathfrak{D}_{(s+1)}$ sera alors un \mathfrak{H} tel que

$$\mathfrak{H} + [\mathfrak{H}, \mathfrak{H}] = \mathfrak{D}_{(s)} .$$

La condition de régularité / singularité est $(\mathfrak{H} \neq \mathfrak{L}_{(s-1)}) / (\mathfrak{H} = \mathfrak{L}_{(s-1)})$.

- (b) On montre un lemme fondamental disant que si \mathfrak{H} et \mathfrak{H}' sont des germes de \mathfrak{G} de corang $s + 1$ tels que

$$\mathfrak{H} + [\mathfrak{H}, \mathfrak{H}] = \mathfrak{H}' + [\mathfrak{H}', \mathfrak{H}'] = \mathfrak{D}_{(s)}$$

et $\mathfrak{H}_x = \mathfrak{H}'_x$, alors \mathfrak{H} et \mathfrak{H}' sont équivalents à travers un difféomorphisme local. On est donc amené à classer les espaces \mathfrak{H}_x intermédiaires entre $\mathfrak{L}_{(s-2),x}$ et $\mathfrak{D}_{(s),x}$ tels que $\mathfrak{H} + [\mathfrak{H}, \mathfrak{H}] = \mathfrak{D}_{(s)}$.³⁹

- (c) Un autre lemme fondamental permet de montrer que, si le rang de $\mathfrak{D}_{(s)} \geq 3$, on peut en fait obtenir ainsi *tous* les espaces intermédiaires. On se ramène ainsi à la droite projective $\mathbb{P}^1_{(s),x}$ (topologiquement un cercle) des droites du plan $\mathfrak{D}_{(s),x} / \mathfrak{L}_{(s-2),x}$.
 - (d) L'action des difféomorphismes va ainsi se traduire par une action linéaire d'un certain groupe $\Gamma_{(s),x}$ sur $\mathbb{P}^1_{(s),x}$. Cette action est induite par les isomorphismes linéaires tangents aux difféomorphismes locaux en x qui préservent la structure $\{\mathfrak{D}_{(s)}, \mathfrak{D}_{(s+1)}, \mathfrak{L}_{(s-2)}, \mathfrak{L}_{(s-1)}\}$.
7. Or les actions d'un sous-groupe Γ du groupe des transformations projectives de la droite projective \mathbb{P}^1 projectivant le plan \mathbb{R}^2 de coordonnées (a, b) sont bien connues et dépendent essentiellement du nombre $\#Fix(\Gamma)$ de *points fixes* de Γ .

- (a) Si $\#Fix(\Gamma) = 1$, Γ agit transitivement en dehors du point fixe α et il y a donc 2 orbites $\{\alpha\}$ et $\mathbb{P}^1 - \{\alpha\}$.

39. Intermédiaires au sens où $\dim(\mathfrak{H}_x) = \dim(\mathfrak{L}_{(s-2),x}) + 1 = \dim(\mathfrak{D}_{(s),x}) - 1$.

- (b) Si $\#Fix(\Gamma) = 2$, et si α et β sont les points fixes, l'action de Γ dépend de la présence ou de l'absence de la symétrie σ de \mathbb{P}^1 projectivisant la réflexion dans \mathbb{R}^2 par rapport à l'une des deux droites α ou β parallèlement à l'autre. Plus précisément, on prend les deux droites α et β comme axes de coordonnées (a, b) et σ correspond à la symétrie $\sigma([a, b]) = [a, -b] = [-a, b]$ (où $[a, b]$ dénote les coordonnées homogènes d'un point de \mathbb{P}^1). Évidemment σ fixe α et β et $\sigma^2 = Id$. Il y a alors 3 cas.
- (i) Si Γ contient Id , σ et au moins un autre élément, alors Γ agit transitivement en dehors des deux points fixes α et β et il y par conséquent trois orbites $\{\alpha\}$, $\{\beta\}$ et $\mathbb{P}^1 - \{\alpha, \beta\}$.
 - (ii) Si $\sigma \notin \Gamma$, alors Γ opère transitivement sur chaque composante connexe de $\mathbb{P}^1 - \{\alpha, \beta\}$ sans les mélanger. Il y a donc 4 orbites $\{\alpha\}$, $\{\beta\}$ et les deux composantes connexes de $\mathbb{P}^1 - \{\alpha, \beta\}$ (rappelons que \mathbb{P}^1 est topologiquement un cercle).
 - (iii) Si $\Gamma = \{Id, \sigma\}$ alors l'espace des orbites est \mathbb{P}^1/σ qui est topologiquement un intervalle et il y a donc une infinité d'orbites.
- (c) Enfin si $\#Fix(\Gamma) \geq 3$, Γ se réduit à l'identité Id car une transformation linéaire du plan laissant trois droites invariantes est l'identité.

8. En appliquant cette classification projective on en arrive à la classification, à difféomorphisme local près, des distributions de Goursat \mathfrak{G} . Il y a toujours un point fixe de $\Gamma_{(s),x}$, nommé $\mathfrak{L}_{(s-1),x}$. Dans le cas de régularité c'est l'unique point fixe.
9. Les auteurs montrent aussi des résultats globaux, par exemple que les feuilletages caractéristiques déterminent les \mathfrak{G} relativement à des petites déformations près : si \mathfrak{G} et \mathfrak{G}' sont des distributions de Goursat sur M assez voisines pour la topologie de Whitney C^{n-1} et si les feuilletages caractéristiques de codimension 3 $L(\mathfrak{G})$ et $L(\mathfrak{G}')$ sont égaux, alors il existe un difféomorphisme global de M échangeant \mathfrak{G} et \mathfrak{G}' .

3.7. Le “Monstre” classifiant

3.7.1. La tour infinie des complétions projectives $\mathbb{P}_{\mathfrak{M}}^s$ des $J^s(\mathbb{R}, \mathbb{R})$.

Dans un texte plus tardif [377], Richard Montgomery et Michail Zhitomirskii montrent que si l'on part de \mathbb{R}^2 et si l'on effectue $s = (n - 2)$ -fois son prolongement de Cartan *projectif*, on obtient une variété $\mathbb{P}_{\mathfrak{M}}^s$ de dimension n munie d'une distribution de Goursat Δ^s *universelle* au sens où toute \mathfrak{G} sur une M de dimension n est *localement* isomorphe, au sens des isomorphismes locaux entre (M, \mathfrak{G}) préservant

leur \mathfrak{G} -structure, à Δ^s en un point de Δ^s .⁴⁰ Autrement dit, *toutes* les formes normales locales possibles des (M, \mathfrak{G}) de dimension $n = s + 2$ sont représentées dans Δ^s . Cela vient du fait que, à chaque étape, $(\mathbb{P}_{\mathfrak{M}}^s, \Delta^s)$ *complète projectivement* la partie *affine* correspondant aux prolongements que sont les espaces de jets $J^s(\mathbb{R}, \mathbb{R})$. Les auteurs appellent la tour infinie des $\mathbb{P}_{\mathfrak{M}}^s$ le “Monster manifold”.

Tout au début de notre parcours, dès la section 2.1 du chapitre 3, nous avons souligné la différence entre l’espace $J^1(\mathbb{R}, \mathbb{R})$ des 1-jets des $f : \mathbb{R} \rightarrow \mathbb{R}$ dérivables et l’espace $C\mathbb{R}^2$ des éléments de contact (a, p) du plan. Si l’on se donne une $f(x)$ et si l’on considère son graphe $\gamma = \{x, f(x)\}$, le 1-jet de f ,

$$j^1 f(x) = \{x, y = f(x), p = f'(x)\}$$

est une application de \mathbb{R} dans $C\mathbb{R}^2$ qui est la relevée legendrienne Γ de γ . Mais $J^1(\mathbb{R}, \mathbb{R})$ est un *ouvert affine* de $C\mathbb{R}^2$ qui exclut les tangentes *verticales* correspondant à $p = \infty$. Pour les intégrer et construire $\pi : C\mathbb{R}^2 \rightarrow \mathbb{R}^2$ de fibre \mathbb{P}^1 comme compactification projective des fibres \mathbb{R} de $J^1(\mathbb{R}, \mathbb{R}) \rightarrow \mathbb{R}^2$, il faut changer les coordonnées locales.

Nous avons également insisté dès cette section sur le fait que le choix des coordonnées était *arbitraire* et n’avait pas de sens géométrique intrinsèque. En revanche $C\mathbb{R}^2$ avec sa structure de contact possède lui une structure géométrique *intrinsèque*. Cela est lié au fait que les contactomorphismes de $C\mathbb{R}^2$ (les difféomorphisme transformant la structure de contact en structure de contact) sont *plus riches* que les relèvements de difféomorphismes du plan qui sont, eux, des contactomorphismes de $C\mathbb{R}^2$ *préservant la fibration* $\pi : C\mathbb{R}^2 \rightarrow \mathbb{R}^2$. Voyons-le sur un exemple.

Si $\Gamma(t)$ non constant est un germe en $t = 0$ d’une courbe intégrale de la distribution de contact de $C\mathbb{R}^2$ alors Γ est partout contactomorphe à la relevée legendrienne d’une courbe d’un plan. En effet, soit $\gamma(t)$ la projection $\pi(\Gamma(t))$ de $\Gamma(t)$ sur $\mathbb{R}_{(x,y)}^2$. Si $\gamma(t)$ n’est pas constante, $\Gamma(t)$ est la relevée legendrienne de $\gamma(t)$. Si en revanche $\gamma(t) = \gamma(0) (= 0)$ est constante, i.e. $\Gamma(t) = \{0, 0, p(t)\}$, $p(t) \in \mathbb{P}_0^1$ la fibre de $C\mathbb{R}^2$ au-dessus de 0, on peut utiliser le contactomorphisme

$$(x, y, p) \rightarrow (x^* = p, y^* = -x, p^* = y - xp) ,$$

i.e. $(x = -y^*, y = p^* - y^*x^*, p = x^*)$. Il s’agit d’un difféomorphisme d’application linéaire tangente $\begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ -p & 1 & -x \end{pmatrix}$ de déterminant -1 qui envoie la 1-forme $\omega = dy - p dx$ définissant la structure de contact sur la 1-forme de contact $\omega^* = dp^* - y^* dx^*$:

$$dy - p dx = dp^* - y^* dx^* - x^* dy^* - x^* (-dy^*) = dp^* - y^* dx^* .$$

40. Dans la suite, nous abrègerons souvent par \mathfrak{G} l’expression “distribution de Goursat \mathfrak{G} ”.

Dans les nouvelles coordonnées, $\Gamma(t) = \{0, 0, p(t)\}$ devient la courbe

$$\Gamma^*(t) = \{x^*(t) = p(t), y^*(t) = -x(t) = 0, p^*(t) = y(t) - x(t)p(t) = 0\}$$

de projection $\gamma^*(t) = \{x^*(t) = p(t), p^*(t) = 0\}$. La relevée de $\gamma^*(t)$ est

$$\widetilde{\Gamma}^*(t) = \left\{ x^*(t) = p(t), p^*(t) = 0, y^*(t) = \frac{dp^*}{dx^*}(t) = \frac{d(y - xp)}{dp}(t) = -x(t) = 0 \right\}$$

c'est-à-dire $\Gamma^*(t)$ et donc $\Gamma^*(t)$ est bien une relevée legendrienne.

On peut ainsi effectuer à chaque étape cette compactification des fibres de la tour infinie des

$$\dots \rightarrow J^s(\mathbb{R}, \mathbb{R}) \rightarrow J^{s-1}(\mathbb{R}, \mathbb{R}) \rightarrow \dots \rightarrow J^1(\mathbb{R}, \mathbb{R}) \rightarrow \mathbb{R}^2$$

et construire la tour infinie \mathfrak{M} (\mathfrak{M} pour "Monster")

$$\dots \rightarrow \mathbb{P}_{\mathfrak{M}}^s \rightarrow \mathbb{P}_{\mathfrak{M}}^{s-1} \rightarrow \dots \rightarrow \mathbb{P}_{\mathfrak{M}}^1 \rightarrow \mathbb{P}_{\mathfrak{M}}^0 = \mathbb{R}^2$$

où, à chaque étape les projections sont des *fibrés en cercles* correspondant à la compactification projective des variables eulériennes successives p, q, r, s, \dots des espaces de jets successifs. Sur chacun des $\mathbb{P}_{\mathfrak{M}}^s$ il y a une distribution $\Delta^s \subset T\mathbb{P}_{\mathfrak{M}}^s$ de rang 2 définie par une condition de tangence. On passe de l'étage $s - 1$ de $\mathbb{P}_{\mathfrak{M}}^{s-1}$ à l'étage s de $\mathbb{P}_{\mathfrak{M}}^s$ en prenant comme points a_s de $\mathbb{P}_{\mathfrak{M}}^s$ les couples (a_{s-1}, ℓ_s) d'un point a_{s-1} de $\mathbb{P}_{\mathfrak{M}}^{s-1}$ et d'une droite ℓ_s de Δ^{s-1} en a_{s-1} . On part de $\Delta^0 = T\mathbb{P}_{\mathfrak{M}}^0 = T\mathbb{R}^2$ et on considère les éléments de contact (a, ℓ) , $a \in \mathbb{R}^2$, ℓ une droite de $T_a\mathbb{R}^2$ (de pente p éventuellement infinie). Cela donne le fibré $\mathbb{P}_{\mathfrak{M}}^1 \rightarrow \mathbb{P}_{\mathfrak{M}}^0 = \mathbb{R}^2$ de fibre $\mathbb{P}_a^1 = \mathbb{P}(T_a\mathbb{R}^2)$. C'est le fibré de contact dont les fibres sont les compactifications projectives de celles de $J^1(\mathbb{R}, \mathbb{R})$ (Heisenberg). Si $\gamma(t) = a_0(t)$ est une courbe dans \mathbb{R}^2 , elle se relève dans $\mathbb{P}_{\mathfrak{M}}^1$ en $\Gamma(t) = ((a_0(t), \ell_1(t)))$ avec $a'_0(t) \in \ell_1(t)$ pour tout t .⁴¹ On peut avoir $a'_0(t) = 0$. Comme nous l'avons vu et revu, $\Gamma(t)$ est une courbe intégrale (relevée legendrienne) de la distribution de contact de $\mathbb{P}_{\mathfrak{M}}^1$ que l'on note $\Delta^1 \subset T\mathbb{P}_{\mathfrak{M}}^1$. On itère cette construction à tous les étages $\pi_s : \mathbb{P}_{\mathfrak{M}}^s \rightarrow \mathbb{P}_{\mathfrak{M}}^{s-1}$ de la tour : une courbe $\Gamma_s(t) = ((a_{s-1}(t), \ell_s(t)))$ dans $\mathbb{P}_{\mathfrak{M}}^s$ est tangente à $\Delta^s \subset T\mathbb{P}_{\mathfrak{M}}^s$ (i.e. une courbe intégrale de Δ^s)

- (i) si $a_{s-1}(t)$ est tangente à $\Delta^{s-1} \subset T\mathbb{P}_{\mathfrak{M}}^{s-1}$ (i.e. une courbe intégrale de Δ^{s-1}), et
- (ii) si $a'_{s-1}(t) \in \ell_s(t)$ pour tout t , $a'_{s-1}(t) = 0$ étant autorisé.⁴²

Si l'on considère l'application linéaire tangente $T\pi_s : T\mathbb{P}_{\mathfrak{M}}^s \rightarrow T\mathbb{P}_{\mathfrak{M}}^{s-1}$ on a

$$T_{(a_{s-1}, \ell_s)}\pi_s : \Delta_{(a_{s-1}, \ell_s)}^s \subset T_{(a_{s-1}, \ell_s)}\mathbb{P}_{\mathfrak{M}}^s \rightarrow T_{a_{s-1}}\mathbb{P}_{\mathfrak{M}}^{s-1}, (a_{s-1}, \ell_s) \mapsto \ell_s .$$

41. Jusqu'ici, nous avons considéré des courbes C^r ou C^∞ . Pour des raisons techniques les auteurs considèrent des germes de courbes *analytiques* (on dit aussi C^ω) qui ont l'avantage d'être égales à leurs développements de Taylor. Cela ne change pas la géométrie ponctuelle puisque tous les jets en un point sont réalisables par des fonctions analytiques.

42. Pour $\mathbb{P}_{\mathfrak{M}}^1$ la condition (i) est vide car $a_0(t)$ est évidemment toujours tangente à $\Delta^0 = T\mathbb{R}^2$.

Les auteurs montrent que toutes ces distributions Δ^s , $s \geq 1$, sont de Goursat et que les $(\mathbb{P}_{\mathfrak{M}}^s, \Delta^s)$ sont *génériquement* de Cartan au sens où les $x \in \mathbb{P}_{\mathfrak{M}}^s$ tels que $(\mathbb{P}_{\mathfrak{M}}^s, \Delta^s)_x$ soit de Cartan forment un *ouvert dense* (des ouverts affines à chaque étage de la tour).

Remarque. Au niveau 1, $\mathbb{P}_{\mathfrak{M}}^1$ est un fibré *trivial* $\mathbb{P}_{\mathfrak{M}}^0 \times \mathbb{P}^1 \rightarrow \mathbb{P}_{\mathfrak{M}}^0 = \mathbb{R}^2$. À partir du niveau 2, les $\mathbb{P}_{\mathfrak{M}}^{s+1} \rightarrow \mathbb{P}_{\mathfrak{M}}^s$ ne sont plus triviaux à cause des fibres projectives. \square

Remarque. Insistons encore sur le fait que la géométrie intrinsèque de la distribution de contact Δ^1 ne tient pas compte de la façon dont les coordonnées (x, y) sont arbitrairement choisies dans \mathbb{R}^2 . \square

On peut alors considérer les “symétries locales”, à savoir les contactomorphismes locaux préservant ces stratifications de structures de contact itérées et, si l’on a un tel φ au niveau $s - 1$, le prolonger au niveau s par

$$\Phi((a_{s-1}, \ell_s)) = (\varphi(a_{s-1}), T_{a_{s-1}}\varphi(\ell_s)) .$$

Les auteurs démontrent alors que, à partir de $\mathbb{P}_{\mathfrak{M}}^1$, l’on obtient ainsi par prolongement *toutes* les symétries locales Φ de niveau s , ce qui, nous l’avons vu, n’est pas le cas du niveau $\mathbb{P}_{\mathfrak{M}}^0$ puisqu’il existe des contactomorphismes de $\mathbb{P}_{\mathfrak{M}}^1$ qui ne sont pas des relèvements de difféomorphismes de $\mathbb{P}_{\mathfrak{M}}^0 = \mathbb{R}^2$. Pour cela ils utilisent la théorie des champs caractéristiques esquissée plus haut. On peut ensuite définir une relation d’équivalence locale entre courbes $\Gamma_s(t)$ de $\mathbb{P}_{\mathfrak{M}}^s$ en disant que $\Gamma_s(t)$ avec $\Gamma_s(0) = a_s$ et $\Lambda_s(t)$ avec $\Lambda_s(0) = b_s$ sont localement équivalentes s’il existe une symétrie locale Φ avec $\Phi(a_s) = b_s$ et un difféomorphisme local δ de $(\mathbb{R}, 0)$ avec $\delta(0) = 0$ (i.e. une reparamétrisation des courbes) tels que

$$\Lambda_s = \Phi \circ \Gamma \circ \delta$$

(on a bien $\Lambda_s(0) = \Phi(\Gamma_s(0)) = \Phi(a_s) = b_s$). Le résultat fondamental est le théorème suivant :

Théorème. Tout germe de distribution de Goursat de rang 2 sur une variété de dimension $n = s + 2$ apparaît à équivalence près dans $(\mathbb{P}_{\mathfrak{M}}^s, \Delta^s)$. La tour infinie $(\mathbb{P}_{\mathfrak{M}}^s, \Delta^s)$ du “Monster” \mathfrak{M} est donc un espace classifiant universel pour les structures locales des \mathfrak{G} . \diamond

Ce théorème de classification est en fait celui des jets *finis* d’ordre s de courbes legendriennes lorsque l’on considère les *points* de $\mathbb{P}_{\mathfrak{M}}^s$, et celui des germes de courbes legendriennes lorsque l’on considère des germes de courbes intégrales dans $\mathbb{P}_{\mathfrak{M}}^s$. On montre plusieurs choses :

1. La projection de toute courbe intégrale de $(\mathbb{P}_{\mathfrak{M}}^s, \Delta^s)$ sur $\mathbb{P}_{\mathfrak{M}}^1$ est legendrienne, i.e. une courbe intégrale de la structure de contact Δ^1 .
2. Deux courbes intégrales *immergées*, c’est-à-dire de vecteur tangent partout non nul, de $(\mathbb{P}_{\mathfrak{M}}^s, \Delta^s)$ sont équivalentes (échangeables par un difféomorphisme

respectant la structure de Δ^s) si et seulement si leurs projections sur $\mathbb{P}_{\mathfrak{M}}^1$ sont contactomorphes.

3. Les opérations de prolongement et de projection sont inverses l’une de l’autre : tout germe de $\gamma(t) = (x(t), y(t))$ dans \mathbb{R}^2 non constante (car sinon, on ne peut pas la relever puisque $x'(t)$ et $y'(t)$ sont $\equiv 0$) et analytique se prolonge à tous les étages en une courbe $\gamma^s(t)$ qui se projette sur $\gamma(t)$.
4. Tout point a_s de $\mathbb{P}_{\mathfrak{M}}^s$ est représentable par le $\gamma^s(0)$ d’un relèvement d’ordre s d’une courbe $\gamma(t)$ du plan \mathbb{R}^2 (pouvant être très singulière en 0). On appelle alors réguliers les points a_s de $\mathbb{P}_{\mathfrak{M}}^s$ qui sont représentables par le $\gamma^s(0)$ d’une $\gamma(t)$ dont la relevée de niveau 1, $\gamma^1(t)$, est régulière.

3.7.2. Le relèvement de singularités.

Pour mieux comprendre géométriquement l’origine des singularités dans les \mathfrak{G} , il faut comprendre ce qui se passe lorsque que l’on relève une courbe $\gamma(t)$ dans une variété de dimension k ayant, par exemple en 0, une singularité, c’est-à-dire un vecteur tangent nul $\gamma'(0) = 0$. L’exemple le plus simple est celui d’une singularité cusp $\gamma(t)$ définie dans un plan $\mathbb{R}_{(u,v)}^2$ par les équations paramétriques

$$\{u(t) = t^2, v(t) = t^3\}$$

(l’équation algébrique est donc $u^3 - v^2 = 0$). Quand on la relève dans $\mathbb{R}_{(u,v,w)}^3$, on obtient la courbe gauche $\Gamma(t)$ d’équations

$$\left\{ u(t) = t^2, v(t) = t^3, w(t) = \frac{v'(t)}{u'(t)} = \frac{3}{2}t \right\} .$$

Si $t \neq 0$ il n’y a aucun problème car $\Gamma(t)$ est transverse à la fibration $\pi : \mathbb{R}_{(u,v,w)}^2 \rightarrow \mathbb{R}_{(u,v)}^2$. En effet,

$$\Gamma'(t) = \left\{ u'(t) = 2t, v'(t) = 3t^2, w'(t) = \frac{3}{2} \right\}$$

et $\Gamma'(t)$ n’est pas tangente à la fibre au-dessus de $\{u(t), v(t)\}$ (i.e. “verticale”) puisque $u'(t), v'(t) \neq 0$. En revanche au point singulier $t = 0$, $\Gamma'(0)$ devient “verticale”, i.e. tangente à la fibre. Cette non transversalité (cette tangence) de la relevée par rapport aux fibres de la projection est l’expression géométrique de la singularité. Cela se voit clairement sur les figures 4 et 5 Il s’agit là d’un principe fondamental que l’on pourrait appeler “le principe de Thom” : les singularités sont la conséquence de défauts de transversalité dans les relations d’incidence de strates d’ensembles stratifiés décrivant les structures géométriques considérées.

Dans $(\mathbb{P}_{\mathfrak{M}}^1, \Delta^1)$ les courbes legendriennes Γ (courbes intégrales de Δ^1) qui relèvent des courbes régulières (sans singularités) γ de $\mathbb{P}_{\mathfrak{M}}^0 = \mathbb{R}^2$ sont transverses aux fibres

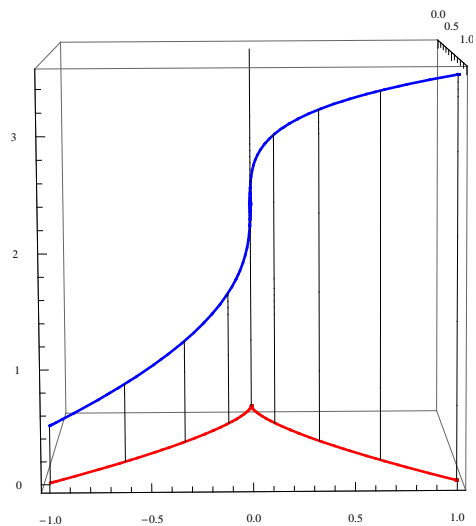


FIGURE 4. Le relèvement de la singularité cusp. La courbe plane rouge possède une singularité cusp à l'origine. Sa relevée bleue est transverse aux fibres de la projection (9 fibres sont représentées) sauf à l'origine où elle est tangente à la fibre.

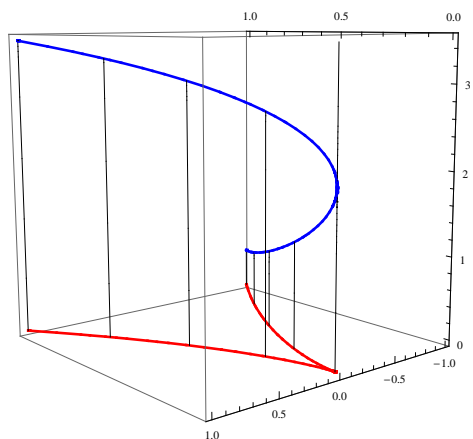


FIGURE 5. Un autre point de vue sur le relèvement de la singularité cusp. Il montre bien la façon dont la relevée bleue est tangente à l'origine à la fibre de la projection.

de

$$\pi_1 : \mathbb{P}_{\mathfrak{M}}^1 \rightarrow \mathbb{P}_{\mathfrak{M}}^0 = \mathbb{R}^2 .$$

Si γ a une singularité, par exemple en 0, il apparaît un défaut de transversalité. Mais, comme nous l'avons vu pour le cusp

$$\gamma(t) = \{x(t) = t^2, y(t) = t^3\},$$

la relevée

$$\Gamma(t) = \gamma^1(t) = \left\{ x(t) = t^2, y(t) = t^3, p(t) = \frac{3}{2}t \right\}$$

peut être quand même *régulière* puisque

$$\Gamma'(t) = \left\{ x'(t) = 2t, y'(t) = 3t^2, p'(t) = \frac{3}{2} \right\}$$

n'est jamais nulle. D'ailleurs, la projection de $\Gamma(t) = \gamma^1(t)$ sur le plan (x, p) est $(x(t) = t^2, p(t) = \frac{3}{2}t)$ c'est-à-dire la parabole tout à fait régulière $x = \frac{4}{9}p^2$. Dans le cas du cusp, on voit bien qu'il n'y a pas d'obstruction à utiliser les formes normales de Cartan $\omega_1 = dy - p dx$:

$$p_1(t) = p(t) = \frac{y'(t)}{x'(t)} = \frac{3}{2}t,$$

$$p_2(t) = q(t) = \frac{p'(t)}{x'(t)} = \frac{3}{4},$$

$$p_3(t) = r(t) = 0, p_s(t) = 0 \text{ pour } s \geq 3.$$

Pour

$$\gamma(t) = \{x(t) = t^2, y(t) = t^4\}$$

tout est régulier car en reparamétrisant par $u = t^2$, $\gamma(t)$ est simplement la parabole $y = x^2$. Il n'y a pas non plus d'obstruction à utiliser les formes normales de Cartan $\omega_1 = dy - p dx$ et $\omega_2 = dp - q dx$:

$$p(t) = \frac{y'(t)}{x'(t)} = 2t^2,$$

$$q(t) = \frac{p'(t)}{x'(t)} = 2, r(t) = 0,$$

$$s(t) = 0 \text{ et tous les suivants .}$$

En revanche, pour

$$\gamma(t) = \{x(t) = t^2, y(t) = t^5\}$$

apparaît la première singularité inéliminable de

$$\gamma^1(t) = \left\{ x(t) = t^2, y(t) = t^5, p(t) = \frac{5}{2}t^3 \right\}$$

car la projection sur le plan (x, p) comporte elle aussi une singularité (un cusp) ainsi que la projection sur le plan (y, p) (un point d'inflexion). On rencontre une

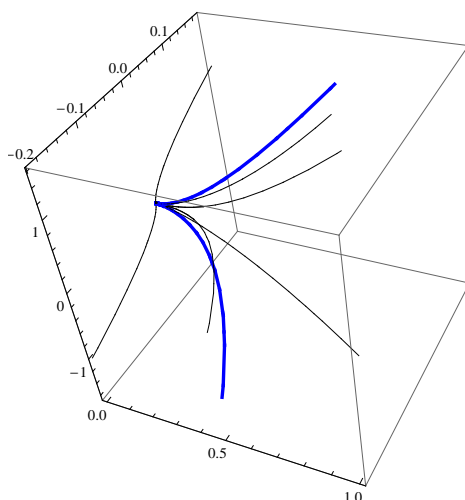


FIGURE 6. Le relèvement $\gamma^1(t) = (x(t) = t^2, y(t) = t^5, p(t) = \frac{5}{2}t^3)$ du cusp d'ordre supérieur $\gamma(t) = (x(t) = t^2, y(t) = t^5)$. On représente en bleu la singularité de $\gamma^1(t)$ en 0 ainsi que ses trois projections sur les trois plans de coordonnées.

obstruction à l'utilisation des formes normales de Cartan $\omega_1 = dy - pdx$, $\omega_2 = dp - qdx$, $\omega_3 = dq - rdx$:

$$p(t) = \frac{y'(t)}{x'(t)} = \frac{5}{2}t^3,$$

$$q(t) = \frac{p'(t)}{x'(t)} = \frac{15}{4}t,$$

$$r(t) = \frac{q'(t)}{x'(t)} = \frac{15}{8} \frac{1}{t} \text{ qui diverge en } t = 0 .$$

On doit donc changer les coordonnées projectives au niveau 3 et prendre pour nouvelle coordonnée $r^*(t) = \frac{x'(t)}{q'(t)} = \frac{8}{15}t$, ce qui correspond non plus à la forme normale de Cartan $\omega_3 = dq - rdx$ mais à la forme normale de Kumpera-Ruiz $\omega'_3 = dx - rdq$ (cf. section 3.4).

La figure 6 représente cette singularité ainsi que ses trois projections sur les trois plans de coordonnées.

3.7.3. Points singuliers verticaux.

Si l'on approfondit le problème du relèvement des singularités, on renoue avec une problématique qui était centrale dans le volume 1 (chapitres 4 et 5) et concernait la notion déterminante de "pinwheel" qui est la clé de l'architecture fonctionnelle de l'aire visuelle primaire V1. Nous avons alors modélisé les pinwheels par des "éclatements". Rappelons ce dont il s'agit. Dans le fibré de contact $\mathbb{P}_{\mathfrak{M}}^1$ la fibre $\mathbb{P}_{a_0}^1$ au-dessus de $a_0 \in \mathbb{R}^2$ est une courbe intégrale de Δ^1 puisque $\Delta_{a_0}^1$ est engendré par les vecteurs ∂_p et $\partial_x + p\partial_y$ et que la fibre correspond simplement à la nullité de la composante relative à $\partial_x + p\partial_y$. Mais cette courbe, dite "verticale", ne correspond à aucun relèvement de courbe. Bien sûr en relevant une courbe $\gamma(t)$ de \mathbb{R}^2 passant par a_0 on peut trouver plusieurs points de la fibre $\mathbb{P}_{a_0}^1$. Pour cela il suffit que $\gamma(t)$ revienne plusieurs fois sur a_0 avec des tangentes différentes, mais cela est un phénomène global correspondant à différentes valeurs du paramètre t . Mais aucune courbe, aussi singulière soit-elle en a_0 ne peut avoir pour relevée une courbe $\gamma^1(t)$ incluse dans la fibre $\mathbb{P}_{a_0}^1$. Cette dernière se projette sur a_0 considérée comme une courbe constante et ne relève pas une direction mais en quelque sorte toutes les directions de \mathbb{R}^2 en a_0 . Elle correspond à la fibre exceptionnelle de l'éclatement de a_0 . Si

$$\gamma_V^1(t) = (a(t) = a_0, p(t) = p_1(t))$$

(V pour vertical) est incluse dans la fibre $\mathbb{P}_{a_0}^1$ avec $\gamma_V^1(0) = (a_0, p(0))$, le germe de $\gamma_V^1(t)$ en 0 ne relève pas une direction mais en quelque sorte un germe de rotation de la direction $p(0)$ autour de a_0 . Mais une courbe verticale $\gamma_V^1(t)$ non constante peut quand même être relevée dans $\mathbb{P}_{\mathfrak{M}}^2$. En effet ses vecteurs tangents en $a_1(t) = \gamma_V^1(t)$ sont $(a'_0 = 0, p'(t))$ et, comme $p'(t)$ n'est pas $\equiv 0$ puisque $\gamma_V^1(t)$ n'est pas constante, même si $p'(t) = 0$ on peut définir la direction tangente $\ell_2(0)$ en 0 en prenant la limite des directions tangentes $\ell_2(t)$ pour t voisin de 0. Si $q(t) = p_2(t)$ est la valeur de la coordonnée q correspondant à $\ell_2(t)$,

$$\gamma_V^1(t) = (a(t) = a_0, p(t) = p_1(t))$$

se relève dans $\mathbb{P}_{\mathfrak{M}}^2$ en

$$\gamma_V^2(t) = (a_0, p(t), q(t) = p'(t)) .$$

Ceci dit, le fait qu'il existe des courbes verticales dans $\mathbb{P}_{\mathfrak{M}}^1$ n'empêche pas tous les points de $\mathbb{P}_{\mathfrak{M}}^1$ d'être réguliers car en tout point (a_0, p_0) passe une relevée legendrienne d'une courbe régulière $a(t)$ de \mathbb{R}^2 passant en a_0 avec la pente p_0 . Mais ce n'est plus le cas à partir du niveau $s = 3$ ($n = 5$) où apparaît une forme normale de Kumpera-Ruiz.

Reprenons la description de la géométrie. Aux points réguliers $\mathbb{P}_{\mathfrak{M}}^3$ la distribution Δ^3 est définie en termes de coordonnées de Cartan par

$$\omega_1 = dy - p dx, \quad \omega_2 = dp - q dx, \quad \omega_3 = dq - r dx .$$

Les $X = (\xi, \eta, \pi, \kappa, \rho) \in \Delta^3$ satisfont

$$\eta = p\xi, \quad \pi = q\xi, \quad \kappa = r\xi$$

et sont donc de la forme

$$X = (\xi, p\xi, q\xi, r\xi, \rho)$$

et les $X_V = (0, 0, 0, 0, \rho)$ sont verticaux. Quand on monte dans les niveaux, on part de $a_0 \in \mathbb{R}^2$. Les points de la fibre de $\mathbb{P}_{\mathfrak{M}}^1$ au-dessus de a_0 correspondent aux pentes p des droites $\ell_{1,p}$ passant par a_0 et le plan de contact $\Delta^1(a_0, p)$ est le plan 1-vertical (i.e. contenant l'axe des p) se projetant sur $\ell_{1,p}$ et qui se "vrille" en colimaçon lorsque p varie. Les points de la fibre de $\mathbb{P}_{\mathfrak{M}}^2$ au-dessus de $a_1 = (a_0, p)$ correspondent aux pentes q des droites $\ell_{2,q}$ de Δ^1 passant par a_1 et le plan $\Delta^2(a_1, q)$ est le plan 2-vertical (i.e. contenant l'axe des q) se projetant sur $\ell_{2,q}$ et qui se "vrille" en colimaçon lorsque q varie. Si donc l'on considère un point q dans cette fibre $\mathbb{P}_{\mathfrak{M},(a_0,p)}^2$, il correspond à la droite q du plan de contact $\Delta^1(a_1) = \Delta^1(a_0, p)$, plan 1-vertical qui se projette sur la droite de pente p passant par a_0 . Si l'on passe au niveau 3, le double vrillage devient un triple vrillage. Les points de la fibre de $\mathbb{P}_{\mathfrak{M}}^3$ au-dessus de $a_2 = (a_1, q) = (a_0, p, q)$ correspondent aux pentes r des droites $\ell_{3,r}$ de Δ^2 passant par a_2 et le plan $\Delta^3(a_1, q, r)$ est le plan 3-vertical (i.e. contenant l'axe des r) se projetant sur $\ell_{3,r}$ et qui se "vrille" en colimaçon lorsque r varie. Si donc l'on considère un point r dans cette fibre $\mathbb{P}_{\mathfrak{M},(a_0,p,q)}^3$, il correspond à la droite de pente r du plan $\Delta^2(a_1, q)$, plan 2-vertical qui se projette sur la droite q du plan de contact $\Delta^1(a_1) = \Delta^1(a_0, p)$, plan 1-vertical qui se projette sur la droite de pente p passant par a_0 . On voit que dans cette suite de projections on finit toujours par arriver à une droite de \mathbb{R}^2 . Cela se voit sur la forme des $X \in \Delta^3$, $X = (\xi, p\xi, q\xi, r\xi, \rho)$. En dehors des $X_V = (0, 0, 0, 0, \rho)$ verticaux, on a $\xi \neq 0$ et donc X ne se projette jamais sur 0 au niveaux inférieurs. Autrement dit, les $X_V = (0, 0, 0, 0, \rho)$, qui sont verticaux à tous les niveaux inférieurs, sont les seuls $X \in \Delta^3$ qui peuvent être verticaux à un niveau inférieur car si un $X \in \Delta^3$ est vertical à un niveau inférieur, alors $\xi = 0$ et X est un X_V .

Regardons maintenant la structure géométrique de $(\mathbb{P}_{\mathfrak{M}}^3, \Delta^3)$ dans le cas de la forme normale de Kumpera-Ruiz

$$\omega_1 = dy - p dx, \quad \omega_2 = dp - q dx, \quad \omega'_3 = dx - r^* dq .$$

Les $X = (\xi, \eta, \pi, \kappa, \rho) \in \Delta^3(a_3)$ satisfont

$$\eta = p\xi, \quad \pi = q\xi, \quad \xi = r^* \kappa$$

et sont donc de la forme

$$X = (r^* \kappa, pr^* \kappa, qr^* \kappa, \kappa, \rho) .$$

Les $X_V = (0, 0, 0, 0, \rho)$ sont verticaux et correspondent à $\kappa = 0$. Mais alors qu'avec ω_3 dans le cas $\xi \neq 0$, l'annulation de p, q ou r n'annulait jamais la projection de X

sur les niveaux inférieurs et qu'il n'y avait donc qu'une direction verticale au-dessus de (a_1, q) , maintenant, avec ω'_3 , la condition $r^* = 0$ implique

$$\xi = \eta = \pi = 0 .$$

Autrement dit le point $r^* = 0$ de la fibre de coordonnée r^* de $\mathbb{P}^3_{\mathfrak{M}}$ au-dessus de $a_2 = (a_1, q) = (a_0, p, q)$, correspond à des vecteurs tangents $X \in \Delta^3(a_3)$ qui se projettent sur 0 dans $\Delta^2(a_2)$, $\Delta^1(a_1)$ et $\Delta^0(a_0)$. Il y a comme avant les X verticaux $X_V = (0, 0, 0, 0, \rho)$ mais il y a aussi, lorsque $r^* = 0$, des

$$X_W = (0, 0, 0, \kappa, \rho)$$

qui ne sont pas verticaux au niveau 3 si $\kappa \neq 0$, mais dont la projection au niveau 2 est, elle, verticale. Cette structure particulière de la tour

$$\Delta^0(a_0), \Delta^1(a_1), \Delta^2(a_2), \Delta^3(a_3 = (a_2, 0))$$

fait de $(a_2, 0)$ un point de niveau 3 en quelque sorte "intrinsèquement" singulier de \mathfrak{M} .

Il en va de même à tous les niveaux $s \geq 3$ (i.e. $\dim \mathbb{P}^s_{\mathfrak{M}} = n \geq 5$). Si

$$a_s = (a_{s-1}, \ell_s), \quad a_{s-1} \in \mathbb{P}^{s-1}_{\mathfrak{M}}, \quad \ell_s \in \Delta^{s-1}(a_{s-1})$$

avec

$$a_{s-1} = (a_{s-2}, \ell_{s-1}), \quad a_{s-2} \in \mathbb{P}^{s-2}_{\mathfrak{M}}, \quad \ell_{s-1} \in \Delta^{s-2}(a_{s-2}) ,$$

on dit que a_s est *vertical* si la direction $\ell_s \in \Delta^{s-1}(a_{s-1})$ est singulière-verticale, i.e. se projette sur 0 dans $\Delta^{s-2}(a_{s-2})$. Au niveau 3, les points verticaux forment une hypersurface lisse \mathcal{V} de $\mathbb{P}^3_{\mathfrak{M}}$. C'est un début de *stratification*. En dehors de \mathcal{V} les points de $\mathbb{P}^3_{\mathfrak{M}}$ sont réguliers et sur \mathcal{V} , qui est de codimension 1, ils sont verticaux.

Au niveau 4 apparaissent des points singuliers dits "*tangentiels*" du type $a_{s+1} = (a_s, \ell_{s+1})$ avec a_s vertical et $\ell_{s+1} \in \Delta^s(a_s)$ tangente à \mathcal{V} . Comme ℓ_{s+1} ne peut pas alors être verticale, un point tangentiel ne peut pas être en même temps vertical : il est au-dessus d'un point fibre vertical d'un étage inférieur. Un point régulier peut se projeter sur un point vertical mais ne peut pas se projeter sur un point tangentiel. On obtient ainsi trois classes de points de \mathfrak{M} codables par les lettres R (régulier), V (vertical), T (tangentiel). Dans $\mathbb{P}^4_{\mathfrak{M}} \rightarrow \mathbb{P}^3_{\mathfrak{M}}$ on a une stratification avec la strate ouverte dense (générique) des points R, deux strates de codimension 1, celle RV des points R de $\mathbb{P}^3_{\mathfrak{M}}$ avec au-dessus (en fibre) un point V de $\mathbb{P}^4_{\mathfrak{M}}$ et celle VR des points V de $\mathbb{P}^3_{\mathfrak{M}}$ avec au-dessus un point R de $\mathbb{P}^4_{\mathfrak{M}}$, ainsi que deux strates de codimension 2, celle VV et celle VT.

On peut alors utiliser ces *points* singuliers de \mathfrak{M} (définis purement géométriquement par la nature des stratifications) pour préciser la nature des points réguliers et singuliers de courbes paramétrées dans les $\mathbb{P}^s_{\mathfrak{M}}$, de leurs relèvements dans des niveaux supérieurs et de leurs projections dans des niveaux inférieurs.

La construction géométrique itérative de \mathfrak{M} permet de mieux comprendre les formes normales de Kumpéra-Ruiz en regardant comment les coordonnées locales doivent s’y adapter. À chaque étage $\mathbb{P}_{\mathfrak{M}}^s$ introduit un axe vertical supplémentaire au-dessus des a_{s-1} de $\mathbb{P}_{\mathfrak{M}}^{s-1}$ avec une nouvelle coordonnée z_s . Celle-ci est une coordonnée affine pour les droites ℓ_s de la 2-distribution $\Delta^{s-1}(a_{s-1})$. Soient (x_{s-1}, y_{s-1}) des coordonnées adaptées dans $\Delta^{s-1}(a_{s-1})$. Alors la contrainte que z_s doit représenter une nouvelle “dérivée cachée” se traduit par le fait que $[dx_{s-1} : dy_{s-1}]$ sont des coordonnées homogènes (i.e. projectives) pour les ℓ_s (i.e. dx_{s-1}, dy_{s-1} sont les composantes du déplacement ds sur ℓ_s) et que $z_s = \frac{dy_{s-1}}{dx_{s-1}}$ ou $z_s = \frac{dx_{s-1}}{dy_{s-1}}$ suivant les cas. D’où la possibilité d’introduire à l’étage s la nouvelle 1-forme

$$\omega_s = dy_{s-1} - z_s dx_{s-1}$$

ou

$$\omega_s = dx_{s-1} - z_s dy_{s-1} .$$

C’est l’origine de la récurrence des formes normales. Par exemple à l’étage 1, si l’on a une droite ℓ_1 de \mathbb{R}^2 , on peut toujours choisir des coordonnées (x, y) de \mathbb{R}^2 telles que $dx \neq 0$ sur ℓ_1 et l’on aura alors $z_1 (= p) = \frac{dy}{dx}$ de coordonnées homogènes $[1 : \frac{dy}{dx}]$ qui annule la 1-forme $\omega_1 = dy - p dx$. À l’étage 2, si ℓ_2 est une droite de $\Delta^1(a_1)$ avec a_1 voisin de 0 se projetant sur a_0 , a_1 est un couple (a_0, p_1) avec p_1 la pente d’une droite ℓ_1 de $T_{a_0}\mathbb{R}^2 \simeq \mathbb{R}^2$. Le plan $\Delta^1(a_1)$ est le plan engendré par la fibre verticale et ℓ_1 . On peut alors choisir ℓ_1 pour axe des x et les coordonnées homogènes des droites de $\Delta^1(a_1)$ deviennent $[dx : dp]$. Si ℓ_2 n’est pas la fibre, $dx \neq 0$ sur ℓ_2 et la nouvelle coordonnée $z_2 = q$ peut s’exprimer comme $z_2 (= q) = \frac{dp}{dx}$ de coordonnées homogènes $[1 : \frac{dp}{dx}]$ qui annule la 1-forme $\omega_2 = dp - q dx$. Δ^2 est alors localement la distribution d’Engel noyau de ω_1 et ω_2 . Si en revanche, ℓ_2 est la fibre (direction critique verticale), alors $dx = 0$ mais $dp \neq 0$ et on peut prendre $z_2 = \frac{dx}{dp}$ de coordonnées homogènes $[\frac{dx}{dp} : 1]$ qui annule la 1-forme $\omega'_2 = dx - q dp$. Δ^2 est alors localement la distribution noyau de ω_1 et ω'_2 . On montre qu’elle est contactomorphe à la distribution d’Engel.

3.7.4. Courbes et points critiques.

Le problème des singularités des courbes paramétrées est que, dans les relèvements successifs, on peut partir de courbes $\gamma_{s-1}(t) = a_{s-1}(t)$ dans $\mathbb{P}_{\mathfrak{M}}^{s-1}$ qui sont bien tangente à $\Delta^{s-1} \subset T\mathbb{P}_{\mathfrak{M}}^{s-1}$ (i.e. sont des courbes intégrales de Δ^{s-1}) mais telles que $a'_{s-1}(t_0) = 0$. Si $a'_{s-1}(t) \neq 0$ partout au voisinage de t_0 et donc même en t_0 , on dit, nous l’avons vu, que $a_{s-1}(t)$ est *immergée*. Dans ce cas, on peut relever $\gamma_{s-1}(t)$ en $\Gamma_s(t) = ((a_{s-1}(t), \ell_s(t)))$ dans $\mathbb{P}_{\mathfrak{M}}^s$ en prenant $\ell_s(t) = \text{Span}\{a'_{s-1}(t)\}$. C’est toujours le cas pour les jets de *graphes* de fonctions $y = f(x)$ parce qu’ils sont du type

$$a_{s-1}(t) = \{x = t, f(x), f'(x), \dots, f^{(s-1)}(x)\}$$

et donc

$$a'_{s-1}(t) = \{1, f'(x), f''(x), \dots, f^{(s)}(x)\}$$

est toujours $\neq 0$. Tous ces points sont réguliers et correspondent à la forme normale de Cartan.

Si $a'_{s-1}(t_0) = 0$ mais si $\gamma_{s-1}(t) = a_{s-1}(t)$ est *non constante*, on peut quand même effectuer le relèvement. Pour cela, on effectue les relèvements pour $t \neq t_0$ et on prend leur limite pour $t \rightarrow t_0$. Elle est bien définie et donne une courbe intégrale $\Gamma_s(t)$ non constante. En effet, soient X et Y des générateurs de Δ^{s-1} localement en $a_{s-1}(t_0)$. Alors

$$a'_{s-1}(t) = \xi(t) X(a_{s-1}(t)) + \eta(t) Y(a_{s-1}(t))$$

avec $\xi(t_0) = \eta(t_0) = 0$. Comme $a_{s-1}(t)$ est non constante, il existe, si l'on prend les développements de Taylor, un r tel que

$$\xi(t) = (t - t_0)^r \xi_1(t), \quad \eta(t) = (t - t_0)^r \eta_1(t)$$

avec $\xi_1(t_0)$ ou $\eta_1(t_0) \neq 0$, par exemple $\eta_1(t_0)$. La limite $\frac{\xi_1(t)}{\eta_1(t)}$ pour $t \rightarrow t_0$ est alors bien définie.

Comme nous l'avons dit, Richard Montgomery et Michail Zhitomirskii montrent (i) que tout point $a_{1+k} \in \mathbb{P}_{\mathfrak{M}}^{1+k}$ est le k -ème relèvement $\gamma^k(0)$ en un point (pris en $t = 0$) d'une courbe legendrienne $\gamma^1(t)$ de $\mathbb{P}_{\mathfrak{M}}^1$ et (ii) que les points réguliers correspondent aux $\gamma^k(0)$ de courbes legendriennes immergées régulières $\gamma^1(t)$ (i.e. $\gamma^{1'}(0) \neq 0$).

Le point (i) montre bien que la tour \mathfrak{M} compactifie projectivement les fibres des projections successives des espaces de jets et que la géométrie de la tour des \mathfrak{G} -distributions Δ^s *géométrise intrinsèquement les développements en jets*. C'est pourquoi, pour la neurogéométrie (car il ne faut jamais perdre de vue cette motivation première), \mathfrak{M} est comme un "réservoir de modèles" *pour les architectures fonctionnelles rétinotopiques*.

La possibilité de singularités est liée à l'existence de courbes intégrales *critiques* des $(\mathbb{P}_{\mathfrak{M}}^s, \Delta^s)$ qui se projettent de façon *constante* sur $\mathbb{P}_{\mathfrak{M}}^1$. La direction ℓ_s de $\Delta^{s-1}(a_{s-1})$ est critique verticale s'il existe un germe de courbe immergée $a_{s-1}(t)$, $a'_s(0) \neq 0$ de projection $a_{s-2}(t)$ *constante* qui est tangente à ℓ_s en a_s . Les directions critiques verticales sont incluses dans les fibres de la projection $\pi_s : \mathbb{P}_{\mathfrak{M}}^{s-1} \rightarrow \mathbb{P}_{\mathfrak{M}}^{s-2}$ et sont du type

$$V_s(t) = (a_{s-1}(t) = a_{s-1}, \ell_s(t)) \text{ avec } a'_{s-1}(t) \equiv 0 .$$

Cela commence dès le niveau 3. Elles peuvent aussi être des prolongements à l'étage s de courbes verticales d'un étage inférieur et, comme nous l'avons vu, on dit qu'elles sont "tangentiellles". Elles sont immergées si $\ell'_s(t) \neq 0$ puisque $V'_s(t) = (0, \ell'_s(t))$. Si $V_s(t)$ est verticale, son prolongement au niveau $s + 1$ donne un point singulier

vertical et ses prolongements aux niveaux $\geq s + 2$ donnent des points singuliers tangentiels.

Cela permet de corrélérer à chaque étage la nature des points (réguliers et critiques, verticaux ou tangentiels) et le comportement de courbes passant par ces points. Un point de \mathfrak{M} est régulier si toutes ses projections sont des points réguliers, autrement dit s'il correspond à une succession de prolongations régulières à partir d'une courbe legendrienne régulière de $\mathbb{P}_{\mathfrak{M}}^1$. Si a_s est régulier, alors la direction verticale est la seule direction critique au-dessus de a_s . Un point $a_s = (a_{s-1}, \ell_s)$ de $\mathbb{P}_{\mathfrak{M}}^s$ avec $a_{s-1} = (a_{s-2}, \ell_{s-1})$ est vertical si $\ell_s \in \Delta^{s-1}$ est une direction verticale de niveau $s - 1$, i.e., comme nous l'avons vu, si elle se projette sur 0 dans $\Delta^{s-2}(a_{s-2})$. Si a_s est critique (vertical ou tangentiel), il existe *deux* directions critiques au-dessus de a_s : la direction verticale et la direction tangentielle.

Revenons à l'exemple du supercusp $\gamma(t) = (x(t) = t^2, y(t) = t^5)$ qui, pour son relèvement dans $\mathbb{P}_{\mathfrak{M}}^3$ nécessite la forme de K-R

$$\omega'_3 = dx - r^*dq .$$

En 0, $r^*(0) = \frac{8}{15}t = 0$ et donc $\gamma^3(0)$ est un point singulier vertical de $\mathbb{P}_{\mathfrak{M}}^3$ appartenant à la strate V. On trouvera dans [573] l'exemple plus sophistiqué de $\gamma(t) = (x(t) = t^3, y(t) = t^7)$.

Il existe un *théorème de désingularisation* disant que l'on peut désingulariser les courbes possédant des points singuliers en effectuant un nombre assez grand de prolongements successifs.

4. Structures de contact globales

4.1. Local/global : une dialectique fondamentale

Nous avons vu que les structures de contact et les structures symplectiques ont des formes normales locales uniques (les formes normales de Darboux standard et universelles) et sont donc toutes localement isomorphes. Elles ne possèdent pas d'invariants locaux. Mais il n'en va pas du tout de même au niveau *global* et leur géométrie globale peut être très compliquée. Nous n'approfondirons pas vraiment ce point car nos modèles ne le nécessitent pas, les structures que nous utilisons étant non seulement localement mais aussi globalement triviales. Mais nous allons quand même en dire un mot car c'est à ce niveau que l'on rencontre de la géométrie moderne et profonde allant bien au-delà des calculs élémentaires que nous utilisons. Et puis cela nous permettra de rendre hommage à d'éminents géomètres comme Mikhaïl Gromov ou Daniel Bennequin (cf. [41]) que nous avons déjà souvent cités et qui ont beaucoup travaillé en neurosciences.

Et enfin l'on nous permettra peut-être une note plus biographique. Dans les années 1970-1980 pendant lesquelles nous travaillions sur la théorie des singularités en géométrie différentielle, ces problèmes faisaient partie de notre pain quotidien.

La découverte qu’une structure de contact était implémentée dans l’architecture en pinwheels de l’aire visuelle primaire $V1$ a été l’événement déclencheur de ce que nous avons proposé d’appeler “neurogéométrie”. La convergence de nos intérêts d’un côté pour la géométrie de contact et, d’un autre côté, pour les neurosciences cognitives nous a motivés pour transférer les méthodes mathématiques de la première aux résultats empiriques de la seconde.

La dialectique du local et du global, non seulement en géométrie mais aussi dans de nombreux autres domaines comme en arithmétique, est l’une des plus profondes et des plus omniprésentes en mathématiques. Elle possède une portée non seulement technique mais également philosophique remarquable. Elle est l’un des thèmes récurrents de la philosophie mathématique d’Albert Lautman (cf. notre étude [414]). Nous en avons présenté plusieurs aspects dans l’item “Locale/Globale” de l’*Enciclopedia Einaudi* [411].

4.2. Contraintes d’existence de structures de contact globales

Nous allons brièvement exposer quelques idées de base qui se sont développées surtout à partir des années 1970. Pour avoir une idée de leur mise en place en contexte, le lecteur pourra consulter par exemple la *IIIe Rencontre de géométrie du Schnepfenried* (10-15 mai 1982) [540] dédiée à Georges Reeb qui ouvrit au début des années 1950 la voie à l’étude des propriétés globales des structures de contact. C’est un excellent instantané de la situation à cette époque charnière.

L’existence d’un champ $\mathcal{K} = \ker(\omega)$ d’hyperplans de contact (i.e. d’une distribution maximale non intégrable d’hyperplans tangents) sur une variété M^{2n+1} de dimension impaire $2n + 1$ n’est possible que si M satisfait certaines contraintes. Si une telle distribution \mathcal{K} existe, elle est définie localement sur des cartes locales U de M par des formes de contact ω_U (isomorphes à la forme normale de Darboux si les U sont assez petits). Mais cela n’implique pas nécessairement, entre autres à cause de problèmes d’orientabilité cohérente des K_x , que les ω_U puissent se recoller en une forme de contact globale ω . Cela est toujours vrai si M est orientable de dimension $4m+1$ mais pas nécessairement vrai si M est de dimension simplement impaire $2n+1$. Une contrainte est que le quotient TM/\mathcal{K} (de dimension $2(2n + 1) - 2n = 2n + 2$) soit orientable (on dit dans ce cas que \mathcal{K} est co-orientée) car il s’agit d’un fibré *en droites* sur M et l’on montre que s’il est orientable il est automatiquement trivial.

En dimension 3 l’orientabilité est la seule contrainte pour les variétés *fermées* (i.e. compactes sans bord). On voit facilement qu’elle est nécessaire. Quand on recolle les ω_U locales sur les $U_{ij} = U_i \cap U_j$ on doit avoir $\omega_{U_i} = \lambda_{ij}\omega_{U_j}$ sur U_{ij} avec $\lambda_{ij} \neq 0$. Mais $\omega_U \wedge d\omega_U$ est une forme volume et définit donc une orientation. Comme $\omega_{U_i} \wedge d\omega_{U_i} = \lambda_{ij}^2 \omega_{U_j} \wedge d\omega_{U_j}$ avec $\lambda_{ij}^2 > 0$, l’orientation définie par les ω_U est partout la même et définit une orientation globale de M . Un théorème de Jean Martinet [355], Robert Lutz [341], William Thurston et Horst Winkelnkemper [535], démontre la

réciroque : si M^3 est orientée et fermée et si \mathcal{K} est un champ de plans de contact, alors il est défini par une 1-forme de contact ω globale.

Une autre façon de voir que l'existence de structures de contact impose des contraintes globales à la géométrie de M est la suivante. Si M^{2n+1} est une variété différentiable, le groupe structural naturel G opérant sur ses plans tangents est celui des applications linéaires tangentes aux changements de coordonnées locales autorisés et, comme ceux-ci sont des difféomorphismes locaux quelconques, G est le groupe linéaire complet $GL(2n+1)$ des automorphismes linéaires de \mathbb{R}^{2n+1} . Mais si une ω globale existe, ω induit une structure symplectique sur les hyperplans K_x et donc, d'après ce que nous avons vu à la section 2, une structure presque-complexe J (maximalement non intégrable) sur \mathcal{K} satisfaisant

$$J^2 = -Id, \quad d\omega(JX, JY) = d\omega(X, Y), \quad d\omega(X, JX) > 0,$$

$d\omega(X, JY) = -d\omega(JX, Y)$ définissant une métrique hermitienne sur \mathcal{K} . La distribution de contact \mathcal{K} devient alors une fibration de fibres isomorphes à \mathbb{C}^n et le groupe G devient réductible au groupe $U(n) \oplus Id$ opérant sur $\mathbb{C}^n \oplus \mathbb{R}$, où $U(n)$ est le groupe des automorphismes linéaires unitaires de \mathbb{C}^n . En effet, de façon plus précise,

- la décomposition de TM en somme directe

$$TM = \mathcal{K} \oplus \mathbb{R}\chi_\omega$$

induite par ω (χ_ω est le champ de Reeb) permet d'abord de réduire le groupe structural G de $GL(2n+1, \mathbb{R})$ à

$$GL(2n, \mathbb{R}) \oplus Id ;$$

- ensuite, le fait que $d\omega$ soit symplectique sur \mathcal{K} permet de réduire G à

$$Sp(2n, \mathbb{R}) \oplus Id ,$$

$Sp(2n, \mathbb{R})$ étant le groupe symplectique sur \mathbb{R}^{2n} ;

- enfin l'existence de J compatible avec ω permet de réduire G à

$$(Sp(2n, \mathbb{R}) \cap GL(n, \mathbb{C})) \oplus Id .$$

- Mais $Sp(2n, \mathbb{R}) \cap GL(n, \mathbb{C}) = U(n)$.

On notera que l'existence d'une structure presque complexe J sur un champ de plans \mathcal{K} est donc une condition plus faible que le fait que \mathcal{K} soit de contact. On dit que \mathcal{K} est "presque" de contact et on se demande à quelles conditions le "presque" peut devenir un "exactement". Par exemple, si M est *ouverte* ou si M est de dimension 3, un théorème de Mikhaïl Gromov (1969) dit que l'existence d'une structure presque complexe garantit l'existence d'une structure de contact.

4.3. Le problème de la classification.

Quand on cherche à classifier les structures de contact, on le fait souvent au moyen de classes d'équivalences définies par des *déformations* et il est important de distinguer différents types et différents contextes de déformations. Par exemple, on peut déformer \mathcal{K} au moyen d'une famille \mathcal{K}_t , $t \in [0, 1]$ ($\mathcal{K}_0 = \mathcal{K}$) en restant dans la classe des champs de plans de contact, ou, plus faiblement, dans la classe des champs de plans définissant une structure presque-complexe, ou, plus faiblement encore, en restant simplement dans la classe des champs de plans pas nécessairement presque-complexes ou de contact. Les résultats sont très différents suivant les cas.⁴³ Les déformations peuvent être de telles "homotopies" \mathcal{K}_t déformant directement \mathcal{K} ou, ce qui est plus contraignant, des "isotopies" déformant l'espace ambiant M tout entier par des difféomorphismes φ_t ($\varphi_0 = Id$) induisant une déformation \mathcal{K}_t de \mathcal{K} . On rencontre beaucoup de subtilités dans ces domaines.

Par exemple, d'après un théorème de John Gray dit "théorème de stabilité" ([221], 1959), les déformations continues (homotopies) \mathcal{K}_t , $t \in [0, 1]$, de distributions de contact sur une variété *fermée* (compacte, sans bord) peuvent toujours être obtenues par isotopies φ_t (i.e. les \mathcal{K}_t sont nécessairement les transformées de \mathcal{K}_0 par des difféomorphismes φ_t de M qui déforment l'identité). Donc, pour une classification à isotopie près, les déformations \mathcal{K}_t de \mathcal{K} sont dans ce cas équivalentes à \mathcal{K} et l'on peut envisager des possibilités de classification par des invariants. Ce théorème est faux pour les formes de contact dont les déformations peuvent complètement (et donc trop) transformer les champs de Reeb.

Par exemple, encore, si M est *ouverte* le théorème de Mikhaïl Gromov (1969) cité plus haut dit qu'une homotopie \mathcal{K}_t de \mathcal{K}_0 à \mathcal{K}_1 de structures de contact dans la classe des structures presque complexes implique l'existence d'une homotopie de \mathcal{K}_0 à \mathcal{K}_1 dans la classe des structures de contact et que donc \mathcal{K}_0 et \mathcal{K}_1 sont homotopes si et seulement si J_0 et J_1 sont homotopes.

4.4. Trajectoires fermées du champ de Reeb et homologie de contact

Un outil important pour analyser la géométrie globale des variétés de contact (M^{2n+1}, ω) est le champ de Reeb χ_ω (chapitre 3, section 4 et section 2.3) défini par $i_{\chi_\omega} d\omega \equiv 0$ et $i_{\chi_\omega} \omega = \omega(\chi_\omega) \equiv 1$. Ce champ partout transverse à \mathcal{K} et égal à $\chi = \frac{\partial}{\partial y}$ pour la forme normale de Darboux, est, nous l'avons vu, le prototype des champs de contact et lève la sous-détermination de \mathcal{K} par rapport à ω . Comme il est partout non nul, il est localement trivial et définit un feuilletage de M par des courbes C^∞ .

43. Par exemple nous parlerons de nœuds dans un instant. On peut aussi les classer en utilisant des déformations. Mais si dans une déformation de nœuds γ_t on permet à certains des γ_t de posséder des croisements (où la ficelle du nœud peut se traverser) alors tous les nœuds deviennent équivalents entre eux et triviaux.

Mais, globalement, il n'a pas de raison d'être trivial. Certes, il l'est dans le modèle de Darboux $\omega = dy - pdx$ sur \mathbb{R}^3 et il définit le feuilletage de \mathbb{R}^3 par les droites parallèles à l'axe des y . Mais il peut avoir une structure très riche.

D'abord, χ_ω peut avoir des trajectoires \mathfrak{p} fermées, donc *périodiques*, et, en dimension 3, celles-ci sont des *nœuds* \mathfrak{n} . La théorie des nœuds, qui est un fleuron de la topologie algébrique, va par conséquent intervenir naturellement et massivement dans la géométrie de contact de dimension 3.⁴⁴ Ensuite, surtout si M est fermée, χ_ω peut avoir des trajectoires non fermées qui s'enroulent de façon compliquée.

Les trajectoires fermées \mathfrak{p} de χ_ω sont essentielles, en particulier pour la raison suivante. Au niveau global, ω étant une 1-forme et les 1-formes étant des entités différentielles destinées être intégrées le long de sous-variétés de dimension 1, il est naturel de considérer les intégrales

$$\mathfrak{A}(\gamma) = \int_\gamma \omega$$

sur les courbes γ fermées, $\gamma : \mathbb{S}^1 \rightarrow M$.⁴⁵ La fonctionnelle

$$\mathfrak{A} : \mathcal{C}^\infty(\mathbb{S}^1, M) \rightarrow \mathbb{R}$$

est l'équivalent d'une "fonctionnelle d'action".⁴⁶ et l'on démontre facilement que si $\gamma = \mathfrak{p}$ est une trajectoire périodique de χ_ω alors elle est un *point critique* de \mathfrak{A} et réciproquement. En effet, soit γ_s une petite déformation régulière de $\gamma_0 = \gamma$, s appartenant à un petit voisinage de $s = 0$ dans \mathbb{R} . L'homotopie γ_s intègre le champ infinitésimal de déformation le long de γ_0 , $X = \frac{d}{ds}\gamma_s|_{s=0}$. On a alors (cf. par exemple Bourgeois [72])

$$\frac{d}{ds}\mathfrak{A}(\gamma_s)\Big|_{s=0} = \frac{d}{ds} \int_{\gamma_s} \omega \Big|_{s=0} = \frac{d}{ds} \int_{\mathbb{S}^1} \gamma_s^*(\omega) \Big|_{s=0} = \int_{\mathbb{S}^1} \gamma^*(\mathcal{L}_X\omega) = \int_\gamma \mathcal{L}_X\omega .$$

Comme $\mathcal{L}_X\omega = i_X d\omega + d(\omega(X))$ et que $\int_\gamma d(\omega(X)) = 0$ (intégrale $\int_\gamma df$ d'une 1-forme exacte sur un cycle), cela implique

$$\frac{d}{ds}\mathfrak{A}(\gamma_s)\Big|_{s=0} = \int_\gamma i_X d\omega .$$

Si $\gamma = \mathfrak{p}$ est une trajectoire fermée du champ de Reeb,

$$d\omega(\dot{\mathfrak{p}}, \bullet) = i_{\dot{\mathfrak{p}}}d\omega = 0$$

44. La topologie algébrique intervient entre autres parce qu'un invariant fondamental d'un nœud \mathfrak{n} dans M^3 est le groupe fondamental $\pi_1(M - \mathfrak{n})$ du complémentaire de \mathfrak{n} dans M .

45. Appelées traditionnellement "lacets" ("loops") en théorie de l'homotopie et "cycles" en théorie de l'homologie.

46. Nous avons rencontré plusieurs fonctionnelles d'action dans les chapitres consacrés aux modèles variationnels.

et donc γ est trivialement un point critique de \mathfrak{A} . Réciproquement si $\gamma = \mathfrak{p}$ est un point critique de \mathfrak{A} on a pour tout X

$$\left. \frac{d}{ds} \mathfrak{A}(\gamma_s) \right|_{s=0} = \int_{\gamma} i_X d\omega = \int_{\gamma} d\omega(X, \dot{\mathfrak{p}}) = 0$$

et donc $d\omega(\dot{\mathfrak{p}}, \bullet) = i_{\dot{\mathfrak{p}}} d\omega = 0$.⁴⁷ Cela signifie que $\dot{\mathfrak{p}}$ est proportionnel à χ_{ω} . Si l'on paramétrise \mathfrak{p} de façon à ce que $\dot{\mathfrak{p}}$ devienne égal à χ_{ω} , alors la *valeur critique* $\mathfrak{A}(\mathfrak{p}) = \int_{\mathfrak{p}} \omega$ de \mathfrak{A} en \mathfrak{p} devient la *période* T de \mathfrak{p} .

Cette remarque conduit à la notion d'*homologie de contact*, introduite par Yakov Eliashberg, Alexander Givental et Helmut Hofer, en particulier dans leur classique *Introduction to Symplectic Field Theory* [167]. Il s'agit d'une théorie profonde et sophistiquée dont le lecteur pourra prendre plus ample connaissance par exemple à travers les textes de Frédéric Bourgeois *Introduction to Contact Homology* [72] et *Lectures on Symplectic and Contact Homology* [73] et l'ouvrage collectif *Contact and Symplectic Topology* [128].

Comme l'explique F. Bourgeois dans [72]

“Contact homology can be considered intuitively as a variant of Morse theory for the action functional.”⁴⁸

On s'inspire en effet de la théorie de Morse pour analyser la fonctionnelle \mathfrak{A} . Pour définir les points critiques \mathfrak{p} non dégénérés, on regarde l'application de Poincaré de premier retour le long de \mathfrak{p} . Si \mathfrak{p} est de période T et si $x \in \mathfrak{p}$, le flot de Reeb R_t intégrant χ_{ω} ramène x sur lui-même au temps T . K_x est par définition transverse à χ_{ω} en x et, sur un voisinage U assez petit de x dans K_x , le flot R_t ramène les $y \in U$ dans U après un temps t voisin de T . On définit ainsi un difféomorphisme de premier retour au voisinage de x dans K_x et la restriction $S_{x,T}$ à K_x de son application linéaire tangente est un automorphisme symplectique car R_t préserve la 2-forme $d\omega$ qui est symplectique sur \mathcal{K} par définition. On montre que \mathfrak{p} est un point critique non dégénéré si et seulement si cette restriction n'a pas de valeur propre égale à 1. La non-dégénérescence est une propriété générique. Quand on suit $R_t(x)$ le long de \mathfrak{p} , $S_{x,t} : K_x \rightarrow K_{R_t(x)}$ est un chemin d'applications symplectiques et on regarde pour quelles valeurs de $t \in]0, T]$ celles-ci peuvent avoir une valeur propre = 1, la condition étant $\det(S_{x,t} - Id) = 0$. Ces valeurs de t sont appelées des “croisements”. Si t_c est un croisement et si $N_{t_c} = \ker(S_{x,t_c} - Id)$ est l'espace propre

47. Pour une 1-forme quelconque, ϖ l'intégrale $\int_{\gamma} \varpi$ est l'intégrale $\int_{\gamma} \varpi(\dot{\gamma}) dt$ le long de γ , supposée paramétrée par t , de la valeur de ϖ appliquée au vecteur vitesse $\dot{\gamma}$.

48. Nous avons évoqué la théorie de Morse dans la section 1.5. Nous l'avons utilisée dans le Vol I, section 5.11.3., pour modéliser des résultats de Nicholas Swindale sur les cartes corticales de direction (cf. aussi [444]) et section 4.6.12. à propos des champs gaussiens. Elle occupe une place centrale dans notre compilation [412] de la théorie de Thom-Mather.

de la valeur propre 1, on peut considérer la forme quadratique sur N_{t_c}

$$Q_{t_c}(n) = d\omega(n, \dot{S}_{x,t_c}) ,$$

t_c étant dit non dégénéré si Q_{t_c} l'est. À partir des différents croisements et des signatures des formes quadratiques associées on peut définir dans le cas de non dégénérescence un *index* de \mathfrak{p} , dit index de Conley-Zehnder, qui est l'analogue de l'index de Morse pour les points critiques non dégénérés.

4.5. Courbes pseudo-holomorphes de Gromov

Un outil puissant introduit par Mikhaïl Gromov au début des années 1980 (cf. par exemple [227]) est celui des courbes “pseudo-holomorphes” ou J -holomorphes. (M, \mathcal{K}, ω) étant une variété de contact, on la symplectise en considérant la variété produit $M \times \mathbb{R}_{(s)} = \widetilde{M}$ munie de la forme

$$\sigma = d(e^s \omega) = e^s ds \wedge \omega + e^s d\omega$$

qui est bien symplectique⁴⁹ et en y prolongeant la structure presque-complexe J définie sur \mathcal{K} en posant

$$J \frac{\partial}{\partial s} = \chi_\omega .$$

Autrement dit, on couple le champ de Reeb χ_ω (qui est partout transverse à \mathcal{K} et permet de décomposer le fibré tangent de M en $TM = \mathcal{K} \oplus \mathbb{R}\chi_\omega$) avec la nouvelle variable s de façon à en faire un couple de variables conjuguées complétant la structure symplectique définie par $d\omega$ sur \mathcal{K} . Cette structure presque-complexe $J_{\widetilde{M}}$ sur \widetilde{M} permet de considérer des “courbes” J -holomorphes⁵⁰, c'est-à-dire des applications holomorphes φ de surfaces de Riemann (Σ, J_Σ) ⁵¹ dans \widetilde{M} , l'holomorphie signifiant la commutation de l'application linéaire tangente $T\varphi$ avec les structures complexes : $T\varphi \circ J_\Sigma = J_{\widetilde{M}} \circ T\varphi$. Du coup, toute la théorie, subtile, profonde et compliquée, des surfaces de Riemann se trouve mobilisée.

En particulier, on étudie les courbes J -holomorphes qui envoient la sphère de Riemann \mathbb{S}^2 épointée des points z_0, z_1, \dots, z_r , dans \widetilde{M} de façon à ce que, à la limite, z_0 (considéré comme limite de cercles concentriques) s'envoie sur une trajectoire périodique \mathfrak{p}_0 de χ_ω située à $s = +\infty$ et z_1, \dots, z_r (considérés aussi comme limites de cercles concentriques) sur des trajectoires périodiques $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ situées à $s = -\infty$ et orientées en sens inverse. La théorie est tout à fait fascinante. Elle est bien exposée dans Frédéric Bourgeois [72].

49. σ est bien symplectique puisque, $\omega \wedge (d\omega)^n$ étant par hypothèse une forme volume sur M , $\sigma^{n+1} = e^{(n+1)s} ds \wedge \omega \wedge (d\omega)^n$ est une forme volume sur \widetilde{M} .

50. Ces courbes sur \mathbb{C} sont des surfaces sur \mathbb{R} .

51. Les surfaces de Riemann étant des courbes sur \mathbb{C} , elles sont par définition munies d'une structure complexe J_Σ .

4.6. La dimension 3 et la théorie des nœuds (Bennequin)

4.6.1. Nœuds et feuilletages caractéristiques.

Évidemment, on a particulièrement étudié le cas d'espaces classiques comme \mathbb{R}^N , \mathbb{S}^N (sphères) ou \mathbb{T}^N (tores), en particulier en dimension 3. En dimension 3, un phénomène fondamental est, nous l'avons dit, l'existence de nœuds \mathbf{n} et il peut a priori y en avoir de deux types complètement opposés : d'une part les nœuds legendriens $\mathbf{n} \parallel \mathcal{K}$ tangents à \mathcal{K} (et qui sont donc des courbes intégrales de \mathcal{K}) et d'autre part les nœuds $\mathbf{n} \pitchfork \mathcal{K}$ transverses à \mathcal{K} comme les trajectoires fermées de χ_ω . Il est donc naturel de voir comment les plans de contact tournent lorsqu'on se déplace le long de tels nœuds.

Par ailleurs, nous avons vu à la section 2.9 du chapitre 3 que si l'on se donne une surface S dans M^3 , elle ne peut jamais être localement (i.e. sur un ouvert, aussi petit soit-il) une surface intégrale de \mathcal{K} puisque celles-ci n'existent pas (eu égard à la non-intégrabilité maximale de \mathcal{K}). Mais cela ne l'empêche pas d'avoir un plan tangent $T_x S$ confondu avec le plan de contact K_x pour certains points x exceptionnels. Génériquement, ces points, dits points *critiques* (ou points singuliers ou encore points caractéristiques), sont isolés. Soit C l'ensemble des points critiques de S . Sur $S - C$, $T_x S$ et K_x sont transverses et leur intersection définit une droite, d'où un champ de droites dit *champ caractéristique* sur S définissant un feuilletage, dit *feuilletage caractéristique*, de $S - C$ par des courbes. Ce champ a pour singularités les points critiques et ces derniers peuvent être génériquement soit des points "elliptiques" (foyers ou nœuds où, localement, toutes les directions sont présentes), soit des points "hyperboliques" (cols).

Si S est orientée, le champ de droites caractéristique est définissable par un champ de vecteurs et les points elliptiques sont alors soit des sources, soit des puits. Ces champs caractéristiques peuvent avoir des trajectoires fermées avec les trajectoires non fermées circulant asymptotiquement entre points critiques et trajectoires fermées. Leurs tableaux peuvent être très compliqués. De nombreux géomètres comme par exemple Frédéric Bourgeois, John Etnyre, Andreas Floer, Hansjörg Geiges, Emmanuel Giroux, Ko Honda, François Laudenbach ou Michael Sullivan, les ont étudiés. On pourra en particulier consulter les excellentes *Introductory Lectures* [172] d'Etnyre.

4.6.2. L'exemple de \mathbb{S}^3 (fibration de Hopf et quaternions).

Un cas particulièrement intéressant est celui de \mathbb{S}^3 qui généralise directement nos modèles de base $\mathbb{V}_J \simeq \mathbb{R}^3$ (Heisenberg) et $\mathbb{V}_S = SE(2) = \mathbb{R}^2 \times \mathbb{S}^1$ (groupe euclidien). On part de l'inclusion de \mathbb{S}^3 dans $\mathbb{C}^2 = \mathbb{R}^4$ comme sphère unité. Si

$$z = (z_1 = x_1 + iy_1, z_2 = x_2 + iy_2)$$

sont les coordonnées sur \mathbb{C}^2 ($z = (x_1, y_1, x_2, y_2)$ sur \mathbb{R}^4), \mathbb{S}^3 est d'équation

$$z_1\bar{z}_1 + z_2\bar{z}_2 = |z_1|^2 + |z_2|^2 = 1 .$$

La structure complexe de \mathbb{C}^2 est donnée par la multiplication par i dans les deux plans complexes z_1 et z_2 , c'est-à-dire pour $Z = (X_1, Y_1, X_2, Y_2) \in T_z\mathbb{C}^2$ par

$$J_z(X_1, Y_1, X_2, Y_2) = (-Y_1, X_1, -Y_2, X_2) ,$$

$J^2 = -1$, et, comme \mathbb{C}^2 est un \mathbb{C} -vectoriel avec donc $T_z\mathbb{C}^2 = \mathbb{C}^2$, J est globalement définie sur \mathbb{C}^2 par $J(z_1, z_2) = (iz_1, iz_2)$, i.e.

$$J(x_1, y_1, x_2, y_2) = (-y_1, x_1, -y_2, x_2) .$$

Dans \mathbb{C}^2 il y a les droites complexes (\mathbb{C} -droites)

$$\mathbb{C}z = \{\alpha(z_1, z_2) : \alpha = a + ib \neq 0 \in \mathbb{C}\}$$

qui sont des plans réels (\mathbb{R} -plans)

$$\{(ax_1 - by_1, bx_1 + ay_1, ax_2 - by_2, bx_2 + ay_2)\} .$$

Par ailleurs \mathbb{S}^3 est donnée par l'équation

$$F(z) = |z_1|^2 + |z_2|^2 = 1 ,$$

les plans tangents $T_x\mathbb{S}^3$ (x dénotant un z de \mathbb{C}^2 qui est sur \mathbb{S}^3) par

$$dF = 2(x_1dx_1 + y_1dy_1 + x_2dx_2 + y_2dy_2) = 2(r_1dr_1 + r_2dr_2) = 0$$

et la distribution de contact \mathcal{K}_0 canonique par

$$\omega_0 = x_1dy_1 - y_1dx_1 + x_2dy_2 - y_2dx_2 .$$

On notera que, comme en coordonnées polaires $xdy - ydx = r^2d\theta$, en coordonnées bipolaires on a

$$\omega_0 = r_1^2d\theta_1 + r_2^2d\theta_2 .$$

Géométriquement, \mathcal{K}_0 est le champ des plans orthogonaux aux fibres de la *fibration de Hopf* (découverte par Heinz Hopf en 1931) $H : \mathbb{S}^3 \rightarrow \mathbb{S}^2$ qui sont des cercles \mathbb{S}^1 plongés de façon un peu compliquée dans \mathbb{S}^3 . Il est évident que toute rotation $e^{i\theta}$ appliquée à la fois dans les deux plans z_1 et z_2 laisse \mathbb{S}^3 invariante. D'où une action naturelle du groupe \mathbb{S}^1 des $e^{i\theta}$ sur \mathbb{S}^3 dont les orbites sont des cercles. La fibration H est la restriction à \mathbb{S}^3 de l'application de Hopf

$$H : \mathbb{C}^2 \rightarrow \mathbb{R}^3 = \mathbb{R} \times \mathbb{C},$$

$$(z_1, z_2) \mapsto (|z_1|^2 - |z_2|^2, 2z_1\bar{z}_2) = (r_1^2 - r_2^2, 2z_1\bar{z}_2) .$$

En effet, si $(z_1, z_2) \in \mathbb{S}^3$,

$$|z_1|^2 + |z_2|^2 = r_1^2 + r_2^2 = 1$$

et le carré de la norme de $H(z_1, z_2)$ est égal à

$$(|z_1|^2 - |z_2|^2)^2 + 4|z_1\bar{z}_2|^2 = (|z_1|^2 + |z_2|^2)^2 = 1$$

et donc $\|H(z_1, z_2)\| = 1$ et $H(z_1, z_2) \in \mathbb{S}^2 \subset \mathbb{R}^3$. Par ailleurs, si

$$\left(|z_1|^2 - |z_2|^2, z_1 \bar{z}_2\right) = \left(|z'_1|^2 - |z'_2|^2, z'_1 \bar{z}'_2\right),$$

alors $|z_1|^2 + |z_2|^2 = |z'_1|^2 + |z'_2|^2 = 1$ et $|z_1|^2 - |z_2|^2 = |z'_1|^2 - |z'_2|^2$ impliquent $|z_1| = |z'_1|$ et $|z_2| = |z'_2|$ soit $z'_1 = e^{i\theta_1} z_1$ et $z'_2 = e^{i\theta_2} z_2$. Mais alors $z_1 \bar{z}_2 = z'_1 \bar{z}'_2$ implique

$$z_1 \bar{z}_2 = e^{i(\theta_1 - \theta_2)} z_1 \bar{z}_2$$

et donc $\theta_1 = \theta_2 \pmod{2\pi}$. (z_1, z_2) et (z'_1, z'_2) sont par conséquent dans la même orbite. Considérons alors une fibre $C = (e^{i\theta} z_1, e^{i\theta} z_2)$ de H avec $z_1 = r_1 e^{i\theta_1}$ et $z_2 = r_2 e^{i\theta_2}$ et calculons comme autrefois avec des infinitésimaux. En partant de (z_1, z_2) le long de C , $dz = (ir_1 e^{i\theta_1} d\theta, ir_2 e^{i\theta_2} d\theta)$. Par ailleurs un $d\xi$ quelconque à partir de (z_1, z_2) est de la forme

$$(ir_1 e^{i\theta_1} d\theta_1 + e^{i\theta_1} dr_1, ir_2 e^{i\theta_2} d\theta_2 + e^{i\theta_2} dr_2).$$

L'orthogonalité $d\xi \perp dz$ s'écrit $d\xi_1 d\bar{z}_1 + d\xi_2 d\bar{z}_2 = 0$. Mais

$$\begin{aligned} d\xi_1 d\bar{z}_1 + d\xi_2 d\bar{z}_2 &= (ir_1 e^{i\theta_1} d\theta_1 + e^{i\theta_1} dr_1) (-ir_1 e^{-i\theta_1} d\theta) \\ &\quad + (ir_2 e^{i\theta_2} d\theta_2 + e^{i\theta_2} dr_2) (-ir_2 e^{-i\theta_2} d\theta) \\ &= r_1^2 d\theta_1 d\theta + r_2^2 d\theta_2 d\theta - i(r_1 dr_1 d\theta + r_2 dr_2 d\theta). \end{aligned}$$

Comme $d\theta$ est la même pour les deux composantes dans l'action du groupe \mathbb{S}^1 on peut factoriser par $d\theta$ et l'éliminer de la condition d'orthogonalité. Et comme par ailleurs $r_1 dr_1 + r_2 dr_2 = 0$ car $d\xi$ est tangent à \mathbb{S}^3 , on obtient en définitive la condition

$$r_1^2 d\theta_1 + r_2^2 d\theta_2 = 0$$

qui n'est rien d'autre que $\omega_0 = 0$. Le plan orthogonal à C est donc bien le plan de contact.

On peut considérer les $T_x \mathbb{S}^3$ comme des sous-espaces affines de \mathbb{C}^2 dont le vecteur normal unitaire N_x est $N_x = (x_1, y_1, x_2, y_2)$. On note que $\omega_0(N_x) = 0$ puisque

$$\omega_0(Z) = x_1 Y_1 - y_1 X_1 + x_2 Y_2 - y_2 X_2$$

et comme

$$dF(JZ) = 2(-x_1 Y_1 + y_1 X_1 - x_2 Y_2 + y_2 X_2) = -2\omega_0(Z)$$

on voit que \mathcal{K}_0 est aussi le noyau de $dF \circ J = -2\omega_0$ ou encore l'intersection

$$T_x \mathbb{S}^3 \cap J(T_x \mathbb{S}^3)$$

de $T_x \mathbb{S}^3$ avec son conjugué, intersection qui est l'unique \mathbb{C} -droite affine (\mathbb{R} -plan J -invariant) incluse dans le $\mathbb{R}^3 = T_x \mathbb{S}^3$. Cette \mathbb{C} -droite est l'orthogonale en $z = x$ (pour le produit hermitien de \mathbb{C}^2) de la \mathbb{C} -droite vectorielle $\mathbb{C}z$. On vérifie immédiatement que ω_0 est bien une structure de contact. En effet

$$\begin{aligned}\omega_0 \wedge d\omega_0 &= 2(x_1 dy_1 - y_1 dx_1 + x_2 dy_2 - y_2 dx_2) \wedge (dx_1 \wedge dy_1 + dx_2 \wedge dy_2) \\ &= 2 \left(\begin{array}{l} -y_2 dx_1 \wedge dy_1 \wedge dx_2 + x_2 dx_1 \wedge dy_1 \wedge dy_2 - y_1 dx_1 \wedge dx_2 \wedge dy_2 \\ + x_1 dy_1 \wedge dx_2 \wedge dy_2 \end{array} \right)\end{aligned}$$

a pour noyau $\text{Span}\{N_x\}$ en tant que 3-forme sur \mathbb{R}^4 et comme N_x est la normale à $T_x\mathbb{S}^3$, $\omega_0 \wedge d\omega_0$ est une forme volume sur \mathbb{S}^3 .

Soit χ_0 le vecteur tangent

$$(-y_1, x_1, , -y_2, x_2) = Jz$$

en $z = x$. Trivialement $\chi_0 \in T_x\mathbb{S}^3$ puisque $dF(\chi_0) = 0$ (i.e. $\langle N_x, \chi_0 \rangle = 0$) et en plus $\omega_0(\chi_0) = 1$. χ_0 satisfait donc la première condition du champ de Reeb. Il satisfait aussi la seconde condition. En effet,

$$\begin{aligned}d\omega_0 &= 2(dx_1 \wedge dy_1 + dx_2 \wedge dy_2) \\ d\omega_0(Z, Z') &= 2 \left(\det \begin{pmatrix} dx_1(Z) & dy_1(Z) \\ dx_1(Z') & dy_1(Z') \end{pmatrix} + \det \begin{pmatrix} dx_2(Z) & dy_2(Z) \\ dx_2(Z') & dy_2(Z') \end{pmatrix} \right) \\ &= 2(X_1 Y_1' - X_1' Y_1 + X_2 Y_2' - X_2' Y_2) .\end{aligned}$$

Donc

$$i_{\chi_0} d\omega_0(Z') = d\omega_0(\chi_0, Z') = 2(x_1 Y_1' - y_1 X_1' + x_2 Y_2' - y_2 X_2')$$

mais si $Z' \in K_x$ on a

$$\omega_0(Z') = 0 = x_1 Y_1' - y_1 X_1' + x_2 Y_2' - y_2 X_2'$$

et donc $i_{\chi_0} d\omega_0 \equiv 0$ sur \mathcal{K} . Les orbites de χ_0 sont les fibres de la fibration H et sont donc toutes fermées.

On peut pousser plus loin la compréhension géométrique de cette structure en introduisant la structure du corps (non commutatif) \mathbb{H}_Q des *quaternions* sur $\mathbb{C}^2 = \mathbb{R}^4$.⁵² En plus de l'unité imaginaire i satisfaisant $i^2 = -1$, on introduit les unités j et k satisfaisant aussi $j^2 = -1$, $k^2 = -1$ avec la relation de commutation $ijk = -1$ (donc $ij = k$, $jk = i$, $ki = j$ et $ji = -ij$, $kj = -jk$, $ij = -ji$). On écrit alors

$$z = z_1 + z_2 j = x_1 + y_1 i + (x_2 + i y_2) j = x_1 + y_1 i + x_2 j + y_2 k.$$

52. Comme nous l'avons déjà noté au chapitre 2, section 8.1.3, un théorème fondamental (Frobenius) dit qu'il n'existe une structure de corps sur \mathbb{R}^n que pour $n = 1$ (nombres réels), $n = 2$ (nombres complexes), $n = 4$ (quaternions d'Hamilton) et $n = 8$ (octonions de Graves et Cayley). Cela est intimement lié au fait (théorème d'Adams) que leurs sphères unité sont les seules sphères parallélisables (i.e. de fibré tangent globalement trivialisable). En particulier il n'existe pas de structure de corps sur \mathbb{R}^3 et c'est l'une des raisons qui conduisit Hamilton à inventer les quaternions. \mathbb{S}^2 n'est pas parallélisable et tout champ de vecteurs tangents sur \mathbb{S}^2 s'annule quelque part (comme on dit, quand on peigne une sphère il y a toujours un épi).

Dire que $z \in \mathbb{S}^3$ signifie alors que z est un quaternion de norme 1 et comme zi , zj , zk sont orthogonaux à z , ils forment une base de $T_z\mathbb{S}^3$. En chaque point $z \in \mathbb{S}^3$, zi est le champ de Reeb et $\text{Span}\{zj, zk\}$ le plan de contact \mathcal{K}_z .

Remarque. On montre que \mathcal{K}_0 est isomorphe, dans le complémentaire de chaque point x de \mathbb{S}^3 , à la distribution de contact standard sur \mathbb{R}^3 . \square

4.7. Le théorème de Bennequin

Sur \mathbb{S}^3 et \mathbb{R}^3 on disposait au début des années 1980 de plusieurs résultats. Nous avons déjà cité les théorèmes de John Gray pour les variétés fermées et de Mikhaïl Gromov pour les variétés ouvertes. Un théorème de Robert Lutz (1971) disait que dans chaque classe d'équivalence d'homotopie de champs de plans (co-orientés) sur \mathbb{S}^3 il existe une structure de contact, ce qui implique l'existence d'une infinité de structures de contact non isomorphes (non isotopes).

Dans un contexte dominé par les travaux de grands spécialistes comme Mikhaïl Gromov, Yakov Eliashberg, Vladimir Arnold, William Thurston ou Alan Weinstein, le grand théorème du début des années 1980 est le théorème de Bennequin disant qu'il existe sur \mathbb{S}^3 (ou sur sa restriction à \mathbb{R}^3) une structure de contact globale \mathcal{K} homotope comme simple champ de plans à la structure standard \mathcal{K}_0 mais *non isomorphe* (non isotope) à \mathcal{K}_0 .⁵³ Dans le cas de \mathbb{S}^3 , le théorème de stabilité de Gray implique alors que \mathcal{K} n'est pas homotope à \mathcal{K}_0 dans la classe des champs de contact mais dans le cas de \mathbb{R}^3 où le théorème de Gray n'est pas valide, \mathcal{K} et \mathcal{K}_0 peuvent être homotopes comme champs de contact. Ce résultat très technique a été prouvé en 1982 par Daniel Bennequin dans sa thèse d'état [40] *Entrelacements et équations de Pfaff* (dont le directeur était Alain Chenciner), exposé dans [41] et a été le sujet d'un séminaire Bourbaki d'Adrien Douady [149]⁵⁴. Il est très bien illustré par Patrick Massot sur le site des *Images des mathématiques* du CNRS [357]. Il a fait l'objet depuis de plusieurs démonstrations différentes, ce qui montre son intérêt. Pour des prolongements, on pourra consulter par exemple le séminaire Bourbaki [316] de François Laudenbach sur la conjecture de Weinstein en dimension 3.

Ce théorème n'est pas trivial pour la raison suivante. Le fait d'être isomorphe pour des structures de contact \mathcal{K} est une condition plus faible que d'être isomorphes pour des formes de contact ω (globales) car ω détermine non seulement \mathcal{K} mais également le champ de Reeb χ_ω .

Une structure de contact non standard (\mathcal{K}, ω) sur \mathbb{R}^3 peut se construire facilement en s'arrangeant pour qu'il existe un disque plongé Δ tel que la trace de \mathcal{K} sur Δ , qui est un champ de directions dans Δ ayant pour trajectoires des courbes legendriennes, ait une courbe intégrale γ fermée (non réduite à un point). Le cercle γ découpe dans

53. Selon l'auteur de l'excellent item "Géométrie de contact" de Wikipedia, "cet énoncé est le théorème fondateur de la topologie de contact moderne".

54. Je me souviens encore très bien de ce beau moment.

Δ un sous-disque dont le bord (γ précisément) est tangent à \mathcal{K} . De tels disques sont dits “vrillés” (“overtwisted”). Or le théorème dit que de tels disques ne peuvent pas exister pour \mathcal{K}_0 .

Un exemple (dû à Erlandsson) de (\mathcal{K}, ω) sur \mathbb{R}^3 possédant un disque vrillé s’obtient facilement en prenant des coordonnées cylindriques (semi-polaires) $x = (r, \theta, z)$ et la 1-forme

$$\omega = \cos(r) dz + r \sin(r) d\theta .$$

Il s’agit bien d’une forme de contact car

$$d\omega = -\sin(r) dr \wedge dz + (r \cos(r) + \sin(r)) dr \wedge d\theta$$

et donc

$$\begin{aligned} \omega \wedge d\omega &= (\cos(r) dz + r \sin(r) d\theta) \wedge (-\sin(r) dr \wedge dz + (r \cos(r) + \sin(r)) dr \wedge d\theta) \\ &= \cos(r) dz \wedge (r \cos(r) + \sin(r)) dr \wedge d\theta + r \sin(r) d\theta \wedge (-\sin(r) dr \wedge dz) \\ &= \left(1 + \frac{\cos(r) \sin(r)}{r}\right) dr \wedge r d\theta \wedge dz = \left(1 + \frac{\sin(2r)}{2r}\right) dr \wedge r d\theta \wedge dz . \end{aligned}$$

Or $dr \wedge r d\theta \wedge dz$ est une forme volume et comme $1 + \frac{\sin(2r)}{2r} \geq 0.782 > 0$, $\omega \wedge d\omega \neq 0$ partout. Les rayons horizontaux sont dans \mathcal{K} puisqu’ils satisfont $d\theta = 0$ et $dz = 0$ et en chaque point $x_0 = (r_0, \theta_0, z_0)$ K_x est engendré par le rayon (r, θ_0, z_0) ($r \geq 0$) (i.e. le vecteur ∂_r) et le vecteur tangent en x_0 au cylindre vertical de rayon r_0 de pente $\frac{dz}{r_0 d\theta} = -\tan(r_0)$ (i.e. le vecteur $\cos(r_0) \partial_\theta - r_0 \sin(r_0) \partial_z$). Le vecteur normal N_{x_0} à K_x est porté par $r_0 \sin(r_0) \partial_\theta + \cos(r_0) \partial_z$ et quant au champ de Reeb χ il est donné en x_0 par

$$\chi_{x_0} = \frac{\sin(r_0)}{r_0 + \cos(r_0) \sin(r_0)} \partial_\theta + \frac{r_0 \cos(r_0) + \sin(r_0)}{r_0 + \cos(r_0) \sin(r_0)} \partial_z .$$

Pour $r_0 = k\pi$, $N_{x_0} = \chi_{x_0} = (-1)^k \partial_z$; pour $r_0 = (2k+1)\frac{\pi}{2}$, $N_{x_0} = (-1)^k r_0 \partial_\theta$ et $\chi_{x_0} = \frac{(-1)^k}{r_0} (\partial_\theta + \partial_z)$.

On voit donc comment les plans de contact K_x tournent le long des rayons horizontaux. Le long d’un rayon, chaque fois que $r = n\pi$ ($n \in \mathbb{N}$) K_x devient horizontal. Si donc on considère le disque horizontal Δ de hauteur $z = 0$ et de rayon $r = \pi$, \mathcal{K} est transverse à Δ sur l’intérieur de $\Delta - \{0\}$ mais se confond avec $T_x \Delta$ sur le bord $\partial \Delta$ et $\partial \Delta$ est donc tangent à \mathcal{K} . La figure 7 illustre ce phénomène. Toutefois ce résultat n’est pas satisfaisant car la situation est hautement dégénérée : pour $r = n\pi$ le plan de contact se confond avec le plan tangent de Δ et le bord $\partial \Delta$ est donc un cercle de points *critiques*. Mais on peut obtenir un cercle horizontal legendrien de \mathcal{K} sans points critiques en considérant par exemple $r = \pi$ sur le parabolôïde P d’équation $z = r^2$. En effet le plan de contact en $x = (r, \theta, z)$ est, nous venons de le voir,

$$\text{Span} \{ \partial_r, \cos(r) \partial_\theta - r \sin(r) \partial_z \} .$$

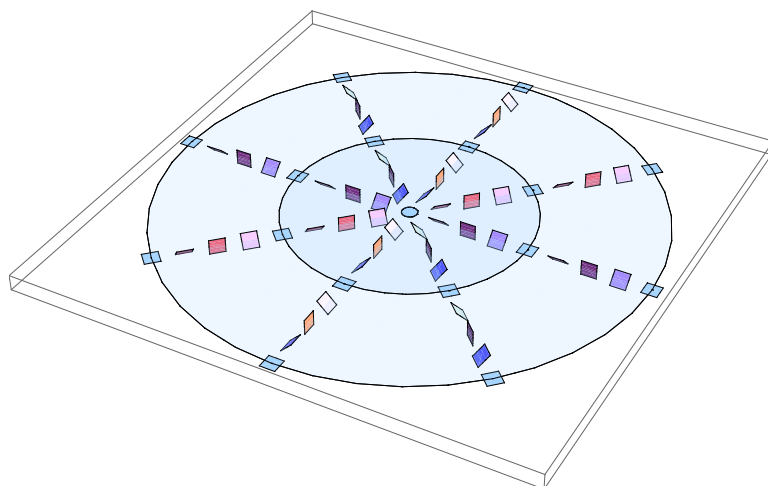


FIGURE 7. Disques vrillés de la forme de contact $\omega = \cos(r) dz + r \sin(r) d\theta$. Les plans de contact tournent le long des rayons horizontaux à partir du plan à l'origine qui est horizontal. Le long des cercles de rayon $r = n\pi$, ils redeviennent horizontaux et ces cercles sont donc des courbes legendriennes fermées. Les disques correspondants sont “vrillés”.

Quant au plan tangent $T_x P$ il est

$$\text{Span} \{ \partial_\theta, \partial_r + 2r\partial_z \} .$$

Pour $z = r$ et $r = \pi$, ces deux plans sont transverses et s'intersectent le long de la droite $\mathbb{R}\partial_\theta$, le cercle C de rayon π et de hauteur π^2 étant bien legendrien sans points critiques de P . On notera à propos de C (nous y reviendrons plus bas) que le long de C le champ de Reeb est $\chi = -\partial_z$ et que donc si l'on déforme C en la déplaçant un peu (de ε) dans la direction de χ qui est toujours transverse à \mathcal{K} , on obtient un cercle $C' = C + \varepsilon\chi$ translaté vers le bas de ε dont le nombre d'entrelacement avec C est nul. De même, toujours le long de C , la normale \tilde{N} à C dans \mathcal{K} (qui est horizontal) est la direction du rayon et donc si l'on déforme C en la déplaçant un peu (de ε) dans la direction de \tilde{N} , on obtient un cercle homothétique $C'' = C + \varepsilon\tilde{N}$ dont le nombre d'entrelacement avec C est également nul.

Or pour la distribution \mathcal{K}_0 sur \mathbb{S}^3 (ou sur sa restriction à \mathbb{R}^3) tous les disques plongés Δ sont non vrillés i.e. “tendus” (“tight”). Pour comparer \mathcal{K} à \mathcal{K}_0 sur \mathbb{R}^3 on peut remarquer que ω_0 est isomorphe à

$$\tilde{\omega} = d\tilde{z} + \tilde{r}^2 d\tilde{\theta}$$

car $\tilde{r}^2 d\tilde{\theta} = \tilde{x}d\tilde{y} - \tilde{y}d\tilde{x}$ et

$$\tilde{\omega} = d\tilde{z} + \tilde{x}d\tilde{y} - \tilde{y}d\tilde{x} = d(\tilde{z} - \tilde{x}\tilde{y}) + 2\tilde{x}d\tilde{y}$$

est la version symétrisée de $\omega_0 = dy - pdx$ (cf. le chapitre 5 et la section 2.2 du chapitre 2 où opère une petite variante de $\tilde{\omega}$), le difféomorphisme les reliant étant

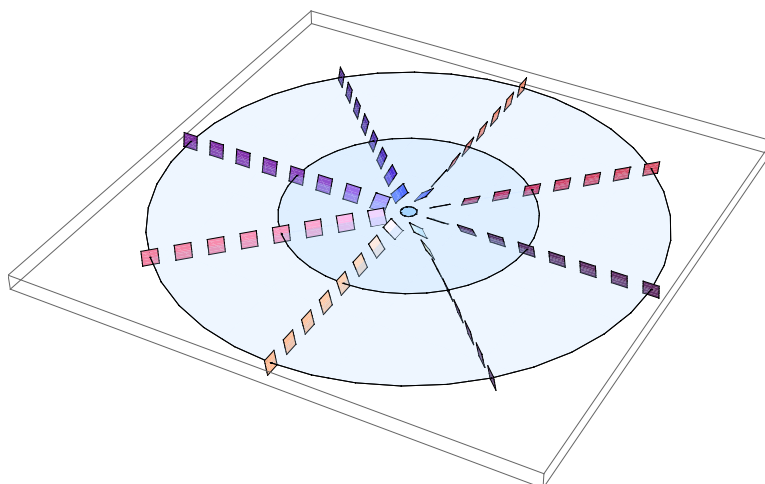


FIGURE 8. Pour la forme de contact $\omega_0 = dz + r^2 d\theta$, les plans de contact tournent le long des rayons horizontaux à partir du plan à l'origine qui est horizontal mais sans jamais redevenir horizontaux. À la limite $r = \infty$, ils deviennent seulement verticaux. Les disques plongés horizontaux sont donc tous "tendus".

$y = \tilde{z} - \tilde{x}\tilde{y}$, $x = \tilde{y}$ et $p = -2\tilde{x}$ (comme le jacobien est partout -2 , il s'agit bien d'un difféomorphisme). Là encore $\tilde{\mathcal{K}}$ est invariant par z -translations et les rayons horizontaux sont dans $\tilde{\mathcal{K}}$. En chaque point $\tilde{\mathcal{K}}$ est engendré par le rayon et le vecteur de pente $\frac{d\tilde{z}}{r d\theta} = -\tilde{r}$ et, en dehors de l'axe des \tilde{z} ($\tilde{r} = 0$), $\tilde{\mathcal{K}}$ n'est nulle part horizontal. Pour $r = \infty$, $\tilde{\mathcal{K}}$ devient vertical. Aucun des disques horizontaux Δ ne peut donc être vrillé. La figure 8 illustre ce phénomène.

On peut facilement généraliser cet exemple à $M = \mathbb{R}^2 \times \mathbb{S}^1$ en remplaçant z par une coordonnée angulaire φ . La forme de contact est alors

$$\omega_1 = d\varphi + r^2 d\theta = d\varphi + xdy - ydx .$$

C'est une compactification des fibres du groupe de Heisenberg non polarisé.⁵⁵

Le théorème de Bennequin dit que, aussi "tordu" soit-il, *aucun* disque plongé ne peut être vrillé pour la structure de contact standard $\omega_0 = dz + r^2 d\theta$ en coordonnées semi-polaires. La démonstration initiale repose sur l'intrication de résultats fins de théorie des nœuds \mathfrak{n} , des entrelacs \mathfrak{e} et des tresses (fermées) \mathfrak{t} dans \mathbb{R}^3 avec ω_0 . Les entrelacs sont constitués de plusieurs nœuds entrelacés et les tresses fermées sont des entrelacs présentés d'une façon adaptée aux coordonnées : on s'arrange pour que les composantes connexes (orientées) de \mathfrak{e} tournent autour de l'axe des z avec un angle θ strictement croissant. Le nombre de tours n s'appelle aussi le nombre de brins. Un nœud (i.e. un entrelacs connexe) a en général plusieurs brins. En projetant un

55. C'est l'analogue de $SE(2) = \mathbb{V}_S$, avec sa 1-forme $\omega_S = -\sin(\theta) dx + \cos(\theta) dy$, qui est une compactification des fibres du groupe de Heisenberg polarisé. \mathbb{V}_J .

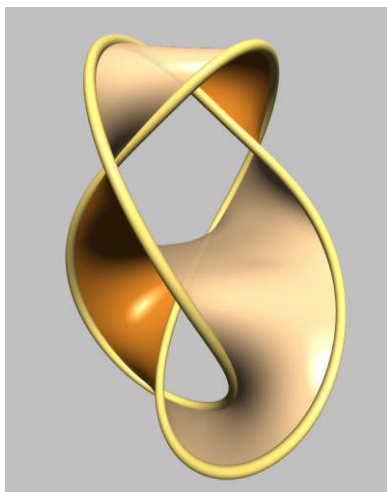


FIGURE 9. Un exemple de surface de Seifert pour le nœud de 8 (image produite par le logiciel *SeifertView* de Jack van Wijk).

nœud sur le plan de base (r, θ) ⁵⁶ et en comptant algébriquement les croisements de la projection, on obtient un nombre $c(\mathbf{n})$.

On peut à partir d'un nœud \mathbf{n} construire des “surfaces de Seifert”, c'est-à-dire des surfaces S (plongées, orientées⁵⁷, compactes, connexes, sans auto-intersection) de bord $\partial S = \mathbf{n}$ (il en existe pour tous les nœuds⁵⁸). La figure 9 en fournit un exemple pour le nœud de 8.

Une telle surface permet de définir un repère mobile R_S positivement orienté le long de \mathbf{n} : si $\nu \in \mathbf{n}$, on prend le vecteur tangent T_ν à \mathbf{n} en ν , le vecteur N_ν normal à T_ν mais tangent à S et pointant vers l'intérieur de S et un troisième vecteur V_ν complétant le repère direct. Si \mathbf{n}_ℓ est un nœud *legendrien*, on peut par ailleurs définir des champs de vecteurs X le long de \mathbf{n}_ℓ qui sont transverses à \mathcal{K} (par exemple le champ de Reeb χ). Comme $T_\nu \in \mathcal{K}_\nu$ on peut considérer la normale $\tilde{N}_\nu \perp T_\nu$ de \mathbf{n}_ℓ dans \mathcal{K}_ν et définir ainsi des repères mobiles $(T_\nu, \tilde{N}_\nu, X_\nu)$ le long de \mathbf{n}_ℓ . Pour mesurer l'écart entre ces deux types repères mobiles, on peut alors déformer \mathbf{n}_ℓ en \mathbf{n}' au moyen de X et regarder comment \mathbf{n}' intersecte S . Génériquement, cela permet

56. En projetant un nœud sur le plan (x, y) pour la structure de contact standard polarisée $dy - pdx$ on obtient des courbes fermées possédant en général des cusps là où \mathbf{n} a une tangente verticale (i.e. parallèle à l'axe des p).

57. Cette contrainte d'orientabilité est très importante. Par exemple la plus simple des surfaces ayant pour bord le nœud de trèfle est un ruban de Möbius, mais ce n'est pas une surface de Seifert car il n'est pas orientable.

58. Il en existe même beaucoup car étant donnée une surface de Seifert S et une surface orientée fermée F quelconque, on peut enlever à S et à F de petits disques Δ_S et Δ_F et recoller F et S le long de leurs bords respectifs $\partial\Delta_S$ et $\partial\Delta_F$.

de définir un nombre d'intersection $B(\mathbf{n}_\ell)$ dit "nombre de Thurston-Bennequin". Il mesure la façon dont les \mathcal{K}_ν tournent relativement à S le long de \mathbf{n}_ℓ car $B(\mathbf{n}_\ell)$ mesure la rotation de \tilde{N}_ν relativement au repère (T_ν, N_ν, V_ν) lorsque ν se déplace le long de \mathbf{n}_ℓ . C'est le nombre d'entrelacement de \mathbf{n}_ℓ et $\mathbf{n}_\ell + \varepsilon\tilde{N}$ (ε assez petit) et aussi celui de $\mathbf{n}_\ell + \varepsilon\chi$.

On définit un autre nombre de rotation et d'entrelacement (index de Maslov) de la façon suivante. À $\nu \in \mathbf{n}_\ell$, on associe la tangente $T_\nu \in \mathcal{K}_\nu$ ($T_\nu \in \mathcal{K}_\nu$ car \mathbf{n}_ℓ est legendrien) et, comme $\mathcal{K}_\nu \simeq \mathbb{R}^2$ et $T_\nu \neq 0$, on obtient une application $\tau : \mathbf{n}_\ell \rightarrow \mathbb{R}^2 - \{0\}$. Soit $r(\mathbf{n}_\ell)$ le nombre de fois que l'image $\tau(\mathbf{n}_\ell)$ tourne autour de 0.

Si maintenant \mathbf{n}_τ est un nœud (orienté) transverse à \mathcal{K} , $\omega_0(T_\nu) \neq 0$ le long de \mathbf{n}_τ et \mathbf{n}_τ est dit ascendant (resp. descendant) si $\omega_0(T_\nu) > 0$ (resp. < 0). Si Z est un champ $\neq 0$ dans \mathcal{K} le long de \mathbf{n}_τ , on peut déformer \mathbf{n}_τ en \mathbf{n}' au moyen de Z et calculer le nombre d'entrelacement $L(\mathbf{n}_\tau)$ de \mathbf{n}_τ et \mathbf{n}' . Il est indépendant de Z .⁵⁹

On montre le résultat suivant établissant un lien entre les deux cas : si \mathbf{n}_ℓ est legendrien, $\mathbf{n}_\tau = \mathbf{n}_\ell \pm \varepsilon\tilde{N}$ est transverse, ascendant pour $-$ et descendant pour $+$, et $L(\mathbf{n}_\tau) = B(\mathbf{n}_\ell) \pm r(\mathbf{n}_\ell)$. On démontre également que si \mathbf{n}_τ est ascendant $L(\mathbf{n}_\tau) = c(\mathbf{n}_\tau) - n$.

Le théorème de Bennequin relie ces nombres à la *caractéristique d'Euler-Poincaré* maximale $\chi_{EP}(\mathbf{n})$ des surfaces de Seifert de \mathbf{n} (donc $\chi_{EP}(S) \leq \chi_{EP}(\mathbf{n})$). Pour en comprendre l'énoncé on a besoin de quelques concepts tout à fait basiques de topologie algébrique. Quand on considère *topologiquement* les surfaces fermées (orientables, compactes, connexes, sans bord) Σ de \mathbb{R}^3 , elles sont toutes homéomorphes à des sphères à g anses, autrement dit à des recollements d'une sphère avec g tores. Elles sont donc caractérisées par ce nombre g , qu'on appelle leur *genre*, et leur caractéristique d'Euler-Poincaré $\chi_{EP}(\Sigma)$, définie homologiquement par la formule d'Euler $S - A + F$ ($S = \#$ sommets, $A = \#$ arêtes, $F = \#$ faces) pour n'importe laquelle de leurs triangulations, satisfait la formule fondamentale

$$2 - \chi_{EP}(\Sigma) = 2g(\Sigma) .$$

Si S est une surface de Seifert de \mathbf{n} , elle a pour bord \mathbf{n} qui est topologiquement un cercle. En recollant un disque à S le long de \mathbf{n} , on obtient une surface fermée de genre g que l'on appelle le genre de S . On a alors

$$2g(S) = 1 - \chi_{EP}(S) .$$

Comme nous l'avons vu, on peut augmenter à volonté le genre de S et c'est donc le genre minimal $g(\mathbf{n})$ (i.e. la $\chi_{EP}(\mathbf{n})$ maximale) qui contient de l'information. $g(\mathbf{n})$

59. De façon générale, l'existence d'un tel champ Z dans \mathcal{K} partout non nul signifie que \mathcal{K} est globalement trivialisable (on dit alors que ω est parallélisable). C'est une contrainte sur M . Elle est satisfaite ici.

s'appelle le genre de \mathbf{n} et l'on a

$$\chi_{EP}(\mathbf{n}) = 1 - 2g(\mathbf{n}) .$$

Théorème (Bennequin, 1982). Pour tout nœud transverse \mathbf{n}_τ ascendant de la structure de contact standard de \mathbb{R}^3 on a $L(\mathbf{n}_\tau) = c(\mathbf{n}_\tau) - n \leq -\chi_{EP}(\mathbf{n}_\tau)$ (i.e. $\leq -\chi_{EP}(S)$ pour toute surface de Seifert S de \mathbf{n}). \diamond

Il implique immédiatement

$$|c(\mathbf{n}_\tau)| - n \leq -\chi_{EP}(\mathbf{n}_\tau)$$

car avec une symétrie on peut changer le signe de c sans changer ni n ni $\chi_{EP}(\mathbf{n}_\tau)$ et l'on a donc également

$$-c(\mathbf{n}_\tau) - n \leq -\chi_{EP}(\mathbf{n}_\tau) .$$

En particulier si \mathbf{n}_τ n'est pas noué alors $\chi_{EP}(\mathbf{n}_\tau) = 1$ et $c(\mathbf{n}_\tau) \leq n - 1 < n$. Si la projection de \mathbf{n}_τ sur le plan de base est un simple cercle on a $n = 1$ et $c = 0$.

Le corollaire le plus important concerne les nœuds legendriens.

Théorème. Pour tout nœud legendrien \mathbf{n}_ℓ de la structure de contact standard de \mathbb{R}^3 on a $B(\mathbf{n}_\ell) \leq -\chi_{EP}(\mathbf{n}_\ell) - r(\mathbf{n}_\ell)$. \diamond

Pour la preuve, on applique la formule ci-dessus

$$B(\mathbf{n}_\ell) \pm r(\mathbf{n}_\ell) = L(\mathbf{n}_\tau) \leq -\chi_{EP}(\mathbf{n}_\tau)$$

en remarquant que, \mathbf{n}_τ étant une ε -déformation de \mathbf{n}_ℓ , $\chi_{EP}(\mathbf{n}_\tau) = \chi_{EP}(\mathbf{n}_\ell)$. Mais si \mathbf{n}_ℓ est non noué, on a $\chi_{EP}(\mathbf{n}_\ell) = 1$ et donc $B(\mathbf{n}_\ell) < 0$. Supposons alors qu'il existe dans $(\mathbb{R}^3, \omega_0, \mathcal{K}_0)$ un disque plongé Δ tel que \mathcal{K}_0 y induise un champ de droites possédant une trajectoire fermée γ (non réduite à un point critique). Cette orbite périodique γ serait alors un nœud legendrien non noué. Mais quand on circule le long de γ , le plan de contact \mathcal{K} ne peut pas accomplir un tour sinon il y aurait un point x où il serait tangent à Δ et x serait un point critique de Δ . Comme $B(\gamma)$ est le nombre d'entrelacement entre γ et $\gamma + \varepsilon\tilde{N}$ (\tilde{N} la normale directe de γ dans \mathcal{K}_0), on a $B(\gamma) = 0$, ce qui contredit le théorème.

Bref, sur la variété *ouverte* \mathbb{R}^3 il existe des structures de contact homotopes non isotopes.

Après ces résultats clés, de nombreuses recherches ont été développées dans de nombreuses directions. Beaucoup de problèmes (petits ou gros) apparaissent tout de suite.

Par exemple, dans sa version "Heisenberg polarisé" $(\mathbb{R}^3, \mathcal{K}_J, \omega_J = dy - pdx)$, la structure de contact standard est celle de notre espace \mathbb{V}_J des 1-jets. Lorsque l'on passe à l'espace $\mathbb{R}^2 \times \mathbb{P}^1 = \mathbb{P}(T\mathbb{R}^2)$ de tous les éléments de contact, on projectivise $T\mathbb{R}^2$, on compactifie les fibres de \mathbb{V}_J et ω_J devient

$$\omega = \cos(\varphi) dy - \sin(\varphi) dx \quad (\varphi \bmod \pi) .$$

On peut alors chercher à prolonger la théorie que nous venons d'évoquer. Mais le champ \mathcal{K}_S n'est plus co-orientable à cause des fibres projectives \mathbb{P}^1 et il faut en tenir compte. Comme ∂_φ fait partie de \mathcal{K} partout, toutes les fibres \mathbb{P}^1 sont des cercles legendriens (mais ne sont pas des relevées legendriennes de courbes dans le plan de base (x, y)).

En prenant le revêtement d'ordre 2 de $\mathbb{R}^2 \times \mathbb{P}^1$ (éléments de contact orientés) on retrouve notre modèle $\mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1 = SE(2)$ qui est orientable et co-orientable avec

$$\omega_S = \cos(\varphi) dy - \sin(\varphi) dx \pmod{2\pi}.$$

Mais $\mathbb{R}^2 \times \mathbb{S}^1$ n'est pas simplement connexe et on peut considérer son revêtement universel qui, lui, est simplement connexe et $\simeq \mathbb{R}^3$ car la variable $\tilde{\varphi}$ relevant la variable angulaire φ varie maintenant dans \mathbb{R} . En relevant ω_S , on obtient

$$\tilde{\omega}_S = \cos(\tilde{\varphi}) dy - \sin(\tilde{\varphi}) dx$$

en coordonnées cartésiennes et

$$\tilde{\omega}_S = \sin(\theta - \tilde{\varphi}) dr + r \cos(\theta - \tilde{\varphi}) d\theta$$

en coordonnées semi-polaires, ce qui redonne ω_J au moyen du changement de variables

$$\tilde{\varphi} = \tilde{\varphi}, \quad x = -(z \sin(\tilde{\varphi}) + p \cos(\tilde{\varphi})), \quad y = z \cos(\tilde{\varphi}) - p \sin(\tilde{\varphi})$$

de jacobien -1 .⁶⁰ En effet, $\tilde{\omega}_S$ devient $dz - pd\tilde{\varphi}$ qui est bien ω_J avec d'autres notations.

Citons quelques résultats. D'après des résultats de Yakov Eliashberg, la seule structure de contact tendue (i.e. sans disques vrillés) sur \mathbb{S}^3 est la structure standard à isotopie près et sur une variété fermée de dimension 3 il n'existe qu'un nombre fini de classes d'homotopie (pour l'homotopie de champs de plans quelconques) de structures de contact tendues. Dans [165], Eliashberg donne une classification complète des structures de contact "vrillées" ("overtwisted", i.e. contenant un disque vrillé) sur les M^3 (orientées, connexes). Citons aussi sa classique synthèse [166] (1992) de l'état des travaux depuis Martinet, de Lutz, Bennequin, Giroux et lui-même sur les variétés de contact 3D. Citons encore le Séminaire Bourbaki [43] de Bennequin consacré à ses travaux.

En 2006, Clifford Taubes a démontré au moyen de la théorie de jauge la conjecture de Weinstein disant que, pour toute variété de contact orientée fermée de dimension 3, le champ de Reeb a une trajectoire fermée (conjecture démontrée pour \mathbb{S}^3 en 1993 par Hofer). C'est tout un nouveau domaine qui s'est développé avec l'homologie de contact, l'homologie de Floer et la théorie symplectique des champs.

60. Le difféomorphisme inverse est $z = y \cos(\tilde{\varphi}) - x \sin(\tilde{\varphi})$, $p = -y \sin(\tilde{\varphi}) + x \cos(\tilde{\varphi})$.

4.8. La méthode des surfaces convexes (Giroux)

Nous voudrions dire un mot de la méthode des surfaces convexes utilisée par Emmanuel Giroux et qui permet de redémontrer les théorèmes de Bennequin et Eliashberg. Comme nous l'avons rappelé plus haut en renvoyant à notre exemple de la section 2.9 du chapitre 3, si l'on se donne une surface S (on suppose S régulière et *fermée* i.e. compacte sans bord) dans une variété de contact (M, \mathcal{K}) de dimension 3 (on suppose M et \mathcal{K} orientées), \mathcal{K} induit un champ de lignes "caractéristique" (et donc un feuilletage de dimension 1) sur S , $TS \cap \mathcal{K} = \mathcal{K}_S$, dont les points critiques sont les points $x \in S$ tels que $T_x S = K_x$. On peut alors étudier \mathcal{K} au voisinage de S et la façon dont \mathcal{K}_S varie lorsqu'on déforme S . Il s'agit donc d'une analyse *semi-locale* intermédiaire entre la structure locale de \mathcal{K} qui est toujours isomorphe à la structure de contact standard (théorème de Darboux) et la structure globale qu'il s'agit de comprendre. On pourra se référer au texte de référence d'Emmanuel Giroux [214] et au mémoire de 2013 [358] de Patrick Massot qui introduit aux méthodes topologiques de Giroux avec de nombreuses belles illustrations.

Cette analyse semi-locale fait intervenir, pour déformer S , la notion de champ de contact C que nous avons défini plus haut dans la section 2.3. Les champs de contact sont définis par le fait qu'ils laissent la distribution \mathcal{K} globalement invariante, ce qui est équivalent à la propriété

$$\mathcal{L}_C \omega = \iota_C d\omega + d(\iota_C \omega) = \lambda \omega$$

où λ est une fonction sur M . Le champ de contact C est alors presque partout transverse à \mathcal{K} en dehors des domaines où il est nul. L'exemple prototypique est le champ de Reeb χ pour lequel $\lambda \equiv 0$ puisque $\iota_\chi d\omega = 0$ et $\iota_\chi \omega \equiv 1$ ce qui implique $d(\iota_C \omega) = 0$.

Un champ de contact C peut être considéré comme une section du fibré de rang 1 TM/\mathcal{K} . Si $\mathcal{K} = \ker(\omega)$, le champ de Reeb χ trivialisait TM et tout champ peut écrire $X = X_{\mathcal{K}} + f\chi$. Le fait d'être de contact détermine alors $X_{\mathcal{K}}$. La fonction f est l'équivalent dans le cas contact de ce qu'est un hamiltonien dans le cas symplectique. La surface de niveau $f = 0$ définit la surface *caractéristique* Σ_C de C qui est le lieu où $C \in \mathcal{K}$ i.e. $\iota_C \omega = \omega(C) = 0$.⁶¹ Σ_C peut être vide comme dans le cas du champ de Reeb χ pour lequel $f \equiv 1$. *Génériquement*, f est une fonction de Morse et $\Sigma_C = f^{-1}(0)$ est régulière. Qui plus est C est tangent à Σ_C . En effet le flot de C préserve évidemment C mais aussi \mathcal{K} puisque C est de contact. Donc le flot

61. Dans l'exemple de \mathbb{V}_J traité plus haut à la section 2.3 on a

$$C = -(\partial_p f) \partial_x + (f - p \partial_p f) \partial_y + (\partial_x f + p \partial_y f) \partial_p$$

et donc là où $f = 0$, C est donné par

$$C = -(\partial_p f) \partial_x - (p \partial_p f) \partial_y + (\partial_x f + p \partial_y f) \partial_p .$$

préserve la relation $C \in \mathcal{K}$ et donc Σ_C . Maintenant, en tant que surface, Σ_C possède son propre feuilletage caractéristique \mathcal{K}_{Σ_C} et C (tangent à Σ_C par hypothèse de généralité) est supporté par \mathcal{K}_{Σ_C} puisque $C \in T\Sigma_C \cap \mathcal{K}$.⁶² Si $x \in M$ est un point singulier (i.e. un point où le champ C s'annule, $C_x = 0$), alors évidemment $x \in \Sigma_C$ puisque $\omega(C) = 0$ en x et $T_x\Sigma_C = K_x$ autrement dit Σ_C est tangente à \mathcal{K} . Qui plus est, la valeur de λ en x est la demi-divergence $\text{div}(C)$ de C . En effet, ω étant une forme de contact $\omega \wedge d\omega$ est une forme volume dV et, par définition de la divergence relativement à cette forme, $\mathcal{L}_C(dV) = \text{div}(C) dV$. Mais, C étant de contact,

$$\mathcal{L}_C\omega = \iota_C d\omega + d(\iota_C\omega) = \lambda\omega$$

et

$$\mathcal{L}_C d\omega = d(\mathcal{L}_C\omega) = d\lambda \wedge \omega + \lambda d\omega .$$

Donc

$$\begin{aligned} \mathcal{L}_C(dV) &= \mathcal{L}_C(\omega \wedge d\omega) = (\mathcal{L}_C\omega) \wedge d\omega + \omega \wedge \mathcal{L}_C d\omega \\ &= \lambda\omega \wedge d\omega + \omega \wedge (d\lambda \wedge \omega + \lambda d\omega) = 2\lambda\omega \wedge d\omega \\ &= 2\lambda dV = \text{div}(C) dV . \end{aligned}$$

On a ainsi une géométrie bien précise d'un champ de contact C avec sa surface caractéristique Σ_C et le feuilletage caractéristique de sa surface caractéristique \mathcal{K}_{Σ_C} lorsque cette dernière est non vide. On va utiliser un tel champ pour déformer la surface S et son champ caractéristique \mathcal{K}_S . D'où une gymnastique géométrique assez subtile. On peut éviter une trop grande complexité en imposant la contrainte de généralité que S (fermée) soit \mathcal{K} -convexe.

La notion de \mathcal{K} -convexité d'une surface S plongée dans (M, \mathcal{K}) est, rappelons-le, qu'il existe un champ de contact C *transverse* à S . Elle fait interférer ce champ C avec le champ caractéristique \mathcal{K}_S de S . Le flot de C sur un intervalle de temps assez petit définit alors une déformation S_t de $S = S_0$ qui est un voisinage tubulaire de S sur lequel \mathcal{K} est t -invariant. On montre, et ce n'est pas du tout évident, que les surfaces \mathcal{K} -convexes fermées sont génériques relativement aux surfaces fermées.

Plus précisément, pour étudier de façon générale les liens entre les déformations de S et la structure de contact \mathcal{K} , on considère d'abord la structure des champs de vecteurs X sur S qui sont portés par le champ caractéristique \mathcal{K}_S . On considère une 2-forme d'aire sur S . Notons-la α . Par dualité relativement à α , les X correspondent à des 1-formes $\beta = \iota_X\alpha$ sur S .⁶³ Soit S_t , un voisinage tubulaire de S avec ses X_t et

62. La notion de "caractéristique" est omniprésente et polysémique dans ce genre de contexte et peut entraîner des confusions de structure si l'on n'y prend garde.

63. Si Y est un champ sur S , $\beta(Y) = \iota_X\alpha(Y) = \alpha(X, Y)$. et donc $\beta(X) = \iota_X\alpha(X) = \alpha(X, X) = 0$.

α_t . Dans ce voisinage une 1-forme de contact ω associée à \mathcal{K} peut alors s'écrire

$$\omega = u_t dt + \beta_t .$$

Il est naturel d'essayer de trouver des coordonnées adaptées et de simplifier la situation.

On considère d'abord la variété $M = S \times \mathbb{R}$ avec $\omega = u_t dt + \beta_t$ de contact. On montre qu'une adaptation du théorème de stabilité de Gray (utilisant le fait que les S_t sont compactes même si $S \times \mathbb{R}$ n'est pas fermée) implique que si deux structures de contact \mathcal{K} et \mathcal{K}' induisent sur S des champs caractéristiques égaux $\mathcal{K}_{S_t} = \mathcal{K}'_{S_t}$ pour tout $t \in \mathbb{R}$ alors \mathcal{K} et \mathcal{K}' sont isotopes. En effet, soient $\omega = u_t dt + \beta_t$ et $\omega' = u'_t dt + \beta'_t$ des 1-formes de contact définissant \mathcal{K} et \mathcal{K}' . Comme $\mathcal{K}_{S_t} = \mathcal{K}'_{S_t}$ on a $\beta'_t = f(t) \beta_t$ avec $f(t) > 0$ puisque les formes de contact (ici sur S) ne sont définies qu'à un facteur > 0 près. On peut donc remplacer ω' par

$$\omega'' = \frac{\omega'}{f(t)} = \frac{u'_t}{f(t)} dt + \beta_t = u''_t dt + \beta_t .$$

Mais alors on peut linéairement interpoler entre ω et ω'' au moyen de

$$\omega_s = ((1-s)u_t + su'_t) dt + \beta_t$$

ce qui permet d'interpoler par \mathcal{K}_s entre \mathcal{K} et \mathcal{K}' . Le théorème de Gray adapté dit alors que cette déformation vient d'une isotopie de M .

En localisant ce résultat au voisinage de S dans M , on montre de la même façon que si $\mathcal{K}_S = \mathcal{K}'_S$ alors \mathcal{K} et \mathcal{K}' sont isotopes dans un voisinage de S .

Si maintenant S est \mathcal{K} -convexe, et en prenant pour S_t le flot d'un champ de contact C transverse à S , la description ci dessus devient t -invariante. On peut donc écrire, au voisinage de S , $\omega = u dt + \beta$ avec la fonction u et la 1-forme

$$\beta = \beta_x dx + \beta_y dy$$

indépendantes de t . La structure reste très riche puisque u et β peuvent être quelconques du moment que ω est de contact, c'est-à-dire que la 2-forme sur S $-ud\beta + \beta \wedge du$ est partout non nulle (cf. plus bas).

En termes de coordonnées locales (x, y, t) où (x, y) sont des coordonnées sur S , u et β ne dépendent que de (x, y) . On peut immédiatement vérifier à la main (c'est toujours un bon exercice) que ∂_t (transverse à S par construction) est un champ de contact de fonction génératrice $\omega(\partial_t) = u$. En effet

$$\mathcal{L}_{\partial_t} = \iota_{\partial_t} d\omega + d(\iota_{\partial_t} \omega) = 0 .$$

Le second terme est du . Quant au premier terme on a, puisque ∂_t est de composantes 0 par rapport à ∂_x et ∂_y ,

$$d\omega(\partial_t, Y) = (du \wedge dt + d\beta)(\partial_t, Y) = du \wedge dt(\partial_t, Y) = -du(Y)$$

$$\text{i.e. } \iota_{\partial_t} d\omega = -du$$

et les deux termes s'éliminent. La surface caractéristique Σ_{∂_t} de ∂_t est le lieu d'annulation de sa fonction génératrice, i.e. $u = 0$. C'est bien le lieu où $\partial_t \in \mathcal{K}$ puisqu'alors $\omega(\partial_t) = u = 0$.

On peut calculer le champ de Reeb $\chi = \chi_x \partial_x + \chi_y \partial_y + \chi_t \partial_t$. Considérons les deux 2-formes sur S ,

$$\begin{aligned} d\beta &= (\partial_x \beta_y - \partial_y \beta_x) dx \wedge dy = A dx \wedge dy \\ \beta \wedge du &= (\beta_x \partial_y u - \beta_y \partial_x u) dx \wedge dy = B dx \wedge dy . \end{aligned}$$

Comme ω est une 1-forme de contact, $\omega \wedge d\omega$ est une forme volume partout non nulle. Mais

$$\omega \wedge d\omega = -ud\beta \wedge dt + \beta \wedge du \wedge dt = (-uA + B) dx \wedge dy \wedge dt$$

et donc $(-uA + B)$ est partout non nul. En écrivant que $\iota_\chi \omega = \omega(\chi) \equiv 1$ et $\iota_\chi d\omega(Y) = 0$ pour tout Y (et donc que les coefficients des composantes de Y doivent être nuls) on trouve

$$\begin{aligned} u\chi_t + \beta_x \chi_x + \beta_y \chi_y &= 1 \\ -A\chi_y - (\partial_x u)\chi_t &= 0 \\ A\chi_x - (\partial_y u)\chi_t &= 0 \\ (\partial_x u)\chi_x + (\partial_y u)\chi_y &= 0 \end{aligned}$$

d'où on déduit

$$\chi_x = \frac{\partial_y u}{uA + B}, \quad \chi_y = -\frac{\partial_x u}{uA + B}, \quad \chi_t = \frac{A}{uA + B} .$$

On remarquera que la condition $(\partial_x u)\chi_x + (\partial_y u)\chi_y = 0$ signifie $du(\chi) = 0$. On remarquera aussi que $\chi = \partial_t$ lorsque u est constante et égale à 1 (ce qui implique $B = 0$ puisque $du = 0$). La surface caractéristique Σ_{∂_t} de ∂_t , $u = 0$, est alors vide. C'est le cas de notre modèle \mathbb{V}_J où $\omega_J = dy - pdx = udt + \beta$ avec β la 1-forme $-pdx$ et $\chi = \partial_y$. La surface S est le plan symplectique (x, p) muni de sa 2-forme canonique $dx \wedge dy$. À l'autre extrême χ est tangent à S (i.e. $\chi_t \equiv 0$) si $A = 0$, autrement dit si $\beta = dh$ est une 1-forme exacte et $\omega = udt + dh$. B est alors le crochet de Poisson $\{h, u\}$.

Ce modèle est donc valable au voisinage de S si S est \mathcal{K} -convexe. Comme.

$$\omega \wedge d\omega = (-ud\beta + \beta \wedge du) \wedge dt$$

est une forme volume, on a $-ud\beta + \beta \wedge du > 0$. Soit X le champ sur S dual de β défini par $\iota_X \alpha = \beta$ où α est la 2-forme d'aire sur S . On a

$$\beta(X) = \iota_X \alpha(X) = \alpha(X, X) = 0$$

et X est donc porté par le feuilletage caractéristique \mathcal{K}_S de S quelle que soit la 2-forme α . En termes de X la condition $-ud\beta + \beta \wedge du > 0$ est équivalente à

$$u \operatorname{div}_\alpha(X) + du(X) < 0.$$

La surface caractéristique Σ_{∂_t} du champ de contact ∂_t d'équation $u = 0$ est régulière et transverse à S . Soit Σ sa trace sur S , c'est-à-dire la ligne de niveau de $u = 0$ sur S . Σ divise S en domaines S^\pm dans lesquels u est $\gtrless 0$. Le long de Σ on a $du(X) < 0$. Mais *en dehors* de Σ on a $u \neq 0$ et on peut remplacer $\omega = udt + \beta$ par $\tilde{\omega} = \frac{\omega}{u} = dt + \frac{\beta}{u} = dt + \tilde{\beta}$. La condition devient dans ce cas $\operatorname{div}(X) < 0$. Mais la divergence dépend de la forme d'aire choisie sur S . Or en modifiant cette forme (et d'abord en la divisant par $|u|$) on peut s'arranger pour que $\operatorname{div}(X)$ devienne $\gtrless 0$ sur S^\pm et s'annule sur Σ , X étant transversal à Σ et pointant de S^+ vers S^- . Bref avec une forme d'aire convenable on peut construire un champ X sur S \mathcal{K} -convexe qui est porté par le feuilletage caractéristique \mathcal{K}_S de S , transversal à Σ et dilatant sur S^+ et contractant sur S^- . Giroux appelle cette situation une *division* de la surface S \mathcal{K} -convexe. Elle caractérise la situation car les feuilletages \mathcal{F} de S de même Σ (i.e. tels qu'il existe une 2-forme d'aire α sur S et un champ X de support \mathcal{F} avec $X \neq 0$ sur $S - \Sigma$, X transversal à Σ et pointant de S^+ vers S^-) sont les feuilletages caractéristiques de structures de contact \mathcal{K}' isotopes à \mathcal{K} . C'est le théorème dit de "flexibilité" de Giroux. Il permet de démontrer (cf. Massot [358]) la généricité de la \mathcal{K} -convexité en utilisant les théorèmes généraux de généricité (Poincaré-Bendixson, Morse-Smale, Peixoto) sur les feuilletages de surfaces.

Avec cette géométrie sophistiquée de la \mathcal{K} -convexité Giroux a réussi à définir un critère pour qu'une structure de contact \mathcal{K} soit tendue, ce qui permet de nouvelles démonstrations des résultats de Bennequin et d'Eliashberg. Ce critère dit que une surface S divisée par Σ a un voisinage tendu si et seulement si, soit

- (i) aucune composante de Σ ne borde un disque dans S , soit
- (ii) S est une sphère et Σ n'a qu'une composante (une courbe régulière fermée γ).

On utilise alors des méthodes de "chirurgie" consistant à déformer une situation initiale Sit_0 générique et bien contrôlée en une autre situation Sit_1 , elle aussi générique et bien contrôlée, au moyen d'une famille Sit_t $t \in [0, 1]$ générique en tant famille et ne présentant donc que des non-généricités bien contrôlées à la traversée de valeurs critiques isolées du paramètre t .⁶⁴ On en tire l'inégalité de Bennequin-Eliashberg, le théorème de Bennequin disant que les structures de contact standard de \mathbb{R}^3 et \mathbb{S}^3 sont tendues ainsi que le théorème d'Eliashberg disant que toutes les structures de contact tendues sur \mathbb{R}^3 ou \mathbb{S}^3 sont isomorphes à la structure de contact standard.

64. Pour les fonctions réelles, le prototype est celui du point d'inflexion éliminant le minimum et le maximum $x = \pm\sqrt{-\frac{t}{3}}$ de $f_t(x) = x^3 + tx$, $t < 0$, en traversant le point d'inflexion $f_0(x) = x^3$.

Une référence historique particulièrement intéressante est la synthèse d'Eliashberg lui-même à l'époque de son théorème (1992) [166].

4.9. Prolongements

Nous arrêterons là cette petite incursion dans le vaste domaine de la géométrie globale des structures de contact. Beaucoup de prolongements ont été explorés. On a par exemple analysé en dimension > 3 les structures de contact “exotiques” sur \mathbb{R}^{2n+1} et les structures symplectiques “exotiques” sur \mathbb{R}^{2n} . Les techniques de théorie des nœuds n'étant plus utilisables il a fallu trouver des méthodes alternatives.

Par ailleurs, tout ce que nous avons évoqué ne concerne que les structures de contact qui globalisent le modèle local $dy - pdx$ des 1-jets. On imagine que les problèmes de globalisation deviennent encore plus compliqués et profonds lorsque l'on s'attaque aux jets d'ordre supérieur en commençant par l'ordre 2, c'est-à-dire aux structures d'Engel, autrement dit aux champs de plans différentiables \mathcal{E} sur une variété de dimension 4 qui sont maximalelement non intégrables et satisfont $\text{rang}[\mathcal{E}, \mathcal{E}] = 3$, $\text{rang}[\mathcal{E}, [\mathcal{E}, \mathcal{E}]] = 4$. Le lecteur désirant se documenter à ce sujet pourra par exemple consulter la thèse [555] de Thomas Vogel où se trouve le théorème que toute variété de dimension 4 parallélisable (i.e. dont le fibré tangent est trivialisable) admet une structure d'Engel orientable (c'est en fait équivalent).

Partie III : Analyse harmonique non commutative et diffusion

Mini *vademecum* d'analyse fonctionnelle

1. Géométrie et analyse fonctionnelle en analyse harmonique

Dans les chapitres précédents, nous avons longuement analysé la géométrie immanente (interne) de l'architecture fonctionnelle de $V1$. Il s'agissait de géométrie pure, la géométrie sous-riemannienne permettant de modéliser la connectivité de l'aire $V1$. Mais nous savons que les neurones de $V1$ opèrent comme des filtres sur le signal optique et que la géométrie de $V1$ doit par conséquent être intimement couplée avec le traitement du signal. Nous voudrions donc faire le lien entre, d'un côté, la théorie des groupes et la géométrie et, de l'autre côté, la théorie du filtrage et l'analyse fonctionnelle.

Remarque terminologique. Nous rencontrons ici encore une fois une homonymie conceptuelle. En neuroscience, une architecture fonctionnelle concerne le design de la connectivité et “fonctionnel” y renvoie à “fonction” au sens de la dialectique entre structure et fonction en biologie. En mathématiques, “fonctionnel” renvoie à des espaces de fonctions, “fonction” signifiant une application $x \mapsto f(x)$ entre deux espaces. □

Une telle synthèse relève de l'*analyse harmonique* y compris de l'analyse harmonique *non commutative* qui approfondit ce que nous avons déjà exposé dans le Vol I à la section 3.4. du chapitre 3 à propos des ondelettes.

Les images rétinienne sont des fonctions $I(x, y)$ bruitées, très irrégulières (des *distributions* au sens de Schwartz, cf. la section 3.3.4. du Vol I) qui sont traitées par des filtres.¹ Le traitement le plus classique est l'analyse de Fourier. Mais nous avons vu au Vol I avec David Marr et Stéphane Mallat que les champs récepteurs des neurones visuels effectuent plutôt une analyse en ondelettes (une analyse de

1. Comme nous l'avons rappelé section 1.1 au tout début du premier chapitre 2, la mesure en (x_0, y_0) d'un signal $I(x, y)$ par un profil récepteur $\varphi_{(x_0, y_0)}(x, y) = \varphi(x - x_0, y - y_0)$ se fait par l'intégrale $\langle \varphi, I \rangle = \int \varphi_{(x_0, y_0)}(x, y) I(x, y) dx dy$ ce qui revient, en termes de distributions, à traiter les profils récepteurs comme des “fonctions test” neuralelement câblées et les signaux comme des distributions pouvant être assez “sauvages”. Jan Koenderink et Luc Florack avaient déjà souligné cette analogie. Dans ce cas ce sont les fonctions test qui opèrent sur les signaux alors que, en général, on considère que ce sont les distributions qui agissent sur les fonctions test. Nous reviendrons sur les distributions à la section 5.

Fourier spatialement localisée et multi-échelle) permettant de détecter des traits géométriques locaux comme les discontinuités. Les profils récepteurs sont localisés à la fois en espace et en fréquence et possèdent des orientations privilégiées, des échelles et éventuellement d'autres propriétés leur permettant d'être de bons détecteurs. Mais les activités des neurones se propagent le long des connexions horizontales, ce qui, géométriquement, correspond à un transport le long des géodésiques de la géométrie sous-riemannienne modélisant cette connectivité horizontale. Dans les théories classiques de l'analyse harmonique le processus de diffusion régi par les géodésiques d'une métrique revient à étudier les solutions de l'opérateur différentiel laplacien et le noyau de la chaleur associé. Il nous faudra donc étudier les laplaciens sous-riemanniens et leurs noyaux de la chaleur.

Tout cela conduit à l'étude des représentations des groupes \mathbb{V}_J et \mathbb{V}_S décrivant l'architecture fonctionnelle (au sens biologique) de $V1$ dans des espaces fonctionnels (au sens mathématique) appropriés. Nous plongeons ainsi le traitement des signaux optiques rétiniens (des images visuelles) dans l'immense univers mathématique de l'analyse harmonique (y compris non commutative). De même que la géométrie sous-riemannienne, le calcul variationnel et la théorie du contrôle nous ont connectés à une tradition plutôt "mécaniste" qui remonte à Newton, Leibniz, les Bernouilli, Euler et leurs successeurs jusqu'à aujourd'hui, de même l'analyse harmonique non commutative va nous connecter à une tradition plutôt "quantique" qui remonte à Heisenberg, Schrödinger, Dirac, von Neumann et Gabor.

Les processus de diffusion sont régis par des équations de la chaleur qui, comme les équations des ondes et les équations de Laplace elles aussi associées au laplacien, sont des équations aux dérivées partielles. Les équations aux dérivées partielles à coefficients *constants* dans les espaces euclidiens \mathbb{R}^n sont invariantes par translations et l'on connaît l'importance pour les résoudre de la transformée de Fourier – c'est-à-dire des représentations unitaires irréductibles des groupes additifs \mathbb{R}^n . Dans nos modèles neurogéométriques, les groupes de Lie qui interviennent crucialement sont des groupes nilpotents comme le groupe de Heisenberg \mathbb{H} ou des groupes résolubles comme $SE(2)$, les grandes équations aux dérivées partielles associées n'étant plus à coefficients constants et l'analyse harmonique devenant non commutative. C'est dans cet immense univers, très technique et en grande partie ouvert dans les années 1920-1930 par la mécanique quantique, que nous plongeons donc maintenant notre effort de modélisation.

Mathématiquement parlant, cet univers est plus récent. Après des travaux pionniers de spécialistes comme Jacques Dixmier dans les années 1950, il se déploie dans les années 1960 et surtout 1970 avec les travaux² d'éminents spécialistes comme Richard Beals, Larry Corwin, Michel Duflo, Gerald Folland, Bernard Gaveau, Peter Greiner, Bernard Helffer, Lars Hörmander, Alexandre Kirillov, Jean Nourrigat,

2. Nous en avons déjà croisé plusieurs.

Charles Rockland, Linda Rothschild, Elias Stein, et beaucoup d'autres. À la fin des années 1970 la théorie était déjà bien installée et, dans un séminaire Bourbaki de 1981 consacré aux équations aux dérivées partielles sur les groupes nilpotents [366], Guy Métivier pouvait affirmer :

“On peut dire que l'idée d'utiliser les opérateurs invariants sur un groupe nilpotent comme modèle dans un certain nombre de problèmes, est maintenant devenue une idée standard.”

L'analyse harmonique relève de la théorie des représentations de groupes dans des espaces fonctionnels. Avant de l'aborder, nous allons d'abord consacrer un chapitre au rappel de quelques notions de bases d'analyse fonctionnelle qui sont “common knowledge” et que l'on trouve dans tous les manuels élémentaires et dans les articles de Wikipedia. Il s'agira d'un mini *vademecum*.

Pour plus de détails, le lecteur intéressé pourra consulter, parmi tant d'autres, le traité classique d'analyse fonctionnelle de Walter Rudin [473].

2. Mesure et intégration

Dans les applications que nous avons en vue, deux problématiques très différentes se trouveront conjuguées. D'une part l'action d'opérateurs différentiels sur des fonctions représentant certaines entités neurophysiologiques, ce qui supposera que les fonctions possèdent les propriétés de différentiabilité qu'exige l'applicabilité de l'opérateur considéré. D'autre part des évaluations numériques de la taille des fonctions (par exemple l'énergie d'un signal) et de leurs différences. Ces “mesures” (au sens naïf) seront globales et définies par certaines intégrales sur le domaine des fonctions. Mais les intégrales ne sont définies qu'à des sous-ensembles de mesure nulle près et donc les valeurs exactes des fonctions en chaque point de leur domaine n'interviendront pas en général puisque les points seront en général de mesure nulle. La dérivabilité n'a alors plus guère de sens. Il faut donc être soigneux lorsqu'on veut conjuguer les deux problématiques.

Commençons par la notion technique de mesure.

2.1. Espaces mesurables

La théorie moderne de la mesure, tenant compte des très subtiles difficultés introduites par la complexité des sous-ensembles d'un ensemble dans le cadre de la théorie des ensembles fondée et approfondie par Georg Cantor à partir de 1874, a été élaborée à partir du tournant du XIX^e-XX^e siècle, en particulier par Émile Borel, Henri Lebesgue, Felix Hausdorff, Constantin Carathéodory, Maurice Fréchet, René Baire, puis plus tard Andreï Kolmogorov pour les mesures de probabilité.

Si X est un ensemble, une mesure ν sur X consiste à sélectionner une classe \mathfrak{M} de sous-ensembles S de X que l'on appellera “mesurables” et à leur attribuer une

mesure (une “taille”) $\nu(S) \in [0, +\infty]$. Pour que l'on aboutisse à une bonne théorie, il faut supposer d'abord que l'ensemble vide \emptyset et l'ensemble total X appartiennent à \mathfrak{M} et que \mathfrak{M} soit stable par complémentation, \cup et \cap *dénombrables* : on dit alors que \mathfrak{M} est une σ -algèbre (ou une “tribu”) de l'algèbre booléenne $\mathcal{P}(X)$ des parties de X . La plus grossière σ -algèbre est évidemment $\mathfrak{M} = \{\emptyset; X\}$ et la plus fine (dite “discrète”) est $\mathfrak{M} = \mathcal{P}(X)$. Si $x \in X$ on peut aussi considérer

$$\mathfrak{M}_x = \{S \subset X | x \in S\} \cup \emptyset.$$

L'atome de x , $\mathfrak{a}(x)$ est

$$\mathfrak{a}(x) = \bigcap_{S \in \mathfrak{M}_x, x \in S} S$$

et les atomes partitionnent X et les S mesurables. Si X est infini, on a aussi la σ -algèbre des S soit finis ou dénombrables soit de complémentaires finis ou dénombrables. Une intersection de σ -algèbres est une σ -algèbre et, étant donnée une classe \mathfrak{B} de parties de X , il existe une plus petite σ -algèbre qui les contient, à savoir l'intersection de toutes les σ -algèbres contenant \mathfrak{B} . On dit que c'est la σ -algèbre engendrée par \mathfrak{B} .

Étant donnée une σ -algèbre \mathfrak{M} sur X , une mesure définie sur (X, \mathfrak{M}) est alors une fonction $\nu : \mathfrak{M} \rightarrow [0, +\infty]$ telle que $\nu(\emptyset) = 0$ et qui est σ -additive, c'est-à-dire telle que

$$\nu\left(\bigsqcup_n S_n\right) = \sum_n \nu(S_n)$$

pour toute union dénombrable de $S_n \in \mathfrak{M}$ *disjoints* entre eux (ce qu'indique le symbole \sqcup d'union disjointe). En particulier si $S \subset S'$ alors $S' = S \sqcup (S' - S)$, donc

$$\nu(S') = \nu(S) + \nu(S' - S)$$

et par conséquent $\nu(S) \leq \nu(S')$. Les S tels que $\nu(S) = 0$ sont dits ν -négligeables et le *support* de ν est le Σ_ν de \mathfrak{M} tel que $\nu(S) = \nu(S \cap \Sigma_\nu)$ pour tout les S de \mathfrak{M} .

Lorsque X est un espace topologique, la classe \mathfrak{M} naturelle est celle \mathfrak{B} des *boréliens* c'est-à-dire des \cup et \cap dénombrables d'ouverts et de fermés.³ Elle est engendrée par la classe des ouverts. Dans le cas des espaces \mathbb{R}^n les ouverts sont engendrés par les parallélépipèdes P de côtés c_i , $i = 1, \dots, n$ de longueur ℓ_i et ceux-ci ont un *volume* $\beta(P) = \prod \ell_i$ que l'on prend pour mesure. La mesure de Borel $\beta(S)$ est alors définie par σ -additivité pour les $S \in \mathfrak{B}$. Elle est par construction invariante par isométrie.

3. Une \cup dénombrable d'ouverts est un ouvert mais une \cap dénombrable d'ouverts n'est pas en général un ouvert. Par exemple dans \mathbb{R} , l' \cap dénombrable des intervalles ouverts $]-\frac{1}{n}, \frac{1}{n}[$ pour n entier ≥ 1 est le point 0 qui est fermé. De même une \cap dénombrable de fermés est fermée mais pas forcément une \cup dénombrable de fermés. Par exemple l' \cup dénombrable des fermés $[\frac{1}{n}, 1 - \frac{1}{n}]$ pour n entier ≥ 1 est l'ouvert $]0, 1[$.

La mesure de Lebesgue (que nous avons utilisée jusqu'ici sans entrer dans les détails) est plus raffinée que la mesure de Borel : elle en est la *complétée* au sens suivant. Nous avons vu que si ν est une mesure, un ensemble ν -mesurable N est dit “négligeable” si $\nu(N) = 0$. Par définition, les N négligeables sont stables par unions dénombrables. Pour la mesure de Borel il existe beaucoup de N β -négligeables : par exemple tous les sous-ensembles de volume = 0, les points, les droites, les courbes, les \cup et \cap dénombrables compliquées de tels objets. La σ -algèbre de Lebesgue \mathfrak{L} complétée de \mathfrak{B} consiste à adjoindre *tous* les sous-ensembles des N β -négligeables et à leur attribuer la mesure 0. Elle contient donc tous les sous-ensembles de \mathbb{R}^n ou \mathbb{C}^n qui sont boréliens à un N β -négligeable près.

Cela ne complexifie pas beaucoup la théorie conceptuellement mais la complexifie en revanche énormément sur le plan technique de la théorie des ensembles. En effet il existe beaucoup de sous-ensembles d'un ensemble infini et on arrive très vite à des paradoxes apparents. Il n'est pas problématique de prouver que tout ensemble dense mais dénombrable de \mathbb{R} (comme l'ensemble des rationnels \mathbb{Q}) est de mesure de Lebesgue = 0 puisque les points sont de \mathfrak{L} -mesure nulle. Il est en revanche plus problématique de prouver qu'il existe des sous-ensembles de \mathbb{R} comme l'ensemble triadique de Cantor qui sont de \mathfrak{L} -mesure nulle tout en ayant la puissance du continu. Cette découverte a été considérée au début de la théorie comme un véritable paradoxe et il a fallu de profondes réflexions pour en comprendre la nature.⁴ Il s'est même posé la question de savoir si la σ -algèbre de Lebesgue \mathfrak{L} n'est pas tout simplement la σ -algèbre maximale $\mathcal{P}(X)$. La réponse est loin d'être évidente puisque qu'il faut en fait utiliser l'axiome du choix pour démontrer l'existence d'un sous-ensemble non \mathfrak{L} -mesurable de \mathbb{R} .

L'excès de \mathfrak{L} par rapport à \mathfrak{B} est évident du point de vue des cardinaux car \mathfrak{B} a la puissance du continu (i.e. le cardinal de \mathbb{R}) alors que \mathfrak{L} a la puissance de $\mathcal{P}(\mathbb{R})$ puisqu'elle contient tous les sous-ensembles de l'ensemble de Cantor qui a la puissance de \mathbb{R} .

On dit que la mesure ν est *finie* si $\nu(X)$ est fini et *σ -finie* si X est recouvrable par une famille dénombrable de S_j mesurables de mesure finie.

Remarque. Dans son texte célèbre de 1906 sur les “Séries trigonométriques et séries de Taylor” [179], Pierre Fatou rappelle l'histoire de la notion de mesure de Cantor à Lebesgue. Il explique que c'est Borel qui a mis en place la première définition rigoureuse et que

4. La théorie des sous-ensembles de \mathbb{R} au-delà des boréliens (dite “théorie descriptive” des ensembles) est extrêmement technique et a constitué l'un des principaux moteurs du développement moderne de la théorie des ensembles après la découverte par Paul Cohen de la méthode du forcing lui ayant permis de démontrer en 1963 l'indépendance de l'hypothèse du continu de Cantor par rapport à la théorie des ensembles ZFC (Zermelo-Fraenkel avec axiome du choix), résultat qui lui valut la médaille Fields en 1966. Le lecteur intéressé pourra consulter le magnifique traité de Patrick Dehornoy [140], ainsi que notre étude [437] et sa bibliographie.

“le problème ainsi posé est susceptible d’une solution unique, sinon pour tous les ensembles que l’on peut concevoir comme existant, du moins pour tous ceux que l’on a pu effectivement nommer”.

Il souligne que les définitions de Cantor et de Borel coïncident pour les sous-ensembles fermés mais ne coïncident pas pour les sous-ensembles dénombrables denses non fermés. Et à propos de la notion de fonction intégrable au sens de Lebesgue, il explique que cette notion

“s’applique à toutes les fonctions discontinues que l’on peut nommer (par exemple à toutes les fonctions représentables analytiquement), au moins quand ces fonctions sont bornées”.

On voit que l’on est encore loin de la théorie des ensembles à la Zermelo-Fraenkel qui donnera un sens précis à “ce que l’on peut effectivement nommer”, mais en ouvrant la boîte de Pandore de difficultés insondables. ⁵ _____ □

Remarque. La théorie des probabilités utilise des mesures telles que $\nu(X) = 1$ qui sont donc finies et a fortiori σ -finies. _____ □

Remarque. On peut généraliser la notion de mesure à des mesures prenant leurs valeurs dans des espaces beaucoup plus généraux que $[0, \infty]$. Nous en rencontrerons plus bas un exemple avec les mesures ayant pour valeurs des projecteurs dans un espace de Hilbert. _____ □

2.2. Intégration

Si (X, \mathfrak{M}, ν) est un espace mesurable, on peut alors définir l’intégrale

$$\int_X f(x) d\nu(x)$$

de certaines fonctions sur X dites *mesurables*, le symbole $d\nu(x)$ exprimant conventionnellement que ν fonctionne comme mesure d’intégration de fonctions sur X . ⁶

On commence par la définition évidente des intégrales pour les fonctions caractéristiques $\chi_S(x)$ des $S \in \mathfrak{M}$:

$$\int_X \chi_S(x) d\nu(x) = \nu(S) .$$

On passe ensuite à des combinaisons linéaires de χ_{S_j} et à des limites en utilisant la propriété de σ -additivité. Si X est un espace topologique, les fonctions \mathfrak{B} -mesurables n’ont pas de raison d’être continues et peuvent être topologiquement assez “sauvages”.

5. Cf. [140] et [437].

6. Le symbole $d\nu(x)$ rappelle trois choses : la mesure ν , la variable x des fonctions sur X et la nature “infinitésimale” de l’élément de mesure en x .

La mesure de Lebesgue sur \mathbb{R} et plus généralement sur \mathbb{R}^n , permet de définir l'intégrale de Lebesgue qui est une vaste généralisation de l'intégrale de Riemann. Pour définir l'intégrale $J = \int_a^b f(x) dx$ d'une fonction $f(x)$ sur l'intervalle $[a, b]$, Riemann considérait d'abord des fonctions en escalier f_E définies par des valeurs f_i sur une décomposition finie de l'intervalle $[a, b]$ en une suite d'intervalles successifs $[a = x_1, x_2], [x_2, x_3], \dots, [x_{n-1}, x_n = b]$ et il posait

$$J_E = \int_a^b f_E(x) dx = \sum_{i=1}^{i=n-1} f_i .$$

Il considérait ensuite pour une fonction $f(x)$ l'inf J_+ (resp. le sup J_-) des J_E pour les approximations par le haut (resp. par le bas) de f par des $f_E(x) \geq f(x)$ (resp. $f_E(x) \leq f(x)$) sur $[a, b]$ de "marches" de plus en plus petites. Il disait que f est intégrable si $J_+ = J_-$ et définissait J par cette valeur commune.

L'intégrale de Lebesgue généralise ces techniques de limite d'approximations mais il y a beaucoup plus de fonctions \mathfrak{L} -mesurables puisque si f est mesurable au sens de Riemann et si h est égale à f presque partout (noté $h = f$ p.p. ou $h =_{p.p.} f$), i.e. sauf sur un ensemble négligeable, alors $\int h = \int f$. Par exemple, si $f = 0$ sauf sur un ensemble de mesure nulle, son intégrale sera nulle. Et, répétons-le, il y a beaucoup d'ensemble de mesure nulle puisque leur classe est stable non seulement par \cup et \cap dénombrables mais aussi par *inclusion*. La classe d'équivalence de f constituée des $h = f$ p.p. ne dit donc rien de précis sur les valeurs exactes $f(x)$ de f . Comme nous l'avons déjà noté, il s'agit là de l'une des difficultés intuitives de la compréhension des espaces fonctionnels définis par de telles classes d'équivalence. Nous allons en évoquer quelques exemples des plus classiques. Mais signalons d'abord un important théorème de comparaison.

2.3. Théorème de Radon-Nykodym-Lebesgue

Nous aurons plus bas à utiliser un résultat fondamental permettant de comparer entre elles des mesures. Soit (X, \mathfrak{M}, ν) un espace (X, \mathfrak{M}) muni d'une mesure ν et ρ une autre mesure sur (X, \mathfrak{M}) . On dit que ρ est "absolument continue" par rapport à ν si $\nu(N) = 0$ implique $\rho(N) = 0$, autrement dit si les ν -négligeables sont aussi ρ -négligeables. Dans ce cas, ρ a une "dérivée" de Radon-Nykodym h par rapport à ν , notée $h = \frac{d\rho}{d\nu}$ qui est une fonction ν -mesurable telle que, pour tout $S \in \mathfrak{M}$,

$$\rho(S) = \int_X \chi_S(x) d\rho(x) = \int_S h(x) d\nu(x) = \int_X \chi_S(x) h(x) d\nu(x)$$

(χ_S étant la fonction caractéristique de S). On dit par ailleurs que ρ et ν sont "étrangères" si leurs supports Σ_ρ et Σ_ν sont disjoints. Le théorème de Radon-Nykodym-Lebesgue dit alors que pour les mesures ν et ρ σ -finies il existe deux mesures ρ' et ρ'' telles que

- (i) $\rho = \rho' + \rho''$,
- (ii) ρ' est absolument continue par rapport à ν , et
- (iii) ρ'' est étrangère à ν .

Pour des mesures finies ν et ρ , ce résultat se démontre en considérant les f ν -mesurables telles que $f d\nu \leq d\rho$, en prenant leur sup, soit h , en montrant que h est ν -mesurable et que, si $\rho' = h d\nu$ est la partie absolument continue de ρ , alors $\rho'' = \rho - \rho'$ est étrangère à ν . On étend ensuite facilement aux mesures σ -finies. Mais on ne peut pas étendre aux mesures quelconques. Le contre-exemple classique est celui de la relation sur les boréliens de \mathbb{R} entre la mesure de Lebesgue $\rho = dx$ et la mesure de comptage $\nu(S) = \#S$ si S est fini ($\#$ symbolise le cardinal) et $\nu(S) = +\infty$ sinon. La mesure ν n'est pas σ -finie car il existe des boréliens qui ne sont pas des unions dénombrables de sous-ensembles finis. Par ailleurs, ρ est absolument continue par rapport à ν car si $\nu(S) = 0$ le sous-ensemble S est vide et donc $\rho(S) = 0$. Et pourtant il n'existe pas de dérivée h telle que $\rho(S) = \int_S h(x) d\nu(x)$ pour tout borélien S car sinon pour tout $S = \{x\}$ on aurait $\rho(\{x\}) = h(x) = 0$ et h serait nulle et donc aussi la mesure de Lebesgue. Dans le cas général où ν et ρ ne sont pas σ -finies, il faut ajouter une troisième mesure ρ''' constituée de masses ponctuelles (des mesures de Dirac).

3. Espaces $L^p(X, \nu)$

Dès que l'on dispose d'un espace X muni d'une mesure ν sur une σ -algèbre \mathfrak{M} , on peut mesurer la "taille globale" d'une fonction $f(x)$ à valeurs dans \mathbb{R} ou \mathbb{C} par des intégrales bien choisies. Comme il faut que la nullité des intégrales implique que f soit nulle presque partout, il faut prendre des intégrales de fonctions ≥ 0 associées à f par exemple des valeurs absolues de puissances de f . Les plus simples de ces espaces sont les espaces \mathbb{R} - ou \mathbb{C} -vectoriels $L^p(X, \nu)$ définis pour $p > 0$ (éventuellement $p = +\infty$) comme les espaces des f mesurables en quelque sorte "de taille finie pour la puissance p ", c'est-à-dire telles que

$$\int_X |f(x)|^p d\nu(x) < +\infty .$$

Pour $p = +\infty$, l'espace $L^\infty(X, \nu)$ est l'espace des f mesurables essentiellement bornées i.e. bornées en dehors d'un sous-ensemble négligeable.⁷

Remarque. Il faut insister encore une fois sur le fait que ces intégrales globales ne dépendent que de la classe d'équivalence des f pour l'équivalence qu'est l'égalité presque partout (i.e. à un ensemble de mesure nulle près) et que les espaces $L^p(X, \nu)$ sont donc des espaces fonctionnels quotientés par cette relation d'équivalence. La

7. Le L de L^p est l'initiale de Lebesgue et c'est Stefan Banach, grand admirateur d'Henri Lebesgue, qui les baptisa ainsi.

valeur exacte de f en un point n'y est donc plus en général une notion pertinente.
□

Il s'agit bien d'espaces vectoriels. Ils sont évidemment stables par multiplications scalaires $f \rightarrow \lambda f$, mais ils sont aussi stables par somme. En effet, comme

$$|f(x) + h(x)| \leq |f(x)| + |h(x)|$$

on a évidemment

$$|f(x) + h(x)| \leq 2 \max(|f(x)|, |h(x)|) .$$

Comme la fonction a^p est croissante pour les $a \geq 0$, concave si $0 < p < 1$ et convexe si $p \geq 1$, on a donc

$$|f(x) + h(x)|^p \leq 2^p \max(|f(x)|^p, |h(x)|^p)$$

et comme $\max(a, b) \leq a + b$ si $a, b \geq 0$, on a

$$|f(x) + h(x)|^p \leq 2^p (|f(x)|^p + |h(x)|^p)$$

Donc

$$\int_X |f(x) + h(x)|^p d\nu(x) < +\infty$$

si $\int_X |f(x)|^p d\nu(x) < +\infty$ et $\int_X |h(x)|^p d\nu(x) < +\infty$ et

$$f + h \in L^p$$

si $f, h \in L^p$.

On définit alors une *quasi-norme* $\|f\|_p$ par

$$\|f\|_p = \left(\int_X |f(x)|^p d\nu(x) \right)^{\frac{1}{p}}$$

et

$$\|f\|_\infty = \text{Supess } |f(x)|$$

où Supess est la borne supérieure "essentielle", i.e. adaptée à l'équivalence $=_{p.p.}$. On vérifie les propriétés d'une quasi-norme :

(i) $\|\lambda f\|_p = (|\lambda|^p)^{\frac{1}{p}} \|f\|_p = |\lambda| \|f\|_p$ (homogénéité),

(ii) $\|f\|_p = 0$ équivaut à $f =_{p.p.} 0$ et donc à $f = 0$ dans L^p puisque les f de L^p sont définis à $=_{p.p.}$ près.

Les espaces L^p sont des espaces *métriques* avec la distance

$$d_p(f, h) = \|f - h\|_p^p$$

pour $0 < p < 1$ et

$$d_p(f, h) = \|f - h\|_p$$

pour $p \geq 1$. On a évidemment

$$d_p(f, h) = 0 \iff f = h \text{ et } d_p(f, h) = d_p(h, f)$$

mais on a aussi l'inégalité triangulaire

$$d_p(f, k) \leq d_p(f, h) + d_p(h, k)$$

$$\int_X |f(x) - k(x)|^p d\nu(x) \leq \int_X |f(x) - h(x)|^p d\nu(x) + \int_X |h(x) - k(x)|^p d\nu(x) .$$

Pour $p < 1$ cela découle du fait que, si $0 < a < b$, alors $(a+b)^p < a^p + b^p$ car la fonction $\varphi(u) = u^p$ est strictement croissante et concave et les fonctions concaves sont sous-additives. En effet, considérons les points $O = (0, \varphi(0) = 0)$, $A = (a, \varphi(a) = a^p)$, $B = (b, \varphi(b) = b^p)$, $C = (a+b, \varphi(a+b) = (a+b)^p)$ puis les cordes OA , AB , BC de pentes respectives

$$\frac{\varphi(a) - \varphi(0)}{a - 0} = \frac{\varphi(a)}{a}, \quad \frac{\varphi(b) - \varphi(a)}{b - a}, \quad \frac{\varphi(a+b) - \varphi(b)}{(a+b) - b} = \frac{\varphi(a+b) - \varphi(b)}{a} .$$

Comme $\varphi(u)$ est C^∞ (en fait continue dérivable suffirait), d'après le théorème classique et basique de (Michel) Rolle (1691), il existe $\alpha \in]0, a[$ tel que $\varphi'(\alpha) = \frac{\varphi(a)}{a}$, $\beta \in]a, b[$ tel que $\varphi'(\beta) = \frac{\varphi(b) - \varphi(a)}{b - a}$, $\gamma \in]b, a+b[$ tel que $\varphi'(\gamma) = \frac{\varphi(a+b) - \varphi(b)}{a}$. On a $\alpha < \beta < \gamma$. Or $\varphi(u)$ étant strictement concave, sa dérivée $\varphi'(u)$ est strictement décroissante et l'on a donc

$$\frac{\varphi(a)}{a} > \frac{\varphi(b) - \varphi(a)}{b - a} > \frac{\varphi(a+b) - \varphi(b)}{a} ,$$

a fortiori

$$\frac{\varphi(a)}{a} > \frac{\varphi(a+b) - \varphi(b)}{a} ,$$

soit

$$\varphi(a+b) < \varphi(a) + \varphi(b) .$$

La figure 1 illustre le propos pour une courbe régulière concave quelconque.

Si $p \geq 1$, la fonction a^p est croissante et *convexe* sur $[0, +\infty)$ et l'argument ne fonctionne plus puisqu'il conduit à $\varphi(a+b) > \varphi(a) + \varphi(b)$. Mais la convexité permet de montrer un résultat plus fort, à savoir que la quasi-norme $\|f\|_p$ satisfait l'inégalité triangulaire

$$\|f + h\|_p \leq \|f\|_p + \|h\|_p$$

et est par conséquent une vraie *norme* (avec la distance associée $d_p(f, h) = \|f - h\|_p$).

Pour développer la théorie des espaces L^p pour $1 < p < +\infty$ (les espaces L^1 et L^∞ peuvent être traités pour eux-mêmes) il faut comprendre la dualité fondamentale qui existe entre les $p, q \in]1, +\infty[$ reliés par la relation fondamentale $\frac{1}{p} + \frac{1}{q} = 1$ (i.e. $q = \frac{p}{p-1}$ et $p = \frac{q}{q-1}$). Si $p \in]1, +\infty[$ et si la relation est satisfaite alors $q \in]1, +\infty[$ et q s'appelle le *conjugué* de p .⁸ La première propriété permettant de prouver que

8. Si $p < 1$, il faudrait que q soit < 0 , ce qui n'est pas admis pour les espaces L^q .

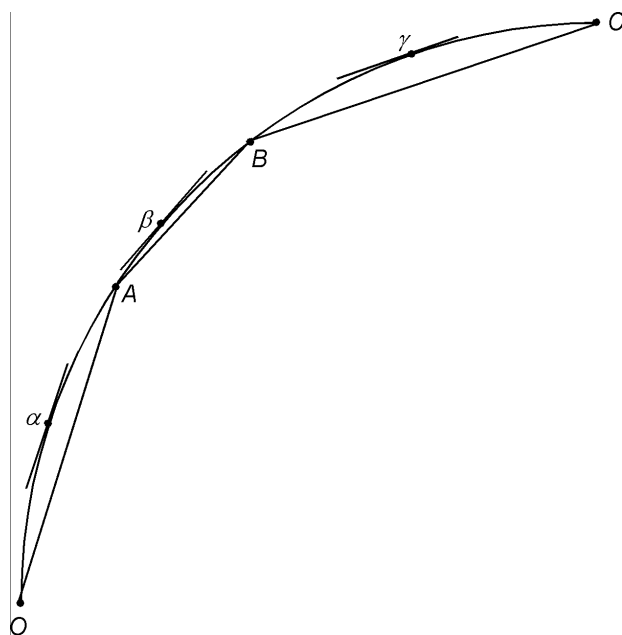


FIGURE 1. Les fonctions régulières concaves quelconques sont sous-additives. Soient 4 points O, A, B, C et les 3 cordes OA, AB, BC . D'après le théorème de Rolle il existe des points intermédiaires α, β, γ de pentes respectives celles des cordes. Comme la fonction est concave, ces pentes sont décroissantes.

dans ce contexte $\|f\|_p$ est bien une norme est la célèbre *inégalité de Hölder* : pour $f \in L^p$ et $h \in L^q$, $fh \in L^1$

$$\int_X |f(x) h(x)| d\nu(x) \leq \left(\int_X |f(x)|^p d\nu(x) \right)^{\frac{1}{p}} \left(\int_X |h(x)|^q d\nu(x) \right)^{\frac{1}{q}}$$

$$\|f \cdot h\|_1 \leq \|f\|_p \|h\|_q .$$

Cette inégalité fonctionnelle se base sur une inégalité élémentaire dite inégalité de Young : pour $\frac{1}{p} + \frac{1}{q} = 1$ et $a, b \geq 0$, on a

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

avec égalité si et seulement si $a^p = b^q$. Si $a^p = b^q = c$, on a trivialement

$$ab = c^{\frac{1}{p}} c^{\frac{1}{q}} = c^{\frac{1}{p} + \frac{1}{q}} = c$$

et

$$\frac{a^p}{p} + \frac{b^q}{q} = c \left(\frac{1}{p} + \frac{1}{q} \right) = c .$$

Pour démontrer le reste de la proposition, on considère la fonction $\psi(u) = ub - \frac{u^p}{p}$. Sa dérivée est $\psi'(u) = b - u^{p-1}$ et s'annule en $u = b^{\frac{1}{p-1}} \geq 0$ (bien défini car $b \geq 0$) et ne s'annule que là. Comme la dérivée seconde est $\psi''(u) = -(p-1)u^{p-2}$ qui vaut $-(p-1)b^{\frac{p-2}{p-1}} \leq 0$ au point critique, cet unique point critique est un maximum et il vaut

$$\begin{aligned} \psi\left(b^{\frac{1}{p-1}}\right) &= \left(b^{\frac{1}{p-1}}\right)b - \frac{\left(b^{\frac{1}{p-1}}\right)^p}{p} = b^{\frac{1}{p-1}+1} - \frac{b^{\frac{p}{p-1}}}{p} = b^{\frac{p}{p-1}} - \frac{b^{\frac{p}{p-1}}}{p} \\ &= b^{\frac{p}{p-1}} \left(1 - \frac{1}{p}\right) = \frac{b^q}{q} . \end{aligned}$$

Donc $\psi(a) = ab - \frac{a^p}{p} \leq \frac{b^q}{q}$ i.e. $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ et il n'y a égalité que si a est le maximum, i.e. si $a = b^{\frac{1}{p-1}}$, soit $a^p = b^{\frac{p}{p-1}} = b^q$.

Soient alors $f \in L^p$ et $h \in L^q$. On peut les supposer $\neq 0$ car sinon l'inégalité de Hölder est triviale. Supposons d'abord $\|f\|_p = \|h\|_q = 1$ et donc $\|f\|_p^p = \|h\|_q^q = 1$. D'après l'inégalité de Young et la linéarité de l'intégrale on a

$$\begin{aligned} \|f.h\|_1 &= \int_X |f(x)h(x)| d\nu(x) \leq \frac{1}{p} \int_X |f(x)|^p d\nu(x) + \frac{1}{q} \int_X |h(x)|^q d\nu(x) \\ &= \frac{1}{p} \|f\|_p^p + \frac{1}{q} \|h\|_q^q = \frac{1}{p} + \frac{1}{q} = 1 \end{aligned}$$

et l'inégalité est satisfaite. Si $f \in L^p$ et $h \in L^q$ sont de normes > 0 quelconques, on les normalise en prenant $\tilde{f} = \frac{f}{\|f\|_p}$ et $\tilde{h} = \frac{h}{\|h\|_q}$ et, en appliquant le résultat précédent à \tilde{f} et \tilde{h} , on obtient $\frac{1}{\|f\|_p \|h\|_q} \|f.h\|_1 \leq 1$, c'est-à-dire l'inégalité de Hölder.

L'inégalité de Hölder permet de démontrer l'*inégalité de Minkowski* qui exprime que $\|f\|_p$ satisfait l'inégalité triangulaire $\|f+h\|_p \leq \|f\|_p + \|h\|_p$ pour $p \geq 1$. Pour $p = 1$, cela découle simplement par linéarité de l'intégration de l'inégalité triangulaire des valeurs absolues :

$$|f(x) + h(x)| \leq |f(x)| + |h(x)| .$$

Pour $p = +\infty$, $|f(x)| \leq \|f\|_\infty$ ν -p.p., donc

$$|f(x) + h(x)| \leq |f(x)| + |h(x)| \leq \|f\|_\infty + \|h\|_\infty \quad \nu\text{-p.p.}$$

et par conséquent

$$\|f+h\|_\infty \leq \|f\|_\infty + \|h\|_\infty .$$

Pour $p \in]1, +\infty[$ et $\|f + h\|_p \neq 0$ (car sinon l'inégalité est triviale), on trouve (en remarquant que $p-1 > 0$ et en utilisant à nouveau $|f(x) + h(x)| \leq |f(x)| + |h(x)|$) :

$$\begin{aligned} \|f + h\|_p^p &= \int_X |f(x) + h(x)|^p d\nu(x) \\ &= \int_X |f(x) + h(x)|^{p-1} |f(x) + h(x)| d\nu(x) \\ &\leq \int_X |f(x) + h(x)|^{p-1} |f(x)| d\nu(x) + \int_X |f(x) + h(x)|^{p-1} |h(x)| d\nu(x) . \end{aligned}$$

Par Hölder, le premier terme satisfait (puisque $q = \frac{p}{p-1}$)

$$\begin{aligned} \int_X |f(x) + h(x)|^{p-1} |f(x)| d\nu(x) &\leq \left(\int_X (|f(x) + h(x)|^{p-1})^q d\nu(x) \right)^{\frac{1}{q}} \\ &\quad \times \left(\int_X |f(x)|^p d\nu(x) \right)^{\frac{1}{p}} \\ &= \left(\int_X (|f(x) + h(x)|^{\frac{(p-1)p}{p-1}}) d\nu(x) \right)^{\frac{p-1}{p}} \|f\|_p \\ &= \|f + h\|_p^{p-1} \|f\|_p . \end{aligned}$$

De même, le deuxième terme satisfait

$$\int_X |f(x) + h(x)|^{p-1} |h(x)| d\nu(x) \leq \|f + h\|_p^{p-1} \|h\|_p$$

et l'on obtient donc en définitive

$$\|f + h\|_p^p \leq \|f + h\|_p^{p-1} \|f\|_p + \|f + h\|_p^{p-1} \|h\|_p$$

soit l'inégalité triangulaire en divisant par $\|f + h\|_p^{p-1}$.

Les espaces L^p sont donc normés par $\|f\|_p$ pour $p \geq 1$. Cela ne signifie pas pour autant qu'ils ressemblent à des espaces \mathbb{R}^n ou \mathbb{C}^n de dimension infinie car la norme *ne dérive pas forcément d'un produit scalaire*. La norme est une mesure de la "longueur" des vecteurs mais n'implique pas en tant que telle une mesure des "angles" entre vecteurs. Nous allons voir que ce n'est que dans le cas $p = 2$ qu'il y a un produit scalaire donnant la norme (espaces de Hilbert).

Un autre théorème fondamental, dit *théorème de Riesz-Fischer*, affirme que les espaces L^p sont *complets* et donc des espaces de Banach (i.e. des espaces vectoriels normés complets). Le théorème est facile à démontrer avec l'inégalité de Minkowski et la théorie élémentaire de la convergence des séries. Soit f_n une suite de Cauchy dans L^p . Par définition, les $\|f_m - f_n\|_p$ deviennent aussi petits que l'on veut ($< \varepsilon$) pour m, n assez grands ($> N_\varepsilon$). Il faut montrer que l'on peut alors construire une sous-suite convergente. Pour cela, on prend des $\varepsilon_k = \frac{1}{2^{k+1}}$ dont la somme converge et

des $N_k = N_{\varepsilon_k}$ associés, puis on somme les f_{N_0} , $f_{N_1} - f_{N_0}$, etc. Les sommes partielles $\sum_{k=0}^{k=K} (f_{N_k} - f_{N_{k-1}})$ (avec par convention $f_{N_{-1}} = 0$) sont égales à f_{N_K} . Mais d'après Minkowski,

$$\left\| \sum_{k=0}^{k=K} |f_{N_k} - f_{N_{k-1}}| \right\|_p \leq \sum_{k=0}^{k=K} \|f_{N_k} - f_{N_{k-1}}\|_p \leq \sum_{k=0}^{k=\infty} \frac{1}{2^{k+1}} < +\infty .$$

Donc les sommes partielles $\sum_{k=0}^{k=K} (f_{N_k} - f_{N_{k-1}})$ étant absolument convergentes convergent simplement vers $f \in L^p$ ν -p.p. et la sous-suite f_{N_k} des f_n converge vers f .

Remarque. Pour $0 < p < 1$, L^p est bien un espace métrique complet pour la distance $d(f, h) = \|f - h\|_p^p$ mais on montre qu'il ne peut pas posséder de norme. Cela est dû au fait qu'il n'est pas localement convexe (0 ne possède pas une base de voisinages U convexes, i.e. tels que si $f, h \in U$ alors le segment $\lambda f + (1 - \lambda) h$, $\lambda \in [0, 1]$, joignant f ($\lambda = 1$) à h ($\lambda = 1$) est entièrement contenu dans U . On a bien pour la quasi-norme $\|f\|_p$ la formule générale

$$\|f + h\|_p \leq 2^{\frac{1-p}{p}} \left(\|f\|_p + \|h\|_p \right)$$

et donc une formule

$$\|f + h\|_p \leq K \left(\|f\|_p + \|h\|_p \right)$$

avec $K \geq 1$. Mais l'inégalité triangulaire exige $K = 1$ alors que pour $p < 1$, $\frac{1-p}{p} > 0$ et donc $2^{\frac{1-p}{p}} > 1$. □

On montre également les résultats clé suivants sur les espaces duaux. Par définition, le dual topologique $(L^p)'$ de L^p est l'ensemble des formes linéaires continues $u : L^p \rightarrow \mathbb{R}$ ou \mathbb{C} (suivant que l'on travaille sur \mathbb{R} ou \mathbb{C}). Si q est le conjugué de p et si $h \in L^q$, alors

$$f \mapsto u_h(f) = \int_X f(x) \overline{h(x)} d\nu(x)$$

est une forme linéaire sur L^p (à cause de la linéarité de l'intégrale en f) qui est intégrable d'après Hölder. On montre que l'application anti-linéaire $u : L^q \rightarrow (L^p)'$, $h \mapsto u_h$ est une *isométrie* et est donc injective puisque si $u_h = 0$, alors $\|u_h\| = 0$, donc $\|h\|_q = 0$ et $h = 0$ dans L^q . Hölder implique

$$|u_h(f)| \leq \|fh\|_1 \leq \|h\|_q \|f\|_p$$

pour tout f de L^p et donc l'inégalité

$$\|u_h\| = \text{Sup} \frac{|u_h(f)|}{\|f\|_p} \leq \|h\|_q .$$

On veut montrer que la borne est atteignable pour une f particulière, ce qui est plus fort que l'égalité de Hölder $\|fh\|_1 = \|h\|_q \|f\|_p$. Pour ce faire, dans le cas

$1 < p < +\infty$, on considère la fonction

$$f(x) = \frac{|h(x)|^q}{\overline{h(x)}}$$

(on suppose $h \neq 0$ car sinon il n'y a rien à démontrer). On a donc $\overline{h(x)}f(x) = |h(x)|^q$ et $|f(x)| = |h(x)|^{q-1}$. Donc, puisque $(q-1)p = (q-1)\frac{q}{q-1} = q$,

$$\int_X |f(x)|^p d\nu(x) = \int_X (|h(x)|^{q-1})^p d\nu(x) = \|h\|_q^q < +\infty$$

et par suite $f \in L^p$ avec $\|f\|_p^p = \|h\|_q^q$. On vérifie aussitôt que

$$\frac{|u_h(f)|}{\|f\|_p} = \frac{\int_X f(x) \overline{h(x)} d\nu(x)}{\|h\|_q^{\frac{q}{p}}} = \frac{\int_X |h(x)|^q d\nu(x)}{\|h\|_q^{\frac{q}{p}}} = \frac{\|h\|_q^q}{\|h\|_q^{\frac{q}{p}}} = \|h\|_q^{q-\frac{q}{p}} = \|h\|_q$$

ce qui montre que la borne est atteinte.

On raffine cette démonstration pour $p = 1$ et $p = +\infty$.

L'application $u : L^q \rightarrow (L^p)'$, $h \mapsto u_h$ étant une isométrie est injective mais elle n'est pas forcément surjective. Dans le cas $p = 2$, comme nous allons le voir plus bas, elle l'est à cause d'un théorème fondamental dû à Frigyes Riesz et Maurice Fréchet. Toutefois, dans le cas $p \neq 2$, elle l'est si la mesure ν est σ -finie⁹ et alors $(L^p)' \simeq L^q$. La démonstration adapte le théorème de Riesz.

On montre ensuite que, pour $p \in]1, +\infty[$, l'espace L^p est *réflexif* au sens suivant. On a une application évidente \mathcal{J} de L^p dans son *bidual* $(L^p)'' = ((L^p)')'$ donnée par $(\mathcal{J}(f))(u) = u(f)$ pour $f \in L^p$ et $u \in (L^p)'$. Un théorème fondamental, le *théorème de Hahn-Banach* (cf. plus bas section 4.4), permet de montrer que \mathcal{J} préserve la norme et est donc *injective*. Dire que L^p est *réflexif* veut dire que \mathcal{J} est *bijective* et donc un isomorphisme entre L^p et son bidual $(L^p)''$.

Pour les $f, h \in L^p \cap L^q$ (par exemple des fonctions bornées de support des sous-ensembles de mesure finie), on pourrait noter $u_h(f) = \langle f, h \rangle_{p,q}$ comme une sorte de produit scalaire. On aimerait bien que $\langle f, f \rangle_{p,q}$ soit réel mais alors il faut définir $u_h(f)$ par $\int_X f(x) \overline{h(x)} d\nu(x)$, donc

$$\langle f, f \rangle_{p,q} = \int_X f(x) \overline{f(x)} d\nu(x) = \int_X |f(x)|^2 d\nu(x)$$

et $h \mapsto u_h$ devient *anti-linéaire* en h . Ce serait encore mieux si l'on pouvait retrouver une norme. Pour cela, il faut $p = q = \frac{1}{2}$ ce qui conduit aux espaces de Hilbert dont la théorie va montrer pourquoi, pour $p \neq 2$, L^p ne peut pas être un Hilbert.

Remarque : espaces de Sobolev. On a souvent besoin pour évaluer la "taille globale" d'une fonction de tenir compte non seulement de ses valeurs mais aussi

9. C'est-à-dire, rappelons-le, si X est recouvrable par une famille dénombrable de S_j mesurables de mesure finie.

de celles de ses dérivées partielles (au sens des distributions)¹⁰ jusqu'à un certain ordre. Cela définit des espaces dits de Sobolev $\mathcal{H}^{p,m}$ des $f \in L^p$ dont toutes les dérivées partielles¹¹ $D^\alpha f$ jusqu'à l'ordre m sont dans L^p , espaces munis de la norme $\left(\sum_{|\alpha| \leq m} \|D^\alpha f\|_p^p\right)^{\frac{1}{p}}$. _____ □

4. Espaces de Hilbert

4.1. Produit scalaire

Lorsque $p = q = 2$, p est son propre conjugué et l'espace $\mathcal{H} = L^2(X, \nu)$ des fonctions à valeurs dans \mathbb{C} de carré intégrable sur l'espace mesurable (X, ν) possède une structure beaucoup plus riche que celle des autres espaces L^p . Cela est dû au fait que la norme $\|f\|_2$ découle d'un *produit scalaire hermitien*

$$\langle f, h \rangle_{\mathcal{H}} = \int_X f(x) \overline{h(x)} d\nu(x) \text{ ,}$$

ce qui signifie que $\langle f, h \rangle_{\mathcal{H}}$ est une application à valeurs dans \mathbb{C} , linéaire en f , anti-linéaire en h , satisfaisant $\langle h, f \rangle_{\mathcal{H}} = \overline{\langle f, h \rangle_{\mathcal{H}}}$ (ce qui implique que $\langle f, f \rangle_{\mathcal{H}}$ est réel) et $\langle f, f \rangle_{\mathcal{H}} > 0$ si $f \neq 0$ (propriété de positivité). En effet

$$\langle f, f \rangle = \int_X f(x) \overline{f(x)} d\nu(x) = \int_X |f(x)|^2 d\nu(x) = \|f\|_{\mathcal{H}}^2 \text{ .}$$

Pour L^2 , l'inégalité de Hölder implique la célèbre *inégalité de Cauchy-Schwarz* qui relie le produit scalaire aux normes : si $f, h \in L^2 = \mathcal{H}$ alors $f.h \in L^1$ et

$$\int_X |f(x) \overline{h(x)}| d\nu(x) \leq \left(\int_X |f(x)|^2 d\nu(x)\right)^{\frac{1}{2}} \left(\int_X |h(x)|^2 d\nu(x)\right)^{\frac{1}{2}}$$

$$|\langle f, h \rangle_{\mathcal{H}}| \leq \|f\|_{\mathcal{H}} \|h\|_{\mathcal{H}} \text{ .}$$

La théorie des espaces de Hilbert est l'une des plus connues de l'analyse fonctionnelle. Le lecteur en trouvera des exposés élémentaires dans de nombreux sites universitaires autres que Wikipedia. Citons par exemple le cours de Bertrand Rémy à l'École Polytechnique ou celui de Joël Merker ou François de Marçay à l'Université d'Orsay[364]. Ils généralisent directement les espaces vectoriels \mathbb{C}^n munis de leur produit scalaire hermitien

$$\langle f, h \rangle = f \cdot \overline{h} = \sum_{j=1}^{j=n} f_j \overline{h_j} = \overline{\langle h, f \rangle}$$

10. Au sens des distributions car les f des L^p n'étant définies qu'à des ensembles ν -négligeables près, on ne peut pas utiliser sans précautions la notion classique de dérivabilité.

11. Dans le cas standard de \mathbb{R}^n et \mathbb{C}^n , si $x = (x_1, \dots, x_n)$ et si $\alpha = (\alpha_1, \dots, \alpha_n)$ est un multi-indice de module $|\alpha| = \sum \alpha_i$, on note $D^\alpha f$ la dérivée partielle d'ordre $|\alpha|$ $\frac{\partial^{|\alpha|} f}{\partial^{\alpha_1} x_1 \dots \partial^{\alpha_n} x_n}$.

où les f_j et h_j sont les composantes respectives de f et h dans la base orthonormée naturelle de \mathbb{C}^n . À un niveau abstrait, ce sont des espaces \mathbb{C} -vectoriels \mathcal{H} munis d'un produit scalaire hermitien $\langle f, h \rangle$, donc d'une norme $\|f\|^2 = \langle f, f \rangle$, complets pour la norme (espaces de Banach) et, nous allons le voir, *isomorphes* à leur dual topologique \mathcal{H}' (l'espace de leurs formes linéaires continues). Si l'on se restreint à leurs sous-espaces vectoriels *fermés* et à leurs applications linéaires *continues*, ils possèdent essentiellement les propriétés des \mathbb{C}^n mais avec toute la richesse supplémentaire due à la possibilité de dimensions infinies. Ils sont au cœur de la mécanique quantique et leur théorie a été considérablement approfondie à cette fin par John von Neumann, entre autres dans son traité classique de 1932 *Mathematical Foundations of Quantum Mechanics* [388].

Comme nous l'avons anticipé, un produit scalaire hermitien est beaucoup plus riche qu'une norme car une norme ne mesure que des "longueurs" de vecteurs alors qu'un produit scalaire mesure aussi des "angles" entre vecteurs. En effet si f et h sont de norme 1, $|\langle f, h \rangle| \leq \|f\|_{\mathcal{H}} \|h\|_{\mathcal{H}} = 1$ (Cauchy-Schwarz) et donc il existe un angle θ tel que $\langle f, h \rangle = \cos(\theta)$. En particulier on peut définir une *orthogonalité* par $\langle f, h \rangle = 0$, i.e. $\theta = \pm \frac{\pi}{2}$. Le produit scalaire fait passer d'une simple métrique à une *géométrie* généralisant toute la richesse de la géométrie classique.

Le fait que la norme provienne d'un produit scalaire équivaut à la "règle du parallélogramme"

$$\|f + h\|^2 + \|f - h\|^2 = 2(\|f\|^2 + \|h\|^2)$$

disant que dans un parallélogramme la somme des carrés des quatre côtés est égale à la somme des carrés des deux diagonales.¹² Notons d'ailleurs que, sur \mathbb{R} , on a

$$\langle f, h \rangle_{\mathbb{R}} = \frac{1}{4} (\|f + h\|^2 - \|f - h\|^2)$$

et que, sur \mathbb{C} , on a

$$\langle f, h \rangle_{\mathbb{C}} = \langle f, h \rangle_{\mathbb{R}} + i \langle f, ih \rangle_{\mathbb{R}} .$$

La règle du parallélogramme explique pourquoi les L^p ne peuvent pas être des espaces de Hilbert pour $p \neq 2$. Pour simplifier prenons la norme $\|f\|_p$ sur \mathbb{C}^n , $\|f\|_p = \left(\sum_{j=1}^{j=n} |f_j|^p \right)^{\frac{1}{p}}$ et prenons $f = (1, 0, 0, \dots, 0)$, $h = (0, 1, 0, \dots, 0)$. On a $\|f + h\|_p^2 + \|f - h\|_p^2 = 2.2^{\frac{2}{p}}$ alors que $2(\|f\|_p^2 + \|h\|_p^2) = 2.2$. On voit qu'il ne peut y avoir égalité que si $p = 2$. On généralise aux L^p quelconques.

12. C'est un résultat enfantin de géométrie du plan euclidien. Soit $ABCD$ un parallélogramme d'angle θ en B . Pour la première diagonale AC on a $AC^2 = AB^2 + BC^2 + 2AB \cdot BC \cos(\theta)$. Pour la deuxième diagonale BD on a $BD^2 = AB^2 + AD^2 + 2AB \cdot AD \cos(\pi - \theta)$. Mais $BC = AD$ et $\cos(\pi - \theta) = -\cos(\theta)$, d'où le résultat.

4.2. Projecteurs

Nous avons dit plus haut que les propriétés des espaces de Hilbert généralisent directement celles valides en dimension finie à condition de se restreindre aux sous-espaces¹³ fermés et aux applications continues. Cela est dû à la raison suivante. Si E est un sous-espace quelconque de \mathcal{H} , le produit scalaire permet de définir son “orthogonal”

$$E^\perp = \{h \mid \langle e, h \rangle = 0, \forall e \in E\}$$

ensemble des h orthogonaux à tous les e de E . On montre facilement que E^\perp est fermé car si h_n est une suite de E^\perp qui converge vers $h \in \mathcal{H}$ alors pour tout $e \in E$, la suite $\langle e, h_n \rangle$, qui converge vers $\langle e, h \rangle$, est identiquement nulle et donc $\langle e, h \rangle = 0$ et $h \in E^\perp$. On en déduit que $E^\perp = \overline{E^\perp}$ et que $(E^\perp)^\perp = \overline{E}$ et que sur les sous-espaces fermés F , l'opération “prendre l'orthogonal” est par conséquent involutive. Si F est fermé, \mathcal{H} se décompose en somme directe $\mathcal{H} = F \oplus F^\perp$.

Remarque. Le fait que E^\perp soit fermé fournit un critère simple de *densité* : E est dense dans \mathcal{H} si et seulement si $E^\perp = \{0\}$. Il est très facile de construire des sous-espaces denses non fermés dans les espaces fonctionnels classiques. Par exemple dans un Hilbert de fonctions approximables par des fonctions C^∞ , le sous-espace des fonctions C^∞ sera dense mais non fermé. \square

À tout sous-espace fermé F on peut associer son *projecteur orthogonal* P_F qui, à tout $f \in \mathcal{H}$ décomposé en $f = f_F + f_{F^\perp}$ dans la décomposition $\mathcal{H} = F \oplus F^\perp$, associe la composante $P_F(f) = f_F$. On a alors $f_{F^\perp} = f - P_F(f)$. La norme $\|f - P_F(f)\|$ est la distance de f à F , c'est-à-dire $\inf_{h \in F} \|f - h\|$.

Les applications linéaires continues entre Hilbert $\varphi : \mathcal{H} \rightarrow \mathcal{F}$ possèdent une norme car il existe une constante C telle que $\|\varphi(f)\|_{\mathcal{F}} \leq C \|f\|_{\mathcal{H}}$. Le plus petit C est la norme $\|\varphi\|$. On peut la définir par $\|\varphi\| = \sup_{\|f\|_{\mathcal{H}} \leq 1} \|\varphi(f)\|_{\mathcal{F}}$. Leurs sous-espaces

noyaux sont fermés¹⁴ et comme le noyau du projecteur P_F est F^\perp , on voit qu'il y a équivalence entre les sous-espaces fermés et les noyaux d'applications linéaires continues.

4.3. Théorème de Riesz et auto-dualité

En ce qui concerne les *formes* linéaires continues $u(f)$ constituant le dual topologique \mathcal{H}' de \mathcal{H} , nous avons vu plus haut que les $u_h(f) = \langle f, h \rangle$ appartiennent à \mathcal{H}' et sont de norme $\|u_h\| = \|h\|$. Un théorème fondamental, le *théorème de représentation de Riesz*, dit que l'on obtient ainsi *tout* \mathcal{H}' et que l'application $h \mapsto u_h$ de \mathcal{H} dans \mathcal{H}' est *bijective*. C'est un isomorphisme anti-linéaire. Il s'agit ainsi de montrer que si $u \in \mathcal{H}'$, il existe un et un seul vecteur h_u de \mathcal{H} tel que $u(f) = \langle f, h_u \rangle$ pour tout

13. Par défaut, “sous-espace” signifie évidemment “sous-espace vectoriel”.

14. $\{0\}$ est fermé et l'image réciproque d'un fermé par une application continue est fermée.

f . L'unicité est évidente car si $\langle f, h_u \rangle = \langle f, h'_u \rangle$ pour tout f , alors $h_u = h'_u$. C'est l'existence de h_u qu'il faut prouver. Le cas $u = 0$ est évident car $\langle f, h \rangle = 0$ pour tout f équivaut à $h = 0$. Supposons donc $u \neq 0$. Son noyau $K = u^{-1}(0)$ est un sous-espace fermé ($\neq \mathcal{H}$) car u est supposée continue et son orthogonal K^\perp est de dimension 1. Soit e_u un élément de norme 1 de K^\perp . On a

$$\mathcal{H} = K \oplus K^\perp = K \oplus \mathbb{C}e_u .$$

On vérifie immédiatement que $u(f) = \langle f, u(e_u)e_u \rangle$ et que l'on peut donc prendre $h_u = u(e_u)e_u$. En effet si $f \in \mathcal{H}$, $f = k + \lambda e_u$ avec $k \in K$ et $\lambda \in \mathbb{C}$. Donc

$$u(f) = u(k + \lambda e_u) = u(k) + \lambda u(e_u) = \lambda u(e_u)$$

puisque $u(k) = 0$. Par ailleurs, puisque $\langle k, e_u \rangle = 0$,

$$\begin{aligned} \langle f, u(e_u)e_u \rangle &= \langle k + \lambda e_u, u(e_u)e_u \rangle = \langle \lambda e_u, u(e_u)e_u \rangle \\ &= \lambda u(e_u) \langle e_u, e_u \rangle = \lambda u(e_u) . \end{aligned}$$

Donc tout $u \in \mathcal{H}'$ est un u_{h_u} et \mathcal{H} est anti-linéairement isomorphe à son dual topologique.

4.4. Théorème de “séparation” de Hahn-Banach

Pour les espaces de Hilbert on peut montrer facilement un cas particulier du théorème de Hahn-Banach que l'on rencontre dans l'analyse convexe dans les Banach. Il s'agit du théorème de “séparation” disant que si E est un sous-espace quelconque de \mathcal{H} , si $\bar{E} \neq \mathcal{H}$ et si $f_0 \notin \bar{E}$ (i.e. si f_0 n'est pas adhérent à E) alors il existe une forme linéaire continue $u \in \mathcal{H}'$ séparant f_0 de E au sens où $u = 0$ sur E alors que $u(f_0) = 1$. Si f_0 est très près de \bar{E} , u variera donc beaucoup en passant de E à f_0 mais restera continue. Considérons l'orthogonal $E^\perp = \bar{E}^\perp$ de E . Comme $f_0 \notin \bar{E}$, $P_{E^\perp}(f_0)$ est $\neq 0$ et l'on peut donc considérer le vecteur

$$h = \frac{P_{E^\perp}(f_0)}{\|P_{E^\perp}(f_0)\|^2} .$$

La forme linéaire $u_h(f) = \langle f, h \rangle$ est continue et $\in \mathcal{H}'$. On a

$$u_h(f_0) = \left\langle f_0, \frac{P_{E^\perp}(f_0)}{\|P_{E^\perp}(f_0)\|^2} \right\rangle$$

mais comme $f_0 = P_{E^\perp}(f_0) + P_{\bar{E}}(f_0)$,

$$\left\langle f_0, \frac{P_{E^\perp}(f_0)}{\|P_{E^\perp}(f_0)\|^2} \right\rangle = \left\langle P_{E^\perp}(f_0), \frac{P_{E^\perp}(f_0)}{\|P_{E^\perp}(f_0)\|^2} \right\rangle = 1 .$$

Donc $u_h(f_0) = 1$. Par ailleurs, si $e \in E$,

$$u_h(e) = \left\langle e, \frac{P_{E^\perp}(f_0)}{\|P_{E^\perp}(f_0)\|^2} \right\rangle = 0$$

puisque e et $P_{E^\perp}(f_0)$ sont orthogonaux.

4.5. Bases hilbertiennes

Lorsque \mathcal{H} est “séparable”, i.e. contient un sous-ensemble dénombrable qui est néanmoins dense, alors il existe une base hilbertienne $\{e_k\}_{k \in \mathbb{N}}$ i.e. un ensemble dénombrable orthonormé de vecteurs e_k qui engendre *topologiquement* \mathcal{H} : le sous-espace engendré par les e_k (i.e. celui des combinaisons linéaires *finies* des e_k) est dense dans \mathcal{H} et tout vecteur f de \mathcal{H} est combinaison linéaire *sommable*

$$f = \sum_{k \in \mathbb{N}} c_k e_k$$

des e_k avec des coefficients $c_k \in \mathbb{C}$ formant une série convergente, la sommabilité étant prise au sens de la norme, et l'on a le “théorème de Pythagore”

$$\|f\|^2 = \sum_{k \in \mathbb{N}} |c_k|^2 .$$

\mathcal{H} est alors vraiment un \mathbb{C}^n avec $n = \infty$. On construit une base hilbertienne en prenant un e_1 de norme 1 dans \mathcal{H} et en considérant la somme directe orthogonale $\mathcal{H} = \mathbb{C}e_1 \oplus^\perp \mathcal{H}_2$ avec $\mathcal{H}_2 = (\mathbb{C}e_1)^\perp$. On prend ensuite un e_2 dans \mathcal{H}_2 et on itère la construction. Si \mathcal{H} est de dimension finie le processus s'arrête à la dimension de \mathcal{H} et si \mathcal{H} est de dimension infinie mais séparable, on obtient à la limite une base hilbertienne (précédé dit de Gram-Schmidt).

Nous avons rencontré dans le Vol I (chapitre 3, section 3.3.1.) des bases hilbertiennes et nous en rencontrerons encore. Par exemple, pour l'espace $L^2\left(\mathbb{R}, e^{-\frac{x^2}{2}} dx\right)$ (que nous rencontrerons à la section 5.10.1 du chapitre 17 à propos de la transformée de Bargmann), une base orthogonale est celle des polynômes d'Hermite

$$H_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} \left(e^{-\frac{x^2}{2}} \right)$$

dont le produit scalaire est $\langle H_k, H_\ell \rangle = k! \sqrt{2\pi} \delta_{k,\ell}$.¹⁵ On en tire immédiatement une base orthonormale. De même pour $L^2(\mathbb{R}^+, e^{-x} dx)$ une base orthonormale est celle

15. On définit souvent les polynômes d'Hermite par la variante

$$\tilde{H}_k(x) = (-1)^k e^{x^2} \frac{d^k}{dx^k} \left(e^{-x^2} \right) = 2^{\frac{k}{2}} H_k(\sqrt{2} x) .$$

des polynômes de Laguerre

$$L_k(x) = \frac{e^x}{k!} \frac{d^k}{dx^k} (x^k e^{-x}).$$

L'omniprésence de ces polynômes en physique est due à cette propriété.

L'existence de bases hilbertiennes permet de comprendre intuitivement les différents types de sous-espaces que l'on peut obtenir. Un sous-espace E sera constitué des éléments de \mathcal{H} dont les coefficients c_k seront restreints par certaines conditions. Si par exemple tous les c_{k_0} des $f \in E$ sont nuls, on obtiendra le sous-espace fermé de codimension 1 qui est l'orthogonal de $\mathbb{C}e_{k_0}$. Avec d'autres contraintes on obtiendra des sous-espaces denses non fermés.

4.6. Non compacité de la boule unité

Notons que si \mathcal{H} est de dimension *infinie* et séparable et si $\{e_k\}_{k \in \mathbb{N}}$ est une base hilbertienne, alors les e_k forment une suite infinie dénombrable de la boule unité fermée $\overline{B}_{\mathcal{H}}(0, 1)$ qui ne peut pas converger puisque $\|e_k - e_\ell\| = \sqrt{2}$ si $k \neq \ell$. Cela montre que $\overline{B}_{\mathcal{H}}(0, 1)$ *n'est pas compacte* et cela non pas parce qu'elle est ouverte ou contient des éléments non bornés mais parce qu'elle est de dimension infinie. Il s'agit d'une source originale de non compacité. D'ailleurs le célèbre *théorème de compacité de Riesz* dit que $\overline{B} = \overline{B}_{\mathcal{H}}(0, 1)$ est compacte si et seulement si \mathcal{H} est de dimension finie. Si \mathcal{H} est de dimension finie, \overline{B} étant fermée et bornée est automatiquement compacte. Si inversement \overline{B} est compacte, on peut considérer son recouvrement ouvert par les boules ouvertes $B_{\mathcal{H}}(f, \frac{1}{2})$ pour $f \in \overline{B}$ et en extraire un recouvrement ouvert fini $B_j = B_{\mathcal{H}}(f_j, \frac{1}{2})$, $j = 1, \dots, N$. Soit F le sous-espace engendré par les f_j . Il est fermé et on montre qu'il est égal à \mathcal{H} ce qui prouve que \mathcal{H} est de dimension finie $\leq N$. Si $F \subsetneq \mathcal{H}$, soit $h \notin F$ et $d = d(h, F) > 0$ (car F fermé). Si k est la projection de h sur F , $d = \|h - k\|$. Comme $d > 0$, $\frac{h-k}{d} \in \overline{B}$ et il existe donc j tel que $\frac{h-k}{d} \in B_j$, i.e. $\|\frac{h-k}{d} - f_j\| < \frac{1}{2}$, i.e.

$$\|h - (k + df_j)\| < \frac{d}{2}.$$

Mais cela est impossible car $k + df_j \in F$ puisque $k, f_j \in F$ et la distance de h à F est $d > \frac{d}{2}$.

4.7. Topologies

Pour les \mathcal{H} fonctionnels il existe plusieurs topologies. D'abord évidemment celle de la convergence en norme disant qu'une suite f_n dans \mathcal{H} tend vers f lorsque $n \rightarrow \infty$ si $\|f_n - f\| \xrightarrow{n \rightarrow \infty} 0$. Mais aussi celle de la convergence "faible" disant que $f_n \xrightarrow[n \rightarrow \infty]{w} f$ si $\langle f_n, h \rangle \xrightarrow{n \rightarrow \infty} \langle f, h \rangle$ pour tout h de \mathcal{H} et celle de la convergence "simple" disant

que $f_n \xrightarrow[n \rightarrow \infty]{s} f$ si $f_n(x) \xrightarrow[n \rightarrow \infty]{} f(x)$ dans \mathbb{C} pour tout $x \in X$. En dimension infinie la convergence faible est “vraiment” faible puisque la boule unité fermée $\overline{B}_{\mathcal{H}}(0, 1)$ est faiblement compacte! En fait, toute suite *bornée* f_n possède une sous-suite convergeant faiblement. Cela est facile à montrer au moyen d’un “argument diagonal”. Soient $f_{n,k} = \langle f_n, e_k \rangle$ les composantes des f_n dans une base hilbertienne $\{e_k\}_{k \in \mathbb{N}}$. Pour chaque k , la suite $f_{n,k}$ est bornée et admet donc une sous-suite convergente. En prenant les e_k les uns après les autres, on construit une suite décroissante de sous-suites convergentes et par passage à la limite une sous-suite $f_{\varphi(n)}$ telle que $\langle f_{\varphi(n)}, e_k \rangle$ soit convergente pour tout k . Par passage à la limite on en déduit que pour tout h de \mathcal{H} , $\langle f_{\varphi(n)}, h \rangle$ converge vers une fonction $u(h)$. On montre que u est une forme linéaire continue et, en appliquant le théorème de représentation de Riesz, on peut écrire $u(h) = \langle f_{\varphi(n)}, h \rangle$ et $f_{\varphi(n)}$ converge faiblement vers f .

On montre d’ailleurs la réciproque à savoir que les $f_n \xrightarrow[n \rightarrow \infty]{w} f$ sont nécessairement bornées. On montre également que la convergence en norme $\|f_n - f\| \xrightarrow[n \rightarrow \infty]{} 0$ équivaut à la convergence faible $f_n \xrightarrow[n \rightarrow \infty]{w} f$ plus la convergence numérique des normes $\|f_n\| \xrightarrow[n \rightarrow \infty]{} \|f\|$.

5. Les distributions de Laurent Schwartz

Dans ce qui suit, nos outils d’analyse fonctionnelle seront pratiquement limités aux espaces L^p et surtout aux espaces de Hilbert L^2 . Toutefois, nous aurons parfois besoin de façon vitale de fonctions plus générales que Laurent Schwartz¹⁶ a appelé des *distributions* et que nous avons déjà évoquées plusieurs fois de façon informelle puisque, répétons-le, l’action des profils récepteurs des neurones visuels sur le signal optique s’interprète comme l’action de fonctions test sur une distribution. L’exemple prototypique en est la distribution de Dirac δ .

La percée accomplie par Schwartz a été de définir rigoureusement en termes d’analyse fonctionnelle, ces fonction utilisées depuis longtemps par les ingénieurs et les physiciens. Par exemple la “fonction” de Dirac $\delta(x)$ est une “fonction” sur \mathbb{R} partout nulle sauf en 0 où elle a une valeur infinie et telle que l’intégrale sur \mathbb{R} , $\int_{\mathbb{R}} \delta(x) dx = 1$. Ces propriétés n’étaient pas rigoureusement définissables dans

16. Laurent Schwartz (1915-2002) a révolutionné l’analyse fonctionnelle et ses liens avec la physique avec sa théorie des distributions (cf. [493]) qui lui valut la médaille Fields en 1950. Il fut Professeur à l’Ecole Polytechnique de 1959 à 1980 à la suite de son beau-père Paul Lévy (l’un des fondateurs de la théorie moderne des probabilités) et, malgré les difficiles problèmes politiques que pouvaient créer ses engagements contre la guerre du Vietnam et la guerre d’Algérie avec une école militaire, il y transforma profondément l’enseignement et y créa le Centre de Mathématiques. Ses cours et ses séminaires sont légendaires. Nous avons eu le privilège d’être son élève, de suivre son séminaire sur la transformée de Gelfand et de faire partie des premiers élèves le l’X rattachés au Centre. Cela détermina notre carrière scientifique.

le cadre de l'analyse fonctionnelle de l'époque. Certes, $\delta(x)$ peut-être considérée comme une limite de gaussiennes, mais cette limite n'est plus une fonction.

Pour les ingénieurs du signal, δ est une "impulsion" en 0. Pour les physiciens, c'est une masse ponctuelle en 0. Mathématiquement, on peut la traiter comme une *mesure* et donc comme une *fonctionnelle* $\delta(A)$ assignant à tout sous-ensemble A de \mathbb{R} la valeur 1 si $0 \in A$ et 0 sinon. Si $\chi_A(x)$ est la fonction caractéristique de A , on a $\delta(A) = \chi_A(0)$ et l'on peut introduire l'écriture fonctionnelle

$$\delta(A) = \int_{\mathbb{R}} \delta(x) \chi_A(x) dx = \chi_A(0) .$$

On essaiera donc de définir la "fonction" $\delta(x)$ par cette formule

$$\int_{\mathbb{R}} \delta(x) f(x) dx = f(0) ,$$

le problème devenant alors de donner un sens rigoureux à cette formule sous des conditions bien définies.

Mais, étant données les propriétés de linéarité des intégrales, on peut considérer δ comme une application linéaire de l'espace fonctionnel \mathcal{F} des f dans \mathbb{R} , $\delta(f) = f(0)$, c'est-à-dire comme une *forme linéaire* sur \mathcal{F} . Si l'espace \mathcal{F} est assez contraint, δ sera *continue* et sera donc un élément de l'espace fonctionnel *dual* topologique de \mathcal{F} , $\delta(f)$ pouvant s'écrire comme une dualité $\langle \delta, f \rangle$. Une fois trouvée cette interprétation fonctionnelle en termes de dualité, on peut alors facilement généraliser et définir une distribution T comme un élément du dual d'un \mathcal{F} bien choisi. Plus \mathcal{F} sera contraint et plus les distributions pourront généraliser les fonctions de façon "sauvage". Si h est une fonction localement intégrable les intégrales $\int_{\mathbb{R}} h(x) f(x) dx$ auront un sens et définiront des distributions T_h dites "régulières", la fonction h étant alors interprétée comme une "densité" de distribution de masse, généralisant l'interprétation de δ comme masse ponctuelle en 0. On voit donc que les distributions T généralisent bien les fonctions h .

Remarque. En fait, on peut considérer intuitivement que, pour des \mathcal{F} bien choisis, les distributions permettent de définir les dérivées de fonctions simplement continues. La régularité normale sera la régularité C^∞ et les fonctions $h \in C^\infty(M)$ seront automatiquement des distributions. Si T est une distribution, on appellera alors *support singulier* de T le fermé complémentaire du plus grand ouvert sur lequel T sera localement représentable par une $h \in C^\infty(M)$. □

La contrainte initiale introduite par Laurent Schwartz était, étant donnée une variété M , de prendre pour \mathcal{F} l'espace $\mathcal{C}_c^\infty(M)$ des fonctions C^∞ à support compact, espace muni de sa topologie naturelle de la convergence uniforme des fonctions et de toutes leurs dérivées. Il les appelait des "fonctions test". C'est cette définition que nous avons introduite dans le Vol I, section 3.3.4., en disant que les profils récepteurs $\varphi_{(x_0, y_0)}(x, y)$ des neurones visuels étaient des "fonctions test" neuralemeent câblées

alors que les images rétinienne $I(x, y)$ (très bruitées et irrégulières) étaient des distributions, les profils récepteurs agissant par dualité comme des filtres $\langle \varphi_{(x_0, y_0)}, I \rangle$.

On peut alors montrer, en utilisant des outils idoines d'analyse fonctionnelle, que le calcul différentiel se généralise bien aux distributions. La dérivée T' d'une distribution T est donnée par la formule

$$\langle T', f \rangle = - \langle T, f' \rangle .$$

Ainsi définies, les distributions appartiennent au dual topologique de l'espace des fonctions test $\mathcal{C}_c^\infty(\mathbb{R}^n)$. Il est clair d'après leur définition même comme dualité $\langle T, f \rangle$ entre l'espace \mathcal{F} des f et son dual topologique \mathcal{F}' qu'un élargissement de \mathcal{F} conduit à restreindre \mathcal{F}' . Un espace particulièrement important est celui $\mathcal{C}_\downarrow^\infty(\mathbb{R}^n)$ des fonctions *rapidement décroissantes à l'infini*, c'est-à-dire décroissant plus vite vers 0 à l'infini que tous les $\frac{1}{\|x\|^N}$ avec N aussi grand que l'on veut. Un exemple typique en est $e^{-\|x\|^2}$. Depuis son introduction, cet espace $\mathcal{C}_\downarrow^\infty(\mathbb{R}^n)$ s'appelle *l'espace de Schwartz* et est noté \mathcal{S} . Sa topologie est plus subtile que celle de $\mathcal{C}_c^\infty(\mathbb{R}^n)$ qui en est un sous-espace dense. Bien que très contraint, il est dense dans les espaces L^p pour $p \geq 1$, $p < \infty$. Son dual topologique \mathcal{S}' est l'espace des distributions dites "tempérées". Nous verrons plus bas à la section 2.1.1 du chapitre 17 que c'est le meilleur espace pour les transformées de Fourier.

On peut encore élargir l'espace \mathcal{F} en considérant des distributions à support *compact* comme δ . Pour δ il suffit que les f soient localement intégrables. Le support d'une distribution T est le fermé complémentaire de l'ouvert U sur lequel T est identiquement nulle. Autrement dit $\langle T, f \rangle = 0$ pour toute $f \in \mathcal{C}_c^\infty(M)$ de support compact $K \subset U$. Leur espace est le dual topologique \mathcal{E}' de l'espace $\mathcal{C}^\infty(M)$ souvent noté \mathcal{E} .

Remarque. Il faut faire attention au fait que, contrairement à ce qui se passe pour les fonctions, le *produit* de distributions est délicat à définir. Cela est possible si T est régulière, i.e. si $T = T_h$ avec h localement intégrable, car alors si f est une fonction test, hf l'est également. On peut donc définir $T_h.S$ par $\langle T_h.S, f \rangle = \langle S, hf \rangle$. Mais si T est quelconque, on ne peut pas définir $T.S$ par $\langle T.S, f \rangle = \langle S, \langle T, f \rangle \rangle$ car il faudrait alors que $\langle T, f \rangle$ soit une fonction test ce qui n'est pas garanti. D'ailleurs si l'on pense que les fonctions test sont, comme les profils récepteurs, des "appareils de mesure" opérant sur les distributions, on voit qu'un produit ne sera bien défini que si l'une des distributions peut être elle-même interprétée comme un "appareil de mesure". □

Le produit opératoire en théorie des distributions est le produit de *convolution*. Il généralise celui des fonctions et peut être introduit dès que l'on dispose sur la variété de base M d'une structure de *groupe* définissant l'équivalent des translations sur \mathbb{R}^n . Rappelons (cf. chapitre 2, section 6.4 à propos des filtres linéaires que sont les profils récepteurs des neurones visuels) que, f et g étant deux fonctions sur \mathbb{R}^n ,

leur produit de convolution $f * g$ est l'intégrale

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y) g(x-y) dy = \int_{\mathbb{R}^n} f(x-y) g(y) dy$$

lorsque celle-ci est définie. Ce produit est linéaire par rapport aux deux fonctions, associatif et commutatif. Il satisfait, lorsque f et g sont dérivables, intégrables et de dérivées intégrables, l'égalité

$$(f * g)'(x) = f'(x) * g(x) = f(x) * g'(x) .$$

Un élément neutre pour ce produit doit satisfaire

$$f(x) = \int_{\mathbb{R}^n} f(y) \delta(x-y) dy ,$$

ce qui implique que $\delta(x)$ soit la distribution de Dirac. Comme celle-ci n'est pas une fonction, on voit que le bon contexte pour le produit de convolution est celui des distributions.

Le produit de convolution se définit d'abord de façon très naturelle avec les fonctions test f . Si T est une distribution, on définit $T * f$ par la formule

$$(T * f)(x) = \langle T(y), f(x-y) \rangle .$$

C'est bien une distribution et même une fonction C^∞ . Si T est à support compact et si $f \in \mathcal{S} = \mathcal{C}_\downarrow^\infty(\mathbb{R}^n)$ alors $T * f$ est également un élément de \mathcal{S} .

Pour définir un produit de convolution $T * R$ avec une autre distribution R il faut être soigneux car cela soulève des subtilités d'analyse fonctionnelle. Si R est à support compact, cela ne pose pas de problème et on généralise simplement la définition pour les fonctions. On pose

$$\langle (T * R)(x), f(x) \rangle = \langle T(x), (\tilde{R} * f)(x) \rangle$$

où la distribution \tilde{R} est définie par

$$\langle \tilde{R}(x), f(x) \rangle = \langle R(x), f(-x) \rangle .$$

Cette définition est cohérente car \tilde{R} étant, comme R , une distribution à support compact, sa convolution avec une fonction test f est encore une fonction test et donc $\langle T(x), (\tilde{R} * f)(x) \rangle$ est bien défini.

On retrouve alors les bonnes propriétés de la convolution : bilinéarité par rapport aux deux fonctions, associativité et commutativité. Elle satisfait l'égalité

$$(T * R)' = T' * R = T * R' .$$

C'est ce que nous avons expliqué dans le Vol I, chapitre 3, section 3.3.4. à propos du critère de "zero crossing" de David Marr qui extrait les bords d'une image au moyen

de profils récepteurs en laplacien de gaussienne ΔG . Si I est le signal considéré comme distribution, la formule est

$$\Delta G * I = \Delta (G * I) ,$$

autrement dit le laplacien du “Gaussian blurring” $G * I$ de I .

Avec δ on peut donc reconstruire immédiatement la dérivation des distributions puisque

$$T' = (\delta * T)' = \delta' * T$$

et, par conséquent, pour tout opérateur différentiel D ,

$$DT = D\delta * T .$$

6. L'exemple des séries de Fourier

L'exemple le plus classique d'analyse hilbertienne est celui intervenant dans les *séries de Fourier* des fonctions périodiques. Nous reviendrons en détail dans la section 2.1.1 du chapitre 17 sur les transformées de Fourier des fonctions définies sur \mathbb{R} . Nous y ferons référence à ce que nous avons déjà exposé dans le Vol I, en particulier avec la *théorie des ondelettes* à la section 5.3. du chapitre 3.

Dans le cas des fonctions *périodiques*, les transformées de Fourier définissent des *séries trigonométriques* et l'analyse des propriétés extrêmement subtiles de convergence de ces séries a été l'un des principaux moteurs de l'émergence et du développement remarquable de l'analyse fonctionnelle au cours du XIX^e siècle. Grâce à de nombreux mathématiciens (à la fois théoriciens et calculateurs de génie) les idées pionnières de Fourier, qui ne pouvaient pas être mathématisées avec toute la rigueur nécessaire à son époque, ont pu être précisées et ont ouvert un immense

continent.¹⁷ L'intérêt des séries de Fourier reste éminent d'une part pour leurs applications innombrables et d'autre part comme exemple pédagogique. Elles constituent une propédeutique à l'analyse fonctionnelle où de très nombreux calculs peuvent être effectués explicitement et facilement illustrés par ordinateur.

Les fonctions périodiques sont le plus souvent définies soit sur le cercle unité \mathbb{S}^1 , et donc de période $T = 2\pi$, soit sur l'intervalle $[0, 1]$, et donc de période $T = 1$. Nous choisirons par commodité $T = 2\pi$ et nous considérerons les f comme préférentiellement définies sur $[0, 2\pi]$ avec $f(0) = f(2\pi)$ ou alternativement sur $[-\pi, \pi]$ avec $f(-\pi) = f(\pi)$.¹⁸ La première remarque est que, contrairement à \mathbb{R} , \mathbb{S}^1 est *compact*, ce qui a de nombreuses conséquences en termes de topologie d'espaces fonctionnels. En particulier les f continues sont bornées (elles atteignent leur maximum et leur minimum car les images continues de compact sont compactes) et sont *uniformément* continues d'après un théorème de 1872 de Eduard Heine (1821-1881).¹⁹

17. Jean-Baptiste Joseph Fourier (21 mars 1768-16 mai 1830) fut l'un des plus grands savants de l'époque révolutionnaire, puis de l'Empire, puis de la Restauration. Auguste Comte le considérait comme un Newton de la thermodynamique. Élève extrêmement brillant il fit partie de l'École Normale de l'An III (1795) où professaient Monge, Lagrange, Laplace, Berthollet, Haüy (il y fut d'ailleurs repéré par Monge qui le soutiendra). Puis il enseigna à l'École Polytechnique (créée en 1794 et, sous ce nom, le 1er septembre 1795). Il joua un rôle clé dans la Campagne de Bonaparte en Égypte en 1798 ; il écrivit la Préface de la *Description de l'Égypte* et son secrétaire était le frère de Champollion qu'il soutint dans ses recherches. Nommé par Napoléon préfet de l'Isère en 1802, il eut une notable vie politique. Sur le plan scientifique, malgré ses importantes responsabilités académiques (il fut nommé Recteur de la Faculté impériale de Grenoble en 1810), il mit un certain temps à être pleinement apprécié à cause des critiques envers un certain manque de rigueur et, surtout, à cause des tourments du changement de régime après l'effondrement de l'Empire. Après une première élection et un premier refus de Louis XVIII en 1817, il fut enfin reçu à l'Académie des sciences en 1818. Il en devint le secrétaire perpétuel en 1822 à la mort de Delambre, puis fut élu à la Royal Society de Londres en 1823 et à l'Académie française en 1826 ; il remplaça aussi Laplace en 1827 comme Président du Conseil scientifique de Polytechnique. Indépendamment de la question de savoir si au niveau physique sous-jacent la chaleur est un fluide – phlogistique, calorique, “fluide igné” – comme le pensait Lavoisier ou une agitation moléculaire comme le pensait Laplace, son grand ouvrage sur la propagation de la chaleur prend forme dans ses mémoires de base de 1807 et 1811 (qui n'aboutiront à une publication qu'entre 1815 et 1826) et paraît sous forme aboutie dans le grand traité de 1822 sur *La Théorie analytique de la chaleur* [189]. Ami de la grande mathématicienne Sophie Germain (surtout connue pour ses travaux sur le théorème de Fermat mais également spécialiste des plaques vibrantes), il influença des mathématiciens et des physiciens comme Lejeune-Dirichlet, Sturm et Navier. Gaston Darboux publia en 1888-1890 une partie de ses œuvres.

18. La différence porte simplement sur les mesures de Lebesgue. Sur $[0, 2\pi]$ elle est dx et sur \mathbb{S}^1 elle est plutôt $\frac{d\theta}{2\pi}$.

19. En effet, comme f est continue, en tout $\theta \in \mathbb{S}^1$, il existe pour tout $\varepsilon > 0$ un intervalle ouvert $U_\delta \ni \theta$ assez petit de taille δ tel que si $\theta' \in U_\delta$ alors $|f(\theta) - f(\theta')| \leq \varepsilon$. La continuité

L'idée initiale des développements en séries trigonométriques, et, plus généralement, de considérer une forme f d'un certain espace de formes \mathcal{F} comme une composition, éventuellement infinie, de formes élémentaires f_k considérées comme particulièrement "harmonieuses" et "parfaites", cette idée est très ancienne. Comme nous l'avons noté plus haut à la section 1.3 du chapitre 4, on peut considérer que son ancêtre est la synthèse effectuée par Ptolémée des trajectoires "errantes" des planètes comme compositions au moyen d'épicycles de trajectoires circulaires parfaites. À l'époque moderne, l'idée acquit sa première formulation précise avec la théorie des cordes vibrantes et la décomposition des sons en harmoniques.

Vers 1750, d'Alembert introduisit l'EDP linéaire hyperbolique du second ordre qu'est l'équation des ondes et Daniel Bernoulli introduisit l'idée de superposition de solutions trigonométriques élémentaires (cf. chapitre 9, section 1). Fourier introduisit l'EDP linéaire parabolique du second ordre qu'est l'équation de la chaleur et en fonda la résolution sur l'idée de superposition et de développement en séries trigonométriques. En révélant l'omniprésence de ce genre d'EDP en physique, il fut l'un des principaux pionniers de l'analyse.

Très tôt, on comprit que les fonctions trigonométriques $e_k = e^{ik\theta}$, $k \in \mathbb{Z}$, étaient une famille "d'harmoniques" formant ce qu'on appela ensuite un "système orthonormal",²⁰ à savoir que

$$\int_{\mathbb{S}^1} e^{ik\theta} \overline{e^{i\ell\theta}} \frac{d\theta}{2\pi} = \langle e_k, e_\ell \rangle = \delta_{k,\ell} .$$

Et l'analyse harmonique commença lorsque l'on se rendit compte que, si l'on voulait exprimer une fonction $f(\theta)$ comme une série trigonométrique formelle

$$T_f(\theta) = \sum_{k \in \mathbb{Z}} c_k e_k$$

(sans se préoccuper de problèmes de convergence), alors les coefficients devaient *formellement* être définis par $c_k = \int_{\mathbb{S}^1} f(\theta) e^{-ik\theta} \frac{d\theta}{2\pi}$ car, formellement, à cause de l'orthonormalité,

$$\int_{\mathbb{S}^1} \left(\sum_{\ell \in \mathbb{Z}} c_\ell e^{i\ell\theta} \right) e^{-ik\theta} \frac{d\theta}{2\pi} = c_k = \langle f, e_k \rangle .$$

uniforme signifie que l'on peut permuter les quantificateurs et passer de $\forall\theta \forall\varepsilon \exists\delta$ à $\forall\varepsilon \exists\delta \forall\theta$. Mais comme les U_δ forment un recouvrement ouvert \mathcal{U} de \mathbb{S}^1 et comme \mathbb{S}^1 est compact, on peut extraire de \mathcal{U} un sous-recouvrement *fini* $\bigcup_{i=1}^n U_{\delta_i} = \mathbb{S}^1$ et alors $\delta = \min \delta_i$ permet de satisfaire la continuité uniforme.

20. On considérerait plutôt au début des séries réelles de cos ou de sin puis on considéra naturellement des séries complexes, la formule de De Moivre $e^{ik\theta} = \cos(k\theta) + i \sin(k\theta)$ permettant des calculs beaucoup plus élégants.

Remarque. Même si l'on ne travaille pas dans $L^2(\mathbb{S}^1)$ nous utiliserons la notation du produit scalaire

$$\int_{\mathbb{S}^1} e^{ik\theta} \overline{e^{i\ell\theta}} \frac{d\theta}{2\pi} = \langle e_k, e_\ell \rangle$$

(toujours bien défini) et

$$c_k = \langle f, e_k \rangle = \int_{\mathbb{S}^1} f(\theta) e^{-ik\theta} \frac{d\theta}{2\pi}$$

(quand il est défini). □

Depuis les travaux de Fourier, les c_k s'appellent les coefficients de Fourier et sont souvent notés conventionnellement $c_k = \widehat{f}(k)$, la transformée de Fourier T_f sur \mathbb{Z} étant alors notée \widehat{f} . Ils ont un sens dès que les intégrales les définissant ont un sens, ce qui conduit à s'interroger d'abord sur les conditions d'intégrabilité des fonctions et ensuite sur la convergence et la sommabilité de séries infinies d'intégrales.

De Riemann à Lebesgue, la bonne notion d'intégrabilité fut, nous l'avons vu plus haut à la section 2, longue à bien définir et se développa parallèlement à la généralisation de la notion de fonction. Avec Riemann on considéra des fonctions bien approximables par le haut et par le bas par des fonctions en escalier avec une limite commune bien définie des approximations. Mais, comme nous l'avons vu, c'est avec Lebesgue que la notion d'intégration trouva sa définition moderne avec la conséquence que les f intégrables, ici les $f \in L^1(\mathbb{S}^1)$, peuvent être très "sauvages" puisqu'elles peuvent être changées de façon arbitraire (purent ensembliste) sur des sous-ensembles de mesure nulle. La théorie de l'intégration développée par Lebesgue entre 1902 et 1910 est intimement liée à ses *Leçons sur les séries trigonométriques* de 1904-1905 au Collège de France (publiées en 1906)

"concernant la possibilité d'utiliser les séries de Fourier pour la représentation des fonctions arbitraires".

On obtint ainsi une correspondance, linéaire puisque $\int_{\mathbb{S}^1}$ est linéaire, dite "transformation de Fourier", entre les f considérées et les familles $T = \{c_k\}_{k \in \mathbb{Z}}$.²¹ Et l'analyse fonctionnelle raffinée naquit de l'étude de cette correspondance qui se révéla être extrêmement compliquée. Le passage de $f(\theta)$ aux $c_k = \widehat{f}(k)$ est "l'analyse harmonique" et le passage inverse de $\{c_k\}_{k \in \mathbb{Z}}$ à une f est la "synthèse harmonique". Comme les $c_k \in \mathbb{C}$ sont de la forme $c_k = \rho_k e^{i\alpha_k}$, on voit que l'idée intuitive naïve est que "toute" fonction périodique (dans un sens restant à définir de "toute") peut s'obtenir comme une somme d'harmoniques à condition de moduler (ρ_k) et déphaser (α_k) ceux-ci de façon idoine. Mais une masse énorme de problèmes apparurent dès qu'on

21. On connaissait déjà bien les développements des fonctions $f \in C^\infty$ en séries de Taylor (cf. la section historique 1 du chapitre 9).

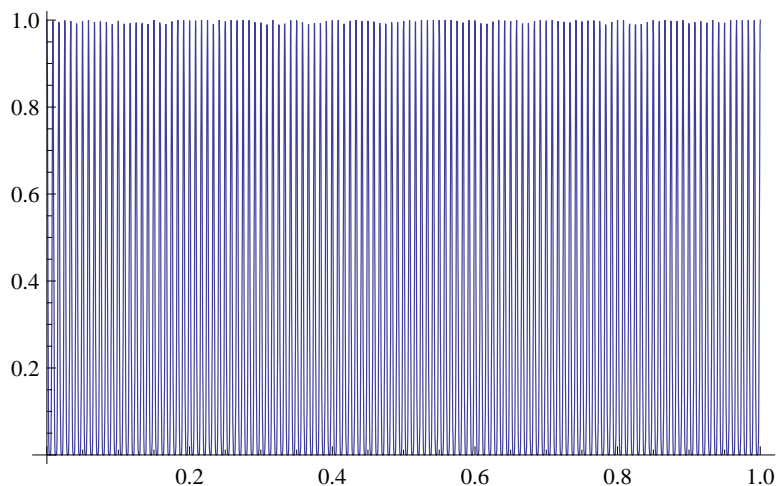


FIGURE 2. La fonction continue $f_{k,j} = \cos(k!\pi x)^{2j}$ sur $[0, 1]$ pour $k = 5, j = 3$. Les $f_{k,j}$ tendent pour $k, j \rightarrow \infty$ vers la fonction de Dirichlet $\delta_{\mathbb{Q}}(x) = 1$ si $x \in \mathbb{Q}$ et $\delta_{\mathbb{Q}}(x) = 0$ si $x \notin \mathbb{Q}$.

essaya d'analyser avec rigueur les propriétés de la transformation T entre certains espaces fonctionnels \mathcal{F} et certains sous-espaces \mathcal{T} de $\mathbb{C}^{\mathbb{Z}}$.

Le problème est facile à comprendre. Considérons la “fonction de Dirichlet” $\delta_{\mathbb{Q}}$ qui est la fonction caractéristique du sous-ensemble \mathbb{Q} de \mathbb{R} , autrement dit $\delta_{\mathbb{Q}}(x) = 1$ si $x \in \mathbb{Q}$ et $\delta_{\mathbb{Q}}(x) = 0$ si $x \notin \mathbb{Q}$. On peut la périodiser si on veut sur $[0, 1]$. Elle est d'une facilité déconcertante à définir et pourtant elle est très “sauvage”. Elle est partout discontinue, elle n'est intégrable-Riemann dans aucun intervalle, elle admet pour périodes tous les nombres rationnels. Et pourtant elle est trivialement bornée, elle est intégrable-Lebesgue ($\delta_{\mathbb{Q}} \in L^1$) d'intégrale = 0 car elle n'est $\neq 0$ que sur \mathbb{Q} qui est un sous-ensemble de mesure nulle de \mathbb{R} ,²² et elle est même la limite explicite des fonctions continues

$$f_{k,j} = \cos(k!\pi x)^{2j}$$

pour $k, j \rightarrow \infty$. La puissance paire $2j$ garantit que $f_{k,j}(x) \in [0, 1]$ et que tous les $f_{k,j}(x) < 1$ pour k constant tendent vers 0. Le facteur de période $k!$ garantit que les $f_{k,j}$ oscillent extrêmement rapidement lorsque k augmente. La figure 2 montre l'exemple $k = 5, j = 3$.

Fourier avait un point de vue plutôt “physicien” de mathématiques appliquées.²³ Il s'intéressait plus à la modélisation mathématique des phénomènes naturels et

22. On voit à quel point l'intégrabilité-Lebesgue généralise l'intégrabilité-Riemann.

23. La tension entre mathématiques “pures” et “appliquées” traverse l'histoire des mathématiques et réapparaît périodiquement. Un exemple est celui des calculs non suffisamment rigoureux des physiciens de la première mécanique quantique comme Dirac qui acquièrent un droit de cité en mathématiques pures après le développement de la théorie des distributions par Laurent Schwartz.

à la bonne résolution numérique des équations qu'à des problèmes profonds de mathématique "pure" comme ceux qui passionnaient les mathématiciens depuis l'Antiquité mais n'avaient guère de sens physique.²⁴ Comme l'a rappelé Jean-Pierre Kahane dans une conférence d'Août 2005 à l'Académie des sciences [275], un mathématicien "pur" comme Jacobi écrivait à Legendre en 1830 après la mort de Fourier que ce dernier s'occupait de "l'explication des phénomènes naturels" mais que

"un philosophe comme lui aurait dû saisir que le but unique de la science, c'est l'honneur de l'esprit humain, et que, sous ce titre, une question de nombres vaut autant qu'une question de système du monde".²⁵

L'évaluation des travaux de Fourier changea à partir du moment où l'on commença à démontrer de façon rigoureuse certains théorèmes de convergence des T_f . À partir de là, les travaux se sont développés dans plusieurs directions :

1. Étudier plus précisément l'influence de la structure des fonctions f (leurs propriétés de continuité, de différentiabilité, la nature de leurs discontinuités ou de leurs oscillations, etc.) sur leurs coefficients de Fourier (par exemple le théorème de Riemann-Lebesgue pour les f intégrables ou le théorème de Cantor-Lebesgue pour les f qui sont C^k).
2. Étudier les propriétés de convergence des séries trigonométriques T et savoir dans quelle mesure elles sont des séries de Fourier T_f de fonctions f d'un certain espace fonctionnel. Le problème est non trivial (contre exemple de Fatou : il existe T convergente qui n'est la T_f d'aucune f intégrable). Si T est une série trigonométrique $T = \sum_{k \in \mathbb{Z}} c_k e^{ik\theta}$, on note $S_n T$ les sommes partielles

$$S_n T = \sum_{|k| \leq n} c_k e^{ik\theta}$$

et l'on étudie la convergence de la suite $S_n T$ ou, ce qui est plus opératoire, des suites de moyennes des $S_n T$ (moyennes de Cesàro chez Féjer, etc.).

3. Étudier les différents types de convergence, en particulier la convergence "ponctuelle" (convergence simple) qui est locale et beaucoup plus contraignante que

24. Savoir que $\sqrt{2}$ est irrationnel ou que π est transcendant ne semble guère avoir d'importance physique évidente puisqu'on peut les approximer aussi bien que l'on veut par des rationnels. Cela a en fait une grande importance mais à un niveau plus profond. Par exemple une droite de pente $\sqrt{2}$ s'enroule densément sur le tore $\mathbb{R}^2/\mathbb{Z}^2$ parce que $\sqrt{2}$ est irrationnel. Plus généralement, dans le "chaos déterministe" on rencontre souvent des systèmes dynamiques dont les trajectoires dépendent de propriétés arithmétiques des conditions initiales. Changer même très peu les conditions initiales change complètement les trajectoires. C'est ce qu'on appelle une "sensitivité" aux conditions initiales.

25. Cf. [275]. Reprise par Hilbert, cette expression de Jacobi "honneur de l'esprit humain" est devenue, comme le dit Kahane, "l'emblème" des mathématiques pures.

les convergences globales en norme L^p (théorème de convergence de Dirichlet pour les f continues C^1 par morceaux, mais contre-exemple de divergence de Du Bois-Reymond pour les f seulement continues, mais aussi, dans ce cas, théorème de convergence de Féjer des moyennes de Cesàro des $S_n T$, théorème de convergence presque partout de Carleson-Hunt pour les $f \in L^p$, $p > 1$).

4. Étudier les “bonnes” situations où il y a équivalence entre les fonctions f et leurs transformées de Fourier $T_f = \widehat{f}$ (espaces L^2).
5. Étudier des contre-exemples “pathologiques” où cette équivalence n’a pas du tout lieu (exemple de Kolmogorov d’une f intégrable pour laquelle T_f ne converge nulle part vers f , découverte de fonctions continues nulle part dérivables, i.e. fractales, etc.).

Pour avancer, il faut avoir conscience de la suite infinie d’espaces fonctionnels emboîtés (par des inclusions topologiques) :

$$\begin{aligned} \mathcal{C}^\infty(\mathbb{S}^1) &\subset \dots \subset \mathcal{C}^k(\mathbb{S}^1) \subset \dots \subset \mathcal{C}^0(\mathbb{S}^1) \subset \\ L^\infty(\mathbb{S}^1) &\subset \dots \subset L^p(\mathbb{S}^1) \subset \dots \subset L^2(\mathbb{S}^1) \subset L^1(\mathbb{S}^1) . \end{aligned}$$

Les inclusions sont topologiques bien que chaque espace soit dense dans le suivant alors que pourtant \mathcal{C}^0 et les L^p sont complets. Il n’y a là aucune contradiction car les topologies de cette suite croissante deviennent de plus en plus grossières. Les \mathcal{C}^k et $L^{q < p}$ ne sont donc pas munis de la topologie induite par celle de L^p mais par une topologie plus fine.

Nous avons vu plus haut les normes L^p . La norme dans \mathcal{C}^0 est

$$\|f\|_{\mathcal{C}^0} = \max_{\theta \in \mathbb{S}^1} |f(\theta)|$$

($|f(\theta)|$ a toujours un maximum car \mathbb{S}^1 est compact), celle dans \mathcal{C}^k est

$$\|f\|_{\mathcal{C}^k} = \max_{\theta \in \mathbb{S}^1} \{ |f(\theta)|, |f'(\theta)|, \dots, |f^{(k)}(\theta)| \} .$$

Les $f \in L^\infty$ sont continues presque partout.

Le premier résultat rigoureux fut, en 1829 (l’année avant la disparition de Fourier), le théorème de convergence ponctuelle de Johann Peter Gustav Lejeune-Dirichlet. Nous allons y revenir. Puis vint le mémoire de thèse de Bernhard Riemann de 1854 “Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe” (publié en 1867 grâce à Richard Dedekind) qui rendit justice à Fourier en comprenant toute l’importance de la transformation $f \longleftrightarrow T_f = \left\{ \widehat{f}(k) \right\}_{k \in \mathbb{Z}}$. En utilisant sa théorie de l’intégration développée à ce propos, Riemann démontra le résultat (généralisé ensuite au *théorème de Riemann-Lebesgue*) que si f est intégrable Riemann (généralisé ensuite à f intégrable Lebesgue, $f \in L^1$) alors ses coefficients de

Fourier

$$c_k = \widehat{f}(k) \xrightarrow{|k| \rightarrow \infty} 0$$

(ce qui n'implique pas du tout que la série formelle T_f converge et, a fortiori, qu'elle converge vers f).²⁶ La réciproque est fautive. Le théorème montre que la correspondance $f \rightarrow T_f$ de L^1 dans l'espace des $\{c_k \in \mathbb{C}\}_{k \in \mathbb{Z}}$ va dans le sous-espace $\mathcal{T}_{\rightarrow 0}$ des $\{c_k\}_{k \in \mathbb{Z}}$ tels que $c_k \xrightarrow{|k| \rightarrow \infty} 0$, mais elle n'est pas surjective. Nous verrons plus bas que

Pierre Fatou a construit un exemple de série T convergente qui ne peut être la série de Fourier d'aucune fonction intégrable, i.e. $T(L^1) \not\subset \mathcal{T}_{cv}$.²⁷ Par ailleurs, Cantor puis Lebesgue montrèrent que $c_k \xrightarrow{|k| \rightarrow \infty} 0$ est une condition nécessaire pour que la

série trigonométrique formelle T de coefficients c_k converge sur un sous-ensemble X de \mathbb{S}^1 aussi petit soit-il à condition qu'il soit de mesure > 0 . On a donc à la fois $T(L^1) \not\subset \mathcal{T}_{\rightarrow 0}$ et $\mathcal{T}_{cv,mes(X)>0} \subset \mathcal{T}_{\rightarrow 0}$.

Le théorème général et basique de Riemann-Lebesgue se démontre en utilisant le fait que $\mathcal{C}^0(\mathbb{S}^1)$ est dense dans $L^1(\mathbb{S}^1)$.²⁸ Soit $g \in \mathcal{C}^0(\mathbb{S}^1)$. Le coefficient

$$\widehat{g}(k) = \int_{-\pi}^{\pi} g(\theta) e^{-ik\theta} \frac{d\theta}{2\pi}$$

ne change pas si on tourne de $\frac{\pi}{k}$ par le changement de variable $\theta = t - \frac{\pi}{k}$ et donc

$$\widehat{g}(k) = \int_{-\pi + \frac{\pi}{k}}^{\pi + \frac{\pi}{k}} g\left(t - \frac{\pi}{k}\right) e^{-ik\left(t - \frac{\pi}{k}\right)} \frac{d\theta}{2\pi}.$$

Mais comme $e^{-ik\left(-\frac{\pi}{k}\right)} = e^{i\pi} = -1$,

$$\widehat{g}(k) = \frac{1}{2} \int_{-\pi}^{\pi} \left(g(\theta) - g\left(\theta - \frac{\pi}{k}\right)\right) e^{-ik\theta} \frac{d\theta}{2\pi}.$$

Et comme g est continue, on peut rendre $|g(\theta) - g\left(\theta - \frac{\pi}{k}\right)|$ aussi petit que l'on veut en prenant k assez grand et par suite $\widehat{g}(k) \xrightarrow{|k| \rightarrow \infty} 0$. Si maintenant $f \in L^1$, on l'approxime par des g_ε continues telles que $\|g_\varepsilon - f\|_1 < \varepsilon$ ce qui implique

$$\left|\widehat{g}_\varepsilon(k) - \widehat{f}(k)\right| \leq \|g_\varepsilon - f\|_1 < \varepsilon$$

et donc $\widehat{f}(k) \xrightarrow{|k| \rightarrow \infty} 0$.

26. Ce théorème a été anticipé par Camille Deffers. Cf. l'article [16] de Silvia Annaratone.

27. Pierre Fatou (1878-1929) fut l'un des fondateurs de la théorie des dynamiques holomorphes où l'on rencontre les belles fractales que sont les ensembles de Julia, théorie relancée dans les années 1980 par Denis Sullivan, Adrien Douady, John Hubbard (cf. le séminaire Bourbaki de Douady [148]), puis ensuite la médaille Fields Jean-Christophe Yoccoz prématurément décédé.

28. Cf. par exemple [363].

Plus généralement, il existe un lien étroit entre les propriétés de régularité (i.e. de continuité et de différentiabilité) de f et la rapidité de la décroissance des $\widehat{f}(k)$. Intuitivement, on voit bien pourquoi : plus les harmoniques de très petite fréquence sont de module évanouissant, plus la fonction sera régulière. On montre que $f \in \mathcal{C}^n$ si les coefficients $\widehat{f}(k)$ décroissent plus vite que $\frac{1}{k^{n+2}}$ et que si $f \in \mathcal{C}^n$ ils décroissent nécessairement plus vite que $\frac{1}{k^n}$. Autrement dit,

$$\mathcal{T}_{\frac{1}{k^{n+2}}} \subset T(\mathcal{C}^n) \subset \mathcal{T}_{\frac{1}{k^n}}.$$

Pour $n \geq 2$ cela implique la convergence absolue de T_f vers f partout puisque la série $\frac{1}{k^2}$ est convergente.²⁹ La preuve est immédiate si on effectue une double intégration par parties, ce qui est possible puisque f est C^2 . Une première intégration par parties donne

$$\widehat{f}(k) = \int_{-\pi}^{\pi} f(\theta) e^{-ik\theta} \frac{d\theta}{2\pi} = \left[f(\theta) \frac{e^{-ik\theta}}{-ik} \right]_{-\pi}^{\pi} + \frac{1}{ik} \int_{-\pi}^{\pi} f'(\theta) e^{-ik\theta} \frac{d\theta}{2\pi}.$$

Le premier terme = 0 et le second donne, par une seconde intégration par parties,

$$\widehat{f}(k) = \frac{1}{ik} \left[f'(\theta) \frac{e^{-ik\theta}}{-ik} \right]_{-\pi}^{\pi} - \frac{1}{k^2} \int_{-\pi}^{\pi} f''(\theta) e^{-ik\theta} \frac{d\theta}{2\pi}.$$

Le premier terme est nul et, comme f'' est continue, donc bornée, $f'' \leq M$, le second terme est en $\frac{1}{k^2}$, $|\widehat{f}(k)| \leq \frac{M}{k^2}$. Cela se généralise immédiatement à $f \in \mathcal{C}^{n \geq 2}$. Si f est seulement C^1 l'argument n'est plus applicable car la série $\frac{1}{k}$, $k \geq 1$, n'est pas convergente.

En fait, on mit assez longtemps à comprendre toute la complexité que pouvaient présenter les séries trigonométriques générales.

On trouva des exemples de fonctions continues “pathologiques” dont la série de Fourier *ne convergeait pas* et pouvait même *diverger* (d'après le théorème de convergence ponctuelle de Dirichlet exposé plus bas, elles ne peuvent pas être C^1 même par morceaux). Le premier, qui fit le tour de l'Europe, est dû à Paul du Bois-Reymond³⁰ en 1873. Il s'agit d'une fonction g continue telle que T_g ne converge

29. La valeur de cette série est le problème dit de Bâle ou de Mengoli (1644). Elle est $\frac{\pi^2}{6}$ d'après un résultat d'Euler de 1741.

30. Membre d'une famille de Huguenots de Berlin venant de Neuchâtel, Paul David Gustave du Bois-Reymond (1831-1889) fut l'élève de Kummer, enseigna à Heidelberg, Fribourg, Tübingen (où il succéda à Henkel et fut le professeur de Hölder), puis Berlin. Il travailla beaucoup sur la complexité de la notion générale de fonction et sur la structure du continu. Inspiré par Weierstrass, avec qui il s'entendait bien, il donna en 1875 un exemple de fonction continue nulle part dérivable. Il s'intéressa aussi beaucoup à la “métaphysique” du continu comme infini actuel et aux paradoxes des infinitésimales. Il avait clairement conscience des difficultés théoriques qui allaient dominer plus tard la question des fondements des mathématiques. Il pensait qu'elles étaient insurmontables et étaient les symptômes des limites de notre intellect.

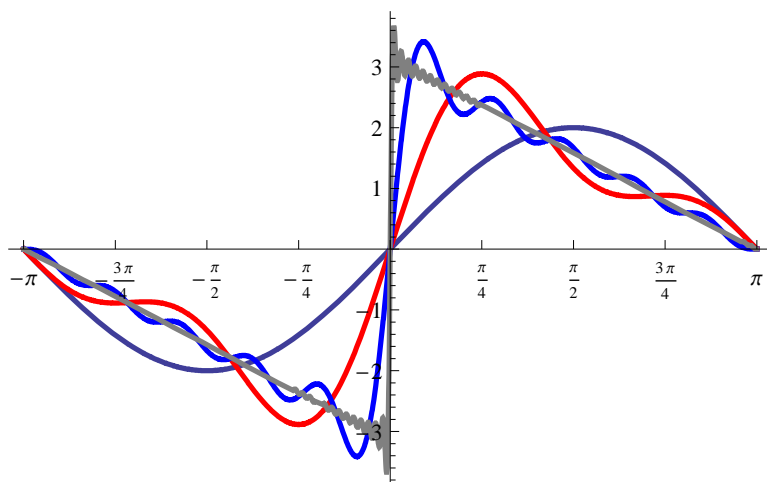


FIGURE 3. Les fonctions 2π -périodiques $\operatorname{im} f_n(\theta) = \sum_{k=1}^{k=n} \frac{2}{k} \sin(k\theta)$ pour $k = 1$ (noir), $k = 3$ (rouge), $k = 10$ (bleu), $k = 100$ (gris).

pas en 0 et même, ce qui est encore plus surprenant, sur un sous-ensemble *dense* de points. Il existe même une sous-suite divergente $S_{n_j}g$ des $S_n g$ telle que $S_{n_j}g(0)$ tende vers $-\infty$ lorsque $n_j \rightarrow \infty$.³¹ À l'époque il s'agissait d'une grande nouveauté. Maintenant, les choses sont plus claires car g est construite comme une fonction fractale.

Reprenons, en l'illustrant, la description de Merker-Marçay [363]. Du Bois-Reymond part des polynômes trigonométriques $f_n(\theta) = \sum_{|k|=1}^{|k|=n} \frac{1}{k} e^{ik\theta}$ (même si $k = 0$ n'est pas dans la somme, c'est le mauvais cas car les $\frac{1}{k}$ tendent bien vers 0 mais la série $\frac{1}{k}$, $k \geq 1$, n'est pas convergente). La valeur $f_n(\theta)$ est imaginaire pure, impaire, égale à $f_n(\theta) = \sum_{k=1}^{k=n} \frac{2i}{k} \sin(k\theta)$ et $f_n(\ell\pi) = 0$. À la limite on obtient une fonction "en dents de scie". La dérivée de $\operatorname{im} f_n(\theta)$ est $\sum_{k=1}^{k=n} 2 \cos(k\theta)$ et $(\operatorname{im} f_n)'(0) = 2n$. La figure 3 en représente quelques exemples.

On montre facilement que les f_n sont de norme bornée $\leq \pi + 4$. On déphase alors les f_n en prenant

$$P_n(\theta) = e^{i2n\theta} f_n(\theta) .$$

La figure 4 montre les parties réelle et imaginaire de $P_{10}(\theta)$. On voit que P_n fonctionne comme une sorte de "paquet d'onde".

On utilise alors le fait que, par construction,

- $S_N(P_n) = P_n$ si $N \geq 3n$,
- $S_N(P_n) = \widetilde{P}_n$ si $N = 2n$ et

31. C'est le début d'une longue histoire menant au théorème de Jean-Pierre Kahane et Yitzhak Katznelson (1965, cf. plus bas).

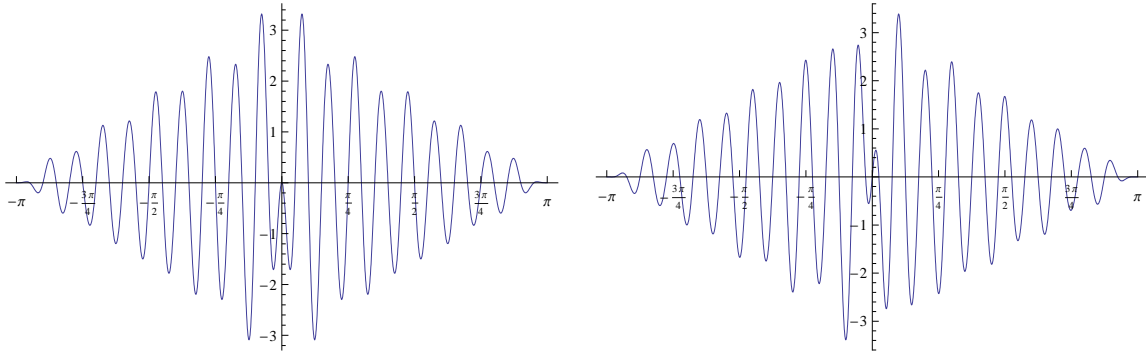


FIGURE 4. Les parties réelle (à gauche) et imaginaire (à droite) de $P_n(\theta) = e^{i2n\theta} f_n(\theta) = e^{i2n\theta} \sum_{|k|=1}^{|k|=n} \frac{1}{k} e^{ik\theta}$ pour $n = 10$.

– $S_N(P_n) = 0$ si $N \leq n - 1$,
 où $\widetilde{P}_n(\theta)$ correspond à la partie $\sum_{k=-n}^{k=-1}$ de f_n , i.e.

$$\widetilde{P}_n(\theta) = e^{i2n\theta} \widetilde{f}_n(\theta) = e^{i2n\theta} \sum_{k=-n}^{k=-1} \frac{1}{k} e^{ik\theta}.$$

En effet P_n ne comprend que les fréquences de n à $3n$ sauf celle $2n$.

On peut changer les P_n d'échelle de façon à les faire tendre en norme vers 0 de façon convergente. L'idée est alors de prendre une suite lacunaire

$$g_J(\theta) = \sum_{j=1}^{j=J} \frac{1}{j^2} P_{n_j}(\theta)$$

avec des n_j choisis astucieusement de façon à ce que les fréquences ne se mélangent pas. Comme $P_n(\theta) = \sum_{|k|=1}^{|k|=n} \frac{1}{k} e^{i(k+2n)\theta}$ ne comprend de fréquences qu'entre n et $3n$, il faut prendre $n_j > 3n_{j-1}$. Et comme $\sum_{j \geq 1} \frac{1}{j^2}$ est convergente (de valeur $\frac{\pi^2}{6}$) et que $|P_n(\theta)| = |f_n(\theta)| \leq \pi + 4$, la série

$$g(\theta) = \sum_{j \geq 1} \frac{1}{j^2} P_{n_j}(\theta)$$

est normalement uniformément convergente et donc $g(\theta)$ est continue. Pour garantir que la série de Fourier T_g de g diverge en 0, il faut préciser les n_j et construire une suite très lacunaire. Considérons $S_N g$ pour $N = 2n_j$ qui ne comprend que les fréquences $|n| \leq N = 2n_j$. Jusqu'à $P_{n_{j-1}}$, les $P_{n_\ell}(\theta)$ satisfont $3n_\ell < 2n_j$ et sont donc entièrement inclus dans $S_N g$. En revanche pour $P_{n_j}(\theta)$ seul $\widetilde{P}_{n_j}(\theta)$ fait partie

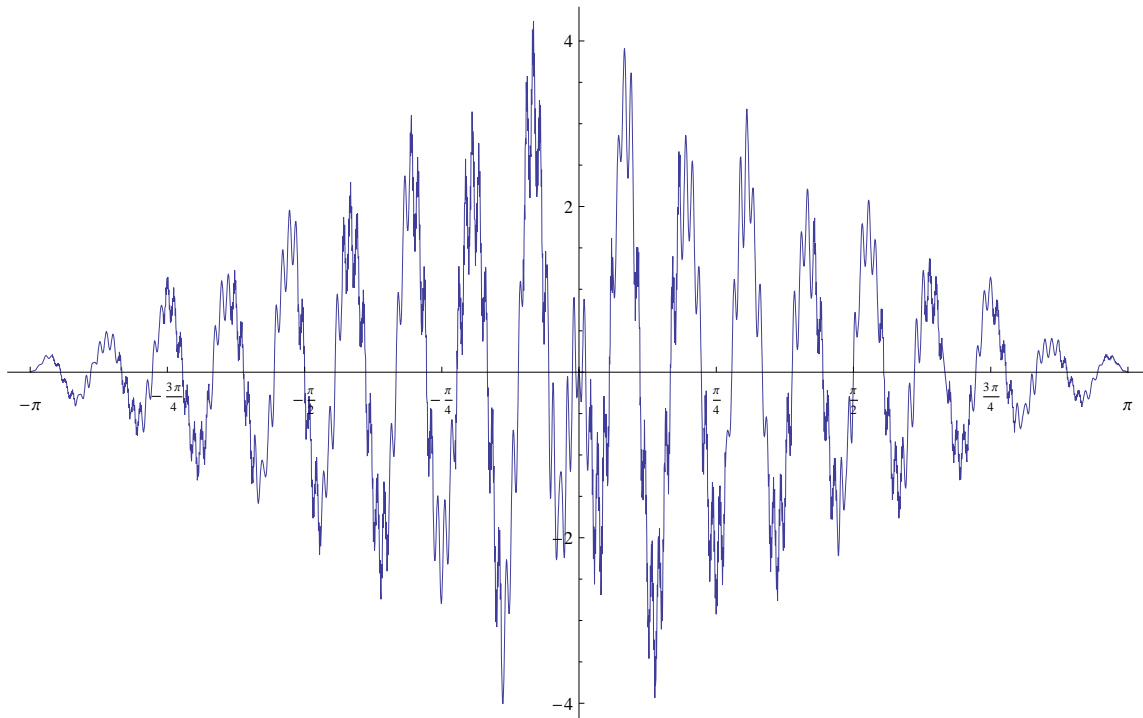


FIGURE 5. La partie réelle de la fonction de du Bois-Reymond $g(\theta) = \sum_{j \geq 1} \frac{1}{j^2} P_{3^{2j}}(\theta) = \sum_{j \geq 1} \frac{1}{j^2} \sum_{\substack{|k|=3^{2j} \\ |k|=1}} \frac{1}{k} e^{i(k+2 \cdot 3^{2j})\theta}$ au niveau $1 \leq j \leq 3$. Il s'agit d'une série trigonométrique de $2 \cdot 3^{2j} = 13122$ termes.

de $S_N g$. On a donc

$$S_N g(\theta) = \sum_{\ell=1}^{\ell=j-1} \frac{1}{\ell^2} P_{n_\ell}(\theta) + \frac{1}{j^2} \widetilde{P}_{n_j}(\theta) .$$

La première somme ne fait pas problème puisqu'elle est de module $\leq \frac{\pi^2}{6}(\pi + 4)$. En revanche, la deuxième somme est en 0 de l'ordre de $-\frac{1}{j^2} \log(n_j)$ et peut donc diverger si on choisit les n_j croissant exponentiellement de façon à contrebalancer la convergence de $\frac{1}{j^2}$. Un choix naturel est alors $n_j = 3^{2j}$, soit

$$g(\theta) = \sum_{j \geq 1} \frac{1}{j^2} P_{3^{2j}}(\theta) = \sum_{j \geq 1} \frac{1}{j^2} \sum_{\substack{|k|=3^{2j} \\ |k|=1}} \frac{1}{k} e^{i(k+2 \cdot 3^{2j})\theta} .$$

Il est représenté (partie réelle) à la figure 5 pour le niveau $J = 3$ qui est déjà computationnellement assez lourd puisque $3^{2j} = 6561$. On y voit très bien la structure fractale de $g(\theta)$.

Plus tard, en 1926, Andrei Kolmogorov donna dans une note [303] aux CRAS (Comptes Rendus de l'Académie des Sciences) un exemple de fonction intégrable $f \in L^1$ dont la série de Fourier T_f diverge *partout*. Ce résultat précisait un résultat analogue de 1923 de divergence presque partout. La construction, assez compliquée, est très bien expliquée dans l'une des "bibles" du sujet (dixit J.-P. Kahane), *Trigonometric Series* [575], d'Antoni Zygmund.³²

On trouva des fonctions continues qui n'étaient dérivables *nulle part* alors qu'on pensait intuitivement que les fonctions continues devaient être dérivables presque partout. Un exemple simple est celui donné par Lebesgue en 1939 de la série très lacunaire

$$f_n(\theta) = \sum_{k=1}^{k=n} \frac{1}{2^k} \sin(2^{k^2} \theta) .$$

Comme la série des $\frac{1}{2^k}$ ($k \geq 1$) est absolument convergente et que \mathbb{S}^1 est compact, les $f_n(\theta)$ convergent vers une fonction continue f mais f n'est dérivable en aucun point. La figure 6 montre f_2 , f_3 et f_4 ainsi que des zooms. On voit que f oscille infiniment partout avec des oscillations de taille infiniment petite mais infiniment "écrasées" l'une sur l'autre. Ce comportement fractal rend impossible toute dérivabilité. On le montre facilement en jouant sur des évaluations de ce que pourrait être la dérivée si elle existait.

Aujourd'hui, de tels exemples sont très bien compris car il s'agit de courbes fractales.

Ceci dit, pour les fonctions f continues les choses ne se passent quand même pas trop horriblement mal car T_f est au moins *unique*. $T_f \equiv 0$ équivaut à $f \equiv 0$ (et donc $T_f = T_g$ équivaut à $f = g$), i.e. $T : \mathcal{C}^0 \rightarrow \mathbb{C}^{\mathbb{Z}}$ est injective. Que $f \equiv 0$ implique $T_f \equiv 0$ est évident. Mais la réciproque n'est pas triviale. Elle a été démontrée et généralisée en 1870-1872 par Georg Cantor dans "Sur l'extension d'un théorème de la théorie des séries trigonométriques" suite à une question posée par Eduard Heine son collègue à Halle. On raisonne par l'absurde. On se ramène au cas f réelle avec $f(0) > 0$ et on construit des *polynômes* trigonométriques $p_n(\theta)$ qui ont la propriété

32. Antoni Zygmund (1900, Varsovie, 1992 Chicago) fut l'un des maîtres de l'analyse harmonique moderne. Brillant mathématicien, il bénéficia en 1929-1930, d'un post-doc lui permettant de se lier avec Godfrey-Harold Hardy à Cambridge et John Edensor Littlewood à Oxford. Professeur à Vilnius dans les années 1930, il dut émigrer en 1940 d'une Pologne et d'une Lituanie dilacérées par les nazis et les communistes et devint l'un des piliers du remarquable département de mathématiques de l'Université de Chicago où travaillèrent entre autres Shiing-Shen Chern, Paul Halmos, Saunders Mac Lane, Irving Segal, Edwin Spanier, Marshall Stone (que nous retrouverons plus bas), André Weil (le nom utopique de "Nancago" associé par Weil à Bourbaki est la contraction de "Nancy" et "Chicago"). Il fut le maître de Józef Marcinkiewicz, Alberto Calderón (avec qui il écrivit de nombreux textes), Elias Stein et Paul Cohen.

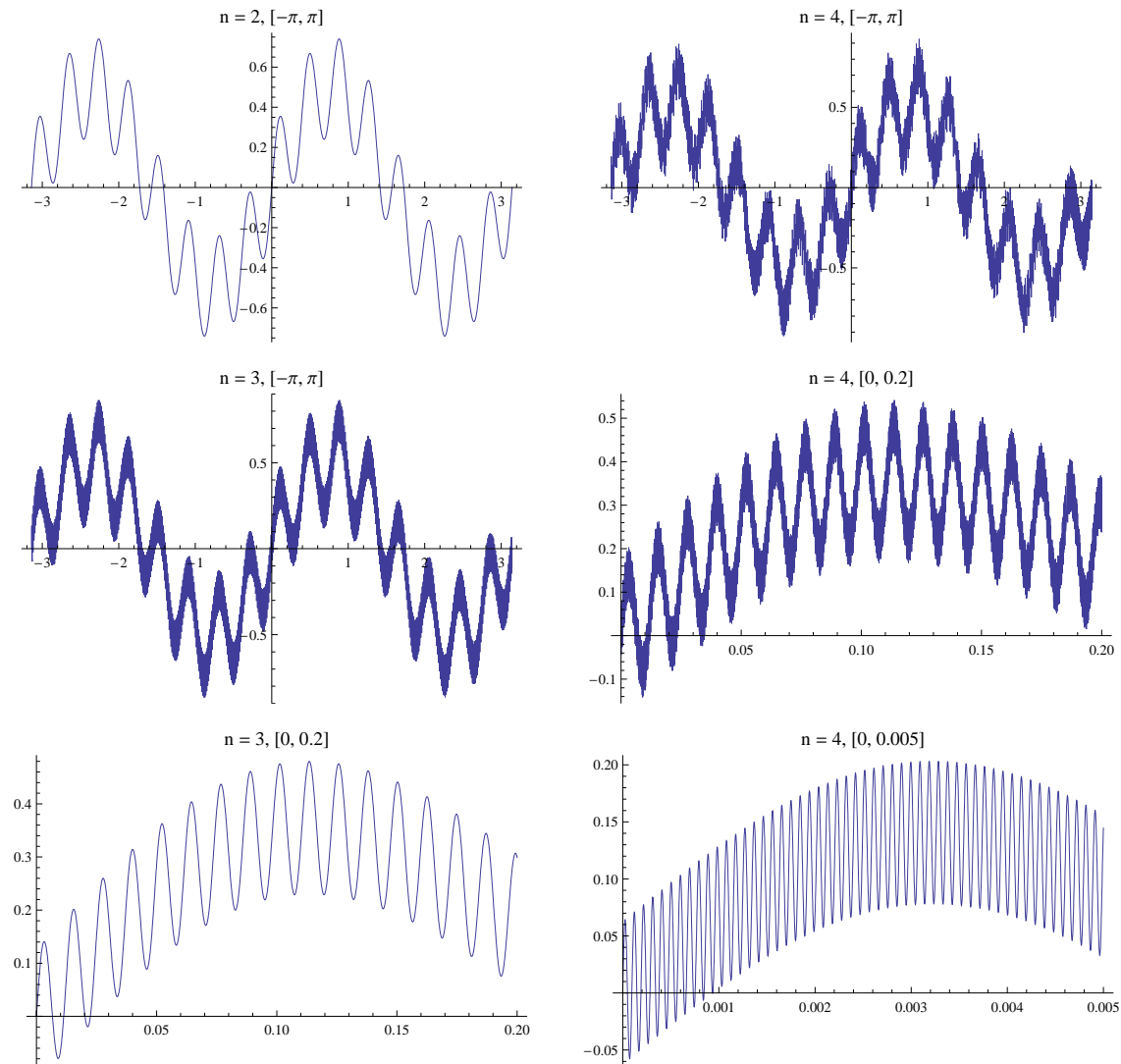


FIGURE 6. Le comportement fractal de la série très lacunaire de Lebesgue $f_n(\theta) = \sum_{k=1}^{k=n} \frac{1}{2^k} \sin(2^{k^2} \theta)$. Les échelles successives sont les $\frac{1}{2^k}$. De haut en bas puis de gauche à droite : (1) $f_2(\theta)$ pour $\theta \in [-\pi, \pi]$. On voit déjà comment chaque partie monotone de $f_1(\theta) = \frac{1}{2} \sin(2\theta)$ devient le support d'une oscillation de type f_1 mais d'échelle inférieure. (2) $f_3(\theta)$ pour $\theta \in [-\pi, \pi]$. On voit que f_2 devient une bande épaisse à l'intérieur de laquelle f_3 oscille à une échelle inférieure. (3) Zoom sur f_3 pour $\theta \in [0, 0.2]$. On voit bien les oscillations de niveau 3. (4) $f_4(\theta)$ pour $\theta \in [-\pi, \pi]$. C'est comme f_3 mais avec des oscillations comportant une échelle de plus. (5) f_4 zoomé sur $\theta \in [0, 0.2]$. est comme f_3 sur $[-\pi, \pi]$. (6) f_4 zoomé sur $\theta \in [0, 0.005]$ rend les oscillations de niveau 4 bien visibles.

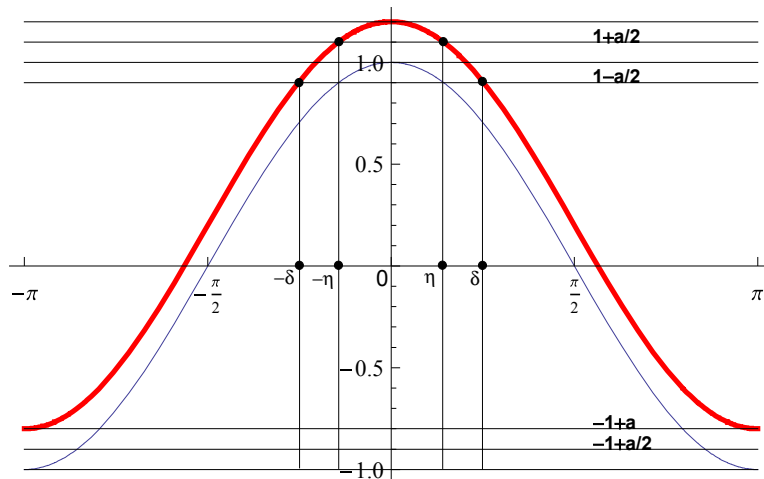


FIGURE 7. La fonction $p(\theta) = a + \cos(\theta)$ pour $a = 0.2$ ($\theta \in [-\pi, \pi]$). $p(\theta) \geq 1 + \frac{a}{2} = 1.1$ sur un petit intervalle ouvert $]-\eta, \eta[$ et $-1 + \frac{a}{2} = -0.9 \leq p(\theta) \leq 0.9$ pour $\theta \in [-\delta, \delta]$, $\delta = \arccos(1 - \frac{3a}{2}) = \arccos(0.7)$.

que $f(0) > 0$ et f continue implique

$$\int_{S^1} f(\theta) p_n(\theta) \frac{d\theta}{2\pi} \xrightarrow{n \rightarrow \infty} \infty .$$

Cela est une contradiction puisque $T_f \equiv 0$ implique que toutes les intégrales

$$\int_{S^1} f(\theta) e^{-ik\theta} \frac{d\theta}{2\pi} = 0$$

alors que les $p_n(\theta)$ sont des combinaisons linéaires finies de $e^{-ik\theta}$. Il est très facile de construire des $p_n(\theta)$ adaptées à f . Comme f est continue en 0, on peut prendre un petit intervalle $|\theta| < \delta$ autour de 0 où $f(\theta)$ reste $\geq \frac{1}{2}f(0)$. On considère alors $p(\theta) = a + \cos(\theta)$, $a > 0$, qui vaut $1 + a$ en 0 et on s'arrange pour que $p(\theta) \geq 1 + \frac{a}{2}$ sur un petit intervalle ouvert $]-\eta, \eta[$ avec $\eta < \delta$ et $|p(\theta)| \leq 1 - \frac{a}{2}$ pour $|\theta| \geq \delta$ (les bornes de θ sont $\pm\pi$). Comme le montre la figure 7, il suffit de prendre a tel que $\delta \geq \arccos(1 - \frac{3a}{2})$.

Il suffit alors de prendre $p_n(\theta) = (p(\theta))^n$ pour obtenir une famille de polynômes trigonométriques conduisant à une contradiction. La figure 8 présente l'évolution des $p_n(\theta)$. On y voit très bien ce qui se passe. Pour $|\theta| < \arccos(1 - a)$, $p(\theta) > 1$ et donc $p_n(\theta) \xrightarrow{n \rightarrow \infty} \infty$, alors que pour $|\theta| > \arccos(1 - a)$, $|p(\theta)| < 1$ et donc $|p_n(\theta)| \xrightarrow{n \rightarrow \infty} 0$ ($p_n(|\arccos(1 - a)|) = 1$).

On décompose alors les intégrales $\int_{-\pi}^{\pi} f(\theta) p_n(\theta) \frac{d\theta}{2\pi}$ en trois parties :

$$\int_{-\pi}^{\pi} = \int_{|\theta| \leq \eta} + \int_{\eta \leq |\theta| \leq \delta} + \int_{\delta \leq |\theta| \leq \pi} .$$

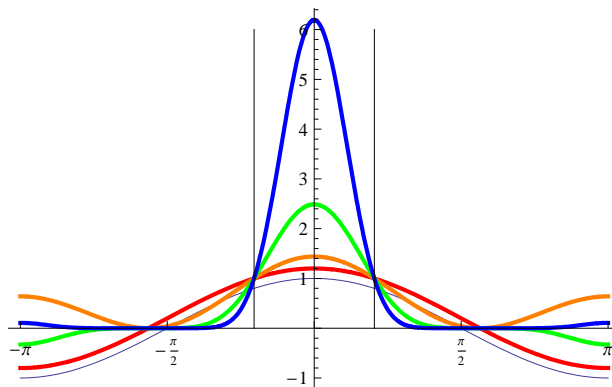


FIGURE 8. Évolution des $p_n(\theta) = (a + \cos(\theta))^n$ pour $a = 0.2$. Pour $|\theta| < \arccos(1 - a) \sim 0.6435$, $p(\theta) > 1$ et $p_n(\theta) \xrightarrow{n \rightarrow \infty} \infty$. Pour $|\theta| > \arccos(1 - a)$, $|p(\theta)| < 1$ et $|p_n(\theta)| \xrightarrow{n \rightarrow \infty} 0$. Évidemment $p_n(|\arccos(1 - a)|) = 1$.

La première intégrale diverge pour $n \rightarrow \infty$ car elle est $\geq 2\eta \frac{f(0)}{2} \left(1 + \frac{a}{2}\right)^n$ et que $1 + \frac{a}{2} > 1$. La seconde intégrale est ≥ 0 car $f(\theta)$ et $p_n(\theta)$ sont ≥ 0 si $|\theta| \leq \delta$. Enfin la troisième intégrale tend vers 0. En effet, f étant continue et \mathbb{S}^1 étant compact $|f|$ est borné par une constante M et donc, puisque $1 - \frac{a}{2} < 1$, $\left| \int_{\delta \leq |\theta| \leq \pi} f(\theta) p_n(\theta) \frac{d\theta}{2\pi} \right| \leq M \left(1 - \frac{a}{2}\right)^n \xrightarrow{n \rightarrow \infty} 0$. En définitive, les intégrales $\int_{-\pi}^{\pi} f(\theta) p_n(\theta) \frac{d\theta}{2\pi}$ tendent bien vers l'infini avec n , ce qui est contradictoire. Donc f est $\equiv 0$.

En particulier, si f est continue et si T_f converge, alors T_f ne peut converger que vers f ponctuellement. Ce sera le cas si par exemple $f \in \mathcal{C}^n$, $n \geq 2$, puisque, dans ce cas, nous avons vu que les c_k décroissent en $\frac{1}{|k|^n}$ et que donc T_f est absolument convergente. Mais pour les espaces plus grands (\mathcal{C}^0 , L^p), cela n'implique pas pour autant que T_f converge ponctuellement vers f . Il faudra attendre le théorème de Carleson disant que si $f \in L^2$, et donc a fortiori si $f \in \mathcal{C}^0$, alors T_f converge presque partout vers f ponctuellement (cf. plus bas).

Remarque. C'est en travaillant autour de ces questions que Cantor fut conduit à la théorie des ensembles. La question générale de l'unicité était de savoir si, étant donnée une série trigonométrique formelle T , le fait qu'elle converge partout vers 0 impliquait que tous ses coefficients c_k étaient nuls. Les sous-ensembles $U \subset \mathbb{S}^1$ ayant la propriété que si $T(\theta) \rightarrow 0$ pour $\theta \in \mathbb{S}^1 - U$ alors $c_k = 0$ pour tout k sont appelés des "ensembles d'unicité". De mesure nulle, ils sont très difficiles à étudier et font intervenir des problèmes d'approximations diophantiennes en théorie des nombres. Pour les comprendre, Cantor fut, entre autres, amené à définir pour un ensemble X de nombres réels, l'ensemble dérivé $X^{(1)}$ de X obtenu en y enlevant les points isolés et à itérer l'opération. Il cherchait à montrer que certains théorèmes restaient vrais même si des hypothèses de convergence et de continuité n'étaient

pas satisfaites sur un X assez petit, c'est-à-dire de n -dérivée $X^{(n)} = \emptyset$ pour un n fini. C'était bien avant les ensembles de mesure nulle de Lebesgue. Il étudia ainsi les $X^{(n)}$ itérés pour eux-mêmes et découvrit que certains X sont si compliqués que, même si l'on itère une infinité de fois cette opération, $X^{(\infty)}$ peut encore posséder des points isolés. Il existe par conséquent des infinis strictement plus grands que l'infini dénombrable et il élaborait la première "arithmétique de l'infini". Il montra très vite, par un argument diagonal, que la puissance du continu (celle de \mathbb{R}) n'est pas dénombrable. Cette épopée de "l'esprit humain" a été sanctifiée par Hilbert et bien analysée par Jean Cavaillès dans ses *Remarques sur la formation de la théorie abstraite des ensembles*.³³ □

Remarque. Nous rencontrerons plusieurs fois des familles de fonctions comme les $p_n(\theta)$ qui tendent vers une distribution comme la distribution de Dirac. Elles sont très faciles à obtenir avec des séries trigonométriques à cause du phénomène fondamental *d'interférences destructrices* dans les phénomènes ondulatoires. Comme nous l'avons déjà noté dans une remarque de la section 9.3 du chapitre 9, lorsque l'on superpose de façon appropriée une infinité d'ondes (ici d'harmoniques), elles peuvent s'annuler réciproquement en dehors d'un "squelette géométrique". Les liens entre optique géométrique (rayons lumineux, caustiques, etc.) et optique ondulatoire ou entre mécanique classique et mécanique quantique sont entièrement fondés sur ce phénomène et le principe dit de la "phase stationnaire". Nous l'avons déjà évoqué dans la section 9.3 du chapitre 9 et nous y reviendrons en détail dans la section 8.7.2 (cf. aussi notre compilation [413] sur les phénomènes critiques). □

Les propriétés fondamentales des espaces de Hilbert L^2 furent découvertes pour les séries trigonométriques. Pierre Fatou démontra en 1906 dans sa thèse *Séries trigonométriques et séries de Taylor* [179] l'égalité de Parseval-Plancherel

$$\langle f, g \rangle = \sum_{k \in \mathbb{Z}} \widehat{f}(k) \widehat{g}(k)$$

et donc

$$\|f\|_{L^2} = \sum_{k \in \mathbb{Z}} |\widehat{f}(k)|^2$$

(théorème de Pythagore) dans le cadre de l'espace de Hilbert L^2 défini au moyen de la mesure de Lebesgue (et pas seulement de l'intégrabilité à la Riemann).³⁴ Le

33. Cf. la thèse [100] de Pierre Cassou-Noguès. Le lecteur intéressé par les problèmes abyssaux de l'hypothèse du continu pourra consulter notre étude [437] sur le platonisme de Kurt Gödel et de Hugh Woodin.

34. Inspiré par des travaux d'Euler, Marc-Antoine Parseval des Chênes (1755-1836) proposa la généralisation du théorème de Pythagore en 1799. Il travaillait sur les EDP linéaires du second ordre hyperboliques et paraboliques avant de s'intéresser à la myrmécologie. Michel Plancherel (1885-1967), éminent mathématicien suisse formé à Göttingen et à Paris et professeur à Fribourg puis à Zurich, généralisa dans les années 1910 la théorie des séries de Fourier

théorème de Fischer-Riesz démontre réciproquement que si $\{c_k\}_{k \in \mathbb{Z}}$ est une suite telle que $\sum_{k \in \mathbb{Z}} |c_k|^2 < \infty$, alors les c_k sont les coefficients d'une $f \in L^2(\mathbb{S}^1)$. La correspondance $f \longleftrightarrow T_f = \widehat{f}$ est donc un isomorphisme *isométrique* entre $L^2(\mathbb{S}^1)$ et l'espace de Hilbert $\ell^2(\mathbb{C})$ des suites $\{c_k\}_{k \in \mathbb{Z}}$ telles que $\sum_{k \in \mathbb{Z}} |c_k|^2 < \infty$. Cela donne une estimation de la vitesse de décroissance des coefficients pour que $f \in L^2$ (nous savons déjà que, comme $f \in L^1$, on a $c_k = \widehat{f}(k) \xrightarrow{|k| \rightarrow \infty} 0$ d'après Riemann-Lebesgue).

Une conséquence immédiate de l'égalité de Parseval est que si $f \in L^2$ alors $T_f = \widehat{f}$ converge nécessairement vers f *au sens de* L^2 . On utilise le fait que \mathcal{C}^0 est dense dans L^2 , que si $g \in \mathcal{C}^0$ alors au niveau des normes $\|g\|_{L^2} \leq \|g\|_{\mathcal{C}^0}$,³⁵ et enfin que les $e^{ik\theta}$ forment une famille totale dans \mathcal{C}^0 (i.e. une famille dont les sommes finies permettent d'approximer aussi bien que l'on veut toute $g \in \mathcal{C}^0$). On écrit alors

$$\|f - S_n T_f\|_{L^2} \leq \|f - g\|_{L^2} + \|g - S_n T_f\|_{L^2} \leq \|f - g\|_{L^2} + \|g - S_n T_f\|_{\mathcal{C}^0} .$$

Comme $\|f\|_{L^2} = \|T_f\|_{L^2}$ par Parseval, on peut trouver des $g \in \mathcal{C}^0$ permettant d'approximer aussi bien que l'on veut f dans L^2 et les $S_n T_f$ dans \mathcal{C}^0 . Donc

$$\lim_{n \rightarrow \infty} \|f - S_n T_f\|_{L^2} = 0 .$$

Mais cette convergence *globale* dans L^2 n'implique évidemment pas a priori une convergence *ponctuelle* (donc maximalement locale) $T_f(\theta) \longrightarrow f(\theta)$ au niveau des valeurs. Certes, si $f \in L^2$, alors il existe une sous-suite $S_{n_k} T_f$ qui converge ponctuellement vers f presque partout. Mais cela n'implique évidemment pas que la suite $S_n T_f$ converge presque partout vers f : il pourrait en effet y avoir d'autres sous-suites soit divergentes, soit convergentes mais convergeant vers une autre valeur que celle de f . En 1913, Nikolai Lusin³⁶ formula la conjecture que les $f \in L^2$ ont une série de Fourier $T_f = \widehat{f}$ qui converge ponctuellement presque partout. En 1966, le grand spécialiste d'analyse harmonique Lennart Carleson (Prix Abel 2006) la démontra (résultat généralisé en 1968 par Richard Hunt aux espaces L^p , $p > 1$). Il s'agit d'un théorème difficile qui ne contredit pas l'exemple "pathologique" de Kolmogorov car ce dernier concerne des fonctions de L^1 . Il implique le résultat pour \mathcal{C}^0 puisque $\mathcal{C}^0 \subset L^2$ topologiquement.

aux espaces de Hilbert généraux, avec également les applications généralisées aux EDP linéaires du second ordre hyperboliques et paraboliques.

35. En effet $\|g\|_{L^2} = \left(\int_{-\pi}^{\pi} |g(\theta)|^2 \frac{d\theta}{2\pi} \right)^{\frac{1}{2}} \leq \max |g(\theta)| = \|g\|_{\mathcal{C}^0}$.

36. Nikolai Lusin (1883-1950) fut l'un des fondateurs (dès 1917) de la théorie "descriptive" des ensembles et de ses applications à l'analyse. Il eut une vie politique difficile à cause du stalinisme.

On peut montrer néanmoins que si f est continue et C^1 par morceaux, alors T_f converge uniformément vers f partout.³⁷ Il s'agit du fameux *théorème de convergence ponctuelle de Dirichlet* (1847).³⁸ De façon plus générale, il s'agit d'un théorème *local* autour d'un point de discontinuité $\theta_0 \in \mathbb{S}^1$ de f . Les hypothèses sont la dérivabilité C^1 à droite et à gauche de f en θ_0 ou la continuité à droite et à gauche et l'existence d'un nombre seulement *fini* d'extrema au voisinage de θ_0 (i.e. f n'oscille pas infiniment au voisinage de 0, ce que l'on appelle la "variation bornée"). La conséquence est dans ce cas que T_f converge en θ_0 et converge qui plus est vers

$$\tilde{f}(\theta_0) = \frac{1}{2} (f(\theta_0)^- + f(\theta_0)^+)$$

(où $f(\theta_0)^-$ et $f(\theta_0)^+$ sont les valeurs limites de f à gauche et à droite de θ_0). $\tilde{f}(\theta_0)$ s'appelle la "régularisée de Dirichlet" au point de discontinuité θ_0 .

Ce théorème est remarquable vu sa date car la notion d'intégration de fonctions un peu "sauvages" n'existait pas encore. Qui plus est, la convergence est très lente lorsqu'il y a une discontinuité. L'exemple standard est celui de la fonction impaire 2π -périodique $h(\theta) = -1$ si $\theta \in]-\pi, 0[$ et $h(\theta) = +1$ si $\theta \in]0, \pi[$ qui est C^∞ sauf en $\{-\pi, 0, \pi\}$ où elle est discontinue.³⁹ Il est trivial de vérifier que $\hat{h}(k) = 0$ si k est pair et $\hat{h}(k) = -\frac{2i}{k\pi}$ si $k = 2\ell + 1$ est impair. Dans ce cas,

$$\hat{h}(-k) e^{-ik\theta} + \hat{h}(k) e^{ik\theta} = \frac{4}{k\pi} \sin(k\theta)$$

et donc

$$T_h(\theta) = \frac{4}{\pi} \sum_{\ell \geq 0} \frac{\sin((2\ell + 1)\theta)}{2\ell + 1}.$$

La figure 9 montre $S_{2n+1}h$ pour $n = 10$ et $n = 100$. On y voit trois choses.

37. C'est également vrai si f est continue et si T_f est absolument convergente.

38. Johann Peter Gustav Lejeune Dirichlet (1805-1859) fut l'un des plus grands mathématiciens de son époque, éminent spécialiste de théorie des nombres (il démontra le dernier théorème de Fermat pour $n = 5$ à vingt ans) et des séries de Fourier. Il généralisa la notion de fonction et étudia par exemple les propriétés d'intégrabilité de la fonction qui porte son nom (cf. plus haut). Venu jeune de Prusse à Paris (1822-1826) pour approfondir ses connaissances mathématiques, il y rencontra Fourier. De retour en Prusse il fut fortement soutenu par Alexander von Humboldt et Gauss et fut professeur à Berlin puis à Göttingen où il succéda à Gauss en 1855. Grand ami de Carl Jacobi et Joseph Liouville, il soutint Ernst Kummer (1810-1893) et fut proche de plus jeunes génies comme Richard Dedekind (1831-1916) qui éditera certaines de ses œuvres ou Bernhard Riemann (1826-1866) qui lui succédera à sa chaire de Göttingen. Marié à Rebecka Mendelssohn (la sœur de Felix) il est un représentant parfait de cette intelligentsia où se croisaient savants et artistes.

39. C'est un avatar de la fonction de Heaviside qui est une primitive de la distribution de Dirac.

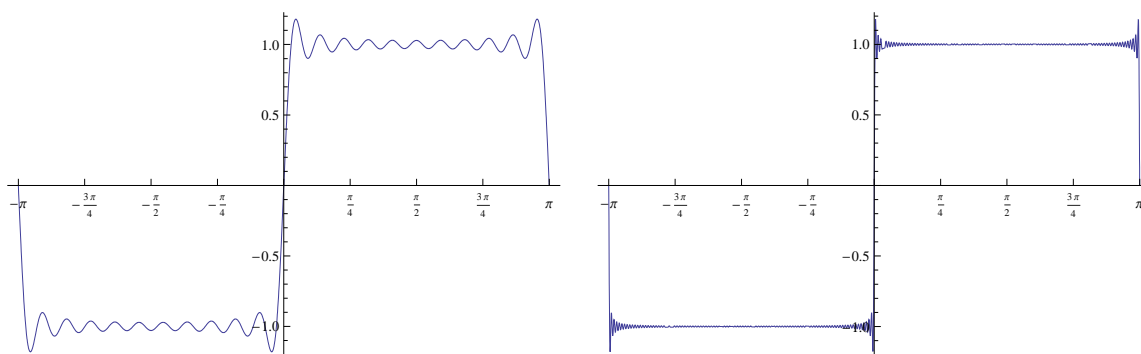


FIGURE 9. La convergence de la série de Fourier de la fonction impaire 2π -périodique $h(\theta) = -1$ si $\theta \in]-\pi, 0[$ et $h(\theta) = +1$ si $\theta \in]0, \pi[$ qui est C^∞ sauf en $\{-\pi, 0, \pi\}$ où elle est discontinue. La figure montre $S_{2n+1}h$ pour $n = 10$ et $n = 100$. La convergence est bonne sauf aux discontinuités où se manifeste le phénomène de Wilbraham-Gibbs-Bôcher.

1. $T_h(\theta)$ converge bien vers $\frac{1}{2}(h(\theta)^- + h(\theta)^+) = 0$ aux discontinuités (voir l'image de gauche).
2. $T_h(\theta)$ converge correctement vers $h(\theta)$ dès que l'on s'éloigne un peu des discontinuités (voir l'image de droite).
3. Mais $T_h(\theta)$ converge beaucoup plus mal vers $h(\theta)$ au voisinage des discontinuités.

Dans le Vol I, chapitre 3, section 4, nous avons expliqué que ce mauvais comportement des transformées de Fourier au voisinage des discontinuités était à l'origine du développement des transformées de Gabor et des transformées en ondelettes. L'analyse de Fourier est fréquentielle et décompose les fonctions en ondes planes qui sont fréquentiellement déterminées et spatialement complètement indéterminées (principe de Heisenberg). Elle est donc très mal adaptée à la synthèse de structures spatiales très localisées comme les discontinuités et les singularités puisque cette information s'y trouve complètement délocalisée dans les coefficients. Si elle est si essentielle dans de nombreux domaines c'est parce que les fonctions trigonométriques $\cos(kt)$ et $\sin(kt)$ sont les fonctions propres de l'opérateur laplacien $\frac{\partial^2 f}{\partial t^2}$.

La mauvaise convergence (3) ci-dessus n'est cependant pas trop "sauvage". Elle correspond à un phénomène a priori énigmatique, dit "phénomène de Gibbs", découvert en 1848 par Henry Wilbraham puis clarifié par Gibbs en 1899 et Bôcher en 1906 (qui lui donna ce nom). Ce phénomène est valable pour les f qui sont C^1 avec un saut $\Delta = f(\theta_0)^+ - f(\theta_0)^-$ en θ_0 (supposons $f(\theta_0)^+ - f(\theta_0)^- > 0$). Il est que, même à la limite, l'amplitude des oscillations est supérieure à Δ , monte à droite jusqu'à $f(\theta_0)^+ + G\Delta$ et descend à gauche jusqu'à $f(\theta_0)^- - G\Delta$ où G est une constante universelle dérivable de l'intégrale $\int_0^\pi \frac{\sin(\theta)}{\theta} d\theta$. Il est dû au fait que

pour approximer la discontinuité il faut de très fortes oscillations qui ne peuvent pas complètement s'éliminer. La convergence de T_f vers f est uniforme sur les intervalles où f est continue C^1 mais pas du tout uniforme en θ_0 . T_f converge mais seulement simplement et vers la régularisée de Dirichlet $\tilde{f}(\theta_0) = \frac{1}{2}(f(\theta_0)^- + f(\theta_0)^+)$.

La démonstration du théorème de Dirichlet se fait en utilisant la notion de "noyau de convolution". L'idée consiste à écrire les sommes partielles $S_n T_f$ sous la forme d'un *produit de convolution* $D_n * f$. On pourra alors intuitivement penser T_f comme la limite des $D_n * f$ pour $n \rightarrow \infty$ et donc comme une convolution $D_\infty * f$ si D_∞ est bien défini.⁴⁰

Supposons que l'on saute plus d'un siècle, que l'on passe de Dirichlet à Schwartz et que l'on puisse montrer (mais ce n'est hélas pas le cas) que $D_\infty = \delta$ la distribution de Dirac. Alors comme δ est l'élément neutre du produit de convolution, on aurait $T_f = D_\infty * f = \delta * f = f$. Les difficultés de sommation et de convergence vers f de T_f peuvent donc se lire sur la limite des D_n . Par définition,

$$D_n(\theta) = \sum_{k=-n}^{k=n} e^{ik\theta} = 1 + 2 \sum_{k=1}^{k=n} \cos(k\theta)$$

et

$$\begin{aligned} S_n f(\theta) &= \sum_{k=-n}^{k=n} \hat{f}(k) e^{ik\theta} = \sum_{k=-n}^{k=n} \left(\int_{-\pi}^{\pi} f(t) e^{-ikt} \frac{dt}{2\pi} \right) e^{ik\theta} = \int_{-\pi}^{\pi} f(t) \sum_{k=-n}^{k=n} e^{ik(\theta-t)} \frac{dt}{2\pi} \\ &= \int_{-\pi}^{\pi} f(t) D_n(\theta-t) \frac{dt}{2\pi} = \int_{-\pi}^{\pi} f(\theta-t) D_n(t) \frac{dt}{2\pi} \\ &= D_n * f. \end{aligned}$$

Si θ est un multiple de 2π , $D_n(2\ell\pi) = 2n + 1$. Sinon, $D_n(t) = \frac{\sin((n+\frac{1}{2})t)}{\sin(\frac{t}{2})}$. En effet (exercice devenu ultra scolaire),

$$\begin{aligned} D_n(t) &= e^{-int} (1 + e^{it} + \dots + e^{i2nt}) = e^{-int} \frac{e^{i(2n+1)t} - 1}{e^{it} - 1} = \frac{e^{i(n+\frac{1}{2})t} - e^{-i(n+\frac{1}{2})t}}{e^{i\frac{t}{2}} - e^{-i\frac{t}{2}}} \\ &= \frac{\sin((n+\frac{1}{2})t)}{\sin(\frac{t}{2})}. \end{aligned}$$

La figure 10 montre quelques exemples de $D_n(t)$. On observe leur forte oscillation.

40. Selon Jean-Pierre Kahane, la notion de noyau de Dirichlet a été anticipé en 1819 par Camille Deflers, mais dans un article insuffisamment rigoureux.

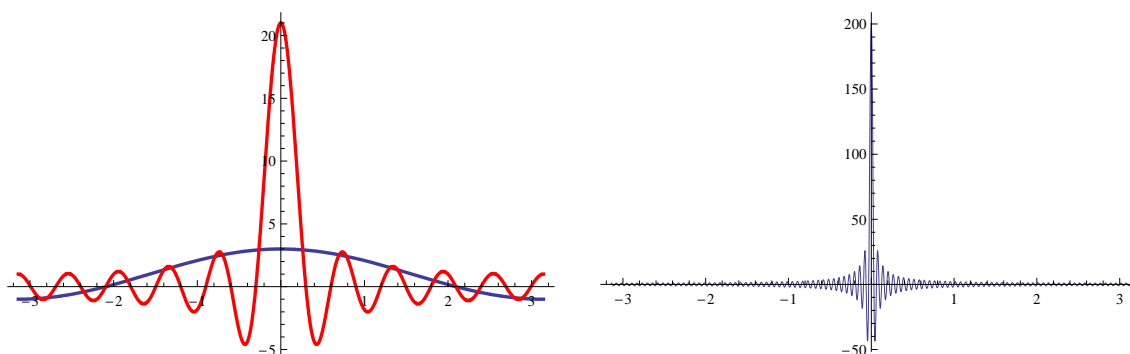


FIGURE 10. Quelques exemples de noyaux de Dirichlet $D_n(t)$. À gauche, $n = 1$ (en noir, de maximum 3) et $n = 10$ (en rouge, de maximum 21). À droite $n = 100$ (de maximum 201).

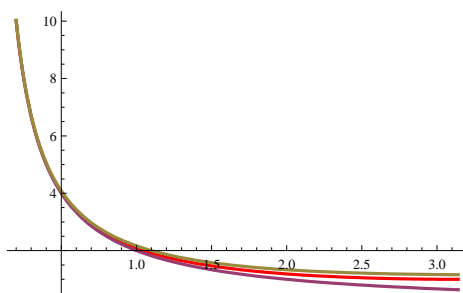


FIGURE 11. Encadrement $\frac{2}{t} \leq \frac{1}{\sin(\frac{t}{2})} \leq \frac{2}{t} + \frac{t}{6}$ de $\frac{1}{\sin(\frac{t}{2})}$ sur $[0, \pi]$.

$D_n(t)$ est une fonction paire (quotient de deux fonctions impaires) et sa moyenne est toujours 1. Elle est très oscillante⁴¹ et de norme L^1

$$\|D_n\|_1 = \int_{-\pi}^{\pi} |D_n(t)| \frac{dt}{2\pi}$$

divergente. Mais la divergence est lente, de type logarithmique. En fait, on montre que lorsque $n \rightarrow \infty$ $\|D_n\|_1$ croît comme $\frac{4}{\pi^2} \log(n)$ à une constante près. Cela se montre en utilisant l’encadrement, illustré à la figure 11,

$$\frac{2}{t} \leq \frac{1}{\sin(\frac{t}{2})} \leq \frac{2}{t} + \frac{t}{6}$$

de $\frac{1}{\sin(\frac{t}{2})}$ sur $[0, \pi]$.

41. Dès $n = 23$, notre version de *Mathematica* détecte un “highly oscillating integrand”.

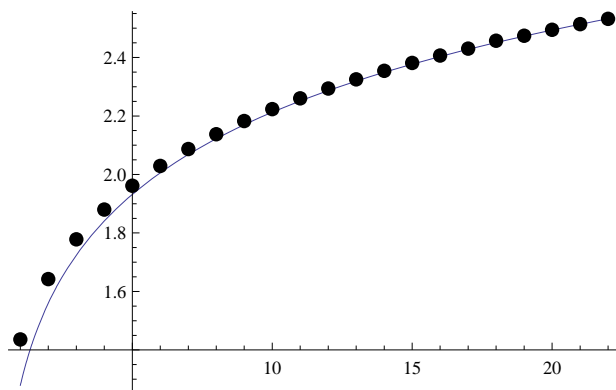


FIGURE 12. La suite des normes $\|D_n\|_1$ des noyaux de Dirichlet de $n = 1$ à $n = 22$ et la courbe $\frac{4}{\pi^2} \log(n) + C$ avec $C = \|D_{22}\|_1 - \frac{4}{\pi^2} \log(22) \sim 1.27947$.

Cet encadrement implique (puisque que D_n est paire)

$$\|D_n\|_1 - 2 \int_0^\pi \frac{2 \left| \sin \left(\left(n + \frac{1}{2} \right) t \right) \right|}{t} \frac{dt}{2\pi} = \|D_n\|_1 - \frac{2}{\pi} \int_0^\pi \frac{\left| \sin \left(\left(n + \frac{1}{2} \right) t \right) \right|}{t} dt \leq \frac{\pi}{12} .$$

Mais par le changement de variable $\left(n + \frac{1}{2} \right) t = u$, $\int_0^\pi \frac{\left| \sin \left(\left(n + \frac{1}{2} \right) t \right) \right|}{t} dt$ devient

$$\int_0^{(n+\frac{1}{2})\pi} \frac{\left| \sin(u) \right|}{u} du$$

La fin de l'intégrale $\int_{n\pi}^{(n+\frac{1}{2})\pi}$ tend vers 0 et il reste, en découpant $\int_0^{n\pi}$ en n intégrales $\int_{k\pi}^{(k+1)\pi}$,

$$\sum_{k=0}^{k=n-1} \int_{k\pi}^{(k+1)\pi} \frac{\left| \sin(u) \right|}{u} du = \sum_{k=0}^{k=n-1} \int_0^\pi \frac{\left| \sin(k\pi + v) \right|}{k\pi + v} dv = \sum_{k=0}^{k=n-1} \int_0^\pi \frac{\sin(v)}{k\pi + v} dv .$$

Le terme $k = 0$ donne le sinus intégral de π , $\text{si}(\pi) \sim 1.85194$ et les autres termes donnent la somme alternée des $\text{si}((k+1)\pi) - \text{si}(k\pi)$. On obtient ainsi une série en $\frac{1}{k}$ qui diverge en $\log(n)$. La figure 12 représente $\|D_n\|_1$ de $n = 1$ à $n = 22$ et y ajuste $\frac{4}{\pi^2} \log(n) + C$ avec $C = \|D_{22}\|_1 - \frac{4}{\pi^2} \log(22) \sim 1.27947$.

En ce sens, les D_n ne constituent pas un “bon” noyau de convolution et, comme l'affirment Joël Merker et François de Marçay ([363], p. 28),

“Là gît la raison profonde pourquoi les théorèmes de convergence pour les séries de Fourier sont difficiles.”

Revenons au théorème de Dirichlet. On a donc

$$S_n f(\theta_0) = D_n * f = \int_{-\pi}^\pi \sin \left(\left(n + \frac{1}{2} \right) t \right) \frac{f(\theta_0 - t)}{\sin \left(\frac{t}{2} \right)} \frac{dt}{2\pi}$$

le terme $\frac{f(\theta_0-t)}{\sin(\frac{t}{2})}$ faisant problème pour $t \rightarrow 0$. Mais, en réarrangeant les termes et en utilisant le fait que D_n est de moyenne 1, on peut écrire $S_n f(\theta_0) - \tilde{f}(\theta_0)$ avec des termes $\frac{f(\theta_0-t)-f(\theta_0^-)}{\sin(\frac{t}{2})}$ et $\frac{f(\theta_0+t)-f(\theta_0^+)}{\sin(\frac{t}{2})}$ et alors les hypothèses d'existence de limites des dérivées à droite et à gauche permettent de conclure que $S_n f(\theta_0) - \tilde{f}(\theta_0) \xrightarrow[n \rightarrow \infty]{} 0$.

Ce théorème fut ensuite globalisé et généralisé aux fonctions continues par morceaux et aux f seulement intégrables mais à variation bornée (Jordan, 1881). Il fournit un moyen puissant de démontrer des égalités connues depuis longtemps et d'en démontrer de nouvelles.

L'exemple scolaire standard est de considérer $f(\theta) = \theta^2$ qui est C^∞ . Ses coefficients de Fourier sont $c_0 = \frac{\pi^2}{3}$ et $c_k = (-1)^k \frac{2}{k^2}$. En 0 tous les $e^{ik0} = 1$ et on obtient donc $\frac{\pi^2}{3} + \sum_{k \geq 1} (-1)^k \frac{4}{k^2} = 0$, c'est-à-dire

$$\sum_{k \geq 1} (-1)^{k+1} \frac{1}{k^2} = \frac{\pi^2}{12}.$$

Pour $x = \pi$, $e^{ik\pi} = (-1)^k$ et donc $\frac{\pi^2}{3} + \sum_{k \geq 1} \frac{4}{k^2} = \pi^2$, c'est-à-dire

$$\sum_{k \geq 1} \frac{1}{k^2} = \frac{1}{4} \left(\pi^2 - \frac{\pi^2}{3} \right) = \frac{\pi^2}{6}$$

(problème de Bâle résolu par Euler en 1741, cf. plus haut)

Notons que les propriétés du noyau de convolution de Dirichlet permettent de démontrer facilement dans un contexte moderne d'analyse fonctionnelle qu'il existe des fonctions continues $f \in \mathcal{C}^0$ telles que T_f soit divergente en au moins un point θ_0 . En effet, pour θ fixé, l'opérateur $\theta \rightarrow S_n f(\theta)$ est un opérateur sur \mathcal{C}^0 muni de la norme sup qui est de norme $\|D_n\|_1$. Or celle-ci tend vers l'infini avec n . D'après le théorème de Banach-Steinhaus (cf. plus bas section 7.1), il existe donc une $f \in \mathcal{C}^0$ telle $\sup_{n \rightarrow \infty} |S_n f(\theta)| = +\infty$. Ce théorème a été fortement généralisé ensuite par Jean-Pierre Kahane et Yitzhak Katznelson en 1965 : pour tout sous-ensemble X Lebesgue-négligeable de \mathbb{S}^1 il existe une f continue telle que sa série de Fourier T_f diverge partout sur X . Il s'agit de la possibilité maximale de divergence puisque d'après le théorème de Carleson, comme $f \in \mathcal{C}^0 \subset L^2$, sa série de Fourier T_f converge ponctuellement presque partout, i.e. en dehors d'un X Lebesgue-négligeable.

Les travaux de Lipót Fejér⁴² entre 1900 et 1904 furent essentiels car ils ont permis d'améliorer considérablement les résultats de convergence en introduisant des méthodes d'approximation par des "bons" noyaux de convolution qui sont, comme

42. Spécialiste d'analyse harmonique, Lipót Fejér (alias Leopold Weiss, 1890-1959) fut un éminent mathématicien hongrois ayant eu comme doctorants des génies comme John von Neumann, Paul Erdős ou George Pólya.

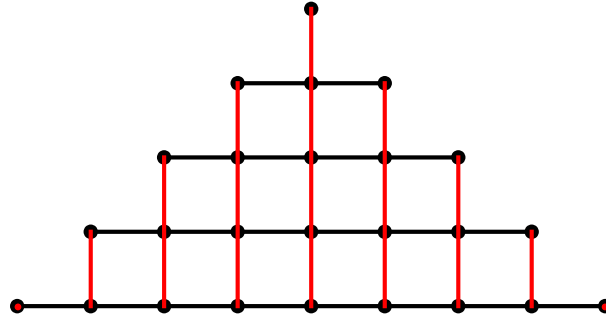


FIGURE 13. Schéma de la double somme $\sum_{n=0}^{n=N-1} \sum_{k=-n}^{k=n}$. Les sommes partielles $\sum_{k=-n}^{k=n}$ s'effectuent le long des lignes horizontales. Mais on peut également sommer d'abord le long des lignes verticales (rouges).

on dit, des “approximations de l'identité” (à l'instar la distribution δ de Dirac comme limite de gaussiennes). Il a introduit et utilisé les limites de Cesàro

$$\sigma_N T = \frac{1}{N} \sum_{n=0}^{n=N-1} S_n T$$

des sommes partielles des séries trigonométriques. De même que les $S_n f$ sont les convolutions $D_n * f$ de f avec les noyaux de Dirichlet D_n , les $\sigma_N f$ sont les convolutions $F_N * f$ de f avec les noyaux de Fejér

$$F_N = \frac{1}{N} \sum_{n=0}^{n=N-1} D_n .$$

Si θ est un multiple de 2π , $D_n(2\ell\pi) = 2n + 1$ et donc

$$F_N(2\ell\pi) = \frac{1}{N} \sum_{n=0}^{n=N-1} (2n + 1) = N .$$

Sinon, on calcule facilement deux expressions de F_N à partir des deux expressions de D_n .

Avec $D_n = \sum_{k=-n}^{k=n} e^{ikt}$, on peut réorganiser les deux sommes conformément à la figure 13. On obtient alors

$$F_N = \frac{1}{N} \sum_{n=0}^{n=N-1} \left(\sum_{k=-n}^{k=n} e^{ik\theta} \right) = \sum_{k=-N}^{k=N} \left(1 - \frac{|k|}{N} \right) e^{ikt} .$$

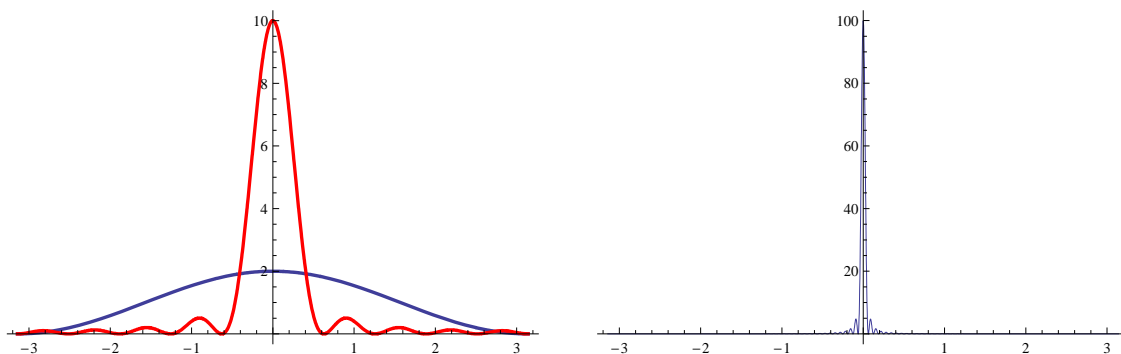


FIGURE 14. Quelques exemples de noyaux de Fejér $F_N(t)$. À gauche, $N = 2$ (en noir, de maximum 2) et $N = 10$ (en rouge, de maximum 10). À droite $N = 100$ (de maximum 100).

Avec $D_n = \frac{\sin((n+\frac{1}{2})t)}{\sin(\frac{t}{2})}$ on trouve alternativement

$$\begin{aligned}
 F_N &= \frac{1}{N} \sum_{n=0}^{n=N-1} D_n = \frac{1}{N} \sum_{n=0}^{n=N-1} \frac{\sin((n+\frac{1}{2})t)}{\sin(\frac{t}{2})} = \frac{1}{N} \frac{1}{\sin(\frac{t}{2})} \sum_{n=0}^{n=N-1} \operatorname{im} e^{i(n+\frac{1}{2})t} \\
 &= \frac{1}{N} \frac{1}{\sin(\frac{t}{2})} \sum_{n=0}^{n=N-1} \operatorname{im} e^{int} e^{i\frac{t}{2}} = \frac{1}{N} \frac{1}{\sin(\frac{t}{2})} \operatorname{im} \left(\frac{e^{iNt} - 1}{e^{it} - 1} e^{i\frac{t}{2}} \right) \\
 &= \frac{1}{N} \frac{1}{\sin(\frac{t}{2})} \operatorname{im} \left(e^{iN\frac{t}{2}} \frac{e^{iN\frac{t}{2}} - e^{-iN\frac{t}{2}}}{e^{i\frac{t}{2}} - e^{-i\frac{t}{2}}} \right) = \frac{1}{N} \frac{1}{\sin(\frac{t}{2})} \operatorname{im} \left(e^{iN\frac{t}{2}} \frac{\sin(\frac{Nt}{2})}{\sin(\frac{t}{2})} \right) \\
 &= \frac{1}{N} \left(\frac{\sin(\frac{Nt}{2})}{\sin(\frac{t}{2})} \right)^2.
 \end{aligned}$$

On voit donc que les F_N sont des fonctions paires ≥ 0 et de moyenne = 1 puisque les D_n sont de moyenne = 1 ($F_1 \equiv 1$), mais également, et cela contrairement aux D_n , de norme $\|F_N\|_1 = 1$ (puisque elles sont ≥ 0 leur norme est leur moyenne). Comme D_n , elles se concentrent sur 0 lorsque $N \rightarrow \infty$ et tendent vers la distribution de Dirac δ . Elles constituent, nous l'avons dit, une "approximation de l'identité" et par conséquent, contrairement aux D_n , de "bons" noyaux de convolution. La figure 14 en donne quelques exemples. On observe encore leur forte oscillation.

Remarque. Nous verrons dans le prochain chapitre 17, section 4, consacré aux "états cohérents" la vaste généralisation de ces concepts de noyau de convolution et d'approximation (ou de "résolution") de l'identité. \square

En utilisant ces "bons" noyaux, Lipót Fejér a pu montrer en 1900 son célèbre théorème généralisant celui de Dirichlet en y suspendant la contrainte C^1 et disant

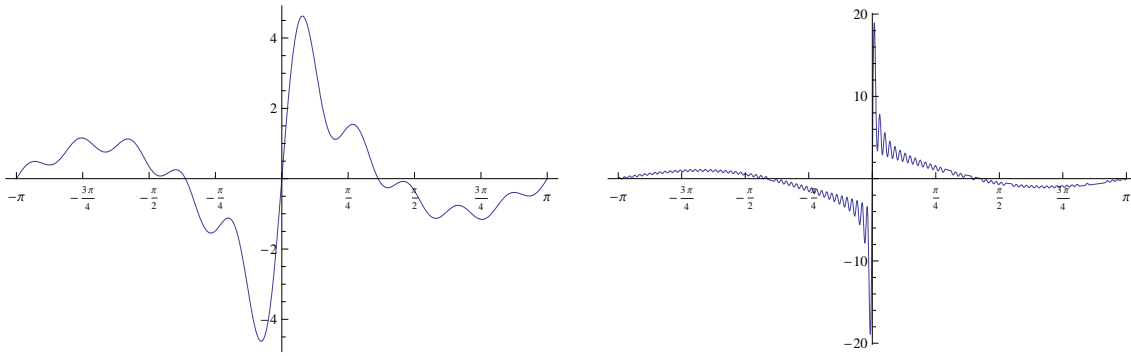


FIGURE 15. Les sommes partielles pour $n = 10$ et $n = 100$ de la série trigonométrique de Fatou $T(\theta) = \sum_{k \geq 2} \frac{\sin(k\theta)}{\log(k)}$ de coefficients $c_k \xrightarrow[|k| \rightarrow \infty]{} 0$ qui ne peut être la série de Fourier T_f d'aucune f intégrable $f \in L^1(\mathbb{S}^1)$.

que si $f \in \mathcal{C}^0(\mathbb{S}^1) \subset L^\infty(\mathbb{S}^1)$, alors la suite des $\sigma_N f$ converge simplement uniformément partout vers f puisqu'elle converge vers la limite des $\sigma_N f = F_N * f$, que la limite des F_N est δ et que $\delta * f = f$. On a $\|\sigma_N f\|_\infty \leq \|f\|_\infty$. Si $f \in L^p(\mathbb{S}^1)$, alors les $\sigma_N f = F_N * f \rightarrow f$ en norme et $\|\sigma_N f\|_p \leq \|f\|_p$. Mais il faut souligner que cela *n'implique pas* que la série de Fourier T_f converge puisque T_f est la limite des $S_n f$ et pas de leurs moyennes de Césàro $\sigma_N f$. D'ailleurs, nous avons vu plus haut des résultats de divergence des T_f pour les f continues.

Les noyaux de Fejér permettent de comprendre le contre-exemple de 1906 de Pierre Fatou exhibant une série trigonométrique T de coefficients $c_k \xrightarrow[|k| \rightarrow \infty]{} 0$ qui n'est la série de Fourier T_f d'aucune $f \in L^1$. Cet exemple explicite,

$$T(\theta) = \sum_{k \geq 2} \frac{\sin(k\theta)}{\log(k)},$$

fait intervenir une décroissance très lente, logarithmique, des c_k . La figure 15 en représente les sommes partielles pour $n = 10$ et $n = 100$. On note que $T(0) = T(\pm\pi) = 0$.

La démonstration est assez subtile (cf. [363]). On montre d'abord (exercice scolaire de trigonométrie) que pour $\theta \neq 2\ell\pi$,

$$\left| \sum_{k=p \geq 2}^{k=q \geq p} \sin(k\theta) \right| \leq \frac{1}{\left| \sin\left(\frac{\theta}{2}\right) \right|}$$

(qui est indépendant de p, q). Un critère de convergence classique dû à Abel permet alors de conclure que $T(\theta)$ est convergente partout. Supposons maintenant que $T(\theta)$

converge vers $g(\theta) \in L^1$, autrement dit que $T = T_g$.⁴³ Comme

$$T(\theta) = \sum_{k \geq 2} \frac{\sin(k\theta)}{\log(k)} = - \sum_{k \leq -2} \frac{1}{2i \log(-k)} e^{-ik\theta} + \sum_{k \geq 2} \frac{1}{2i \log(k)} e^{ik\theta},$$

on voit que les trois premiers coefficients de Fourier de g sont nuls : $\widehat{g}(0) = \widehat{g}(\pm 1) = 0$. On utilise ensuite le fait que $\widehat{g}(0) = 0$ implique que la primitive de g ,

$$G(\theta) = \int_0^\theta g(t) dt,$$

qui est une fonction continue,⁴⁴ est 2π -périodique.⁴⁵ Les formules standard des transformées de Fourier disent alors que $\widehat{G}(k) = \frac{1}{ik} \widehat{g}(k)$ si $k \neq 0$. Pour $\widehat{G}(0)$ on trouve, toujours parce que $\widehat{g}(0) = 0$,⁴⁶ $\widehat{G}(0) = -\frac{1}{2\pi} \int_0^{2\pi} tg(t) dt$.

On regarde alors les approximations de $G(\theta)$ obtenues par convolution avec les noyaux de Fejér $F_N = \sum_{k=-N}^{k=N} \left(1 - \frac{|k|}{N}\right) e^{ikt}$, autrement dit les

$$\sigma_N G(\theta) = F_N * G(\theta) = \sum_{k=-N}^{k=N} \left(1 - \frac{|k|}{N}\right) \widehat{G}(k) e^{ik\theta}.$$

Comme G est continue, on sait d'après le théorème de Fejér que la suite des $\sigma_N G$ converge simplement uniformément partout vers G . En $\theta = 0$ on obtient

$$\begin{aligned} \sigma_N G(0) &= \widehat{G}(0) + \sum_{1 \leq k \leq N} \left(1 - \frac{k}{N}\right) \left(\frac{1}{ik} \widehat{g}(k) - \frac{1}{ik} \widehat{g}(-k)\right) \\ &= \widehat{G}(0) + \sum_{1 \leq k \leq N} \frac{1}{ik} \widehat{g}(k) - \sum_{1 \leq k \leq N} \frac{1}{N} \left(\frac{1}{i} \widehat{g}(k) - \frac{1}{i} \widehat{g}(-k)\right). \end{aligned}$$

Mais comme g est intégrable, les $\widehat{g}(k) \xrightarrow[k \rightarrow \pm\infty]{} 0$ d'après le théorème de Riemann-Lebesgue, et donc leurs moyennes de Cesàro tendent aussi vers 0. Donc, à la limite,

43. On note g plutôt que f car nous allons considérer une primitive G de g et que la notation F entrerait en conflit avec les noyaux de Fejér F_N .

44. Les primitives de fonctions intégrables sont continues.

45. $G(\theta + 2\pi) = \int_0^{\theta+2\pi} g(t) dt = \int_0^\theta g(t) dt + \int_\theta^{\theta+2\pi} g(t) dt = G(\theta) + \int_0^{2\pi} g(t) dt$ (car g est 2π -périodique) $= G(\theta) + 2\pi \widehat{g}(0) = G(\theta)$ (car $\widehat{g}(0) = 0$).

46. $\widehat{G}(0) = \int_0^{2\pi} G(\theta) \frac{d\theta}{2\pi} = \int_0^{2\pi} \left(\int_0^\theta g(t) dt\right) \frac{d\theta}{2\pi}$. Mais on peut intervertir les intégrales. Faire varier θ de 0 à 2π et pour chaque valeur de θ faire varier t de 0 à θ équivaut à faire varier t de 0 à 2π et pour chaque valeur de t faire varier θ de t à 2π . Donc

$$\widehat{G}(0) = \frac{1}{2\pi} \int_0^{2\pi} \left(\int_t^{2\pi} g(t) d\theta\right) dt = \frac{1}{2\pi} \int_0^{2\pi} g(t) (2\pi - t) dt = -\frac{1}{2\pi} \int_0^{2\pi} tg(t) dt$$

$$\text{car } \frac{1}{2\pi} \int_0^{2\pi} g(t) (2\pi) dt = 2\pi \widehat{g}(0) = 0.$$

on obtient

$$\widehat{G}(0) + \sum_{1 \leq k} \frac{1}{ik} \widehat{g}(k) = G(0) = 0 ,$$

autrement dit

$$\sum_{1 \leq k} \frac{1}{k} \widehat{g}(k) = -i\widehat{G}(0) .$$

Mais il y a contradiction puisque la série

$$\sum_{1 \leq k} \frac{1}{k} \widehat{g}(k) = \frac{1}{2i} \sum_{1 \leq k} \frac{1}{k \log(k)}$$

diverge alors que $\widehat{G}(0)$ est fini. On voit que la clé de la preuve est que les coefficients c_k de T tendent bien vers 0 à l'infini mais suffisamment lentement pour que la série $\frac{c_k}{k}$ reste *divergente*. On voit ainsi apparaître (à l'opposé de ce que nous avons vu plus haut avec les séries de Fourier en $\frac{c_k}{k^n}$ $n \geq 2$ garantissant des bonnes propriétés de convergence des T_f pour les f régulières) la nécessité d'analyser des séries convergent trop lentement en tenant compte de l'ordre de décroissance. C'est tout un nouveau champ à explorer.

Mais nous arrêterons ici cette incursion dans le monde classique des séries de Fourier et nous allons continuer notre mini *vademecum* d'analyse fonctionnelle par quelques indications sur la théorie des opérateurs. Cette montée vers l'abstraction nous fera passer du contexte assez calculatoire du XIX^e siècle à un contexte beaucoup plus structural.

7. Les opérateurs et leur décomposition spectrale

7.1. Opérateurs continus et bornés

Les opérateurs que nous rencontrerons généraliseront les opérateurs en dimension finie. Ils seront souvent des opérateurs linéaires *continus* $T : \mathcal{F} \rightarrow \mathcal{G}$ entre espaces de Banach. Quand ils seront partout définis ils seront bornés⁴⁷ et posséderont une norme : l'image de la boule unité $B_{\mathcal{F}}(0, 1)$ sera bornée dans \mathcal{G} et il existera une constante C telle que $\|T(f)\|_{\mathcal{G}} \leq C \|f\|_{\mathcal{F}}$ pour tout $f \in \mathcal{F}$ et l'inf des C donnera la norme

$$\|T\| = \sup_{\|f\|=1} \|T(f)\| < \infty .$$

C'est dire que l'image $T(B_{\mathcal{F}}(0, 1))$ admettra $B_{\mathcal{G}}(0, \|T\|)$ comme boule circonscrite et que T sera Lipschitzienne puisque

$$\|T(f) - T(h)\|_{\mathcal{G}} \leq \|T\| \|f - h\|_{\mathcal{F}} .$$

47. Ce qui ne veut évidemment pas dire borné au niveau de leurs valeurs puisque si $T(f) \neq 0$, $\|T(\lambda f)\|_{\mathcal{G}} = |\lambda| \|T(f)\|_{\mathcal{G}}$ ($\lambda \in \mathbb{C}$) peut être aussi grand que l'on veut.

Leur espace est traditionnellement noté $\mathcal{L}(\mathcal{F}, \mathcal{G})$ et $\mathcal{L}(\mathcal{F})$ lorsque $\mathcal{F} = \mathcal{G}$. $\mathcal{L}(\mathcal{F})$ est une *algèbre* car on peut y composer les opérateurs. Avec la norme $\|T\|$, l'espace vectoriel $\mathcal{L}(\mathcal{F}, \mathcal{G})$ est lui-même un espace de Banach, unitaire lorsque $\mathcal{F} = \mathcal{G}$ car alors $Id_{\mathcal{F}}$ est l'identité de l'algèbre $\mathcal{L}(\mathcal{F})$. C'est dire qu'une suite convergente en norme d'opérateurs bornés converge vers un opérateur borné.

En dimension infinie il est très facile de construire des opérateurs non bornés. Par exemple pour des $f(x)$ dérivables sur \mathbb{R} , l'opérateur $T(f(x)) = xf'(x)$ n'est pas borné. En effet $T(x^n) = nx^n$ et il est donc impossible d'avoir un C indépendant de n tel que $\|nx^n\| \leq C\|x^n\|$. Mais, comme nous le verrons plus bas, on peut restreindre T à un sous-espace $\mathcal{D}(T)$.

La théorie des opérateurs (surtout sur les Banach particuliers que sont les Hilbert) s'est mise en place vers la fin des années 1920 avec tout un ensemble de résultats de Stefan Banach, Hugo Steinhaus, Julius Schauder, Hans Hahn et beaucoup d'autres grands analystes.

Les opérateurs les plus simples sont les opérateurs *compacts* K qui sont par définition des limites (au sens de la norme) d'une suite d'opérateurs de rang fini⁴⁸. Ils sont caractérisés par le fait que l'image $K(\overline{B_{\mathcal{F}}}(0, 1))$ de la boule unité fermée de \mathcal{F} est relativement compacte (i.e. de fermeture compacte) dans \mathcal{G} . C'est une propriété très forte car rappelons qu'en dimension infinie les $\overline{B_{\mathcal{F}}}(0, 1)$ et $\overline{B_{\mathcal{G}}}(0, 1)$ ne sont pas compacts (cf. plus haut).⁴⁹ Les opérateurs compacts ont de très bonnes propriétés généralisant directement ce qui se passe en dimension finie. Si $\mathcal{F} = \mathcal{G}$, ils forment un idéal \mathcal{K} de l'algèbre $\mathcal{L}(\mathcal{F})$, i.e. KT et TK sont compacts pour tout opérateur $T \in \mathcal{L}(\mathcal{F})$.

Pour familiariser le lecteur avec le genre de logique à l'œuvre dans ces éléments d'analyse fonctionnelle, donnons quelques exemples classiques de théorèmes de structure des opérateurs bornés entre Banach, théorèmes démontrés vers la fin des années 1920 dans les références évoquées plus haut. Il s'agit simplement d'esquisses de preuves.

Théorème de Banach-Schauder. Si $T \in \mathcal{L}(\mathcal{F}, \mathcal{G})$, est *surjectif*, alors c'est une application *ouverte*⁵⁰ et l'image $T(B_{\mathcal{F}}(0, 1))$ de la boule unité de \mathcal{F} est encadrée par deux boules de \mathcal{G} . ◇

En effet, par linéarité, cela est équivalent à ce que la boule unité $B_{\mathcal{G}}(0, 1)$ de \mathcal{G} soit contenue dans l'image $T(B_{\mathcal{F}}(0, R))$ d'une boule de \mathcal{F} de rayon R assez grand.

48. A est de rang fini s'il est une combinaison linéaire finie de formes linéaires continues.

49. À propos de la mécanique quantique où les grandeurs sont interprétées comme des opérateurs, Alain Connes a expliqué (cf. par exemple [125]) que les opérateurs compacts permettent d'interpréter des grandeurs *infinitésimales*.

50. L'image réciproque d'un ouvert par une application continue est un ouvert. Mais l'image directe d'un ouvert n'a pas de raison d'être un ouvert (il suffit de prendre une application constante envoyant tout l'espace source sur un point fermé). Lorsque les images directes de tous les ouverts sont des ouverts on dit que l'application est ouverte.

Comme T est surjectif, \mathcal{G} est l'union des fermetures \mathcal{G}_n des images des boules de rayon n de \mathcal{F} , $B_{\mathcal{F}}(0, n)$:

$$\mathcal{G} = \bigcup_{n \in \mathbb{N}} \mathcal{G}_n = \bigcup_{n \in \mathbb{N}} \overline{T(B_{\mathcal{F}}(0, n))} .$$

Comme \mathcal{G} est un Banach l'un de ces fermés, par exemple \mathcal{G}_N , est d'intérieur non vide car dans un Banach une union de fermés d'intérieur vide est d'intérieur vide (propriété de Baire, cf. ci-dessous) et on ne pourrait pas obtenir \mathcal{G} . Donc un des \mathcal{G}_n contient une petite boule $B_{\mathcal{G}}(h, \varepsilon)$, $h \in \mathcal{G}_N$ et par conséquent un autre \mathcal{G}_N contient une petite boule $B_{\mathcal{G}}(0, \varepsilon)$ (on peut prendre $N = 2n$). Par linéarité, cela implique qu'il existe un r assez grand tel que

$$B_{\mathcal{G}}(0, 1) \subset \overline{T(B_{\mathcal{F}}(0, r))} .$$

Donc

$$B_{\mathcal{G}}\left(0, \frac{1}{2^n}\right) \subset \overline{T\left(B_{\mathcal{F}}\left(0, \frac{r}{2^n}\right)\right)}$$

et comme \mathcal{F} est complet,

$$B_{\mathcal{G}}(0, 1) \subset T(B_{\mathcal{F}}(0, 2r)) .$$

La propriété de Baire est équivalente par passage aux complémentaires au fait que toute intersection dénombrable d'ouverts denses reste dense (mais pas ouverte en général). Elle est satisfaite par tous les espaces métriques complets, donc par les Banach. Cela est facile à montrer. Soit U_k , $k \geq 1$, une famille dénombrable d'ouverts denses et V un autre ouvert quelconque (non vide). Il faut montrer qu'il existe un f commun à V et $\bigcap_k U_k$. Comme U_1 est un ouvert dense, $U_1 \cap V$ est un ouvert non vide. Il contient donc (car l'espace est métrique) une boule (ouverte) $B_1(f_1, 2r_1)$ de rayon $2r_1$ assez petit et par suite la boule fermée $\overline{B_1}(f_1, r_1)$. On peut supposer $r_1 \leq 1$. Comme $B_1(f_1, r_1)$ est ouverte, $U_2 \cap B_1(f_1, r_1)$ est non vide et il existe de même un $f_2 \in U_2 \cap B_1(f_1, r_1)$ et une boule fermée $\overline{B_2}(f_2, r_2) \subset U_2 \cap B_1(f_1, r_1)$ de rayon $r_2 \leq \frac{1}{2}$. En itérant le processus, on construit une suite de boules fermées

$$\overline{B_{k+1}}(f_{k+1}, r_{k+1}) \subset U_{k+1} \cap B_k(f_k, r_k)$$

de rayons $r_{k+1} \leq \frac{1}{k+1}$. Pour chaque k , les f_m pour $m \geq k$ sont dans B_k avec des distances $\leq \frac{2}{k}$. Elles forment donc une suite de Cauchy dans B_k et, comme \mathcal{F} est complet, convergent vers un élément $f \in B_k$. Ce f est commun à tous les $U_k \cap V$ et donc commun à V et $\bigcap_k U_k$.

Un corollaire immédiat du théorème de Banach-Schauder est le

Théorème de Banach-Baire. Si T est bijectif, alors l'inverse T^{-1} existe et est continu-borné. C'est un isomorphisme de Banach qui est topologiquement un homéomorphisme. _____ \diamond

Théorème de Banach-Steinhaus. Une suite T_k d'opérateurs bornés est *uniformément* bornée sur la boule $B_{\mathcal{F}}(0, 1)$ si et seulement si elle est *simplement* bornée. Mais elle n'est pas forcément convergente en norme. \diamond

“Uniformément” implique “simplement” est trivial. C'est la réciproque qu'il faut montrer. Soit T_k une telle suite. Soit F_n l'ensemble des $f \in \mathcal{F}$ telles que tous les $\|T_k(f)\|_{\mathcal{G}}$ soient bornés par n :

$$F_n = \{f \in \mathcal{F} \mid \|T_k(f)\|_{\mathcal{G}} \leq n, \forall k\} .$$

Donc $F_n = \bigcap_k F_{n,k}$ avec

$$F_{n,k} = \{f \in \mathcal{F} \mid \|T_k(f)\|_{\mathcal{G}} \leq n\} .$$

Mais les $F_{n,k}$ sont les images inverses par les $\|T_k(\bullet)\|_{\mathcal{G}}$ des fermés $[0, n]$ et comme les T_k sont continus les $F_{n,k}$ sont des fermés. Par suite, F_n étant une intersection de fermés est lui-même fermé. Mais l'union dénombrable des F_n est le sous-ensemble sur lequel la suite des T_k est simplement bornée et donc, par hypothèse, cette réunion est \mathcal{F} tout entier. D'après la propriété de Baire, l'un des F_n, F_{n_0} , doit être d'intérieur non vide. Il existe par conséquent une petite boule $B_{\mathcal{F}}(a, \varepsilon) \subset F_{n_0}$. Si alors $u \in B_{\mathcal{F}}(0, 1)$ est un vecteur unitaire de \mathcal{F} , on a

$$\|T_k(u)\|_{\mathcal{G}} = \frac{1}{\varepsilon} \|T_k(\varepsilon u)\|_{\mathcal{G}} \leq \frac{1}{\varepsilon} \|T_k(a)\|_{\mathcal{G}} + \frac{1}{\varepsilon} \|T_k(a + \varepsilon u)\|_{\mathcal{G}} .$$

Mais $a + \varepsilon u \in B_{\mathcal{F}}(a, \varepsilon)$ puisque u est unitaire et, comme $B_{\mathcal{F}}(a, \varepsilon) \subset F_{n_0}$, la norme $\|T_k(\bullet)\|_{\mathcal{G}}$ est bornée par n_0 sur $B_{\mathcal{F}}(a, \varepsilon)$. Par conséquent, $\|T_k(u)\|_{\mathcal{G}} = \frac{2n_0}{\varepsilon}$ et la suite T_k est uniformément bornée sur $B_{\mathcal{F}}(0, 1)$.

7.2. Opérateurs non bornés et fermés

On doit souvent utiliser des opérateurs non bornés T définis seulement sur un sous-espace. Le plus grand sous-espace $\mathcal{D}(T)$ sur lequel T est défini et satisfait les conditions voulues pour son application considérée, s'appelle alors le “domaine de définition” de T . Il s'agit là d'une notion assez subtile qui mêle des conditions mathématiques nécessaires à des conditions qui peuvent venir des phénomènes empiriques que l'on souhaite modéliser en utilisant T . Si par exemple T fait intervenir la dérivée continue $\frac{\partial^2 f}{\partial x^2}$ pour des f sur $[0, 1]$, il faudra nécessairement que $f \in \mathcal{C}^2([0, 1])$. Mais si l'on travaille sur des cordes vibrantes et si f modélise la vibration de la corde, on voudra que la corde soit clampée à ses extrémités, i.e. que f satisfasse en plus $f(0) = f(1) = 0$. Cette contrainte au bord (condition aux limites) est essentielle au choix du domaine de définition $\mathcal{D}(T)$. Quand on décomposera f en harmoniques par analyse de Fourier, on en fera une superposition de fonctions trigonométriques dont la période dépendra fondamentalement de la longueur ℓ de la corde (ici $\ell = 1$). Le spectre (discret) de T dépendra donc des contraintes modélisatrices s'ajoutant

à la contrainte de différentiabilité C^2 . Nous reviendrons plusieurs fois sur ce point délicat.

Parmi les opérateurs non bornés, il en existe qui restent assez “bons” : les opérateurs *fermés*. Cela signifie que si f_n est une suite d'éléments du domaine $\mathcal{D}(T)$, si $f_n \xrightarrow{n \rightarrow \infty} f$ dans \mathcal{F} et si $T(f_n)$ converge dans \mathcal{G} , alors $f \in \mathcal{D}(T)$ et $T(f_n) \xrightarrow{n \rightarrow \infty} T(f)$ dans \mathcal{G} . On peut aussi dire que le *graphe* de T ,

$$\text{Gr}(T) = \{(f, T(f)) \in \mathcal{D}(T) \times \mathcal{G}\} ,$$

est *fermé* dans $\mathcal{F} \times \mathcal{G}$. Il est alors naturel de dire que T est “fermable” si la fermeture de son graphe est le graphe d'un opérateur \overline{T} qui sera dit la “fermeture” de T . On pourra dans ce cas considérer des sous-ensembles \mathcal{D}' de $\mathcal{D}(T)$ tels que $\overline{T} \upharpoonright_{\mathcal{D}'} = \overline{T}$.

Les opérateurs non bornés fermés sont “bons” parce qu'ils possèdent des propriétés comme

- (i) si $\mathcal{D}(T) = \mathcal{F}$ alors T est borné,
- (ii) leurs noyaux $\ker(T)$ sont des sous-espaces fermés de \mathcal{F} ,
- (iii) lorsqu'ils sont injectifs leur inverse est fermé,
- (iv) ils restent fermés si on leur ajoute un opérateur borné,
- (v) on peut en faire l'analyse spectrale et en définir le calcul fonctionnel (i.e. définir les $F(T)$ pour des fonctions F suffisamment “bonnes”).

Nous en verrons plus bas plusieurs exemples. Beaucoup d'opérateurs *différentiels* sont de ce type. Par exemple si $\mathcal{F} = \mathcal{G} = \mathcal{C}^0([0, 1])$ est l'espace des fonctions continues sur l'intervalle $[0, 1]$, $D(f(x)) = f'(x)$ défini sur le domaine $\mathcal{D}(D) = \mathcal{C}^1([0, 1])$ des f continûment dérivables n'est pas borné mais il est fermé (nous y reviendrons plus bas). Il est même la fermeture de sa restriction à $\mathcal{C}^\infty([0, 1])$.

7.3. Des opérateurs symétriques aux opérateurs auto-adjoints

7.3.1. L'opérateur adjoint.

Si \mathcal{H} est un espace de Hilbert et si T est un opérateur non borné de domaine $\mathcal{D}(T)$ *dense*, alors on peut définir un adjoint T^* satisfaisant

$$\langle T(f), h \rangle = \langle f, T^*(h) \rangle$$

pour $f \in \mathcal{D}(T)$ et $h \in \mathcal{D}(T^*)$. On définit $T^*(h)$ comme l'élément de \mathcal{H} , unique s'il existe, tel que pour tout $f \in \mathcal{D}(T)$ la formule $\langle T(f), h \rangle = \langle f, T^*(h) \rangle$ soit vérifiée. D'après le théorème de Hahn-Banach et le théorème de représentation de Riesz, $T^*(h)$ existe si et seulement si il existe une constante C telle que, pour tout $f \in \mathcal{D}(T)$, $|\langle T(f), h \rangle| \leq C \|f\|_{\mathcal{H}}$. Cela caractérise le domaine $\mathcal{D}(T^*)$.

7.3.2. Opérateurs auto-adjoints en dimension finie.

Des opérateurs particulièrement importants dans de nombreux domaines, en particulier en physique quantique, sont les opérateurs dits “auto-adjoints” qui sont

égaux à leur adjoint, c'est-à-dire qui satisfont

$$\langle T(f), h \rangle = \langle f, T(h) \rangle .$$

Il sont associés à la possibilité d'effectuer des *mesures numériques* car leurs valeurs propres sont *réelles*. En effet si $f \neq 0$ est vecteur propre de T pour la valeur propre λ ,

$$\langle T(f), f \rangle = \langle \lambda f, f \rangle = \lambda \langle f, f \rangle = \langle f, T(f) \rangle = \langle f, \lambda f \rangle = \bar{\lambda} \langle f, f \rangle$$

et donc $\lambda = \bar{\lambda}$ puisque $\langle f, f \rangle = \|f\|^2 \neq 0$. En fait $\langle f, T(f) \rangle$ est réel puisque

$$\langle f, T(f) \rangle = \overline{\langle T(f), f \rangle} = \overline{\langle f, T(f) \rangle} .$$

Cette notion apparemment simple est subtile lorsque \mathcal{H} est de dimension infinie.

En dimension finie, il n'y a pas de difficultés. La matrice de T dans toute base orthonormale est une matrice *hermitienne*, c'est-à-dire égale à la transposée de sa complexe conjuguée. En effet, dans une telle base $\{e_i\}$, si $T(e_i) = \sum_j T_{ij}e_j$ et si $f = \sum_k f_k e_k$, $h = \sum_l h_l e_l$, on a (avec la convention de sommation d'Einstein)

$$\begin{aligned} \langle T(f), h \rangle &= \langle f_k T_{kj} e_j, h_l e_l \rangle = f_k T_{kj} \bar{h}_j \\ \langle f, T(h) \rangle &= \langle f_k e_k, h_j T_{jl} e_l \rangle = f_k \bar{T}_{jk} h_j \end{aligned}$$

et donc $f_k T_{kj} \bar{h}_j = f_k \bar{T}_{jk} h_j$ pour tous f_k et h_j , soit $T_{kj} = \bar{T}_{jk}$.

On montre que T est *diagonalisable* au moyen d'une transformation unitaire U , "unitaire" signifiant que $UU^* = U^*U = I$. Les valeurs propres λ constituent le *spectre* $\text{Spec}(T) \subset \mathbb{C}$ de T et il existe une *décomposition spectrale* de T , c'est-à-dire une décomposition orthogonale⁵¹ de \mathcal{H} en les sous-espaces propres \mathcal{H}_λ associés aux λ , \mathcal{H}_0 étant le noyau de T . Par définition, T agit sur les \mathcal{H}_λ simplement comme l'*opérateur de multiplication* par λ .

7.3.3. Opérateurs symétriques.

En dimension infinie la situation est beaucoup moins simple. Pour les opérateurs bornés partout définis on généralise directement le cas de dimension finie. Mais il existe beaucoup d'opérateurs intéressants, et même fondamentaux, en particulier des opérateurs différentiels sur des espaces de Hilbert de fonctions, qui sont des opérateurs non bornés définis seulement sur un sous-espace $\mathcal{D}(T)$ de \mathcal{H} . Quitte à restreindre T au sous-espace fermé $\overline{\mathcal{D}(T)}$ de \mathcal{H} , on peut supposer $\mathcal{D}(T)$ *dense*. La généralisation directe du concept d'auto-adjonction est alors celle de *symétrie* : T est dit "symétrique" si $\langle T(f), h \rangle = \langle f, T(h) \rangle$ pour tous $f, h \in \mathcal{D}(T)$. Le problème est que l'adjoint T^* peut avoir un domaine de définition $\mathcal{D}(T^*)$ *plus grand* que $\mathcal{D}(T)$.

51. Si f et h sont des vecteurs propres de valeurs propres λ, μ différentes alors f et g sont orthogonaux. En effet $\langle T(f), h \rangle = \lambda \langle f, h \rangle = \langle f, T(h) \rangle = \bar{\mu} \langle f, h \rangle$ et comme λ, μ sont *réelles*, $\bar{\mu} = \mu$ et puisque $\lambda \neq \mu$, cela implique $\langle f, h \rangle = 0$.

Le problème peut s'expliquer très facilement sur un exemple simple que l'on trouve dans tous les textes d'introduction au sujet. Étant donné l'intervalle $[0, 1]$ de \mathbb{R} , considérons sur le Hilbert $L^2([0, 1])$ l'opérateur différentiel de dérivée seconde $T = \frac{d^2}{dx^2}$ sur un sous-espace admissible (mais à définir précisément) $\mathcal{D}(T)$. Il est trivialement non borné car, la norme de x^n ($n \geq 2$) étant $\frac{1}{\sqrt{2n+1}}$, et celle de

$$\frac{d^2 x^n}{dx^2} = n(n-1)x^{n-2}$$

étant $\frac{n(n-1)}{\sqrt{2n-3}}$, sa norme $\|T\|$ est supérieure à la limite pour $n \rightarrow \infty$ de

$$\frac{n(n-1)}{\sqrt{2n-3}} \left(\frac{1}{\sqrt{2n+1}} \right)^{-1},$$

limite qui est infinie. Regardons alors à quelles conditions T peut être symétrique. L'outil de base est l'intégration par parties. On remarque que $(f\bar{h})' = f''\bar{h} + f'\bar{h}'$ et que donc

$$\langle T(f), h \rangle = \int_0^1 f''\bar{h} = \int_0^1 (f\bar{h})' - \int_0^1 f'\bar{h}' = [f\bar{h}]_0^1 - \int_0^1 f'\bar{h}'.$$

De même $(f\bar{h}')' = f\bar{h}'' + f'\bar{h}'$ et donc

$$\langle f, T(h) \rangle = \int_0^1 f\bar{h}'' = \int_0^1 (f\bar{h}')' - \int_0^1 f'\bar{h}' = [f\bar{h}']_0^1 - \int_0^1 f'\bar{h}'.$$

On voit que la symétrie sera assurée si les termes de bord $[f\bar{h}]_0^1$ et $[f\bar{h}']_0^1$ sont toujours égaux sur $\mathcal{D}(T)$, autrement dit si certaines *conditions aux bord* (ou conditions aux limites) sont satisfaites. Mais il y a plusieurs choix possibles, ce qui rend le problème non trivial lorsqu'on veut faire une théorie générale.

Remarque. Le fait qu'il y ait plusieurs choix entraîne que le domaine de définition $\mathcal{D}(T)$ de T fait partie intégrante de la définition de T . Comme nous le verrons plus bas, ce choix a une incidence fondamentale sur la nature des *vecteurs propres* de T . □

Un premier choix, notons le T_0 , peut être d'imposer

$$g(0) = g(1) = g'(0) = g'(1) = 0$$

comme conditions aux limites pour tous les $g \in \mathcal{D}(T_0)$.⁵² Les termes de bord s'annulent et T_0 est donc symétrique. On prendra alors pour $\mathcal{D}(T_0)$ les $g \in L^2([0, 1])$ au moins C^2 (les C^∞ suffisent car elles forment un sous-ensemble dense), telles que $g'' \in L^2([0, 1])$ et telles que les conditions aux limites soient satisfaites. Mais si T_0 est ainsi défini, on voit que l'annulation des termes de bord est assurée par les conditions aux limites de f seule et que l'on n'a pas besoin de conditions aux limites pour h . Cela signifie que l'adjoint T_0^* de T_0 est bien égal à T_0 sur $\mathcal{D}(T_0)$ mais possède un domaine de définition $\mathcal{D}(T_0^*)$ plus large que $\mathcal{D}(T_0)$.

On peut prendre des conditions aux limites plus faibles, par exemple

$$g(0) = ag(1), \quad g'(0) = bg'(1) \quad (a, b \in \mathbb{C}) .$$

On a alors

$$\begin{aligned} [f'\bar{h}]_0^1 &= f'(1)\bar{h}(1)(1 - b\bar{a}) \\ [f\bar{h}']_0^1 &= f(1)\bar{h}'(1)(1 - a\bar{b}) \end{aligned}$$

et donc si $a\bar{b} = 1$, la symétrie est garantie. On obtient ainsi des extensions symétriques T_{ab} de T_0 qui sont toutes *différentes* entre elles. En effet supposons que T_{ab} et $T_{a'b'}$ soient identiques. Il faudrait à la fois

$$a\bar{b} = 1, \quad a'\bar{b}' = 1, \quad a'\bar{b} = 1, \quad a\bar{b}' = 1,$$

ce qui implique $a = a'$ et $b = b'$. On observe ainsi une multiplication d'extensions symétriques de T_0 incompatibles entre elles.

Ceci dit, il existe toujours une extension symétrique *canonique* d'un opérateur symétrique T , qui est sa *fermeture* \bar{T} . Comme on dit, un opérateur symétrique est toujours "fermable". Cela signifie que la fermeture de son graphe $Gr(T) = \{(f, Tf) \in \mathcal{D}(T) \oplus \mathcal{H}\}$ dans $\mathcal{H} \oplus \mathcal{H}$ est le graphe $Gr(\bar{T})$ d'un opérateur \bar{T} étendant T . On montre l'identité remarquable $\bar{T} = T^{**}$.

La notion de graphe permet d'ailleurs de définir géométriquement de façon élégante l'adjoint T^* de T . Soit J la structure symplectique de $\mathcal{H} \oplus \mathcal{H}$ définie par $J(f, g) = (-g, f)$. On montre que le graphe $Gr(T^*)$ de T^* est le complément orthogonal de $JGr(T)$,

52. On remarquera que ces conditions de périodicité ne sont pas du tout satisfaites par les x^n .

On remarquera aussi que l'on peut approximer aussi bien que l'on veut d'autres conditions aux limites avec ces conditions périodiques. Il suffit de faire suffisamment varier des f périodiques dans des petits intervalles $[0, \varepsilon]$ et $[1 - \varepsilon, 1]$ et de tenir compte du fait que la différence entre deux éléments $f, g \in L^2([0, 1])$ est mesurée par une intégrale globale sur $[0, 1]$ et non pas par des valeurs ponctuelles. Si par exemple on considère la fonction constante 1 sur $[0, 1]$, on peut l'approximer par des f_ε qui partent de $x = 0$ avec une tangente horizontale, remontent à 1 sur $[0, \varepsilon]$, restent égales à 1 sur $[\varepsilon, 1 - \varepsilon]$ et replongent vers 0 sur $[1 - \varepsilon, 1]$ en atteignant 0 en $x = 1$ avec une tangente horizontale. On a alors $\|f_\varepsilon - 1\| < 2\varepsilon \xrightarrow{\varepsilon \rightarrow 0} 0$.

$$\begin{aligned} Gr(T^*) &= (JGr(T))^\perp \\ &= \left\{ (g, h) \in \mathcal{H} \oplus \mathcal{H} \mid \langle (g, h), (-T(f), f) \rangle = -\langle g, T(f) \rangle + \langle h, f \rangle = 0 \right. \\ &\quad \left. \text{pour tout } f \in \mathcal{D}(T) \right\}. \end{aligned}$$

Si T est symétrique et si $g \in \mathcal{D}(T)$, on a

$$-\langle g, T(f) \rangle + \langle h, f \rangle = -\langle T(g), f \rangle + \langle h, f \rangle = \langle -T(g) + h, f \rangle$$

et, comme $\mathcal{D}(T)$ est dense, la condition $\langle -T(g) + h, f \rangle = 0$ pour tout $f \in \mathcal{D}(T)$ implique $-T(g) + h = 0$, et donc $h = T(g)$ ce qui correspond bien au fait que, si T est symétrique, $T^* = T$ sur $\mathcal{D}(T)$. Pour démontrer qu'un opérateur T symétrique est fermable, que $Gr(\overline{T}) = \overline{Gr(T)}$ et que $\overline{T} = T^{**}$, on utilise le fait que l'orthogonal d'un sous-espace est toujours fermé et le théorème dit *du graphe fermé* (analogue au théorème de Banach-Schauder évoqué plus haut) disant que les applications linéaires qui sont partout définies et dont le graphe est fermé sont continues.⁵³

7.3.4. Opérateurs auto-adjoints.

On dit que T est "auto-adjoint" (a.a.) non seulement si T est symétrique mais si de plus il y a *égalité des domaines* $\mathcal{D}(T^*) = \mathcal{D}(T)$, ce qui est une condition forte. Conformément à la tradition, nous noterons A un tel opérateur. Par exemple sur $L^2(\mathbb{R})$ l'opérateur de multiplication par x , $A(f) = xf$, défini sur

$$\mathcal{D}(T) = \{f \in \mathcal{H} \mid xf \in \mathcal{H}\}$$

est a.a. Il est trivialement symétrique et on vérifie que $\mathcal{D}(T) = \mathcal{D}(T^*)$ car il n'y a que des conditions d'intégrabilité et pas de conditions aux bords.

Si T est symétrique sans être auto-adjoint, alors la situation n'est pas bonne et le théorème de décomposition spectrale ci-dessous n'est plus valable.

7.3.5. Opérateurs essentiellement auto-adjoints.

On dit que T est "essentiellement auto-adjoint" (e.a.a.) si T est symétrique et admet une *unique* extension a.a. Pour illustrer ce concept, considérons l'exemple particulièrement simple de $T(f) = af'$ sur $L^2([0, 1])$ avec un domaine admissible $\mathcal{D}(T)$ et demandons-nous à quelles conditions il peut être symétrique, e.a.a. ou a.a. La relation de symétrie est $\langle T(f), h \rangle = \langle f, T(h) \rangle$, i.e. $\langle af', h \rangle = \langle f, ah' \rangle$ ou $a \langle f', h \rangle = \bar{a} \langle f, h' \rangle$. L'outil est encore l'intégration par parties. On a

$$(afh)' = af'h + afh'$$

et par conséquent

$$[afh]_0^1 = a \langle f', h \rangle + a \langle f, h' \rangle.$$

On voit ainsi que la symétrie va dépendre de deux choses :

53. Dans l'autre sens, une application linéaire continue a toujours un graphe fermé.

- (i) des conditions aux limites garantissant que le terme de bord $[af\bar{h}]_0^1$ s'annule et
(ii) d'une valeur de a garantissant que $a \langle f', h \rangle = \bar{a} \langle f, h' \rangle$.

Pour satisfaire (ii) on pose $a = -\bar{a}$ ce qui donne

$$[af\bar{h}]_0^1 = \langle af', h \rangle + \langle f, \bar{a}h' \rangle = \langle af', h \rangle - \langle f, ah' \rangle = \langle T(f), h \rangle - \langle f, T(h) \rangle = 0 .$$

Un cas simple est $a = -i$ et $T(f) = -if'$. Nous rencontrerons à nouveau cet opérateur a.a. dans la section 1.3 du chapitre suivant (17) comme générateur infinitésimal du groupe à 1-paramètre d'opérateurs unitaires que sont les translations. Il s'agit de l'opérateur "moment" en mécanique quantique, l'autre opérateur fondamental, l'opérateur de "position", étant la multiplication par x . Nous avons vu à la section 5.4 du chapitre 5 où ces opérateurs étaient notés \hat{p} et \hat{q} et où intervenait la constante de Planck, que le commutateur $[\hat{q}, \hat{p}] = i\hbar I$ définit l'algèbre de Lie de Heisenberg.

Regardons les propriétés de T .⁵⁴

1. Si l'on pose $\mathcal{D}(T) = \mathcal{C}^\infty([0, 1])$ (pas de conditions aux limites), T n'est pas symétrique car le terme de bord n'a pas de raison de s'annuler.
2. Si l'on pose $\mathcal{D}(T) = \{f \in \mathcal{C}^\infty([0, 1]) \mid f(0) = f(1) = 0\}$ (double condition aux limites périodique), T est symétrique mais il n'est pas a.a. car les conditions aux limites sont trop fortes. Celles sur f seule suffisent à éliminer le terme de bord. On montre que

$$\mathcal{D}(T^*) = \{f \in L^2([0, 1]) \mid f', f'' \in L^2([0, 1])\} .$$

T n'est même pas e.a.a. car

$$\mathcal{D}(\bar{T}) = \{f \in L^2([0, 1]) \mid f', f'' \in L^2([0, 1]), f(0) = f(1) = 0\} .$$

et \bar{T} n'est pas a.a. car $\mathcal{D}(\bar{T}^*) = \mathcal{D}(T^*) \neq \mathcal{D}(\bar{T})$.

3. Avec des conditions aux limites plus faibles

$$\mathcal{D}(T) = \{f \in \mathcal{C}^\infty([0, 1]) \mid f(0) = f(1)\}$$

alors $\mathcal{D}(T^*) = \mathcal{D}(\bar{T})$ et T devient e.a.a.

Ces différences se lisent intuitivement sur les *vecteurs propres* envisageables. À un niveau formel, les fonctions $f = e^{\mu x}$, $\mu \in \mathbb{C}$, sont des vecteurs propres possibles de T pour la valeur propre $\lambda = -i\mu$.⁵⁵ Dans le cas (1) il n'y a pas de conditions aux limites et tout $\lambda \in \mathbb{C}$ est valeur propre, autrement dit le spectre de T est \mathbb{C} tout

54. Cf. l'excellent article de Wikipedia sur les opérateurs e.a.a.

55. Remarquons que les $e^{\mu x}$ sont dans $L^2([0, 1])$ car leur norme au carré est $e^{2\Re(\mu)} \int_0^1 dx = e^{2\Re(\mu)}$. Mais ce n'est plus le cas dans $L^2(\mathbb{R})$ où $e^{2\Re(\mu)} \int_{\mathbb{R}} dx$ est infinie. $e^{\mu x}$ est alors une sorte de vecteur propre "généralisé" ("non normalisable" comme disent les physiciens) qui satisfait la propriété d'un vecteur propre tout en n'appartenant pas au Hilbert. Nous allons y revenir.

entier. Dans le cas (2) (conditions aux limites fortes), $f(0) = f(1) = 0$ implique $f = 0$ et il n'y a donc pas de vecteurs propres du tout. Dans le cas (3) (conditions aux limites équilibrées), il faut seulement $1 = e^\mu$ et donc $\mu = 2\pi in$, ce qui donne une base propre orthonormale constituée des $e^{2\pi inx}$ de valeurs propres $\lambda = -i\mu = 2\pi n$.

7.3.6. Conditions de symétrie et d'adjonction.

On voit ainsi que le choix de $\mathcal{D}(T)$ a une incidence déterminante sur le spectre de T . Il existe d'ailleurs des conditions simples sur les valeurs propres d'un opérateur A pour savoir s'il est symétrique, e.a.a. ou a.a. Elles tournent autour de la possibilité pour A d'avoir ou non $\pm i$ comme valeur propre et concerne donc $A \pm iI$ et $A^* \pm iI$ et reposent sur la notion de *transformée de Cayley* d'un opérateur.⁵⁶ La transformée de Cayley est au départ l'isomorphisme conforme

$$W : \mathbb{C} - \{-i\} \rightarrow \mathbb{C} - \{1\}$$

$$z \mapsto W(z) = w = \frac{z - i}{z + i}.$$

Si $z = x$ est réel, alors

$$|W(x)|^2 = W(x) \overline{W(x)} = \frac{x - i}{x + i} \frac{x + i}{x - i} = 1$$

et donc W transforme \mathbb{R} en $\mathbb{S}^1 - \{1\}$. W est une transformation conforme du demi-plan hyperbolique de Poincaré $= \{z \in \mathbb{C} | \Im(z) > 0\}$ en l'intérieur du disque unité. Sur $\mathbb{C} - \{1\}$ son inverse est

$$W^{-1}(w) = i \frac{1 + w}{1 - w}.$$

On généralise cette définition aux opérateurs en partant de la définition

$$W(A)(A + iI) = A - iI$$

et en visant un résultat du genre : A a.a. (ce qui généralise $x \in \mathbb{R}$) équivaut à $W(A)$ unitaire (ce qui généralise $W(x) \in \mathbb{S}^1$).

1. On montre d'abord que si A est symétrique, il existe un seul opérateur linéaire⁵⁷

$$W(A) : \text{Im}(A + iI) \rightarrow \text{Im}(A - iI)$$

56. Arthur Cayley (1821-1895) a été le prototype du grand mathématicien pur en Angleterre. Élève brillantissime au King's College de Londres, puis au Trinity College de Cambridge (où il entra à 17 ans), cumulant les prix, il se forma en mathématiques en étudiant les maîtres de l'École française dont nous avons parlé (Monge, Legendre, Laplace, Cauchy, Fourier). Sa carrière se fera à Cambridge où il est nommé en 1863 et cumulera les honneurs académiques. Il est l'un des fondateurs de l'algèbre linéaire (matrices, déterminants, etc.) avec ses applications à la géométrie, de la théorie des invariants et de la théorie abstraite des groupes. Il travailla aussi en géométrie projective et en théorie des graphes.

57. $\text{Im}(T)$ est l'image de l'opérateur T .

tel que

- (i) $W(A)(Af + if) = Af - if$ pour tout $f \in \text{Dom}(A)$,
- (ii) $W(A)$ est isométrique sur son domaine et
- (iii) $\text{Im}(I - W(A))$ est dense dans \mathcal{H} .

Réciproquement, si W possède ces propriétés, il existe un seul

$$A(W) : \text{Im}(I - W) \rightarrow \text{Im}(I + W)$$

tel que $A(h - Wh) = i(h + Wh)$ pour tout $h \in \text{Dom}(W)$ et A est symétrique.

2. On montre ensuite que A est e.a.a. si et seulement si $\text{Im}(A + iI)$ et $\text{Im}(A - iI)$ sont denses dans \mathcal{H} . Cela équivaut au fait que $\pm i$ ne sont pas des valeurs propres de A^* .
3. Enfin on montre que, effectivement, A est a.a. si et seulement si $W(A)$ est unitaire (ne possédant pas la valeur propre 1).

Pour le détail des preuves, le lecteur intéressé pourra se référer au chapitre 13 du traité classique d'analyse fonctionnelle de Walter Rudin [473] déjà cité. Si A est symétrique de domaine dense, soit $\text{Im}^\pm = \text{Im}(A \pm iI)$. On a

$$(\text{Im}^\pm)^\perp = \ker(A^* \mp iI) .$$

En effet, comme A est symétrique, et donc de valeurs propres réelles, $Af \pm if \neq 0$ pour tout $f \in \text{Dom}(A)$. Qui plus est, comme A symétrique et $f \in \text{Dom}(A)$,

$$\begin{aligned} \|Af \pm if\|^2 &= \|Af\|^2 + \|f\|^2 + \langle \pm if, Af \rangle + \langle Af, \pm if \rangle \\ &= \|Af\|^2 + \|f\|^2 \pm i \langle f, Af \rangle \mp i \langle Af, f \rangle \\ &= \|Af\|^2 + \|f\|^2 \pm i \langle f, Af \rangle \mp i \langle f, Af \rangle \\ &= \|Af\|^2 + \|f\|^2 \end{aligned}$$

et par conséquent, $\{Af \pm if\}$ est *isométrique au graphe* $\{f, Af\}$ de A .

Si alors $h \neq 0 \in (\text{Im}^+)^\perp$, on a par hypothèse $\langle Af + if, h \rangle = 0$ et donc

$$\langle f, A^*h - ih \rangle = 0$$

pour tout $f \in \text{Dom}(A)$ et, comme $\text{Dom}(A)$ est dense, $A^*h - ih = 0$. Il en va de même pour $(\text{Im}^-)^\perp$. On voit ainsi que le complémentaire orthogonal de Im^\pm est constitué des vecteurs propres de A^* de valeurs propres $\pm i$. D'où la décomposition

$$\text{Dom}(A^*) = \text{Dom}(\overline{A}) \oplus \ker(A^* - iI) \oplus \ker(A^* + iI) .$$

Par conséquent le fait que A soit e.a.a., et donc que $\overline{A} = A^*$ avec $\text{Dom}(\overline{A}) = \text{Dom}(A^*)$, oblige h à être nul puisque A^* ne peut avoir $\pm i$ comme valeur propre.

7.4. Décomposition spectrale des opérateurs auto-adjoints

Disons maintenant quelques mots sur la possibilité de généraliser aux opérateurs auto-adjoints A en dimension infinie la notion de décomposition spectrale d'un opérateur auto-adjoint en dimension finie. Nous avons vu plus haut que les A a.a. ont des valeurs propres réelles car si $Af = \lambda f$ ($f \neq 0$), alors

$$\begin{aligned}\langle Af, f \rangle &= \langle \lambda f, f \rangle = \lambda \langle f, f \rangle = \lambda \|f\|^2 \\ &= \langle f, A^* f \rangle = \langle f, Af \rangle = \langle f, \lambda f \rangle = \bar{\lambda} \langle f, f \rangle = \bar{\lambda} \|f\|^2\end{aligned}$$

et donc $\lambda = \bar{\lambda}$ puisque $\|f\|^2 \neq 0$. Ces opérateurs généralisent les opérateurs diagonalisables et on pourrait donc penser qu'il existe une base orthonormale propre de \mathcal{H} permettant de diagonaliser A en décomposant \mathcal{H} en sous-espaces propres. Rappelons que dans le cas où A est opérateur linéaire sur un \mathbb{C} -vectoriel V de dimension finie n , les valeurs propres λ sont les racines du polynôme caractéristique de degré n ,

$$\Pi_A(\lambda) = \det(A - \lambda I) ,$$

car s'il existe un vecteur propre $v \neq 0$ tel que $Av = \lambda v$, alors $(A - \lambda I)v = 0$, $(A - \lambda I)$ possède un noyau non trivial et donc (puisque'on est en dimension finie) son déterminant est nul. Comme on travaille sur \mathbb{C} qui est algébriquement clos, il y a toujours une racine λ_1 de $\Pi_A(\lambda)$ (il y en a même n comptées avec multiplicité). Si V est muni d'un produit scalaire hermitien, il existe donc un vecteur propre v_1 que l'on peut normaliser en le remplaçant par $\frac{v_1}{\|v_1\|}$. Grâce au produit scalaire sur V , $V_1 = \mathbb{C}v_1$ possède un espace orthogonal V_1^\perp et ce dernier est stable par A si A est a.a.. En effet, si $w \in V_1^\perp$,

$$\begin{aligned}\langle Aw, v_1 \rangle &= \langle w, Av_1 \rangle \quad (\text{car } A \text{ est a.a.}) = \langle w, \lambda_1 v_1 \rangle \\ &= \bar{\lambda}_1 \langle w, v_1 \rangle = 0 \quad (\text{car } w \in V_1^\perp)\end{aligned}$$

et donc $Aw \in V_1^\perp$. En restreignant A à V_1^\perp , on trouve alors une seconde valeur propre λ_2 (qui peut être égale à λ_1 car le polynôme caractéristique $\Pi_A(\lambda)$ peut évidemment avoir des racines multiples). En itérant n fois le processus, on obtient une base propre orthonormée de A a.a. qui est en quelque sorte la base "la mieux adaptée" à l'opérateur A . L'ensemble des valeurs propres λ_i différentes (de multiplicité m_i égale à leur multiplicité comme racine de $\Pi_A(\lambda)$) constitue le *spectre* $\text{Spec}(A)$ de A ($\text{Spec}(A) \subset \mathbb{R}$ car A est a.a.) et si les V_{λ_i} sont les espaces propres des λ_i (de dimensions respectives les m_i), V se décompose en la somme directe orthogonale

$$V = \bigoplus_{\lambda_i \in \text{Spec}(A)}^\perp V_{\lambda_i} .$$

Soit alors P_{λ_i} le projecteur orthogonal de V sur V_{λ_i} . A est la somme directe des P_{λ_i} :

$$A = \bigoplus_{\lambda_i \in \text{Spec}(A)} \lambda_i P_{\lambda_i}$$

et si $F(x)$ est une fonction sur \mathbb{R} (polynôme, série entière) telle que $F(A)$ soit bien définie, on a

$$F(A) = \bigoplus_{\lambda_i \in \text{Spec}(A)} F(\lambda_i) P_{\lambda_i}.$$

C'est ce que l'on appelle la "décomposition spectrale" de A . Elle décompose A en une somme d'opérateurs de multiplication par un scalaire et permet de définir le "calcul fonctionnel" des $F(A)$.

Remarque. C'est ce modèle qui sert d'inspiration à la *théorie de la mesure* en mécanique quantique. Les états du système S considéré sont les vecteurs de V , les grandeurs à valeurs réelles (dites "observables") que l'on mesure sont représentées par des opérateurs a.a. A et si le résultat de la mesure de A pour S dans l'état v donne λ , cela signifie que v a été *projeté* sur l'espace propre V_λ par P_λ . C'est ce que l'on appelle le "collapse" ou la "réduction du paquet d'onde" induit par le processus de mesure. Le "problème de la mesure" est de savoir si l'on peut ramener cet axiome, qui s'ajoute à l'évolution de S donnée par l'équation de Schrödinger de S , à l'évolution du système $S + \tilde{A}$ composé de S et de l'appareil \tilde{A} mesurant A , cette dernière évolution étant régie par l'équation de Schrödinger de $S + \tilde{A}$. Cela permettrait alors d'éliminer l'axiome de réduction par projection. Une immense littérature à la fois physico-mathématique et philosophique lui a été consacrée, des origines de la théorie quantique à aujourd'hui, débats où se croisent le paradoxe d'Einstein-Podolsky-Rosen, les variables cachées (Bohm), l'impossibilité de variables cachées locales (inégalités de Bell et expérience d'Aspect), les phénomènes d'intrication, la "décohérence" du monde macroscopique (Zeh, Joos, Zurek, Haroche), le collapse "spontané".⁵⁸ □

Mais le modèle de la dimension finie ne se prolonge pas directement aux espaces fonctionnels de dimension infinie comme les espaces de Hilbert L^2 définis sur des variétés différentiables munies d'une mesure. La théorie reste essentiellement la même pour les opérateurs *compacts* car l'argument reposant sur l'existence de valeurs propres et de vecteurs propres reste valide. Mais ce n'est plus le cas en général pour les A a.a. non compacts et non bornés seulement densément définis, et a fortiori pour les T seulement symétriques. Toutefois, pour les opérateurs a.a. il existe une généralisation de la décomposition spectrale exprimée par un théorème fondamental de Marshall Stone du début des années 1930.⁵⁹ Elle peut s'étendre aux opérateurs

58. Cf. par exemple l'ouvrage [489] de Maximilian Schlosshauer.

59. Nous avons croisé le nom de Marshall Harvey Stone (1903-1969) dans la section 6 sur les séries de Fourier à propos d'Antoni Zygmund. Élève de George Birkhoff il fut professeur à Harvard puis à Chicago où il redonna tout son prestige au département de mathématiques. Spécialiste des opérateurs dans les espaces de Hilbert et de leur analyse spectrale (son traité [515] de 1932 est un classique), il est aussi le co-auteur du théorème de Stone-von Neumann dont nous parlerons longuement dans le chapitre suivant 17, section 5.1. Il est également connu pour son approfondissement du théorème de Weierstrass sur les approximations uniformes de

e.a.a. mais pas aux opérateurs symétriques non e.a.a. Il y a donc beaucoup de subtilités dans les généralisations de la notion de décomposition spectrale et nous allons maintenant en dire un mot.

Souvent, étant donné un A , on peut facilement trouver des vecteurs propres envisageables, mais qui *ne sont pas* des éléments de \mathcal{H} . Ce sont des vecteurs propres en quelque sorte “généralisés” qui correspondent à des valeurs propres “généralisées” et, pour les prendre en compte il faut raffiner notablement les formalismes qui sont opératoires en dimension finie.

Ces difficultés peuvent se comprendre très intuitivement sur l'exemple du Hilbert $\mathcal{H} = L^2(\mathbb{R}, dx)$. Des fonctions généralisées comme les distributions de Dirac $\delta_\lambda(x) = \delta(x - \lambda)$ n'appartiennent pas à \mathcal{H} .⁶⁰ De même des fonctions de phase $f(x) = e^{i\varphi(x)}$ de valeur absolue constante = 1 ne sont pas de carré intégrable sur \mathbb{R} (bien qu'elles le soient sur un intervalle borné) et n'appartiennent donc pas à \mathcal{H} . Considérons alors l'opérateur a.a. de “position” $Qf(x) = xf(x)$. Il est trivial que si $Qf(x) = \lambda f(x)$ avec $f \neq 0$ dans $L^2(\mathbb{R})$ alors $f(x)$ doit être partout nulle sauf en λ puisque si $f(x) \neq 0$ alors $xf(x) = \lambda f(x)$ implique $x = \lambda$. Mais en même temps, comme il faut que f soit un élément non nul de $L^2(\mathbb{R})$, il faut que l'intégrale $\int_{\mathbb{R}} |f(x)|^2 dx$ soit $\neq 0$. Cela signifie que $f(x)$ est comme une distribution de Dirac attribuant un poids (ou une “masse”) non nulle au point λ et le poids 0 partout ailleurs. Effectivement, si l'on applique la définition de Q aux distributions de Dirac $\delta_\lambda(x)$, on obtient trivialement $Q\delta_\lambda(x) = \lambda\delta_\lambda(x)$. En effet δ_λ opère sur les fonctions test $h(x)$ par

$$\delta_\lambda(h(x)) = \int_{\mathbb{R}} \delta(x - \lambda) h(x) dx = h(\lambda)$$

et donc $Q\delta_\lambda$ opère par

$$Q\delta_\lambda(h(x)) = \int_{\mathbb{R}} x\delta(x - \lambda) h(x) dx = \lambda h(\lambda) = \lambda\delta_\lambda(h(x))$$

et par conséquent

$$Q\delta_\lambda(x) = \lambda\delta_\lambda(x) .$$

Mais comme $\delta_\lambda \notin \mathcal{H} = L^2(\mathbb{R})$, ce *n'est pas* un vecteur propre de Q opérant sur \mathcal{H} . Autrement dit Q *n'a pas* de valeur propre au sens strict bien que tout $\lambda \in \mathbb{R}$ soit une valeur propre “généralisée”.

fonctions continues réelles sur des intervalles compacts de \mathbb{R} par des polynômes (théorème de Stone-Weierstrass, cf. plus bas chapitre 17, section 3.1) ainsi que pour ses travaux sur les algèbres de Boole.

60. Comme $\delta_\lambda(x) = 0$ en dehors de λ et que $\{\lambda\}$ est de mesure nulle pour la mesure de Lebesgue dx , la classe de δ_λ serait celle de 0 si δ_λ appartenait à \mathcal{H} . Si ce n'est pas le cas c'est que les $f \in \mathcal{H}$ peuvent certes être très irrégulières mais sont des fonctions. Or $\delta_\lambda(x)$ étant infinie en λ , elle n'est pas une fonction mais, répétons-le encore une fois, une distribution au sens de Laurent Schwartz.

Toutefois, ceci n'empêche pas de considérer le Hilbert de dimension 1 $\mathcal{H}_\lambda = \mathbb{C}\delta_\lambda$ comme un espace propre de la valeur propre généralisée λ . Or, on peut interpréter l'expression $f(\lambda)$ comme la formulation "positionnelle" de la fonction abstraite $f \in \mathcal{H}$, à savoir la formulation qui *évalue* f en λ . En termes des distributions $\delta_\lambda(x)$ l'intégrale $f(\lambda) = \int_{\mathbb{R}} f(x) \delta_\lambda(x) dx$ signifie que la valeur $f(\lambda)$ de f au point λ est la "mesure" de f par δ_λ , autrement dit la "projection généralisée" $P_\lambda(f)$ de f sur l'espace propre généralisé \mathcal{H}_λ . *Si l'on savait* définir rigoureusement une sorte de "somme directe orthogonale"

$$\begin{aligned}\mathcal{H} &= \boxplus_{\lambda \in \mathbb{R}} \mathcal{H}_\lambda, \\ f &= \boxplus_{\lambda \in \mathbb{R}} f(\lambda) \delta_\lambda,\end{aligned}$$

on pourrait alors donner un sens à la "décomposition spectrale"

$$Q = \boxplus_{\lambda \in \mathbb{R}} \lambda P_\lambda$$

en une somme $\boxplus_{\lambda \in \mathbb{R}}$ d'opérateurs de multiplication par un scalaire.⁶¹ Mais cette idée doit être remaniée si l'on veut que les \mathcal{H}_λ soient des vrais sous-espaces de \mathcal{H} .

De même, considérons l'opérateur a.a. de "moment" $D = -i\partial_x$ sur $L^2(\mathbb{R})$ (dont nous avons parlé plus haut sur $L^2([0, 1])$ avec des conditions au bord). Pour tout $\lambda \in \mathbb{R}$ un vecteur propre évident est $e_p(x) = e^{ipx}$ avec sa valeur propre p . Mais les $e_p(x)$ étant de module constant = 1 ils ne sont pas de carré intégrable (il ne sont pas "normalisables"). Et pourtant, la transformée de Fourier $f(x) = \int_{\mathbb{R}} e^{ipx} \widehat{f}(p) dp$ ressemble bien à une décomposition spectrale (une diagonalisation) de A .

Pour préciser les choses, il faut d'abord faire la différence entre les *valeurs propres* de A et les éléments du spectre de A dits *valeurs spectrales*. L'ensemble des premières peut être vide alors que le spectre est \mathbb{R} tout entier. Un réel $\lambda \in \mathbb{R}$ est une valeur spectrale, i.e. appartient au spectre $\text{Spec}(A)$ de A , si l'inverse de $A - \lambda I$ n'est pas partout défini. Si λ est une valeur propre, automatiquement $\lambda \in \text{Spec}(A)$ puisque $A - \lambda I$ a un noyau de dimension ≥ 1 et n'est pas inversible (si f est un vecteur propre de λ alors $(A - \lambda I)f = 0$). Mais en dimension infinie un opérateur comme $(A - \lambda I)$ peut être injectif (i.e. il n'existe pas de $f \neq 0$ telle que $(A - \lambda I)f = 0$ et donc λ n'est pas valeur propre) sans être pour autant inversible (et dans ce cas λ est une valeur spectrale). C'est le cas de $Qf(x) = xf(x)$ de l'exemple précédent. Q est injectif car $Qf(x) = 0$ implique $f(x) = 0$ si $x \neq 0$ et donc, dans la mesure où changer f au point 0 ne change pas f comme élément de \mathcal{H} , $f = 0$ dans \mathcal{H} .⁶²

61. Nous utilisons le symbole \boxplus plutôt que le symbole \oplus^\perp utilisé plus haut car ici la somme n'est encore qu'une opération hypothétique.

62. Répétons que c'est là qu'intervient la différence clé entre les éléments f de \mathcal{H} et les distributions de Dirac. Le fait que $f = 0$ dans \mathcal{H} signifie simplement que $\int_{\mathbb{R}} |f(x)|^2 dx = 0$. Mais cela n'implique pas du tout que f soit identiquement nulle; f peut avoir des valeurs non

Q n'a pas de valeurs propres relativement à \mathcal{H} mais seulement des valeurs propres "généralisées" alors que, en revanche, tout $\lambda \in \mathbb{R}$ appartient au spectre. En effet, dire que $Q - \lambda I$ est inversible signifie que pour tout $f \in \mathcal{H}$ il existe $h \in \mathcal{H}$ tel que $(Q - \lambda I)^{-1} f = h$, i.e. $f = (Q - \lambda I)h$, i.e. $f(x) = (x - \lambda)h(x)$. Cela impose la contrainte $f(\lambda) = 0$ et donc $(Q - \lambda I)^{-1}$ n'est pas partout défini. C'est parce que $\text{Spec}(Q) = \mathbb{R}$ que Q peut être un opérateur mesurant la position d'une particule quantique.

Il en va de même pour l'opérateur $D = -i\partial_x$. Ses vecteurs propres généralisés $e_p(x) = e^{ipx}$ n'appartenant pas à \mathcal{H} il n'existe pas de valeurs propres. Néanmoins, $\text{Spec}(D) = \mathbb{R}$ tout entier. En effet dire que $D - pI$ est inversible signifie que pour tout $f \in \mathcal{H}$ il existe $h \in \mathcal{H}$ tel que $(D - pI)^{-1} f = h$, i.e. $f = (D - pI)h$, i.e. $f(x) = -i\partial_x h(x) - ph(x)$. Pour $f = 0$ cela impose $h = e_p \notin \mathcal{H}$ et donc $(D - pI)^{-1}$ n'est pas partout défini. C'est parce que $\text{Spec}(D) = \mathbb{R}$ que D peut être un opérateur mesurant le moment d'une particule quantique.

Comme nous l'avons vu, l'ensemble R_A complémentaire de $\text{Spec}(A)$ est l'ensemble des λ tels que $A - \lambda I$ soit inversible. $(A - \lambda I)^{-1} = R_\lambda$ s'appelle alors la "résolvante" de A . Comme le fait d'être inversible est une propriété ouverte⁶³, R_A est un ouvert de \mathbb{R} et le spectre $\text{Spec}(A)$ est donc un fermé de \mathbb{R} .

Si le cas des opérateurs a.a. *compacts* généralise directement le cas de la dimension finie c'est parce que, si A est compact, le fait que $A - \lambda I$ soit inversible équivaut au fait que $A - \lambda I$ soit injective. Les valeurs spectrales sont des valeurs propres et l'on a simplement une infinité dénombrable de valeurs propres λ_n . Un théorème – le théorème de Hilbert-Schmidt – dit alors que cet ensemble est borné par la plus grande valeur propre et s'accumule vers 0 et seulement vers 0. À ces valeurs propres se trouve associée une base propre orthonormée φ_n dans laquelle A s'écrit $Af = \sum_n \lambda_n \langle f, \varphi_n \rangle \varphi_n$.

7.5. Vecteurs propres approximatifs et théorème de Stone

La bonne stratégie pour définir correctement la décomposition spectrale de A a.a. consiste à utiliser non pas des vecteurs propres "généralisés" qui n'appartiennent pas à \mathcal{H} mais des états propres "approximatifs" qui, eux, appartiennent à \mathcal{H} mais avec une valeur propre "approximative". Par exemple, dans le cas précédent restreint à l'intervalle $[0, 1]$, on peut remplacer les Dirac δ_λ par un espace \mathcal{G}_λ d'approximations gaussiennes de δ_λ , $G_\lambda \in \mathcal{H}$, centrées sur λ et très "piquées" (i.e. de variance très faible) qui soit stable par $Qf(x) = xf(x)$. Les QG_λ seront alors très proches de λG_λ . En considérant suffisamment d'espaces propres approximatifs \mathcal{G}_λ de ce type et

nulles. En revanche δ_λ est nulle partout sauf en λ alors que son intégrale est = 1, ce qui est impossible pour $f \in \mathcal{H}$.

63. En dimension finie c'est la condition $\det(A - \lambda I) = \Pi_A(\lambda) \neq 0$ qui est trivialement une condition ouverte.

des projecteurs associés P_λ , on pourra espérer obtenir une décomposition spectrale de A de la forme

$$A = \int_{\mathbb{R}} \lambda d\sigma_A(\lambda)$$

pour une certaine *mesure* σ_A sur les boréliens \mathfrak{B} de \mathbb{R} qui soit de support $\text{Spec}(A)$ et à valeurs projecteurs (PVM : “projection-valued measure”), intégrale qui généraliserait la somme

$$A = \bigoplus_{\lambda_i \in \text{Spec}(A)} \lambda_i P_{\lambda_i}$$

où σ_A est la somme des mesures de Dirac placées aux λ_i avec pour valeur en λ_i le projecteur P_{λ_i} .⁶⁴ C’est le contenu du *théorème de décomposition spectrale de Stone*.

Pour pouvoir l’énoncer, il faut d’abord définir ce que signifie pour un opérateur a.a. A de posséder un spectre “simple”. On considère le groupe à 1-paramètre

$$U(t) = e^{-itA}$$

d’opérateurs unitaires (unitaires car A est a.a.). Soit ξ un vecteur de \mathcal{H} et \mathcal{H}_ξ le sous-espace fermé engendré par les $U(t)\xi$. On dit que ξ est un vecteur “cyclique” si $\mathcal{H}_\xi = \mathcal{H}$ et que le spectre de A est “simple” s’il existe un ξ cyclique.⁶⁵

Théorème spectral de Stone (version simple). Si A est a.a. de spectre simple, il existe toujours une mesure spectrale σ_A sur les boréliens \mathfrak{B} de $\mathbb{R}_{(\lambda)}$ (le \mathbb{R} de variable λ), qui est à valeurs projecteurs et de support $\text{Spec}(A)$, telle que $A = \int_{\mathbb{R}} \lambda d\sigma_A(\lambda)$ ($= \int_{\text{Spec}(A)} \lambda d\sigma_A(\lambda)$). Qui plus est, on peut définir un *calcul fonctionnel* au sens où, si $F : \mathbb{R} \rightarrow \mathbb{C}$ est une fonction \mathfrak{B} -mesurable, alors $F(A) = \int_{\mathbb{R}_{\lambda}} F(\lambda) d\sigma_A(\lambda)$. Cette PVM est unique. La décomposition spectrale vient d’une application *unitaire* K entre $L^2(\mathbb{R}_{(\lambda)}, d\sigma_A)$ et \mathcal{H} telle que, si $\varphi(\lambda) \in L^2(\mathbb{R}_{(\lambda)}, d\sigma_A)$ et si $K\varphi(\lambda) \in \mathcal{D}(A)$, alors $K^{-1}AK\varphi(\lambda) = \lambda\varphi(\lambda)$. \diamond

En transférant les opérations de A dans $\mathcal{D}(A)$ à des opérations de multiplication dans $L^2(\mathbb{R}_{(\lambda)}, d\sigma_A)$, le théorème spectral encode l’action *fonctionnelle* de A dans la *géométrie* de sa mesure spectrale.

La mesure σ_A associe à tout borélien \mathfrak{b} de $\mathbb{R}_{(\lambda)}$ un projecteur $\sigma_A(\mathfrak{b}) = P_{\mathfrak{b}}$ de \mathcal{H} avec les mêmes contraintes que celles des mesures normales :

- (i) $P_{\emptyset} = 0$ et $P_{\mathfrak{b}} = 0$ pour tout \mathfrak{b} disjoint de $\text{Spec}(A)$,
- (ii) $P_{\mathbb{R}} = P_{\text{Spec}(A)} = Id_{\mathcal{H}}$,
- (iii) $P_{\mathfrak{b}}$ et $P_{\mathfrak{b}'}$ sont orthogonaux si \mathfrak{b} et \mathfrak{b}' sont disjoints,
- (iv) $P_{\cup \mathfrak{b}_n} = \bigoplus P_{\mathfrak{b}_n}$ pour toute \cup dénombrable de \mathfrak{b}_n disjoints (σ -additivité), et

64. Comme σ_A est de support $\text{Spec}(A)$, les intégrales $\int_{\mathbb{R}} F(\lambda) d\sigma_A(\lambda)$ sont les mêmes que les intégrales $\int_{\text{Spec}(A)} F(\lambda) d\sigma_A(\lambda)$.

65. C’est en quelque sorte l’opposé d’un vecteur propre de A car si $A\xi = \lambda\xi$, $U(t)\xi = e^{-it\lambda\xi}$ et \mathcal{H}_ξ est le cercle faisant tourner ξ dans $\mathbb{C}\xi$.

(v) pour tout $f, h \in \mathcal{H}$, $\mathfrak{b} \mapsto \langle P_{\mathfrak{b}}(f), h \rangle_{\mathcal{H}}$ est une mesure classique $\sigma_{A,f,h}$ sur $\text{Spec}(A)$ à valeurs dans \mathbb{C} .

En partant de

$$\int_{\text{Spec}(A)} \chi_{\mathfrak{b}}(\lambda) d\sigma_A(\lambda) = \sigma_A(\mathfrak{b}) = P_{\mathfrak{b}}$$

pour les fonctions caractéristiques $\chi_{\mathfrak{b}}$ des \mathfrak{b} et en généralisant par combinaisons linéaires, limites, densité, on peut alors intégrer comme pour les mesures normales et, pour toute fonction \mathfrak{B} -mesurable $\varphi(\lambda)$, définir le projecteur P_{φ} . On montre immédiatement que les projecteurs commutent, que l'adjoint $(P_{\varphi})^* = P_{\bar{\varphi}}$ et que

$$\langle P_{\varphi}(f), h \rangle_{\mathcal{H}} = \int_{\text{Spec}(A)} \varphi(\lambda) d\sigma_{A,f,h}(\lambda) .$$

Remarque. Cette mesure σ_A est a priori très différente de la mesure de Lebesgue dx . Dans le cas de dimension finie elle a pour support l'ensemble fini des valeurs propres. Dans le cas des opérateurs compacts elle est composée d'atomes ce qui explique que l'on puisse directement généraliser la théorie en dimension finie. \square

Esquissons brièvement la preuve du théorème de Stone (cf. Taylor [524]). On considère le groupe à 1-paramètre d'opérateurs unitaires $U(t) = e^{-itA}$. Soit ξ un vecteur cyclique de \mathcal{H} avec $\mathcal{H}_{\xi} = \mathcal{H}$. Le théorème repose sur la transformée de Fourier de la fonction $\zeta(t) = \langle e^{-itA}\xi, \xi \rangle$. Comme $U(t)$ est unitaire, $\zeta(t)$ est bornée par $\|\xi\|^2$ et donc $\zeta(t) \in L^{\infty}(\mathbb{R}_t)$. Cela implique que sa transformée de Fourier $\widehat{\zeta}(\tau)$ est une distribution tempérée définissant une mesure $\widehat{\zeta}(\tau)$ sur $\mathbb{R}_{(\tau)}$ telle que

$$\langle e^{-itA}\xi, \xi \rangle = \int_{\mathbb{R}} e^{it\tau} d\widehat{\zeta}(\tau) .$$

C'est à partir de $\widehat{\zeta}(\tau)$ que l'on construit σ_A . Considérons par exemple $D = -i\partial_x$ dans $L^2(\mathbb{R}, dx)$ et donc

$$U(t) = e^{-itD} = e^{-t\partial_x} .$$

Nous verrons plus bas à la section 1.3 du chapitre 17 que $U(t)$ est l'opérateur de translation

$$U(t)f(x) = f(x-t) ,$$

dont le générateur infinitésimal est $\mathfrak{t} = -\partial_x = -iD$. On remarque que si δ_0 est la distribution de Dirac en 0, $U(t)\delta_0 = \delta_t$ et que par conséquent, si l'on faisait comme si les δ_t appartenait à \mathcal{H} , δ_0 serait un vecteur cyclique. Cela n'est pas le cas, mais on peut trouver des approximations aussi bonnes que l'on veut de δ_0 (par exemple des gaussiennes assez étroites), qui, elles, sont des vecteurs cycliques ξ . Considérons un tel vecteur $\xi(x)$ dont la transformée de Fourier $\widehat{\xi}(\omega)$ soit partout $\neq 0$ (c'est

possible car celle de δ_0 est, à normalisation près, la fonction constante = 1). On a

$$\zeta(t) = \langle \xi, e^{-itD} \xi \rangle = \int_{\mathbb{R}} \xi(x) \overline{\xi(x-t)} dx = \overline{\xi^-} * \xi$$

(avec $\xi^-(x) = \xi(-x)$)⁶⁶. et donc

$$\widehat{\zeta}(\tau) = \widehat{\xi}(\omega) \overline{\widehat{\xi}(\omega)} = \left| \widehat{\xi}(\omega) \right|^2 .$$

On peut alors mieux comprendre la différence entre valeurs propres et valeurs spectrales. La mesure d'intégration $d\sigma_A(\lambda)$ étant une mesure borélienne sur \mathbb{R} , on peut la comparer à la mesure de Lebesgue $d\lambda$. La généralisation du théorème de Radon-Nykodym exposée plus haut dans la section 2.1 dit que $d\sigma_A(\lambda)$ comprend une partie *discrète* de masses ponctuelles (des Dirac) correspondant aux valeurs propres et à leurs projecteurs et une partie continue se décomposant elle-même en une partie absolument continue par rapport à dx (au sens où $\sigma_{A,x} = \sigma_{A,f,h}$ pour $f(x) = h(x) = x$ est une mesure classique absolument continue par rapport à dx) et une partie singulière étrangère à dx (au sens où $\sigma_{A,x}$ est une mesure classique étrangère par rapport à dx). Cette tripartition de σ_A correspond à une décomposition de \mathcal{H} en somme directe de trois sous-espaces sur lesquels A agit des trois façons correspondantes.

Le théorème de Stone peut être relié à l'intuition initiale. En effet si Σ est une mesure sur \mathbb{R} , on peut considérer la PVM σ sur \mathbb{R} consistant à prendre une famille Σ -mesurable de Hilbert \mathcal{H}_λ paramétrée par $\lambda \in \mathbb{R}$, à construire l'intégrale (généralisant une somme directe)

$$\mathcal{H}_{\Sigma, \mathbb{R}} = \int_{\mathbb{R}}^{\oplus} \mathcal{H}_\lambda d\Sigma(\lambda)$$

et à définir $\sigma(S)$ comme le projecteur de $\mathcal{H}_{\Sigma, \mathbb{R}}$ sur

$$\mathcal{H}_{\Sigma, S} = \int_S^{\oplus} \mathcal{H}_\lambda d\Sigma(\lambda) .$$

Un théorème dit alors que toute PVM est *unitairement équivalente* à une telle PVM. Les \mathcal{H}_λ sont des espaces propres généralisés mais ils ne sont des sous-espaces de \mathcal{H} que si $\{\lambda\}$ est de Σ -mesure *non nulle*. Il pourra donc y avoir des points isolés du spectre correspondant à de vraies valeurs propres, des parties continues du spectre, tout cela intriqué de façon éventuellement compliquée.

Si le spectre de A n'est pas simple, i.e. s'il n'existe pas de vecteur cyclique ξ , on peut néanmoins considérer un \mathcal{H}_ξ , écrire $\mathcal{H} = \mathcal{H}_\xi \oplus^\perp \mathcal{H}_\xi^\perp$ et, comme \mathcal{H}_ξ^\perp est $U(t)$ -invariant, prendre un $(\mathcal{H}_\xi^\perp)_{\xi'}$ et continuer. On décompose ainsi \mathcal{H} en \oplus^\perp de \mathcal{H}_ξ , on

66. Rappelons que pour deux fonctions $u(x)$ et $v(x)$, le produit de convolution $u * v(t) = \int u(t-x)v(x) dx$ fait intervenir $t-x = -(x-t)$.

applique le théorème aux \mathcal{H}_ξ , et on obtient la forme générale du théorème spectral de Stone :

Théorème spectral de Stone (version générale). Si A est auto-adjoint, il existe toujours

(i) une mesure σ_A sur les boréliens \mathfrak{B} d'un espace Λ σ -compact (union dénombrable de compacts), qui est à valeurs projecteurs et

(ii) une fonction mesurable réelle $a(\lambda)$ sur Λ ,

telles que A soit équivalent à la multiplication par $a(\lambda)$. Autrement dit, il existe une application unitaire K entre $L^2(\Lambda, d\sigma_A)$ et \mathcal{H} , telle que si $\varphi(\lambda) \in L^2(\Lambda, d\sigma_A)$ et si $K\varphi(\lambda) \in \mathcal{D}(A)$, alors $K^{-1}AK\varphi(\lambda) = a(\lambda)\varphi(\lambda)$. \diamond

On peut d'ailleurs vérifier directement la réciproque, à savoir que, si l'on se donne un espace mesuré $(\Lambda, d\sigma)$ et une fonction mesurable réelle $a(\lambda)$, alors l'opérateur M_a de multiplication par $a(\lambda)$ dans $L^2(\Lambda, d\sigma)$ est auto-adjoint de domaine

$$\text{Dom}(M_a) = \{f \in L^2(\Lambda, d\sigma) \mid af \in L^2(\Lambda, d\sigma)\}$$

dense.⁶⁷

1. Densité de $\text{Dom}(M_a)$. Soit $f \in L^2(\Lambda, d\sigma)$. On considère les restrictions f_n de f aux fermés $F_n = \{\lambda \mid |a(\lambda)| \leq n\}$, i.e. $f_n = f\chi_{F_n}$ où χ_{F_n} est la fonction caractéristique de F_n . Par construction, $|af_n| \leq n|f|$ et donc $af_n \in L^2$ puisque $f \in L^2$. Par conséquent les $f_n \in \text{Dom}(M_a)$. Mais les f_n convergent simplement vers f et, comme $|f_n(\lambda)| \leq |f(\lambda)|$ pour tout λ , le *théorème de convergence dominée de Lebesgue* dit exactement que les f_n convergent vers f dans L^2 .
2. Fermeture de M_a . Soit f_n une suite de $\text{Dom}(M_a)$ convergeant vers $f \in L^2$ dans L^2 et supposons que les $M_a(f_n)$ convergent vers $g \in L^2$ dans L^2 . La réciproque du théorème de convergence dominée de Lebesgue implique qu'une sous-suite f_ν des f_n converge vers f simplement et donc que af_ν converge simplement vers af avec $af = g$ presque partout. Par conséquent $f \in \text{Dom}(M_a)$ et $M_a(f) = g$.
3. Auto-adjonction de M_a . M_a est symétrique car a est une fonction réelle. Pour démontrer que M_a est auto-adjoint on utilise la transformée de Cayley

$$W(M_a) = \frac{M_a - iI}{M_a + iI}$$

introduite plus haut et on montre que $W(M_a)$ est unitaire. Comme M_a est fermé, il suffit en fait de montrer que $\pm i$ ne sont pas des valeurs propres de M_a^* . Soit donc $f \in \ker(M_a^* - iI)$, i.e. $M_a^*(f) = if$. Il faut montrer que $f = 0$ dans L^2 . Pour tout $g \in \text{Dom}(M_a)$ on a

$$\langle M_a^*(f), g \rangle = \langle f, M_a(g) \rangle$$

67. Cf. le panorama [309] d'Olivier Lablée.

et donc

$$\langle if, g \rangle = \langle f, -ig \rangle = \langle f, ag \rangle ,$$

soit $\langle f, (a - i)g \rangle = 0$. Mais comme $\text{Dom}(M_a)$ est dense et que a est réelle, cela implique $f = 0$ dans L^2 . Idem pour $f \in \ker(M_a^* + iI)$.

8. Quantification géométrique

Nous allons maintenant faire un détour par ce que l'on appelle la "quantification géométrique" dans les formalismes de la mécanique quantique. Nous voudrions en effet brièvement expliciter ces architectures géométriques de la quantification car elles ressemblent beaucoup aux structures que nous avons introduites et développées pour nos modèles neurogéométriques. Certes les phénomènes naturels en jeu n'ont rien à voir entre eux mais les formalismes mathématiques mobilisés présentent des similitudes partielles.⁶⁸

Cela n'est pas un hasard. Les modèles neurogéométriques font intervenir des fibrations, des lagrangiens et des hamiltoniens, des structures de contact et des structures symplectiques, toutes structures du même ordre que celles que l'on rencontre en mécanique classique. Mais ces structures modélisent des connectivités d'architectures fonctionnelles (au sens neurophysiologique) et constituent donc l'*infrastructure géométrique* d'une analyse du signal de type *ondelettes* au moyen de profils récepteurs. On est donc en présence d'une géométrie qui "pilote" une analyse fonctionnelle (au sens mathématique) de signaux au moyen de décompositions spectrales d'opérateurs dans des espaces fonctionnels.

Or c'est précisément une telle complémentarité entre géométrie et analyse que la mécanique quantique a dû développer de façon extrêmement approfondie et innovante depuis sa fondation eu enrichissant de façon inouïe les formalismes classiques. Il est donc légitime et pertinent de faire un détour par ces innovations sophistiquées pour voir à quels transferts de formalismes ils peuvent conduire en neurogéométrie.

8.1. L'obstruction initiale

8.1.1. Les conditions de Dirac.

68. Ce détour par la quantification géométrique ne saurait évidemment être un détour par la mécanique quantique en tant que telle avec ses formalismes et ses problèmes philosophiques. Pour les lecteurs intéressés par ces questions, nous nous permettons de renvoyer aux travaux du *Collège de Physique et de Philosophie* (<http://cphi2.org>) fondé en 2009 par l'éminent physicien et philosophe des sciences Bernard d'Espagnat avec Michel Bitbol, Hervé Zwirn et nous-mêmes. Bernard d'Espagnat a dirigé pendant plusieurs années un séminaire à l'Institut de France dont les interventions ont été regroupées dans le volume *Le monde quantique. Les débats philosophiques de la physique quantique* [135]. Nous nous permettons de citer également notre étude [427] dans le débat avec Bernard d'Espagnat *Physique et Réalité*.

La théorie de la décomposition spectrale des opérateurs auto-adjoints sur un espace de Hilbert s'est considérablement développée sous la poussée de la mécanique quantique. Très vite, en particulier sous l'impulsion de Dirac, on s'est demandé comment on pouvait *quantifier* les systèmes mécaniques classiques en partant de ce que nous avons vu à la section 5.4 du chapitre 5 à propos du groupe de Heisenberg pour les opérateurs de position $\widehat{q}\psi(q) = q\psi(q)$ et de moment $\widehat{p}\psi(q) = -i\hbar\frac{\partial}{\partial q}\psi(q)$, $\psi(q) \in L^2(\mathbb{R}, \mathbb{C})$ étant la fonction d'onde dans la représentation de Schrödinger.

La procédure de quantification reviendrait alors à trouver, étant donné un système hamiltonien défini sur une variété symplectique (Σ, Θ) de dimension $2n$,⁶⁹ une correspondance \mathbb{R} -linéaire systématique $\widehat{\bullet} : \mathcal{C}^\infty(\Sigma, \mathbb{C}) \rightarrow \mathcal{L}(\mathcal{H})$ associant à toute fonction $f \in \mathcal{C}^\infty(\Sigma, \mathbb{C})$ un opérateur \widehat{f} sur un Hilbert \mathcal{H} de façon à ce que les conditions de Dirac suivantes soient satisfaites

$$(D1) \quad \widehat{1} = Id_{\mathcal{H}};$$

$$(D2) \quad \widehat{\{f, g\}} = \frac{i}{\hbar} [\widehat{f}, \widehat{g}] \quad (\text{i.e. } [\widehat{f}, \widehat{g}] = -i\hbar\widehat{\{f, g\}});^{70}$$

$$(D3) \quad \widehat{p} = -i\hbar\frac{\partial}{\partial q} \text{ pour des coordonnées canoniques ((D2) est donc trivialement le cas pour le moment } \widehat{p} \text{ et la position } \widehat{q} \text{ interprétée comme multiplication par } q), \text{ une variante étant } \widehat{p} = i\hbar\frac{\partial}{\partial q};$$

$$(D4) \quad (\widehat{f})^* = \widehat{\bar{f}} \text{ et donc si } f \text{ est réelle (i.e. est une observable), } \widehat{f} \text{ est auto-adjoint.}$$

Un guide heuristique est que, d'après la théorie spectrale exposée plus haut section 7.5 à propos du théorème de Stone, un opérateur auto-adjoint est en quelque sorte une fonction réelle sur son spectre.

Remarque. Nous nous restreindrons en grande partie dans cette section au cas particulier des espaces de phases standard $\Sigma = T^*M$ que sont les fibrés cotangents d'espaces de configuration M , et même au cas $\Sigma = T^*\mathbb{R}^n$ de $M = \mathbb{R}^n$ avec ses coordonnées canoniques $q = (q_j), p = (p_k), j, k = 1, \dots, n$. Les constructions et les calculs peuvent se généraliser puisqu'ils restent les mêmes localement dans les cartes locales symplectiques. Mais le passage du local au global peut faire intervenir des conditions, en particulier la satisfaction de propriétés *cohomologiques*, lorsque M est compacte et non simplement connexe (nous en verrons des exemples, entre autres celui d'orbites de représentations coadjointes de groupes de Lie, dans la section 8.5).

□

On part donc de la comparaison entre les cas classiques et les cas quantiques.

69. Dans cette section, pour tenir compte de son contexte très spécifique (mécanique quantique), nous noterons Θ plutôt que Ω la 2-forme symplectique et nous noterons θ la 1-forme du potentiel symplectique tel que (du moins localement) $d\theta = \Theta$.

70. Une variante est $\widehat{\{f, g\}} = \frac{1}{i\hbar} [\widehat{f}, \widehat{g}]$.

8.1.2. *Cas classique.*

Commençons par quelques rappels de formules que nous avons déjà souvent utilisées. Dans les calculs qui vont suivre on va utiliser les conventions de signes suivantes. D'abord la 2-forme symplectique sera (en notation condensée ⁷¹)

$$\Theta = dp \wedge dq$$

dérivée extérieure de la 1-forme fondamentale $\theta = pdq$ (dite aussi dans ce contexte physique “potentiel symplectique”) et donc

$$\Theta(X, Y) = \det \begin{pmatrix} dp(X) & dq(X) \\ dp(Y) & dq(Y) \end{pmatrix} = X_p Y_q - X_q Y_p$$

pour deux vecteurs tangents $X = X_q \partial_q + X_p \partial_p$ et $Y = Y_q \partial_q + Y_p \partial_p$. ⁷² Si

$$H : \Sigma \rightarrow \mathbb{R}$$

est un hamiltonien sur Σ , il engendre le champ hamiltonien de son “gradient symplectique”

$$X_H = (\partial_p H) \partial_q - (\partial_q H) \partial_p$$

défini par $i_{X_H} \Theta = -dH$ (cf. la section 3 du chapitre 12). En effet

$$\begin{aligned} i_{X_H} \Theta(Y) &= \Theta(X_H, Y) = X_{H,p} Y_q - X_{H,q} Y_p = -(\partial_q H) Y_q - (\partial_p H) Y_p \\ &= -dH(Y) . \end{aligned}$$

Le crochet de Poisson entre deux fonctions est défini par ⁷³

$$\begin{aligned} \{H, f\} &= \Theta(X_H, X_f) = i_{X_f} i_{X_H} \Theta = i_{X_H} df \\ &= -dH(X_f) = -X_f(H) = df(X_H) = X_H(f) \\ &= X_{H,p} X_{f,q} - X_{H,q} X_{f,p} = -(\partial_q H)(\partial_p f) - (\partial_p H)(-\partial_q f) \\ &= (\partial_p H)(\partial_q f) - (\partial_q H)(\partial_p f) . \end{aligned}$$

Notons d'ailleurs que si X_f est considéré comme un opérateur de dérivation sur des fonction $\psi(q, p)$ et g comme l'opérateur de multiplication par g , on a

$$[X_f, g] = X_f(g) = \{f, g\}$$

puisque

$$[X_f, g] \psi = X_f(g\psi) - gX_f(\psi) = X_f(g) \psi + gX_f(\psi) - gX_f(\psi) = X_f(g) \psi .$$

71. “Condensée” au sens où pdq signifie $\sum_j p_j dq_j$, etc.

72. Certains auteurs préfèrent utiliser la 2-forme opposée $dq \wedge dp$ et le potentiel symplectique $\theta = -pdq$. Il existe plusieurs variantes notationsnelles entre les auteurs et en particulier entre les mathématiciens et les physiciens : la forme $\Theta = \pm dp \wedge dq$ et le potentiel θ , les conditions de quantification avec i ou $-i$, et, pour les angles, $\varphi \in [0, 2\pi]$ ou $2\pi\varphi$, $\varphi \in [0, 1]$.

73. Une autre convention de signe utilise $i_{X_H} \Theta = -dH$ pour définir X_H et on alors $\{f, H\} = dH(X_f) = X_f(H)$.

Le crochet $\{f, g\}$ est une dérivation de l'algèbre $\mathcal{C}^\infty(\Sigma)$ qui satisfait la règle de Leibniz pour le produit des fonctions :

$$\{f, gh\} = \{f, g\}h + g\{f, h\} .$$

Mais il définit aussi une structure d'algèbre de Lie sur $\mathcal{C}^\infty(\Sigma)$ car

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0 .$$

$\mathcal{C}^\infty(\Sigma)$ est donc une algèbre de fonctions qui est en plus une algèbre de Lie, cette double structure en faisant ce que l'on appelle une *algèbre de Poisson*.

On dit alors qu'un ensemble d'observables $\{f_j\}$ est *complet* si les seules g commutant avec toutes les f_j sont les constantes.⁷⁴

On notera que d'après la formule d'homotopie de Cartan (chapitre 2, section 7.2.12)

$$\mathcal{L}_X\Theta = i_X d\Theta + d(i_X\Theta) = d(i_X\Theta)$$

puisque $d\Theta = 0$. Par conséquent si X laisse Θ invariante, alors $\mathcal{L}_X\Theta = 0$, $d(i_X\Theta) = 0$ et la 1-forme $i_X\Theta$ est *fermée*. Mais elle n'est pas nécessairement *exacte*. Si $X = X_H$ est un champ hamiltonien elle est exacte puisque $i_{X_H}\Theta = -dH$. Il s'agit en fait d'une équivalence. Cela montre que le groupe $H^1(\Sigma, \mathbb{R})$ de la cohomologie de de Rham de Σ , à savoir le quotient des 1-formes fermées par les 1-formes exactes qui ne dépend que de la structure différentiable, peut se récupérer au niveau symplectique comme le quotient des X laissant Θ invariante par les X_H .

L'hamiltonien H engendre ainsi un flot hamiltonien $\exp(tX_H) = \Phi_t$ préservant la forme symplectique Θ de Σ et par conséquent la forme volume de Liouville $\Lambda = \frac{1}{n!}\Theta^{\wedge n}$. Si $f : \Sigma \rightarrow \mathbb{R}$ est alors une observable donnée, son évolution le long des trajectoires $x(t) = \Phi_t(x(0))$ de X_H est

$$f(x(t)) = f_t(x(0)) = \Phi_t^* f(x(0))$$

et donc

$$\frac{d}{dt}\Phi_t^*(f) = \Phi_t^* (\{H, f\}) ,$$

soit

$$\frac{d}{dt}f_t = \{H, f_t\}$$

puisque H est constant le long des trajectoires.⁷⁵ Par conséquent, si H est invariant par X_f alors f est une intégrale première du flot de X_H (une "constante du mouvement") et le crochet de Poisson s'annule.

74. Les accolades $\{f_j\}$ symbolisent un ensemble et les accolades $\{f, g\}$ symbolisent un crochet. Les notations conventionnelles sont parfois ambiguës... Le contexte désambiguïse.

75. L'observable f est donnée et donc fixée en tant que fonction mais ses valeurs évoluent le long des trajectoires $x(t)$.

S'il existe n intégrales premières f_j (dont H) indépendantes et en involution (i.e. commutantes au sens où les crochets de Poisson $\{f_j, f_k\}$ sont tous nuls), le système hamiltonien est dit "intégrable" et, si les mouvements sont bornés, il existe alors, d'après un célèbre théorème de Vladimir Arnold, des variables canoniques "actions-angles" telles que l'hamiltonien H ne dépendent que des actions. D'après les équations de Hamilton, les variables d'action sont donc constantes le long des mouvements puisque les dérivées partielles de H par rapport aux variables angulaires sont nulles et les variables angulaires décrivent des mouvements quasi-périodiques sur des n -tores invariants de l'espace des phases.

8.1.3. Cas quantique.

Côté quantique, si \widehat{H} est un hamiltonien indépendant du temps⁷⁶, alors l'évolution à partir de $t = 0$ est donnée par le groupe à un paramètre sur \mathbb{R} d'opérateurs unitaires $U(t) = \exp(-\frac{i}{\hbar}t\widehat{H})$ qui satisfait l'équation de Schrödinger

$$i\hbar \frac{\partial}{\partial t} U(t) = \widehat{H} \circ U(t)$$

(i.e. $\frac{\partial}{\partial t} U(t) = -\frac{i}{\hbar} \widehat{H} \circ U(t)$).⁷⁷ Dans la représentation de Schrödinger où les états $\psi \in \mathcal{H}$ varient⁷⁸ mais où les observables sont des opérateurs fixes, on a $\psi(t) = U(t)\psi(0)$ qui intègre l'équation de Schrödinger

$$i\hbar \frac{\partial}{\partial t} \psi(t) = \widehat{H} \psi(t)$$

(i.e. $\frac{\partial}{\partial t} \psi(t) = -\frac{i}{\hbar} \widehat{H} \psi(t)$). Dans la représentation de Heisenberg où, au contraire, les états $\psi \in \mathcal{H}$ sont fixes mais où les observables \widehat{f} sont des opérateurs variables, ces dernières évoluent par conjugaison avec $U(t)$, i.e.

$$\widehat{f}(t) = U(t)^{-1} \circ \widehat{f}(0) \circ U(t)$$

et

$$\frac{\partial}{\partial t} \widehat{f}(t) = \frac{i}{\hbar} [\widehat{H}, \widehat{f}(t)]$$

76. Pour les systèmes mécaniques classiques, $H = \frac{p^2}{2m} + V(q)$ (en coordonnées canoniques (q, p)), et on obtient donc $\widehat{H} = -\frac{\hbar^2}{2m} \Delta + V(\widehat{q})$.

77. Si \widehat{H} dépend du temps, les choses sont plus compliquées.

78. Les vecteurs d'états ne sont en fait définis qu'à une phase près et, physiquement, ce sont les "rayons" qui définissent l'état du système.

est l'équation d'évolution.⁷⁹

D'où la condition (D2) $\widehat{\{f, g\}} = \frac{i}{\hbar} [\widehat{f}, \widehat{g}]$ imposée par la comparaison classique/quantique entre $\frac{d}{dt} f_t = \{H, f_t\}$ et $\frac{d}{dt} \widehat{f}(t) = \frac{i}{\hbar} [\widehat{H}, \widehat{f}(t)]$.

8.1.4. Comment passer du commutatif au non commutatif ?

Un premier lien évident entre les deux cas est la correspondance des crochets au niveau des champs hamiltoniens : l'application linéaire $f \mapsto X_f$ de $\mathcal{C}^\infty(\Sigma, \mathbb{R})$ muni du crochet $\{f, g\}$ dans $\mathfrak{X}(\Sigma)$ (l'espace des champs de vecteurs sur Σ interprétés comme opérateurs de dérivation⁸⁰) muni du crochet $[X, Y]$ est un morphisme d'algèbres de Lie :

$$X_{\{f, g\}} = [X_f, X_g]$$

comme on peut le vérifier facilement.⁸¹ On a donc un lien entre $\mathcal{C}^\infty(\Sigma, \mathbb{R})$ muni du crochet $\{\bullet, \bullet\}$ et une algèbre d'opérateurs. On pourrait alors penser prendre pour Hilbert \mathcal{H} l'espace $L^2(\Sigma, \mathbb{C})$ et associer à une observable $f \in \mathcal{C}^\infty(\Sigma)$ l'opérateur $\widehat{f} = -i\hbar X_f$ ce qui permettrait trivialement de satisfaire les propriétés (D2) et (D3).

79. Symboliquement, en traitant les compositions comme des produits et en tenant compte du fait que, par définition, U commute avec \widehat{H} , on a

$$\begin{aligned} \frac{\partial}{\partial t} \widehat{f}(t) &= -\frac{\partial}{\partial t} (U(t)) U(t)^{-2} \circ \widehat{f}(0) \circ U(t) + U(t)^{-1} \circ \widehat{f}(0) \circ \frac{\partial}{\partial t} (U(t)) \\ &= -\left(-\frac{i}{\hbar} \widehat{H} \circ U(t)\right) \circ U(t)^{-2} \circ \widehat{f}(0) \circ U(t) + U(t)^{-1} \circ \widehat{f}(0) \circ \left(-\frac{i}{\hbar} \widehat{H} \circ U(t)\right) \\ &= \frac{i}{\hbar} \left(\widehat{H} \circ U(t)^{-1} \circ \widehat{f}(0) \circ U(t) - U(t)^{-1} \circ \widehat{f}(0) \circ \widehat{H} \circ U(t)\right) = \\ &= \frac{i}{\hbar} \left(\widehat{H} \circ \widehat{f}(t) - \widehat{f}(t) \circ \widehat{H}\right) = \frac{i}{\hbar} [\widehat{H}, \widehat{f}(t)]. \end{aligned}$$

80. Cf. chapitre 2, section 6.3.2.

81. Il est immédiat de le vérifier à la main.

- (i) $\{f, g\} = (\partial_p f)(\partial_q g) - (\partial_q f)(\partial_p g)$,
- (ii) $X_f = (\partial_p f) \partial_q - (\partial_q f) \partial_p$, $X_g = (\partial_p g) \partial_q - (\partial_q g) \partial_p$,
- (iii) $[X_f, X_g] = X_f X_g - X_g X_f$.

Donc

$$\begin{aligned} [X_f, X_g] &= ((\partial_p f) \partial_q X_g - (\partial_q f) \partial_p X_g) - ((\partial_p g) \partial_q X_f - (\partial_q g) \partial_p X_f) \\ &= (\partial_p f) [((\partial_{qp}^2 g) \partial_q + (\partial_p g) \partial_{q^2}^2) - ((\partial_{q^2}^2 g) \partial_p + (\partial_q g) \partial_{qp}^2)] \\ &\quad - (\partial_q f) [((\partial_{p^2}^2 g) \partial_q + (\partial_p g) \partial_{pq}^2) - ((\partial_{pq}^2 g) \partial_p + (\partial_q g) \partial_{p^2}^2)] \\ &\quad - (\partial_p g) [((\partial_{qp}^2 f) \partial_q + (\partial_p f) \partial_{q^2}^2) - ((\partial_{q^2}^2 f) \partial_p + (\partial_q f) \partial_{qp}^2)] \\ &\quad + (\partial_q g) [((\partial_{p^2}^2 f) \partial_q + (\partial_p f) \partial_{pq}^2) - ((\partial_{pq}^2 f) \partial_p + (\partial_q f) \partial_{p^2}^2)]. \end{aligned}$$

Mais cela ne conduit pas très loin puisque si f est constante alors X_f est nul⁸² et donc si le crochet de Poisson $\{f, g\} = c$ est constant alors le crochet de Lie $[X_f, X_g] = 0$ est nul et X_f et X_g commutent. C'est le cas des fonctions coordonnées q_j et p_k dont le crochet est

$$\{q_j, p_k\} = \sum_i (\partial_{p_i} q_j) (\partial_{q_i} p_k) - (\partial_{q_i} q_j) (\partial_{p_i} p_k) = -\delta_{jk}$$

et qui forment donc une algèbre de Heisenberg. En revanche les $X_{q_j} = -\partial_{p_j}$ et les $X_{p_k} = \partial_{q_k}$ ne forment pas l'algèbre de Heisenberg puisque les dérivées partielles ∂_{p_j} et ∂_{q_k} commutent. En fait le groupe exponentiant les X_{q_j} et les X_{p_k} est celui des translations de \mathbb{R}^{2n} respectivement par rapport à p_j et q_k , $X_1 = 0$ agissant trivialement. On voit que le problème de base est d'obtenir, conformément à (D1), (D2), $\widehat{\{f, g\}} = c.Id_{\mathcal{H}} = \frac{i}{\hbar} [\widehat{f}, \widehat{g}]$ et pour cela il faut rajouter des dimensions de façon à ce que les constantes puissent y agir de façon non triviale. Le prix du passage à des observables non commutatives est une augmentation de dimension liée à une extension centrale du groupe des translations comme groupe de symétries de l'espace des phases (cf. plus bas section 8.2).

Ce problème se traduit par un mauvais comportement général de l'application moment de l'action symplectique de groupes de Lie G sur (Σ, Θ) . Nous avons vu à la section 3.2 du chapitre 12 que si (Σ, Θ) est une variété symplectique sur laquelle un groupe de Lie G agit symplectiquement par $(g \in G, x \in \Sigma) \mapsto g(x) \in \Sigma$, alors tout $\gamma \in \mathcal{G}$ (\mathcal{G} algèbre de Lie de G) engendre, au moyen de l'application linéaire tangente en $(e \in G, x \in \Sigma)$, un champ $\gamma_\Sigma(x)$ sur Σ . L'application $\gamma \mapsto \gamma_\Sigma$ est un morphisme d'algèbres de Lie, autrement dit $[\gamma, \eta]_\Sigma = [\gamma_\Sigma, \eta_\Sigma]$. La 1-forme $i_{\gamma_\Sigma} \Theta$ est fermée (γ_Σ est un champ localement hamiltonien) et si elle est de plus exacte alors $i_{\gamma_\Sigma} \Theta = dF_\gamma$ pour une certaine fonction F_γ . Si F_γ existe pour tout γ , autrement dit si l'action de G sur (Σ, Θ) est non seulement symplectique mais qui plus est hamiltonienne, on

On vérifie alors que les coefficients des opérateurs de dérivées secondes s'annulent et qu'il reste simplement

$$\begin{aligned} [X_f, X_g] &= (\partial_p f) [(\partial_{qp}^2 g) \partial_q - (\partial_{q^2}^2 g) \partial_p] - (\partial_q f) [(\partial_{p^2}^2 g) \partial_q - (\partial_{pq}^2 g) \partial_p] \\ &\quad - (\partial_p g) [(\partial_{qp}^2 f) \partial_q - (\partial_{q^2}^2 f) \partial_p] + (\partial_q g) [(\partial_{p^2}^2 f) \partial_q - (\partial_{pq}^2 f) \partial_p] \\ &= [(\partial_p f) (\partial_{qp}^2 g) - (\partial_q f) (\partial_{p^2}^2 g) - (\partial_p g) (\partial_{qp}^2 f) + (\partial_q g) (\partial_{p^2}^2 f)] \partial_q \\ &\quad [- (\partial_p f) (\partial_{q^2}^2 g) + (\partial_q f) (\partial_{pq}^2 g) + (\partial_p g) (\partial_{q^2}^2 f) - (\partial_q g) (\partial_{pq}^2 f)] \partial_p \\ &= \partial_p [(\partial_p f) (\partial_q g) - (\partial_q f) (\partial_p g)] \partial_q - \partial_q [(\partial_p f) (\partial_q g) - (\partial_q f) (\partial_p g)] \partial_p \\ &= \partial_p \{f, g\} \partial_q - \partial_q \{f, g\} \partial_p \\ &= X_{\{f, g\}} . \end{aligned}$$

82. Le noyau de $f \mapsto X_f$ est exactement l'espace des f constantes car si $X_f = (\partial_p f) \partial_q - (\partial_q f) \partial_p = 0$ alors $\partial_p f = \partial_q f = 0$ et f est constante.

peut alors définir l'application moment $j : \Sigma \rightarrow \mathcal{G}^*$ par $j(x)(\gamma) = F_\gamma(x)$ pour tout $x \in \Sigma$ et les champs $\gamma_\Sigma(x) = X_{F_\gamma}(x)$ sont *globalement* hamiltoniens. Le problème est que les fonctions $F_\gamma(x)$ ne forment pas en général une sous-algèbre de Lie de l'algèbre de Poisson $\mathcal{C}^\infty(\Sigma, \mathbb{R})$ pour le crochet $\{f, g\}$. On montre en effet (cf. [353] et [509]) que $\{F_\gamma, F_\eta\} = F_{[\gamma, \eta]} + c(\gamma, \eta)$ où $c(\gamma, \eta)$ est une fonction constante sur Σ formant un "cocycle symplectique" de \mathcal{G} pour son action coadjointe, c'est-à-dire une forme bilinéaire antisymétrique satisfaisant l'identité de Jacobi

$$c([\gamma, \delta], \eta) + c([\delta, \eta], \gamma) + c([\eta, \gamma], \delta) = 0 .$$

Quand ce cocycle n'est pas nul, il faut donc considérer au moins la sous-algèbre de Lie engendrée par les F_γ , $\gamma \in \mathcal{G}$, et la fonction constante 1, autrement dit une *extension d'algèbres de Lie* de \mathcal{G} par \mathbb{R} .

Remarque. C'est ce que nous avons fait dès le début de cet ouvrage avec notre modèle de base des 1-jets (x, y, p) des courbes planes. L'espace des (x, p) est l'espace des phases $T^*\mathbb{R}$ muni de sa structure symplectique canonique et $y = \int p dx$ permet d'étendre cette structure symplectique à une structure de contact, celle du groupe de Heisenberg. □

On notera par ailleurs que le choix $\widehat{f} = -i\hbar X_f$ se comporte très mal par rapport au produit des fonctions : il ne le transforme pas en composition des opérateurs. Il est adapté au crochet de Poisson des fonctions et non pas à leur produit. En effet X_f est un opérateur différentiel du premier ordre et donc $X_{f^2} \neq (X_f)^2$ puisque $(X_f)^2$ est du deuxième ordre.

8.2. Préquantification

Trouver une bonne quantification est donc très loin d'être évident et plusieurs solutions ont été proposées, les plus connues étant celle d'Hermann Weyl (1927) et celle de la *quantification géométrique* (1960). On peut les compléter par les approches d'Eugene Wigner (qui introduisit les "fonctions de Wigner" en 1932), de José Moyal (qui approfondit à la fin des années 1940 la mécanique quantique dans l'espace des phases dans ses analogies avec la mécanique statistique classique et introduisit, indépendamment de Hilbrand Groenewold, le $*$ -produit non commutatif $f * g$ ⁸³) ou de Richard Feynman. Nous allons d'abord nous intéresser à celle de la quantification géométrique et reviendrons sur celle de Weyl dans la section 9 à propos des opérateurs pseudo-différentiels.

83. $f * g$ est une \hbar -déformation de fg telle que $[f, g] = f * g - g * f$ soit une \hbar -déformation de $i\hbar\{f, g\}$.

Fondée par Jean-Marie Souriau⁸⁴ (cf. [509]), la quantification géométrique s'est puissamment développée à partir de 1960 grâce aux efforts conjugués et aux travaux convergents de grands géomètres comme Vladimir Arnold (cf. [20]), Hans Duistermaat (cf. [155]), Victor Guillemin (cf. [237]), Lars Hörmander (cf. [256]), Alexandre Kirillov (cf. [288], [289]), Bertram Kostant⁸⁵ (cf. [305]), Jean Leray (cf. [325]), Viktor Maslov (cf. [356]), Alan Weinstein (cf. [560]), et beaucoup d'autres. Nous avons déjà rencontré plusieurs de ces grands créateurs dans ce qui précède. Nous nous appuyons sur leurs travaux ainsi que sur des exposés synthétiques comme le séminaire [353] de Charles Michel Marle et le cours [57] de Matthias Blau. Le lecteur pourra se référer aussi à des traités classiques comme celui de Nicholas Woodhouse [569].

La première idée développée, celle dite de la *préquantification*, consiste à représenter les observables classiques f par des opérateurs \hat{f} auto-adjoints sur des Hilbert \mathcal{H} de sections de carré intégrable de *fibrés* de base Σ munis de structures appropriées.

C'est la structure d'algèbre de Lie de $(\mathcal{C}^\infty(\Sigma), \{\bullet, \bullet\})$ que l'on transformera en algèbre de Lie d'opérateurs sur \mathcal{H} . Et en plus d'une correspondance $f \rightarrow \hat{f}$ il faudra aussi construire le *produit scalaire* sur \mathcal{H} qui est un élément essentiel de sa structure. On voudra aussi satisfaire une certaine condition "d'irréductibilité" s'ajoutant aux conditions de Dirac (D1)-(D4) :

(D5 = Irr) tout ensemble complet $\{f_j\}$ d'observables se transforme en un ensemble complet d'opérateurs $\{\hat{f}_j\}$, ce qui signifie que l'algèbre de Poisson des f est représentée dans \mathcal{H} de façon irréductible.

Nous allons voir que satisfaire directement les cinq conditions (D1)-(D5) est trop difficile et que trouver des bons compromis est extrêmement technique. Qui plus est, une fois obtenues de telles constructions ayant pour base la géométrie symplectique et les fibrations, il faut ensuite tenir compte du fait qu'en mécanique symplectique on travaille dans des *espaces de phases* de dimension $2n$ de coordonnées canoniques (p, q) alors qu'en mécanique quantique on travaille avec des fonctions $\psi(q)$ ou $\psi(p)$ définies sur les *espaces de configurations* et ne dépendant que de n variables. Il faut donc trouver un moyen pour *diviser par 2* le nombre de variables. Cela s'effectue au moyen de *polarisations*.

84. Jean-Marie Souriau (1922-2012) était ancien élève de l'ENS et doctorant d'André Lichnerowicz. Grand prix Jaffé de l'Académie des Sciences et grand prix scientifique de la Ville de Paris, il introduisit non seulement l'application moment et la quantification géométrique mais aussi la difféologie ainsi que la thermodynamique des groupes de Lie et la théorie géométrique de l'information. Depuis 2013, des conférences GSI sur la *Geometric Science of Information* sont organisées régulièrement par Frédéric Barbaresco spécialiste de ces questions (cf. [24]).

85. Bertram Kostant (1928-2017) était un doctorant d'Irving Segal. À Princeton, à Berkeley et au MIT, il a été un éminent spécialiste de la théorie de la représentation des groupes de Lie et de la géométrie symplectique.

Pour $T^*\mathbb{R}^n$, l'algèbre de Lie des fonctions coordonnées q_j et p_k contient les constantes et est l'algèbre de Heisenberg avec les relations de commutation

$$\{q_j, p_k\} = -\delta_{j,k}.$$

Par exponentiation on obtient le groupe de Heisenberg. La quantification donne d'après (D2)

$$[\widehat{q}_j, \widehat{p}_k] = i\hbar\delta_{j,k}Id$$

et, dans ce cas, le théorème fondamental de Stone-von Neumann que nous étudierons en détail plus bas à la section 5 du chapitre 17 implique, qu'à équivalence unitaire près, on ait $\widehat{q}_j = q_j$ (multiplication par q_j) et $\widehat{p}_k = -i\hbar\partial_{q_k}$ (condition D3) agissant sur des fonctions d'ondes $\psi(q)$.

Dans ce cas, d'autres observables se quantifient correctement. Par exemple l'énergie cinétique est, à un facteur près (en notation condensée) p^2 et sa quantification donne bien le laplacien Δ par rapport aux coordonnées q_j dans l'équation de Schrödinger : $\widehat{p}^2 = -\hbar^2\Delta$. On vérifie que la condition (D2) est bien satisfaite pour les commutations de \widehat{p}^2 avec \widehat{q} (évidemment, $[\widehat{p}^2, \widehat{p}] = 0 = -i\hbar\{p^2, p\}$ puisque p commute avec toutes ses puissances)

$$\begin{aligned} [\widehat{p}^2, \widehat{q}] &= -\hbar^2\Delta(q\psi(q)) + q\hbar^2\Delta(\psi(q)) \\ &= -\hbar^2(q\partial_{q^2}^2\psi(q) + 2\partial_q\psi(q)) + q\hbar^2\partial_{q^2}^2\psi(q) \\ &= -2\hbar^2\partial_q\psi(q) = -2i\hbar\widehat{p} \\ \{p^2, q\} &= 2p \\ -i\hbar\widehat{\{p^2, q\}} &= -2i\hbar\widehat{p} = [\widehat{p}^2, \widehat{q}] . \end{aligned}$$

Mais on constate tout de suite qu'il faut faire appel à des heuristiques pour définir la quantification d'observables un tout petit peu plus compliquées. Les observables quadratiques forment une sous-algèbre de Poisson avec les relations de commutation

$$\{pq, p^2\} = -2p^2, \{pq, q^2\} = -2q^2, \{p^2, q^2\} = 4pq .$$

Comment quantifier par exemple pq (cf. Blau [57])? Poser $\widehat{pq} = \widehat{p}\widehat{q}$ ne convient certainement pas car $pq = qp$ alors que $\widehat{p}\widehat{q} \neq \widehat{q}\widehat{p}$. Cette non commutativité oblige à ordonner les produits et à poser plutôt

$$\widehat{pq} = \frac{1}{2}(\widehat{p}\widehat{q} + \widehat{q}\widehat{p}) = \widehat{qp} .$$

On vérifie alors que l'on a bien

$$[\widehat{pq}, \widehat{p}^2] = -i\hbar\widehat{\{pq, p^2\}} = -i\hbar(-2)\widehat{p}^2 = 2i\hbar(-i\hbar)^2\partial_{q^2}^2 = -2i\hbar^3\partial_{q^2}^2 .$$

En effet

$$\begin{aligned} [\widehat{p}q, \widehat{p}^2] &= \frac{1}{2} (\widehat{p}q + \widehat{q}p) \widehat{p}^2 - \frac{1}{2} \widehat{p}^2 (\widehat{p}q + \widehat{q}p) \\ &= \frac{1}{2} (\widehat{p}q\widehat{p}^2 + \widehat{q}p\widehat{p}^2 - \widehat{p}^2\widehat{p}q - \widehat{p}^2\widehat{q}p) \\ &= \frac{1}{2} (\widehat{p} [\widehat{q}, \widehat{p}^2] + [\widehat{q}, \widehat{p}^2] \widehat{p}) = 2i\hbar\widehat{p}^2 . \end{aligned}$$

Mais pour les fonctions *cubiques* de gros problèmes apparaissent ce qui amène à se poser la question de savoir quelles observables classiques sont quantifiables de façon naturelle.

Remarque. Pour $\Sigma = T^*\mathbb{R}^n$ les translations en q et p agissent transitivement et leur groupe est un groupe de symétrie naturel. Le flot hamiltonien Φ_t régissant l'évolution du système est un groupe à 1-paramètre de symplectomorphismes. L'algèbre de Heisenberg (l'algèbre de Poisson des q et p) est une extension centrale de l'algèbre commutative des translations et l'on a donc un groupe G d'ensemble sous-jacent $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$, autrement dit une extension de $\mathbb{R}^n \times \mathbb{R}^n$ par \mathbb{R} , qui est un groupe de symétrie universel pour les mouvements. On voudrait que, lors du processus de quantification, ce groupe de symétrie induise une représentation unitaire dans le Hilbert \mathcal{H} considéré qui associe à tout $\lambda \in \mathbb{R}$ l'opérateur unitaire de multiplication par un facteur de phase $\exp(\frac{i}{\hbar}\lambda)$ (cf. [353]). Cela se généralise au cas où Σ est un espace homogène sous l'action d'un groupe de Lie G , la G -action étant définie par les champs hamiltoniens X_{f_j} d'un ensemble complet d'observables $\{f_j\}$. On cherche alors une représentation irréductible de leur algèbre de Poisson dans un Hilbert. Mais si un équivalent du théorème de Stone-von Neumann n'est pas disponible, alors il y aura plusieurs quantifications possibles (cf. [57], p.16). \square

Explicitons les bases de la préquantification. Comme nous voulons conserver la notation $f \rightarrow \widehat{f}$ pour la "vraie" quantification, nous noterons $f \rightarrow \widetilde{f}$ la préquantification. Nous avons vu que poser $\widetilde{f} = -i\hbar X_f$ ne convient pas car cela satisfait les conditions de Dirac (D2)⁸⁶, (D3) et (D4) mais pas (D1) $\widehat{1} = Id_{\mathcal{H}}$. On peut alors penser à $\widetilde{f} = -i\hbar X_f + f$ qui garantit trivialement $\widehat{1} = Id_{\mathcal{H}}$ mais ce choix ne satisfait pas (D2) puisque, en utilisant

$$[X_f, g] = X_f(g) = \{f, g\}$$

86. Avec les réserves quant au comportement par rapport au produit des fonctions.

et $[f, g] = 0$ (les opérateurs de multiplication commutent), on a

$$\begin{aligned}
[\tilde{f}, \tilde{g}] &= [-i\hbar X_f + f, -i\hbar X_g + g] \\
&= [-i\hbar X_f, -i\hbar X_g] + [-i\hbar X_f, g] + [f, -i\hbar X_g] + [f, g] \\
&= -\hbar^2 [X_f, X_g] - i\hbar \{f, g\} + i\hbar \{g, f\} \\
&= -\hbar^2 X_{\{f, g\}} - 2i\hbar \{f, g\} \\
-i\hbar \widetilde{\{f, g\}} &= -i\hbar (-i\hbar X_{\{f, g\}} + \{f, g\}) = -\hbar^2 X_{\{f, g\}} - i\hbar \{f, g\}
\end{aligned}$$

et qu'il existe donc un $-i\hbar \{f, g\}$ d'écart entre les deux expressions.

Il est remarquable (l'harmonie des calculs est toujours remarquable) que l'on puisse corriger cet écart au moyen du potentiel symplectique $\theta = pdq$ en prenant

$$\tilde{f} = -i\hbar X_f - pdq(X_f) + f = -i\hbar X_f - p\partial_p f + f$$

le terme correcteur venant du potentiel symplectique enlevant $p\partial_p f$ et le terme $pdq(X_f) - f$ étant le *lagrangien* associé par *transformée de Legendre* (cf. la section 2.1 du chapitre 12) à l'hamiltonien f .

On a toujours (D1) et pour (D2) on a maintenant la bonne égalité :

$$\begin{aligned}
[\tilde{f}, \tilde{g}] &= [-i\hbar X_f + (f - p\partial_p f), -i\hbar X_g + (g - p\partial_p g)] \\
&= [-i\hbar X_f, -i\hbar X_g] + [-i\hbar X_f, (g - p\partial_p g)] + [f - p\partial_p f, -i\hbar X_g] \\
&\quad + [f - p\partial_p f, g - p\partial_p g] \\
&= -\hbar^2 [X_f, X_g] - i\hbar \{f, g\} + i\hbar X_f (p\partial_p g) - i\hbar \{f, g\} \\
&\quad - i\hbar X_g (p\partial_p f) \\
&= -\hbar^2 X_{\{f, g\}} - 2i\hbar \{f, g\} + i\hbar \{f, p\partial_p g\} - i\hbar \{g, p\partial_p f\} \\
&= -\hbar^2 X_{\{f, g\}} - 2i\hbar \{f, g\} + i\hbar \{f, g\} + i\hbar p\partial_p \{f, g\} \\
&= -\hbar^2 X_{\{f, g\}} - i\hbar \{f, g\} + i\hbar p\partial_p \{f, g\} \\
-i\hbar \widetilde{\{f, g\}} &= -i\hbar (-i\hbar X_{\{f, g\}} - p\partial_p \{f, g\} + \{f, g\}) \\
&= -\hbar^2 X_{\{f, g\}} + i\hbar p\partial_p \{f, g\} - i\hbar \{f, g\} .
\end{aligned}$$

On a alors

$$\begin{cases} \tilde{p} = -i\hbar X_p - pdq(X_p) + p = -i\hbar \partial_q - p\partial_p p + p = -i\hbar \partial_q \\ \tilde{q} = -i\hbar X_q - pdq(X_q) + q = i\hbar \partial_p + q \end{cases}$$

et donc $\tilde{q}\psi(q) = q\psi(q)$ pour les $\psi(q)$ ne dépendant *que de q* (car alors $i\hbar \partial_p \psi(q) = 0$).

Cette solution est clairement beaucoup plus satisfaisante, mais elle ne l'est pas encore vraiment. En effet,

1. l'opérateur ∂_p commute avec les \tilde{q} et les \tilde{p} sans être pour autant constant, ce qui viole la condition (D5) d'irréductibilité,
2. outre le fait que ce choix de \tilde{f} se comporte évidemment toujours aussi mal du côté du produit des fonctions, il transforme les fonctions $\psi(q)$ ne dépendant que de q en fonctions dépendant de q et de p . Par exemple $\tilde{p}^2 = -2i\hbar p \partial_q - p^2$.

Pour essayer de surmonter ces difficultés et généraliser le cas $\Sigma = T^*\mathbb{R}^n$, les spécialistes (suivant un modèle proposé par Koopman, Van Hove et Segal entre 1930 et 1960, puis pleinement élaboré par Kostant [305] et Souriau [509]) ont proposé la procédure suivante que, comme nous l'avons dit, nous allons expliciter en suivant l'excellent exposé de Charles Michel Marle [353] et le non moins excellent cours de Matthias Blau [57]. Le lecteur pourra aussi consulter le traité de Woodhouse [569] déjà cité ainsi que le traité classique de Viktor Maslov [356].

Comme les fonctions d'onde quantiques sont à valeurs complexes (propriété *fondamentale* qui permet la modélisation des interférences quantiques), on introduit un fibré (localement trivial) *en droites complexes* $\delta : L \rightarrow \Sigma$ au-dessus de $\Sigma = T^*\mathbb{R}^n$ dont les fibres sont isomorphes à \mathbb{C} muni de l'action du groupe $\mathbb{C}^* = \mathbb{C} - \{0\}$ agissant par multiplications. Cette construction permet de définir les fonctions d'état à valeurs complexes $s(q, p)$ que sont les sections de L , sections qui auront ensuite vocation à être interprétées comme des fonctions d'onde.

Remarque. L est globalement trivial pour $\Sigma = T^*\mathbb{R}^n$ mais il ne l'est pas pour une variété symplectique générale. Les classes d'isomorphisme de ces fibrés L forment un groupe (appelé le groupe de Picard de Σ) pour le produit tensoriel $L_1 \otimes_{\mathbb{C}} L_2$. L'élément neutre (l'unité) est le produit direct $\Sigma \times \mathbb{C}$ et l'inverse L^{-1} de L est le dual L^* . En utilisant les rudiments de cohomologie de Čech à valeurs dans des faisceaux que nous avons introduits aux sections 7.2.9 et 9 (surtout la sous-section 9.4) du chapitre 2, on peut montrer qu'il est isomorphe au groupe de cohomologie $\check{H}^2(\Sigma, \mathbb{Z})$ à valeurs dans le faisceau constant \mathbb{Z} .⁸⁷ En effet lorsque l'on recolle des trivialisations locales $U_j \times \mathbb{C}$ de L au-dessus d'un recouvrement ouvert localement fini $\mathcal{U} = \{U_j\}$ de Σ (on suppose tous les U_j connexes), les fonctions de transition forment un 1-cocycle de Σ à valeurs dans le "faisceau" \mathcal{O}_{Σ}^* des germes de fonctions $\mathcal{C}^{\infty}(\Sigma, \mathbb{C})$ partout $\neq 0$. On considère alors les suites

$$0 \rightarrow \mathbb{Z}(U_j) \rightarrow \mathcal{O}_{\Sigma}(U_j) \xrightarrow{\exp} \mathcal{O}_{\Sigma}^*(U_j) \rightarrow 0$$

où $\mathbb{Z}(U_j)$ est l'ensemble des fonctions constantes à valeurs entières, $\mathcal{O}_{\Sigma}(U_j)$ (resp. $\mathcal{O}_{\Sigma}^*(U_j)$) celui des fonctions de $\mathcal{C}^{\infty}(U_j, \mathbb{C})$ (resp. $\mathcal{C}^{\infty}(U_j, \mathbb{C}^*)$) et $\exp(f) = e^{2\pi i f}$. L'existence de ces suites, l'utilisation du nerf du recouvrement et le passage à la limite inductive sur des recouvrements de plus en plus fins permettent de définir la

87. Pour les fibrés en droites réelles, le groupe de cohomologie classifiant est simplement $\check{H}^2(\Sigma, \mathbb{Z}/2\mathbb{Z})$.

suite exacte “courte” de faisceaux

$$0 \rightarrow \mathbb{Z} \rightarrow \mathcal{O}_\Sigma \xrightarrow{\text{exp}} \mathcal{O}_\Sigma^* \rightarrow 0$$

et la suite exacte “longue” de cohomologie

$$\dots \rightarrow \check{H}^p(\Sigma, \mathcal{O}_\Sigma) \rightarrow \check{H}^p(\Sigma, \mathcal{O}_\Sigma^*) \xrightarrow{\partial} \check{H}^{p+1}(\Sigma, \mathbb{Z}) \rightarrow \check{H}^{p+1}(\Sigma, \mathcal{O}_\Sigma) \rightarrow \dots$$

Mais comme $\check{H}^p(\Sigma, \mathcal{O}_\Sigma) = 0$ si $p \geq 1$, on en déduit $\check{H}^1(\Sigma, \mathcal{O}_\Sigma^*) \simeq \check{H}^2(\Sigma, \mathbb{Z})$. \square

L'idée de base est alors d'introduire sur L la *connexion* ∇ (cf. la section 3.7 du chapitre 6) qui corrige la dérivation des sections s de L par les champs de vecteurs ξ sur L ainsi que, dualement, la différentielle d des sections de L , cela au moyen du potentiel symplectique θ (qui était plus haut pdq). Si ξ est un champ de vecteurs tangents sur L et si s est une section de L , ∇ est définie par les formules :

$$\begin{aligned} \nabla_\xi s &= \xi(s) - \frac{i}{\hbar} \theta(\xi) s = ds(\xi) - \frac{i}{\hbar} \theta(\xi) s \\ \nabla_\xi &= \xi - \frac{i}{\hbar} \theta(\xi) \\ \nabla s &= ds - \frac{i}{\hbar} s \theta \\ \nabla &= d - \frac{i}{\hbar} \theta . \end{aligned}$$

On dérive la connexion ∇ de la 1-forme de connexion α à valeurs dans \mathbb{C} sur $L^* = L - \{\text{section nulle}\}$ qui est *invariante* par l'action de \mathbb{C}^* dans les fibres et dont la valeur sur les fibres de L^* (qui par définition sont identifiables à \mathbb{C}^*) est

$$\frac{1}{2\pi i} \frac{dz}{z}$$

(où z est la variable complexe sur les fibres \mathbb{C}^*). La différentielle logarithmique $\frac{dz}{z}$ est bien définie sur L^* où $z \neq 0$ et elle est bien invariante par $z \rightarrow \lambda z, \lambda \in \mathbb{C}^*$.⁸⁸ Sur la base Σ du fibré $\delta : L \rightarrow \Sigma$ on choisit, avec le facteur approprié, la 1-forme θ du potentiel symplectique et l'on obtient la 1-forme

$$\alpha = \frac{1}{i} \frac{dz}{z} - \frac{1}{\hbar} \theta .$$

On notera que si l'on restreint α au sous-fibré en cercles des $z = e^{i\varphi}$, on a $\alpha = d\varphi - \frac{1}{\hbar} \theta$. La dérivée extérieure est $d\alpha = -\frac{1}{\hbar} \Theta$ et par conséquent la *courbure* de α , qui est par définition la 2-forme R sur Σ telle que $\delta^* R = d\alpha$, est tout simplement $-\frac{1}{\hbar} \Theta$. Autrement dit, à des facteurs près, on considère la 1-forme fondamentale qu'est le potentiel symplectique de Σ et on lui ajoute dans les fibres la 1-forme logarithmique $\frac{dz}{z} = d \log(z)$. Aux facteurs près, α est donc la *forme de contact*

88. Nous allons voir comment la connexion est, quant à elle, définie sur tout L .

canonique sur $\mathbb{C}^* \times \Sigma$. Cela signifie que l'on considère z comme un facteur $z = e^\zeta$ ($\frac{dz}{z} = d\zeta$) produit d'un facteur d'échelle $e^{\Re(\zeta)}$ et d'un facteur de phase $e^{\Im(\zeta)}$.

Remarque. On voit que cette construction est analogue à ce qui se passe pour la structure de contact \mathbb{V}_J du point de vue de la dualité projective (chapitre 3, section 2.2) où (x, p) est le plan symplectique et y la fibre et où la 1-forme de contact est $\omega_J = dy - p dx$. Nous avons vu à la section 2 du chapitre 6 comment ces constructions interviennent de façon générale. \square

Comme dans les modèles de jets, la 1-forme de contact α définit une distribution $\ker(\alpha)$ d'hyperplans "horizontaux" de dimension $2n$ sur $\mathbb{C}^* \times \Sigma$ dont une base est donnée (en notation condensée) dans le cas standard $\theta = pdq$ par les $2n$ vecteurs tangents ∂_p et $\partial_q - \frac{1}{i\hbar} pz \partial_z$ de crochet de Lie

$$\left[\partial_p, \partial_q - \frac{1}{i\hbar} pz \partial_z \right] = -\frac{1}{i\hbar} z \partial_z .$$

Ces hyperplans sont les hyperplans "horizontaux" de la connexion sur les sections $s : \Sigma \rightarrow L^*$ qui est définie, pour tout champ de vecteurs $\xi = (\xi_q, \xi_p)$ sur Σ (à valeurs dans le complexifié $T_{\mathbb{C}}\Sigma$ de $T\Sigma$), par la formule

$$\nabla_{\xi} s = i(s^* \alpha)(\xi) s$$

où $s^* \alpha$ est le pull back sur Σ de α par $s : \Sigma \rightarrow L^*$. Ce pull-back est quant à lui défini de la façon suivante. La section $s : \Sigma \rightarrow L^*$ est une application $(q, p) \mapsto (q, p, z = s(q, p))$ qui possède une application linéaire tangente

$$Ds : T_{(q,p)}\Sigma \rightarrow T_{(q,p,s(q,p))}L^*$$

de matrice $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \partial_q s & \partial_p s \end{pmatrix}$ et si $\xi = (\xi_q, \xi_p) \in T_{(q,p)}\Sigma$,

$$Ds(\xi) = (\xi_q, \xi_p, (\partial_q s) \xi_q + (\partial_p s) \xi_p) = (\xi, ds(q, p)(\xi) = \xi(s(q, p))) \in T_{(q,p,s(q,p))}L^*$$

(où dans la dernière expression, $\xi(s(q, p))$, le champ ξ défini sur Σ opère comme dérivation sur la fonction $s(q, p)$ définie sur Σ). Par définition $(s^* \alpha)(\xi) = \alpha(Ds(\xi))$ et donc

$$(s^* \alpha)(\xi) = \frac{1}{i} \frac{ds(\xi)}{s} - \frac{1}{\hbar} \theta(\xi) .$$

On retrouve bien ainsi la connexion $\nabla_{\xi} s$:

$$\begin{aligned}\nabla_{\xi} s &= i((s^* \alpha)(\xi)) s = i \left(\frac{1}{i} \frac{ds(\xi)}{s} - \frac{1}{\hbar} \theta(\xi) \right) s \\ &= \left(\frac{ds(\xi)}{s} - \frac{i}{\hbar} \theta(\xi) \right) s \\ &= ds(\xi) - \frac{i}{\hbar} \theta(\xi) s.\end{aligned}$$

Il est alors immédiat de vérifier que cette connexion possède bien les propriétés caractéristiques décrites à la section 3.7 du chapitre 6 :

- (i) elle est \mathbb{C} -linéaire en ξ et s ,
- (ii) pour $f : \Sigma \rightarrow \mathbb{C}$, $\nabla_{f\xi} s = f \nabla_{\xi} s$ ($\mathcal{C}^{\infty}(\Sigma)$ -linéarité) et
- (iii) $\nabla_{\xi}(fs) = f \nabla_{\xi} s + df(\xi)$ (règle de dérivation de Leibniz).

On peut d'ailleurs utiliser cette dernière propriété pour étendre $\nabla_{\xi} s$ à toutes les sections de L (et pas seulement de L^*). En effet, on a d'abord

$$\begin{aligned}\nabla_{f\xi} s &= i(s^* \alpha)(f\xi) s = \left(\frac{ds(f\xi)}{s} - \frac{i}{\hbar} \theta(f\xi) \right) s \\ &= f \nabla_{\xi} s\end{aligned}$$

car $\frac{ds(f\xi)}{s} = \frac{(f\xi)(s)}{s} = f \frac{\xi(s)}{s}$ et $\theta(f\xi) = f\theta(\xi)$. Et ensuite

$$\begin{aligned}\nabla_{\xi}(fs) &= d(fs)(\xi) - \frac{i}{\hbar} \theta(\xi) fs \\ &= (fds(\xi) + sdf(\xi)) - \frac{i}{\hbar} \theta(\xi) fs \\ &= fds(\xi) - f \left(\frac{i}{\hbar} \theta(\xi) \right) s + \frac{df(\xi)}{f} fs \\ &= f \left(ds(\xi) - \frac{i}{\hbar} \theta(\xi) s \right) + df(\xi) s \\ &= f \nabla_{\xi} s + df(\xi) s.\end{aligned}$$

Si $\gamma(t) = (q(t), p(t))$ est une courbe dans Σ , on peut ainsi la relever horizontalement dans L^* à partir de $(q(0), p(0), z(0))$. On obtient une relevée lagrangienne qui est une courbe horizontale $\Gamma(t) = (q(t), p(t), z(t))$ de vecteur tangent $(\dot{q}(t), \dot{p}(t), \dot{z}(t))$ satisfaisant l'équation différentielle

$$\alpha \left(\frac{d\Gamma(t)}{dt} \right) = 0 = \frac{1}{i} \frac{\dot{z}(t)}{z(t)} - \frac{1}{\hbar} p(t) \dot{q}(t)$$

de solution $z(t) = z(0) \exp \left(\frac{i}{\hbar} \int_{\gamma} p dq \right)$.

Si $\xi = (\xi_q = a\partial_q, \xi_p = b\partial_p)$ est un champ sur Σ on peut également le relever (relevé legendrien) en un champ horizontal ξ^{L^*} sur L^*

$$\xi^{L^*} = -\frac{1}{i\hbar}zap\partial_z + \xi .$$

Les champs ∂_p et $\partial_q - \frac{1}{i\hbar}pz\partial_z$ formant une base des plans “horizontaux” sont donc simplement les relevés legendriens $\partial_p^{L^*}$ et $\partial_q^{L^*}$ de ∂_p et ∂_q . Si $f \in \mathcal{C}^\infty(\Sigma, \mathbb{R})$, le relevé du champ hamiltonien $X_f = (\partial_p f) \partial_q - (\partial_q f) \partial_p$ est

$$X_f^{L^*} = -\frac{1}{i\hbar}zp(\partial_p f) \partial_z + (\partial_p f) \partial_q - (\partial_q f) \partial_p .$$

La donnée de L et ∇ (ou α) permet alors de redéfinir la préquantification par la formule :

$$\begin{aligned} \tilde{f}(s) &= -i\hbar\nabla_{X_f}s + fs = -i\hbar X_f(s) - \theta(X_f)s + fs \\ &= -i\hbar\{f, s\} - \theta(X_f)s + fs . \end{aligned}$$

On retrouve évidemment (D1) $\hat{1} = Id$, ainsi que (D3) $\hat{p} = -i\hbar\partial_q$ dans le cas standard puisque $\partial_p p = 1$, $-p(\partial_p p)s + ps = 0$ et $\{p, s\} = \partial_q s$. Pour vérifier la condition (D2) $[\tilde{f}, \tilde{g}] = -i\hbar\{f, g\}$ nous allons refaire le calcul effectué plus haut de façon à faire apparaître la courbure $R_\nabla(X, Y)$ de la connexion ∇ . D’après la formule de la section 3.7 du chapitre 6 sur la géométrie des connexions,

$$R_\nabla(\xi, \eta) = [\nabla_\xi, \nabla_\eta] - \nabla_{[\xi, \eta]} .$$

Le calcul donne

$$\begin{aligned} [\nabla_\xi, \nabla_\eta]s &= \left(\xi - \frac{i}{\hbar}\theta(\xi)\right) \left(\eta - \frac{i}{\hbar}\theta(\eta)\right) s - \left(\eta - \frac{i}{\hbar}\theta(\eta)\right) \left(\xi - \frac{i}{\hbar}\theta(\xi)\right) s \\ &= [\xi, \eta]s - \frac{i}{\hbar}(\theta(\xi)\eta(s) - \theta(\eta)\xi(s)) - \frac{i}{\hbar}(\xi(\theta(\eta)s) - \eta(\theta(\xi)s)) \\ &= [\xi, \eta]s - \frac{i}{\hbar}(\theta(\xi)\eta(s) - \theta(\eta)\xi(s)) \\ &\quad - \frac{i}{\hbar}(\theta(\eta)\xi(s) + s\xi(\theta(\eta)) - \theta(\xi)\eta(s) - s\eta(\theta(\xi))) \\ &= [\xi, \eta]s - \frac{i}{\hbar}s(\xi(\theta(\eta)) - \eta(\theta(\xi))) \\ &= [\xi, \eta]s - \frac{i}{\hbar}s(\theta[\xi, \eta] + \Theta(\xi, \eta)) \\ &= \nabla_{[\xi, \eta]} - \frac{i}{\hbar}\Theta(\xi, \eta) \end{aligned}$$

et donc

$$R_{\nabla}(\xi, \eta) = [\nabla_{\xi}, \nabla_{\eta}] - \nabla_{[\xi, \eta]} = -\frac{i}{\hbar} \Theta(\xi, \eta) .$$

La courbure de la connexion est ainsi, à un facteur près, la forme symplectique Θ de Σ elle-même. C'est la condition maîtresse de "l'harmonie" des calculs.⁸⁹

On vérifie alors (D2) "harmonieusement" en utilisant

$$\nabla_{X_f} g = g \nabla_{X_f} + dg(X_f) = g \nabla_{X_f} + X_f(g) .$$

Cela donne :

$$\begin{aligned} [\tilde{f}, \tilde{g}] &= -i\hbar \nabla_{X_f} (-i\hbar \nabla_{X_g} + g) + f (-i\hbar \nabla_{X_g} + g) + i\hbar \nabla_{X_g} (-i\hbar \nabla_{X_f} + f) \\ &\quad - g (-i\hbar \nabla_{X_f} + f) \\ &= -\hbar^2 (\nabla_{X_f} \nabla_{X_g} - \nabla_{X_g} \nabla_{X_f}) - i\hbar (\nabla_{X_f} g - \nabla_{X_g} f) - i\hbar (f \nabla_{X_g} - g \nabla_{X_f}) \\ &\quad + (fg - gf) \\ &= -\hbar^2 (\nabla_{X_f} \nabla_{X_g} - \nabla_{X_g} \nabla_{X_f}) - i\hbar (g \nabla_{X_f} + X_f(g) - (f \nabla_{X_g} + X_g(f))) \\ &\quad - i\hbar (f \nabla_{X_g} - g \nabla_{X_f}) \\ &= -\hbar^2 (\nabla_{X_f} \nabla_{X_g} - \nabla_{X_g} \nabla_{X_f}) - i\hbar (X_f(g) - X_g(f)) \\ &= -\hbar^2 (\nabla_{X_f} \nabla_{X_g} - \nabla_{X_g} \nabla_{X_f}) - 2i\hbar \{f, g\} \\ &= -\hbar^2 [\nabla_{X_f}, \nabla_{X_g}] - 2i\hbar \{f, g\} \\ &= -\hbar^2 \left(\nabla_{[X_f, X_g]} - \frac{i}{\hbar} \Theta(X_f, X_g) \right) - 2i\hbar \{f, g\} \\ &= -\hbar^2 \nabla_{[X_f, X_g]} + i\hbar \Theta(X_f, X_g) - 2i\hbar \{f, g\} \\ &= -\hbar^2 \nabla_{[X_f, X_g]} - i\hbar \{f, g\} \\ &= -\hbar^2 \nabla_{X_{\{f, g\}}} - i\hbar \{f, g\} \\ &= -i\hbar \widetilde{\{f, g\}} . \end{aligned}$$

Les opérateurs \tilde{f} agissent de façon auto-adjointe sur les sections $s \in C^\infty$ à support compact de L munies du produit hermitien $\langle s_1, s_2 \rangle = \int_{\Sigma} s_1 \bar{s}_2 d\lambda$ où $d\lambda$ est la $2n$ -forme volume de Liouville $\frac{1}{n!} \Theta^{\wedge n}$. On peut étendre par complétion cette action au Hilbert $L^2(L)$ des sections s de carré intégrable.

89. Certains auteurs (par exemple [57]) définissent la courbure de ∇ par $\Omega(X, Y) = iR_{\nabla}(X, Y)$ et trouvent donc $\Omega(X, Y) = \frac{1}{\hbar} \omega(X, Y)$.

8.3. Polarisation

La préquantification qui vient d'être définie agit sur des sections $s(q, p)$ de L qui dépendent des $2n$ variables (q, p) . Mais cela pose évidemment un gros problème car ces variables conjuguées sont commutatives alors que les opérateurs associés par la quantification ne devraient pas l'être, leur non-commutativité s'exprimant par les relations d'incertitude de Heisenberg. Comme des fonctions $s(q, p)$ peuvent être de support aussi petit que l'on veut, les \tilde{q} , \tilde{p} de la préquantification peuvent violer les relations d'incertitude.

Il s'agit en quelque sorte d'une "revanche" de la mécanique lagrangienne sur la mécanique hamiltonienne. On passe à la mécanique hamiltonienne en traitant les moments p comme de nouvelles variables indépendantes et en introduisant la structure symplectique forçant leur interprétation comme dérivées.⁹⁰ Si l'on veut quantifier cette approche hamiltonienne il faut introduire les nouvelles variables p sans supposer qu'elles commutent avec les variables q .

D'où l'idée de considérer comme espaces de Hilbert des espaces L^2 non plus de sections générales de la base $\Sigma = T^*\mathbb{R}^n$ mais des espaces L^2 associées à des ensembles maximaux d'observables quantiques *commutantes*, c'est-à-dire, côté classique, à des ensembles complets de n fonctions f_j indépendantes *en involution*, c'est-à-dire de crochet de Poisson $\{f_j, f_{k \neq j}\} = 0$. De telles fonctions définissent des *distributions lagrangiennes* sur Σ . Cela conduit à la seconde idée force de la quantification géométrique qui est celle de *polarisation*. Elle va permettre de retrouver des opérateurs auto-adjoints opérant sur des fonctions d'état $\psi(q)$ ou $\phi(p)$ dépendant de seulement n variables.

8.3.1. Polarisation verticale de $T^*\mathbb{R}^n$.

Revenons au fait que la variété symplectique $\Sigma = T^*\mathbb{R}^n$ est naturellement fibrée par la submersion $\pi : T^*\mathbb{R}^n \rightarrow \mathbb{R}^n_{(q)}$, $(q, p) \mapsto q$,⁹¹ submersion dont les fibres $T^*\mathbb{R}^n = \mathbb{R}^n_{(p)}$ sont des sous-variétés lagrangiennes.⁹² La fibration en droites complexes $\delta : L \rightarrow T^*\mathbb{R}^n$ est donc *doublement* fibrée,

$$L \xrightarrow{\delta} T^*\mathbb{R}^n \xrightarrow{\pi} \mathbb{R}^n_{(q)} .$$

90. C'est le principe "philosophique" que nous suivons depuis le début de cet ouvrage à partir de la structure de contact de l'espace des 1-jets des courbes du plan.

91. Rappelons que nous notons quand besoin est le \mathbb{R}^n de coordonnées x par $\mathbb{R}^n_{(x)}$.

92. Rappelons, cf. chapitre 15 section 2.2.3 et chapitre 12, section 6, qu'une sous variété de dimension moitié de (Σ, Θ) est dite "lagrangienne" si Θ s'y annule identiquement. C'est évidemment le cas ici pour les fibres puisque les vecteurs tangents aux fibres sont de la forme $X = (0, X_p)$ et que donc $\Theta(X, Y) = X_p Y_q - X_q Y_p = X_p \cdot 0 - 0 \cdot Y_p = 0$. Localement, toute sous-variété lagrangienne est le lieu d'annulation de n observables f_j indépendantes en involution. Rappelons aussi qu'on appelle *isotrope* une sous-variété sur laquelle Θ s'annule identiquement. Une sous-variété lagrangienne est donc une sous-variété isotrope de dimension maximale n .

Si alors p_1 et p_2 sont deux points de la même fibre $T_q^*\mathbb{R}^n$, on peut regarder leurs fibres $\delta^{-1}(p_1)$ et $\delta^{-1}(p_2)$ pour la fibration δ et les *identifier* par transport parallèle parce que la courbure $R_\nabla = -\frac{i}{\hbar}\Theta$ de la connexion ∇ y est nulle et qu'elles sont simplement connexes. Si $\gamma(t) = (q = \text{cst}, p(t))$, $t \in [0, 1]$, est une courbe de p_1 à p_2 dans $T_q^*\mathbb{R}^n$ à q constant, sa relevée legendrienne à partir de $z(0)$ au-dessus de (q, p_1) est

$$z(t) = z(0) \exp\left(\frac{i}{\hbar} \int_\gamma pdq\right) = z(0)$$

puisque $dq = 0$ le long de $\gamma(t)$. L'identification des fibres $\delta^{-1}(p_1)$ et $\delta^{-1}(p_2)$ est donc triviale puisque $z(1)$ au-dessus de (q, p_1) reste égal à $z(0)$. On obtient ainsi un fibré

$$\delta^\pi = \pi \circ \delta : L^\pi \rightarrow \mathbb{R}_{(q)}^n$$

dont les fibres $(\delta^\pi)^{-1}(q)$ sont les fibres $(\delta)^{-1}(q, p \in T_q^*\mathbb{R}^n)$ toutes identifiées entre elles et dont les sections s^π sont les sections de $\delta : L \rightarrow T^*\mathbb{R}^n$ qui sont ∇ -*constantes* le long des fibres $\mathbb{R}_{(p)}^{*n}$ de $\pi : T^*\mathbb{R}^n \rightarrow \mathbb{R}_{(q)}^n$, c'est-à-dire telles que $\nabla_\xi s = 0$ pour tout champ $\xi = a\partial_p$ tangent aux fibres $T_q^*\mathbb{R}^n \simeq \mathbb{R}_{(p)}^{*n}$ de π . Comme

$$\nabla_\xi s = ds(\xi) - \frac{i}{\hbar} pdq(\xi) s,$$

cette condition signifie $a\partial_p s = 0$ pour tout a et donc $\partial_p s = 0$, i.e.

$$s^\pi(q) = s(q, p(q)) = s(q).$$

On obtient ainsi la polarisation verticale $\mathfrak{P}_{\text{vert}}$ de δ par rapport à π .

On peut alors réduire le formalisme de la préquantification à la fibration δ^π en considérant des fonctions f telles que X_f laisse la fibration invariante. Pour cela, il faut et il suffit que la projection $(\partial_p f) \partial_q$ de $X_f = (\partial_p f) \partial_q - (\partial_q f) \partial_p$ sur la base $\mathbb{R}_{(q)}^n$ soit indépendante de p , autrement dit que les $\partial_{p_j} f$ soient indépendantes des p_k . On doit donc se restreindre à des f *linéaires* en p de la forme

$$f(q, p) = r(q) + t(q)p.$$

Pour de telles f ,

$$X_f = t(q) \partial_q - (\partial_q r(q) + p \partial_q t(q)) \partial_p$$

et la préquantification \tilde{f} devient une vraie quantification \hat{f} agissant sur les $s^\pi(q)$ par

$$\begin{aligned} \hat{f}(s^\pi) &= -i\hbar \{f, s^\pi\} - p(\partial_p f) s^\pi + f s^\pi \\ &= -i\hbar \{f, s^\pi\} - pt(q) s^\pi + r(q) s^\pi + t(q) ps^\pi \\ &= -i\hbar \{f, s^\pi\} + r(q) s^\pi \\ &= -i\hbar t(q) \partial_q s^\pi + r(q) s^\pi \end{aligned}$$

puisque $\{f, s^\pi\} = t(q) \partial_q s^\pi$. On obtient bien

- (i) $\widehat{1} = Id$ puisque pour la fonction 1 $r(q) = 1$ et $t(q) = 0$,
- (ii) $\widehat{q} = q$ puisqu'alors $r(q) = q$ et $t(q) = 0$, et
- (iii) $\widehat{p} = -i\hbar \partial_q$ puisque $r(q) = 0$ et $t(q) = 1$.

On peut étendre par complétion la construction précédente des $s^\pi(q)$ à support compact aux $s^\pi(q) \in L^2\left(\mathbb{R}_{(q)}^n\right)$ avec le produit scalaire

$$\langle s_1^\pi, s_2^\pi \rangle_{vert} = \int_{\mathbb{R}_{(q)}^n} s_1^\pi \overline{s_2^\pi} dq .$$

La quantification au moyen de la polarisation verticale correspond donc à la *représentation de Schrödinger*.

8.3.2. Polarisation horizontale de $T^*\mathbb{R}^n$.

La variété symplectique $\Sigma = T^*\mathbb{R}^n = \mathbb{R}_{(q)}^n \times \mathbb{R}_{(p)}^{*n}$ est également naturellement fibrée par la submersion duale $\rho : T^*\mathbb{R}^n \rightarrow \mathbb{R}_{(p)}^{*n}$, $(q, p) \mapsto p$ dont les fibres, notées $\mathbb{R}_{(q \downarrow p)}^n$, sont des sous-variétés lagrangiennes.⁹³ La fibration en droites complexes $\delta : L \rightarrow T^*\mathbb{R}^n$ est donc elle aussi *doublement* fibrée,

$$L \xrightarrow{\delta} T^*\mathbb{R}^n \xrightarrow{\rho} \mathbb{R}_{(p)}^{*n} .$$

Si alors q_1 et q_2 sont deux points de la même fibre $\mathbb{R}_{(q \downarrow p)}^n$, on peut considérer leurs fibres $\delta^{-1}(q_1)$ et $\delta^{-1}(q_2)$ pour la fibration δ et, comme précédemment, les *identifier* par transport parallèle parce que la courbure $R_\nabla = -\frac{i}{\hbar} \Theta$ de la connexion ∇ y est nulle et qu'elles sont simplement connexes. Si $\gamma(t) = (q(t), p = \text{cst})$, $t \in [0, 1]$, est une courbe de q_1 à q_2 dans $\mathbb{R}_{(q \downarrow p)}^n$ à p constant, sa relevée legendrienne à partir de $z(0)$ au-dessus de (q_1, p) est

$$z(t) = z(0) \exp\left(\frac{i}{\hbar} \int_\gamma pdq\right) = z(0) \exp\left(\frac{i}{\hbar} p(q(t) - q_1)\right) .$$

L'identification des fibres $\delta^{-1}(q_1)$ et $\delta^{-1}(q_2)$ est donc non triviale et $z(0)$ au-dessus de (q_1, p) est identifié à

$$z(1) = z(0) \exp\left(\frac{i}{\hbar} p(q_2 - q_1)\right)$$

au-dessus de (q, p_2) .

On obtient ainsi un fibré

$$\delta^\rho = \rho \circ \delta : L^\rho \rightarrow \mathbb{R}_{(p)}^{*n}$$

93. Les vecteurs tangents aux fibres $\mathbb{R}_{(q \downarrow p)}^n$ sont de la forme $X = (X_q, 0)$ et donc $\Theta(X, Y) = X_p Y_q - X_q Y_p = 0 \cdot Y_q - X_q \cdot 0 = 0$.

dont les fibres $(\delta^\rho)^{-1}(p)$ sont les fibres $(\delta)^{-1}(q \in \mathbb{R}_{(q \downarrow p)}^n, p)$ toutes identifiées entre elles et dont les sections s^ρ sont les sections de $\delta : L \rightarrow T^*\mathbb{R}^n$ qui sont ∇ -*constantes* le long des fibres de $\rho : T^*\mathbb{R}^n \rightarrow \mathbb{R}_{(p)}^n$, c'est-à-dire telles que $\nabla_\xi s = 0$ pour tout champ $\xi = b\partial_q$ tangent aux fibres $\mathbb{R}_{(q \downarrow p)}^n$ de ρ . Comme

$$\nabla_\xi s = ds(\xi) - \frac{i}{\hbar} pdq(\xi) s ,$$

cette condition signifie $b\partial_q s - \frac{i}{\hbar} pbs = 0$ pour tout b et donc $\partial_q s - \frac{i}{\hbar} ps = 0$, i.e. $\frac{\partial_q s}{s} = \frac{i}{\hbar} p$, équation de solution

$$s^\rho(q, p) = e^{\frac{i}{\hbar} p \cdot q} \sigma^\rho(p)$$

où q n'intervient que dans un *facteur de phase*. Mais en fait ce facteur est celui de l'identification des fibres et ces sections satisfont bien la contrainte d'identification pour p fixé :

$$\begin{aligned} s^\rho(q_2, p) &= e^{\frac{i}{\hbar} p \cdot q_2} \sigma^\rho(p) = e^{\frac{i}{\hbar} p(q_2 - q_1)} e^{\frac{i}{\hbar} p \cdot q_1} \sigma^\rho(p) \\ &= e^{\frac{i}{\hbar} p(q_2 - q_1)} s^\rho(q_1, p) . \end{aligned}$$

C'est donc $\sigma^\rho(p)$ qui est la section génératrice.

On obtient ainsi une seconde polarisation de δ , la polarisation horizontale \mathfrak{P}_{hor} , cette fois par rapport à ρ , avec le produit scalaire

$$\langle s_1^\rho, s_2^\rho \rangle_{hor} = \int_{\mathbb{R}_{(p)}^n} \sigma_1^\rho \overline{\sigma_2^\rho} dp .$$

On peut dans ce nouveau contexte reprendre le passage de la préquantification à la quantification en tenant compte des spécificités :

- faut-il travailler avec les $s^\rho(q, p) = e^{\frac{i}{\hbar} p \cdot q} \sigma^\rho(p)$ ou seulement avec les $\sigma^\rho(p)$?
- faut-il garder le potentiel symplectique pdq ou utiliser le dual $-qdp$ qui est mieux adapté à la polarisation horizontale ?

Les deux polarisations sont complètement transverses l'une à l'autre et, comme on pouvait s'y attendre, leur dualité correspond à la *transformée de Fourier* au sens suivant. On a d'un côté des $s^\pi(q, p) = s^\pi(q)$ dont le produit scalaire est

$$\langle s_1^\pi, s_2^\pi \rangle = \int_{\mathbb{R}_q^n} s_1 \overline{s_2} dq$$

et de l'autre côté des $s^\rho(q, p) = e^{\frac{i}{\hbar} p \cdot q} \sigma^\rho(p)$ dont le produit scalaire est

$$\langle s_1^\rho, s_2^\rho \rangle = \int_{\mathbb{R}_p^n} \sigma_1^\rho \overline{\sigma_2^\rho} dp .$$

Il existe un opérateur unitaire U transformant les s^π en s^ρ en conservant les produits scalaires. En effet, à un facteur près,

$$\int_{\mathbb{R}^n} e^{\frac{i}{\hbar}p \cdot q} \sigma^\rho(p) dp = \widehat{\sigma}^\rho(q)$$

et réciproquement,⁹⁴ on peut associer à $s^\pi(q)$ la section $s^\rho(q, p) = e^{\frac{i}{\hbar}p \cdot q} \widehat{s}^\pi(p)$.

8.3.3. Polarisation kähleriennes.

Les polarisations verticales \mathfrak{P}_{vert} et horizontales \mathfrak{P}_{hor} peuvent se généraliser aux fibrés cotangents $\Sigma = T^*M$ sous certaines conditions cohomologiques liées à la topologie de M puis à des variétés symplectiques plus générales. Mais pour certaines quantifications il est nécessaire d'avoir recours à des polarisations définies à partir du *complexifié* $T^{\mathbb{C}}\Sigma$ du fibré tangent à Σ . Cela est le cas lorsque l'on a trois structures qui interfèrent de façon compatible sur Σ :

- (i) une structure symplectique $\Theta(X, Y)$,
- (ii) une structure complexe (produit scalaire hermitien) $\langle X, Y \rangle$ et
- (iii) une métrique $g(X, Y)$.

Nous allons en dire un mot car cela recoupe plusieurs éléments de structure que nous avons déjà rencontrés.

Dès le début de notre présentation de la structure du groupe de Heisenberg au chapitre 3, section 3.4, nous avons expliqué les liens très étroits entre les structures symplectiques, métriques et complexes sur une variété symplectique (Σ, Θ) de dimension $2n$. En prenant une base symplectique comme orthonormée on définit une métrique riemannienne g . On a alors

$$\Theta(X, Y) = g(X, J(Y))$$

pour un endomorphisme (unique) J des espaces tangents satisfaisant $J^2 = -I$. Le plan $\text{Span}\{X, Y\}$ est donc "isotrope" (i.e. $\Theta(X, Y) = 0$) si et seulement si $J(X)$ et Y sont g -orthogonaux.

La compatibilité de J avec Θ et g s'exprime par l'invariance

$$\begin{aligned} \Theta(J(X), J(Y)) &= \Theta(X, Y) \text{ ,} \\ g(J(X), J(Y)) &= g(X, Y) \text{ .} \end{aligned}$$

Elle confirme que $g(X, Y)$ est bien symétrique puisque, avec $Z = J(Y)$,

94. Ne pas confondre les notations. $\widehat{\sigma}^\rho(q)$ et $\widehat{s}^\pi(p)$ sont évidemment ici des transformées de Fourier et non pas des opérateurs.

$$\begin{aligned} g(X, Z) &= \Theta(X, -J(Z)) = \Theta(X, Y) \text{ ,} \\ g(Z, X) &= \Theta(Z, -J(X)) = \Theta(J(Y), -J(X)) = \Theta(J(X), J(Y)) \\ &= \Theta(X, Y) \text{ .} \end{aligned}$$

Si la métrique g est définie positive, on dit que la structure est “kählerienne”. Les espaces tangents $T_{q,p}\Sigma$ à Σ étant de dimension paire $2n$, ils sont isomorphes à \mathbb{C}^n et on peut choisir l’isomorphisme de façon à ce que J soit la multiplication par i (la rotation de $\frac{\pi}{2}$). On peut alors définir un produit scalaire hermitien par

$$\langle X, Y \rangle = g(X, Y) + i\Theta(X, Y)$$

dont la partie réelle (symétrique) donne la métrique et la partie imaginaire (anti-symétrique) donne la 2-forme symplectique. On a $\langle J(X), Y \rangle = i\langle X, Y \rangle$ car

$$\begin{aligned} \langle J(X), Y \rangle &= g(J(X), Y) + i\Theta(J(X), Y) = g(Y, J(X)) + i\Theta(J(X), Y) \\ &= \Theta(Y, X) + ig(J(X), J(Y)) = ig(X, Y) - \Theta(X, Y) \\ &= i\langle X, Y \rangle \text{ .} \end{aligned}$$

On peut ainsi considérer les espaces tangents $T_{x=(q,p)}\Sigma$ comme des \mathbb{C} -vectoriels, $\alpha = a + ib$ agissant sur $X \in T_{x=(q,p)}\Sigma$ par $\alpha X = aX + bJ(X)$.

Dans le modèle standard $\Sigma = T^*\mathbb{R}^n$, J est donné par

$$\begin{cases} J(\partial_{q_k}) = \partial_{p_k} \\ J(\partial_{p_k}) = -\partial_{q_k} \end{cases}$$

puisque la métrique est définie par la base orthonormée des ∂_{q_k} et ∂_{p_k} . Par exemple,

$$g(\partial_{q_k}, \partial_{q_k}) = -\Theta(\partial_{q_k}, J(\partial_{q_k})) = -\Theta(\partial_{q_k}, \partial_{p_k}) = \Theta(\partial_{p_k}, \partial_{q_k}) = 1 \text{ .}$$

On a ainsi trois groupes de structures géométriques qui interviennent : le groupe orthogonal $O(2n)$ des isométries de \mathbb{R}^{2n} , le groupe $GL(n, \mathbb{C})$ des automorphismes \mathbb{C} -linéaires de \mathbb{C}^n , le groupe $Sp(2n)$ des automorphisme symplectiques de \mathbb{R}^{2n} . Leur intersection est le groupe unitaire $U(n)$ de \mathbb{C}^n . C’est cette triple structure symplectique – métrique – (presque)complexe qui permet alors de définir des polarisations kähleriennes.

Comme nous l’avons vu également à la section 3.4 du chapitre 3, on peut introduire des vecteurs tangents à Σ complexes $X + iY$ (c’est-à-dire des vecteurs appartenant au complexifié $T^{\mathbb{C}}\Sigma = T\Sigma \otimes_{\mathbb{R}} \mathbb{C}$ de $T\Sigma$)⁹⁵ et utiliser alors J pour définir une structure de *Cauchy-Riemann*.

Pour ce faire, on considère, répétons-le, le fibré en droites complexes de fibres

$$\mathcal{K}_{\mathbb{C},x} = \{X - iJ(X) | X \in T_x\Sigma\}$$

95. Σ est de dimension $2n$, $T\Sigma$ de dimension $4n$ et $T^{\mathbb{C}}\Sigma$ de dimension $8n$.

au-dessus de Σ et on traite $J(X)$ non plus comme la rotation de $\frac{\pi}{2}$ de X dans $T_x\Sigma$ mais comme une composante dans $iT_x\Sigma$. On a d'un côté

$$\mathcal{K}_{\mathbb{C},x} \cap \overline{\mathcal{K}_{\mathbb{C},x}} = \{0\}$$

puisque $X - iJ(X) = Y + iJ(Y)$ implique à la fois $Y = X$ et $Y = -X$ et, d'un autre côté,

$$\mathcal{K}_{\mathbb{C},x} \oplus \overline{\mathcal{K}_{\mathbb{C},x}} = T_x^{\mathbb{C}}\Sigma$$

puisque pour tout $Z \in T_x\Sigma$ ($a + ib$) Z s'écrit de façon unique sous la forme

$$(a + ib)Z = X - iJ(X) + Y + iJ(Y)$$

avec $X + Y = aZ$ et $-J(X) + J(Y) = bZ$. Le fibré $\mathcal{K}_{\mathbb{C}}$ est en fait le fibré des espaces propres de J (étendu à $T^{\mathbb{C}}\Sigma$ par linéarité) pour la valeur propre i . En effet

$$J(X - iJ(X)) = J(X) - iJ^2(X) = J(X) + iX = i(X - iJ(X)) .$$

Comme nous l'avons vu, toujours à la section 3.4 du chapitre 3, $\mathcal{K}_{\mathbb{C}}$ est *intégrable* au sens où $[\mathcal{K}_{\mathbb{C}}, \mathcal{K}_{\mathbb{C}}] \subset \mathcal{K}_{\mathbb{C}}$. C'est en fait un sous-fibré *lagrangien* de $T^{\mathbb{C}}\Sigma$ puisque

$$\begin{aligned} \Theta(X - iJ(X), Y - iJ(Y)) &= \Theta(X, Y) - i\Theta(X, J(Y)) - i\Theta(J(X), Y) \\ &\quad - \Theta(J(X), J(Y)) \\ &= \Theta(X, Y) - ig(X, -Y) - ig(J(X), J(Y)) \\ &\quad - \Theta(X, Y) \\ &= \Theta(X, Y) + ig(X, Y) - ig(X, Y) - \Theta(X, Y) \\ &= 0 . \end{aligned}$$

Pour construire une quantification à partir de ces données on introduit alors, comme dans la section 5.5 du chapitre 9 à propos des conditions de Cauchy-Riemann caractérisant les fonctions holomorphes $f(z)$, des coordonnées locales complexes

$$z_k = \frac{1}{\sqrt{2}}(p_k + iq_k), \quad \bar{z}_k = \frac{1}{\sqrt{2}}(p_k - iq_k)$$

redonnant

$$p_k = \frac{1}{\sqrt{2}}(z_k + \bar{z}_k), \quad q_k = -\frac{i}{\sqrt{2}}(z_k - \bar{z}_k)$$

en sens inverse.⁹⁶ Dans le cas standard $\Sigma = T^*\mathbb{R}^n$ où ces coordonnées sont globales, on a (en notation condensée)

$$\begin{aligned} dz \wedge d\bar{z} &= \frac{1}{\sqrt{2}} d(p_k + iq_k) \wedge \frac{1}{\sqrt{2}} d(p_k - iq_k) \\ &= \frac{-i}{2} (dp \wedge dq) + \frac{i}{2} (dq \wedge dp) \\ &= -i\Theta, \end{aligned}$$

soit

$$\Theta = i(dz \wedge d\bar{z})$$

et donc un potentiel symplectique adapté à ces coordonnées est $\theta_K = -i\bar{z}dz$ puisque

$$d\theta_K = -id\bar{z} \wedge dz = i(dz \wedge d\bar{z}) = \Theta.$$

Les fonctions holomorphes $f(q, p) = u(q, p) + iv(q, p) = f(z)$ sont caractérisées par $\partial_{\bar{z}}f = 0$, autrement dit par les conditions de Cauchy-Riemann $\partial_p u = \partial_q v$, $\partial_q u = -\partial_p v$. On considère alors la polarisation \mathfrak{P}_{hol} engendrée par les $X_{\bar{z}}$. Comme

$$\begin{aligned} \partial_{\bar{z}}s &= (\partial_q s)(\partial_{\bar{z}}q) + (\partial_p s)(\partial_{\bar{z}}p) = \frac{1}{\sqrt{2}}(i\partial_q + \partial_p)s \\ X_{\bar{z}} &= (\partial_p \bar{z})\partial_q - (\partial_q \bar{z})\partial_p = \frac{1}{\sqrt{2}}(\partial_q + i\partial_p) = i\partial_{\bar{z}}, \end{aligned}$$

se restreindre aux sections du fibré en droite ∇ -constantes de long de \mathfrak{P}_{hol} consiste à se restreindre aux sections $s(z)$ telles que $\partial_{\bar{z}}s = 0$, i.e. aux sections holomorphes. On étend par complétion au Hilbert L^2 des fonctions holomorphes de carré intégrable pour le produit scalaire

$$\langle f, g \rangle = \int_{\mathbb{C}^n} f(z) \overline{g(z)} e^{-\frac{|z|^2}{\hbar}} d^n z$$

défini par la mesure gaussienne $e^{-\frac{|z|^2}{\hbar}} d^n z$.

8.3.4. Noyaux BKS.

En résumé, partant de la préquantification $\tilde{f} = -i\hbar\nabla_{X_f} + f$, il faut chercher quels opérateurs \tilde{f} peuvent opérer naturellement sur les sections polarisées d'une polarisation \mathfrak{P} et être acceptés comme des \hat{f} . Pour cela il faut que $\nabla_{\xi}s = 0$ pour

96. Là encore il y a plusieurs variantes suivant les auteurs. On peut échanger z_k et \bar{z}_k , prendre les $q_k + ip_k$ plutôt que les $p_k + iq_k$, et enfin prendre des facteurs asymétriques 1 et $\frac{1}{2}$ plutôt que des facteurs symétriques $\frac{1}{\sqrt{2}}$ et $\frac{1}{\sqrt{2}}$.

tout $\xi \in \mathfrak{P}$ implique $\nabla_\xi \widehat{f}(s) = 0$ pour tout $\xi \in \mathfrak{P}$. Comme

$$\begin{aligned} \nabla_\xi \widehat{f}(s) &= -i\hbar \nabla_\xi (\nabla_{X_f} s) + \nabla_\xi (f s) = -i\hbar \nabla_\xi \nabla_{X_f} s + f \nabla_\xi s + df(\xi) s \\ &= -i\hbar [\nabla_\xi, \nabla_{X_f}] s - i\hbar \nabla_{X_f} \nabla_\xi s + f \nabla_\xi s + df(\xi) s \end{aligned}$$

et que, par définition de la courbure de ∇ , on a

$$[\nabla_\xi, \nabla_{X_f}] - \nabla_{[\xi, X_f]} = -\frac{i}{\hbar} \Theta(\xi, X_f) = -\frac{i}{\hbar} df(\xi) ,$$

on voit que

$$\begin{aligned} \nabla_\xi \widetilde{f}(s) &= -i\hbar \left(\nabla_{[\xi, X_f]} - \frac{i}{\hbar} df(\xi) \right) s - i\hbar \nabla_{X_f} \nabla_\xi s + f \nabla_\xi s + df(\xi) s \\ &= -i\hbar \nabla_{[\xi, X_f]} s - i\hbar \nabla_{X_f} \nabla_\xi s + f \nabla_\xi s \end{aligned}$$

et que, par conséquent, si $\nabla_\xi s = 0$,

$$\nabla_\xi \widetilde{f}(s) = -i\hbar \nabla_{[\xi, X_f]} s .$$

La condition sur f qu'est $\nabla_\xi \widetilde{f}(s) = 0$ pour tout $\xi \in \mathfrak{P}$, équivaut donc à $[\xi, X_f] \in \mathfrak{P}$ pour tout $\xi \in \mathfrak{P}$, autrement dit à la condition que le flot hamiltonien de f laisse le feuilletage lagrangien \mathfrak{P} *invariant*.

Pour \mathfrak{P}_{vert} les $\xi \in \mathfrak{P}$ sont les ∂_{p_k} et la condition $[\partial_{p_k}, X_f] \in \mathfrak{P}$ pour $k = 1, \dots, n$ équivaut, nous l'avons vu, à $f(q, p) = r(q) + t(q)p$. Mais cette restriction à des observables f linéaires en p n'est pas satisfaisante car beaucoup d'observables importantes ne sont pas linéaires en p , à commencer par les hamiltoniens classiques qui comportent en général un terme d'énergie cinétique $\frac{p^2}{2m}$ quantifiés par des laplaciens et non pas par des opérateurs du premier ordre. Ces hamiltoniens ne laissent donc pas du tout la polarisation \mathfrak{P}_{vert} invariante! Pour pallier cette difficulté, il faut notablement complexifier la construction et accepter des f dont le flot hamiltonien Φ_t fait évoluer la polarisation \mathfrak{P} en une famille \mathfrak{P}_t . On cherchera alors des sections paramétrées s_t qui seront polarisées par rapport à \mathfrak{P}_t et appartiendront à une famille de Hilbert \mathcal{H}_t . La mise au point de cette idée due à Robert Blattner, Bertram Kostant et Shlomo Sternberg a conduit à la méthode dite des *noyaux BKS*. Elle se heurte à de nombreuses difficultés techniques qui dépassent le cadre de cet ouvrage introductif.

8.3.5. Stone-von Neumann : de Schrödinger à Fourier et Bargmann.

Les trois polarisations \mathfrak{P}_{vert} , \mathfrak{P}_{hor} , \mathfrak{P}_{hol} conduisent à trois représentations irréductibles différentes du groupe de Heisenberg dans trois Hilbert différents \mathcal{H}_{vert} , \mathcal{H}_{hor} , \mathcal{H}_{hol} . D'après le théorème de Stone-von Neumann (que nous allons bientôt présenter dans la section 5 du chapitre 17) elles sont unitairement équivalentes. La transformation de la représentation de Schrödinger "position" (\mathcal{H}_{vert}) en la

représentation “moment” \mathcal{H}_{hor} est la transformée de Fourier. Le passage à \mathcal{H}_{hol} correspond quant à lui à la *transformée de Segal-Bargmann* (appelée aussi transformée de Fock en mécanique quantique) sur laquelle nous reviendrons à la section 5.10.1 du chapitre 17 (cf. Bargmann [29], [30], Folland [187], Hall [241]). La généralisation de la transformée de Bargmann au groupe euclidien $SE(2)$ a été développée par Davide Barbieri, Giovanna Citti, Gonzalo Sanguinetti et Alessandro Sarti (cf. [27], [28]) et appliquée à la neurogéométrie et à l’analyse d’images (cf. chapitre 17, section 8.3).

La transformée de Bargmann de $f(q) \in L^2(\mathbb{R})$ est (en dimension 1, à un facteur de normalisation près et avec $\hbar = 1$)

$$(Bf)(z) = \int_{\mathbb{R}} f(q)e^{\sqrt{2}qz - \frac{1}{2}q^2 - \frac{1}{2}z^2} dq .$$

Elle est une isométrie de $L^2(\mathbb{R})$ sur un sous-espace fermé \mathcal{B} – dit espace de Segal-Bargmann ou espace de Fock – de $L^2(\mathbb{C}, e^{-|z|^2} dz)$ muni de la mesure gaussienne $e^{-|z|^2} dz$.⁹⁷ Un théorème fondamental de Segal-Bargmann dit alors que \mathcal{B} est exactement l’espace

$$\mathfrak{H}(\mathbb{C}) \cap L^2(\mathbb{C}, e^{-|z|^2} dz)$$

des fonctions holomorphes sur \mathbb{C} (espace $\mathfrak{H}(\mathbb{C})$) de carré intégrable pour $e^{-|z|^2} dz$.

8.4. Demi-formes et demi-densités

Malgré leurs limites, les polarisations précédentes $\mathfrak{P}_{vert}, \mathfrak{P}_{hor}, \mathfrak{P}_{hol}$ peuvent paraître bien construites mais elles recèlent en fait une difficulté majeure ayant trait aux principes de la construction du produit scalaire hermitien du Hilbert associé $\mathcal{H}_{vert}, \mathcal{H}_{hor}, \mathcal{H}_{hol}$. Ce point est fort bien expliqué par Matthias Blau [57]. On part du fibré avec connexion (L, ∇) muni d’une polarisation \mathfrak{P} qui est une fibration lagrangienne $\lambda : \Sigma \rightarrow \Lambda$ de Σ de fibres Σ_λ et l’on construit le Hilbert $\mathcal{H}_{\mathfrak{P}}$ à partir des sections \mathfrak{P} -polarisées s de L , i.e. constantes sur les fibres Σ_λ ($\nabla_\xi s = 0$ pour tout $\xi \in \mathfrak{P}$). Mais pour définir correctement le produit scalaire de $\mathcal{H}_{\mathfrak{P}}$ comme un espace L^2 il faut définir correctement la mesure d’intégration. Or la mesure naturelle donnée sur Σ est la mesure de Liouville associée à la forme symplectique Θ . Mais elle ne peut pas convenir. En effet la polarisation feuilletée Σ en divisant sa dimension par 2 et les sections constantes sur les feuilles ne seront pas en général de carré intégrable par rapport à la mesure de Liouville.

97. Dans la section 5.10.1 du chapitre 17, nous rencontrerons la variante

$$(Bf)(z) = A \int_{\mathbb{R}} f(t)e^{2\pi tz - \pi t^2 - \frac{\pi}{2}z^2} dt$$

avec la mesure gaussienne $e^{-\pi|z|^2} dz$ qui s’obtient par changement de variables $z \rightarrow \sqrt{\pi}z$ et $q \rightarrow \sqrt{2\pi}t$.

Pour la polarisation verticale $\pi : T^*\mathbb{R}^n \rightarrow \mathbb{R}^n$, nous avons utilisé une mesure induite sur une section Σ_0 de λ transverse aux fibres de façon à pouvoir définir $\langle s_1, s_2 \rangle_{\mathfrak{P}}$ par $\langle s_1, s_2 \rangle_{\mathfrak{P}} = \int_{\Sigma_0} s_1 \overline{s_2} d\mu$. La fibration lagrangienne $\lambda : \Sigma \rightarrow \Lambda$ était la projection $\pi : T^*\mathbb{R}^n \rightarrow \mathbb{R}^n_{(q)}$ avec pour fibres les $T^*\mathbb{R}^n$, les s polarisées étaient les $s(q)$, la section Σ_0 était le $\mathbb{R}^n_{(q)}$ de base de coordonnées q , la mesure μ était la mesure de Lebesgue $d^n q = dq_1 \cdots dq_n$ et on prenait $\langle s_1, s_2 \rangle_{vert} = \int_{\mathbb{R}^n_{(q)}} s_1 \overline{s_2} d^n q$. Mais pour aller plus loin, il faut en quelque sorte pouvoir “couper en deux” la mesure μ et en attribuer la $\sqrt{\bullet}$ à chaque section s_1 et s_2 .

Pour préciser cette idée intuitive on introduit le nouveau concept, au prime abord un peu sophistiqué, de “demi-densité”. Explicitons-la, toujours pour la polarisation verticale, en suivant Blau [57]

Les mesures d’intégration à valeurs complexes sur la base $\mathbb{R}^n_{(q)}$ de coordonnées q sont les multiples de la mesure de Lebesgue $d^n q$ par un facteur $\mu \neq 0$. Par changement de coordonnées B , une telle μ est multipliée par le Jacobien $J(B)$ (cf. chapitre 2, section 7.2.6). On considère ainsi au-dessus de $\mathbb{R}^n_{(q)}$ les sections μ du fibré V des formes volume à valeurs complexes. Les demi-densités sont les sections du fibré \sqrt{V} . On considère alors comme éléments du Hilbert de la quantification, non plus seulement des s^π de L^π au-dessus de la base $\mathbb{R}^n_{(q)}$ mais des couples (s^π, ν) avec le produit scalaire $\langle s_1^\pi, s_2^\pi \rangle = \int_{\mathbb{R}^n_{(q)}} s_1 \overline{s_2} (\nu_1 \overline{\nu_2}) d^n q$. Cela revient à tensoriser le fibré L par le pull-back $\pi^*(\sqrt{V})$ du fibré \sqrt{V} (où π est la projection $\pi : T^*\mathbb{R}^n \rightarrow \mathbb{R}^n_{(q)}$) et à travailler dans $L \otimes \pi^*(\sqrt{V})$. On choisit alors comme dérivation covariante sur $\pi^*(\sqrt{V})$ la dérivation

$$\nabla_\xi(\sqrt{\mu}) = \frac{1}{2} \frac{1}{\sqrt{Det}} \nabla_\xi(\mu)$$

(règle de dérivation d’une $\sqrt{\bullet}$) avec $\nabla_\xi(\mu) = i_\xi d\mu$.⁹⁸ On a maintenant

$$\nabla_\xi(s^\pi \otimes \nu) = \nabla_\xi(s^\pi) \nu + s^\pi \nabla_\xi(\nu)$$

et les $s^\pi \otimes \nu$ ∇ -constants le long de la polarisation \mathfrak{P} satisfont $\nabla_\xi(s^\pi \otimes \nu) = 0$ pour tout $\xi \in \mathfrak{P}$, ce qui introduit un *terme correcteur*. La formule de la quantification devient

$$\widehat{f}(s^\pi \otimes \nu) = \widehat{f}(s^\pi) \nu - i\hbar s \nabla_{X_f}(\nu) .$$

On peut traiter μ comme une “densité”, c’est-à-dire comme une application qui associe à chaque repère \mathbf{e} de l’espace tangent $T_q\mathbb{R}^n$ une mesure $\mu_{\mathbf{e}} = \mu(\mathbf{e}) d^n q$, où le facteur $\mu(\mathbf{e}) \in \mathbb{C}$ se transforme sous l’action d’un changement de base $B \in GL(n, \mathbb{R})$ de $\mathbb{R}^n_{(q)}$, en $B(\mu(\mathbf{e})) = \det(B) \mu(\mathbf{e})$. Une “demi-densité” est alors une fonction $\nu(\mathbf{e})$

98. Ici $d\mu$ est la différentielle de la section μ (et non pas la mesure d’intégration associée à μ).

satisfaisant $B(\nu(\mathbf{e})) = \det(B)^{\frac{1}{2}} \nu(\mathbf{e})$. Mais la $\sqrt{\bullet}$ étant une fonction multivoque il faut choisir sa détermination. Cela fait passer des transformations linéaires aux transformations “métalinéaires” et pour les structures symplectiques aux structures “métaplectiques”. Une telle construction introduit des “corrections métaplectiques” qui sont nécessaires pour expliquer des phénomènes subtils comme l’indice de Maslov dans les intégrales oscillantes, par exemple celles intervenant dans les caustiques. Nous n’entrons pas dans ce problème trop technique.

8.5. Conditions discrètes de quantification

Ces constructions peuvent être appliquées localement dans le cas d’une variété symplectique (Σ, Θ) quelconque avec des polarisations qui sont des feuilletages lagrangiens. Mais pour pouvoir les globaliser en recollant leurs données sur des cartes locales certaines conditions doivent être satisfaites, en particulier la *condition d’intégrabilité* de Weil disant que la classe de cohomologie de la courbure $\frac{1}{2\pi\hbar}\Theta$ de α est *entière*. Cela permet de retrouver sur la base de fondements purement géométriques les règles de quantification à la Bohr-Sommerfeld d’observables comme l’énergie, règles introduites heuristiquement au tout début de la mécanique quantique pour expliquer les phénomènes expérimentaux comme les raies du spectre d’énergie de l’hydrogène.

Les règles de quantification de la première mécanique quantique sont bien expliquées par Joseph Keller dans son article de 1958 [285]. Niels Bohr introduisit l’idée en 1913 avec son modèle planétaire de l’atome d’hydrogène où l’électron décrit des orbites circulaires autour du proton-noyau. Arnold Sommerfeld le précisa en introduisant des orbites elliptiques et en rendant compte de la structure fine des raies spectrales sous l’action d’un champ magnétique (cf. son traité de référence de 1919 *Structure atomique et lignes spectrales*), puis Albert Einstein le généralisa en 1917.

Ensuite, une première explication géométrique fut donnée pour les systèmes *intégrables*. Il existe dans ce cas (cf. plus haut le théorème d’Arnold) des variables canoniques actions-angles, l’hamiltonien ne dépendant que des actions et les n variables d’action correspondant donc à n intégrales premières indépendantes en involution. S’ils sont bornés, les mouvements se font sur des tores invariants correspondant chacun à une valeur des actions (et donc de l’énergie qui n’est fonction que des actions). Ce sont les actions qui sont quantifiées au moyen du quantum d’action \hbar . La contrainte est que l’intégrale de la 1-forme pdq le long d’un circuit C dans l’espace des phases doit être (à un facteur et une constante près caractéristiques du système) un multiple de \hbar , autrement dit de la forme

$$\int_C pdq \sim n\hbar .$$

En 1926 Léon Brillouin [80], Hendrik Kramers [307] et Gregor Wentzel précisèrent les choses avec une méthode appelée depuis l'approximation BKW.⁹⁹ En s'inspirant du fait que la mécanique quantique est à la mécanique classique ce que l'optique ondulatoire est à l'optique géométrique, ils approximèrent les solutions de l'équation de Schrödinger $\psi(q, t)$ par des $\psi_0(q, t)$ de la forme

$$\psi_0(q, t) = A(q, t) \exp\left(\frac{i}{\hbar} S(q, t)\right)$$

avec S une fonction d'“action”.¹⁰⁰ Si S est alors une fonction “multivaluée”, dans la mesure où $\psi_0(q, t)$ doit être bien définie (univaluée), il faut que l'indétermination ΔS sur la valeur de S soit telle que $\frac{1}{\hbar} \Delta S = 2\pi n$. Par conséquent $\Delta S = n\hbar$. Cela s'applique au cas $\Delta S = \oint \nabla S \cdot ds$ qui est la circulation du gradient ∇S le long d'un lacet. Et si $\nabla S \cdot ds = pdq$ on retrouve la condition $\oint pdq = n\hbar$. Plus précisément, comme on a dans le cas d'un hamiltonien classique indépendant du temps

$$H\left(q, \frac{\partial S}{\partial q}\right) = -\frac{\partial S}{\partial t}$$

(équation de Hamilton-Jacobi) et que $H\left(q, \frac{\partial S}{\partial q}\right) = E$ le long des trajectoires (conservation de l'énergie) on a $\frac{\partial S}{\partial t} = -E$ et donc

$$S = S_0 + \int_0^t pdq - Et$$

le long des trajectoires. Mais $\int pdq$ n'est définie qu'à $\oint pdq$ près, pdq étant une 1-forme fermée mais pas nécessairement exacte si l'espace des phases n'est pas simplement connexe.

Si l'amplitude A est également multivaluée alors la condition devient

$$\Delta S = n \left(\hbar + i \frac{\log(A)}{2\pi} \right) .$$

Si par exemple A n'est définie qu'au signe près, $\Delta \log(A) = -i\pi$ et

$$\Delta S = n \left(\hbar + \frac{1}{2} \right) .$$

99. WKB en Amérique.

100. Nous avons rencontré des approximations de ce type dans la section 6.1 du chapitre 12 consacrée aux caustiques et à l'approximation semi-classique en optique ondulatoire. Nous y reviendrons plus bas dans la section 8.7 consacrée aux solutions asymptotiques et aux approximations semi-classiques en général. Nous y reviendrons également dans la section 10 du chapitre 17 consacrée à la diffusion sous-riemannienne.

Mais ces règles de quantification, d'abord *ad hoc* puis reformulées plus géométriquement, rencontraient de grandes difficultés qui ne furent résolues qu'avec l'avènement de la "vraie" mécanique quantique de Schrödinger et Heisenberg avec le modèle de l'atome d'hydrogène de Wolfgang Pauli de 1925. Puis on relia ensuite tout cela aux "approximations semi-classiques" et aux "solutions asymptotiques" de l'équation de Schrödinger (cf. Keller [285] et, plus bas, la section 8.7 et la section 10 du chapitre 17).

Comme l'explique Charles Marle ([353], p. 21) on peut facilement comprendre pourquoi la quantification de l'énergie associée à la contrainte d'une classe de cohomologie entière est nécessaire en général. Si H est l'hamiltonien du système considéré et si E est une valeur régulière de H , les mouvements se font le long des trajectoires de X_H sur l'hypersurface $\Sigma_H = H^{-1}(E)$ de codimension 1. Dans les bons cas les trajectoires (qui sont de dimension 1) constitueront une variété $\widetilde{\Sigma}_H$ de dimension paire $(n-1) - 1 = n-2$ et la 2-forme symplectique Θ pourra passer au quotient et définira une 2-forme symplectique $\widetilde{\Theta}_H$ sur $\widetilde{\Sigma}_H$. Mais pour pouvoir construire comme ci-dessus un fibré en droites complexes sur $\widetilde{\Sigma}_H$ muni d'une connexion de courbure

$$\frac{\widetilde{\Theta}_H}{2\pi\hbar},$$

il faut, nous l'avons vu, que la classe de cohomologie de $\frac{\widetilde{\Theta}_H}{2\pi\hbar}$ soit entière, ce qui exige la quantification discrète de E .

8.6. Exemples

Donnons maintenant quelques exemples de ces formalismes.

8.6.1. Le fibré cotangent du cercle.

Comme premier exemple, considérons le fibré cotangent $\Sigma = T^*\mathbb{S}^1 \xrightarrow{\pi} \mathbb{S}^1$ du cercle unité \mathbb{S}^1 . Il s'agit d'un fibré cotangent, mais la base \mathbb{S}^1 est compacte non simplement connexe et a un groupe fondamental isomorphe à \mathbb{Z} (nombre de tours effectués par le lacet).

$T^*\mathbb{S}^1$ est trivial et isomorphe au cylindre. Il n'existe que deux classes de fibrés en droites réelles sur \mathbb{S}^1 : le fibré trivial (orientable) et le ruban de Möbius (non orientable).¹⁰¹ Pour le voir on considère \mathbb{S}^1 comme le cercle unité $x^2 + y^2 = 1$ du plan \mathbb{R}^2 ($(x, y) = (r, \varphi)$ en coordonnées polaires) et on considère les vecteurs tangents en (r, φ) de composantes $\{-\lambda \sin(\varphi), \lambda \cos(\varphi)\}$, $\lambda \in \mathbb{R}$. Un vecteur tangent en (r, φ) est donc défini par λ et $T\mathbb{S}^1 \simeq \mathbb{S}^1 \times \mathbb{R}$. Le fibré cotangent $T^*\mathbb{S}^1$ est le fibré dual de $T\mathbb{S}^1$ et il est également trivial et isomorphe à un cylindre.

101. C'est un cas très élémentaire d'un célèbre théorème des années 1930 dû à Eduard Stiefel et Hassler Whitney.

Mais, comme nous l'avons vu dans les remarques des sections 7.2.9 et 7.2.11 du chapitre 2, l'angle φ n'est pas une variable canonique globale sur \mathbb{S}^1 car elle n'est pas périodique et n'est définie que modulo 2π .¹⁰² Cela implique une *discrétisation* du moment conjugué quantifié \widehat{p} . En effet \widehat{p} correspond bien à $-i\hbar\partial_\varphi$ mais à condition de tenir compte du fait que ce sont les fonctions $\cos(\varphi)$, $\sin(\varphi)$ et p qui forment un système complet d'observables sur $T^*\mathbb{S}^1$. Les relations de commutation de Poisson sont :

$$\begin{cases} \{p, \sin(\varphi)\} = \cos(\varphi) \\ \{p, \cos(\varphi)\} = -\sin(\varphi) \\ \{\cos(\varphi), \sin(\varphi)\} = 0 \end{cases}$$

et l'on voit que p et $\sin(\varphi)$ ne suffisent pas pour obtenir un ensemble de fonctions fermé par crochet de Poisson. Il faut obligatoirement y adjoindre $\cos(\varphi)$.

Appliquons alors les procédures de préquantification et de quantification.

Pour la préquantification, la forme symplectique est $\Theta = dp \wedge d\varphi$ et le potentiel symplectique naturel est $\theta = pd\varphi$. Le fibré en droites complexes $\delta : L \longrightarrow T^*\mathbb{S}^1 = \Sigma$ est trivial et la connexion est $\nabla = d - \frac{i}{\hbar}\theta$.

Pour la quantification, prenons d'abord la polarisation *verticale* \mathfrak{P}_{vert} engendrée par ∂_p ce qui donne le Hilbert $L^2(\mathbb{S}^1, \mathbb{C})$ des fonctions périodiques à valeurs complexes. Le spectre de $\widehat{p} = -i\hbar\partial_\varphi$ est défini par

$$\widehat{p}\psi(\varphi) = \lambda\psi(\varphi) .$$

Les solutions sont du type

$$\psi(\varphi) = e^{\mu\varphi} ,$$

$\mu \in \mathbb{C}$ pour la valeur propre

$$\lambda = -i\hbar\mu .$$

Mais comme $\psi(\varphi)$ est 2π -périodique, il faut que $e^{2\pi\mu} = 1$ et donc $\mu = ni$ et donc

$$\lambda = n\hbar .$$

Autrement dit, le spectre de \widehat{p} est $\hbar\mathbb{Z}$.

Mais θ n'étant défini qu'à une jauge proportionnelle à $d\varphi$ près et $d\varphi$ étant fermée mais non exacte, on obtient un continuum de connexions possibles qui sont non équivalentes

$$\nabla^\mu = d - \frac{i}{\hbar}\theta + i\nu d\varphi ,$$

102. Néanmoins les dérivées par rapport à φ peuvent avoir un sens global car la dérivée des constantes $2k\pi$ est nulle. De même la 1-forme $d\varphi$ possède un sens global (mais n'est pas une différentielle exacte).

$\nu \in [0, 1)$.¹⁰³ Ces quantifications sont non équivalentes car $\widehat{p} = -i\hbar\partial_\varphi$ devient $\widehat{p}^\nu = -i\hbar\partial_\varphi + \hbar\nu$ dont le spectre est $\hbar(\mathbb{Z} + \nu)$ et ne peut donc pas être égal à $\hbar\mathbb{Z}$ si $\nu \in (0, 1)$.

Prenons maintenant pour polarisation la polarisation *horizontale* \mathfrak{P}_{hor} engendrée par ∂_φ . Les sections polarisées doivent satisfaire, puisque $d\varphi(\partial_\varphi) = 1$ et $d\psi(\partial_\varphi) = \partial_\varphi\psi$,

$$\begin{aligned}\nabla_{\partial_\varphi}^\mu \psi(p, \varphi) &= \partial_\varphi\psi(p, \varphi) - \frac{i}{\hbar}p\psi(p, \varphi) + i\nu\psi(p, \varphi) = 0 \\ \partial_\varphi\psi(p, \varphi) &= \frac{i}{\hbar}(p - \hbar\nu)\psi(p, \varphi) .\end{aligned}$$

Les solutions sont de la forme

$$\psi(p, \varphi) = \exp\left(\frac{i}{\hbar}(p - \hbar\nu)\varphi\right)\phi(p) .$$

Pour que $\psi(p, \varphi)$ soit périodique en φ il faut donc

$$p - \hbar\nu = n\hbar$$

afin d'avoir

$$\exp\left(\frac{i}{\hbar}(p - \hbar\nu)\varphi\right) = \exp(in\varphi) .$$

D'où la condition de Bohr-Sommerfeld sur la variable canonique classique p .

Bref, la polarisation horizontale a des feuilles qui sont des cercles et ne sont donc pas simplement connexes. Cela implique que la quantification impose une discrétisation à la Bohr-Sommerfeld de la variable canonique p et donc que les fonctions $\psi(p, \varphi)$ soient en fait des *distributions* dont le support soit restreint aux valeurs admissibles de p .

Cela est un phénomène général lorsque les feuilles de \mathfrak{P} ne sont pas simplement connexes.

8.6.2. Les orbites de la représentation co-adjointe de $SE(2)$.

La quantification de cylindres intervient de façon naturelle pour le groupe résoluble des déplacements du plan $SE(2)$ qui est le produit semi-direct $G = \mathbb{R}^2 \rtimes SO(2)$ de notre modèle \mathbb{V}_S avec la loi de composition

$$(s, r_\varphi) \circ (q, r_\theta) = (s + r_\varphi(q), r_{\varphi+\theta})$$

(cf. la section 7 du chapitre 5). En effet les orbites de la représentation coadjointe de G ont génériquement une géométrie de cylindres symplectiques à la Kirillov liée à la structure de produit semi-direct.

103. Cf. [57] p. 24. L'interprétation structurale de ce fait est de nature cohomologique. C'est le groupe de cohomologie $H^1(\Sigma, U(1))$ qui classe les connexions ∇^μ .

Reprenons les notations de la section 3 du chapitre 11.¹⁰⁴ Soient $g = (q, r_\theta)$ les éléments de G , q étant un vecteur de \mathbb{R}^2 et $r_\theta \in SO(2)$ une rotation d'angle θ .¹⁰⁵ Dans la base (\varkappa, τ) de \mathcal{G} , $\varkappa = (\varkappa_1, \varkappa_2)$ avec

$$\varkappa_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \varkappa_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

et $\tau = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ (la matrice $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ étant celle de la multiplication par i des vecteurs du plan, i.e. la rotation de $\frac{\pi}{2}$), la matrice de Ad_g est

$$Ad_g = \begin{pmatrix} e^{i\theta} & -i.q \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & q_2 \\ \sin(\theta) & \cos(\theta) & -q_1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Dans la base duale (\varkappa^*, τ^*) de \mathcal{G}^* , la matrice de Ad_g^* est celle de

$$Ad_{g^{-1}} = \begin{pmatrix} e^{-i\theta} & i.e^{-i\theta}.q \\ 0 & 1 \end{pmatrix}$$

opérant, par dualité, à droite sur des vecteurs lignes et non plus à gauche sur des vecteurs colonnes. Si $\varsigma^* = \lambda^*.\varkappa^* + \mu^*.\tau^* \in \mathcal{G}^*$ on a donc

$$Ad_g^*(\varsigma^*) = (\lambda_1^*, \lambda_2^*, \mu^*) \begin{pmatrix} \cos(\theta) & \sin(\theta) & q_1 \sin(\theta) - q_2 \cos(\theta) \\ -\sin(\theta) & \cos(\theta) & q_1 \cos(\theta) + q_2 \sin(\theta) \\ 0 & 0 & 1 \end{pmatrix},$$

soit

$$\lambda^*.e^{-i\theta}.\varkappa^* + (\lambda^*.ie^{-i\theta}.q + \mu^*)\tau^*.$$

Nous avons vu, toujours à la section 3 du chapitre 11 que la représentation coadjointe de \mathcal{G} dans \mathcal{G}^* est le morphisme $ad^* : \mathcal{G} \rightarrow \text{End } \mathcal{G}^*$ qui associe à tout vecteur tangent $\gamma = (\delta q, \delta \theta) \in \mathcal{G}$ l'endomorphisme ad_γ^* de \mathcal{G}^* qui associe lui-même au covecteur $\varsigma^* = (\lambda^*, \mu^*) \in \mathcal{G}^*$ le covecteur $ad_\gamma^*(\varsigma^*) \in \mathcal{G}^*$ dont la valeur sur le vecteur tangent $\eta = (\delta s, \delta \varphi) \in \mathcal{G}$ est

$$ad_\gamma^*(\varsigma^*)(\eta) = -\varsigma^*([\gamma, \eta]) = \lambda_1^*(\delta s_2 \delta \theta - \delta q_2 \delta \varphi) - \lambda_2^*(\delta s_1 \delta \theta - \delta q_1 \delta \varphi).$$

Soit alors $\varsigma^* = (\lambda^* \neq 0, \mu^*) \in \mathcal{G}^*$. L'orbite co-adjointe $\mathcal{O}_{\varsigma^*}$ est le cylindre ayant pour base le cercle C_{λ^*} du plan \varkappa^* (le cercle centré sur l'origine et passant par λ^*) qui est l'orbite de l'action de $SO(2)$ paramétré par θ à partir de λ^* , et pour

104. Nous reprenons aussi à titre de rappel quelques calculs déjà effectués.

105. Vu le contexte il n'y a pas de confusion possible entre l'angle θ et la 1-forme symplectique $\theta = pdq$.

génératrice l'axe $\mathbb{R}\tau^*$ de coordonnée p (dont les valeurs sont les μ^*).¹⁰⁶ Ce cylindre $C_{\lambda^*} \times \mathbb{R}\tau^*$ s'identifie à l'espace des phases $T^*C_{\lambda^*}$ de l'espace de configuration $C_{\lambda^*} \simeq \mathbb{S}_{(\theta)}^1$. L'espace tangent à \mathcal{O}_{ζ^*} en ζ^* est par définition l'ensemble des $ad_{\gamma}^*(\zeta^*)$ pour $\gamma \in \mathcal{G}$. On aura $ad_{\gamma}^*(\zeta^*) = 0$ si pour tout $\eta = (\delta s, \delta \varphi) \in \mathcal{G}$ on a

$$ad_{\gamma}^*(\zeta^*)(\eta) = -\zeta^*([\gamma, \eta]) = \lambda_1^*(\delta s_2 \delta \theta - \delta q_2 \delta \varphi) - \lambda_2^*(\delta s_1 \delta \theta - \delta q_1 \delta \varphi) = 0 .$$

Comme $ad_{\gamma}^*(\zeta^*) = 0$ signifie que γ appartient au stabilisateur \mathcal{G}_{ζ^*} de ζ^* sous l'action coadjointe de \mathcal{G} sur \mathcal{G}^* , on a donc $T_{\zeta^*}\mathcal{O}_{\zeta^*} \simeq \mathcal{G}/\mathcal{G}_{\zeta^*}$. Mais

$$ad_{\gamma}^*(\zeta^*)(\eta) = -\zeta^*([\gamma, \eta]) .$$

La 2-forme symplectique de Kirillov-Kostant-Souriau Θ_{ζ^*} sur \mathcal{O}_{ζ^*} est donnée par

$$\Theta_{\zeta^*}(ad_{\gamma}^*(\zeta^*), ad_{\eta}^*(\zeta^*)) = -\zeta^*([\gamma, \eta]) = ad_{\gamma}^*(\zeta^*)(\eta) .$$

On vérifie aisément qu'elle est G -invariante et fermée. Elle est bien non dégénérée sur l'orbite \mathcal{O}_{ζ^*} puisque

$$\Theta_{\zeta^*}(ad_{\gamma}^*(\zeta^*), ad_{\eta}^*(\zeta^*)) = ad_{\gamma}^*(\zeta^*)(\eta) = 0$$

pour tout η signifie $ad_{\gamma}^*(\zeta^*) = 0$ et donc précisément $\gamma \in \mathcal{G}_{\zeta^*}$. Or le quotient $\mathcal{G}/\mathcal{G}_{\zeta^*}$ est le plan tangent à l'orbite $T_{\zeta^*}\mathcal{O}_{\zeta^*}$.

Nous avons vu à la section 3.2 du chapitre 12 le formalisme de *l'application moment*. Il s'applique ici de façon particulièrement simple. Rappelons que si (Σ, Θ) est une variété symplectique sur laquelle un groupe de Lie G agit symplectiquement, alors tout $\gamma \in \mathcal{G}$ engendre un champ γ_{Σ} sur Σ , l'application $\gamma \mapsto \gamma_{\Sigma}$ étant un morphisme d'algèbres de Lie. La 1-forme $i_{\gamma_{\Sigma}}\Theta$ est fermée et si elle est de plus exacte alors $i_{\gamma_{\Sigma}}\Theta = dF_{\gamma}$ pour une certaine fonction F_{γ} . Si F_{γ} existe pour tout γ , on peut alors définir l'application moment $j : \Sigma \rightarrow \mathcal{G}^*$ par $j(x)(\gamma) = F_{\gamma}(x)$ pour tout $x \in \Sigma$.

Ici, l'application moment est tout simplement l'inclusion $j : \mathcal{O}_{\zeta^*} \hookrightarrow \mathcal{G}^*$. En effet, si $\gamma \in \mathcal{G}$, $\gamma_{\mathcal{O}_{\zeta^*}}$ est le champ $ad_{\gamma}^*(\zeta^*)$, où $ad_{\gamma}^*(\zeta^*)$ est un vecteur tangent à \mathcal{O}_{ζ^*} en ζ^* . On regarde alors $i_{\gamma_{\mathcal{O}_{\zeta^*}}}\omega_{\zeta^*}$. Comme

$$\Theta_{\zeta^*}(ad_{\gamma}^*(\zeta^*), ad_{\eta}^*(\zeta^*)) = ad_{\gamma}^*(\zeta^*)(\eta) = -\zeta^*([\gamma, \eta]) ,$$

on a

$$i_{\gamma_{\mathcal{O}_{\zeta^*}}}\Theta_{\zeta^*} = -\zeta^*([\gamma, \bullet]) ,$$

ce qui montre que $i_{\gamma_{\mathcal{O}_{\zeta^*}}}\Theta_{\zeta^*}$ redonne $\gamma_{\mathcal{O}_{\zeta^*}}$.

Comme nous l'avons vu à propos du fibré cotangent du cercle $T^*\mathbb{S}^1 \xrightarrow{\pi} \mathbb{S}^1$, la quantification des orbites coadjointes de $SE(2)$ va introduire une discrétisation. Nous allons en donner d'autres variantes.

106. Dans la section 6.3 du chapitre 5 (modèle \mathbb{V}_J), les orbites co-adjointes étaient notées Ω et leur forme de Kirillov σ_{Ω} .

8.6.3. *La fibration lagrangienne du plan épointé.*

Le cylindre $\mathbb{S}^1 \times \mathbb{R}^+$ est diffeomorphe au plan \mathbb{R}^2 épointé de l'origine (noté $\mathring{\mathbb{R}}^2$) si l'on identifie le cercle de hauteur $p > 0$ du cylindre avec le cercle centré sur l'origine de rayon $r = p$ de $\mathring{\mathbb{R}}^2$. Pour le cylindre $\mathbb{S}^1 \times \mathbb{R}$, on peut aussi transformer le moment $p \in \mathbb{R}$ en échelle e^p et identifier le cercle de hauteur p du cylindre avec le cercle de rayon $r = e^p$ de $\mathring{\mathbb{R}}^2$. En coordonnées polaires (r, φ) , $r \in]0, \infty)$, la forme symplectique $\Theta = dx \wedge dy$ de $\mathring{\mathbb{R}}^2$ devient $\Theta = r dr \wedge d\varphi$. Les cercles centrés sur l'origine sont des sous-variétés lagrangiennes puisque $dr = 0$ et donc $\Theta = 0$ et il en va de même des rayons pour lesquels $d\varphi = 0$. On a

$$\partial_\varphi = x\partial_y - y\partial_x = X_H$$

avec $H = \frac{1}{2}r^2$.¹⁰⁷

La polarisation \mathfrak{P} associée est le feuilletage lagrangien de ces cercles défini par $X_H(f) = \partial_\varphi f = 0$. La connexion est

$$\nabla_{X_H}(s) = X_H(s) - \frac{i}{\hbar} H d\varphi(X_H)s = \partial_\varphi(s) - \frac{i}{\hbar} H s .$$

Les sections s ∇ -constantes sur les feuilles de \mathfrak{P} doivent donc satisfaire l'EDP

$$\partial_\varphi(s) - \frac{i}{\hbar} H s = 0$$

et sont donc de la forme

$$\exp\left(\frac{i}{\hbar} \frac{1}{2} r^2 \varphi\right) \sigma(r) .$$

107. Repères calculatoires élémentaires :

$$\begin{aligned} x &= r \cos(\varphi), y = r \sin(\varphi), \\ dx &= dr \cos(\varphi) - r \sin(\varphi) d\varphi, dy = dr \sin(\varphi) + r \cos(\varphi) d\varphi, \\ dx \wedge dy &= (dr \cos(\varphi) - r \sin(\varphi) d\varphi) \wedge (dr \sin(\varphi) + r \cos(\varphi) d\varphi) \\ &= dr \cos(\varphi) \wedge r \cos(\varphi) d\varphi - r \sin(\varphi) d\varphi \wedge dr \sin(\varphi) \\ &= (\cos^2(\varphi) + \sin^2(\varphi)) r dr \wedge d\varphi = r dr \wedge d\varphi, \\ \partial_H f &= (\partial_r f) \frac{dr}{dH} = \frac{1}{r} \partial_r f, \\ X_f &= (\partial_H f) \partial_\varphi - (\partial_\varphi f) \partial_H = \frac{1}{r} (\partial_r f) \partial_\varphi - (\partial_\varphi f) \frac{1}{r} \partial_r = \frac{1}{r} ((\partial_r f) \partial_\varphi - (\partial_\varphi f) \partial_r), \\ X_H &= \partial_\varphi, X_\varphi = -\frac{1}{r} \partial_r, [X_H, X_\varphi] = 0 \\ \{f, g\} &= X_f(g) = \frac{1}{r} ((\partial_r f) (\partial_\varphi g) - (\partial_\varphi f) (\partial_r g)), \\ \{H, \varphi\} &= \frac{1}{r} ((\partial_r H) (\partial_\varphi \varphi) - (\partial_\varphi H) (\partial_r \varphi)) = 1 . \end{aligned}$$

Pour qu'elles soient périodiques de période 2π en φ il faut donc quantifier $\frac{1}{2}r^2$ à la Bohr-Sommerfeld par

$$\frac{1}{2}r^2 = nh ,$$

$n \in \mathbb{Z}$. En fait il faut rajouter une “correction métaplectique” que nous allons expliciter en remarquant que ce modèle de \mathbb{R}^2 est celui de l'oscillateur harmonique avec l'hamiltonien $H = \frac{1}{2}r^2 = \frac{1}{2}(p^2 + q^2)$.

8.6.4. L'oscillateur harmonique.

La polarisation holomorphe \mathfrak{P}_{hol} est particulièrement bien adaptée à la quantification de l'oscillateur harmonique dans la représentation “énergie”. Nous avons plusieurs fois évoqué l'oscillateur harmonique depuis la section 5.4 du chapitre 5. Utilisons ici les coordonnées complexes $z = \frac{1}{\sqrt{2}}re^{i\varphi}$, $\bar{z} = \frac{1}{\sqrt{2}}re^{-i\varphi}$.¹⁰⁸ L'hamiltonien classique est $H = \frac{1}{2}(p^2 + q^2) = z\bar{z} = |z|^2 = \frac{1}{2}r^2$. Le champ hamiltonien est $X_H = i(z\partial_z - \bar{z}\partial_{\bar{z}}) = \partial_\varphi$ et par conséquent

$$\begin{aligned} [X_H, \partial_{\bar{z}}] &= i(z\partial_z - \bar{z}\partial_{\bar{z}})\partial_{\bar{z}} - i\partial_{\bar{z}}(z\partial_z - \bar{z}\partial_{\bar{z}}) \\ &= i(z\partial_z\partial_{\bar{z}} - \bar{z}\partial_{\bar{z}}\partial_{\bar{z}} - z\partial_z\partial_z + \bar{z}\partial_{\bar{z}}\partial_{\bar{z}} + \partial_{\bar{z}}) \\ &= i\partial_{\bar{z}} . \end{aligned}$$

Cela montre que le flot hamiltonien de X_H laisse la polarisation \mathfrak{P}_{hol} invariante puisque \mathfrak{P}_{hol} est engendrée par les $\partial_{\bar{z}}$ et que donc si les $\partial_{\bar{z}}f = 0$ alors les $[X_H, \partial_{\bar{z}}]f = 0$, autrement dit $[X_H, \partial_{\bar{z}}] \in \mathfrak{P}_{hol}$.

La quantification de z et \bar{z} donne $\widehat{z}s = zs$, $\widehat{\bar{z}}s = \hbar\partial_zs$ et donc celle de $H = z\bar{z}$ donne

$$\widehat{H}s = \frac{1}{2}(\widehat{z}\widehat{\bar{z}} + \widehat{\bar{z}}\widehat{z})s = \hbar\left(z\partial_z + \frac{1}{2}\right)s .$$

La mesure gaussienne fait que \widehat{z} et $\widehat{\bar{z}}$ sont adjoints, c'est-à-dire que

$$\begin{aligned} \langle zf, g \rangle &= \int_{\mathbb{C}^n} zf(z)\overline{g(z)}e^{-\frac{|z|^2}{\hbar}}d^nz \\ &= \langle f, \hbar\partial_zg \rangle = \int_{\mathbb{C}^n} f(z)\overline{\hbar\partial_zg(z)}e^{-\frac{|z|^2}{\hbar}}d^nz \\ &= \hbar \int_{\mathbb{C}^n} f(z)\partial_{\bar{z}}\overline{g(z)}e^{-\frac{|z|^2}{\hbar}}d^nz . \end{aligned}$$

Pour le vérifier on fait une intégration par parties en utilisant

$$\partial_{\bar{z}}f(z)\overline{g(z)}e^{-\frac{|z|^2}{\hbar}} = \left(f(z)\partial_{\bar{z}}\overline{g(z)} + g(z)\partial_{\bar{z}}f(z) - \frac{z}{\hbar}f(z)g(z)\right)e^{-\frac{|z|^2}{\hbar}} .$$

108. Le jacobien du changement de variables et son inverse sont :

$$\begin{pmatrix} \partial_r z & \partial_\varphi z \\ \partial_r \bar{z} & \partial_\varphi \bar{z} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\varphi} & ire^{i\varphi} \\ e^{-i\varphi} & -ire^{-i\varphi} \end{pmatrix}, \begin{pmatrix} \partial_z r & \partial_{\bar{z}} r \\ \partial_z \varphi & \partial_{\bar{z}} \varphi \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{-i\varphi} & e^{i\varphi} \\ -\frac{i}{r}e^{-i\varphi} & \frac{i}{r}e^{i\varphi} \end{pmatrix} .$$

En intégrant sur \mathbb{C}^n , comme l'intégrale du terme de gauche donne 0 par le théorème de Stokes et comme $\partial_{\bar{z}}f(z) = 0$, on trouve bien la relation cherchée

$$0 = \int_{\mathbb{C}^n} f(z) \overline{\partial_{\bar{z}}g(z)} e^{-\frac{|z|^2}{\hbar}} d^n z - \frac{1}{\hbar} \int_{\mathbb{C}^n} z f(z) g(z) e^{-\frac{|z|^2}{\hbar}} d^n z .$$

Les fonctions propres de \widehat{H} sont les fonctions holomorphes satisfaisant

$$\hbar \left(z \partial_z + \frac{1}{2} \right) s(z) = \lambda s(z)$$

et donc les z^n ($n \geq 0$) avec la valeur propre

$$\lambda = \hbar \left(n + \frac{1}{2} \right) ,$$

le $\frac{\hbar}{2}$ pour $n = 0$ correspondant à “l'énergie du vide” et venant de la symétrisation de \widehat{H} . Avec

$$\widehat{H}s(z) = \widehat{z}\widehat{z}s(z) = \hbar z \partial_z s(z)$$

on aurait $\lambda = \hbar n$ et avec

$$\widehat{H}s(z) = \widehat{z}\widehat{z}s(z) = \hbar \partial_z (zs(z))$$

on aurait $\lambda = \hbar(n + 1)$.

Ce terme supplémentaire $\frac{\hbar}{2}$ peut s'interpréter plus profondément comme une “correction métaplectique” manifestant le rôle des demi-densités dans la quantification. Nous avons vu que la connexion ∇ sur le fibré des densités μ est donnée par $\nabla_X \mu = i_X d\mu = X(\mu)$. C'est une connexion “plate”, c'est-à-dire de courbure $[\nabla_X, \nabla_Y] - \nabla_{[X, Y]}$ nulle. En effet,

$$[\nabla_X, \nabla_Y] \mu = \nabla_{[X, Y]} \mu = [X, Y](\mu) .$$

De façon générale, la correction métaplectique est définie de la façon suivante. Soit $\mathbf{X} = \{X_j\}$ n champs hamiltoniens engendrant la polarisation \mathfrak{P} . Si f préserve \mathfrak{P} , i.e. si $f \in \mathfrak{P}$, alors il existe une matrice A_f telle que $[X_f, \mathbf{X}] = A_f \mathbf{X}$. Le terme de correction est alors $-\frac{1}{2} i \hbar \operatorname{tr}(A)$. Ici \mathfrak{P} est $\partial_{\bar{z}}$, H préserve \mathfrak{P} et $[X_H, \partial_{\bar{z}}] = i \partial_{\bar{z}}$. Donc $\operatorname{tr}(A) = i$ et le terme correctif est $-\frac{1}{2} i \hbar i = \frac{\hbar}{2}$.

Nous verrons plus loin dans le chapitre 17 sur les états cohérents, dans la section 5 sur l'analyse harmonique non commutative, et plus précisément dans la section 5.10.1 sur la représentation de Segal-Bargmann et aussi dans la section 5.10.2, comment on retrouve purement algébriquement la façon dont cette quantification de l'oscillateur harmonique opère sur des fonctions d'ondes holomorphes à travers les opérateurs de “création” et d’“annihilation” que sont \widehat{z} et $\widehat{\bar{z}}$.

8.7. Approximations semi-classiques

8.7.1. La quantification de Weyl.

Après avoir présenté ces quelques exemples simples, nous voudrions esquisser la façon dont la quantification géométrique acquiert une *représentation fonctionnelle*, c'est-à-dire la façon dont les opérateurs quantiques peuvent s'écrire comme des *intégrales*. La plus connue de ces représentations est la *quantification de Weyl* qui consiste, étant donnée une fonction $a(x, \xi)$ sur le fibré cotangent $T^*\mathbb{R}^n = \mathbb{R}^n_{(x)} \times \mathbb{R}^n_{(\xi)}$, à lui associer l'opérateur qu'est l'intégrale oscillante

$$a^W u(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\xi \cdot (x-y)} a\left(\frac{x+y}{2}, \xi\right) u(y) dy d\xi.$$

On remarquera que si $a(x, \xi) = 1$ alors

$$\begin{aligned} a^W u(x) &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\xi \cdot (x-y)} u(y) dy d\xi = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\xi \cdot x} e^{-i\xi \cdot y} u(y) dy d\xi \\ &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\xi \cdot x} \widehat{u}(\xi) d\xi = u(x) \end{aligned}$$

et donc $1^W = Id$. On voit que ce résultat est dû au fait que $e^{i\xi \cdot (x-y)}$ étant le produit $e^{i\xi \cdot x} e^{-i\xi \cdot y}$ on peut utiliser $e^{-i\xi \cdot y}$ pour effectuer une transformée de Fourier et $e^{i\xi \cdot x}$ pour faire une transformée de Fourier inverse. De même si $a(x, \xi) = x$ (en dimension 1)

$$\begin{aligned} a^W u(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{i\xi \cdot (x-y)} \frac{x+y}{2} u(y) dy d\xi = \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi \cdot x} \frac{x}{2} \left(\int_{\mathbb{R}} e^{-i\xi \cdot y} u(y) dy \right) d\xi + \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi \cdot x} \left(\int_{\mathbb{R}} e^{-i\xi \cdot y} \frac{y}{2} u(y) dy \right) d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi \cdot x} \frac{x}{2} \widehat{u}(\xi) d\xi + \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi \cdot x} \frac{1}{2i} \left(\frac{\partial}{\partial \xi} \widehat{u}(\xi) \right) d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi \cdot x} \frac{x}{2} \widehat{u}(\xi) d\xi + \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi \cdot x} \frac{1}{2i} \left(\frac{\partial}{\partial \xi} \widehat{u}(\xi) \right) d\xi \\ &= \frac{x}{2} u(x) + \frac{x}{2} u(x) = xu(x) \end{aligned}$$

et donc $x^W =$ multiplication par x . En plusieurs dimensions $x_j^W =$ multiplication par x_j .

De même encore si $a(x, \xi) = \xi$ (toujours en dimension 1),

$$\begin{aligned} a^W u(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{i\xi \cdot (x-y)} \xi u(y) dy d\xi = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi \cdot x} \xi \left(\int_{\mathbb{R}} e^{-i\xi \cdot y} u(y) dy \right) d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\xi \cdot x} \xi \widehat{u}(\xi) d\xi = \widehat{\xi \widehat{u}(\xi)} = -i \frac{\partial}{\partial x} u(x) \end{aligned}$$

et donc

$$\xi^W = -i \frac{\partial}{\partial x} u(x) .$$

En plusieurs dimensions $\xi_j^W = -i \frac{\partial}{\partial x_j} u(x)$.

En introduisant un facteur \hbar dans la phase et en compensant cette introduction par un facteur de l'intégrale on peut faire que les opérateurs a^W satisfassent aux règles de quantification. Les a^W sont donc des représentations fonctionnelles des opérateurs quantiques associés aux quantités classiques. Ceci dit, le comportement de ces opérateurs par rapport aux produits des fonctions n'est pas classique. On a $a^W \circ b^W = (a \bullet b)^W$ où $a \bullet b$ est le *produit de Moyal* (aussi dit de *Weyl–Groenewold*) des symboles a et b . Ce produit est égal au produit normal ab au premier ordre en \hbar mais comprend des termes (compliqués) en \hbar d'ordre supérieur qui le rend *non commutatif*. C'est une déformation non commutative du produit classique. On obtient alors la règle de quantification appropriée $[a^W, b^W] = -i\hbar \{a, b\}^W$ mais seulement au deuxième ordre en \hbar .

Lorsque $\hbar \rightarrow 0$ on obtient ainsi des approximations semi-classiques du quantique. Il est donc pertinent d'explorer la façon dont on peut calculer ce type d'intégrales oscillantes, en particulier au moyen de développements asymptotiques en \hbar .

Pour cela, nous allons commencer par revenir sur l'exemple qui a servi historiquement de modèle, celui des caustiques en optique, car l'optique ondulatoire est la quantification de l'optique géométrique et l'on connaît les représentations fonctionnelles des solutions de l'équation des ondes au voisinage des caustiques. Dans une certaine mesure la mécanique quantique est une immense généralisation de l'optique et beaucoup de ses formalismes plongent leurs racines dans ceux développés pour l'optique dès le XIX^e siècle.

8.7.2. Retour sur les caustiques en optique.

Dans la section 6.1 du chapitre 12 consacrée aux caustiques et aux solutions asymptotiques de l'équation des ondes dans \mathbb{R}^3 , nous avons vu comment le symbole de l'équation des ondes interprété comme l'hamiltonien $1 - \|\xi\|^2$ permet de reconstruire toute l'optique géométrique (rayons, fronts d'ondes, équation eikonale, solutions lagrangiennes, caustiques, etc.). Nous allons maintenant considérer la façon dont une "chair" ondulatoire peut se greffer sur ce "squelette" géométrique.

Commençons par un retour sur cette infrastructure géométrique décrite dans la section 6.1 du chapitre 12. On cherche des solutions de l'équation des ondes

$$Df(x, t) = \frac{\partial^2 f(x, t)}{\partial t^2} - \Delta = 0$$

qui est de symbole $\tau^2 - \|\xi\|^2$ (cf. plus bas section 9.1). Les ondes planes

$$a(k, \omega) e^{i(k \cdot x - \omega t + \varphi_0)},$$

où $k \neq 0$ est un vecteur d'onde (un covecteur dans l'espace dual de celui des vecteurs x), ω une pulsation, φ_0 une phase initiale et $a(k, \omega)$ une amplitude, sont des solutions évidentes si $\omega = \|k\|$. Elles se propagent à la vitesse $\frac{\omega}{\|k\|} = 1$. Comme D est un opérateur linéaire, le principe de superposition est valide et donc les intégrales

$$u(x, t) = \int a(k) e^{i(k \cdot x - \|k\|t + \varphi_0)} dk$$

sont aussi des solutions.

Ceci dit, résoudre l'équation avec des conditions initiales données est très difficile. Par exemple une solution élémentaire $\psi(x, t)$ satisfaisant $D\psi(x, t) = \delta(x, t)$ (distribution de Dirac à l'origine) est une distribution

$$\psi(x, t) = \frac{1}{2\pi} H(t) \delta(t^2 - \|x\|^2)$$

dont le support est sur le cône de lumière positif $t = \|x\|$ (cf. Gårding [204]).

Pour étudier la façon dont l'optique ondulatoire redonne l'optique géométrique, on sépare, rappelons-le, les variables temporelles et spatiales et l'on cherche plutôt des solutions ayant une fréquence τ donnée, c'est-à-dire de la forme

$$F(x, t) = e^{i\tau t} f(x),$$

$f(x)$ étant une amplitude satisfaisant $D_\tau f(x) = 0$, où

$$D_\tau = \tau^2 + \Delta,$$

avec pour condition initiale la donnée d'une fonction $f_0(x)$ sur le front d'onde source S_0 . L'approximation de l'optique géométrique correspondant à une fréquence τ infinie, on recherche des solutions asymptotiques

$$f_\tau(x) = A_\tau(x) e^{i\tau\varphi(x)}$$

des équations $D_\tau f_\tau = \varepsilon_\tau$ où ε_τ est une fonction à décroissance rapide en τ .¹⁰⁹ La fonction $\varphi(x)$ est une phase spatiale, et l'amplitude $A_\tau(x)$ admet par hypothèse un

109. Dans cette problématique et dans ses généralisations, les fonctions à décroissance rapide permettent de passer au "squelette géométrique" de la situation en les considérant comme ~ 0 .

développement asymptotique en τ de la forme :

$$A_\tau(x) \sim \tau^\mu \left[A_0(x) + \frac{1}{\tau} A_1(x) + \cdots + \frac{1}{\tau^k} A_k(x) + \cdots \right]$$

avec $A_0 \not\equiv 0$. En remarquant alors que, comme f_τ doit être solution de $D_\tau f_\tau = \varepsilon_\tau$ avec ε_τ à décroissance rapide, les coefficients de ce développement asymptotique doivent tous être identiquement nuls, on obtient l'équation eikonale $1 - |\nabla\varphi|^2 \equiv 0$ et les équations de transport pour les coefficients A_k .

En passant au fibré cotangent et au formalisme hamiltonien avec l'hamiltonien $H(x, \xi) = 1 - \|\xi\|^2$ (le symbole principal¹¹⁰) le gradient $\nabla\varphi$ devient la 1-forme $\xi = d\varphi$ et l'équation eikonale devient l'équation de Hamilton-Jacobi

$$H(x, d\varphi) = 1 - \|d\varphi\|^2 = 0 .$$

Le graphe Λ_φ de $d\varphi$ sur l'ouvert U de \mathbb{R}^3 où la phase spatiale φ est définie est une sous-variété lagrangienne de T^*U transverse aux fibres de la projection canonique $\pi : T^*U \rightarrow U$. Pour tenir compte des caustiques, on doit considérer des solutions lagrangiennes Λ qui ne sont plus forcément transverses et dont le contour apparent par projection donne précisément les caustiques, la question devenant alors celle de leur représentation fonctionnelle sur ces lieux singuliers.

Rappelons par conséquent, quitte à nous répéter, quelle est l'origine de l'obstruction à une représentation fonctionnelle

$$f_\tau(x) = A_\tau(x) e^{i\tau\varphi(x)}$$

due à l'existence de caustiques. La phase spatiale $\varphi(x)$ s'identifie à une longueur de chemin optique régie par un principe variationnel (principe de Fermat). Dans \mathbb{R}^3 plusieurs rayons (plusieurs géodésiques) peuvent passer par un même point x . Si α est un paramétrage local de S_0 , $\varphi(x)$ dépend aussi de α et les origines sur S_0 des rayons passant par x sont données par $\frac{\partial\varphi}{\partial\alpha} = 0$, *i.e.* par les points critiques de $\varphi(x, \alpha)$ à x constant (ce qui signifie que les rayons sont orthogonaux au front d'onde initial S_0 puisque l'orthogonalité s'exprime par la minimisation du chemin optique de x à S_0). Il y a caustique lorsqu'il y a coalescence de deux rayons passant par x , *i.e.* lorsqu'un des points critiques α devient *dégénéré*. L'existence de caustiques est donc lié à la *multiformité* de l'action φ . En passant dans l'espace des phases $T^*\mathbb{R}^3$, l'introduction des coordonnées conjuguées ξ lève l'obstruction et "désingularise" la situation. Les bicaractéristiques ne se coupent plus et l'on peut construire des solutions lagrangiennes globales. Les caustiques se retrouvent alors par projection de l'espace des phases $T^*\mathbb{R}^3$ sur l'espace de configuration \mathbb{R}^3 .

110. Nous verrons plus bas à la section 9 que le symbole complet de l'équation des ondes est $\tau^2 - \|\xi\|^2$. Ici $\tau = 1$ car la variable t est prise en charge par le facteur oscillant $e^{i\tau t}$.

En ce qui concerne maintenant les représentations fonctionnelles locales des solutions lagrangiennes, l'idée fondamentale est due à Viktor Maslov. Elle consiste à chercher des solutions asymptotiques $f_\tau(x)$ qui ne sont plus simplement de type $A_\tau(x)e^{i\tau\varphi(x)}$ mais des sommes localement finies d'intégrales oscillantes définissant des patterns de diffraction :

$$I(x, \tau) = \left(\frac{\tau}{2\pi}\right)^{p/2} \int e^{i\tau\varphi(x, \alpha)} a_\tau(x, \alpha) d\alpha$$

où α parcourt \mathbb{R}^p , où $(\tau/2\pi)^{p/2}$ est un facteur de renormalisation et où $a_\tau(x, \alpha)$, supposé être de support compact en α , admet un développement asymptotique

$$a_\tau(x, \alpha) = \sum_{j=0}^{\infty} \tau^{\mu-j} a_j(x, \alpha)$$

analogue à celui des $A_\tau(x)$. Le raccord entre ces représentations fonctionnelles et les solutions lagrangiennes définies plus haut se fait à travers le principe fondamental dit de la *phase stationnaire* qui explique les interférences destructrices faisant passer de l'optique ondulatoire à l'optique géométrique.¹¹¹ Selon ce principe, une intégrale oscillante se concentre lorsque τ tend vers l'infini sur le lieu critique V_φ où la phase $\varphi_x(\alpha) = \varphi(x, \alpha)$ est stationnaire (i.e. où $\frac{\partial\varphi_x(\alpha)}{\partial\alpha} = 0$), cette concentration étant définie à équivalence à une fonction à décroissance rapide près.

Lorsque les points critiques $\alpha_1, \dots, \alpha_r$ de $\varphi_x(\alpha)$ sont non dégénérés, l'intégrale oscillante $\int e^{i\tau\varphi(x, \alpha)} a_\tau(x, \alpha) d\alpha$ est approximée par

$$\sum_{k=1}^r \frac{\tau^{\mu-p/2}}{|\det H_k|^{1/2}} a_\tau(x, \alpha_k) e^{i(\tau\varphi(x, \alpha_k) + \frac{\pi s}{4})}$$

où H_j est le hessien de φ_x en α_j et s sa signature.¹¹² C'est bien une somme de termes de la forme $f_\tau(x) = A_\tau(x)e^{i\tau\varphi(x)}$ avec $A_\tau(x) \approx \sum_{j=0}^{\infty} \tau^{\nu-j} A_j(x)$.

Lorsque des points critiques sont dégénérés (i.e. sur la caustique) certains des $\det H_j$ s'annulent et cette approximation diverge. Mais l'expression intégrale reste valable. La méthode BKW (ou WKB) de la phase stationnaire consiste à introduire des paramètres internes supplémentaires α pour la phase spatiale φ . D'après un théorème d'Hörmander, on peut ainsi représenter localement toute solution lagrangienne et donc localement toute caustique.

Soit alors Λ une solution lagrangienne globale de projection $\pi(\Lambda)$, de lieu critique Σ et de caustique (de contour apparent) $C = \pi(\Sigma)$.

111. Nous avons déjà évoqué ce principe à la section 9.3 du chapitre 9 et à la section 6 ci-dessus.

112. Cette formule est subtile et difficile à obtenir. Le terme $\frac{\pi s}{4}$ a été introduit par Maslov.

1. Si $x \notin \pi(\Lambda)$, les intégrales oscillantes sont à décroissance rapide en τ . $\mathbb{R}^3 - \pi(\Lambda)$ est la zone d'ombre.
2. Si $x \in \pi(\Lambda) - C$, les points critiques de $\varphi_x(\alpha)$ sont non dégénérés. Les intégrales oscillantes décroissent comme $\left(\frac{1}{\tau}\right)^{p/2}$ et, étant donné le facteur de normalisation $\left(\frac{\tau}{2\pi}\right)^{p/2}$, on peut éliminer α et représenter localement Λ par une phase purement spatiale. $\pi(\Lambda) - C$ est la zone de lumière.
3. Si $x \in C$ et si α est un point de Σ au-dessus de x (un point critique dégénéré de $\varphi_x(\alpha)$) alors l'intégrale oscillante $I(x, \tau)$ décroît comme $\left(\frac{1}{\tau}\right)^{p/2-\beta}$ où β est le degré de la singularité (α, q) . Si Λ est structurellement stable, le calcul de $I(x, \tau)$ se ramène à celui d'une intégrale oscillante de type

$$J(y, \tau) = \int e^{i\tau\psi(y, \alpha)} f(\alpha) d(\alpha)$$

où $\psi(y, \alpha)$ est un *déploiement universel* de la singularité (α, x) . On obtient ainsi des patterns de diffraction associés aux singularités des caustiques. Les développements asymptotiques ne sont plus en τ^μ mais en $\tau^\mu(\log(\tau))^\nu$.

- (i) Dans le cas, le plus simple, celui d'une singularité pli (singularité de codimension 1), on retrouve ainsi l'intégrale oscillante

$$J(y, \tau) = \int e^{i\tau(\alpha^3/3 + y\alpha)} d\alpha$$

(où $\alpha \in \mathbb{R}$: corang 1) introduite dès 1838 par George Airy dans [11] pour exprimer l'intensité lumineuse au voisinage d'une caustique.¹¹³

- (ii) Dans le cas d'une singularité cusp, on retrouve l'intégrale oscillante

$$J(y, \tau) = \int e^{i\tau(\alpha^4/4 + y_1\alpha^2 + y_2\alpha)} d\alpha$$

étudiée par Trevor Pearcey (1919–1998) en 1946 dans [403] (cf. figure 16).

- (iii) Dans le cas de l'ombilic elliptique, les travaux de Michael Berry [52] manifestent un accord tout aussi spectaculaire entre la théorie et l'expérience.

113. George Biddell Airy (1801–1892) fut un éminent spécialiste des phénomènes lumineux (arc-en-ciel) et de l'astronomie et des télescopes. Lauréat de nombreux prix, il fut président de la Royal Society qu'il rejoignit en 1836.

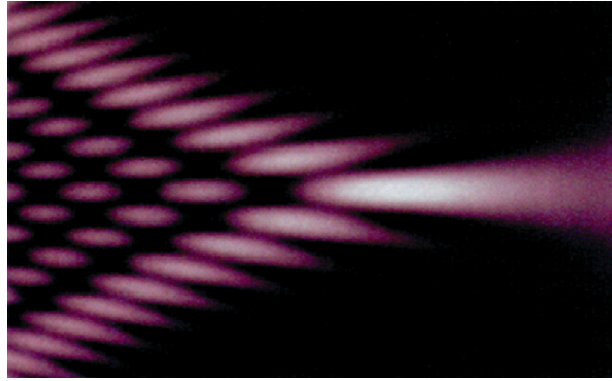


FIGURE 16. Franges de diffraction de Pearcey au voisinage d'un cusp d'une caustique. (Image de M. Berry)

8.7.3. De l'optique à la mécanique quantique.

Comme nous l'avons souvent dit, la mécanique classique est une optique généralisée où l'action joue le rôle de longueur du chemin optique, le principe de moindre action celui du principe de Fermat, l'équation de Hamilton-Jacobi celui du principe de Huygens, les trajectoires celui des rayons et l'impulsion celui de la vitesse normale des fronts d'ondes. "L'optique ondulatoire" associée est la mécanique quantique et c'est pourquoi la théorie des caustiques a grandement anticipé historiquement sur les formalismes de la mécanique quantique.

La méthode BKW (ou WKB, vue plus haut section 8.5) de la phase stationnaire et des approximations semi-classiques que sont les développements asymptotiques en $1/\hbar$ lorsque la constante de Planck $\hbar \rightarrow 0$ ($1/\hbar$ jouant le rôle de τ) en est un exemple. Elle consiste à généraliser ces formalismes à des espaces de configuration qui sont des espaces fonctionnels de chemins et à calculer des intégrales oscillantes, dites intégrales de chemin de Feynman

$$K = \int_{\Gamma} e^{iS(\gamma)/\hbar} d\gamma$$

où $d\gamma$ est une différentielle de chemin (intégration dans l'espace fonctionnel des chemins¹¹⁴) et S une fonctionnelle d'action définie sur les chemins γ . Le principe de la phase stationnaire dit ici que, dans la limite semi-classique $\hbar \rightarrow 0$, l'intégrale de Feynman K se concentre sur les chemins classiques γ_c qui sont solutions du principe variationnel classique $\delta S(\gamma) = 0$ (équations d'Euler-Lagrange). Comme pour le cas des caustiques, lorsqu'un chemin classique γ_c est un point critique non dégénéré de l'action S on peut calculer facilement K à partir de ce principe. Lorsque γ_c est un

114. Il est très délicat de donner un sens rigoureux à ces intégrales car il faut donner un sens à la mesure $d\gamma$.

point critique dégénéré, il faut introduire les déploiements universels de singularités. On pourra se référer entre autres à Dangelmayr-Güttinger-Veit [132].

L'interprétation de la mécanique classique comme "squelette" de la mécanique quantique dans l'approximation semi-classique a permis de développer un immense domaine, celui du "calcul" de Weyl et du "calcul de Kohn-Nirenberg", à savoir celui des opérateurs pseudo-différentiels. Nous allons maintenant en dire un mot.

9. Opérateurs pseudo-différentiels (OΨD)

9.1. Généralités

Parmi les opérateurs, les opérateurs différentiels *linéaires* (OD) jouent un rôle central et la complexité de leur théorie générale est immense.¹¹⁵ Sans pouvoir entrer vraiment dans les détails, nous voudrions quand même en esquisser certains aspects.

Sur M une variété différentiable. Un opérateur différentiel linéaire s'écrit en coordonnées locales $x = (x_i)_{i=1,\dots,n}$ sous la forme

$$P = \sum_{\alpha, 0 \leq |\alpha| \leq m} a_\alpha(x) \frac{\partial^\alpha}{\partial x^\alpha}$$

les $a_\alpha(x)$ étant des fonctions coefficients et les α des multi-indices $\alpha = (\alpha_i \geq 0)_{i=1,\dots,n}$ de degré $|\alpha| = \alpha_1 + \dots + \alpha_n$. Il est linéaire en les dérivées partielles $\frac{\partial^\alpha}{\partial x^\alpha}$. Le nombre m (qui est l'ordre maximal des dérivations intervenant dans P) est "l'ordre" de P .¹¹⁶

Certains cas sont typiques. Par exemple :

- (i) $M = \mathbb{R}^n$ et les coefficients a_α sont *constants*, c'est-à-dire invariants par translation. Ce cas sert de modèle aux approximations locales d'un P quelconque au voisinage d'un point x_0 : on considère l'opérateur à coefficients constants $a_\alpha(x_0)$ sur l'espace tangent $T_{x_0}M$.
- (ii) P est *homogène* d'ordre m : $P = \sum_{\alpha, |\alpha|=m} a_\alpha(x) \frac{\partial^\alpha}{\partial x^\alpha}$.

La question est en général de trouver les solutions de l'EDP

$$Pf(x) = u(x)$$

avec des conditions aux limites appropriées. Des EDP de base sont

– l'équation de Laplace

$$\Delta f(x) = 0$$

avec des conditions au bord de Dirichlet ou de Neumann,

115. Les opérateurs différentiels non linéaires soulèvent des problèmes d'une extrême difficulté et nous ne pouvons pas les aborder ici.

116. Si la référence à P est nécessaire, on peut noter m_P ou (P, m) .

– l'équation de Poisson

$$\Delta f(x) = u(x)$$

avec le même genre de conditions aux limites (puisque les solutions sont définies à une solution de l'équation de Laplace près),

– l'équation de la chaleur

$$\frac{\partial f(x, t)}{\partial t} = \Delta f(x, t)$$

avec $f(x, 0) = u(x)$ (nous reviendrons sur ces équations à la section 9 du chapitre 17),

– l'équation des ondes

$$\frac{\partial^2 f(x, t)}{\partial t^2} = \Delta f(x, t)$$

avec $f(x, 0) = u(x)$ (équation que nous avons déjà discutée à propos des caustiques dans la section 6.1 du chapitre 12 et ci-dessus),

– l'équation de Schrödinger

$$i\hbar \frac{\partial \psi(q, t)}{\partial t} = H\psi(q, t) = \left(-\frac{\hbar^2}{2m} \Delta + V(x) \right) \psi(q, t)$$

où $\psi(q, t) \in L^2(\mathbb{R}^n)$ est la fonction d'onde d'une particule quantique de masse m , $V(x)$ le potentiel et

$$H = -\frac{\hbar^2}{2} \Delta + V(x)$$

l'opérateur hamiltonien associé à l'énergie totale = énergie cinétique + énergie potentielle. Nous avons déjà rencontré plusieurs fois les opérateurs quantiques, en particulier à la section 5.4 du chapitre 5 et dans la section précédente 8.

Tout de suite on voit apparaître des conditions à préciser.

1. Travaille-t-on globalement dans \mathbb{R}^n ou localement puis par recollement de cartes locales dans M ? Nous nous placerons dans le cas \mathbb{R}^n pour simplifier.
2. Quelles sont les propriétés de différentiabilité des fonctions coefficients ? En général on suppose les $a(x)$ différentiables C^∞ .
3. Quelles sont d'autres conditions qu'il serait naturel d'imposer aux $a(x)$ pour avoir de "bonnes" théories ? Nous allons voir qu'il intervient de façon naturelle de nombreuses contraintes de *croissance* à l'infini.

Le problème de base est évidemment que pour résoudre une EDP de type $Pf = u$, il faut savoir *inverser* l'opérateur P afin de pouvoir écrire $f = P^{-1}u$. L'idée qui s'impose immédiatement est de considérer les transformées de Fourier $f(x) \rightsquigarrow \widehat{f}(\xi)$.

Nous reviendrons de façon plus précise sur ces transformées en termes d'analyse harmonique à la section 2.1.1 du chapitre 17.¹¹⁷ Rappelons simplement ici qu'elles correspondent à l'analyse "fréquentielle" des $f(x)$ comme superpositions d'ondes planes $e^{i\xi \cdot x} : \widehat{f}(\xi) = \int_{\mathbb{R}^n} e^{-i\xi x} f(x) dx$ et que, sous des conditions que nous expliciterons plus bas, on peut les inverser au moyen de la formule $f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\xi x} \widehat{f}(\xi) d\xi$.

Ce qui est essentiel dans le présent contexte est la dualité entre dérivation et multiplication disant que

$$\widehat{\frac{\partial}{\partial x_j} f(x)}(\xi) = i\xi_j \cdot \widehat{f}(\xi) .$$

Cette propriété fondamentale conduit d'ailleurs souvent à remplacer les $\frac{\partial}{\partial x_j}$ par les opérateurs $D_{x_j} = -i\frac{\partial}{\partial x_j}$ dont la transformée de Fourier est la multiplication par $-i^2\xi_j = \xi_j$. En remplaçant les coefficients a_α par $i^{|\alpha|}a_\alpha$, on peut écrire $Pf(x) = \sum_{\alpha, |\alpha| \leq m} a_\alpha(x) D_x^\alpha f(x)$.¹¹⁸ Si l'on écrit alors que les $D_x^\alpha f(x)$ sont les transformées de Fourier inverses des $\xi^\alpha \widehat{f}(\xi)$ (si les conditions d'inversion sont satisfaites) et si l'on intervertit les \sum et les \int on obtient

$$\begin{aligned} Pf(x) &= \frac{1}{(2\pi)^n} \sum_{\alpha, |\alpha| \leq m} a_\alpha(x) \int_{\mathbb{R}^n} e^{i\xi \cdot x} \xi^\alpha \widehat{f}(\xi) d\xi \\ &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\xi \cdot x} \left(\sum_{\alpha, |\alpha| \leq m} a_\alpha(x) \xi^\alpha \right) \widehat{f}(\xi) d\xi \quad (\text{OD}) \\ &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\xi \cdot x} \sigma(x, \xi) \widehat{f}(\xi) d\xi \\ &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\xi \cdot x} \sigma(x, \xi) \left(\int_{\mathbb{R}^n} e^{-i\xi y} f(y) dy \right) d\xi \\ &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\xi \cdot (x-y)} \sigma(x, \xi) f(y) dy d\xi \end{aligned}$$

où $\sigma(x, \xi) = \sum_{\alpha, |\alpha| \leq m} a_\alpha(x) \xi^\alpha$ est un polynôme en ξ de degré m égal à l'ordre de P .¹¹⁹ Ce polynôme $\sigma(x, \xi)$ s'appelle le "symbole" de l'OD P et son terme homogène

117. En général, la transformée de Fourier est notée $\widehat{f}(\omega)$, la variable ω duale de x étant pensée comme une fréquence. Ici, la notation ξ renvoie plutôt à une interprétation comme covecteur (cf. plus bas).

118. Il est donc naturel de considérer que les coefficients a_α peuvent être à valeurs complexes.

119. $\sigma(x, \xi)$ peut être noté $\sigma_P(x, \xi)$ lorsque la référence à P est nécessaire.

d'ordre m en ξ ,

$$\sigma_m(x, \xi) = \sum_{\alpha, |\alpha|=m} a_\alpha(x) \xi^\alpha$$

s'appelle le symbole "principal" de P .¹²⁰ On remarquera que

$$\sigma(x, \xi) = e^{-i\xi \cdot x} P(e^{i\xi \cdot x})$$

(il est facile de le vérifier).¹²¹ Si les coefficients $a_\alpha(x)$ sont constants, $\sigma(x, \xi)$ ne dépend plus de x .

On notera que si $P = I$ est l'identité (qui est de degré 0) $\sigma_I \equiv \sigma_{I,0} = 1$ et que si P et Q sont des ODs d'ordres respectifs m et r , leur produit PQ est d'ordre $m+r$ et que le symbole principal $\sigma_{PQ, m+r}$ est simplement le produit des symboles principaux $\sigma_{PQ, m+r} = \sigma_{P, m} \sigma_{Q, r}$.¹²² Le symbole σ_{PQ} a quant à lui une forme beaucoup plus compliquée. Si $\sigma(x, \xi)$ et $\tau(x, \xi)$ sont les symboles respectifs de P et Q , alors PQ est équivalent (par la règle de Leibniz) à l'opérateur donné par le symbole produit *non commutatif* :

$$\sigma(x, \xi) \bullet \tau(x, \xi) = \sum_{\alpha} \frac{1}{\alpha!} \frac{\partial^\alpha}{\partial \xi^\alpha} \sigma(x) D_x^\alpha \tau(x, \xi) .$$

Pour le laplacien $-\Delta$ de la variable x , $\sigma(\xi) = \sigma_2(\xi) = \|\xi\|^2$ est une forme quadratique définie positive, d'où la qualification d'"elliptique". Pour l'équation des ondes $\frac{\partial^2 f(x,t)}{\partial t^2} = \Delta f(x,t)$, $\sigma(\tau, \xi) = \sigma_2(\tau, \xi) = -\tau^2 + \|\xi\|^2$ qui s'annule sur les ξ -sphères de rayon τ . D'où la qualification d'"hyperbolique". Pour l'équation de la chaleur $\frac{\partial f(x,t)}{\partial t} = \Delta f(x,t)$, le symbole principal est celui de $-\Delta$, $\sigma_2(\xi, \tau) = \|\xi\|^2$. Mais le symbole complet est $\sigma(\xi, \tau) = i\tau + \|\xi\|^2$ et l'opérateur est qualifié de "parabolique".

La formule (OD) admet deux généralisations tout à fait naturelles.

- (i) La première consiste à généraliser le symbole $\sigma(x, \xi)$ en une application C^∞ qui n'est plus nécessairement polynomiale en ξ et peut être une *série* (dont

120. Si la référence à P est nécessaire, on le notera $\sigma_{P, m}$.

121. Comme $D_{x_j}(e^{i\xi \cdot x}) = \xi_j e^{i\xi \cdot x}$, $P(e^{i\xi \cdot x}) = \sum_{\alpha, |\alpha| \leq m} a_\alpha(x) \xi^\alpha e^{i\xi \cdot x}$ et donc $e^{-i\xi \cdot x} P(e^{i\xi \cdot x}) = \sum_{\alpha, |\alpha| \leq m} a_\alpha(x) \xi^\alpha = \sigma(x, \xi)$.

122. C'est trivial puisque les dérivations d'ordre maximal dans PQf ne peuvent porter que sur f et pas sur les coefficients de P ou de Q .

on tiendra compte des propriétés de convergence). On obtient ainsi la notion “d’opérateur pseudo-différentiel” de Joseph Kohn et Louis Nirenberg¹²³ (OΨD)¹²⁴

$$Pf(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\xi \cdot (x-y)} \sigma(x, \xi) f(y) dy d\xi$$

et encore plus généralement

$$Pf(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\Phi(x,y,\xi)} \sigma(x, \xi) f(y) dy d\xi$$

où la phase $\Phi(x, y, \xi)$ n’est plus forcément une onde plane. Cette généralisation permet de développer de nombreux calculs et en particulier d’inverser sous certaines conditions des OD en utilisant le fait que les inverses de polynômes peuvent (en dehors des 0) se développer en séries (comme $\frac{1}{1-\xi} = \sum_{n \geq 0} \xi^n$ si $|\xi| < 1$).

(ii) La seconde consiste à introduire des opérateurs

$$Pf(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\Phi(x,y,\xi)} \sigma(x, y, \xi) f(y) dy d\xi$$

où le symbole $\sigma(x, y, \xi)$ dépend aussi de la variable y . On les appelle des “opérateurs intégraux de Fourier” (OIF).

Sur des bases établies par Alberto Calderón et Antoni Zygmund en 1957, ces généralisations furent développées dans les années 1960 par Lars Hörmander, Louis Nirenberg et Joseph Kohn puis, parmi d’autres, par Elias Stein, Gerald Folland, Linda Rothschild, Richard Beals, Peter Greiner, Viktor Maslov, Bernard Helffer ou Johannes Sjöstrand.

L’utilisation rigoureuse de ces généralisations soulève un grand nombre de problèmes techniques d’analyse fonctionnelle que nous ne pouvons pas aborder ici. De nombreux outils d’analyse sont mobilisés, des espaces fonctionnels comme les espaces de Sobolev, des méthodes de convergence et de calcul de bornes, etc. Le lecteur intéressé pourra se référer aux multiples ouvrages spécialisés sur ce sujet. Pour notre part, nous nous référerons en particulier à l’ouvrage classique [501] de Mikhail Shubin, au cours [273] de Mark Joshi et au “panorama” d’Olivier Lablée [309]. Citons aussi les traités de référence *Théorie des perturbations et méthodes asymptotiques* de Viktor Maslov et *Semi-classical Analysis* [238] de Victor Guillemin et Shlomo

123. Pour une introduction aux travaux de Louis Nirenberg (1925-2020, prix Abel 2015), le lecteur pourra consulter la note de Robert Kohn [301]. Nirenberg a aussi travaillé sur les EDP semi-linéaires et non linéaires. Pour avoir l’idée d’un exemple, le lecteur pourra consulter ses travaux avec Henri Berestycki sur la propagation de fronts d’onde dans des cylindres infinis (cf. [47]).

124. Cf. l’article de référence [300].

Sternberg et le non moins classique *Harmonic Analysis in Phase Space* de Gerald Folland [187].

Toutefois, tout en restant très élémentaire, nous allons essayer de donner un aperçu des outils théoriques et des méthodes d'analyse utilisés. Notons d'emblée que les symboles $\sigma(x, \xi)$ peuvent être interprétés comme des fonctions sur le fibré cotangent $T^*\mathbb{R}^n$ (et plus généralement T^*M si l'on travaille sur une variété M) et donc comme des *hamiltoniens*. La géométrie symplectique de l'espace des phases T^*M intervient ainsi en arrière fond des problèmes d'analyse de résolution d'EDP, ce qui ouvre à ce que l'on appelle "l'analyse microlocale" qui n'est pas seulement locale dans la base \mathbb{R}^n au voisinage d'un point x mais également locale dans les fibres $T_x^*\mathbb{R}^n$ au voisinage d'un covecteur ξ . C'est tout un nouveau domaine à explorer qui s'est ainsi ouvert.

Dans la théorie des OΨD il y a d'abord tout un ensemble d'hypothèses que l'on fait en général sur les P, σ, f, u dans l'expression $Pf = u$ avec P de symbole σ . Citons en quelques unes :

- (a) Les contraintes classiques sur le symbole $\sigma(x, \xi)$ sont qu'il doit être C^∞ en x et ξ et à croissance "tempérée" en ξ pour $\|\xi\| \rightarrow \infty$. Cela signifie que l'on doit éviter des croissances trop fortes comme les croissances exponentielles. On demande également que $\sigma(x, \xi)$ soit à variation "lente" en x . Ces deux conditions s'expriment par le fait qu'il existe des majorants

$$\begin{cases} |\partial_\xi^\alpha \sigma(x, \xi)| \leq C (1 + \|\xi\|)^{m-|\alpha|} \\ |\partial_x^\alpha \sigma(x, \xi)| \leq C' (1 + \|\xi\|)^{m'} \end{cases}$$

pour tout α .¹²⁵ C'est nécessaire pour contrôler les (semi-)normes de la situation dans les espaces fonctionnels qu'on a besoin d'utiliser.

- (b) Si l'on veut pouvoir garder l'idée qu'un OΨD puisse être d'ordre m , il faut changer de définition par rapport à celle des OD. On utilise les conditions ci-dessus pour définir m par la contrainte (toujours satisfaites par les OD)

$$|\partial_\xi^\alpha \partial_x^\beta \sigma(x, \xi)| \leq C (1 + \|\xi\|)^{m-|\alpha|}$$

pour tous α et β et tout x dans un compact K avec une constante dépendant de α, β et K , m étant minimal. Comme on peut faire entrer dans la constante le maximum de $|\partial_\xi^\alpha \partial_x^\beta \sigma(x, \xi)|$ sur $K \times \Gamma$ où Γ est un compact de l'espace des ξ , on voit que la contrainte contrôle l'ordre de croissance en ξ de σ et de ses dérivées

125. Certains auteurs utilisent $\sqrt{1 + \|\xi\|^2}$ au lieu de $(1 + \|\xi\|)$. C'est équivalent pour $\|\xi\|$ assez grand. On notera que $(1 + \|\xi\|)^p$ est un polynôme de degré p en $\|\xi\|$ avec pour coefficients les coefficients du binôme et qu'on peut donc majorer la norme de tout polynôme de degré p en ξ à coefficients constants par des $C(1 + \|\xi\|)^p$. Il en va de même si les coefficients d'un tel polynôme dépendent de x mais restent bornés.

pour $\xi \rightarrow \infty$. Si $m > 0$, c'est une croissance analogue à celle d'un polynôme d'ordre m en ξ avec une croissance en $(1 + \|\xi\|)^{m-|\alpha|}$ pour $m - |\alpha| \geq 0$ et une décroissance en $\frac{1}{(1+\|\xi\|)^{-m+|\alpha|}}$ pour $m - |\alpha| < 0$. On voit que si $\sigma(x, \xi)$ est de support *compact* en ξ alors σ est d'ordre $-\infty$. On note traditionnellement S^m la classe des symboles C^∞ d'ordre m . On a évidemment $S^{-\infty} = \bigcap_m S^m$. On montre que $\sigma(x, \xi) \in S^m$ si et seulement si les $a_\alpha(x)$ sont *bornés* ainsi que toutes leurs dérivées. Si $\sigma \in S^m$ et $\tau \in S^n$ le produit $\sigma\tau \in S^{m+n}$. Si $\sigma \in S^m$ on a aussi $\partial_\xi^\alpha \partial_x^\beta \sigma(x, \xi) \in S^{m-|\alpha|}$.

- (c) On peut renforcer cette définition dans les cas *classiques* et utiliser le fait que si l'OD classique P est d'ordre m , $\sigma(x, \xi)$ est une somme de termes

$$\sigma_{m-j}(x, \xi) = \sum_{\alpha, |\alpha|=m-j} a_\alpha(x) \xi^\alpha,$$

$j = 0, \dots, m$, *homogènes* de degré $m - j$ en ξ au sens où

$$\sigma_{m-j}(x, \lambda\xi) = \lambda^{m-j} \sigma_{m-j}(x, \xi)$$

pour $\lambda > 0$. Cela peut se généraliser en définissant la classe $O\Psi D_{cl}^m$ des OΨD d'ordre m "classiques" par la formule ?? où $\sigma(x, \xi)$ est une somme infinie de $\sigma_{m-j}(x, \xi) C^\infty$ pour $\xi \neq 0$ et homogènes de degré $m - j$.¹²⁶

- (d) On suppose en général que les OΨD satisfont une condition technique d'avoir un support "propre".¹²⁷ Si K est un compact de $\mathbb{R}^n_{(x)}$ il existe un compact associé K' tel que pour toute distribution $f(x)$ de support dans K $Pf(x)$ soit de support dans K' et si $f(x)$ s'annule sur K' alors $Pf(x)$ s'annule sur K .¹²⁸ Sous cette condition, on obtient des opérateurs continus qui préservent les fonctions C^∞ à support compact et plus généralement les fonctions C^∞ . En fait ces opérateurs sont *pseudo-locaux* au sens où ils ne créent pas de nouvelles singularités et restreignent les supports singuliers, le support singulier de Pf étant inclus dans celui de f .

Plusieurs lignes théoriques de l'analyse se croisent autour de ces questions.

1. Il y a d'abord la transformée de Fourier $f(x) \rightsquigarrow \widehat{f}(\xi)$ qui, d'après le théorème de Parseval-Plancherel est (aux facteurs de renormalisation près) une isométrie

126. On ne peut pas imposer aux $\sigma_{m-j}(x, \xi)$ d'être C^∞ pour $\xi = 0$ car on veut pouvoir avoir des ξ^α d'exposant < 0 .

127. Une application continue $f : X \rightarrow Y$ entre espaces topologiques est dite "propre" si l'image réciproque $f^{-1}(K)$ de tout compact de Y est un compact de X (cela signifie intuitivement qu'il n'y a pas de fibre de f non compacte).

128. Rappelons que le support d'une fonction sur un espace topologique est la fermeture de l'ensemble où elle ne s'annule pas, que le support d'une distribution est le complémentaire du plus grand ouvert sur lequel elle est nulle et que le support singulier d'une distribution est le complémentaire du plus grand ouvert sur lequel elle est C^∞ .

entre le Hilbert $L^2\left(\mathbb{R}_{(x)}^n\right)$ des $f(x)$ et le Hilbert $L^2\left(\mathbb{R}_{(\xi)}^n\right)$ des $g(\xi)$. Cet isomorphisme se restreint à un automorphisme topologique du sous-espace dense qu'est l'espace de Schwartz $\mathcal{S}(\mathbb{R}^n)$ des $f(x) \in C^\infty$ qui sont à décroissance rapide ainsi que toutes leurs dérivées.¹²⁹ Et par dualité de $\mathcal{S}(\mathbb{R}^n)$ avec l'espace des distributions tempérées $\mathcal{S}'(\mathbb{R}^n)$,¹³⁰ la transformée de Fourier définie par $\langle \widehat{T}, f \rangle = \langle T, \widehat{f} \rangle$ pour tout $f \in \mathcal{S}(\mathbb{R}^n)$ est aussi un automorphisme de $\mathcal{S}'(\mathbb{R}^n)$. Tout cela est une conséquence des relations de dualité entre la décroissance à l'infini et la régularité pour les f et leurs transformées de Fourier : si f est C^∞ à décroissance rapide à l'infini ainsi que toutes ses dérivées ($f \in \mathcal{S}(\mathbb{R}^n)$), alors c'est aussi le cas pour \widehat{f} . Comme nous l'avons dit, nous y reviendrons plus en détail dans la section 2.1.1 du chapitre 17 et expliquerons alors mieux la dualité $\widehat{\frac{\partial}{\partial x_j} f}(x)(\xi) = i\xi_j \widehat{f}(\xi)$.

2. Le fait que le groupe à 1-paramètre d'opérateurs unitaires de $L^2(\mathbb{R}, dx)$, $U(t) = e^{-itD} = e^{-t\partial_x}$ de générateur infinitésimal $\mathfrak{t} = -\partial_x = -iD$ associé à l'opérateur auto-adjoint $D = -i\partial_x$ (cf. la section 7.5 sur le théorème spectral de Stone) est l'opérateur de translation $U(t)f(x) = f(x-t)$. Nous y reviendrons à la section 1.3 du chapitre 17.
3. Les formalismes de la quantification de la mécanique classique. Nous les avons déjà rencontrés à la section 5.4 du chapitre 5 à propos du groupe de Heisenberg en mécanique quantique et nous les avons détaillés dans la section précédente 8.
4. Les intégrales oscillantes de forme générale $\int e^{i\varphi(x,\alpha)} a(x,\alpha) d\alpha$.
5. Les OIF.

Une fois introduite la notion d'OΨD l'idée de base pour *inverser* un OD classique $Pf = u$ est de chercher un OΨD Q tel que $PQ = QP = I$. Il faut donc que $\sigma_{PQ,0} = 1 = \sigma_{P,m}\sigma_{Q,-m}$. L'idée est de partir de la représentation de u pour donner

129. Cf. plus haut la section 5. Rappelons que cela signifie que, pour tout multi-indice α et tout polynôme $p(x)$ de degré aussi grand que l'on veut, $p(x) \frac{\partial^\alpha}{\partial x^\alpha} f(x)$ est borné à l'infini, i.e. $\|p(x) \frac{\partial^\alpha}{\partial x^\alpha} f\|_\infty < \infty$.

130. Un résultat important dit qu'une distribution T sur \mathbb{R}^n est tempérée si et seulement si $T = \frac{\partial^\alpha}{\partial x^\alpha} \left((1 + \|x\|^2)^N g(x) \right)$ pour un $N \in \mathbb{N}$ et une $g(x)$ continue bornée. Comme g est seulement continue, elle peut être très pathologique au niveau des dérivées au sens des distributions.

un sens à la formule

$$u(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\xi \cdot (x-y)} u(y) dy d\xi$$

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\xi \cdot (x-y)} \frac{1}{\sigma(x, \xi)} u(y) dy d\xi .$$

On voit alors immédiatement qu'une propriété clé est de savoir si $\sigma(x, \xi)$ admet ou non des zéros. Si $\sigma(x, \xi) \neq 0$ partout si $\xi \neq 0$, P est dit *elliptique* et dans ce cas la théorie des OΨD peut être bien développée (avec beaucoup de technicités d'analyse fonctionnelle). Sinon P est dit *hyperbolique* et il faut enrichir la théorie en introduisant en plus les OIF.

Remarque. Une limite de la définition de Kohn-Nirenberg des OΨD est que la non-commutativité des opérateurs s'accorde mal avec la commutativité des produits et des convolutions des fonctions. Un grand avantage du calcul de Weyl évoqué à la section 8.7.1 est de résoudre ce problème. \square

9.2. Intégrales oscillantes

On voit que dans tous ces formalismes de représentations fonctionnelles interviennent constamment des intégrales oscillantes de forme générale

$$\Psi(x) = \int e^{i\varphi(x, \alpha)} a(x, \alpha) d\alpha$$

qui généralisent les transformées de Fourier. On peut en faire la théorie générale pour des α qui ne sont pas forcément des ξ dans $T_x^* \mathbb{R}^n$ et peuvent appartenir à un \mathbb{R}^p . Pour que ces intégrales aient un sens il faut certaines conditions sur la phase $\varphi(x, \alpha)$ et le symbole $a(x, \alpha)$. Ces contraintes seront très fortes si l'on veut que $\Psi(x)$ ait de bonnes propriétés de régularité mais on peut les affaiblir en demandant que $\Psi(x)$ soit simplement une *distribution*, c'est-à-dire que l'on puisse définir

$$\langle \Psi, g \rangle = \int \Psi(x) g(x) dx = \int \int e^{i\varphi(x, \alpha)} a(x, \alpha) g(x) d\alpha dx$$

pour une classe convenable de fonctions test g , comme les fonctions C^∞ à support compact (distributions générales) ou les fonctions C^∞ à décroissance rapide ainsi que toutes leurs dérivées (distributions tempérées). Les conditions sur le symbole $a(x, \alpha)$ sont typiquement $a(x, \alpha) \in S^m$. On montre que si $m + k < -p$ ($p =$ la dimension du \mathbb{R}^p des α) l'intégrale oscillante $\Psi(x)$ est une fonction de classe C^k . Celle sur la phase $\varphi(x, \alpha)$ est en général qu'elle soit C^∞ (en dehors de l'origine), sans point critique total (x_c, α_c) sur les *deux* multi-variables et éventuellement qu'elle possède des propriétés d'homogénéité en α comme $\varphi(x, \lambda\alpha) = \lambda\varphi(x, \alpha)$.

En revanche $\Psi(x)$ est singulière en x_0 s'il existe une valeur α_0 de α telle que $a(x_0, \alpha_0) \neq 0$ et que α_0 soit un point critique de $\varphi_{x_0}(\alpha) = \varphi(x_0, \alpha)$ (i.e. $\partial_\alpha \varphi(x, \alpha) =$

0 en (x_0, α_0)). Si la condition d'homogénéité est satisfaite alors l'ensemble critique C_φ des points critiques des $\varphi_x(\alpha)$ est *conique*. Le *lieu singulier* S_φ de φ (considérée comme phase) est la projection de C_φ sur $\mathbb{R}_{(x)}^n$.

On montre que le *support singulier* de la distribution $\Psi(x)$ est inclus dans S_φ et que donc $\Psi(x)$ est C^∞ en dehors de S_φ c'est-à-dire sur le lieu régulier de φ . On voit également que si $a(x, \xi)$ s'annule sur un voisinage conique de C_φ alors $\Psi(x)$ est une fonction C^∞ .

Prenons l'exemple de la phase $(x - y) \cdot \xi$. Le produit du sous-espace $x = y$ par $\xi = 0$ est une sous variété complètement critique. Comme cela est interdit par définition pour les phases on suppose $\xi \neq 0$. On voit alors immédiatement que si $\xi \neq 0$, $\partial_\xi((x - y) \cdot \xi) = (x - y)$ est nul si et seulement si $x = y$ et que dans ce cas $(x - y) \cdot \xi = 0$. Les intégrales oscillantes $\Psi(x) = \int e^{i(x-y) \cdot \xi} a(x, \xi) d\xi$ ne peuvent être singulières que sur la diagonale $x = y$.

9.3. Opérateurs elliptiques

Si P est elliptique, classique et de support propre et si $\sigma \in S^m$ alors $\sigma^{-1} \in S^{-m}$. Pour inverser P on part alors d'un OΨD d'ordre $-m$, Q_0 , de symbole principal $\sigma_{Q_0, -m} = \sigma_{P, m}^{-1}$. On a $\sigma_{PQ_0, 0} = \sigma_{I, 0}$ et donc $PQ_0 - I$ est d'ordre -1 . En itérant cette construction, on construit un OΨD Q qui est un inverse (à droite) approximatif de P , la différence $PQ - I$ étant un OΨD d'ordre $-\infty$ (et donc ~ 0). On écrit formellement $PQ_0 - I = R$, c'est-à-dire $PQ_0 = I + R$ et en développant formellement $\frac{I}{I+R}$ en série on obtient

$$\begin{aligned} PQ_0(I - R + R^2 - R^3 + \dots) &\approx I \\ P(Q_0(I - R + R^2 - R^3 + \dots)) &\approx I \\ Q_0(I - R + R^2 - R^3 + \dots) &\approx P^{-1} \end{aligned}$$

$Q = Q_0(I - R + R^2 - R^3 + \dots) \approx P^{-1}$ s'appelle une *parametrix* de P . Elle donne une approximation $\pi(x)$ de la solution fondamentale $Pf = \delta$ (la distribution de Dirac à l'origine) qui, lorsqu'on la connaît, permet de résoudre l'EDP $Pf = u$ par convolution.¹³¹ Elle est solution de $P\pi(x) = \delta(x) + w(x)$ avec $w(x)$ une fonction C^∞ à support compact contenu dans un petit voisinage de l'origine. Elle est en général beaucoup plus facile à calculer que la solution fondamentale et fournit de nombreuses informations sur elle.

Il existe un résultat fondamental sur les opérateurs elliptiques.

Lemme de régularité elliptique de Weyl. Si P est un OΨD d'ordre m , elliptique, classique et de support propre, alors si $Pf(x)$ (f une distribution) est

131. Nous reviendrons sur ce concept dès la section 2 du chapitre 17 sur la diffusion et les états cohérents.

C^∞ , c'est que f est déjà C^∞ et, plus généralement, les support singuliers de f et Pf sont les mêmes. \diamond

Nous approfondirons de façon technique les opérateurs elliptiques dans le chapitre suivant (17) consacré à l'analyse harmonique et aux processus de diffusion. Nous exposerons l'équation de diffusion et le noyau de la chaleur pour la diffusion euclidienne (section 9.1) et la diffusion sous-riemannienne dans nos deux modèles de référence \mathbb{V}_J et \mathbb{V}_S (sections 10.1.1 et 10.1.5). Nous verrons à quel point le calcul des solutions fondamentales est compliqué.

9.4. Opérateurs hyperboliques

Si P est hyperbolique le symbole principal $\sigma_{P,m}$ possède des zéros réels. L'exemple de base reste celui de l'équation des ondes. L'équation $\sigma_{P,m} = 0$ définit la "variété caractéristique" $\Sigma(P)$ de P et l'on cherche alors à inverser P en dehors de $\Sigma(P)$.

Nous avons vu que les solutions $f(x)$ des $Pf = u$ doivent être considérées comme des distributions. Il est par conséquent important de relier les singularités de f avec celles des Pf . On est amené à appeler "front d'onde" de f l'intersection $WF(f)$ des $\Sigma(P)$ de tous les P qui *régularisent* f i.e. qui sont tels que Pf soit C^∞ . Il est inclus dans

$$\mathbb{R}_{(x)}^n \times (\mathbb{R}_{(\xi)}^n - \{0\}) = T^*\mathbb{R}^n - \{\text{section } 0\} .$$

Sa projection sur $\mathbb{R}_{(\xi)}^n - \{0\}$ est le support singulier de $f(x)$. On définit aussi le front d'onde $WF(P)$ de P comme, ainsi que l'explique Joshi [273],

"the directions in which P kills all singularities as it is of order $-\infty$."

Par conséquent si P est d'ordre $-\infty$ alors $WF(P)$ est vide.

Un théorème dit que si P est d'ordre m , classique et de support propre et si f est une distribution alors

$$WF(Pf) \subset WF(P) \cap WF(f)$$

(propriété de *microlocalité* de P) et

$$WF(f) \subset WF(Pf) \cup \Sigma(P) .$$

Si P est elliptique $WF(P) = \emptyset$ et donc

$$WF(f) = WF(Pf) .$$

Si Pf est C^∞ , par définition de $\Sigma(P)$,

$$WF(Pf) \subset WF(f) \subset \Sigma(P) .$$

Un théorème fondamental d'Hörmander (1970) dit alors que $WF(f) - WF(Pf)$ est une réunion d'intégrales du système hamiltonien associé à σ_m . Autrement dit on a une situation analogue à celle de l'optique géométrique.

9.5. Opérateurs intégraux de Fourier (OIF)

Pour les opérateurs hyperboliques, ce sont plutôt les opérateurs intégraux de Fourier (OIF) qui sont utilisés. Ils sont de la forme

$$Pf(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\Phi(x,y,\xi)} \sigma(x,y,\xi) f(y) dy d\xi$$

où la phase $\Phi(x,y,\xi)$ n'est pas forcément une onde plane et où le symbole $\sigma(x,y,\xi)$ dépend aussi de la variable y . Autrement dit, ce sont des OΨD généralisés. P transforme les $f \in \mathcal{C}_\downarrow^\infty$ en distributions Pf .

Leur *noyau* (dit noyau de Schwartz) est l'intégrale oscillante

$$K(x,y) = \int_{\mathbb{R}^n} e^{i\Phi(x,y,\xi)} \sigma(x,y,\xi) d\xi$$

qui est C^∞ en dehors du lieu critique C_Φ de la phase Φ . Par conséquent si $\sigma(x,y,\xi)$ est nul sur un voisinage de C_Φ le noyau est C^∞ . Si P est à support propre, i.e. si les projections du support de K sur $\mathbb{R}_{(x)}^n$ et $\mathbb{R}_{(y)}^n$ sont propres (c'est le cas si le support du noyau est propre), alors P transforme les $f \in \mathcal{C}_\downarrow^\infty$ en $Pf \in \mathcal{C}_\downarrow^\infty$ et aussi les $f \in \mathcal{C}^\infty$ en $Pf \in \mathcal{C}^\infty$.

9.6. Analyse microlocale

En utilisant des symboles $\sigma(x,\xi)$ qui dépendent non seulement de la position x mais aussi du moment (de la "direction") ξ on peut étudier le comportement local de l'opérateur au voisinage d'un point *et* d'une direction. Cette localisation en quelque sorte du second degré est qualifiée de "microlocale". L'analyse microlocale a été introduite par Mikio Sato, Lars Hörmander et Viktor Maslov à la fin des années 1969.

Par exemple la micro-ellipticité d'un OΨD classique d'ordre m signifie que

$$\sigma_m(x_0, \xi_0) \neq 0$$

et que donc, localement en x_0 , P est elliptique dans la direction ξ_0 . C'est équivalent au fait qu'il existe un OΨD classique Q d'ordre $-m$ tel que

$$(x_0, \xi_0) \notin WF(PQ - I) \text{ et } (x_0, \xi_0) \notin WF(QP - I) .$$

9.7. Opérateurs pseudo-différentiels sur les variétés de Heisenberg

Tout naturellement, les spécialistes des groupes de Lie et de géométrie sous-riemannienne ont cherché à généraliser la théorie des OΨD à ces structures géométriques. On peut citer en particulier, après les travaux pionniers d'Elias Stein, l'ouvrage de référence de Michael Taylor en 1986 [524] ainsi que celui de 1988 de Richard Beals

et Peter Greiner [35] qui traite le cas des variétés de Heisenberg au sens de la section 8 du chapitre 6.

Considérons une variété de Heisenberg (M, \mathcal{K}) (M de dimension n et \mathcal{K} un champ d'hyperplans de dimension $n - 1$). On considère des opérateurs différentiels, par exemple des sous-laplaciens du type $P = \sum_{j=1}^{j=n-1} X_j^2$ + termes d'ordre 1 et 0, où les X_j sont $n - 1$ champs de vecteurs sur \mathcal{K} indépendants. L'opérateur P et le champ \mathcal{K} s'interdéfinissent. Pour simplifier, on considérera seulement les P d'ordre 2, $P = \sum_{j=1}^{j=n-1} X_j^2$, sans termes de degré 1 ou 0. La théorie peut aller du cas complètement intégrable des champs commutant jusqu'aux divers type de champs non commutants qui engendrent TM .

Il faut trouver un équivalent des symboles principaux alors que les espaces tangents $T_a M$ sont maintenant des groupes de Carnot de niveau 2, $\mathfrak{N}_a M$, obtenus par nilpotentisation en $a \in M$. La généralisation est la suivante (cf. l'Introduction du Beals-Greiner [34] et le mémoire [452] de Raphaël Ponge qui approfondit les travaux précurseurs de Stein, Beals, Greiner et Taylor).

Le fibré tangent gradué $\mathfrak{N}M$ s'écrit $\mathcal{K} \oplus TM/\mathcal{K}$, TM/\mathcal{K} pouvant être vu comme un fibré transverse à \mathcal{K} si l'on a par exemple une structure de contact (champ de Reeb) ou une structure métrique (champ orthogonal à \mathcal{K}). Dans la suite les X_0 représenteront un élément de TM/\mathcal{K} ou un élément de TM transverse à \mathcal{K} . Le crochet de Lie induit une forme, dite de Levi¹³²,

$$L : \mathcal{K} \times \mathcal{K} \rightarrow TM/\mathcal{K} ,$$

donnée par

$$L(X, Y) = [X, Y] \text{ modulo } \mathcal{K}.$$

Si e_j , $j = 1, \dots, n - 1$ est une base de \mathcal{K} et e_0 un générateur de TM/\mathcal{K} , L est la matrice antisymétrique $(n - 1) \times (n - 1)$ $L = (L_{jk})$ définie par

$$[e_j, e_k] = L_{jk} e_0 .$$

Elle est nulle dans le cas intégrable. Lorsqu'elle est non nulle, son rang (nécessairement pair puisque L est une matrice réelle antisymétrique) décompose localement la géométrie de (M, \mathcal{K}) entre une partie intégrable de type \mathbb{R}^p et une partie complètement non intégrable de type Heisenberg \mathbb{H}^q avec $p + 2q + 1 = n$, le rang étant alors $2q$.

Soit $a \in M$. Les vecteurs tangents de $\mathfrak{N}_a M$ se décomposent en $X = X_0 + X_{\mathcal{K}}$ avec une composante $X_{\mathcal{K}}$ dans \mathcal{K} et une composante X_0 transverse. On définit la structure de groupe de $\mathfrak{N}_a M$ par

$$(X_0 + X_{\mathcal{K}}) \cdot (Y_0 + Y_{\mathcal{K}}) = X_0 + Y_0 + \frac{1}{2} L(X_{\mathcal{K}}, Y_{\mathcal{K}}) + X_{\mathcal{K}} + Y_{\mathcal{K}} .$$

132. Nous avons rencontré un avatar de cette forme dans un contexte plus compliqué dans la section 3.4 du chapitre 3.

On peut identifier $\mathfrak{N}_a M$ à son algèbre de Lie $\mathfrak{n}_a M$ avec le crochet de Lie

$$[X_0 + X_{\mathcal{K}}, Y_0 + Y_{\mathcal{K}}] = L(X_{\mathcal{K}}, Y_{\mathcal{K}})$$

en prenant l'identité pour fonction exponentielle et, par abus de notation, en identifiant $X \in \mathfrak{n}_a M$ à $\exp(X) \in \mathfrak{N}_a M$. La loi de groupe ne fait alors que traduire la formule de Campbell-Hausdorff

$$\exp(X) \exp(Y) = \exp\left(X + Y + \frac{1}{2}[X, Y]\right).$$

Comme la nilpotentisation de (M, \mathcal{K}) en a , $\mathfrak{N}_a M$, est l'analogie de l'espace tangent $T_a M$ on cherche à appliquer aux opérateurs différentiels sur (M, \mathcal{K}) la méthode classique de "gel des coefficients", c'est-à-dire d'approximation locale d'un opérateur différentiel P en $a \in \mathbb{R}^n$ par l'opérateur à coefficients *constants* qui a pour coefficients la valeur de ceux de P en a . Comme la constance signifie en fait l'*invariance* par les translations de \mathbb{R}^n , on va approximer $P = \sum_{j=1}^{j=n-1} X_j^2$ sur (M, \mathcal{K}) en a par le champ invariant par translations de $\mathfrak{N}_a M$ qui vient de la valeur de sa dérivée en a comme élément de $\mathfrak{n}_a M$. Pour le calculer prenons une carte locale $U \subset \mathbb{R}^n$ de M au voisinage de a , une base composée des e_j , $j = 1, \dots, n-1$ (base de \mathcal{K}) et de e_0 (générateur de TM/\mathcal{K}) et les coordonnées locales associées $(x_j) = (x_0, x_{\mathcal{K}} = (x_1, \dots, x_{n-1}))$ avec $x_j(a) = 0$. La loi de groupe de $\mathfrak{N}_a M$ s'écrit (v^t est le transposé du vecteur v)

$$\begin{aligned} (x_0, x_{\mathcal{K}}) \cdot (y_0, y_{\mathcal{K}}) &= \left(x_0 + y_0 + \frac{1}{2}L_{jk}x_j y_k, x_{\mathcal{K}} + y_{\mathcal{K}}\right) \\ &= \left(x_0 + y_0 + \frac{1}{2}x_{\mathcal{K}}^t L y_{\mathcal{K}}, x_{\mathcal{K}} + y_{\mathcal{K}}\right). \end{aligned}$$

On peut s'exercer à revérifier à la main que c'est bien une loi de groupe.¹³³

133. Son élément neutre est évidemment $(x_0 = 0, x_{\mathcal{K}} = 0)$ et l'inverse de $(x_0, x_{\mathcal{K}})$ est $(-x_0 + \frac{1}{2}x_{\mathcal{K}}^t L x_{\mathcal{K}}, -x_{\mathcal{K}})$. La loi est bien associative car

$$\begin{aligned} ((x_0, x_{\mathcal{K}}) \cdot (y_0, y_{\mathcal{K}})) \cdot (z_0, z_{\mathcal{K}}) &= \left(x_0 + y_0 + \frac{1}{2}x_{\mathcal{K}}^t L y_{\mathcal{K}}, x_{\mathcal{K}} + y_{\mathcal{K}}\right) \cdot (z_0, z_{\mathcal{K}}) \\ &= \left((x_0 + y_0) + z_0 + \frac{1}{2}x_{\mathcal{K}}^t L y_{\mathcal{K}} + \frac{1}{2}(x_{\mathcal{K}}^t + y_{\mathcal{K}}^t) L z_{\mathcal{K}}, (x_{\mathcal{K}} + y_{\mathcal{K}}) + z_{\mathcal{K}}\right) \\ (x_0, x_{\mathcal{K}}) \cdot ((y_0, y_{\mathcal{K}}) \cdot (z_0, z_{\mathcal{K}})) &= (x_0, x_{\mathcal{K}}) \cdot \left(y_0 + z_0 + \frac{1}{2}y_{\mathcal{K}}^t L z_{\mathcal{K}}, y_{\mathcal{K}} + z_{\mathcal{K}}\right) \\ &= \left(x_0 + (y_0 + z_0) + \frac{1}{2}y_{\mathcal{K}}^t L z_{\mathcal{K}} + \frac{1}{2}x_{\mathcal{K}}^t L (y_{\mathcal{K}} + z_{\mathcal{K}}), x_{\mathcal{K}} + (y_{\mathcal{K}} + z_{\mathcal{K}})\right) \\ x_{\mathcal{K}}^t L y_{\mathcal{K}} + (x_{\mathcal{K}}^t + y_{\mathcal{K}}^t) L z_{\mathcal{K}} &= y_{\mathcal{K}}^t L z_{\mathcal{K}} + x_{\mathcal{K}}^t L (y_{\mathcal{K}} + z_{\mathcal{K}}) = x_{\mathcal{K}}^t L y_{\mathcal{K}} + x_{\mathcal{K}}^t L z_{\mathcal{K}} + y_{\mathcal{K}}^t L z_{\mathcal{K}}. \end{aligned}$$

La translation (à gauche) dans $\mathfrak{N}_a M$ a donc pour application linéaire tangente

$$\begin{pmatrix} 1 & \frac{1}{2} \sum_{j=1}^{n-1} L_{j1} x_j & \cdots & \frac{1}{2} \sum_{j=1}^{n-1} L_{j(n-1)} x_j \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 \cdots & 0 \\ 0 & 0 & & 1 \end{pmatrix}.$$

Les champs invariants par translation dans $\mathfrak{N}_a M$ qui traduisent les $\partial_{x_j} \in \mathfrak{n}_a M$ au point a (qui est de coordonnées $x_j(a) = 0$) sont donc (en utilisant la notation de Beals-Greiner)

$$\begin{aligned} X_0^a &= \partial_{x_0} \\ X_k^a &= \partial_{x_k} + \left(\frac{1}{2} \sum_{j=1}^{n-1} L_{jk} x_j \right) \partial_{x_0}. \end{aligned}$$

On obtient ainsi ce que Beals et Greiner ont appelé des “model operators”

$$P^a = \sum_{j=1}^{j=n-1} (X_j^a)^2$$

“à coefficients constants” sur le nilpotentisé $\mathfrak{N}_a M$ tangent à (M, \mathcal{K}) en a .

Les propriétés d’homogénéité de (M, \mathcal{K}) confèrent un poids 1 aux vecteurs de \mathcal{K} et un poids 2 aux vecteurs transverses à \mathcal{K} , autrement dit, si $X = X_0 + X_{\mathcal{K}}$,

$$\lambda X = \lambda X_{\mathcal{K}} + \lambda^2 X_0.$$

C’est dans cette propriété d’homogénéité codant la nilpotentisation de M que réside la différence principale d’avec le cas classique.

On construit les transformées de Fourier dans les fibres \mathfrak{N}_a comme dans le cas classique par $\widehat{f}(\xi) = \int_{\mathfrak{N}_a} e^{-ix \cdot \xi} f(x) dx$ avec les covecteurs $\xi \in \mathfrak{N}_a^*$ satisfaisant la même loi d’homogénéité

$$\lambda \xi = \lambda \xi_{\mathcal{K}} + \lambda^2 \xi_0.$$

On considère alors des symboles dans les fibres $\mathfrak{N}_a^* - \{0\}$, $p(\xi) \in \mathcal{C}^\infty(\mathfrak{N}_a^* - \{0\})$, qui sont homogènes de degré m au sens où $p(\lambda \xi) = \lambda^m p(\xi)$ pour $\lambda > 0$. Les opérateurs différentiels associés, qui sont par construction invariants par translation du groupe dans la fibre, c’est-à-dire “à coefficients constants”, effectuent des convolutions par ces symboles comme dans le cas classique.

Pour passer aux opérateurs différentiels sur (M, \mathcal{K}) on passe au fibré de groupes de Carnot $\mathfrak{N}M$ et on recolle ce qui se passe dans les fibres. Par exemple pour des sous-laplaciens du type $P = \sum_{j=1}^{j=n-1} X_j^2$, si l’on travaille avec les opérateurs $D_k = -i\partial_{x_k}$ on considèrera d’abord les symboles $\sigma_j(x, \xi)$ des $-iX_j$, puis des symboles homogènes

$p(x, \lambda\xi) = \lambda^m p(x, \xi)$ (où $\lambda\xi = \lambda\xi_{\mathcal{K}} + \lambda^2\xi_{\mathcal{H}}$) et on étudiera les opérateurs différentiels

$$Pf = \frac{1}{(2\pi)^n} \int e^{ix \cdot \xi} p(x, \sigma(x, \xi)) \widehat{f}(\xi) d\xi .$$

On cherchera à inverser P et à trouver une parametrix avec un développement asymptotique du symbole et du noyau.¹³⁴

Comme dans le cas classique, il y a de nombreux problèmes techniques d'analyse qui doivent être résolus pour développer une théorie correcte. L'une des principales différences avec le cas classique est que la composition des opérateurs ne correspond plus à l'opération *algébrique* de produit au niveau des symboles principaux. Elle est plus compliquée. Comme l'explique très bien Nancy Stanton qui a fait le compte-rendu de Beals-Greiner [34] pour le Bulletin de l'American Mathematical Society [513],

“The approximating group is constructed from the operator P and may vary from point to point on M . The operator P is then approximated by a left invariant operator on the group, a model operator, and this operator has a left invariant inverse which gives a good first approximation to the parametrix for P . In the classical pseudodifferential operator calculus, the principal symbol of the composition of two operators is given by the product of the two principal symbols. This is just the Fourier transform of the convolution of the two approximating translation invariant operators. In the Beals-Greiner calculus, the principal symbol of the composition is the Fourier transform of the convolution of the approximating left invariant operators. In general, this does not correspond to an algebraic operation on the symbols, so the resulting calculus is much more complicated than the classical one.”

9.8. Passage à l'analyse harmonique

Tous ces résultats profonds d'analyse fonctionnelle sur les opérateurs conduisent vers *l'analyse harmonique* générale, domaine immense dans lequel nous allons maintenant timidement nous aventurer. Nous allons voir comment la géométrie et l'algèbre de nos modèles neurogéométriques \mathbb{V}_J et \mathbb{V}_S se relie à l'analyse du signal rétinien au moyen d'*états cohérents* (profils récepteurs) et comment les représentations de ces groupes de Lie, leur analyse harmonique non commutative et leur diffusion sous-riemannienne permet de commencer à comprendre les extraordinaires propriétés du cortex visuel primaire.

134. Rappelons qu'un symbole $p(x, \xi)$ est la transformée de Fourier d'un noyau $K(x, z)$ par rapport à une de ses variables (l'autre variable fonctionnant comme paramètre) $p(x, \xi) = \int e^{-i\xi \cdot z} K(x, z) dz$.

En un mot, il s'agit d'approfondir mathématiquement la dialectique entre *la géométrie* des architectures fonctionnelles (connectivité) et *l'analyse fonctionnelle* des états cohérents (profils récepteurs).

Analyse harmonique non commutative, états cohérents et diffusion sous-riemannienne

1. Représentations de groupes

Après ce mini *vademecum* d'analyse fonctionnelle, nous allons passer à l'analyse harmonique sur des groupes nettement plus compliqués que \mathbb{S}^1 (analyse de Fourier), et en particulier non compacts et non commutatifs puisque c'est le cas de nos deux exemples de base \mathbb{V}_J et \mathbb{V}_S . Pour cela, nous devons d'abord approfondir la notion de *représentation de groupe* car il s'agit du concept clé.

1.1. Généralités

1.1.1. Un concept fondamental.

Les structures géométriques neuralemement implémentées que nous avons étudiées pour décrire les connectivités neuronales des architectures fonctionnelles sont essentiellement décrites en termes

- (i) de groupes de Lie (en particulier \mathbb{V}_J et \mathbb{V}_S) exponentiant leurs algèbres de Lie et
- (ii) d'entités différentielles, comme des champs de vecteurs tangents ou des formes différentielles, invariantes sous une action naturelle de ces groupes.

Pour comprendre comment les connectivités neuronales peuvent construire des percepts à partir d'inputs rétiniens, il faut comprendre comment ces groupes peuvent agir sur des espaces fonctionnels \mathcal{H} de fonctions modélisant d'une façon ou d'une autre des signaux d'input et des états d'activité d'aires neuronales.

Plus précisément, les profils récepteurs des neurones d'un certain type constituent une famille ψ_g de fonctions d'un espace fonctionnel approprié \mathcal{H} , famille paramétrée par les éléments g du groupe pertinent. Le traitement du signal I consiste alors en la famille des *mesures* de I fournie par les ψ_g , ces mesures étant les produits scalaires $\langle I, \psi_g \rangle_{\mathcal{H}}$ dans \mathcal{H} lorsque \mathcal{H} est un espace de Hilbert muni d'un tel produit scalaire (cf. la section 4 du chapitre 16).

On se trouve ainsi naturellement plongé dans un univers mathématique immense, celui des représentations de groupes de Lie. Une fois définie la notion, des problèmes qui se posent immédiatement sont ceux des espaces fonctionnels admissibles, de la réductibilité de leurs représentations et de leur structure, qui dépend à la fois de la structure du groupe et de celle de l'espace fonctionnel. Le domaine est si vaste qu'il

est impossible d'en présenter ici ne serait-ce que des rudiments. En accord avec notre méthode pédagogique, nous donnerons quelques définitions de base et préciserons les choses à partir des cas particuliers qui interviennent dans nos modèles.

1.1.2. Définition basique.

La définition la plus basique est la suivante. Étant donné un groupe G et un ensemble X muni d'une certaine structure, une "représentation" de G est un morphisme de groupes $\rho : G \rightarrow \text{Aut}(X)$ de G dans le groupe des automorphismes de X . Cela signifie que G est un groupe de symétrie de la structure de X . On dit aussi que X est un G -espace ou une G -structure.

L'action de G sur X est souvent une action à gauche et $\rho(g)(x)$ peut s'écrire gx avec $g'(gx) = (g'g)x$, autrement dit $\rho(g') \circ \rho(g) = \rho(g'g)$ et ρ est un isomorphisme. Mais G peut aussi agir à droite et $\rho(g)(x)$ peut s'écrire xg avec $(xg)g' = x(gg')$, autrement dit $\rho(g') \circ \rho(g) = \rho(gg')$ et ρ est un anti-isomorphisme. Souvent X est un espace de fonctions $f : M \rightarrow \mathbb{R}$ ou \mathbb{C} avec G opérant (à gauche) sur l'espace de base M . Dans ce cas, G opère à gauche sur les $f(x)$ par $(gf)(x) = f(g^{-1}x)$ car

$$\begin{aligned} (g'(gf))(x) &= gf(g'^{-1}x) = f(g^{-1}(g'^{-1}x)) = f((g^{-1}g'^{-1})x) \\ ((g'g)f)(x) &= f((g'g)^{-1}x) = f((g^{-1}g'^{-1})x) = (g'(gf))(x) . \end{aligned}$$

C'est du côté des représentations dans des espaces fonctionnels que les applications physiques sont les plus nombreuses et remontent, comme nous en avons vu des exemples au chapitre précédent (16), à la décomposition en harmoniques des sons produits par des éléments vibrants (cordes, tambours, tuyaux d'orgue, etc.) et, pour la diffusion de la chaleur, à la théorie de Fourier. Derrière ces décompositions il y a le fait que les fonctions $f \in \mathcal{F}$ que l'on considère appartiennent à un espace fonctionnel \mathcal{F} sur lequel un groupe G opère naturellement. À G se trouvent associées des fonctions très particulières φ_i (par exemple les fonctions trigonométriques décrivant les harmoniques) et l'on cherche à représenter toutes les $f \in \mathcal{F}$ comme des sommes ou des intégrales de φ_i . De telles décompositions sont évidemment d'un intérêt majeur pour trouver les f qui sont solution d'un problème (par exemple une équation différentielle) dépendant de façon naturelle de G .

Historiquement, la théorie s'est développée à partir de deux sources :

- d'un côté des groupes commutatifs très élémentaires (\mathbb{S}^1, \mathbb{R}) mais représentés dans des espaces fonctionnels pouvant être subtils (par exemple, pour les séries de Fourier traitées à la section 6 du précédent chapitre (16), l'espace des fonctions égales à la somme trigonométrique qu'est leur série de Fourier) ;
- d'un autre côté des groupes finis éventuellement compliqués mais représentés dans des espaces élémentaires (des espaces vectoriels de dimension finie).

Puis on étudia les groupes de Lie et la façon dont intervenaient dans la nature de leurs représentations leurs propriétés de nilpotence, de résolubilité, de simplicité, ou de semi-simplicité. Très vite, on se mit donc à étudier les représentations des groupes finis abéliens ($\mathbb{Z}/n\mathbb{Z}$), des groupes finis non abéliens (groupes de permutation), des groupes abéliens compacts (le cercle \mathbb{S}^1 , les tores), des groupes abéliens non compacts (\mathbb{R}^n : transformées de Fourier), des groupes compacts non abéliens (groupes de rotation semi-simples $SO(n)$, $n \geq 3$, groupes spinoriels, etc.), des groupes non abéliens non compacts (groupes résolubles des déplacements euclidiens $SE(n)$ comme produits semi-directs $\mathbb{R}^n \rtimes SO(n)$), groupe de Poincaré (résoluble), les groupes $SL(n, \mathbb{R})$ (semi-simples), le groupe de Lorentz (semi-simple), les groupes symplectiques (semi-simples), etc. Les mécaniques classique et relativiste conduisirent à approfondir les représentations des groupes de relativité fondamentaux : groupe de Galilée, groupe de Lorentz, groupe de Poincaré, ainsi que les groupes des symplectomorphismes des espaces de phases. Et la mécanique quantique conduisit également à investiguer les représentations de groupes comme le groupe de Heisenberg, les groupes de symétrie interne comme $SU(2)$, etc.

Remarque. Dans cet ouvrage nous ne pouvons pas entrer dans ce que l'on appelle la "théorie des catégories" où, dans un univers de la théorie des ensembles, on considère tous les ensembles munis de structures d'un même type (les groupes, les espaces vectoriels, les variétés différentiables, etc.) reliés par des morphismes de structure c'est-à-dire des applications compatibles aux structures. Mais il est clair que les G -espaces (à gauche) forment une catégorie dont les objets sont les G -espaces X, Y , etc. et dont les morphismes sont les applications $\varphi : X \rightarrow Y$ telles que $\varphi(gx) = g\varphi(x)$. Nous y reviendrons à la section 6.3 à propos des représentations induites des groupes de Lie. □

1.1.3. Groupes finis.

La théorie des représentations s'efforce de généraliser aux groupes généraux, en particulier aux groupes de Lie, des résultats anciens (fin XIX^e - début XX^e) démontrés pour les groupes finis par d'éminents spécialistes comme Richard Dedekind (1831-1916), Ferdinand Georg Frobenius¹ (1849-1917) et son disciple Issai Schur (1875-1941)², ou William Burnside (1852-1927). Ces groupes ainsi que leurs représentations linéaires peuvent être très compliqués mais leur structure est accessible algébriquement.³ Il n'est pas pertinent pour notre propos de résumer leur

-
1. Dont nous avons déjà longuement parlé à propos des formes différentielles et du problème de Pfaff.
 2. Schur est particulièrement connu pour ses travaux sur les représentations polynomiales de $GL(n, \mathbb{C})$. Nous rencontrerons plusieurs fois son célèbre lemme.
 3. La classification des groupes finis simples est l'un des sommets des mathématiques des années 1950-1980. Elle inclut entre autres le théorème de Feit-Thompson disant que tout groupe fini d'ordre impair est résoluble et donc, dans la mesure où un groupe simple fini non commutatif ne peut pas être résoluble, tout groupe simple fini non commutatif est d'ordre pair.

théorie (qui est à elle seule tout un univers) mais nous ferons néanmoins quelques remarques à leur sujet.

Les représentations linéaires les plus simples de G sont les morphismes de groupes $\rho : G \rightarrow GL(V)$ de G dans le groupe (de Lie) des automorphismes linéaires de V où V est un espace vectoriel de dimension finie sur \mathbb{R} ou \mathbb{C} . L'image $\rho(G)$ est ainsi un sous-groupe *fini* de $GL(V)$ (et donc de matrices une fois choisie une base de V). Les morphismes de représentations, aussi dits "entrelacements", entre $\rho_1 : G \rightarrow GL(V_1)$ et $\rho_2 : G \rightarrow GL(V_2)$ sont les morphismes d'espaces vectoriels $\theta : V_1 \rightarrow V_2$ commutant avec les ρ_i au sens où, pour tout $g \in G$, $\theta \circ \rho_1(g) = \rho_2(g) \circ \theta$, le diagramme suivant est commutatif :

$$\begin{array}{ccc} V_1 & \xrightarrow{\rho_1(g)} & V_1 \\ \theta \downarrow & & \downarrow \theta \\ V_2 & \xrightarrow{\rho_2(g)} & V_2 \end{array}$$

Un morphisme est un isomorphisme si θ est un isomorphisme.

La finitude de G simplifie énormément la théorie car on peut sans problème *moyenner* sur G . En effet, si τ est une représentation dans un vectoriel \mathcal{V} , on peut considérer l'opérateur

$$P(v) = \frac{1}{|G|} \sum_{g \in G} \tau(g)v$$

($|G|$ est le cardinal de G). C'est un projecteur sur le sous-espace \mathcal{W} de \mathcal{V} fixé par G . $P(\mathcal{V})$ est bien fixé par G parce que

$$P(g'v) = \frac{1}{|G|} \sum_{g \in G} \tau(g'g)v = P(v)$$

(car $g'g$ parcourt G). Réciproquement, P est bien surjectif sur \mathcal{W} , car si v est fixé par G , alors

$$P(v) = \frac{1}{|G|} \sum_{g \in G} v = \frac{|G|}{|G|} v = v .$$

Par ailleurs P est un projecteur car

$$\begin{aligned} P^2(v) &= P\left(\frac{1}{|G|} \sum_{g \in G} \tau(g)v\right) = \frac{1}{|G|} \sum_{g \in G} P(\tau(g)v) \\ &= \frac{1}{|G|} \sum_{g \in G} \left(\frac{1}{|G|} \sum_{g' \in G} \tau(g')\tau(g)v\right) = \frac{1}{|G|^2} \sum_{g, g' \in G} \tau(g'g)v \\ &= \frac{|G|}{|G|^2} \sum_{g \in G} \tau(g)v = P(v) . \end{aligned}$$

Si V est un \mathbb{C} -vectoriel, prenons alors pour \mathcal{V} l'espace $\mathcal{Q}(V)$ des formes hermitiennes sur V . Si $Q \in \mathcal{Q}(V)$, alors la moyenne $\overline{Q} = \frac{1}{|G|} \sum_{g \in G} \rho(g)^* Q \rho(g)$ est également définie ≥ 0 et en plus G -invariante par construction. Par rapport à \overline{Q} , la représentation devient *unitaire*, ce qui implique automatiquement qu'elle est complètement réductible. En effet, si W est un sous-espace G -invariant, son orthogonal W^\perp est également G -invariant et donc $V = W \oplus W^\perp$ est une décomposition G -invariante. V est par conséquent une somme directe orthogonale de représentations unitaires irréductibles (unirreps).⁴

On montre alors qu'il existe, à isomorphisme près, un nombre fini k de représentations irréductibles (qui sont donc des unirreps) ρ_j , $j = 1, \dots, k$, où k est le nombre de *classes de conjugaison* de G (si $g \in G$, sa classe de conjugaison est l'ensemble des hgh^{-1} pour $h \in G$). Le *dual de Pontryagin* \widehat{G} de G , c'est-à-dire l'ensemble des classes d'équivalence des représentations irréductibles de G , est donc décrit par l'ensemble des ρ_j . Soient d_j les dimensions des V_j sur lesquels les ρ_j sont définies. On montre que les d_j divisent $|G|$ et surtout que l'on a l'identité de Burnside

$$\sum_{j=1}^{j=k} d_j^2 = |G| .$$

Si G est commutatif, $hgh^{-1} = g$, les classes de conjugaison sont simplement les éléments, $k = |G|$ et les d_j sont tous $= 1$.

Ce théorème fondamental se démontre en considérant la représentation régulière (à droite) $\mathcal{R}(g')$ de G dans $L^2(G)$ l'espace des fonctions $\varphi : G \rightarrow \mathbb{C}$ muni du produit scalaire hermitien

$$\langle \varphi, \psi \rangle = \sum_{g \in G} \varphi(g) \overline{\psi(g)}$$

(la somme ne pose pas de problème puisque G est fini). La dimension de $L^2(G)$ sur \mathbb{C} est $|G|$ puisque $L^2(G) \simeq \mathbb{C}^G$. Par définition, $(\mathcal{R}(g')\varphi)(g) = \varphi(gg')$. On utilise alors le théorème disant que, pour toute représentation ρ de G , la dimension de l'espace $\text{Hom}_G(\mathcal{R}, \rho)$ des entrelacements de \mathcal{R} dans ρ est la dimension de ρ . Si donc ρ est l'une des ρ_j , on a

$$\dim(\text{Hom}_G(\mathcal{R}, \rho_j)) = d_j .$$

Mais toute représentation T de G étant complètement réductible, on a toujours $T = \sum_{j=1}^{j=k} t_j \rho_j$ les t_j étant des multiplicités et T agissant sur la somme directe correspondante des V_j qui est de dimension $\sum_{j=1}^{j=k} t_j d_j$. On montre que si $S = \sum_{j=1}^{j=k} s_j \rho_j$, alors la dimension de $\text{Hom}_G(T, S)$ est $\sum_{j=1}^{j=k} t_j s_j$. Pour $T = \mathcal{R}$, on obtient

$$\dim(\text{Hom}_G(\mathcal{R}, \rho_j)) = r_j = d_j$$

4. L'expression "représentation unitaire irréductible" est souvent abrégée en "unirrep".

et donc

$$\dim (L^2 (G)) = |G| = \sum_{j=1}^{j=k} d_j \dim (\rho_j) = \sum_{j=1}^{j=k} d_j^2 .$$

À toute représentation ρ de G est associé ce que l'on appelle un "caractère" $\chi_\rho (g)$ de G qui est la *trace* $\text{tr} (\rho (g))$ des $\rho (g) \in GL (V)$ (cf. la section 3 du chapitre 7 pour la notion de trace). Comme

$$\begin{aligned} \chi_\rho (hgh^{-1}) &= \text{tr} (\rho (hgh^{-1})) = \text{tr} (\rho (h) \rho (g) \rho (h)^{-1}) = \text{tr} (\rho (h)^{-1} \rho (h) \rho (g)) \\ &= \text{tr} (\rho (h^{-1}h) \rho (g)) = \text{tr} (\rho (e_G) \rho (g)) = \text{tr} (\rho (g)) = \chi_\rho (g) , \end{aligned}$$

(e_G étant l'élément neutre de G) les caractères sont constants sur les classes de conjugaison, ce que l'on appelle des fonctions "centrales". Leur égalité caractérise l'équivalence des représentations, autrement dit $\rho \simeq \tau$ si et seulement si $\chi_\rho = \chi_\tau$. Pour la représentation régulière (à droite) \mathcal{R} agissant sur les fonctions $\varphi : G \rightarrow \mathbb{C}$ de $L^2 (G)$ par $(\mathcal{R} (g') \varphi) (g) = \varphi (gg')$, les δ_g (l'équivalent des Dirac) définies par

$$\begin{cases} \delta_g (g) = 1 \\ \delta_g (h) = 0 \text{ si } h \neq g \end{cases}$$

forment une base orthonormée de $L^2 (G)$ et, relativement à cette base, $\mathcal{R} (g')$ agit par translations. La trace $\chi_{\mathcal{R}} (g')$ est donc le cardinal de l'ensemble des points fixes de $\mathcal{R} (g')$ et vaut $|G|$ si $g = e_G$ et 0 si $g \neq e_G$. En termes de caractères, la formule précédente

$$\dim (\text{Hom}_G (T, S)) = \sum_{j=1}^{j=k} t_j s_j$$

s'écrit simplement

$$\dim (\text{Hom}_G (T, S)) = \langle \chi_T, \chi_S \rangle_{L^2(G)} ,$$

ce qui fournit un critère d'irréductibilité : ρ est irréductible si et seulement si $\|\chi_\rho\|_{L^2(G)} = 1$.

1.1.4. Groupes de Lie.

Les représentations les plus communes et les plus simples à analyser des groupes de Lie sont les représentations dites "linéaires" où V est un espace vectoriel de dimension finie sur \mathbb{R} ou \mathbb{C} et $\rho : G \rightarrow GL (V)$ un morphisme de groupes de G dans le groupe de Lie des automorphismes linéaires de V qui est en même temps une application différentiable puisque la structure de variété différentiable fait partie de la structure de groupe de Lie. Pour cela, il est suffisant que ρ soit *continu*.⁵ De même,

5. On montre que, ρ étant un morphisme de groupes de Lie, sa continuité implique sa différentiabilité. En effet le graphe Γ de ρ dans $G \times GL (V)$ qui relève G de façon homéomorphe dans $GL (V)$ au moyen de ρ est un sous-groupe *fermé* et, par conséquent, d'après le théorème de Cartan-von Neumann évoqué aux sections 2 et 9 du chapitre 5 (montrant la réciproque du

pour une algèbre de Lie \mathcal{G} , une représentation sera un morphisme d'algèbres de Lie dans l'algèbre de Lie $\mathfrak{gl}(V)$ d'un $GL(V)$. Si $\rho : G \rightarrow GL(V)$ est une représentation de G , l'application linéaire tangente $T_e\rho$ (qui existe puisque ρ est C^∞) induit une représentation de l'algèbre de Lie \mathcal{G} de G dans $\mathfrak{gl}(V)$.

Les notions de morphisme (dit aussi "entrelacement") et d'isomorphisme (dit aussi "équivalence") de représentations $\rho : G \rightarrow GL(V)$ et $\sigma : G \rightarrow GL(W)$ s'imposent d'elles-mêmes. Il s'agit, comme plus haut pour les groupes finis, d'une application linéaire $\theta : V \rightarrow W$ qui commute avec ρ et σ , c'est-à-dire telle que, pour tout $g \in G$, on ait le diagramme commutatif $\theta \circ \rho(g) = \sigma(g) \circ \theta$

$$\begin{array}{ccc} V & \xrightarrow{\rho(g)} & V \\ \theta \downarrow & & \downarrow \theta \\ W & \xrightarrow{\sigma(g)} & W \end{array}$$

Il y a équivalence lorsque φ est un isomorphisme. Dans ce cas, puisque θ^{-1} existe, la relation de commutation devient la relation de conjugaison $\sigma(g) = \theta \circ \rho(g) \circ \theta^{-1}$.

Mais ces représentations *ne suffisent pas*, même pour les cas les plus simples comme ceux que nous avons rencontrés. En effet, dès l'avènement de la mécanique quantique où ce groupe joue un rôle crucial, Hermann Weyl avait tout de suite noté pour l'algèbre de Lie de Heisenberg définie par la relation de commutation

$$[\hat{q}, \hat{p}] = i\hbar I,$$

qu'il ne peut pas en exister de représentations de dimension finie dans une algèbre de Lie $\mathfrak{gl}(V)$. En effet la trace de $\rho([\hat{q}, \hat{p}])$ est nulle⁶ alors que la trace de $i\hbar I = i\hbar \dim(V)$. D'où la nécessité d'utiliser des représentations de dimension *infinie*. Cela soulève immédiatement toute une série de difficultés relevant de l'analyse fonctionnelle : quels espaces fonctionnels utiliser, quelles conditions imposer aux opérateurs $\rho(g)$, etc. Nous verrons que l'on travaille souvent dans des *espaces de Hilbert* avec des opérateurs *auto-adjoints* et des transformations *unitaires*.

Pour comprendre la structure des représentations (linéaires) il faut d'abord comprendre celle des représentations irréductibles qui ne sont pas des représentations sur un sous-espace vectoriel strict de V . Si l'on appelle "sous-représentation" de ρ la restriction de ρ à un sous-vectoriel W stable par ρ , cela signifie que ρ n'a pas d'autres sous-représentations que celles sur $\{0\}$ et V . Ensuite, on peut facilement faire des sommes directes

$$\rho \oplus \rho' : G \rightarrow GL(V \oplus V')$$

fait que tout sous-groupe de Lie d'un groupe de Lie est fermé), un sous-groupe *de Lie*. Cela permet de montrer que le relèvement est un difféomorphisme et comme ρ est le composé de ce difféomorphisme par la projection $G \times GL(V) \rightarrow GL(V)$ qui est C^∞ , ρ est lui-même C^∞ .

6. La trace $\text{tr}(A)$ d'une matrice étant la somme de ses éléments diagonaux, si A et B sont deux matrices on a toujours $\text{tr}(AB) = \text{tr}(BA)$.

de représentations $\rho : G \rightarrow GL(V)$ et $\rho' : G \rightarrow GL(V')$ en posant

$$(\rho \oplus \rho')(g) = \rho(g) \oplus \rho'(g) .$$

La théorie pourrait alors être considérablement simplifiée si l'on pouvait montrer que toute représentation ρ est *complètement* réductible, au sens où elle serait décomposable en somme directe de représentations irréductibles. Cela signifierait que tout sous-espace G -invariant de V aurait un supplémentaire également G -invariant. Mais ce n'est hélas pas du tout le cas ! ρ peut être indécomposable en somme directe sans être pour autant irréductible.⁷ Cela est facile à comprendre. Considérons le cas le plus trivial d'une représentation d'algèbres de Lie ρ du \mathbb{R} -vectoriel \mathbb{R} (qui est trivial comme algèbre de Lie) dans $\mathfrak{gl}(V)$. Soit $A = \rho(1)$. Comme ρ est un morphisme de \mathbb{R} -vectoriels,

$$\rho(x) = \rho(x1) = x\rho(1) = xA$$

et la décomposition de ρ va correspondre à la décomposition de A . Si le sous-vectoriel W de V est un espace propre de A de valeur propre λ , alors $\rho(x)(w) = xA(w) = \lambda xw$ pour tout $w \in W$ et W est donc la sous-représentation λxI_W . La décomposition spectrale de A , et donc les racines du polynôme caractéristique $\Pi_A(\lambda) = \det(A - \lambda I)$, jouent un rôle clé. Comme on est sur \mathbb{R} , ces racines sont soit des λ réels (les valeurs propres réelles de A) soit des complexes ζ (et vont alors par paires $(\zeta, \bar{\zeta})$ de complexes conjugués). Les premières vont donner des sous-représentations irréductibles W de dimension réelle 1 avec $A \upharpoonright_W = \lambda I_{\mathbb{R}}$ et les secondes des sous-représentations irréductibles W de dimension réelle 2 avec, en identifiant W à \mathbb{C} , $A(z) = \zeta z$.⁸ L'exemple le plus basique de ce dernier cas est celui de $\zeta = i$, A agissant sur \mathbb{R}^2 par la matrice de Pauli $i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Si l'action de A se décompose en somme directe de telles actions propres, i.e. si A est *diagonale* (sur \mathbb{C}), alors ρ est complètement réductible. Mais, si A n'est pas diagonale, les choses sont plus compliquées et il faut étudier en détail la structure interne de A .

7. Un exemple classique est la représentation τ de \mathbb{R} agissant par translations dans l'espace \mathcal{P}_k des polynômes $P(x)$ de degré $\leq k : \tau(t)P(x) = P(x+t)$. $\mathcal{P}_{k-1} \subsetneq \mathcal{P}_k$ est évidemment un sous-espace \mathbb{R} -invariant. Mais il ne possède aucun supplémentaire \mathbb{R} -invariant car le sous-espace des $\mathbb{R}x^k$ ne l'est pas.

8. Autrement dit, si $z = x + iy \in W$ et si $\zeta = re^{i\theta}$, A agit par $A \begin{pmatrix} x \\ y \end{pmatrix} = r \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$.

Dans le cas où G est *compact*, la situation est “bonne” parce que l’image $\rho(G)$ est un sous-groupe compact de $GL(V)$,⁹ ce qui implique qu’il existe une métrique euclidienne sur V qui soit $\rho(G)$ -invariante. Si W est une sous-représentation, alors l’orthogonal W^\perp est également une sous-représentation (à cause de la $\rho(G)$ -invariance de la métrique) et $V = W \oplus W^\perp$. Cela implique que ρ est complètement réductible.

Toutefois, dans les cas qui nous intéressent, G n’est pas compact.

1.2. Représentations dans des espaces de Hilbert

1.2.1. Représentations générales.

Dans de nombreux cas que nous traitons, le groupe G est un groupe de Lie de dimension finie opérant sur l’espace de Hilbert $\mathcal{H} = L^2(X, \nu)$ des fonctions de carré intégrable sur un espace X muni d’une mesure ν . Les représentations $\pi(g)$ considérées sont alors supposées *continues* et à valeurs dans l’espace $\mathcal{L}(\mathcal{H})$ des opérateurs *continus* et *bornés*.

Soit donc π une telle représentation sur $\mathcal{H} = L^2(X, \nu)$ venant d’une action de G sur X à travers la définition

$$\pi(g)f(x) = f(g^{-1}x) .$$

G agit sur X au moyen d’automorphismes de structure éléments de $\text{Aut}(X)$, par exemple des permutations si X est simplement un ensemble, des homéomorphismes si X est un espace topologique, des difféomorphismes si X est une variété différentiable, des isométries si X est un espace métrique, etc. Nous avons rencontré des translations, des rotations, des changements d’échelle, des modulations par des fréquences, mais le cadre est infiniment plus riche. D’après la définition, $\pi(g)$ est une représentation au sens où :

$$(\pi(g') \circ \pi(g))f(x) = f(g'^{-1}(g^{-1}x)) = f((gg')^{-1}x) = \pi(gg')f(x) .$$

Remarque. Si G opère à gauche sur X par gx alors G opère sur les fonctions $f(x)$ au moyen de $f(g^{-1}x)$. Mais on peut aussi utiliser l’automorphisme involutif d’inversion de G ($g \mapsto g^{-1}$) et faire opérer G sur X par $g^{-1}x$. G opère alors sur les $f(x)$ au moyen de $f(gx)$ et l’on obtient la formule plus habituelle

$$(\pi(g') \circ \pi(g))f(x) = f(g'(gx)) = f((g'g)x) = \pi(g'g)f(x) .$$

On peut aussi continuer à faire opérer G à gauche sur X par gx mais en le faisant opérer à droite sur les fonctions $f(x)$ au moyen de $f(xg)$. On aura encore la formule habituelle

$$(\pi(g') \circ \pi(g))f(x) = f(xgg') = \pi(gg')f(x) .$$

Tout cela est affaire de convention et conduit à des variantes dans les formules. \square

9. Rappelons en effet que l’image d’un compact par une application continue est compacte.

On suppose qu'il existe une certaine invariance de la mesure $d\nu(x)$ de \mathcal{H} par rapport à l'action de G :

- soit une vraie invariance définie par $d\nu(gx) = d\nu(x)$ pour tout $g \in G$,
- soit une quasi-invariance disant que $d\nu(gx) = C(g) d\nu(x)$ avec $C(g) > 0$ et dans ce cas on considère la représentation

$$\pi(g) f(x) = \sqrt{C(g^{-1})} f(g^{-1}x) .$$

Ainsi définies, les représentations $\pi(g)$ sont automatiquement unitaires, i.e.

$$\|\pi(g) f\|_{\mathcal{H}}^2 = \|f\|_{\mathcal{H}}^2 .$$

En effet, comme les g agissent par automorphismes sur X ,

$$\begin{aligned} \|\pi(g) f\|_{\mathcal{H}}^2 &= \int_X |\pi(g) f(x)|^2 d\nu(x) = \int_X C(g^{-1}) |f(g^{-1}x)|^2 d\nu(x) \\ &= \int_X C(g^{-1}) |f(g^{-1}x)|^2 \frac{d\nu(g^{-1}x)}{C(g^{-1})} = \int_X |f(g^{-1}x)|^2 d\nu(g^{-1}x) \\ &= \int_X |f(x')|^2 d\nu(x') = \|f\|_{\mathcal{H}}^2 . \end{aligned}$$

Souvent également la représentation $\pi(g)$ introduit des facteurs de phase de type $e^{i\phi(g,x)}$ comme les facteurs d'ondes planes $e^{i\omega x}$ dans les transformées de Fourier. Cela permet de tenir compte d'éléments d'analyse fréquentielle. Nous en rencontrerons plusieurs exemples.

1.2.2. *Extension des représentations.*

Supposons que l'on sache définir correctement l'espace $\mathcal{C}_\downarrow^\infty(G)$ des fonctions C^∞ sur G à décroissance rapide à l'infini.¹⁰ Si π est une représentation comme ci-dessus, elle s'étend automatiquement à une action de $\mathcal{C}_\downarrow^\infty(G)$ sur \mathcal{H} par la formule

$$\pi(\varphi) f = \int_G \varphi(g) \pi(g) f dg .$$

En effet, si $\varphi(g) \in \mathcal{C}_\downarrow^\infty(G)$, $\pi(\varphi(g))$ est bien un opérateur borné et si les $\pi(g)$ sont uniformément bornés (par exemple si π est unitaire) alors $\pi(\varphi(g))$ peut même s'étendre aux $\varphi(g) \in L^1(G)$. Cette extension importante est une représentation de

10. Si par exemple G est un groupe de Carnot (réel, de dimension n) connexe et simplement connexe (cf. la section 8.1 du chapitre 6), c'est un groupe nilpotent stratifié dont l'algèbre de Lie \mathcal{G} admet une graduation $\mathcal{G} = \bigoplus_{i=1}^{i=m} \mathcal{G}_i$ en somme directe de sous-espaces \mathcal{G}_i de dimension $n_i = \dim(\mathcal{G}_i)$ tels que $\mathcal{G}_{i+1} = [\mathcal{G}_1, \mathcal{G}_i]$ et $[\mathcal{G}_1, \mathcal{G}_m] = 0$. On peut alors identifier G à \mathcal{G} par l'application exponentielle et transporter à G des notions comme celles de polynôme ou de fonction à décroissance rapide définissables dans le vectoriel \mathcal{G} .

l'algèbre de convolution des fonctions (cf. par exemple [524]). En effet, comme la mesure de Haar dg est invariante, on a, pour $h \in G$, la formule

$$\begin{aligned} \pi(h) \circ \pi(\varphi) f &= \pi(h) \left(\int_G \varphi(g) \pi(g) f dg \right) = \int_G \varphi(g) \pi(h) \pi(g) f dg \\ &= \int_G \varphi(g) \pi(hg) f dg = \int_G \varphi(h^{-1}g') \pi(g') f dg' . \end{aligned}$$

Soit alors

$$\pi(\psi) \circ \pi(\varphi) f = \int_G \psi(g') \pi(g') \left(\int_G \varphi(g) \pi(g) f dg \right) dg' .$$

On voudrait écrire cette intégrale double comme une

$$\pi(\theta) f = \int_G \theta(g) \pi(g) f dg .$$

Pour cela il faut que $\theta(g)$ soit lui-même une intégrale. Mais

$$\pi(g') \left(\int_G \varphi(g) \pi(g) f dg \right) = \int_G \varphi(g'^{-1}g'') \pi(g'') f dg''$$

et donc

$$\pi(\psi) \circ \pi(\varphi) f = \int_G \psi(g') \left(\int_G \varphi(g'^{-1}g'') \pi(g'') f dg'' \right) dg'$$

ce qui est bien une formule de type $\int_G \theta(g'') \pi(g'') f dg''$ avec

$$\theta(g'') = \int_G \psi(g') \varphi(g'^{-1}g'') dg'$$

ce qui est exactement le produit de convolution $\psi * \varphi$.

Cela montre que si $\varphi(g) \in \mathcal{C}_\downarrow^\infty(G)$ les vecteurs $h = \pi(\varphi) f$ (pour tout $f \in \mathcal{H}$) sont ce que l'on appelle des "vecteurs C^∞ ", c'est-à-dire tels que $\pi(g)h$ soit une fonction C^∞ de G à valeurs dans \mathcal{H} .

1.3. Représentations unitaires de \mathbb{R}

1.3.1. Groupes à 1-paramètre et générateurs infinitésimaux.

Le plus simple exemple de représentation unitaire dans un Hilbert \mathcal{H} est celui d'une représentation unitaire du groupe additif de \mathbb{R} . Cela signifie qu'on se donne un groupe à 1-paramètre d'opérateurs unitaires $U(t)$ de \mathcal{H} , autrement dit

$$\left\{ \begin{array}{l} (a) U(t) \circ U(t') = U(t') \circ U(t) = U(t+t') , \\ (b) U(0) = Id_{\mathcal{H}}, U(-t) = U(t)^{-1} , \\ (c) U(t)^{-1} = U(t)^* , \\ (d) U(t)^{-1} = U(-t) = U(t)^* . \end{array} \right.$$

(b) est une propriété de représentation qui implique que le groupe soit commutatif et (c) est la propriété d'unitarité¹¹. La condition de continuité (dite de continuité forte) est que si $t_n \xrightarrow{n \rightarrow \infty} t$ dans \mathbb{R} , alors $U(t_n) f \xrightarrow{n \rightarrow \infty} U(t) f$ dans \mathcal{H} pour tout f de \mathcal{H} . Elle implique que pour $|t| \leq 1$, les $U(t)$ sont bornés en norme : il existe une borne $M \geq 1$ telle que $\|U(t)\| \leq M$ pour $|t| \leq 1$ et donc, pour tout t , $\|U(t)\| \leq Me^{K|t|}$ pour un certain K .

Un exemple typique est celui des translations $T(t) f(x) = f(x - t)$ pour $\mathcal{H} = L^2(\mathbb{R})$ le Hilbert des fonctions sur \mathbb{R} à valeurs dans \mathbb{C} qui sont de carré intégrable ($\|f\|^2 = \int_{\mathbb{R}} |f(x)|^2 dx < \infty$).¹² $T(t)$ préserve les normes puisque la mesure de Lebesgue dx est invariante par translation (mesure de Haar). La continuité forte vient du fait qu'elle est satisfaite sur des sous-espaces *denses* de \mathcal{H} comme par exemple l'espace $\mathcal{C}_\downarrow^\infty(\mathbb{R})$ des fonctions C^∞ à décroissance rapide à l'infini.

Remarque. Nous ne développerons pas ici les lemmes d'analyse permettant de montrer ce genre de résultats au moyen d'estimations bien choisies. Le lecteur intéressé trouvera toutes les informations souhaitées dans nos items bibliographiques et dans les (excellents) articles de Wikipedia. □

On peut alors exprimer un tel groupe à 1-paramètre par l'exponentielle $e^{t\mathbf{u}}$ d'un générateur infinitésimal \mathbf{u} exactement comme, dans la section 9 du chapitre 5, nous avons exprimé les éléments g d'un groupe de Lie G comme une exponentielle e^X d'un vecteur X de l'algèbre de Lie \mathcal{G} . D'ailleurs toute représentation unitaire ρ d'un groupe de Lie induira une foule de tels sous groupes $U(t) = \rho(e^{tX})$ pour $X \in \mathcal{G}$. Pour cela, on considère simplement la dérivée $\mathbf{u}f$ de $U(t) f$ en $t = 0$:

$$\mathbf{u}f = \left. \frac{d}{dt} U(t) f \right|_{t=0} = \lim_{\varepsilon \rightarrow 0} \left(\frac{U(\varepsilon) f - f}{\varepsilon} \right) .$$

On a alors $U(t) = e^{t\mathbf{u}}$ et, comme $U(t)^* = U(-t)$, le générateur infinitésimal satisfait $\mathbf{u}^* = -\mathbf{u}$. On dit qu'il est "anti-hermitien" ("skew-adjoint" en anglais). Cela équivaut au fait que $A = i\mathbf{u}$ soit *auto-adjoint* ou encore "hermitien" (i.e. $A = A^*$) puisque

$$A^* = (i\mathbf{u})^* = -i\mathbf{u}^* = i\mathbf{u} = A .$$

On écrit par conséquent

$$U(t) = e^{-itA} (\mathbf{u} = -iA) .$$

Par exemple, pour le groupe $T(t)$ des translations,

$$\mathbf{t}f = \left. \frac{d}{dt} T(t) f \right|_{t=0} = \left. \frac{d}{dt} f(x - t) \right|_{t=0} = -\frac{\partial f}{\partial x}$$

11. Rappelons que l'adjoint U^* d'un opérateur U est défini par $\langle Uf, h \rangle_{\mathcal{H}} = \langle f, U^*h \rangle_{\mathcal{H}}$ où $\langle \bullet, \bullet \rangle_{\mathcal{H}}$ est le produit scalaire de \mathcal{H} .

12. Souvent, on prend plutôt $T(t) f(x) = f(x + t)$. La forme $f(x - t)$ correspond à l'action à gauche des translations sur les fonctions et $f(x + t)$ à l'action à droite.

(noté aussi $-\partial_x f$) et l'on a $T(t) = e^{-t\partial_x}$. La possibilité d'appliquer t exige que f soit dérivable.¹³ On peut vérifier par une intégration par parties que $-\partial_x$ satisfait $(-\partial_x)^* = -(-\partial_x) = \partial_x$ pour les f dérivables s'annulant à l'infini, par exemple les $f \in \mathcal{C}_\downarrow^\infty(\mathbb{R})$. Comme

$$\int_{-\infty}^{\infty} \partial_x (f(x) \overline{h(x)}) dx = 0 = \int_{-\infty}^{\infty} (\partial_x f)(x) \overline{h(x)} dx + \int_{-\infty}^{\infty} f(x) \overline{(\partial_x h)(x)} dx$$

on a

$$\begin{aligned} \langle -\partial_x f, h \rangle &= - \int_{-\infty}^{\infty} (\partial_x f)(x) \overline{h(x)} dx = \int_{-\infty}^{\infty} f(x) \overline{(\partial_x h)(x)} dx \\ &= \langle f, \partial_x h \rangle = \langle f, (-\partial_x)^* h \rangle \end{aligned}$$

et donc

$$(-\partial_x)^* = \partial_x = -(-\partial_x) .$$

Par conséquent

$$D = -i\partial_x = \frac{1}{i}\partial_x$$

est auto-adjoint et l'on peut écrire

$$T(t) = e^{-itD} .$$

Nous avons déjà rencontré plusieurs fois cet opérateur, en particulier à la section 7.3.5 du chapitre 16.

La réciproque est vraie : si $\mathbf{u}^* = -\mathbf{u}$ est défini sur un sous-espace $\mathcal{D}(\mathbf{u})$ dense, alors il engendre un groupe à 1-paramètre $U(t)$ d'opérateurs unitaires. Cela est lié au théorème de décomposition spectrale.

Évidemment, on rencontre aussitôt un problème d'analyse fonctionnelle, à savoir que \mathbf{u} n'a pas de raison d'être partout défini. Son domaine de définition $\mathcal{D}(\mathbf{u})$ est celui où la dérivée $\frac{d}{dt}U(t)f|_{t=0}$ est bien définie. Mais on peut montrer facilement un certain nombre de propriétés.¹⁴

1. \mathbf{u} est un opérateur *fermé*, autrement dit si $f_n \in \mathcal{D}(\mathbf{u})$ est une suite dans $\mathcal{D}(\mathbf{u})$ qui converge vers f (i.e. $f_n \xrightarrow{n \rightarrow \infty} f$) dans \mathcal{H} et si $\mathbf{u}f_n$ converge vers h dans \mathcal{H} , alors $f \in \mathcal{D}(\mathbf{u})$ et $\mathbf{u}f = h$. Cela découle du fait que $\mathcal{D}(\mathbf{u})$ est stable par l'action de $U(t)$.
2. $\mathbf{u}U(t)f = \frac{d}{dt}U(t)f$.

13. Nous avons vu que lorsqu'on travaille dans un espace fonctionnel comme L^p , il faut prendre la dérivée au sens des distributions puisque les éléments de L^p ne sont définis qu'à égalité $=_{p.p.}$ près. Mais la contrainte est alors que la dérivée appartienne à L^p .

14. Cf. par exemple le beau traité de Michael Taylor *Noncommutative Harmonic Analysis* [524] que nous avons déjà cité plusieurs fois et auquel nous nous référerons encore souvent.

3. $\mathcal{D}(\mathbf{u})$ est *dense* dans \mathcal{H} . En effet, soit $f \in \mathcal{H}$. On construit une suite $f_\delta \in \mathcal{D}(\mathbf{u})$ d'éléments de $\mathcal{D}(\mathbf{u})$ convergeant vers f dans \mathcal{H} . Il suffit de prendre $f_\delta = \frac{1}{\delta} \int_0^\delta U(t) f dt$ car (cf. [524])

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \left(\frac{U(\varepsilon) f_\delta - f_\delta}{\varepsilon} \right) &= \lim_{\varepsilon \rightarrow 0} \left(\frac{U(\varepsilon) \int_0^\delta U(t) f dt - \int_0^\delta U(t) f dt}{\varepsilon \delta} \right) \\ &= \lim_{\varepsilon \rightarrow 0} \left(\frac{\int_0^\delta U(\varepsilon) U(t) f dt - \int_0^\delta U(t) f dt}{\varepsilon \delta} \right) \\ &\quad \text{par linéarité de } U(\varepsilon) \\ &= \lim_{\varepsilon \rightarrow 0} \left(\frac{\int_0^\delta U(\varepsilon + t) f dt - \int_0^\delta U(t) f dt}{\varepsilon \delta} \right) \\ &\quad \text{car } U(\varepsilon) U(t) = U(\varepsilon + t) \\ &= \lim_{\varepsilon \rightarrow 0} \left(\frac{\int_\varepsilon^{\delta+\varepsilon} U(t') f dt' - \int_0^\delta U(t) f dt}{\varepsilon \delta} \right) \\ &\quad \text{par changement de variable } t + \varepsilon \rightarrow t' \\ &= \lim_{\varepsilon \rightarrow 0} \left(\frac{\int_\delta^{\delta+\varepsilon} U(\varepsilon + t) f dt - \int_0^\varepsilon U(t) f dt}{\varepsilon \delta} \right) \\ &\quad \text{par } \int_\varepsilon^{\delta+\varepsilon} - \int_0^\delta = \int_\varepsilon^\delta + \int_\delta^{\delta+\varepsilon} - \int_0^\varepsilon - \int_\varepsilon^\delta \\ &= \frac{U(\delta) f - f}{\delta} \end{aligned}$$

et comme cette limite existe, on a bien $f_\delta \in \mathcal{D}(\mathbf{u})$ pour tout δ . Par exemple pour $\mathbf{t} = -\partial_x$ dans l'exemple des translations, $\mathcal{D}(\mathbf{t})$ contient $\mathcal{C}_\downarrow^\infty(\mathbb{R})$ qui est dense dans $L^2(\mathbb{R})$.

Si $U(t)$ est un groupe à 1-paramètre, il existe une majoration de type

$$U(t) \leq M e^{K|t|}$$

et tout λ de partie réelle $\Re(\lambda) > K$ appartient à la résolvante $R_{\mathbf{u}}$ de \mathbf{u} et

$$(\mathbf{u} - \lambda I)^{-1} f = - \int_0^{+\infty} e^{-\lambda t} U(t) f dt .$$

Il en va de même pour $\Re(\lambda) < -K$ et

$$(\mathbf{u} - \lambda I)^{-1} f = \int_{-\infty}^0 e^{-\lambda t} U(t) f dt .$$

Cette formule définit bien un opérateur borné $R_\lambda = (\mathbf{u} - \lambda I)^{-1}$ car, pour $t > 0$ et λ réel $> K$, $e^{-\lambda t} e^{Kt} = e^{(K-\lambda)t}$ est d'exposant $(K - \lambda)t$ négatif. Elle est évidente pour $f \in \mathcal{D}(\mathbf{u})$ et, comme $\mathcal{D}(\mathbf{u})$ est dense dans \mathcal{H} , se prolonge à \mathcal{H} par des arguments de continuité et de densité.

1.3.2. Représentations de \mathcal{G} et de G .

Nous venons de voir le lien entre une représentation unitaire de \mathbb{R} dans \mathcal{H} , $U(t) = e^{t\mathbf{u}} = e^{-itA}$ ($\mathbf{u} = -iA$) avec \mathbf{u} un opérateur anti-hermitien fermé de domaine de définition $\mathcal{D}(\mathbf{u})$ dense dans \mathcal{H} . Le groupe $U(t)$ est l'exponentiation de \mathbf{u} . Soit alors G un groupe de Lie de dimension n , \mathcal{G} son algèbre de Lie, X_j , $j = 1, \dots, n$, une base de \mathcal{G} , et supposons-nous donnée une représentation ρ de \mathcal{G} dans \mathcal{H} au moyen de n opérateurs anti-hermitiens $\mathbf{u}_j = \rho(X_j)$ de domaines $\mathcal{D}(\mathbf{u}_j)$ denses satisfaisant les relations de commutations de \mathcal{G} . Pour chaque X_j le groupe à un paramètre e^{tX_j} de G est représenté dans \mathcal{H} par le groupe à 1-paramètre d'opérateurs unitaires $e^{t\mathbf{u}_j} = e^{-itA_j}$. La question est de savoir si l'on peut construire ainsi, quitte à introduire des conditions et des restrictions, une représentation unitaire au moins locale de G . Le problème est délicat (cf. Taylor [524]). Pour exemplifier le type de preuves dans ce domaine, donnons-en une idée sommaire.

1. Si $X = \sum_{j=1}^{j=n} a_j X_j \in \mathcal{G}$ et si

$$\mathbf{u} = \rho(X) = \sum_{j=1}^{j=n} a_j \mathbf{u}_j,$$

il faut pouvoir définir

$$e^{t\mathbf{u}} f = e^{t\rho(X)} f = \rho(e^{tX}) f$$

pour $f \in \mathcal{H}$ par la série

$$e^{t\mathbf{u}} f = \sum_{k \geq 0} \frac{t^k}{k!} \mathbf{u}^k f,$$

au moins pour $|t|$ assez petit (par exemple $|t| \leq 1$). Pour cela, certaines conditions de convergence doivent être remplies. On considère d'abord des éléments $f \in \mathcal{H}$ particuliers, appelés *analytiques*, relativement aux opérateurs \mathbf{u}_j . Ils sont définis par la propriété

- (i) que f appartienne à tous les domaines de définition $\mathcal{D}(\mathbf{u}^\alpha)$ de produits quelconques \mathbf{u}^α de \mathbf{u}_j (α est une suite d'indices j) et
- (ii) qu'il existe deux constantes finies C_1 et C_2 telles que les normes $\|\mathbf{u}^\alpha f\|$ soient majorées avec le facteur constant C_1 par les $(C_2 |\alpha|)^{|\alpha|}$.

Cela garantit que la série $e^{\mathbf{u}} f$ est absolument convergente pour $\sum_{j=1}^{j=n} |a_j| < C$, C dépendant de C_2 . On peut donc définir $e^{\mathbf{u}} f$ (et donc les $e^{t\mathbf{u}} f$, $|t| \leq 1$) pour les f analytiques.

2. Or un théorème (difficile) de Harish-Chandra, Nelson et Gårding dit qu'il existe toujours une constante C_2 telle que les f analytiques relativement aux \mathfrak{u}_j et à C_2 soient *denses* dans \mathcal{H} . Par des arguments de densité et de continuité on peut donc dans ce cas définir $e^{\mathfrak{u}}f$ pour tout f .

3. Pour construire une représentation locale ρ_{loc} de G , on suppose que les vecteurs de base X_j appartiennent à un voisinage V_0 de 0 assez petit pour que l'exponentielle $\exp : \mathcal{G} \rightarrow G$ soit un difféomorphisme et pour contenir les $X = \sum_{j=1}^{j=n} a_j X_j$ avec $\sum_{j=1}^{j=n} |a_j| < C$. Il faut montrer que la définition $e^{\mathfrak{u}}f$ fournit bien une représentation locale de G , i.e. que l'on a bien

$$\rho_{\text{loc}}(gg') f = \rho_{\text{loc}}(g) \rho_{\text{loc}}(g') f .$$

Cela se démontre au moyen de la formule de Baker-Campbell-Hausdorff.

4. On peut ensuite *globaliser* à G , du moins si G est *simplement connexe*. On part du fait que tout $g \in G$ peut s'écrire comme un produit fini $\prod_{\ell=1}^{\ell=N} g_\ell$ de g_ℓ proches de 0, $g_\ell \in V_0 = \exp(V_0)$. En augmentant N on peut prendre les g_ℓ aussi proches de 0 que l'on veut. On peut donc poser

$$\rho(g) f = \left(\prod_{\ell=1}^{\ell=N} \rho_{\text{loc}}(g_\ell) \right) f .$$

5. Tout le problème, fort délicat, est alors de montrer que le $\rho(g) f$ ainsi obtenu est *indépendant* de la décomposition $g = \prod_{\ell=1}^{\ell=N} g_\ell$. Il suffit bien sûr de le montrer pour l'élément neutre e de G en tenant compte de $\rho(g) = Id_{\mathcal{H}}$. Si $e = \prod_{\ell=1}^{\ell=N} g_\ell$, on veut donc montrer que

$$\left(\prod_{\ell=1}^{\ell=N} \rho_{\text{loc}}(g_\ell) \right) f = f .$$

Pour ce faire, on connecte la boucle des éléments e, g_ℓ, e de façon à en faire un lacet et, comme G est *simplement connexe*, on peut le *contracter* sur e par une homotopie paramétrée par $s \in [0, 1]$ de façon à obtenir des arcs $g_\ell(s)$ joignant en étoile les g_ℓ à e ($g_\ell(0) = g_\ell$ et $g_\ell(1) = e$) avec $\prod_{\ell=1}^{\ell=N} g_\ell(s) = e$ pour tout $s \in [0, 1]$. Pour s assez proche de 1, le lacet sera assez petit pour que l'on ait

$$\left(\prod_{\ell=1}^{\ell=N} \rho_{\text{loc}}(g_\ell(s)) \right) f = f$$

car, localement sur V_0 , ρ_{loc} est bien une représentation et

$$\begin{aligned} \left(\prod_{\ell=1}^{\ell=N} \rho_{\text{loc}}(g_\ell(s)) \right) f &= \rho_{\text{loc}} \left(\prod_{\ell=1}^{\ell=N} g_\ell(s) \right) f = \rho_{\text{loc}}(e) f = Id_{\mathcal{H}} f \\ &= f . \end{aligned}$$

6. Le pas crucial de la démonstration consiste alors à montrer que l'on peut faire cette construction *analytiquement* à la fois côté G et côté \mathcal{H} , avec des lacets, des contractions et des arcs $g_\ell(s)$ analytiques tels que $\left(\prod_{\ell=1}^{\ell=N} \rho_{\text{loc}}(g_\ell(s))\right) f = f$ soit une fonction analytique de s pour f analytique. Par *prolongement analytique*, on peut alors en déduire que $\left(\prod_{\ell=1}^{\ell=N} \rho_{\text{loc}}(g_\ell(s))\right) f = f$ pour *tout* $s \in [0, 1]$, donc pour $s = 0$, et que par conséquent $\left(\prod_{\ell=1}^{\ell=N} \rho_{\text{loc}}(g_\ell)\right) f = f$ pour les f analytiques. Comme ces derniers sont denses, on en déduit, encore une fois avec des arguments de densité et de continuité, que $\left(\prod_{\ell=1}^{\ell=N} \rho_{\text{loc}}(g_\ell)\right) f = f$ pour tout f .

7. Enfin on montre que $\rho(g)$ ainsi défini est bien une représentation, i.e. $\rho(gg') = \rho(g)\rho(g')$.

2. Représentations de carré intégrable et états cohérents

Les représentations que nous avons rencontrées dans le Vol I, en particulier avec la théorie des ondelettes à la section 5.3. du chapitre 3 relèvent de ce qu'on appelle la théorie des *états cohérents* (cf. par exemple Kalisa, Torresani [277]). Retournons-y un instant en nous focalisant sur ce concept unificateur.

2.1. Retour sur les exemples du Volume I

2.1.1. Transformée de Fourier.

Nous avons traité des séries de Fourier dans la section 6 du précédent chapitre (16). Nous allons maintenant revenir sur les transformées de Fourier de fonctions définies non plus sur le compact \mathbb{S}^1 mais sur \mathbb{R} . Nous en avons déjà un peu parlé à propos des opérateurs pseudo-différentiels à la section 9 de ce chapitre 16. Comme \mathbb{R} n'est pas compact, les conditions aux limites peuvent inclure des conditions "à l'infini".

On considère donc des fonctions $f(x) \in \mathcal{H} = L^2(\mathbb{R})$ de \mathbb{R} à valeurs dans \mathbb{C} qui sont de carré intégrable, i.e. telles que $\int_{\mathbb{R}} |f(x)|^2 dx < \infty$, ce qui est une contrainte beaucoup plus forte que $\int_{\mathbb{S}^1} |f(x)|^2 dx < \infty$. Ce sont en quelque sorte des signaux d'énergie finie. $\mathbb{R}_{(x)}$ (l'indice x spécifie la variable) est muni de la mesure de Lebesgue dx . La transformée de Fourier des $f(x) \in L^2(\mathbb{R})$ correspond, comme pour les séries de Fourier, à leur analyse "fréquentielle". Cela signifie que l'on se donne des fonctions correspondant à des fréquences pures, autrement dit des *ondes planes* $\psi_\omega(x) = e^{i\omega x}$ et, pour chaque fréquence ω , on "mesure" le "poids" (le coefficient) de ψ_ω dans f au moyen du produit scalaire

$$\langle f, \psi_\omega \rangle_{\mathcal{H}} = \int_{\mathbb{R}} f(x) \psi_\omega^*(x) dx = \int_{\mathbb{R}} f(x) e^{-i\omega x} dx .$$

Cette mesure définit le coefficient de Fourier $\widehat{f}(\omega)$ de f pour la fréquence ω et la fonction $\widehat{f}(\omega)$ à valeurs complexes définie sur le $\mathbb{R}_{(\omega)}$ de variable ω est la transformée de Fourier. On peut utiliser la notation parlante $\mathbb{R}_{(\omega)} = \widehat{\mathbb{R}_{(x)}}$ si besoin est.

Remarque. Pour définir $\widehat{f}(\omega)$, il suffit de supposer que f est intégrable. Mais pour que toutes les propriétés explicitées ci-dessous soit bien définies il faut supposer que f est de carré intégrable. On commence par les $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ et on généralise (c'est un peu technique) à $L^2(\mathbb{R})$. \square

L'analyse fréquentielle est fondamentale parce que, comme nous l'avons souligné dans le Vol I, les ondes planes sont des *fonctions propres* de tout filtrage linéaire \mathcal{L} sur les $f(x)$. En effet, si $\theta(x) = \mathcal{L}(\delta(x))$ est la réponse impulsionnelle de \mathcal{L} ($\delta(x)$ = distribution de Dirac), on a, puisque \mathcal{L} est linéaire et $f(x) = \int_{\mathbb{R}} f(x-u)\delta(u)du = f * \delta$, $\mathcal{L}(f(x)) = \int_{\mathbb{R}} f(x-u)\mathcal{L}(\delta(u))du$, c'est-à-dire la convolution $\mathcal{L}(f(x)) = \int_{\mathbb{R}} f(x-u)\theta(u)du = f * \theta$. Donc $\mathcal{L}(e^{i\omega x}) = \int_{\mathbb{R}} e^{i\omega(x-u)}\theta(u)du$ qui est $e^{i\omega x}$ avec pour facteur la transformée de Fourier de la réponse impulsionnelle $\int_{\mathbb{R}} e^{-i\omega u}\theta(u)du = \widehat{\theta}(\omega)$.

Remarque. Pour mieux faire apparaître le fait que la transformée de Fourier est une *dualité*, rappelons que $f(x)$ est en fait une représentation "positionnelle" de la fonction *abstraite* $f \in \mathcal{H}$, à savoir une représentation qui *évalue* f en x . Cela peut s'exprimer au moyen des distributions de Dirac $\delta_x(u)$ (attribuant le poids (ou la "masse") 1 au point x et le poids 0 partout ailleurs) par l'intégrale

$$f(x) = \int_{\mathbb{R}} f(u)\delta_x(u)du .$$

L'évaluation $f(x)$ est aussi une "mesure" de f , à savoir précisément celle de la valeur de f au point x . \square

Comme nous l'avons déjà noté plusieurs fois, dans le Vol I et dans les chapitres précédents, nous n'approfondissons pas le problème technique d'analyse fonctionnelle dû à ce que ni les ondes planes $e^{i\omega x}$ (en particulier la fonction constante 1) ni leurs transformées de Fourier (distributions de Dirac) n'appartiennent à $L^2(\mathbb{R})$. Il est lié au fait que les distributions δ_x sont parfaitement définies en position et totalement indéfinies en fréquence parce que toutes les ondes planes y sont présentes avec le même poids mais se détruisent par interférence destructives¹⁵ sauf au point x , alors que, dualement, les ondes planes sont parfaitement définies en fréquence et totalement indéfinies en position.

Souvent, on écrit les ondes planes $e^{2\pi i\omega x}$ en remplaçant la fréquence par la pulsation ou fréquence angulaire, ce qui conduit à introduire des facteurs dans les formules.

15. Rappelons (cf. chapitre 16, section 6) que le phénomène d'interférences destructives est subtil et fondamental. Nous avons vu plusieurs fois qu'il intervient dans les "intégrales oscillantes" (pour des précisions, cf. aussi [413] et sa bibliographie) et établit un lien entre structures géométriques et phénomènes ondulatoires à travers le principe dit de la "phase stationnaire".

Souvent également, on normalise la transformée de Fourier des $f(x) \in L^2(\mathbb{R}_{(x)})$ en utilisant les ondes planes $\varphi_\omega(x) = \frac{1}{\sqrt{2\pi}}e^{i\omega x}$. L'analyse fréquentielle devient alors

$$\mathcal{F}_f(\omega) = \widehat{f}(\omega) = \langle f, \varphi_\omega \rangle_{\mathcal{H}} = \int_{\mathbb{R}} f(x) \varphi_\omega^* dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx \in L^2(\mathbb{R}_\omega)$$

la synthèse s'effectuant par la transformée inverse

$$f(x) = \int_{\mathbb{R}} \mathcal{F}_f(\omega) \varphi_\omega(x) d\omega = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathcal{F}_f(\omega) e^{i\omega x} d\omega .$$

Si $\widehat{f}(\omega)$ est définie au moyen des ψ_ω (i.e. sans le facteur $\frac{1}{\sqrt{2\pi}}$) alors dans la transformée inverse le facteur $\frac{1}{\sqrt{2\pi}}$ doit être remplacé par $\frac{1}{2\pi}$.

La formule d'inversion vient des *relations d'orthogonalité* entre ondes planes. Si ω et ω' sont deux fréquences, le produit scalaire (au sens large, éventuellement infini) a pour valeur, puisque $\int_{\mathbb{R}} e^{i\alpha x} dx = k\delta(\alpha)$,

$$\langle e^{i\omega x}, e^{i\omega' x} \rangle = \int_{\mathbb{R}} e^{i\omega x} e^{-i\omega' x} dx = \int_{\mathbb{R}} e^{i(\omega - \omega')x} dx = k\delta(\omega - \omega')$$

et est donc $= 0$ si $\omega \neq \omega'$ (orthogonalité) et $= \infty$ si $\omega = \omega'$ (l'orthogonalité ne peut donc pas se prolonger en orthonormalité puisque les ondes planes sont de norme infinie). À cause de ces relations, on a (en vérifiant que l'on peut inverser l'ordre des intégrations)

$$\begin{aligned} \widehat{\widehat{f}}(x) &= k' \int_{\mathbb{R}_\omega} k'' \left(\int_{\mathbb{R}_u} f(u) e^{-i\omega u} du \right) e^{i\omega x} d\omega = k' k'' \int_{\mathbb{R}_u} \left(\int_{\mathbb{R}_\omega} e^{i\omega(x-u)} d\omega \right) f(u) du \\ &= k' k'' \int_{\mathbb{R}_u} 2\pi \delta(x-u) f(u) du = 2\pi k' k'' f(x) \end{aligned}$$

les facteurs de normalisation étant tels que $2\pi k' k'' = 1$ (le choix symétrique ci-dessus est $k' = k'' = \frac{1}{\sqrt{2\pi}}$).

On montre alors que si $f(x) \in L^2(\mathbb{R}_{(x)})$ on a $\widehat{f}(\omega) \in L^2(\mathbb{R}_{(\omega)})$ et que l'application $f \mapsto \widehat{f}$ est (aux facteurs de renormalisation près) une *isométrie* entre ces deux espaces de Hilbert. C'est le *théorème de Parseval-Plancherel*. Notons que $\widehat{f}(\omega) \in L^2(\mathbb{R}_{(\omega)})$ s'interprète comme $\int_{\mathbb{R}_\omega} |\widehat{f}(\omega)|^2 d\omega < \infty$, autrement dit comme

$$\int_{\mathbb{R}_\omega} |\langle f, \psi_\omega \rangle_{\mathcal{H}}|^2 d\omega < \infty .$$

Jusqu'ici, la structure algébrique de $\mathbb{R}_{(x)}$ n'est pas intervenue. Or $G = (\mathbb{R}_{(x)}, +)$ est un groupe additif (autrement dit abélien) et sa mesure dx est une mesure de Haar (i.e. invariante par translations). Cette structure joue en fait un rôle essentiel. Pour le voir, revenons aux ondes planes $\psi_\omega(x) = e^{i\omega x}$. Comme fonctions de $\mathbb{R}_{(x)}$ dans \mathbb{C} ,

ce sont des morphismes du groupe additif $G = (\mathbb{R}_x, +)$ dans le groupe multiplicatif des nombres complexes de module 1. En effet,

$$\psi_\omega(x + x') = e^{i\omega(x+x')} = e^{i\omega x} e^{i\omega x'} = \psi_\omega(x) \psi_\omega(x') .$$

Les ondes planes sont des *caractères* du groupe G paramétrés par les ω . Mais l'espace $\mathbb{R}_{(\omega)}$ des paramètres est lui-même muni d'une structure de groupe $\widehat{G} = (\mathbb{R}_{(\omega)}, +)$ et ce groupe agit multiplicativement sur les caractères $\psi_\omega(x)$ par

$$\psi_{\omega+\omega'}(x) = e^{i(\omega+\omega')x} = e^{i\omega x} e^{i\omega' x} = \psi_\omega(x) \psi_{\omega'}(x) .$$

La transformée de Fourier des $f(x) \in \mathcal{H} = L^2(\mathbb{R}_{(x)})$ correspond à cette dualité entre les groupes additifs $G = (\mathbb{R}_{(x)}, +)$ et $\widehat{G} = (\mathbb{R}_{(\omega)}, +)$. Les caractères ψ_ω agissent multiplicativement sur les fonctions $h(x) \in L^2(\mathbb{R}_{(x)})$ par

$$\pi(\omega) h(x) = \psi_\omega(x) h(x) = e^{i\omega x} h(x)$$

et l'on obtient ainsi une représentation de $\widehat{G} = (\mathbb{R}_{(\omega)}, +)$ dans \mathcal{H} . Le fait que $\widehat{f}(\omega) \in L^2(\mathbb{R}_{(\omega)})$, i.e.

$$\int_{\mathbb{R}_\omega} |\langle f, \psi_\omega \rangle_{\mathcal{H}}|^2 d\omega < \infty ,$$

s'exprime en disant (cf. plus bas) que la représentation $\pi(\omega)$ est "de carré intégrable".

La synthèse de $f(x)$ s'interprète alors de la façon suivante :

- (i) On considère la fonction constante de base $\varphi_0(x) = \frac{1}{\sqrt{2\pi}}$ (en laissant de côté le fait que $\varphi_0(x) \notin \mathcal{H}$).
- (ii) On fait opérer sur cette fonction la représentation π . On obtient ainsi la famille des ondes planes

$$\pi(\omega) \varphi_0(x) = \frac{1}{\sqrt{2\pi}} e^{i\omega x} = \varphi_\omega(x) .$$

- (iii) On considère les mesures

$$\mathcal{F}_f(\omega) = \langle f, \varphi_\omega \rangle_{\mathcal{H}} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx$$

donnant les coefficients de Fourier. Ces coefficients (ne dépendant pas de x) sont des éléments de $L^2(\mathbb{R}_{(\omega)}) = L^2(\widehat{G})$, cette affirmation ayant un sens car \widehat{G} est muni d'une mesure de Haar invariante, à savoir la mesure de Lebesgue $d\omega$.

- (iv) Enfin, on effectue la synthèse

$$f(x) = \int_{\widehat{G}} \mathcal{F}_f(\omega) \varphi_\omega(x) d\omega = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathcal{F}_f(\omega) e^{i\omega x} d\omega .$$

Cet exemple fondateur d'analyse harmonique a été généralisé de façon extraordinaire à un immense ensemble de situations et a engendré l'analyse harmonique sur des classes entières de groupes (y compris non commutatifs) et d'espaces homogènes quotients de groupes par des sous-groupes non normaux. Nous allons en voir plusieurs exemples. Toutefois, avant que d'y venir faisons quelques remarques.

1. \widehat{G} s'appelle le *dual de Pontryagin* du groupe G . La bonne façon de le comprendre est que c'est *l'ensemble des unirreps* de G . En effet les unirreps ρ de $G = (\mathbb{R}_{(x)}, +)$ doivent agir de façon unitaire sur un \mathbb{C} -espace vectoriel de dimension 1 et donc de la forme $\mathbb{C}e$ où e est un élément de base. Les $\rho(x)$ dans la base $\{e\}$ sont des nombres complexes et l'unitarité impose qu'ils soient de module 1, i.e. soient des éléments de \mathbb{S}^1 . Comme il faut que ρ soit un morphisme du groupe G dans le groupe multiplicatif de \mathbb{S}^1 , ρ doit être un caractère et il existe par conséquent un $\omega \in \mathbb{R}_{(\omega)}$ tel que $\rho(x) = e^{i\omega x} = \psi_\omega(x)$. Ce dual \widehat{G} est muni d'une mesure, la mesure de Plancherel, invariante sous l'action de G (nous y reviendrons à propos d'autres exemples).
2. Le fait que \widehat{G} soit lui aussi un groupe est lié au fait que le groupe G est *abélien*. Un théorème général montre en effet que les unirreps des groupes abéliens constituent un groupe, la mesure invariante étant une mesure de Haar.
3. L'isomorphisme entre G et \widehat{G} est exceptionnel. Même dans les cas abéliens les plus simples ce n'est pas le cas. Si l'on considère par exemple les fonctions périodiques, c'est-à-dire les fonctions $f(\theta)$ sur \mathbb{S}^1 , les transformées de Fourier deviennent des séries de Fourier et les ω sont des entiers relatifs. Autrement dit le dual de Pontryagin de \mathbb{S}^1 est le groupe discret \mathbb{Z} . Un théorème général dit d'ailleurs que le dual de Pontryagin d'un groupe *compact* est nécessairement *discret*.
4. Dans le cas de la transformée de Fourier, il existe de nombreuses relations remarquables entre certaines opérations sur les fonctions et des opérations duales sur leurs transformées de Fourier. Les plus importantes sont (formulées dans la variante des $\psi_\omega(x)$, i.e. $k' = 1$ et $k'' = \frac{1}{2\pi}$) :

- (a) La dualité entre produit de convolution et produit normal :

$$\widehat{(f * g)}(x)(\omega) = \widehat{f}(\omega) \cdot \widehat{g}(\omega)$$

et

$$\widehat{f(x) \cdot g(x)}(\omega) = \frac{1}{2\pi} \widehat{(f * g)}(\omega).$$

- (b) La dualité entre dérivation et multiplication (si de bonnes conditions sont vérifiées, comme f dérivable et f et f' intégrables) :

$$\widehat{f'(x)}(\omega) = i\omega \cdot \widehat{f}(\omega)$$

et (si par exemple $x.f(x)$ est intégrable)

$$\widehat{f'}(\omega)(x) = -ix.f(x) .$$

Cela signifie que les opérateurs $D = \frac{1}{i} \frac{d(\bullet)}{dx} = -i \frac{d(\bullet)}{dx}$ (cf. plus haut section 1.3.1) et $\omega.(\bullet)$ sont duaux de Fourier et que si $\Pi(X)$ est un polynôme

$$\mathcal{F}\Pi(D) = \Pi(\omega) \mathcal{F} .$$

Cela explique ce que nous avons vu à la section 5.4 du chapitre 5 à propos du groupe de Heisenberg en mécanique quantique. Dans la représentation $\psi(q) \in L^2(\mathbb{R}, \mathbb{C})$ des fonctions d'ondes, l'opérateur de position \widehat{q} agit par multiplication par $q : \widehat{q}(\psi(q)) = q.\psi(q)$. Pour l'opérateur moment il en va de même, et dans la représentation $\widetilde{\psi}(p)$ de $\psi : \widehat{p}(\widetilde{\psi}(p)) = p.\widetilde{\psi}(p)$. Mais $\widetilde{\psi}(p)$ doit être proportionnelle à $\widehat{\psi}(p) : \widetilde{\psi}(p) = \alpha \widehat{\psi}(p)$. Donc, dans la représentation $\psi(q)$, $\widehat{p}(\widetilde{\psi}(p)) = \alpha p.\widehat{\psi}(p)$ devient

$$\alpha \frac{1}{i} \frac{\partial \psi(q)}{\partial q} = -i\alpha \frac{\partial \psi(q)}{\partial q} ,$$

ce qui est bien la formule quantique avec $\alpha = \hbar$.

- (c) La dualité entre translations et modulations par des facteurs de phase : $f(x+a)$ $\widehat{f}(\omega) = e^{i\omega a} \widehat{f}(\omega)$ et $e^{ix\alpha} f(x)$ $\widehat{f}(\omega - \alpha)$.
- (d) La dualité entre dilatations et contractions : $f(\lambda x)$ $\widehat{f}(\omega) = \frac{1}{\lambda} \widehat{f}(\frac{\omega}{\lambda})$ pour $\lambda > 0$ un facteur d'échelle.
- (e) Il existe tout un ensemble de relations entre les propriétés de localisation et de régularité des f et celles des \widehat{f} . Si par exemple, f est à support compact K , il faut que dans l'intégrale donnant la synthèse de f à partir des coefficients $\widehat{f}(\omega)$ des ondes planes, les interférences destructives annulent f en dehors de K . Pour cela il faut une infinité de coefficients et le support de \widehat{f} n'est pas borné. Cela se généralise aux comportements à décroissance rapide des f à l'infini. En fait si f est C^∞ à décroissance rapide à l'infini, \widehat{f} l'est aussi.
- (f) Par ailleurs on voit que les propriétés de parité de f ont des conséquences automatiques sur les \widehat{f} car

$$\widehat{f}(\omega) = \int_0^\infty (f(x)e^{-i\omega x} + f(-x)e^{i\omega x}) dx .$$

Si par exemple f est paire, alors

$$\widehat{f}(\omega) = 2 \int_0^\infty f(x) \cos(\omega x) dx$$

est également paire. Si de plus f est réelle (resp. imaginaire pure), alors \widehat{f} est également réelle (resp. imaginaire pure). De même, si f est impaire alors

$$\widehat{f}(\omega) = -2i \int_0^\infty f(x) \sin(\omega x) dx$$

est également impaire. Si de plus f est réelle (resp. imaginaire pure), alors \widehat{f} est imaginaire pure (resp. réelle).

- (g) Ces propriétés remarquables sont intimement liées au fait que la transformée de Fourier \mathcal{F} commute avec la représentation régulière \mathcal{R} de $G = (\mathbb{R}_{(x)}, +)$ qui agit sur les $f(x)$ par les translations

$$\mathcal{R}(a) f(x) = f(x + a) .$$

En fait, \mathcal{F} fournit une représentation spectrale de $\mathcal{R}(a)$ et l'entrelace avec des opérateurs de simple multiplication. En effet

$$\mathcal{F}^{-1} \mathcal{R}(a) \mathcal{F}(\varphi(\omega)) = \mathcal{F}^{-1} \widehat{\varphi}(x + a) = e^{i\omega a} \varphi(\omega) .$$

- (h) Ces propriétés expliquent aussi l'extraordinaire utilité de la transformée de Fourier pour résoudre des EDP linéaires. Prenons l'exemple standard de l'équation de la chaleur ($t \geq 0$)

$$\frac{\partial}{\partial t} f(t, x) = \Delta f$$

(le laplacien Δ est ici simplement $\frac{\partial^2}{\partial x^2}$) avec la condition initiale $f(0, x) = \delta(x)$ caractérisant la "solution élémentaire" aussi dite "solution fondamentale". Par transformée de Fourier sur x on la transforme en l'équation très simple

$$\frac{\partial}{\partial t} \widehat{f}(t, \omega) = -|\omega|^2 \widehat{f}(t, \omega)$$

avec $\widehat{f}(0, \omega) = \frac{1}{\sqrt{2\pi}}$ dont la solution évidente est

$$\widehat{f}(t, \omega) = \frac{1}{\sqrt{2\pi}} e^{-t|\omega|^2} .$$

Par transformée de Fourier inverse, on obtient

$$f(t, x) = \frac{1}{\sqrt{2\pi}} \int e^{-t|\omega|^2} e^{i\omega x} d\omega$$

soit, cette intégrale étant classique,

$$f(t, x) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{|x|^2}{4t}} .$$

Ce “noyau de la chaleur” exprime la diffusion du Dirac δ , nous y reviendrons à la section 5.10.7.

- (i) De façon générale, de nombreux opérateurs différentiels vont se ramener par entrelacement à des synthèses d’opérateurs de multiplication sur les composantes irréductibles de la représentation régulière.
- 5. Tout ce schéma de la transformée de Fourier se généralise en dimension n en remplaçant les produits ωx par les produits scalaires $\langle \omega, x \rangle$ et les facteurs $\frac{1}{2\pi}$ par $\frac{1}{(2\pi)^n}$.
- 6. Sur le plan conceptuel, dans le cas des groupes *commutatifs*, la transformée de Fourier échange les opérateurs de multiplication par des fonctions avec des opérateurs de translations munis du produit de convolution. Dans le cas non commutatif, les translations engendrent une algèbre non commutative que la transformée de Fourier échange en multiplications par des fonctions qui ne sont plus scalaires mais à valeurs opérateurs (cf. Kirillov [290]).
- 7. Enfin, disons qu’il est particulièrement intéressant de généraliser la transformée de Fourier aux distributions. Nous avons défini rigoureusement celles-ci à la section 5 du chapitre 16. Si $T(x)$ est une distribution (par exemple sur \mathbb{R} pour simplifier) il serait naturel de définir $\widehat{T}(\omega)$ par

$$\langle \widehat{T}(\omega), \varphi(\omega) \rangle = \langle T(x), \widehat{\varphi}(x) \rangle .$$

Avec des fonctions test dans $\mathcal{C}_c^\infty(\mathbb{R})$ à support compact on se heurte très vite à des difficultés. C’est pourquoi il est beaucoup plus simple de prendre les fonctions test dans l’espace de Schwartz $\mathcal{S} = \mathcal{C}_\downarrow^\infty(\mathbb{R})$ et donc de considérer des distribution *tempérées* $T \in \mathcal{S}'$. L’avantage de \mathcal{S} est que si $f \in \mathcal{S}_{(x)}$ alors $\widehat{f} \in \mathcal{S}_{(\omega)}$. La formule d’inversion est valable dans \mathcal{S} ainsi que la formule de Parseval-Plancherel, et qui plus est, \mathcal{S} est stable par produit et par produit de convolution. Toutes les propriétés de la transformée de Fourier se généralisent alors très bien et l’on retrouve immédiatement des résultats que nous avons déjà rencontrés. Par exemple

$$\langle \widehat{\delta}(\omega), \varphi(\omega) \rangle = \langle \delta(x), \widehat{\varphi}(x) \rangle = \widehat{\varphi}(0) = \int_{\mathbb{R}} \varphi(\omega) d\omega = \langle 1, \varphi(\omega) \rangle$$

et donc $\widehat{\delta} = 1$. De même $\widehat{1} = \delta$. Si T est non seulement tempérée mais à support compact, alors \widehat{T} est régulière et définie par la densité $C^\infty \langle T(x), e^{-i\omega \cdot x} \rangle$. La formule $\widehat{\delta'(x)}(\omega) = i\omega \cdot \widehat{\delta}(\omega) = i\omega 1$ découle des définitions :

$$\begin{aligned} \langle \widehat{\delta'(x)}(\omega), \varphi(\omega) \rangle &= \langle \delta'(x), \widehat{\varphi(\omega)}(x) \rangle = - \langle \delta(x), \widehat{\varphi(\omega)}'(x) \rangle \\ &= \langle i\omega, \varphi(\omega) \rangle . \end{aligned}$$

2.1.2. *Transformée de Gabor.*

Nous avons vu également dans le Vol I, chapitre 3, section 3.4., qu'il est utile d'analyser et de synthétiser les signaux $f(x) \in \mathcal{H} = L^2(\mathbb{R})$ avec la transformée de Fourier *spatialement localisée* qu'est la transformée de Gabor. On se donne une "fenêtre" réelle $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ (donc $g = \bar{g}$) que l'on translate le long de l'axe des x et que l'on module par des ondes planes. L'état cohérent est donné par les fonctions $g_{(a,\omega)}(x) = e^{i\omega x} g(x-a)$ (atomes fréquence-espace) et les coefficients de Gabor sont les fonctions

$$G_f(a, \omega) = \langle f, g_{(a,\omega)} \rangle_{\mathcal{H}} = \int_{\mathbb{R}} f(x) e^{-i\omega x} g(x-a) dx \in L^2(\mathbb{R}^2) .$$

On montre que la synthèse s'effectue par la transformée inverse (bien définie car la "condition d'admissibilité" de g , $\|g\|^2 < \infty$, est réalisée puisque $g \in L^2(\mathbb{R})$)

$$f(x) = \frac{1}{2\pi \|g\|^2} \int_{\mathbb{R}} G_f(a, \omega) e^{i\omega x} g(x-a) dad\omega$$

et que l'on obtient ainsi une isométrie entre $L^2(\mathbb{R}_{(x)})$ et $L^2(\mathbb{R}_{(a,\omega)}^2)$.

La transformée de Gabor est paramétrée par $(a, \omega) \in \mathbb{R}_{(a,\omega)}^2$. Même si a et ω n'ont pas le même statut puisque ce sont des variables conjuguées, dans la mesure où il existe une structure de groupe additif naturelle sur $G_0 = \mathbb{R}_{(a,\omega)}^2$ avec une mesure invariante $dad\omega$ (la mesure de Lebesgue), il est néanmoins naturel de se demander s'il existe une représentation unitaire $\pi_{(a,\omega)}$ de $G_0 = \mathbb{R}_{(a,\omega)}^2$ dans \mathcal{H} satisfaisant

$$(\pi_{(a,\omega)} h)(x) = h_{(a,\omega)}(x) = e^{i\omega x} h(x-a)$$

pour tout h . Comme la condition d'admissibilité $\|g\|^2 < \infty$ est vérifiée, cette représentation serait alors de carré intégrable. On rencontre toutefois une difficulté. Supposons que $\pi_{(a,\omega)}$ soit une représentation du groupe additif G_0 . Il faudrait vérifier que $\pi_{(a',\omega')} \circ \pi_{(a,\omega)} \stackrel{?}{=} \pi_{(a'+a,\omega'+\omega)}$. Mais on a

$$\begin{aligned} (\pi_{(a',\omega')} \circ \pi_{(a,\omega)} h)(x) &= e^{i\omega' x} h'(x-a') \text{ avec } h'(x) = (\pi_{(a,\omega)} h)(x) = e^{i\omega x} h(x-a) \\ &= e^{i\omega' x} e^{i\omega(x-a')} h(x-a'-a) = e^{-i\omega a'} e^{i(\omega'+\omega)x} h(x-(a'+a)) \\ &= \boxed{?} \pi_{(a'+a,\omega'+\omega)} = e^{i(\omega'+\omega)x} h(x-(a'+a)) . \end{aligned}$$

On constate alors que tout marche bien au facteur de phase $e^{-i\omega' a}$ près qui est présent dans $\pi_{(a',\omega')} \circ \pi_{(a,\omega)}$ mais pas dans $\pi_{(a'+a,\omega'+\omega)}$. On peut résoudre cette difficulté de deux façons. Soit en considérant la "représentation" $\pi_{(a,\omega)}$ comme "projective", i.e. définie à un facteur de phase près, soit en complétant $\pi_{(a,\omega)}$ de façon à en faire une "vraie" représentation.

Dans le premier cas, la situation générale des représentations projectives d'un groupe G est que l'on ait

$$\pi_{g'g} = e^{i\varphi(g',g)} \pi_{g'} \pi_g ,$$

$\varphi(g',g)$ étant une phase définie sur $G \times G$. L'associativité du produit de G impose alors à φ la condition "cohomologique" d'être un *2-cocycle* de $G \times G$ à valeurs dans \mathbb{R} . En effet l'égalité des $\pi_{g''(g'g)}$ avec les $\pi_{(g''g')g}$:

$$\begin{aligned} \pi_{g''(g'g)} &= e^{i\varphi(g'',g'g)} \pi_{g''} \pi_{g'g} = e^{i\varphi(g'',g'g)} \pi_{g''} e^{i\varphi(g',g)} (\pi_{g'} \pi_g) \\ \pi_{(g''g')g} &= e^{i\varphi(g''g',g)} \pi_{g''g'} \pi_g = e^{i\varphi(g''g',g)} e^{i\varphi(g'',g')} (\pi_{g''} \pi_{g'}) \pi_g \end{aligned}$$

impose

$$\varphi(g'',g'g) + \varphi(g',g) = \varphi(g''g',g) + \varphi(g'',g') .$$

Dans notre cas $e^{i\omega a'}$ où

$$\varphi((a',\omega'),(a,\omega)) = \omega a' ,$$

on a bien

$$(\omega + \omega') a'' + \omega a' = \omega (a'' + a') + \omega' a'' .$$

Dans le deuxième cas, vu la forme des formules, la loi de composition des translations a doit rester l'addition. C'est donc la loi de composition des fréquences ω qu'il faut modifier, mais sans ajouter pour autant aucun facteur contenant x puisque le bilan de ces termes est correct. Il faut donc garder la composition

$$(a,\omega) \cdot (a',\omega') = (a + a', \omega + \omega') .$$

La réponse consiste à introduire une *troisième* variable $\phi \in [0, 2\pi]$ (un facteur de phase) à identifier les (a,ω) à des $(a,\omega,\phi = 0)$ et à rajouter le facteur de phase $e^{-i\phi}$ dans la représentation de façon à avoir

$$(\pi_{(a,\omega,\phi)} h)(x) = h_{(a,\omega,\phi)}(x) = e^{-i\phi} e^{i\omega x} h(x - a) .$$

On peut alors introduire sur les (a,ω,ϕ) une loi de composition de type

$$(a',\omega',\phi') \cdot (a,\omega,\phi) = (a' + a, \omega' + \omega, \phi' + \phi + \rho(a,\omega'))$$

ajoutant un facteur de phase $e^{-i\rho(a,\omega')}$ même pour $\phi = \phi' = 0$. On élargit ainsi le groupe additif initial G_0 des (a,ω) à un groupe plus grand G . Pour avoir une représentation, il faut

$$\begin{aligned}
(\pi_{(a',\omega',\phi')} \circ \pi_{(a,\omega,\phi)} h)(x) &= e^{-i\phi'} e^{i\omega'x} h'(x - a') \\
&\text{avec } h'(x) = (\pi_{(a,\omega,\phi)} h)(x) = e^{-i\phi} e^{i\omega x} h(x - a) \\
&= e^{-i\phi'} e^{i\omega'x} e^{-i\phi} e^{i\omega(x-a')} h(x - a' - a) \\
&= e^{-i(\phi'+\phi)} e^{i(\omega'+\omega)x} e^{-i\omega a'} h(x - (a' + a)) \\
&= \boxed{?} \pi_{(a'+a,\omega'+\omega,\phi'+\phi+\rho(a,\omega'))} \\
&= e^{-i(\phi'+\phi+\rho(a,\omega'))} e^{i(\omega'+\omega)x} h(x - (a' + a)) .
\end{aligned}$$

L'égalité sera vérifiée si $\rho(a, \omega') = \omega a'$. Pour $\phi = \phi' = 0$, on a

$$(a', \omega', 0) \cdot (a, \omega, 0) = (a' + a, \omega' + \omega, \omega a')$$

et il ne reste que le facteur de phase $e^{-i\omega a'}$, c'est-à-dire précisément le facteur qu'il fallait obtenir. En termes matriciels, la loi de composition est la multiplication des

matrices $\begin{pmatrix} 1 & a & \phi \\ 0 & 1 & \omega \\ 0 & 0 & 1 \end{pmatrix}$ et donc la loi de composition des (a, ω) est devenue la multiplication des $\begin{pmatrix} 1 & a & 0 \\ 0 & 1 & \omega \\ 0 & 0 & 1 \end{pmatrix}$ au lieu de la simple addition.

L'introduction du facteur de phase ϕ est nécessaire parce que les actions de a et ω respectivement par translation et modulation fréquentielle *ne commutent pas*. Soient T_a l'opérateur de translation $T_a h(x) = h(x - a)$ et F_ω l'opérateur de modulation fréquentielle $F_\omega h(x) = e^{i\omega x} h(x)$. On a

$$\begin{aligned}
(T_a \circ F_\omega) h(x) &= T_a (e^{i\omega x} h(x)) = e^{i\omega(x-a)} h(x - a) \\
(F_\omega \circ T_a) h(x) &= F_\omega (h(x - a)) = e^{i\omega x} h(x - a)
\end{aligned}$$

et donc $T_a \circ F_\omega = e^{-i\omega a} F_\omega \circ T_a$. D'où la nécessité du facteur de phase. Ce facteur "twiste" la covariance.

On remarquera que le produit que nous venons d'introduire pour que l'état cohérent de Gabor dérive d'une représentation de groupe n'est rien d'autre que celui du groupe de *Heisenberg polarisé* \mathbb{H}_{pol} introduit dans la section 3 du chapitre 5 :

$$(x', y', p') \cdot (x, y, p) = (x' + x, y' + y + p'x, p' + p) .$$

Il suffit d'identifier x à ω (moment), p à a (position) et y à ϕ . Ce n'est pas étonnant car nous avons vu avec la transformée de Fourier la dualité entre les modulations par des facteurs de phase et les translations spatiales. Nous reviendrons en détail sur ce point plusieurs fois dans la section 5.

Mais il y a pourtant deux différences. Dans l'interprétation de \mathbb{H}_{pol} comme \mathbb{V}_J , x et y sont des variables spatiales de même type (coordonnées du plan) ce qui n'est pas le cas de a et ϕ . Contrairement à y , ϕ est une phase définie modulo 2π . Ensuite, on ne paramétrise les $g_{(a,\omega)}$ que par les (a, ω) . Comme nous venons de le voir et comme nous l'avons déjà vu à la section 3.2 du chapitre 5, les $(a, \omega, 0)$ ne constituent pas un sous-groupe de G (autrement dit G_0 n'est pas un sous-groupe de G) mais sont identifiables au groupe quotient $G/Z \simeq G_0$ de G par le sous-groupe normal qu'est son centre $Z = \mathbb{S}_{(\phi)}^1$. Ce quotient isomorphe à G_0 fonctionne, nous l'avons vu, comme un "espace de phases" symplectique (avec a et ω comme variables conjuguées) que l'on "relève" dans G au moyen de la section "plate" σ_0 de la projection canonique $\pi : G \rightarrow G_0$ qui sélectionne dans chaque fibre $(a, \omega, \mathbb{S}_{(\phi)}^1)$ de G au-dessus de (a, ω) l'élément $(a, \omega, 0)$, l'image $\sigma_0(G_0)$ n'étant pas un sous-groupe.

Il s'agit d'un exemple standard de la procédure de passage au quotient que nous commenterons plus bas. La transformée de Gabor est la représentation $\pi_{(a,\omega,\phi)}$ du groupe de Heisenberg *modulo son centre* Z . Cet exemple admet des généralisations à d'autres groupes nilpotents. Le lecteur intéressé pourra se référer à l'article [225] de 2020 de Karlheinz Gröchenig.

2.1.3. Transformée en ondelettes.

Pour les ondelettes standard de Grossmann et Morlet où les fréquences ω sont remplacées par l'échelle σ , on considère l'état cohérent

$$\varphi_{(a,\sigma)}(x) = \frac{1}{\sqrt{\sigma}} \varphi_0 \left(\frac{x-a}{\sigma} \right)$$

satisfaisant la condition d'admissibilité

$$C_{\varphi_0} = \int_{\mathbb{R}^+} |\widehat{\varphi_0}(\xi)|^2 \frac{d\xi}{\xi} < \infty$$

(normalisable à $C_{\varphi_0} = 1$) et on synthétise les f de $L^2(\mathbb{R})$ au moyen de l'identité de Calderón

$$f(x) = \frac{1}{C_{\varphi_0}} \int_{\mathbb{R}_+^* \times \mathbb{R}} T_f(a, \sigma) \varphi_{(a,\sigma)}(x) \frac{d\sigma}{\sigma} \frac{da}{\sigma}$$

avec $T_f(a, \sigma) = \langle f, \varphi_{a,\sigma} \rangle$. Cette formule est bien définie si la condition d'admissibilité est satisfaite et est une isométrie de $L^2(\mathbb{R})$ sur un sous-espace fermé \mathcal{H}_{φ_0} de $L^2(\mathbb{R}^2, \frac{d\sigma}{\sigma} \frac{da}{\sigma})$. Comme dans le Vol I, nous renvoyons au grand traité de Stéphane Mallat [348] *A Wavelet Tour of Signal Processing* (1998).

Le groupe associé à cet état cohérent est le groupe *affine* A ¹⁶ produit semi-direct du groupe des translations spatiales $a \in (\mathbb{R}, +)$ par le groupe multiplicatif des

16. Traditionnellement, le groupe affine est surnommé " $ax+b$ " en référence à l'équation d'une droite affine. C'est pourquoi on note souvent dans la littérature l'échelle a et la translation b .

échelles $\sigma \in (\mathbb{R}_+^*, \times)$ avec la loi de composition¹⁷ $(a', \sigma') \cdot (a, \sigma) = (a' + \sigma'a, \sigma'\sigma)$. L'élément neutre est $(0, 1)$ et l'inverse de $g = (a, \sigma)$ est $g^{-1} = (-\frac{a}{\sigma}, \frac{1}{\sigma})$. L'action de \mathbb{R}_+^* sur $\mathbb{R}_{(a)}$ possède trois orbites : $\mathcal{O}_0 = \{0\}$, l'orbite $\mathcal{O}_+ =]0, \infty[$ (par exemple de 1) et l'orbite $\mathcal{O}_- =]-\infty, 0[$ (par exemple de -1). Une représentation matricielle commode est

$$g = (a, \sigma) = \begin{pmatrix} \sigma & a \\ 0 & 1 \end{pmatrix} .$$

Ce groupe A agit sur $\mathbb{R}_{(x)}$ par $gx = \sigma x + a$ et $g^{-1}x = \frac{1}{\sigma}x - \frac{a}{\sigma} = \frac{x-a}{\sigma}$ et, si $dx = d\nu(x)$ est la mesure de Lebesgue de $\mathbb{R}_{(x)}$, on a $d\nu(gx) = \sigma d\nu(x)$ et donc $C(g) = \sigma > 0$. La formule

$$\pi(g) f(x) = \sqrt{C(g^{-1})} f(g^{-1}x)$$

donne par conséquent

$$\pi(a, \sigma) f(x) = \frac{1}{\sqrt{\sigma}} f\left(\frac{x-a}{\sigma}\right) .$$

La mesure de Haar invariante à gauche est $\frac{d\sigma}{\sigma} \frac{da}{\sigma}$.¹⁸ A agit sur les $f(x) \in \mathcal{H} = L^2(\mathbb{R}, dx)$ par la représentation $(\pi_{(a,\sigma)} f)(x) = \frac{1}{\sqrt{\sigma}} f\left(\frac{x-a}{\sigma}\right)$. Il s'agit bien d'une représentation¹⁹ car

$$\begin{aligned} (\pi_{(a',\sigma')} \circ \pi_{(a,\sigma)} f)(x) &= \frac{1}{\sqrt{\sigma'}} \frac{1}{\sqrt{\sigma}} f\left(\frac{\frac{x-a}{\sigma} - a'}{\sigma'}\right) = \frac{1}{\sqrt{\sigma'\sigma}} f\left(\frac{x - (a + \sigma a')}{\sigma'\sigma}\right) \\ &= \pi_{(a+\sigma a', \sigma'\sigma)} f(x) \\ &= \pi_{(a,\sigma) \cdot (a',\sigma')} f(x) . \end{aligned}$$

En fait, comme l'on se restreint aux $\sigma > 0$, cette représentation est irréductible sur l'espace de Hardy \mathfrak{H}_+ des $f(x) \in \mathcal{H} = L^2(\mathbb{R}, dx)$ dont la transformée de Fourier $\widehat{f}(\xi)$ est nulle pour $\xi \leq 0$. Si l'on se permet des $\sigma \neq 0 \in \mathbb{R}^*$, la représentation est définie sur \mathcal{H} tout entier et, si la condition d'admissibilité supplémentaire

$$\int_{\mathbb{R}_+} |\widehat{\varphi}_0(-\xi)|^2 \frac{d\xi}{\xi} < \infty$$

17. On notera évidemment l'analogie formelle avec le groupe $SE(2)$ qui est lui aussi un produit semi-direct, celui des translations q et des rotations r_θ du plan, dont la loi est $(s, r_\varphi) \circ (q, r_\theta) = (s + r_\varphi(q), r_{\varphi+\theta})$ (cf. section 2.4 du chapitre 3).

18. La mesure sur la composante $\sigma > 0$ est $\frac{d\sigma}{\sigma}$ parce que les échelles agissent multiplicativement. C'est $\log(\sigma)$ qui agit additivement avec la mesure de Lebesgue. C'est donc $d\log(\sigma) = \frac{d\sigma}{\sigma}$ la bonne mesure multiplicative. Dans l'action à gauche, l'échelle σ' agit sur la position a , d'où l'élément de mesure $\frac{da}{\sigma}$. En revanche, dans l'action à droite σ n'agit pas sur a' . C'est pourquoi la mesure de Haar invariante à droite est seulement $\frac{d\sigma}{\sigma} da$: le groupe affine n'est pas unimodulaire.

19. On peut élargir \mathbb{R}_+^* à \mathbb{R}^* en considérant si besoin est des échelles négatives. L'action du groupe est alors $\pi_{(a,\sigma)} f(x) = \frac{1}{\sqrt{|\sigma|}} f\left(\frac{x-a}{\sigma}\right)$.

est satisfaite (cf. [547]), se décompose en la somme directe des deux sous-représentations définies sur les espaces de Hardy \mathfrak{H}_+ et \mathfrak{H}_- (les $f(x)$ dont la transformée de Fourier $\widehat{f}(\xi)$ est nulle pour $\xi \geq 0$).

L'action duale $\widehat{\pi}_{(a,\sigma)}$ de $\pi_{(a,\sigma)}$ sur les transformées de Fourier $\widehat{f}(\xi)$ a la forme

$$\widehat{\pi}_{(a,\sigma)}\widehat{f}(\xi) = \sqrt{\sigma}\widehat{f}(\sigma\xi)e^{-ia\xi}.$$

2.1.4. *Synthèses des méthodes.* Les méthodes temps/fréquence de Fourier et Gabor d'un côté et les méthodes temps/échelle des ondelettes de l'autre côté sont des méthodes d'analyse qui ont été considérablement développées. D'où l'intérêt de les assembler pour obtenir des méthodes très souples et très puissantes. Le lecteur intéressé pourra consulter à ce propos l'ouvrage technique [255] de Jeffrey Hogan et Joseph Lakey *Time-Frequency and Time-Scale Methods*.

2.1.5. *Profils récepteurs.*

Enfin, nous avons aussi rencontré le cas des profils récepteurs des neurones simples de V1, c'est-à-dire d'ondelettes planes $\varphi(x, y)$ spatialement localisées et directionnelles (qu'il s'agisse de Gabor ou de dérivées de gaussiennes) sur lesquelles le groupe $SE(2)$ (avec éventuellement adjonction de l'échelle) opère naturellement.

2.2. Le problème de la discrétisation

Nous allons expliquer comment les états cohérents que nous venons d'évoquer, états dépendant de façon *continue* de leurs paramètres, sont utilisés pour l'analyse harmonique en relation avec la théorie des représentations de groupes. Mais nous voudrions d'abord dire un mot sur le problème de leur *discrétisation*. En effet, pour l'effectivité des calculs, il est évidemment computationnellement utile (et même indispensable) de discrétiser les transformées utilisées.

De nombreux travaux sur la discrétisation de nos modèles neurogéométriques ont été menés à bien. Nous pouvons citer par exemple ceux de Davide Barbieri [26] sur la reconstruction d'image à partir d'une itération convergente de traitements discrétisés

“based on the reproducing kernel arising from the group structure”.

Nous résumerons aussi dans la section 11 les algorithmes sous-riemanniens semi-discrets de Jean-Paul Gauthier pour l'inpainting.

Mais nous voudrions ici commenter un article de Matilde Marcolli²⁰ et Vasiliki Liontou sur la construction de trames de Gabor discrètes directement liées à la structure de contact de notre modèle \mathbb{V}_S .

20. Matilde Marcolli est une spécialiste de géométrie non commutative et de physique mathématique. Elle s'intéresse aussi maintenant aux neurosciences.

2.2.1. *Les trames de Gabor.*

Il existe bien sûr des versions discrètes des transformées continues (Fourier, Gabor, ondelettes). Par exemple pour la transformée de Fourier la version discrète consiste à partir d'un échantillonnage du "signal" $f(x_j) = f_j$, $j = 1, \dots, N$ et à lui associer la transformée

$$\widehat{f}_k = \sum_{j=1}^{j=N} f_j e^{-\frac{2\pi i}{N} k j} .$$

Il en va de même pour les autres transformées. De nombreuses méthodes numériques ont été développées pour accélérer les calculs comme la "transformée de Fourier rapide" (FFT, "fast Fourier transform"). Mais nous voudrions ici nous focaliser plutôt sur la qualité d'approximation des transformées continues par un échantillonnage de leurs valeurs.

Évidemment, la façon la plus simple d'approximer un élément $f \in \mathcal{H}$ dans un Hilbert (séparable) \mathcal{H} est de le développer sur une base hilbertienne orthonormée $\{e_n\}$ en une série convergente $f = \sum_n c_n e_n$. Mais la base doit être adaptée au problème envisagé et il est souvent plus commode d'utiliser une base de Riesz, c'est-à-dire la transformée $\{r_n\}$ d'une base orthonormée $\{e_n\}$ par un opérateur borné inversible et d'inverse borné U . Cela est équivalent au fait que la famille $\{r_n\}$ soit complète ($\overline{\text{Span}} \{r_n\} = \mathcal{H}$) et qu'il existe des constantes $0 < A \leq B < \infty$ telles que pour toute suite $\{c_n\}$ de carré sommable ($\sum_n |c_n|^2 < \infty$) on ait

$$A \sum_n |c_n|^2 \leq \left\| \sum_n c_n r_n \right\|^2 \leq B \sum_n |c_n|^2 .$$

Il est souvent encore plus commode dans certains cas d'utiliser des systèmes générateurs "surcomplets" $\{s_n\}$ (c'est-à-dire engendrant \mathcal{H} mais avec redondance) qui sont bien adaptés au problème envisagé. Le développement $f = \sum_n c_n s_n$ existe toujours mais n'est plus unique. Dans ce cas il faut quand même contrôler la redondance de façon à ce qu'elle soit quand même minimale par rapport aux contraintes qui la motivent. Mais on voudrait aussi savoir dans quelle mesure la discrétisation d'une transformée engendre un système $\{s_n\}$ permettant de bonnes synthèses des $f \in \mathcal{H}$. C'est un problème difficile.

Par exemple pour la transformée de Gabor les atomes fréquence-espace

$$g_{(a,\omega)}(x) = e^{i\omega x} g(x - a), (a, \omega) \in \mathbb{R}_{(a,\omega)}^2$$

sont définis, nous venons de le voir, par une "fenêtre" réelle $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ et l'analyse d'un "signal" $f \in L^2(\mathbb{R})$ s'effectue par les coefficients de Gabor

$$G_f(a, \omega) = \langle f, g_{(a,\omega)} \rangle_{\mathcal{H}} = \int_{\mathbb{R}} f(x) e^{-i\omega x} g(x - a) dx \in L^2(\mathbb{R}^2) .$$

La synthèse du “signal” s’effectue quant à elle par la transformée inverse

$$f(x) = \frac{1}{2\pi \|g\|^2} \int_{\mathbb{R}} G_f(a, \omega) e^{i\omega x} g(x - a) da d\omega .$$

On obtient ainsi une isométrie entre $L^2(\mathbb{R}_{(x)})$ et $L^2(\mathbb{R}_{(a,\omega)}^2)$.

La question est donc de savoir avec quel type de fenêtre g et quelle discrétisation de $\mathbb{R}_{(a,\omega)}^2$ on peut néanmoins obtenir une synthèse correcte. Ainsi s’introduit naturellement la question subtile des *trames* ou repères de Gabor (*Gabor frames*, aussi appelées *Weyl-Heisenberg frames*) dont nous voudrions dire un mot. Le lecteur intéressé par un approfondissement pourra consulter l’ouvrage de référence *Foundations of Time-Frequency Analysis* [224] de Karlheinz Gröchenig.

Pour discrétiser $\mathbb{R}_{(a,\omega)}^2$ le plus simple est de se donner des mailles a_0 et ω_0 et de considérer le réseau $\Lambda = \Lambda_a \times \Lambda_\omega = \{m(a_0, 0) + n(0, \omega_0)\}_{m,n \in \mathbb{Z}}$. Si $\lambda \in \Lambda$, on peut noter g_λ l’atome $g_{(a,\omega)} = g_{(ma_0, n\omega_0)}$ et considérer $\sum_{\lambda \in \Lambda} |\langle f, g_\lambda \rangle|_{L^2(\mathbb{R})}^2$. La condition de définition d’une trame de Gabor est alors qu’il existe des constantes $A, B > 0$ telles que pour toute $f \in L^2(\mathbb{R})$ on ait l’encadrement uniforme

$$A \|f\|_{L^2(\mathbb{R})}^2 \leq S(f) = \sum_{\lambda \in \Lambda} |\langle f, g_\lambda \rangle|_{L^2(\mathbb{R})}^2 \leq B \|f\|_{L^2(\mathbb{R})}^2 .$$

L’idéal serait évidemment $A = B$ mais le fait qu’il existe $A, B > 0$ valides pour tout $L^2(\mathbb{R})$ reste une contrainte très forte. La contrainte

$$\sum_{\lambda \in \Lambda} |\langle f, g_\lambda \rangle|_{L^2(\mathbb{R})}^2 \leq B \|f\|_{L^2(\mathbb{R})}^2$$

pout toute $f \in L^2(\mathbb{R})$ définit ce que l’on appelle un système de Bessel. Elle est la moins forte et elle est équivalente au fait que l’application

$$T : \ell^2 \rightarrow \mathcal{H}, \{c_\lambda\} \mapsto \sum_{\lambda \in \Lambda} c_\lambda g_\lambda = f$$

(on suppose qu’on s’est donné une énumération de Λ) est un opérateur *bien défini et borné* de l’espace ℓ^2 des suites de carré sommable dans \mathcal{H} . L’adjoint $T^* : \mathcal{H} \rightarrow \ell^2$ est alors $f \mapsto \{\langle f, g_\lambda \rangle\}$.

La condition

$$A \|f\|_{L^2(\mathbb{R})}^2 \leq \sum_{\lambda \in \Lambda} |\langle f, g_\lambda \rangle|_{L^2(\mathbb{R})}^2$$

est la plus contraignante. Si $\{g_\lambda\}$ est une trame de Gabor alors l’opérateur

$$S = TT^* : \mathcal{H} \rightarrow \mathcal{H}, f \mapsto \{\langle f, g_\lambda \rangle\} \mapsto \sum_{\lambda \in \Lambda} \langle f, g_\lambda \rangle g_\lambda$$

s’appelle l’opérateur de la trame. Il est borné, positif, surjectif, inversible. Cela implique d’ailleurs que T lui-même soit surjectif. Il admet un “quasi-inverse” au sens

où si \tilde{T} est la restriction de T au complément orthogonal K^\perp de son noyau $K \subset \ell^2$, \tilde{T} possède un inverse borné \tilde{T}^{-1} . On a alors (cf. l'article [119] de Ole Christensen)

$$\sum_{\lambda \in \Lambda} |\langle f, g_\lambda \rangle|_{L^2(\mathbb{R})}^2 \geq \frac{1}{\|\tilde{T}^{-1}\|^2} \|f\|^2 = \frac{1}{\|S^{-1}\|} \|f\|^2$$

et l'on peut donc choisir $A = \frac{1}{\|S^{-1}\|}$.²¹ En fait on peut montrer que $A = \frac{1}{\|S^{-1}\|}$ et $B = \|S\|$ sont les bornes *optimales*.

Quand la fenêtre g est donnée, la caractérisation des réseaux Λ satisfaisant la contrainte est un problème extrêmement difficile. Il a été résolu par exemple en dimension 1 pour les gaussiennes $g = e^{-\alpha x^2}$ par Yurii Lyubarskii, Kristian Seip et Robert Wallstén [494] grâce à l'utilisation de la transformée de Bargmann et ensuite pour les fonctions d'Hermite qui les généralisent par Karlheinz Gröchenig et Yurii Lyubarskii [226].²² La condition pour obtenir une trame est une condition de *densité* $a_0 \omega_0 \leq 1$ et si $a_0 \omega_0 = 1$ on obtient une base de Riesz.

Donnons une idée des calculs car ils montrent techniquement pourquoi le cas où g est une gaussienne est si privilégié. La transformée de Bargmann permet en effet d'établir l'équivalence entre la condition de trame et les ensembles d'unicité de fonctions holomorphes satisfaisant une certaine condition de croissance. Nous reviendrons plus bas à la section 5.10.6 sur ces calculs dans le cas de la transformation de Gabor *continue*, en relation avec la représentation du groupe de Heisenberg associée et la transformée de Bargmann.

Remarque sur les notations. Comme nous l'avons souvent indiqué, il existe plusieurs variantes des diverses transformées et représentations que nous mobilisons. Par exemple les phases sont définies modulo 1 (fréquences) ou modulo 2π (pulsations). En général autour d'un problème donné le milieu des spécialistes concernés utilise une variante bien définie. Dans le cas qui nous intéresse ici les atomes de Gabor sont souvent notés $g_{(x,\xi)}(t) = e^{2\pi i \xi t} g(t-x)$ (ξ est une fréquence) alors que

21. En effet, $\tilde{T}^{-1}f = \left\{ \left(\tilde{T}^{-1}f \right)_\lambda \right\}_{\lambda \in \Lambda}$ et $f = T\tilde{T}^{-1}f = \sum_{\lambda \in \Lambda} \left(\tilde{T}^{-1}f \right)_\lambda g_\lambda$. Calculons alors formellement comme dans [119] $\|f\|^4 = \langle f, f \rangle^2 = \left| \left\langle \sum_{\lambda \in \Lambda} \left(\tilde{T}^{-1}f \right)_\lambda g_\lambda, f \right\rangle \right|^2$. Cela vaut $\left| \sum_{\lambda \in \Lambda} \left(\tilde{T}^{-1}f \right)_\lambda \langle g_\lambda, f \rangle \right|^2$. Cette expression est de la forme $\left| \sum_{\lambda \in \Lambda} A_\lambda B_\lambda \right|^2$ et d'après l'inégalité de Cauchy-Schwartz dans ℓ^2 est $\leq \left(\sum_{\lambda \in \Lambda} |A_\lambda|^2 \right) \left(\sum_{\lambda \in \Lambda} |B_\lambda|^2 \right)$. Donc $\|f\|^4 \leq \left(\sum_{\lambda \in \Lambda} \left| \left(\tilde{T}^{-1}f \right)_\lambda \right|^2 \right) \left(\sum_{\lambda \in \Lambda} |\langle g_\lambda, f \rangle|^2 \right)$. Mais $\sum_{\lambda \in \Lambda} \left| \left(\tilde{T}^{-1}f \right)_\lambda \right|^2 \leq \|\tilde{T}^{-1}\|^2 \|f\|^2$. En simplifiant par $\|f\|^2$ on obtient donc $\|f\|^2 \leq \|\tilde{T}^{-1}\|^2 \left(\sum_{\lambda \in \Lambda} |\langle g_\lambda, f \rangle|^2 \right)$ qui est l'inégalité affirmée.

22. Pour le lecteur qui consulterait ce dernier article notons que les auteurs utilisent les variantes $g_{(a,\omega)}(x) = e^{2\pi i \omega(x-a)} g(x-a)$ et $(Bf)(z) = 2^{\frac{1}{4}} \int_{\mathbb{R}} f(t) e^{2\pi i t z - \pi t^2 - \frac{\pi}{2} z^2} dt$ (cf. section 5.10.1).

nous avons utilisé plus haut la variante $g_{(a,\omega)}(x) = e^{i\omega x}g(x-a)$ (ω est une pulsation). Nous passerons aux fréquences car c'est imposé par la variante choisie de la transformée de Bargmann utilisée plus bas à la section 5.10.1. \square

Nous partons donc d'atomes de Gabor

$$g_{(a,\xi)}(x) = e^{2\pi i \xi x} g(x-a)$$

où ξ est une fréquence et g une *gaussienne* $g(x) = e^{-\pi x^2}$. Nous posons $z = a + i\xi$ (donc $g_{(a,\xi)} = g_z$) et nous discrétisons en restreignant z à un réseau Λ . Si $f(x) \in L^2(\mathbb{R})$ sa transformée de Bargmann est (cf. ci-dessus chapitre 8, section 8.3.5 et ci-dessous section 5.10.1)

$$(Bf)(z) = F(z) = 2^{\frac{1}{4}} \int_{\mathbb{R}} f(t) e^{2\pi t z - \pi t^2 - \frac{\pi}{2} z^2} dt .$$

Rappelons que Bf est une isométrie de $L^2(\mathbb{R})$ sur l'espace de Fock ou de Segal-Bargmann \mathcal{F} des fonctions *holomorphes* de carré intégrable pour la mesure gaussienne $e^{-\pi|z|^2}$. On remarque alors que l'opérateur de Gabor

$$(V_g f)(z) = \langle f, g_z \rangle$$

satisfait

$$2^{\frac{1}{4}} (V_g f)(z) = e^{i\pi \xi a} (Bf)(z) e^{-\frac{\pi|z|^2}{2}} .$$

En effet

$$2^{\frac{1}{4}} (V_g f)(z) = 2^{\frac{1}{4}} \int_{\mathbb{R}} f(t) e^{2\pi i \xi t} e^{-\pi(t-a)^2} dt$$

ce qui donne comme puissance dans l'exponentielle

$$2\pi i \xi t - \pi(t-a)^2 = 2\pi i \xi t - \pi t^2 - \pi a^2 + 2\pi t a .$$

D'un autre côté

$$e^{i\pi \xi a} (Bf)(z) e^{-\frac{|z|^2}{2}} = 2^{\frac{1}{4}} \int_{\mathbb{R}} f(t) e^{i\pi \xi a} e^{-\frac{\pi|z|^2}{2}} e^{2\pi t z - \pi t^2 - \frac{\pi}{2} z^2} dt$$

ce qui donne comme puissance dans l'exponentielle

$$\begin{aligned} & i\pi \xi a - \frac{\pi|z|^2}{2} + 2\pi t z - \pi t^2 - \frac{\pi}{2} z^2 \\ &= i\pi \xi a - \frac{\pi a^2}{2} - \frac{\pi \xi^2}{2} + 2\pi t a + 2\pi i t \xi - \pi t^2 - \frac{\pi}{2} a^2 + \frac{\pi}{2} \xi^2 - \frac{\pi}{2} 2ia\xi \\ &= -\pi a^2 + 2\pi t a + 2\pi i t \xi - \pi t^2 . \end{aligned}$$

On a bien l'égalité.

Cette égalité

$$2^{\frac{1}{4}} (V_g f)(z) = e^{i\pi \xi a} (Bf)(z) e^{-\frac{\pi|z|^2}{2}}$$

est remarquable car elle relie le problème des trames de Gabor pour les gaussiennes à des propriétés d'espaces de fonctions *holomorphes*.

Soit alors l'action suivante sur les fonctions $F(w)$ d'une autre variable complexe w :

$$\chi(z) F(w) = e^{i\pi\xi a} e^{\pi w z} F(w - \bar{z}) e^{-\frac{\pi|z|^2}{2}} .$$

Un calcul analogue, sur lequel nous reviendrons en détail plus bas à la section 5.10.6 en termes de représentations continues, montre que χ et B sont "entrelacées" au sens où

$$\chi(z) (Bf(x))(w) = B(f_z(x))(w) .$$

En effet

$$\chi(z) (Bf(x))(w) = e^{i\pi\xi a} e^{\pi w z} e^{-\frac{\pi|z|^2}{2}} 2^{\frac{1}{4}} \int_{\mathbb{R}} f(t) e^{2\pi t(w - \bar{z}) - \pi t^2 - \frac{\pi}{2}(w - \bar{z})^2} dt$$

ce qui donne comme exposant de l'exponentielle

$$\begin{aligned} & i\pi\xi a + \pi w z - \frac{\pi|z|^2}{2} + 2\pi t(w - \bar{z}) - \pi t^2 - \frac{\pi}{2}(w - \bar{z})^2 \\ &= 2\pi i\xi a + 2\pi w a - \pi a^2 - \frac{\pi}{2}w^2 - \pi t^2 + 2\pi t w - 2\pi t a + 2\pi i t \xi \\ &= 2\pi i\xi (a + t) + 2\pi (a + t) w - \pi (a + t)^2 - \frac{\pi}{2}w^2 . \end{aligned}$$

D'un autre côté, en appliquant le changement de variable $t \rightsquigarrow t - a$,

$$\begin{aligned} B(f_z(x))(w) &= 2^{\frac{1}{4}} \int_{\mathbb{R}} e^{2\pi i \xi t} f(t - a) e^{2\pi t w - \pi t^2 - \frac{\pi}{2}w^2} dt \\ &= 2^{\frac{1}{4}} \int_{\mathbb{R}} e^{2\pi i \xi (t+a)} f(t) e^{2\pi (t+a) w - \pi (t+a)^2 - \frac{\pi}{2}w^2} dt \end{aligned}$$

ce qui donne comme exposant de l'exponentielle

$$2\pi i \xi (t + a) + 2\pi (t + a) w - \pi (t + a)^2 - \frac{\pi}{2}w^2 .$$

Il y a bien égalité.

La conséquence de cet entrelacement est que la condition de trame peut se traduire en une condition sur les fonctions holomorphes de \mathcal{F} , c'est-à-dire de carré intégrable pour la mesure gaussienne $e^{-|z|^2}$. Elles satisfont la contrainte

$$|F(z)| \leq e^{\pi|z|^2} \|F\|$$

(cf. plus bas la section 5.10.1). La nouvelle condition est que le réseau $\bar{\Lambda}$ soit un *ensemble d'échantillonnage* (sampling set) pour \mathcal{F} c'est-à-dire qu'il existe des

constantes $0 < A \leq B < \infty$ telles que

$$A \|F\|^2 \leq \sum_{z_n \in \bar{\Lambda}} |F(z_n)|^2 \leq B \|F\|^2 .$$

Cela est équivalent à ce que $\bar{\Lambda}$ soit un *ensemble d'unicité* pour le sous-espace \mathcal{F}^∞ de \mathcal{F} qui est l'espace des $F(z) \in \mathcal{F}$ telles que

$$\|F\|_\infty = \sup e^{\frac{\pi|z|^2}{2}} |F(z)| < \infty$$

(i.e. $|F(z)| \leq e^{\frac{\pi|z|^2}{2}} \|F\|_\infty$). Cette dernière condition signifie que si $F(z) \in \mathcal{F}^\infty$ s'annule sur les points de $\bar{\Lambda}$ elle est alors *identiquement nulle*.

En dimension 1, un autre cas proche de celui des gaussiennes a été résolu par Augustus Janssen et Thomas Strohmer [272]. Il s'agit de celui de la sécante hyperbolique

$$g(x) = \operatorname{sech}(x) = \frac{1}{\cosh(x)} = \frac{2}{e^x + e^{-x}}$$

pour le réseau $\Lambda = (a_0\mathbb{Z} \times \omega_0\mathbb{Z})$ de $\mathbb{R}_{(a,\omega)}^2$. Là encore la condition de trame est une condition de densité $a_0\omega_0 < 1$. Si $g(x) = C \operatorname{sech}(\alpha x)$, $\alpha, C > 0$, $\|g\|^2 = C^2 \frac{2}{\alpha}$ et la normalisation $\|g\| = 1$ impose $C = \sqrt{\frac{\alpha}{2}}$. Souvent on prend $\alpha = \pi$ et donc $g(x) = \sqrt{\frac{\pi}{2}} \operatorname{sech}(\pi x)$. Dans ce cas, Markus Faulhuber et Irina Shafkulovska [180] ont montré que, pour $a_0\omega_0 = \frac{1}{n}$, $n \geq 2$, les bornes A et B optimales sont obtenue pour le réseau carré $a_0 = \omega_0 = \frac{1}{\sqrt{n}}$. Les calculs sont compliqués.

Remarque. On peut développer des techniques analogues de discrétisation pour les ondelettes. Cf. par exemple Mallat [348], chapitre VII. Mais le problème est nettement plus compliqué car si l'on regarde les groupes qui sont en arrière plan du problème, nous avons vu que pour les Gabor il s'agit du groupe de Heisenberg alors que pour les ondelettes il s'agit du groupe affine. Or le premier est unimodulaire alors que le second ne l'est pas, ce qui complique les choses. □

Il existe donc quelques résultats en dimension 1 pour $L^2(\mathbb{R})$. Mais il n'existe que peu de résultats un peu précis pour $L^2(\mathbb{R}^d)$ en dimension $d > 1$. Ainsi que l'affirmait Karlheinz Gröchenig en 2011 :

“By contrast, in higher dimensions next to nothing is known about the question which lattices $\Lambda \subseteq \mathbb{R}^{2d}$ generate a Gabor frame $\mathcal{G}(\varphi, \Lambda)$. Gabor frames in higher dimensions seem to be on a different level of difficulty and are completely uncharted territory.”

2.2.2. *Trames de Gabor dans \mathbb{V}_S (Liontou-Marcolli).*

D'où l'intérêt de l'article [336] de Vasiliki Liontou et Matilde Marcolli “Gabor Frames from Contact Geometry in Models of the Primary Visual Cortex” (2022)

montrant comment, dans le cas de notre modèle \mathbb{V}_S de $V1$, on peut obtenir des trames de Gabor. Nous allons le résumer brièvement.

Les auteurs posent d'emblée la question de la discrétisation qui est imposée par le caractère discret de l'implémentation neuronale.

“While there is a rich bibliography on the representation of receptive profiles by continuous time-frequency signal representations, the question that arises is whether the functional geometry of the visual mechanism directly incorporates a choice of a discrete sampling set suitable for decoding and encoding visual stimuli.” (p. 3)

Les auteurs partent de l'action de $E(2)$ sur $M = \mathbb{V}_S = \mathbb{R}^2 \times \mathbb{S}^1$ (coordonnées $(x, y, \theta \in [0, \pi])$) et du champ de profils récepteurs $\varphi_{(x,y,\theta)}$ obtenu à partir d'une ondelette mère φ_0 et ils se proposent de construire une trame de Gabor directement à partir de la géométrie de l'architecture fonctionnelle de \mathbb{V}_S , c'est-à-dire de sa structure de contact. Comme ils l'expliquent

“Our key observation (...) is the fact that the combined presence of the contact structure on the Legendrian circle bundle and a complex structure on the base surface determines an associated bundle of framed lattices, which in turn provide the required discrete sampling set for the Gabor frames.” (p. 2)

Ils tiennent compte du fait que la surface rétinienne S_r n'est pas plane, mais un morceau d'une surface de Riemann projeté sur $V1$ par une application conforme (la voie rétinogéniculocorticale), considèrent la coordonnée locale complexe $z = x + iy$ et interprètent \mathbb{V}_S comme le fibré en sphères $M = \mathbb{S}(T^*S_r)$ du fibré cotangent T^*S_r de S_r (la variable angulaire θ est donc l'orientation des covecteurs de norme 1 de T^*S_r). Ils prennent aussi en considération la structure presque complexe sur le fibré tangent TS_r et travaillent dans le fibré \mathcal{E} qui est le pull-back de TS_r par la projection canonique $\pi : M \rightarrow S_r$ ainsi que dans son dual \mathcal{E}^* . Cette double structure

“provide(s) a natural choice of a framed lattice (a lattice together with the choice of a basis) on the bundle $\mathcal{E} \oplus \mathcal{E}^*$ over the contact 3-manifold M .” (p. 3)

Ils construisent un système de filtres de Gabor et montrent que, par simple changement d'échelle, on peut réaliser la si contraignante condition de trame ;.

“In terms of the geometric model of the $V1$ visual cortex, this shows that the contact geometry directly determines the signal analysis, the Gabor frames property, and the observed shape of the receptor profiles of the $V1$ neurons.” (p. 4)

L'idée est que si l'on considère la structure de contact de M avec sa 1-forme ω et son champ de Reeb R , alors, en utilisant la structure complexe de S_r , on peut construire une autre forme de contact ω^J avec son champ de Reeb R^J , J étant

l'opérateur de multiplication par i satisfaisant donc $J^2 = -1$.²³ Rappelons qu'en termes de coordonnées (x, y, θ) on a

$$\omega = \cos(\theta) dx + \sin(\theta) dy ,$$

les plans de contact K sont engendrés par $\{\partial_\theta, -\sin(\theta) \partial_x + \cos(\theta) \partial_y\}$ et le champ de Reeb est le champ

$$R = \cos(\theta) \partial_x + \sin(\theta) \partial_y$$

satisfaisant $\omega(R) = \cos^2(\theta) + \sin^2(\theta) = 1$. On a aussi

$$\omega^J = -\sin(\theta) dx + \cos(\theta) dy ,$$

les plans de contact K_J étant engendrés par $\{\partial_\theta, \cos(\theta) \partial_x + \sin(\theta) \partial_y\}$ et le champ de Reeb étant

$$R^J = -\sin(\theta) \partial_x + \cos(\theta) \partial_y .$$

Les deux plans de contact K et K_J s'intersectent le long de ∂_θ , et R (resp. R^J) est legendrien pour ω (resp. ω^J). Ces trois vecteurs tangents ∂_θ, R et R^J définissent une base naturelle de TM .

Comme la surface S_r est différente de \mathbb{R}^2 , le lien de sa géométrie avec l'analyse du signal doit être un peu clarifié. Les auteurs passent donc des variables locales non linéaires (x, y) de S_r (avec $z = x + iy$) aux variables linéaires (v_1, v_2) dans TS_r muni du produit scalaire induit par la structure riemannienne de S_r .

“Thus, we think of the retinal signal as a collection of compatible signals in the planes $T_{(x,y)}S_r$, as (x, y) varies in S_r .” (p. 7)

Un signal est donc maintenant une fonction réelle dans $L^2(\mathcal{E}, \mathbb{R})$ et il s'agit de l'analyser.

“For signals on manifolds there is in general no good construction of associated filters for signal analysis, although partial results exist involving splines discretization, diffusive wavelets, or special geometries such as spheres and conformally flat manifolds. One of our goals here is to show that geometric modelling of the visual cortex in terms of contact geometry and the description of receptive fields in terms of Gabor frames suggest a general way of performing signal analysis on a specific class of contact manifolds.” (p. 9).

Les auteurs reprennent alors les bases de la théorie des trames de Gabor que nous venons d'exposer. On est en dimension 2 et les filtres de Gabor sur \mathbb{R}^2 sont construits à partir des opérateurs $\rho(\lambda)$ sur $L^2(\mathbb{R}^2)$ paramétrés par

$$\lambda = (a, \xi) \in \mathbb{R}^{2 \times 2}$$

23. Nous utilisons la notation ω^J pour éviter la confusion avec ω_J , la forme de contact de notre modèle \mathbb{V}_J . Quand nous travaillons dans \mathbb{V}_J ou \mathbb{V}_S nous notons si besoin est ω_J et ω_S les formes de contact correspondantes. Ici l'on travaille dans \mathbb{V}_S et ω_S est notée simplement ω .

et donnés par

$$\rho(\lambda) = \exp(2\pi i \langle a, \xi \rangle) T_a M_\xi$$

avec

$$\begin{cases} T_a h(t) = h(t - a) & \text{(translation)} \\ M_\xi h(t) = \exp(2\pi i \langle \xi, t \rangle) h(t) & \text{(modulation)} \end{cases}$$

liés par la relation de commutation

$$T_s M_\xi = \exp(-2\pi i \langle s, \xi \rangle) M_\xi T_s .$$

Si le réseau est $\Lambda = AZ^{2 \times 2}$ ($A \in GL(\mathbb{R}, 4)$), le système de Gabor induit par la fenêtre g (l'ondelette mère) est $\mathcal{G} = \{\rho(\lambda)g\}_{\lambda \in \Lambda}$ et les $f \in L^2(\mathbb{R}^2)$ sont analysées par les coefficients $\{\langle f, \rho(\lambda)g \rangle\}_{\lambda \in \Lambda}$.

La question est de savoir si l'on peut généraliser ce cas classique à M muni de sa structure de contact :

“it is natural to consider the question of whether there is a lattice, directly determined by the geometric model of $V1$, with respect to which the receptor profiles are organized into a Gabor frame system.” (p. 10)

Les auteurs passent donc au fibré \mathcal{E} . Avec $v = (v_1, v_2) \in T_{(x,y)}S_r \simeq \mathbb{R}^2$ et $\eta = (\eta_1, \eta_2) \in T_{(x,y)}^*S_r \simeq \mathbb{R}^2$ dans les bases (∂_x, ∂_y) et (dx, dy) , ils considèrent d'abord une fenêtre Φ_0 sur $TS_r \oplus T^*S_r$ de la forme

$$\Phi_{0,(x,y)}(v, \eta) = \exp\left(-v^t A_{(x,y)} v - i \langle \eta, v \rangle_{(x,y)}\right)$$

où A est une forme quadratique définie positive bien choisie. Ils restreignent Φ_0 au fibré

$$TS_r \oplus \mathbb{S}(T^*S_r) = TS_r \oplus M = \mathcal{E}$$

et en dérivent une fenêtre sur \mathcal{E} . Cela leur permet de passer à la structure symplectique de dimension 4 de $TS_r \oplus T^*S_r$ et à la structure de contact de dimension 3 sous-jacente. Ils prennent alors les filtres

$$\psi_{0,(x,y,\theta)}(v) = \exp\left(-v^t A_{(x,y)} v - i \langle \eta_\theta, v \rangle_{(x,y)}\right)$$

avec $\eta_\theta = (w \cos(\theta), w \sin(\theta))$ un covecteur d'angle θ .

Pour obtenir une trame de Gabor, il faut choisir des réseaux Λ particuliers permettant de vérifier la condition de trame. C'est là que les auteurs utilisent la structure de contact et les champs de Reeb. Ils notent que les 1-formes ω and ω^J définissent une base adaptée à la structure de contact de M et à la structure complexe de S_r et prennent les réseaux $\mathbb{Z}R + \mathbb{Z}R^J$ et, dualement, $\mathbb{Z}\omega + \mathbb{Z}\omega^J$ sur \mathcal{E} et \mathcal{E}^* . Il les réarrange en $\Lambda = \mathbb{Z}R + \mathbb{Z}\omega$ et $\Lambda_J = \mathbb{Z}R^J + \mathbb{Z}\omega^J$. Ils obtiennent ainsi un système de Gabor pour $L^2(\mathcal{E}_{(x,y,\theta)})$. La question est maintenant de vérifier la condition de trame. C'est évidemment la plus difficile et la plus technique. Elle passe là encore par l'utilisation de la transformée de Bargmann.

La vérification est délicate mais, en utilisant les résultats en dimension 1 que nous avons exposés, les auteurs arrivent à l'effectuer. Si la matrice A est diagonale $A = \text{diag}(\alpha, \beta)$ ($\alpha, \beta > 0$), le système de Gabor se découple en un produit de deux systèmes de Gabor de dimension 1 indépendants, celui de $g(t) = e^{-\alpha t^2 - it}$ et le réseau \mathbb{Z}^2 et celui de $h(t) = e^{-\beta t^2}$. La condition de trame n'est pas satisfaite car il faudrait $s(\Lambda) < 1$ ce qui n'est pas le cas pour \mathbb{Z}^2 puisque $s(\mathbb{Z}^2) = 1$. Mais en rescalant le modèle d'un facteur $b_{(x,y,\theta)} < 1$ qui est le quotient du rayon d'injectivité r_{inj} de S_r par le rayon empirique r_{max} venant des contraintes neurophysiologiques on peut obtenir $s(\Lambda) < 1$.²⁴

Lorsque la matrice A n'est pas diagonale la situation est plus complexe mais l'on peut également réaliser la condition de trame.

Dans une dernière partie, les auteurs envisagent les paramètres additionnels de notre modèle \mathbb{V}_S :

- (i) l'échelle, ce qui conduit à la symplectisation de M (dimension 4 et espace de Engel) ;
- (ii) la phase, ce qui conduit à la contactisation de cette symplectisation (dimension 5).

Ils appliquent les mêmes méthodes et montrent que la condition de trame peut encore être obtenue si le modèle est correctement mis à l'échelle.

3. Vers l'analyse harmonique généralisée

À partir de ces exemples, et avant que d'en venir aux états cohérents de façon un peu plus précise, faisons quelques remarques sur l'analyse harmonique générale d'un groupe de Lie. Elle consiste à décrire les représentations irréductibles de G , dont les classes d'équivalence constituent, rappelons-le, le dual de Pontryagin \widehat{G} de G , et à décomposer la représentation régulière de G en représentations irréductibles. Rappelons également que la représentation régulière à gauche Λ de G est définie sur les $F(h) \in L^2(G, dg)$ par $\Lambda_g F(h) = F(g^{-1}h)$. La représentation régulière à droite \mathcal{R} est définie par $\mathcal{R}_g F(h) = F(hg)$. Il y a deux classes de groupes qui généralisent des résultats connus depuis bien longtemps : celle des groupes compacts et celle des groupes abéliens.

3.1. Groupes de Lie compacts

L'analyse harmonique des groupes compacts généralise directement celle des groupes finis de la section 1.1.3. Si $\rho : G \rightarrow \text{Aut}(V_\rho)$ est une représentation, alors, par moyennage sur G , on peut toujours construire une forme hermitienne $\langle \cdot, \cdot \rangle$ G -invariante sur V_ρ qui en fait une représentation *unitaire*. ρ est alors complètement

24. Le rayon d'injectivité est le rayon maximal r tel que l'exponentielle $\exp(x, y) : T_{(x,y)}S \rightarrow S$ soit un difféomorphisme sur la boule $B(0, r)$ de $T_{(x,y)}S$.

réductible puisque si W est un sous-espace G -invariant, W^\perp l'est également et donc V_ρ est décomposable en somme directe orthogonale de sous-représentations. Les éléments de \widehat{G} sont donc des unirreps que la compacité de G force à être de dimension finie $d_\rho < +\infty$. \widehat{G} est un espace discret. Un théorème de base à leur propos est le théorème dit de Peter-Weyl.²⁵ Soit $\rho \in \widehat{G}$ et, pour $v, w \in V_\rho$, soit $C_{\rho,v,w}(g)$ la fonction sur G (continue car ρ est toujours continue par hypothèse) définie par

$$C_{\rho,v,w}(g) = \langle \rho(g)v, w \rangle .$$

Les $C_{\rho,v,w}$ sont des fonctions de G dans \mathbb{C} .²⁶ Comme G est compact, les $C_{\rho,v,w} \in L^2(G, dg)$. Le théorème de Peter-Weyl affirme d'abord que, lorsque l'on varie ρ, v, w , les $C_{\rho,v,w}$ sont denses dans $L^2(G)$. C'est une conséquence du théorème de Stone-Weierstrass qui dit que sur un espace topologique compact X une sous-algèbre \mathfrak{A} de l'algèbre $\mathcal{C}^0(X)$ des fonctions continues (la topologie étant celle de la convergence uniforme car X est compact) est dense si et seulement si elle sépare les points, autrement dit si pour tout $x \neq y$ de X il existe une $f \in \mathfrak{A}$ telle que $f(x) \neq f(y)$.²⁷ La séparation est triviale dans notre cas.

Si maintenant l'on fixe ρ et si l'on fait varier v, w , les $C_{\rho,v,w}$ engendrent un sous-espace $L_\rho^2(G)$ de fonctions "élémentaires" qui est G -invariant et dont les $C_{\rho,v,w}$, une fois normalisées en se restreignant à une base orthonormée de V_ρ , forment une base orthonormée. On montre que la restriction de la représentation régulière (à droite) $\mathcal{R}(g)$ de G à $L_\rho^2(G)$ est constituée de $d_\rho = \dim(V_\rho)$ copies de ρ et le théorème de Peter-Weyl dit que $L^2(G) \simeq \bigoplus_{\rho \in \widehat{G}} L_\rho^2(G)$ et donc chaque unirrep $\rho \in \widehat{G}$ est contenue dans la représentation régulière avec la multiplicité d_ρ . Cela signifie que la mesure de Plancherel est constituée des masses de Dirac $d_\rho \delta$ pour chaque "point" ρ de \widehat{G} . L'application la plus connue est celle du groupe des rotations $G = SO(3)$ conduisant aux *harmoniques sphériques*.

La transformée de Fourier s'exprime alors de la façon suivante. Si $\rho \in \widehat{G}$ est un unirrep, son caractère $\chi_\rho(g)$ est facile à définir car les $\rho(g)$ sont des matrices dans une base $\{e_j\}$ de V_ρ et

$$\chi_\rho(g) = \text{tr}(\rho(g)) = \sum_{j=1}^{j=d_\rho} \langle \rho(g)e_j, e_j \rangle .$$

25. Démontré en 1927 par Hermann Weyl et son élève Fritz Peter, ce théorème généralise des résultats connus depuis Frobenius, Schur et Burnside.

26. Les $C_{v,w}(g) = \langle \rho(g)v, w \rangle$ sont de la forme $L \circ \rho$ avec L une forme linéaire sur $\text{End}(V_\rho)$. En effet, si $\varphi \in \text{End}(V_\rho)$, $L(\varphi(v))$ est une forme linéaire sur V_ρ (i.e. un élément de V_ρ^* et donc, étant donnée la structure hermitienne de V_ρ , il existe un $w \in V_\rho$ tel que $L(\varphi(v)) = \langle \varphi(v), w \rangle$).

27. Weierstrass avait démontré le théorème pour $X = [a, b]$ un intervalle fermé borné (donc compact) de \mathbb{R} et \mathfrak{A} l'algèbre des polynômes. C'était un théorème d'approximations polynomiales des fonctions continues sur $[a, b]$.

Donc $\chi_\rho \in L^2_\rho(G)$. Si maintenant $f \in L^2(G)$, soit f_ρ sa projection sur $L^2_\rho(G)$. On montre que

$$f_\rho(g) = d_\rho((\chi_\rho * f)(g)) = d_\rho \int \chi_\rho(gg'^{-1}) f(g') dg'.$$

On a $f = \bigoplus_{\rho \in \widehat{G}} f_\rho$ où l'égalité et la convergence sont prises dans $L^2(G)$ mais deviennent celles des valeurs en tout g si f est "assez régulière".

3.2. Groupes de Lie abéliens

Une autre classe pour laquelle la théorie est sans difficulté est celle des groupes de Lie *abéliens non compacts* qui généralise directement la transformée de Fourier classique. Toutes les unirreps sont de dimension 1, i.e. à valeurs dans \mathbb{C} et, comme elles sont unitaires sont en fait à valeurs $\chi(g) \in \mathbb{S}^1 \subset \mathbb{C}$. Ce sont donc des caractères. Si G est compact on obtient un développement en série de Fourier des éléments de $L^2(G)$. Mais si G n'est pas compact ces caractères ne sont pas de carré intégrable puisqu'ils sont de valeur absolue constante $|\chi(g)| = 1$ et $L^2(G)$ n'est plus décomposable en somme directe \oplus mais seulement en intégrale \int^\oplus . Toutefois ces caractères forment un groupe abélien et le dual de Pontryagin \widehat{G} est un groupe²⁸ (avec la topologie de la convergence uniforme sur les compacts et pour mesure de Plancherel $dP(\chi)$ la mesure de Haar $d\chi$), ce qui n'est pas du tout le cas en général.

Si $f \in L^1(G) \cap L^2(G)$ (i.e. si f est intégrable et de carré intégrable) on peut définir pour tout $\chi \in \widehat{G}$ l'intégrale

$$\widehat{f}(\chi) = \int_G f(g) \overline{\chi(g)} dg$$

car $\chi(g) \in \mathbb{C}$. C'est la transformée de Fourier, dont on montre qu'elle est une isométrie entre $L^2(G, dg)$ et $L^2(\widehat{G}, d\chi)$. Si f est "assez régulière", il y a une formule d'inversion

$$f(g) = \int_{\widehat{G}} \widehat{f}(\chi) \chi(g) d\chi$$

et comme $\widehat{f}(\chi) \chi(g) = (\chi * f)(g)$ elle s'écrit, comme pour le cas compact,

$$f(g) = \int_{\widehat{G}} (\chi * f)(g) d\chi.$$

On trouve ainsi la décomposition

$$L^2(G) = \int_{\widehat{G}}^\oplus \mathbb{C} \chi d\chi.$$

28. Le produit des caractères est le produit des valeurs et l'inverse d'un caractère est son conjugué.

3.3. Séries de Fourier

Nous retrouvons ainsi le cas le plus simple, à la fois compact et abélien, à savoir celui du groupe multiplicatif \mathbb{S}^1 des nombres complexes de module 1 traité dans la section des séries de Fourier (6) du chapitre 16. Si θ est l'angle coordonnée de \mathbb{S}^1 , le dual est le groupe des caractères $\chi_n(\theta) = e^{in\theta}$ et la décomposition de

$$L^2(\mathbb{S}^1) = \bigoplus_{\chi_n} \mathbb{C}\chi_n$$

correspond au développement des f continues périodiques en séries de Fourier,

$$f(\theta) = \sum_n \widehat{f}(n) e^{in\theta}$$

avec

$$\widehat{f}(n) = \int_{\mathbb{S}^1} f(\theta) e^{-in\theta} d\theta = \langle f, \chi_n \rangle .$$

Le dual du groupe abélien continu compact \mathbb{S}^1 est, répétons-le, le groupe abélien discret \mathbb{Z} .

4. Etats cohérents

4.1. Idée de la théorie générale

Donnons maintenant une idée de la théorie générale des états cohérents, concept dont nous avons déjà rencontré un exemple à propos des séries de Fourier (chapitre 16, section 6) avec les noyaux de convolution de Féjer en tant qu'approximations de l'identité. Parmi les nombreuses introductions à ce sujet, citons par exemple l'excellent panorama de Sayed Twareque Ali, Jean-Pierre Antoine, Jean-Pierre Gazeau et Ute Mueller [547].

1. On veut analyser des signaux considérés comme vecteurs d'un espace de Hilbert \mathcal{H} qui est en général un $L^2(X)$ pour un certain espace avec mesure X (ici $L^2(\mathbb{R})$ ou $L^2(\mathbb{V}_J)$ ou encore $L^2(\mathbb{V}_S)$).

2. On dispose d'un groupe G séparable localement compact²⁹ agissant irréductiblement et unitairement sur \mathcal{H} par une représentation π (unirrep). Comme G est

29. G séparable localement compact signifie (i) que G admet un sous-ensemble dense au plus dénombrable (séparabilité), (ii) que G admet en tout point des voisinages compacts (compacité locale). Cela est évident pour tous les groupes que nous considérons. Mais il faut le préciser dans des situations plus générales, par exemple lorsque l'on considère des groupes comme les groupes de difféomorphismes qui sont de dimension infinie.

localement compact, on montre qu'il existe une mesure de Haar $d\mu(g)$ sur G , c'est-à-dire une mesure invariante par translation (à gauche) qui est l'analogue de la mesure de Lebesgue dx sur \mathbb{R} .³⁰

3. On dispose d'une ondelette "mère" $\varphi_0(x) \in \mathcal{H}$, $\varphi_0 \neq 0$, qui est bien localisée à la fois dans l'espace "direct" X (des positions, des orientations, et éventuellement des échelles, etc.) et aussi dans l'espace "de Fourier" correspondant, et qui approxime bien la minimisation des relations d'incertitudes associées à la non-commutativité de l'algèbre de Lie \mathcal{G} de G .³¹

4. On prend l'orbite $\Phi = \{\varphi_g\}_{g \in G}$ de φ_0 sous l'action de G (i.e. $\varphi_g(x) = (\pi(g)(\varphi_0))(x)$) et l'on suppose que φ_0 est "admissible" au sens où

$$\int_G |\langle f, \varphi_g \rangle_{\mathcal{H}}|^2 d\mu(g) < \infty$$

pour tout $f \in \mathcal{H}$, avec $\langle \bullet, \bullet \rangle_{\mathcal{H}}$ le produit scalaire sur \mathcal{H} et $d\mu(g)$ la mesure de Haar sur G .³² Pour cela, il suffit que

$$C_{\varphi_0} = \int_G |\langle \varphi_0, \varphi_g \rangle_{\mathcal{H}}|^2 d\mu(g) < \infty .$$

On dit alors que la représentation π est de *carré intégrable* et on obtient un *état cohérent*. Si φ_0 est admissible, toutes les φ_g sont admissibles. Évidemment, si G est compact, toutes les représentations sont de carré intégrable.

5. L'ensemble \mathcal{A} des φ admissibles est bien sûr G -invariant et comme la représentation est irréductible, il est soit nul soit dense dans \mathcal{H} . Dans le premier cas π n'est pas de carré intégrable. On montre que \mathcal{A} est tout le Hilbert \mathcal{H} si et seulement si G est unimodulaire.

6. L'existence d'un état cohérent $\Phi = \{\varphi_g\}$ implique l'existence de ce que l'on appelle une "résolution de l'identité" et d'un "noyau reproduisant". Comme $C_{\varphi_0} < \infty$ on peut normaliser φ_0 de façon appropriée. Posons alors

$$K(g, g') = \langle \varphi_{g'}, \varphi_g \rangle_{\mathcal{H}} .$$

30. Pour les groupes de Lie G , l'existence de la mesure de Haar (à gauche) est due au fait que l'on peut prendre une n -forme ($n = \dim(G)$) sur $T_e G$ et la translater à gauche, ce qui donne une mesure L -invariante. Idem pour la mesure de Haar à droite. Dans le cas général la preuve d'existence et unicité est plus technique et a été clarifiée par André Weil. Alfred Haar (1885-1933) est également connu pour l'"ondelette de Haar" (1909) considérée comme le premier exemple d'ondelette.

31. C'est d'ailleurs un problème intéressant et ouvert que de savoir, pour une représentation d'un groupe non commutatif G donné, quelles sont les ondelettes mères qui minimisent, si elles existent, les relations de Heisenberg. Cf. plus bas.

32. Notons que l'intégrale \int_G a pour intégrande des produits scalaires qui sont eux-mêmes des intégrales \int_X lorsque \mathcal{H} est un $L^2(X)$. De nombreuses formules sont obtenues en montrant que l'on peut inverser l'ordre des intégrations.

On a $K(g, g) = \|\varphi_g\|_{\mathcal{H}}^2 > 0$, $K(g', g) = K(g, g')^*$ et la relation fondamentale

$$\int_G K(g, g'') K(g'', g') d\mu(g'') = K(g, g') .$$

Cette relation découle du fait que l'intégrale sur G des opérateurs dans \mathcal{H} de projection Π_g sur les φ_g , i.e.

$$A = \int_G \Pi_g d\mu(g)$$

est l'*identité* à cause de la normalisation de φ_0 . En notation "physique" à la Dirac, ces projecteurs s'écrivent $\Pi_g = |\varphi_g\rangle \langle \varphi_g|$, les produits scalaires s'écrivent $\langle \varphi_g | \varphi_{g'} \rangle$ et l'on a

$$\begin{aligned} \int_G K(g, g'') K(g'', g') d\mu(g'') &= \int_G \langle \varphi_{g''} | \varphi_g \rangle \langle \varphi_{g'} | \varphi_{g''} \rangle d\mu(g'') \\ &= \langle \varphi_{g'} | \left(\int_G |\varphi_{g''}\rangle \langle \varphi_{g''}| d\mu(g'') \right) | \varphi_g \rangle \\ &= \langle \varphi_{g'} | \varphi_g \rangle = K(g, g') . \end{aligned}$$

Le noyau K généralise les fonctions de Dirac rencontrées dans la transformée de Fourier. Dans ce cas (sans tenir compte des constantes de normalisation), les $\varphi_\omega(x)$ sont les ondes planes $e^{i\omega x}$,

$$K(\omega, \omega') = \int_{\mathbb{R}_x} e^{i\omega'x} e^{-i\omega x} dx = \delta(\omega' - \omega)$$

et l'on a bien

$$\int_{\mathbb{R}_{\omega''}} \delta(\omega'' - \omega) \delta(\omega' - \omega'') d\omega'' = \delta(\omega' - \omega) .$$

7. On effectue l'analyse harmonique des signaux f de \mathcal{H} au moyen de cet état cohérent. Cela permet de représenter les signaux comme superpositions de fonctions élémentaires et donc, dans notre contexte neurogéométrique, de les mesurer neurophysiologiquement par des profils récepteurs à des fins de compression et d'analyse géométrique. Soit

$$T_f(g) = \langle f, \varphi_g \rangle_{\mathcal{H}} \in L^2(G)$$

la transformée de f . Les $T_f(g)$ sont bien dans $L^2(G)$ car π est de carré intégrable. La formule générale de la synthèse de $f \in \mathcal{H} = L^2(X)$ est

$$f(x) = \int_G T_f(g) \varphi_g(x) d\mu(g) .$$

On notera bien la structure de cette formule. On veut analyser des fonctions $f(x)$ et l'on dispose d'ondelettes $\varphi_g(x)$ paramétrées par $g \in G$. La fonction $f(x)$ est analysée en terme des $T_f(g)$ et reconstruite, synthétisée, comme une intégrale sur

G en considérant que x est fixé, les coefficients $T_f(g)$ étant les poids respectifs des ondelettes $\varphi_g(x)$ dans la synthèse de $f(x)$. C'est pourquoi l'état cohérent

$$\Phi = \{\varphi_g\}_{g \in G}$$

est appelé une "résolution de l'identité". Il permet de représenter les $f \in \mathcal{H}$ par des transformées $T_f(g) = \langle f, \varphi_g \rangle_{\mathcal{H}} \in L^2(G)$, les $\langle f, \varphi_g \rangle$ étant les "mesures" du signal f fournies par les profils récepteurs φ_g .

8. La transformée

$$\begin{aligned} T : \mathcal{H} = L^2(X) &\longrightarrow L^2(G), \\ f &\mapsto T_f \end{aligned}$$

est une isométrie sur son image \mathcal{H}_{φ_0} qui est un fermé de $L^2(G)$. Elle permet la "synthèse" de f par transformée inverse.

9. Un élément $u \in L^2(G)$ est dans l'image \mathcal{H}_{φ_0} , i.e. il existe $f \in \mathcal{H}$ tel que $u = T_f$, si et seulement si u satisfait la propriété de "reproduction"

$$u(g) = \int_G K(g, g') u(g') d\mu(g') .$$

En effet

$$\begin{aligned} \int_G K(g, g') T_f(g') d\mu(g') &= \int_G \langle \varphi_{g'}, \varphi_g \rangle_{\mathcal{H}} \langle f, \varphi_{g'} \rangle_{\mathcal{H}} d\mu(g') \\ &= \int_G \langle f, \varphi_{g'} \rangle_{\mathcal{H}} \langle \varphi_{g'}, \varphi_g \rangle_{\mathcal{H}} d\mu(g') \\ &= \int_G \langle f | \varphi_{g'} \rangle \langle \varphi_{g'} | \varphi_g \rangle d\mu(g') \quad (\text{notation à la Dirac}) \\ &= \langle f | \left(\int_G |\varphi_{g'} \rangle \langle \varphi_{g'} | d\mu(g') \right) | \varphi_g \rangle \\ &= \langle f, \varphi_g \rangle = T_f(g) . \end{aligned}$$

Il faut noter que les T_f sont de "vraies" fonctions et même des fonctions C^∞ , ce qui n'est pas le cas des éléments de $L^2(G)$, qui sont par définition des classes d'équivalence (au sens de la mesure) de fonctions de carré intégrable. Cela vient du fait que le noyau reproduisant permet d'évaluer les $u(g) \in \mathcal{H}_\varphi$ au moyen de la formule

$$u(g) = \int_G K(g, g') u(g') d\mu(g') = \langle u, K(*, g) \rangle_{L^2(G)} .$$

Qui plus est, le noyau K régularise les fonctions et les rend C^∞ , et par conséquent si u satisfait la formule ci-dessus, u est C^∞ . On obtient par cette technique une "représentation fonctionnelle" du Hilbert \mathcal{H} .

10. On notera que le noyau K est en fait un *noyau de convolution* sur G . En effet, on a

$$\begin{aligned} K(g, g') &= \langle \varphi_{g'}, \varphi_g \rangle_{\mathcal{H}} = \langle \pi(g') \varphi_0, \pi(g) \varphi_0 \rangle_{\mathcal{H}} \\ &= \langle \varphi_0, \pi(g')^* \circ \pi(g) \varphi_0 \rangle_{\mathcal{H}} \text{ par adjonction} \\ &= \langle \varphi_0, \pi(g')^{-1} \circ \pi(g) \varphi_0 \rangle_{\mathcal{H}} \text{ par unitarité} \\ &= \langle \varphi_0, \pi(g'^{-1}g) \varphi_0 \rangle_{\mathcal{H}} \text{ car } \pi \text{ est une représentation .} \end{aligned}$$

11. Ces travaux sur les états cohérents se complètent d'une recherche sur les formes optimales des profils récepteurs φ_g . En mécanique quantique, cette problématique est très ancienne. Elle consiste à trouver des états φ_g minimisant les relations d'incertitude de Heisenberg canoniquement associées au groupe considéré. Ces états optimaux sont en quelque sorte les états quantiques les plus proches possibles d'états classiques et permettent de mieux comprendre la façon dont la réalité quantique modifie la réalité classique.

Nous avons vu à la section 3.6.2. du chapitre 3 du Vol I que la statistique des images naturelles imposait de fortes contraintes à ces profils. D'autres travaux convergents ont montré que les architectures fonctionnelles et les champs d'association matérialisant le principe gestaltiste de bonne continuation reflètent les propriétés statistiques des lignes et des bords dans les images naturelles. Par exemple dans [502], Mariano Sigman *et al.* ont confirmé que

“the geometry of the pattern of interactions in primary visual cortex parallels the interactions of oriented segments in natural scenes.” (p.1939)

Leur méthode expérimentale consiste à mesurer la corrélation des orientations des bords entre un point origine 0 quelconque et un autre point a sur un corpus de $N = 4.000$ images naturelles. Dans sa thèse [485], Gonzalo Sanguinetti a montré que les résultats de Sigman convergent de façon remarquable avec les modèles neurogéométriques de bonne continuité.

Par ailleurs, s'inspirant du fait que les fonctions de Gabor ont été introduites en mécanique quantique comme les fonctions suffisamment bien localisées en position et en fréquence pour optimiser les relations d'incertitude de Heisenberg elles-mêmes associées à la non commutativité du groupe de Heisenberg, Citti, Sarti et leurs doctorants Davide Barbieri et Sanguinetti ont montré que des profils récepteurs optimaux pouvaient être déduits de la structure du groupe $SE(2)$ (cf. [28]).

Les états cohérents $\Phi = \{\varphi_g\}$ permettent ainsi de représenter les $f \in \mathcal{H}$ par des transformées

$$T_f(g) = \langle f, \varphi_g \rangle_{\mathcal{H}} \in L^2(G) .$$

Or c’est précisément ce que font les profils récepteurs des neurones (simples) d’une aire visuelle comme $V1$ ou $V2$, les $\langle f, \varphi_g \rangle_{\mathcal{H}}$ étant les “mesures” du signal f fournies par les profils récepteurs φ_g .

12. La *covariance* des mesures est garantie par le fait que les ondelettes φ_g de l’état cohérent appartiennent à une sous-représentation de la représentation régulière (à gauche) Λ de G dans $L^2(G, dg)$, représentation qui consiste, nous l’avons vu, à faire opérer les translations à gauche de G sur les fonctions sur G , autrement dit à associer à tout $g \in G$ l’automorphisme de $L^2(G)$:

$$\Lambda(g) = \Lambda_g \mapsto (F(h) \mapsto F(g^{-1}h)) .$$

En effet, la transformée en ondelettes T

$$\begin{aligned} T : \mathcal{H} &\rightarrow L^2(G) \\ f &\mapsto T_f \end{aligned}$$

“entrelace” la représentation π avec la représentation régulière Λ de G sur $L^2(G)$ au sens où le diagramme suivant est commutatif, $T \circ \pi = \Lambda \circ T$, car $f \mapsto T_{gf}$ peut s’obtenir de deux façons (on utilise l’égalité $\pi(h^{-1})\pi(g) = \pi((g^{-1}h)^{-1})$) :

$$\begin{array}{ccc} \mathcal{H} & \xrightarrow{T} & L^2(G) \\ f(x) & \longmapsto & T_f(h) = \langle f(x), \varphi_h(x) \rangle_{\mathcal{H}} \\ \pi(g) \downarrow & & \downarrow \Lambda(g) \\ gf(x) = f(g^{-1}x) & \longmapsto & \begin{aligned} T_f(g^{-1}h) &= \langle f(x), \varphi_{g^{-1}h}(x) \rangle_{\mathcal{H}} \\ &= \langle f(g^{-1}x), \varphi_h(x) \rangle_{\mathcal{H}} = T_{gf}(h) \end{aligned} \end{array}$$

13. En fait, on peut généraliser cette notion d’état cohérent en utilisant la formule d’inversion

$$\mathfrak{A}f(x) = \int_G T_f(g) \varphi_g(x) d\mu(g) = \int_G \langle f, \varphi_g \rangle_{\mathcal{H}} \varphi_g(x) d\mu(g)$$

pour définir un opérateur sur \mathcal{H} . Pour les états cohérents au sens strict décrits ci-dessus, \mathfrak{A} est l’*identité*. Mais on peut affaiblir cette propriété en demandant que \mathfrak{A} soit suffisamment bien “inversible” par exemple borné et d’inverse borné. Il y a là tout un riche champ d’analyse qui s’ouvre. Comme le note Bruno Torrèsani dans [539], il faut que G soit “assez gros” pour que l’on puisse retrouver assez de f et en même temps “pas trop gros” pour que \mathfrak{A} soit borné.

14. Dans de nombreux cas, on peut généraliser la construction d’états cohérents par un *passage au quotient* naturel et approprié (nous en avons vu un exemple plus haut avec la transformée de Gabor). Il existe un sous-groupe naturel H de G tel qu’en paramétrant les φ_g par les classes d’équivalence \tilde{g} des g dans G/H on arrive en fait à une construction satisfaisante. Pour cela il faut trouver une “bonne” section

$\sigma : G/H \longrightarrow G$ de la projection canonique $G \longrightarrow G/H$ qui permette de définir alors $\varphi_{\tilde{g}}$ par $\varphi_{\sigma(\tilde{g})}$ et l'opérateur \mathfrak{A}^σ par $\int_{G/H} \langle f, \varphi_{\tilde{g}} \rangle_{\mathcal{H}} \varphi_{\tilde{g}} d\mu(\tilde{g})$.

15. Souvent, cette construction passe par la corrélation entre la représentation π et une orbite $Ad_G^*(\zeta_\pi^*) = \mathcal{O}_\pi^*$ de la représentation co-adjointe de G fonctionnant comme un espace des phases. Cela signifie que π agit infinitésimalement à travers une représentation $d\pi$ de l'algèbre de Lie \mathcal{G} et que cette action infinitésimale est reconstructible à partir de $\zeta_\pi^* \in \mathcal{G}^*$. G opère sur \mathcal{G} à travers la représentation adjointe Ad_g et sur \mathcal{G}^* à travers la représentation co-adjointe Ad_g^* conformément à l'adjonction

$$\langle Ad_g^*(\zeta^*), X \rangle = \langle \zeta^*, Ad_{g^{-1}}(X) \rangle .$$

Le sous-groupe H de G est alors le stabilisateur de ζ_π^* et $\mathcal{O}_\pi^* \simeq G/H$ permet de paramétrer un état cohérent.

4.2. Précisions sur quelques exemples

Revenons aux ondelettes. La condition d'admissibilité

$$C_{\varphi_0} = \int_{\mathbb{R}^+} |\widehat{\varphi_0}(\xi)|^2 \frac{d\xi}{\xi} < \infty$$

exprime que

$$\int_A \langle \varphi_{(a,\sigma)}, \varphi_0 \rangle \frac{d\sigma da}{\sigma \sigma} < \infty ,$$

autrement dit que la représentation $\pi_{(a,\sigma)}$ est de carré intégrable et que l'ondelette mère φ_0 définit un état cohérent. Si φ_0 est assez régulière, par exemple intégrable et de carré intégrable, cela signifie que $\widehat{\varphi_0}(0) = 0$ autrement dit que $\varphi_0(x)$ est de moyenne $\int_{\mathbb{R}} \varphi_0(x) dx$ nulle. On a une résolution de l'identité (en notations de Dirac)

$$\int_A |\varphi_{(a,\sigma)}\rangle \langle \varphi_{(a,\sigma)}| \frac{d\sigma da}{\sigma \sigma} = Id$$

avec un noyau reproduisant

$$K((a, \sigma), (a', \sigma')) = \langle \varphi_{(a',\sigma')}, \varphi_{(a,\sigma)} \rangle .$$

On peut généraliser de plusieurs façons naturelles cette théorie basique des ondelettes. On peut par exemple considérer des translations $a \in \mathbb{R}^n$ et des multi-échelles σ définies par des matrices $n \times n$ appartenant à un groupe approprié Σ . G est alors le produit $\mathbb{R}^n \times \Sigma$ muni de la multiplication des matrices

$$g = (a, \sigma) = \begin{pmatrix} \sigma & a \\ 0 & 1 \end{pmatrix}$$

où maintenant a est un n -vecteur colonne et 0 un n -vecteur ligne. Les *espaces de Hardy* sont alors les $f(x) \in L^2(\mathbb{R}_{(x)}^n)$ dont la transformée de Fourier $\widehat{f}(\xi)$ a son support dans une orbite du groupe d'échelle Σ . Si cette orbite est ouverte et si Σ

agit librement sur elle, alors, comme l'ont montré David Bernier et Keith Taylor dans [49], on peut directement généraliser le cas $n = 1$ (cf. [539]).

On peut aussi ne garder qu'une seule échelle $\sigma \in \mathbb{R}_+^*$ mais introduire les rotations de $\mathbb{R}_{(x)}^n$.

Si besoin est, on peut également ajouter aux ondelettes des fréquences comme dans la transformée de Fourier, ce qui revient à considérer une transformée de Gabor avec des échelles puisque l'on a maintenant des fréquences ω (Fourier et Gabor), des translations a d'une fenêtre spatiale (Gabor, ondelettes) et des échelles σ (ondelettes). La loi de groupe est

$$(a', \sigma', \omega') \cdot (a, \sigma, \omega) = \left(a' + \sigma' a, \sigma' \sigma, \omega' + \frac{\omega}{\sigma'} \right) .$$

Les transformées sont alors les

$$(\pi_{(a, \sigma, \omega)} f)(x) = \frac{1}{\sqrt{\sigma}} e^{i\omega(x-a)} f\left(\frac{x-a}{\sigma}\right) .$$

Mais l'on rencontre le même problème qu'avec la transformée de Gabor. Les $\pi_{(a, \sigma, \omega)}$ ne constituent pas une représentation du groupe G_0 des (a, σ, ω) . On a en effet

$$\begin{aligned} (\pi_{(a', \sigma', \omega')} \circ \pi_{(a, \sigma, \omega)} h)(x) &= \frac{1}{\sqrt{\sigma'}} e^{i\omega'(x-a')} h'\left(\frac{x-a'}{\sigma'}\right) \\ &\quad \text{avec } h'(x) = \frac{1}{\sqrt{\sigma}} e^{i\omega(x-a)} h\left(\frac{x-a}{\sigma}\right) \\ &= \frac{1}{\sqrt{\sigma' \sigma}} e^{i\omega'(x-a')} e^{i\omega\left(\frac{x-(a'+\sigma'a)}{\sigma'}\right)} h\left(\frac{x-(a'+\sigma'a)}{\sigma' \sigma}\right) \\ &= ? \pi_{(a'+\sigma'a, \sigma' \sigma, \omega' + \frac{\omega}{\sigma'})} \\ &= \frac{1}{\sqrt{\sigma' \sigma}} e^{i(\omega' + \frac{\omega}{\sigma'})(x-(a'+\sigma'a))} h\left(\frac{x-(a'+\sigma'a)}{\sigma' \sigma}\right) . \end{aligned}$$

Là encore on s'aperçoit que la deuxième expression comprend le terme supplémentaire $e^{-i\omega' \sigma' a}$. Pour obtenir un tel terme, il faut introduire un facteur de phase $e^{-i\phi}$, $\phi \in \mathbb{S}_{(\phi)}^1$, et considérer donc le groupe de Heisenberg (polarisé) *affine* G des $(a, \sigma, \omega, \phi)$ avec la loi de composition

$$(a', \sigma', \omega', \phi') \cdot (a, \sigma, \omega, \phi) = \left(a' + \sigma' a, \sigma' \sigma, \omega' + \frac{\omega}{\sigma'}, \phi' + \phi + \omega' \sigma' a \text{ modulo } (2\pi) \right) .$$

Avec des $\phi = 0$ on obtient alors une représentation de G_0 en tant que *quotient* de G par le sous-groupe des phases $\mathbb{S}_{(\phi)}^1$. Ce sous-groupe des $(0, 1, 0, \phi)$ est le centre Z de G . G_0 peut être "relevé" dans G au moyen de la section "plate"³³ Σ_0 de la

33. Nous n'utilisons pas le symbole naturel σ car il est déjà utilisé pour l'échelle.

projection canonique $\pi : G \rightarrow G_0$ qui sélectionne dans chaque fibre $(a, \sigma, \omega, \mathbb{S}_{(\phi)}^1)$ de G au-dessus de (a, σ, ω) l'élément $(a, \sigma, \omega, 0)$. Mais $\Sigma_0(G_0)$ n'est pas un sous-groupe de G puisque justement G est construit de façon à ce que, même lorsque $\phi = \phi' = 0$, le produit donne une phase $\omega'\sigma'a$. En termes matriciels, la loi de composition de G

est la multiplication des matrices $\begin{pmatrix} 1 & \sigma\omega & \phi \\ 0 & \sigma & a \\ 0 & 0 & 1 \end{pmatrix}$ et donc la loi de composition des (a, σ, ω) est devenue la multiplication des matrices $\begin{pmatrix} 1 & \sigma\omega & 0 \\ 0 & \sigma & a \\ 0 & 0 & 1 \end{pmatrix}$ avec $\phi = 0$.

La représentation

$$(\pi_{(a, \sigma, \omega, \phi)} f)(x) = \frac{1}{\sqrt{\sigma}} e^{-i\phi} e^{i\omega(x-a)} f\left(\frac{x-a}{\sigma}\right)$$

du groupe de Heisenberg (polarisé) affine est irréductible. Mais elle *n'est pas* de carré intégrable car, en quelque sorte, G est "trop gros" pour l'espace de représentation $\mathcal{H} = L^2(\mathbb{R})$. On trouvera dans [547] et [539] des méthodes pour pallier ce problème et définir correctement des états cohérents. En utilisant la méthode des orbites de la représentation co-adjointe de G , on peut par exemple

(i) quotienter G par le sous-groupe G_1 des phases et des échelles, en remarquant que, comme

$$(0, \sigma', 0, \phi') \cdot (0, \sigma, 0, \phi) = (0, \sigma'\sigma, 0, \phi + \phi') ,$$

il s'agit bien d'un sous-groupe et

(ii) relever dans G l'espace des phases symplectique qu'est le quotient $G/G_1 \simeq G_0$ au moyen de sections appropriées Σ_1 de la projection canonique $\pi : G \rightarrow G_0$ qui sélectionnent un élément dans chaque fibre $(a, \mathbb{R}_+^*, \omega, \mathbb{S}_{(\phi)}^1)$ de G au-dessus de (a, ω) . On peut en particulier prendre une section "plate" pour les phases (i.e. $\phi = 0$) et faire de la fréquence ω une fonction de l'échelle σ en préservant la structure du groupe. Un choix standard est

$$\omega_\lambda(\sigma) = \lambda \left(\frac{1}{\sigma} - 1 \right) .$$

La composition donne en effet pour les $\omega_\lambda(\sigma)$

$$\begin{aligned} \omega'_\lambda(\sigma') + \frac{\omega_\lambda(\sigma)}{\sigma'} &= \lambda \left(\frac{1}{\sigma'} - 1 \right) + \frac{\lambda}{\sigma'} \left(\frac{1}{\sigma} - 1 \right) = \lambda \left(\frac{1}{\sigma'\sigma} - 1 \right) \\ &= \omega_\lambda(\sigma'\sigma) . \end{aligned}$$

On peut aussi quotienter G par le sous-groupe à 1-paramètre

$$H = \{(0, \sigma, \omega_\lambda(\sigma), 0)\}$$

et travailler avec cet espace homogène $X_\lambda = G/H_\lambda$. Dans le demi-plan des $(\sigma > 0, \omega)$, le graphe de $\omega_\lambda(\sigma)$ est une hyperbole équilatère que l'on peut paramétrer par $\sigma > 0$ ou par $\omega \in]-\lambda, +\infty[$ (si $\lambda > 0$). Les coordonnées naturelles de X_λ sont la translation a (qui n'a pas été quotientée) et, par exemple, la fréquence ω . Il faut alors construire une "bonne" section $\Sigma : X_\lambda \rightarrow G$ permettant de construire un état cohérent et une résolution de l'identité. On trouvera dans [547] une analyse des "bonnes" sections de la forme

$$\Sigma(a, \omega) = (a, \sigma(a, \omega), \omega) .$$

Avec $\sigma(a, \omega) = \frac{1}{\gamma\omega + \delta}$, $\gamma, \delta \in \mathbb{R}$ et $\omega \neq -\frac{\delta}{\gamma}$ si $\gamma \neq 0$, on obtient des résolutions de l'identité. Si $\gamma = 0$ et $\delta > 0$, σ est constante et l'on retrouve simplement la transformée de Gabor scalée. Mais avec d'autres fonctions $\sigma(a, \omega)$ bien choisies on peut quand même obtenir des états cohérents comme l'a montré Bruno Torrèsani dans [538].

4.3. Les représentations régulières et quasi-régulières

4.3.1. Représentations quasi-régulières des produits semi-directs.

Nous avons vu que la représentation régulière (à gauche) Λ d'un groupe de Lie G fait opérer G sur les $F(h) \in L^2(G, dg)$ par $\Lambda_g F(h) = F(g^{-1}h)$.³⁴ Les opérateurs Λ_g sont de norme 1 mais $\|\Lambda_g - \Lambda_{g'}\| = 2$ si $g \neq g'$ ($\|\Lambda_g - \Lambda_{g'}\| \leq 2$ par inégalité triangulaire et la borne est atteignable). La continuité forte est évidente sur le sous-espace dense $\mathcal{C}_\downarrow^\infty(G)$ et donc sur $L^2(G)$. Elle se décompose, nous l'avons vu, en représentations irréductibles et possède une structure très riche.

Dans le cas où le groupe G est un produit semi-direct, par exemple de la forme $G = \mathbb{R}_{(a)}^n \rtimes \Gamma_{(\gamma)}$, Γ opérant sur $\mathbb{R}_{(a)}^n$ par un morphisme de groupes $\gamma \mapsto \tilde{\gamma} \in GL(n)$ et le produit sur G étant donné pour les $g = (a, \gamma) \in G$ par

$$gg' = (a, \gamma) \cdot (a', \gamma') = (a + \tilde{\gamma}(a'), \gamma\gamma') ,$$

on peut définir une représentation que l'on appelle "quasi-régulière". La mesure de Haar de G est

$$d\mu(g) = \frac{1}{|\det(\tilde{\gamma})|} da d\mu(\gamma)$$

("det" signifie "déterminant") où da est la mesure de Lebesgue sur $\mathbb{R}_{(a)}^n$ et $d\mu(\gamma)$ la mesure de Haar sur Γ . La représentation quasi-régulière $\Lambda_\times : G \rightarrow \text{Aut}\left(L^2\left(\mathbb{R}_{(a)}^n\right)\right)$ (et non pas $L^2(G)$) est définie par

$$\Lambda_\times(a, \gamma) f(b) = \frac{1}{\sqrt{|\det(\tilde{\gamma})|}} f(\tilde{\gamma}^{-1}(b - a)) .$$

34. La représentation régulière à droite \mathcal{R} fait opérer G par $\mathcal{R}_g F(h) = F(hg)$.

On vérifie immédiatement qu'il s'agit bien d'une représentation. En effet, d'un côté

$$\begin{aligned} \Lambda_{\times}(a, \gamma) \circ \Lambda_{\times}(a', \gamma') f(b) &= \Lambda_{\times}(a, \gamma) \left(\frac{1}{\sqrt{|\det(\tilde{\gamma}')|}} f(\tilde{\gamma}'^{-1}(b - a')) \right) \\ &= \frac{1}{\sqrt{|\det(\tilde{\gamma})|}} \frac{1}{\sqrt{|\det(\tilde{\gamma}')|}} f(\tilde{\gamma}^{-1}(\tilde{\gamma}'^{-1}(b - a') - a)) . \end{aligned}$$

D'un autre côté,

$$\Lambda_{\times}(a' + \tilde{\gamma}'(a), \gamma'\gamma) f(b) = \frac{1}{\sqrt{|\det(\tilde{\gamma}'\gamma)|}} f(\tilde{\gamma}'\gamma^{-1}(b - a' - \tilde{\gamma}'(a))) .$$

Le déterminant étant multiplicatif les facteurs sont les mêmes dans les deux formules. Par ailleurs $\tilde{\gamma}^{-1}\tilde{\gamma}'^{-1} = \tilde{\gamma}'\gamma^{-1}$ et donc le terme en $b - a'$ est le même. Enfin $\tilde{\gamma}'\gamma^{-1}\tilde{\gamma}'(a) = \tilde{\gamma}^{-1}\tilde{\gamma}'^{-1}(a) = \tilde{\gamma}^{-1}(a)$ et le terme en $-a$ est également le même. On a donc bien

$$\Lambda_{\times}(a, \gamma) \circ \Lambda_{\times}(a', \gamma') = \Lambda_{\times}((a', \gamma') \cdot (a, \gamma)) .$$

Qui plus est, la représentation est *unitaire* grâce au facteur de normalisation $\frac{1}{\sqrt{|\det(\tilde{\gamma})|}}$.

Il est facile de calculer la transformée de Fourier de $\widehat{\Lambda}_{\times}$ sur $\widehat{\mathbb{R}}_{(a)}^n = \mathbb{R}_{(\omega)}^n$. C'est une représentation $\widehat{\Lambda}_{\times} : G \rightarrow \text{Aut}(L^2(\mathbb{R}_{(\omega)}^n))$ satisfaisant

$$\widehat{\Lambda}_{\times}(a, \gamma) \widehat{f}(\omega) = \sqrt{|\det(\tilde{\gamma})|} e^{-2\pi i \langle \omega, a \rangle} \widehat{f}(\gamma^*(\omega)) .$$

En effet, elle s'obtient

- (i) en passant de $\widehat{f}(\omega)$ à $f(a)$ par transformée de Fourier inverse,
- (ii) en appliquant $\Lambda_{\times}(a, \gamma)$,
- (iii) en prenant la transformée de Fourier.

Pour (i), on a

$$f(b) = N \int e^{2\pi i \langle b, \omega \rangle} \widehat{f}(\omega) d\omega$$

(N est le facteur de normalisation choisi). Pour (ii), on obtient

$$\Lambda_{\times}(a, \gamma) f(b) = \frac{N}{\sqrt{|\det(\tilde{\gamma})|}} \int e^{2\pi i \langle \tilde{\gamma}^{-1}(b-a), \varpi \rangle} \widehat{f}(\varpi) d\varpi$$

comme fonction de b . On en dérive pour (iii)

$$\frac{1}{N} \frac{N}{\sqrt{|\det(\tilde{\gamma})|}} \int db e^{-2\pi i \langle \omega, b \rangle} \int e^{2\pi i \langle \tilde{\gamma}^{-1}(b-a), \varpi \rangle} \widehat{f}(\varpi) d\varpi$$

comme fonction de ω . Mais comme par dualité

$$\langle \tilde{\gamma}^{-1}(b-a), \varpi \rangle = \langle b-a, (\tilde{\gamma}^{-1})^*(\varpi) \rangle,$$

(iii) s'écrit $\frac{1}{\sqrt{|\det(\tilde{\gamma})|}} \int \int e^{-2\pi i \langle b, \omega \rangle} e^{2\pi i \langle b-a, (\tilde{\gamma}^{-1})^*(\varpi) \rangle} \widehat{f}(\varpi) d\varpi db$, soit

$$\frac{1}{\sqrt{|\det(\tilde{\gamma})|}} \int \int e^{-2\pi i \langle a, (\tilde{\gamma}^{-1})^*(\varpi) \rangle} \widehat{f}(\varpi) e^{2\pi i \langle b, -\omega + (\tilde{\gamma}^{-1})^*(\varpi) \rangle} d\varpi db.$$

Mais

$$\int e^{2\pi i \langle b, -\omega + (\tilde{\gamma}^{-1})^*(\varpi) \rangle} db = \delta(-\omega + (\tilde{\gamma}^{-1})^*(\varpi))$$

impose $\varpi = \tilde{\gamma}^*(\omega)$. Ce changement de variable introduit un $|\det(\tilde{\gamma}^*)| = |\det(\tilde{\gamma})|$ et l'on obtient donc en définitive

$$\sqrt{|\det(\tilde{\gamma})|} e^{-2\pi i \langle a, \omega \rangle} \widehat{f}(\tilde{\gamma}^*(\omega)).$$

On retrouve le fait que par Fourier les translations spatiales se transforment en modulations par des facteurs de phase.

4.3.2. *L'exemple de $\mathbb{V}_J = \mathbb{H}_{pol}$.*

Pour $\mathbb{V}_J = \mathbb{R}_{(a)}^2 \rtimes \mathbb{R}_{(p)}$ ($a = (x, y)$), c'est p qui agit sur $\mathbb{R}_{(a)}^2$ par

$$\tilde{p}(x, y) = (x, y + px).$$

La matrice de $\tilde{p} = \begin{pmatrix} 1 & 0 \\ p & 1 \end{pmatrix}$ est de $\det(\tilde{p}) = 1$ et

$$\begin{aligned} \Lambda_{\times}((x, y), p) f(x', y') &= f(\tilde{p}^{-1}(x' - x, y' - y)) \\ &= f(x' - x, y' - y - p(x' - x)). \end{aligned}$$

Pour $\widehat{\Lambda}_{\times}$, on utilise le fait que

$$\langle \tilde{p}(x, y), (\xi, \eta) \rangle = \langle (x, y), \tilde{p}^*(\xi, \eta) \rangle$$

implique

$$\tilde{p}^*(\xi, \eta) = (\xi + p\eta, \eta)$$

et que par conséquent

$$\widehat{\Lambda}_{\times}((x, y), p) \widehat{f}(\xi, \eta) = e^{-2\pi i(x\xi + y\eta)} \widehat{f}(\xi + p\eta, \eta).$$

4.3.3. *L'exemple de $\mathbb{V}_S = SE(2)$.*

Dans le cas de $\mathbb{V}_S = SE(2) = \mathbb{R}_{(a)}^2 \rtimes \mathbb{S}_{(\theta)}^1$, $\mathbb{S}_{(\theta)}^1$ agit sur le plan $\mathbb{R}_{(a)}^2$ par les rotations $\tilde{\theta} = r_\theta$ d'angle θ qui sont de déterminant $\det(\tilde{\theta}) = 1$. On a donc

$$\begin{aligned} \Lambda_{\rtimes}((x, y), \theta) f(x', y') &= f(\tilde{\theta}^{-1}(x' - x, y' - y)) \\ &= f\left(\begin{matrix} (x' - x) \cos(\theta) + (y' - y) \sin(\theta), \\ -(x' - x) \sin(\theta) + (y' - y) \cos(\theta) \end{matrix}\right). \end{aligned}$$

Pour $\widehat{\Lambda}_{\rtimes}$, on utilise le fait que

$$\langle \tilde{\theta}(x, y), (\xi, \eta) \rangle = \langle (x, y), \tilde{\theta}^*(\xi, \eta) \rangle$$

implique

$$\tilde{\theta}^*(\xi, \eta) = (\xi \cos(\theta) + \eta \sin(\theta), -\xi \sin(\theta) + \eta \cos(\theta)) = \widetilde{(-\theta)}(\xi, \eta)$$

et que par conséquent

$$\widehat{\Lambda}_{\rtimes}((x, y), \theta) \widehat{f}(\xi, \eta) = e^{-2\pi i(x\xi + y\eta)} \widehat{f}(\xi \cos(\theta) + \eta \sin(\theta), -\xi \sin(\theta) + \eta \cos(\theta)).$$

4.4. **Unirreps de carré intégrable de produits semi-directs**

Pour les produits semi-directs Davide Barbieri et Giovanna Citti ont développé dans [27] et [28] une méthode originale pour construire des unirreps de carré intégrable de $G = \mathbb{R}_{(a)}^n \rtimes \Gamma(\gamma)$. L'idée de base est de restreindre la représentation quasi-régulière

$$\widehat{\Lambda}_{\rtimes} : G \longrightarrow \text{Aut}(L^2(\mathbb{R}_{(\omega)}^n))$$

sur l'espace de Fourier $\widehat{\mathbb{R}_{(a)}^n} = \mathbb{R}_{(\omega)}^n$ à des orbites de l'action de G sur $\mathbb{R}_{(\omega)}^n$. Si $\omega_0 \in \mathbb{R}_{(\omega)}^n$, son orbite $\mathcal{O}_{\omega_0} = \Gamma(\omega_0)$ est isomorphe au quotient Γ/S_{ω_0} de Γ par le stabilisateur de ω_0 . Notons γ_{ω_0} ses éléments (qui sont donc des classes d'équivalence modulo S_{ω_0}). On montre que l'espace de Hilbert $L^2(\mathbb{R}_{(\omega)}^n, d\omega)$ se décompose en la "somme directe" des Hilberts $L^2(\mathcal{O}_{\omega_0} d\mu_{\omega_0})$ où la mesure $d\mu_{\omega_0}$ sur \mathcal{O}_{ω_0} est induite par la mesure de Haar $d\mu(\gamma)$ sur Γ .

Remarque. La définition de cette "somme directe" soulève des problèmes subtils d'analyse fonctionnelle. Il s'agit en fait d'une intégrale de Hilberts paramétrés par un espace Σ sélectionnant de façon correcte un ω_0 dans chaque orbite, Σ étant munie d'une mesure $d\sigma$ convenable. Étant donné un ω , on note $\Sigma(\omega)$ l'élément de son orbite sélectionné par Σ . □

Si l'on restreint $\widehat{\Lambda}_x$ à \mathcal{O}_{ω_0} , on obtient une représentation ρ_{ω_0} de G dans le Hilbert $\mathcal{H}_{\omega_0} = L^2(\mathcal{O}_{\omega_0}, d\mu_{\omega_0})$ au moyen de la formule ("det $_{\omega_0}$ " à la place de "det" tient compte de la restriction à l'orbite)

$$\rho_{\omega_0}(g = (a, \gamma)) \psi(\gamma'_{\omega_0}) = \sqrt{\left| \det_{\omega_0}(\tilde{\gamma}) \right|} e^{-2\pi i \langle \tilde{\gamma}'^*(\omega_0), a \rangle} \psi((\gamma'\gamma)_{\omega_0}) .$$

On montre que l'on obtient ainsi des unirreps de carré intégrable. Si φ_0 est admissible, on peut alors construire un état cohérent $\varphi_g = \rho_{\omega_0}(g) \varphi_0$ et analyser les $\psi(\gamma_{\omega_0}) \in \mathcal{H}_{\omega_0}$ au moyen des

$$T_{\psi}(g) = \langle \psi, \varphi_g \rangle_{\mathcal{H}_{\omega_0}} = \sqrt{\left| \det_{\omega_0}(\tilde{\gamma}) \right|} \int_{\mathcal{O}_{\omega_0}} \psi(\gamma'_{\omega_0}) e^{2\pi i \langle \tilde{\gamma}'^*(\omega_0), a \rangle} \overline{\varphi((\gamma'\gamma)_{\omega_0})} d\mu_{\omega_0}(\gamma'_{\omega_0}) ,$$

la synthèse s'effectuant au moyen de la formule

$$\psi(\gamma_{\omega_0}) = \frac{1}{C_{\Phi_0}} = \int_G T_{\psi}(g) \varphi_g(\gamma_{\omega_0}) d\mu(g)$$

avec $C_{\varphi_0} = \int_G |\langle \varphi_0, \varphi_g \rangle_{\mathcal{H}}|^2 d\mu(g)$.

Il est alors intéressant de calculer, comme l'a fait Barbieri, la transformée de Fourier de $T_{\psi}(g)$ par rapport au domaine de Fourier $\mathbb{R}_{(a)}^n$, ce qui fait passer d'une distribution $T_{\psi}(a, \gamma)$ à une distribution $\widehat{T}_{\psi}(\omega, \gamma)$. Étant donnée la forme de $T_{\psi}(a, \gamma)$, le terme

$$e^{-2\pi i \langle \omega, a \rangle} e^{2\pi i \langle \tilde{\gamma}'^*(\omega_0), a \rangle} da$$

va introduire un $\delta(\omega - \tilde{\gamma}'^*(\omega_0))$ lors de l'intégration sur a et l'on trouve par conséquent

$$\widehat{T}_{\psi}(\omega, \gamma) = \sqrt{|\det_{\omega_0}(\tilde{\gamma})|} \int_{\mathcal{O}_{\omega_0}} \psi(\gamma'_{\omega_0}) \overline{\varphi((\gamma'\gamma)_{\omega_0})} \delta(\omega - \tilde{\gamma}'^*(\omega_0)) d\mu_{\omega_0}(\gamma'_{\omega_0}) .$$

Mais la distribution $\delta(\omega - \tilde{\gamma}'^*(\omega_0))$ s'interprète en disant

- (i) que ω doit être sur l'orbite \mathcal{O}_{ω_0} autrement dit que l'on doit avoir $\Sigma(\omega) = \omega_0$ et
- (ii) que l'élément γ' doit être celui qui envoie ω_0 sur ω .

Donc

$$\delta(\omega - \tilde{\gamma}'^*(\omega_0)) = \delta(\Sigma(\omega) - \omega_0) \delta(\omega - \tilde{\gamma}'^*(\omega_0)) .$$

Si l'on note $\gamma^{\omega_0 \rightarrow \omega}$ cet élément on obtient, en utilisant $\delta(\omega - \tilde{\gamma}'^*(\omega_0))$ dans l'intégrale, la formule

$$\widehat{T}_{\psi}(\omega, \gamma) = \sqrt{|\det_{\omega_0}(\tilde{\gamma})|} \psi(\gamma^{\omega_0 \rightarrow \omega}) \overline{\varphi((\gamma^{\omega_0 \rightarrow \omega} \gamma)_{\omega_0})} \delta(\Sigma(\omega) - \omega_0) .$$

Le résultat principal est alors que, si l'ondelette mère $\varphi_0 \in \mathcal{H}_{\omega_0}$ est de norme 1, $T_{\psi}(g)$ est une isométrie de l'espace des $\psi \in \mathcal{H}_{\omega_0} = L^2(\mathcal{O}_{\omega_0}, d\mu_{\omega_0})$ dans un sous

espace $L^2(\Gamma/S_{\omega_0}, \mathfrak{H}_{\omega_0})$ de $L^2(G)$, nommément celui des distributions qui appartiennent à $L^2(\Gamma/S_{\omega_0})$ pour la variable γ et à l'espace des transformées de Fourier de $L^2(\mathcal{O}_{\omega_0})$ pour la variable a .

4.5. Généralisations

Dans certains cas, comme nous allons le voir, pour obtenir une théorie satisfaisante, il est nécessaire d'indexer les états cohérents non pas par le groupe G lui-même mais par un de ses quotients $W = G/H$ par un sous-groupe fermé approprié. Si H n'est pas un sous-groupe normal de G , W n'est pas un groupe mais un espace homogène. On doit alors utiliser une section appropriée $\sigma : W \rightarrow G$ de la projection canonique $\Pi : G \rightarrow G/H$, c'est-à-dire sélectionner dans chaque classe d'équivalence w de W un élément $\sigma(w)$ tel que $\sigma(w)H = w$. L'état cohérent $\Phi = \{\varphi_w\}$ devient alors la famille $\varphi_w = \pi(\sigma(w))\varphi_0$. On relativise la condition d'admissibilité en travaillant sur W (i.e. modulo H). La covariance reste encore garantie par l'action de G sous-jacente.

Remarque. Un exemple élémentaire de cette situation est celui de la mécanique quantique où les fonctions d'ondes ψ sont définies seulement à un facteur de phase près et où l'on considère par conséquent des représentations "projectives". Nous en avons rencontré plus haut un exemple à propos des représentations de Gabor scalées. Cette procédure consiste à prendre pour H le sous-groupe laissant φ_0 invariante à un facteur de phase près. On a donc pour $h \in H$, $\pi(h)\varphi_0 = e^{i\alpha(h)}\varphi_0$ avec $\alpha(h)$ une fonction $H \rightarrow \mathbb{R}$ faisant de $\chi(h) = e^{i\alpha(h)}$ un caractère de H . \square

De façon encore plus générale, on peut étudier les situations où les φ_w sont paramétrées par un espace $(W, d\mu(w))$ muni d'une mesure $d\mu(w)$, les coefficients $T_f(w)$ étant donnés comme précédemment par les $\langle f, \varphi_w \rangle_{\mathcal{H}}$. On peut alors généraliser les concepts de "résolution de l'identité", de "noyau reproduisant" et de "représentation fonctionnelle" des \mathcal{H} en disant que l'intégrale sur W des opérateurs dans \mathcal{H} de projection Π_w sur les φ_w , i.e. $\mathfrak{A} = \int_W \Pi_w d\mu(w)$, est un opérateur "proche" de l'identité de \mathcal{H} au sens où il est positif, borné, auto-adjoint avec un inverse densément défini. Si \mathfrak{A}^{-1} est lui aussi borné on peut alors normaliser les φ_w par $\frac{1}{\sqrt{\mathfrak{A}}}$ (ce qui a un sens puisque $\frac{1}{\mathfrak{A}}$ est > 0 comme \mathfrak{A}) et se ramener à une résolution de l'identité.

5. Analyse harmonique et groupe de Heisenberg

Étant donné l'isomorphisme entre \mathbb{V}_J (groupe de Heisenberg polarisé \mathbb{H}_{pol}) et le groupe de Heisenberg non polarisé \mathbb{H} , l'analyse harmonique sur \mathbb{V}_J est la même que celle sur \mathbb{H} , bien connue en mécanique quantique et incroyablement étudiée, approfondie et généralisée.

Cela est très important. De même que la neurogéométrie plonge les modèles de connectivité neuronale dans l'univers mathématique de la géométrie de contact et

de la géométrie sous-riemannienne, de même elle plonge le traitement des signaux visuels dans l'univers physico-mathématique de *l'analyse harmonique non commutative* dominante en mécanique quantique.

L'analyse harmonique non commutative est une immense généralisation de la transformée de Fourier. Comme le formule très bien Michael Taylor dans l'Introduction de [524],

“The basic method of noncommutative harmonic analysis, generalizing the use of the Fourier transform, is to synthesize operators on a space on which a Lie group has a unitary representation from operators on irreducible representations spaces.”

On peut ainsi transférer au traitement du signal visuel par V1 de nombreux résultats de l'analyse harmonique sur \mathbb{H} (cf. par exemple les excellents ouvrages de Gerald Folland [187] : *Harmonic Analysis in Phase Space* et Vladimir Kisil [293] : *Wavelets in Applied and Pure Mathematics*, et aussi [292]). En particulier on peut appliquer le théorème fondamental de Stone–von Neumann.

5.1. Représentations de Schrödinger (Stone–von Neumann)

Nous avons déjà noté à la section 5.4 du chapitre 5 qu'au tout début de la mécanique quantique, Heisenberg était arrivé à la conclusion que les opérateurs auto-adjoints (les “observables”) de position $\hat{q} = Q$ et de moment $\hat{p} = P$ satisfaisaient la relation de commutation $[P, Q] = -i\hbar I$ (i.e. $[Q, P] = i\hbar I$). La représentation standard de cette algèbre de Lie était celle de Schrödinger où ces opérateurs agissent sur la fonction d'onde $\psi(q)$ du système quantique considéré respectivement comme la multiplication par q et l'opérateur de dérivation auto-adjoint $-i\hbar \frac{\partial}{\partial q}$.³⁵

Dans ce contexte, la version initiale du théorème de Stone–von Neumann, inspirée à Marshall Stone par les travaux de Hermann Weyl dans son *Gruppentheorie une Quantenmechanik* de 1928 [565],³⁶ disait que toute représentation irréductible unitaire ρ de \mathbb{H} dans un espace de Hilbert \mathcal{H} , si elle n'est pas triviale sur le centre Z de \mathbb{H} , est nécessairement unitairement équivalente (à travers un isomorphisme isométrique d'espaces de Hilbert $\mathcal{H} \simeq L^2(\mathbb{R}, \mathbb{C})$) à une représentation de Schrödinger sur les $\psi(q) \in L^2(\mathbb{R}, \mathbb{C})$. Cette représentation s'écrit traditionnellement en physique :

$$\rho_h(x, y, p)\psi(q) = e^{2i\pi hy} e^{2i\pi(hpD+xQ)}\psi(q)$$

où les facteurs $2i\pi$ et h viennent du formalisme quantique et où les opérateurs sont $D = \partial/\partial q$ ($P = -i\hbar D$) et $Q =$ multiplication par q , avec la relation de commutation

35. Rappelons qu'il n'existe pas de représentation de dimension finie puisque, en dimension finie n , $\text{trace}([P, Q]) = 0$ alors que $\text{trace}(i\hbar I) = i\hbar n$.

36. Article de Stone du 10 janvier 1930, voir la synthèse historique de Jonathan Rosenberg [472].

$[D, Q] = 1$. Comme $e^{2i\pi hp \partial / \partial q} \psi(q) = \psi(q + hp)$,³⁷ elle s'écrit aussi

$$\rho_h(x, y, p) \psi(q) = e^{2i\pi hy + 2i\pi xq + i\pi hp x} \psi(q + hp) \quad (\text{R1})$$

le terme $e^{i\pi hp x} = e^{\frac{1}{2}(2i\pi hp x)[D, Q]}$ venant de la formule de Baker-Campbell-Hausdorff. Donc ρ_h agit sur les $\psi(q)$ par translations $q \rightarrow q + hp$ et multiplications par des facteurs $e^{2i\pi xq}$ de module 1, avec des coefficients $e^{2i\pi hy + i\pi hp x}$. Il faut faire attention au fait que ce sont x et p les variables conjuguées et que y fonctionne comme une troisième variable de statut différent (elle parcourt le centre Z) dont le moment conjugué est *l'échelle* dans le groupe élargi dit de Weyl-Heisenberg affine.

La structure des représentations unitaires irréductibles est une conséquence d'un corollaire du lemme de Schur qui dit que si $\rho_1 : G \rightarrow GL(V_1)$ et $\rho_2 : G \rightarrow GL(V_2)$ sont deux représentations irréductibles de dimension finie d'un groupe G et si φ est une application linéaire de V_1 dans V_2 qui commute avec l'action de G (autrement dit si φ est un morphisme entrelaçant ρ_1 et ρ_2), alors soit φ est inversible, soit $\varphi = 0$. Si ρ est irréductible et unitaire et non triviale sur le centre Z , elle doit envoyer Z sur le cercle unité de \mathbb{C} et on doit donc avoir

$$\rho(0, y, 0) \psi(q) = e^{2i\pi hy} \psi(q)$$

pour un certain $h \in \mathbb{R}$. Si en revanche ρ est triviale sur Z alors elle se factorise par $\mathbb{H}/Z \simeq \mathbb{R}^2$ et, toujours d'après le lemme de Schur, elle est alors nécessairement de dimension 1 et de la forme

$$\rho(x, p) \psi(q) = e^{2i\pi(ax+bp)} \psi(q)$$

($a, b \in \mathbb{R}$).

Il faut être conscient du fait qu'un théorème d'équivalence de ce genre est techniquement tout à fait non trivial car les représentations considérées du groupe de Heisenberg sont de dimension nécessairement *infinie* et que les opérateurs P et Q sont nécessairement non bornés,³⁸ ce qui introduit de délicats problèmes de domaine de définition. D'où l'intérêt de la version *intégrale* des relations de Heisenberg obtenue par exponentiation.

5.2. Forme intégrale de Weyl

La version intégrale, due à Weyl, des relations de Heisenberg est plus maniable car elle permet d'éviter les difficultés liées au fait que P et Q sont nécessairement non bornés. Elle consiste à considérer dans \mathcal{H} les sous-groupes à 1-paramètre unitaires

37. Il s'agit toujours, avec de nouvelles variables, de l'opérateur de translation $T(t)f(x) = f(x-t)$, $T(t) = e^{-itD}$, $D = -i\partial_x$ (cf. section 1.3 ci-dessus et section 7.3.5 du chapitre 16).

38. Par exemple, si $\psi(q)$ est un vecteur propre de Q de valeur propre λ , $q\psi(q) = \lambda\psi(q)$, alors $(q+r)\psi(q+r) = \lambda\psi(q+r)$ et donc $q\psi(q+r) = (\lambda-r)\psi(q+r)$ ce qui montre que la translatée $\psi(q+r)$ est vecteur propre de Q de valeur propre $\lambda-r$. Le spectre de Q est \mathbb{R} tout entier et par conséquent, Q n'est pas borné.

$U(t) = \exp(itQ)$ et $V(s) = \exp(isP)$ (on pose $\hbar = 1$ pour simplifier et l'on a donc $[Q, P] = iI$). Leur relation de commutation s'obtient en appliquant la formule de Baker-Campbell-Hausdorff :

$$\begin{aligned} V(s)U(t) &= \exp(isP)\exp(itQ) = \exp(itQ)\exp(isP)\exp(-[itQ, isP]) \\ &= \exp(itQ)\exp(isP)\exp(-itisi) \\ &= U(t)V(s)\exp(its) \end{aligned}$$

Le théorème de Stone-von Neumann dit alors qu'il existe une transformation unitaire $\varphi : \mathcal{H} \xrightarrow{\simeq} L^2(\mathbb{R})$ telle que

$$\begin{cases} \varphi \circ U(t) \circ \varphi^{-1} = \exp(itq) \\ \varphi \circ V(s) \circ \varphi^{-1} = \exp\left(s\frac{\partial}{\partial q}\right) \end{cases}$$

avec, comme nous l'avons vu plus haut,

$$\exp\left(s\frac{\partial}{\partial q}\right)\psi(q) = \psi(q+s) .$$

Remarque. Nous avons déjà rencontré plus haut à la section 2.1.2 une variante de la forme intégrale de Weyl à propos de la transformée de Gabor. Si T_a est l'opérateur de translation

$$T_a h(x) = h(x-a)$$

et F_ω l'opérateur de modulation fréquentielle

$$F_\omega h(x) = e^{i\omega x} h(x) ,$$

on a

$$T_a \circ F_\omega = e^{-i\omega a} F_\omega \circ T_a .$$

5.3. Versions du théorème de Stone-von Neumann

Transformons ces formules classiques pour passer du cas \mathbb{H} en mécanique quantique au cas \mathbb{V}_J en neurogéométrie. Nous remplacerons $\psi(q)$ par $u(s)$ et h par λ pour ne plus évoquer par nos notations les fonctions d'onde de Schrödinger. D'après ce que nous avons vu à la section 5.3 du chapitre 5, on passe de \mathbb{H} à \mathbb{V}_J en passant des coordonnées exponentielles canoniques de première espèce à celle de deuxième espèce, autrement dit en faisant les substitutions $x \rightsquigarrow x$, $y + \frac{1}{2}xp \rightsquigarrow y$, $p \rightsquigarrow p$. La formule devient donc

$$\rho_\lambda(x, y, p)u(s) = e^{2i\pi\lambda y + 2i\pi x s} u(s + \lambda p) . \quad (\text{R2})$$

On peut ensuite par une homothétie de 2π sur les coordonnées (x, y, p) de \mathbb{V}_J , éliminer le facteur 2π , ce qui donne

$$\rho_\lambda(x, y, p)u(s) = e^{i(\lambda y + xs)}u(s + \lambda p) . \quad (\text{R3})$$

On trouve souvent cette forme dans la littérature avec $\lambda = \hbar$. On peut enfin utiliser le changement de variable $p \rightsquigarrow \frac{p}{\lambda}$, $x \rightsquigarrow \lambda x$ qui est un automorphisme $\beta(v)$ de \mathbb{V}_J puisque

$$\begin{aligned} \beta(v.v') &= \left(\lambda(x+x'), y+y' + \frac{\lambda}{\lambda}px', \frac{p+p'}{\lambda} \right) \\ &= \left(\lambda(x+x'), y+y' + px', \frac{p+p'}{\lambda} \right) \\ \beta(v).\beta(v') &= \left(\lambda x, y, \frac{p}{\lambda} \right) \cdot \left(\lambda x', y', \frac{p'}{\lambda} \right) \\ &= \left(\lambda x + \lambda x', y+y' + \frac{p}{\lambda}\lambda x', \frac{p}{\lambda} + \frac{p'}{\lambda} \right) \\ &= \left(\lambda(x+x'), y+y' + px', \frac{p+p'}{\lambda} \right) . \end{aligned}$$

Cela permet de se ramener à la formule

$$\begin{aligned} \rho_\lambda : \mathbb{V}_J &\rightarrow \mathcal{U}(\mathcal{H}) \\ v &\mapsto \rho_\lambda(v) : \mathcal{H} &\rightarrow \mathcal{H} \\ & &u(s) \mapsto e^{i\lambda(y+xs)}u(s+p) . \end{aligned} \quad (\text{R4})$$

Remarque. On peut évidemment échanger les variables conjuguées x et p et considérer les représentations $u(r) \mapsto e^{i\lambda(y+pr)}u(r+x)$. \square

Bref, les briques constituant les unirreps de \mathbb{V}_J dans un Hilbert \mathcal{H} sont ainsi les $\rho(x)$ agissant par multiplication e^{ixs} , les $\rho(p)$ agissant par translation $u(s+p)$ et les $\rho(y)$ agissant par un facteur de phase e^{iy} , λ étant un *paramètre d'échelle* caractéristique de ρ_λ . Et le théorème de Stone-von Neumann implique que toutes les unirreps de \mathbb{V}_J dans un Hilbert qui ne sont pas triviales sur le centre Z (l'axe y) sont de dimension infinie et équivalentes à une représentation de Schrödinger $\rho_\lambda(x, y, p)$ ($\lambda \neq 0$) sur $\mathcal{H} = L^2(\mathbb{R}, \mathbb{C})$, ρ_λ étant bien non triviale sur $Z = \{(0, y, 0)\}$ puisque

$$\rho_\lambda(0, y, 0)u(s) = e^{i\lambda y}u(s) .$$

Insistons encore une fois sur le fait que la variable y agit comme un facteur de phase à travers la multiplication par le caractère $e^{i\lambda y}$ et que ce facteur est nécessaire dans la mesure où les actions de x et p sur les fonctions ne commutent pas. Si l'on travaille modulo ce facteur, on obtient alors des représentations projectives.

En ce qui concerne les unirreps triviales sur le centre Z , elles sont de dimension 1 et sont les caractères

$$\rho_{\mu,\nu}(x, y, p) u(s) = e^{i(\mu x + \nu p)} u(s) .$$

Leur théorie se ramène à la transformée de Fourier.

La conséquence de tous ces résultats est que le dual de Pontryagin $\widehat{\mathbb{H}}_{pol}$ du groupe de Heisenberg polarisé $\mathbb{H}_{pol} = \mathbb{V}_J$ est l'union de la famille à un paramètre des ρ_λ ($\lambda \neq 0$) de dimension infinie et de la famille à deux paramètres des $\rho_{\mu,\nu}$ unidimensionnelles. Son espace de paramétrage est donc $(\mathbb{R} - \{0\}) \cup \mathbb{R}^2$ la limite des ρ_λ pour $\lambda \rightarrow 0$ étant d'une certaine façon les $\rho_{\mu,\nu}$. Cette limite est un peu difficile à intuitionner. Pour $\lambda \rightarrow 0$, la représentation $\rho_\lambda(x, y, p) : u(s) \mapsto e^{i\lambda(y+xs)} u(s+p)$ tend intuitivement vers la représentation régulière $\rho_0 : u(s) \mapsto u(s+p)$ de $\mathbb{R}_{(p)}$. Mais, dans l'espace des représentations de \mathbb{H}_{pol} , celle-ci n'est pas du tout irréductible. Elle est même très réductible et se décompose en les $\rho_{\mu,\nu}$, ce qui correspond, comme nous le verrons, au fait que l'orbite co-adjointe associée qui est un *plan* pour $\lambda \neq 0$ se décompose pour $\lambda = 0$ en *points* (μ, ν) . En fait, ρ_0 est une intégrale de $\rho_{\mu,\nu}$ et, plus précisément, $\rho_0 = \int_{\mathbb{R}} \rho_{0,\nu} d\nu$.

Remarque. Tous ces résultats sur le groupe de Heisenberg se généralisent de façon remarquable aux groupes de Carnot présentés à la section 8 du chapitre 6. Le lecteur intéressé pourra consulter à ce sujet le grand ouvrage de référence [61] d'Andrea Bonfiglioli, Ermanno Lanconelli et Francesco Uguzzoni. □

5.4. Irréductibilité des représentations de Schrödinger

Avant d'esquisser la preuve du théorème de Stone-von Neumann, vérifions que la représentation unitaire

$$\rho_\lambda(v) = e^{i\lambda(y+xs)} u(s+p)$$

($\lambda = 1$ pour simplifier) est bien irréductible. On utilise le lemme de Schur, déjà cité plus haut, disant que l'irréductibilité équivaut au fait que si A est un opérateur dans $\mathcal{L}(\mathcal{H})$ commutant avec tous les $\rho_\lambda(v)$ alors A est *scalaire*. Supposons donc d'abord que A commute avec les multiplications : $Ae^{ixs} = e^{ixs} A$. On utilise Fourier en considérant les $v(s)$ comme les transformées de Fourier inverses $v(s) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixs} \widehat{v}(x) dx$ et on en conclut

$$\begin{aligned} v(s) Au(s) &= \frac{1}{\sqrt{2\pi}} \int e^{ixs} \widehat{v}(x) Au(s) dx = \frac{1}{\sqrt{2\pi}} \int \widehat{v}(x) Ae^{ixs} u(s) dx \\ &= Av(s) u(s) . \end{aligned}$$

En appliquant cette égalité à des fonctions caractéristiques bien choisies, on en déduit qu'il existe une fonction $\alpha(s) \in L^\infty(\mathbb{R})$ telle que $Au(s) = \alpha(s) u(s)$. Mais supposons ensuite que A commute aussi avec les translations et que donc $\alpha(s)$ est invariante par translations. Cela implique que $\alpha(s)$ est une constante α et donc

$Au(s) = \alpha u(s)$ agit bien de façon scalaire. D'après le lemme de Schur, A est irréductible.

5.5. Reconstruction du groupe de Heisenberg

Nous venons d'expliciter les unirreps du groupe de Heisenberg. Vérifions, pour le fun et à titre d'exercice scolaire, qu'en sens inverse on peut reconstruire le groupe \mathbb{H}_{pol} à partir de groupes unitaires à 1-paramètres de $L^2(\mathbb{R}_{(s)})$. Cela précisera ce que nous avons vu à la section 2.1.2 à propos de la transformée de Gabor.

Le premier groupe à 1-paramètre est celui des translations

$$T_p u(s) = u(s+p) ,$$

$u(s) \in L^2(\mathbb{R}_s)$.³⁹ L'inverse $(T_p)^{-1}$ de T_p est T_{-p} et l'adjoint de T_p satisfait

$$\langle T_p u, v \rangle = \langle u, T_p^* v \rangle = \int_{\mathbb{R}_s} u(s+p) \overline{v(s)} ds$$

ce qui, par le changement de variable trivial $s+p = s'$ donne

$$\int_{\mathbb{R}_{s'}} u(s') \overline{v(s'-p)} ds' = \langle u, T_{-p} v \rangle .$$

Donc

$$T_{-p} = (T_p)^{-1} = T_p^*$$

et T_p est bien unitaire. Nous avons vu à la section 1.3 que son générateur infinitésimal est l'opérateur anti-hermitien ∂_s et que donc

$$T_p = e^{p\partial_s} = e^{ipD_s}$$

avec D_s l'opérateur auto-adjoint $D_s = -i\partial_s$.

Le second groupe à 1-paramètre est celui des multiplications par un facteur de phase (caractère)

$$M_x u(s) = e^{ixs} u(s) .⁴⁰$$

Son générateur infinitésimal est la multiplication par is . Les produits satisfont la relation intégrale de Weyl

$$T_p \circ M_x = e^{ixp} M_x \circ T_p$$

dont la version infinitésimale est la relation de Heisenberg

$$[pD_s, xsI] = -ipxI .$$

39. Comme nous l'avons déjà noté plusieurs fois, l'on peut, si l'on préfère, prendre $T_p u(s) = u(s-p)$ en utilisant l'automorphisme involutif de $\mathbb{R}_{(s)}$, $s \mapsto -s$. C'est ce que nous avons fait à la section 2.1.2.

40. Dans la section 2.1.2 les notations étaient T_a et F_ω .

La relation de Weyl conduit ainsi à regarder de façon générale des produits de la forme $e^{iy}e^{ixs}e^{ipD_s}$ (ci-dessus on a le cas $y = xp$). Cela donne immédiatement la loi de groupe puisque

$$e^{iy_1}e^{ix_1s}e^{ip_1D_s}e^{iy_2}e^{ix_2s}e^{ip_2D_s} = e^{i(y_1+y_2)}e^{ix_1s} \left(e^{ip_1D_s}e^{ix_2s} \right) e^{ip_2D_s}$$

car les y agissent de façon scalaire

$$\begin{aligned} &= e^{i(y_1+y_2)}e^{ix_1s} \left(e^{ix_2p_1}e^{ix_2s}e^{ip_1D_s} \right) e^{ip_2D_s} \\ &= e^{i(y_1+y_2+x_2p_1)}e^{i(x_1+x_2)s}e^{i(p_1+p_2)D_s} \end{aligned}$$

par Weyl et donc

$$(y_1, x_1, p_1)(y_2, x_2, p_2) = (y_1 + y_2 + p_1x_2, x_1 + x_2, p_1 + p_2) .$$

5.6. Esquisse de preuve du théorème

La preuve du théorème de Stone-von Neumann repose sur la théorie des opérateurs élaborée par von Neumann ⁴¹ ⁴², le théorème de Stone et les profonds travaux

41. Reparcourir la vie intellectuelle géniale de John von Neumann (1903-1957), c'est reparcourir une grande partie de l'histoire scientifique et politique du XX^e siècle, des années 1920 à la guerre froide. Issu d'une famille aisée de l'intelligentsia hongroise, enfant prodige, formé entre autres à Göttingen où il subit une influence forte et durable de Hilbert, lié à Werner Heisenberg, Robert Oppenheimer et Kurt Gödel, il émigre en 1930 et se fixe à Princeton. C'est l'un des grands savants de cette période. Ce qui nous intéresse ici est son travail sur la théorie des opérateurs, démarré dès 1926 et exposé dans son traité classique de 1932 *Mathematische Grundlagen der Quantenmechanik* [388]. Dans ce domaine, après Hilbert et Banach, von Neumann a édifié l'analyse fonctionnelle des algèbres non commutatives d'opérateurs qui est indispensable à la physique quantique. Mais von Neumann a aussi fondé avec Birkhoff le champ de la logique quantique où les sous-ensembles sur lesquels telle ou telle propriété est vraie sont des sous-espaces fermés d'un espace de Hilbert (l'espace des états du système considéré). Les opérations de la syntaxe logique s'interprètent de façon spécifique dans cette sémantique hilbertienne, ce qui fait que la logique quantique est très différente de la logique classique. Dès les années 1920, le jeune von Neumann joua un rôle important en théorie des ensembles en introduisant au moyen de "l'axiome de fondation" une hiérarchisation des procédés de construction des ensembles et en introduisant, avec Paul Bernays (disciple de Hilbert) et Kurt Gödel, la notion de "classes" (trop grosses pour être traitées comme des ensembles sans contradiction). Et il ne faut pas oublier que von Neumann est l'un des principaux concepteurs (après Turing, John Atanasoff, Konrad Zuse, puis l'E.N.I.A.C. de John W. Mauchly et John P. Eckert en 1945) des ordinateurs modernes, machines dont "l'architecture de von Neumann" donne un sens technique beaucoup plus opérationnel aux machines de Turing universelles abstraites. Son rapport de 1945 sur l'E.D.V.A.C. a lancé cette épopée techno-scientifique. Un autre aspect des recherches de von Neumann qui croise les thèmes de la neurogéométrie sont sa participation aux célèbres Conférences Macy (de 1946 à 1950) où, avec entre autres Walter Pitts, Warren McCulloch, Norbert Wiener et Julian Bigelow, se mirent en place la cybernétique et la théorie des réseaux de neurones à l'intersection des mathématiques appliquées, de la physique statistique, de l'informatique et de la neurobiologie. Pour une réflexion approfondie sur l'importance des Conférences Macy, cf. l'ouvrage de Jean-Pierre Dupuy [160]. Dans un domaine proche, celui des automates cellulaires, von Neumann est celui qui, avec sa théorie des automates auto-reproducteurs (développée en partie avec Stanislaw Ulam entre 1948 et 1952), a le mieux anticipé la découverte du code génétique. Et enfin, il ne faut pas oublier la fondation de la théorie des jeux en économie mathématique avec Oskar Morgenstern.

42. Il faut en particulier souligner toute l'importance des algèbres de von Neumann W . Elles sont des algèbres de Banach particulières d'opérateurs bornés T dans des Hilberts \mathcal{H} , algèbres que l'on souhaite pouvoir interpréter de façon naturelle comme des algèbres d'observables portant sur des systèmes physiques. Techniquement, elles contiennent l'identité, sont stables pour l'involution (l'adjonction) $T \leftrightarrow T^*$ issue de la conjugaison dans \mathbb{C} (ce sont des $*$ -algèbres) et leur norme satisfait $\|T^*T\| = \|TT^*\| = \|T\|^2$. Elles sont les duales d'un "pré-dual" $W_* \subset W^*$ (i.e. une sous-algèbre de l'algèbre des formes linéaires continues sur W) qui est une algèbre de Banach. Elles sont fermées pour la topologie faible qu'elles héritent de cette dualité (qui est plus faible que la topologie de la norme $\|T\| = \sup \{|\lambda| \mid \lambda \in \text{Spec}(T)\}$) et elles contiennent beaucoup de *projecteurs*, propriété indispensable en théorie quantique où

de George Mackey⁴³. Elle utilise les notions techniques de “système d'imprimitivité” et de “représentation induite” sur lesquelles nous reviendrons plus en détail dans la section 6. Des références de base sont Stone [514], von Neumann [387], Weyl [565], Mackey [342]. Elle est assez difficile mais essayons quand même d'en donner une vague idée. Le lecteur intéressé pourra par exemple consulter le texte de Veeravalli Varadarajan [551] et l'ouvrage de Michael Taylor [524] (déjà souvent cité).

5.6.1. *Systèmes d'imprimitivité.*

La notion de système d'imprimitivité (SI) est ancien et remonte à la théorie des représentations des groupes G finis. Soit $\rho : G \rightarrow GL(V)$ une représentation du groupe fini G dans un espace vectoriel V de dimension finie. ρ opère par permutations sur l'ensemble \mathfrak{W} des sous-espaces W de V par les $\rho(g)W$. Un SI est un sous-ensemble \mathfrak{X} de \mathfrak{W} constitué de W linéairement indépendants, permutés par G et décomposant V en somme directe $V = \bigoplus_{W \in \mathfrak{X}} W$. Le SI est dit “transitif” si l'action de G sur \mathfrak{X} est transitive. L'exemple le plus évident est la représentation régulière

les mesures sont des projections (“réduction du paquet d'ondes”) sur les sous-espaces propres des opérateurs considérés. La condition de l'existence d'un pré-dual W_* est lié à la volonté de pouvoir interpréter les opérateurs T de W comme opérant sur les “états” Ψ d'un système physique associé au moyen d'une dualité $\Psi(T)$ donnant la valeur moyenne (au sens des probabilités quantiques) de l'observable T mesurée lorsque le système est dans l'état Ψ . Les deux exemples de base sont les suivants. (i) Les $L^\infty(X)$ où X est un espace mesurable (exemple des W commutatives) et dans ce cas $L^\infty(X)_*$ est l'espace des mesures sur X (à valeurs dans \mathbb{C}). Le fait que les W aient beaucoup de projecteurs se voit alors clairement dans le fait que, si W est commutative, on peut construire un espace mesurable X tel que $W \simeq L^\infty(X)$. En effet, l'algèbre des projecteurs P_W de W est assez riche pour être l'algèbre de Boole mesurable des parties d'un ensemble X lui-même assez riche pour redonner tout W . (ii) Les $\mathcal{B}(\mathcal{H})$ (tous les opérateurs bornés sur un Hilbert, $M_n(\mathbb{C})$ en dimension finie) qui permettent d'obtenir toutes les W comme sous-algèbres de von Neumann. Dans ce cas de $\mathcal{B}(\mathcal{H})$, les états Ψ de W_* sont en correspondance avec les vecteurs ψ de \mathcal{H} et $T(\Psi) = \langle \psi, T(\psi) \rangle_{\mathcal{H}}$. Pour ces W non commutatives, les projecteurs permettent de construire un espace mesurable “non commutatif”. Les propriétés des projecteurs de certaines de ces algèbres ont conduit von Neumann à l'idée de nouvelles géométries dont la dimension peut varier de façon *continue*. La classification des W repose sur la classification des “facteurs” (les W maximalelement non commutatives dont le centre se réduit à \mathbb{C}) et a été effectuée par Francis Joseph Murray et John von Neumann avant d'être achevée, pour les cas les plus difficiles, par Alain Connes, ce qui a ouvert la voie à la géométrie non commutative.

43. George Whitelaw Mackey (1916-2006) fut un doctorant de Marshall Stone à Harvard en 1942 et fit essentiellement carrière dans cette université. Ses concepts de “système d'imprimitivité” et de “représentation induite” ont joué un rôle décisif dans l'évolution de la théorie des représentations, en particulier des produits semi-directs. Ses travaux sur les représentations des groupes de Lie nilpotents utilisent la “méthode des orbites” de Kirillov. Son ouvrage de 1976, *The Theory of Unitary Group Representations* [343], est une référence.

de G sur l'espace $\mathcal{H} = L^2(G) \simeq \mathbb{C}^G$ des fonctions $\varphi : G \rightarrow \mathbb{C}$ avec le produit scalaire

$$\langle \varphi, \psi \rangle = \sum_{g \in G} \varphi(g) \overline{\psi(g)} .$$

En effet si W_g est le sous espace $\mathbb{C}\chi(\{g\})$ engendré par la fonction caractéristique $\chi(\{g\})$ de g , $L^2(G) = \bigoplus_{g \in G} W_g$ qui sont bien linéairement indépendants et permutés par G puisque $g'W_g = W_{g'g}$.

Soit alors G_0 un sous-groupe de G et ρ_0 une représentation de G_0 . La représentation ρ est dite “induite” par ρ_0 s'il existe un SI transitif (ρ, \mathfrak{X}) et $W_0 \in \mathfrak{X}$ tel que G_0 soit le sous-groupe laissant W_0 invariant (i.e. G_0 est l'ensemble des g tels que $gW_0 \subseteq W_0$) et ρ_0 est la restriction de ρ à W_0 . Cela signifie que “l'induction” est l'opération duale de la “restriction”. Nous y reviendrons plus en détails à la section 6.3. Cette “induction” est bien définie. En effet si ρ est induite par ρ_0 , son caractère $\chi_\rho(g) = \text{tr}(\rho(g))$ est la moyenne sur G_0 du caractère χ_{ρ_0} de ρ_0 sur les $g_0 \in G_0$ conjugués de g :

$$\chi_\rho(g) = \text{tr}(\rho(g)) = \frac{1}{|G_0|} \sum_{h \in G, h^{-1}gh = g_0 \in G_0} \chi_{\rho_0}(g_0) .$$

Lorsque l'on généralise aux groupes de Lie et à leurs unirreps dans des Hilberts \mathcal{H} , il faut passer, comme nous allons le voir, de décompositions en somme directe $V = \bigoplus_{W \in \mathfrak{X}} W$ à des mesures σ définies sur un espace M muni de sa σ -algèbre borélienne \mathfrak{B} et d'une mesure μ sur laquelle G opère de façon continue (mesurable peut suffire) par $g\mu(B) = \mu(g^{-1}B)$, σ étant une mesure à valeurs dans les projecteurs de \mathcal{H} (PVM). Un SI (G, M, σ) sera alors la donnée d'une ρ unitaire (bien sûr fortement continue) $\rho : G \rightarrow \mathcal{U}(\mathcal{H})$ (le groupe unitaire de \mathcal{H}) et d'une PVM σ sur M satisfaisant, pour tout borélien $B \in \mathfrak{B}$ de M ,

$$\rho(g) \sigma(B) \rho(g)^{-1} = \sigma(g(B)) .$$

Il sera dit “irréductible” si les seuls sous-espaces fermés de \mathcal{H} invariants par les $\rho(g)$ et les $\sigma(B)$ sont $\{0\}$ et \mathcal{H} .

Si μ est σ -finie (au sens de la section 2.1) et G -invariante (i.e. $\mu(g^{-1}B) = \mu(B)$ pour tout borélien B), si $\mathcal{H} = L^2(M, \mu)$ et si $\sigma_{res}(B)$ est le projecteur de restriction des fonctions de \mathcal{H} à B (i.e. la multiplication des fonctions par la fonction caractéristique de B), alors, pour la représentation régulière $(\rho_{reg}(g)f)(x) = f(g^{-1}x)$ ($x \in M$), $(\rho_{reg}, \sigma_{res})$ est un SI.

Pour avancer dans la théorie, supposons que G agisse *transitivement* sur M et que μ soit une mesure *invariante* par G . Étant donné un Hilbert \mathcal{H}^\bullet , on peut alors définir une représentation unitaire $\rho_G : G \rightarrow \mathcal{U}(\mathcal{H}_0)$, où \mathcal{H}_0 est le Hilbert des $\psi : M \rightarrow \mathcal{H}^\bullet$

appropriées⁴⁴, à condition de se donner un *cocycle unitaire* $C : M \times G \rightarrow \mathcal{U}(\mathcal{H}^\bullet)$, le terme cocycle signifiant que $C(x, e) = Id_{\mathcal{H}}$ et

$$C(x, g'g) = C(x, g) C(gx, g') .$$

La représentation unitaire ρ_G est définie par

$$\rho_G(g) \psi(x) = C(g^{-1}x, g) \psi(g^{-1}x)$$

pour les $\psi : M \rightarrow \mathcal{H}^\bullet$ et l'on vérifie que la condition de cocycle signifie simplement que ρ_G est bien une représentation. En effet

$$\begin{aligned} \rho_G(g'g) \psi(x) &= C\left((g'g)^{-1}x, g'g\right) \psi\left((g'g)^{-1}x\right) \\ &\quad \text{par définition} \\ &= C\left((g'g)^{-1}x, g\right) C\left(g(g'g)^{-1}x, g'\right) \psi\left((g'g)^{-1}x\right) \\ &\quad \text{par la propriété de cocycle} \\ &= C\left((g'g)^{-1}x, g\right) C(g^{-1}x, g') \psi\left((g'g)^{-1}x\right) . \end{aligned}$$

Par ailleurs, on a, puisque $\rho_G(g') \psi(x) = C(g'^{-1}x, g') \psi(g'^{-1}x)$,

$$\begin{aligned} \rho_G(g) \rho_G(g') \psi(x) &= \rho_G(g) \left(C(g'^{-1}x, g') \psi(g'^{-1}x) \right) \\ &= C(g^{-1}g'^{-1}x, g) \left(C(g'^{-1}x, g') \psi(g'^{-1}x) \right) \end{aligned}$$

et donc

$$\rho_G(g'g) = \rho_G(g) \rho_G(g') .$$

Si $\sigma_{res}(B)$ est alors, comme précédemment, le projecteur de restriction des fonctions de \mathcal{H}_0 au borélien B de M , (ρ_G, σ_{res}) est un SI.

Nous en détaillerons plus bas un exemple dans les sections 6 et surtout 6.3, avec $M = G/H$ l'espace homogène associé à un sous-groupe fermé de G . Dans ce cas il y a une bijection préservant l'irréductibilité entre les classes d'équivalence de représentations unitaires ρ_H de H et les SI ci-dessus. On peut par conséquent construire des représentations induites ρ_G à partir des ρ_H .

5.6.2. La preuve.

Le centre Z de \mathbb{H}_{pol} est la droite des y et toute unirrep ρ de \mathbb{H}_{pol} dans un Hilbert \mathcal{H} doit agir scalairement sur Z . Si ρ est triviale sur Z , elle est en fait une unirrep de $\mathbb{H}_{pol}/Z = \mathbb{R}_{(x,p)}^2$ qui est un groupe abélien et elle donc de dimension 1. Cela donne les

$$\rho_{\mu,\nu}(x, y, p) u(s) = e^{i(\mu x + \nu p)} u(s) .$$

44. En analogie avec la somme directe $L^2(G) = \bigoplus_{g \in G} W_g$ du cas fini, on peut interpréter \mathcal{H}_0 comme $\int_M^\oplus \mathcal{H}_x^\bullet d\mu(x)$ (où tous les \mathcal{H}_x^\bullet sont des exemplaires de \mathcal{H}).

Si ρ n'est pas triviale sur Z , alors, restreinte à Z , elle est un caractère

$$\rho(0, y, 0) u(s) = e^{i\lambda y} u(s) .$$

Ce λ va paramétrer ρ et en faire une ρ_λ . En utilisant les dilatations de \mathbb{H}_{pol} (cf. la section 8 du chapitre 6) on peut se ramener au cas $\lambda = 1$. Soit alors $V(x) = \rho_1(x, 0, 0)$ et $W(p) = \rho_1(0, 0, p)$ les groupes à 1-paramètre unitaires restreignant ρ_1 aux axes x et p de \mathbb{H}_{pol} . $V(x)$ et $W(p)$ sont des représentations unitaires respectivement de $\mathbb{R}_{(x)}$ et de $\mathbb{R}_{(p)}$. Puisque $(x, y, p) = (0, y, 0)(x, 0, 0)(0, 0, p)$, on a $\rho_1(x, y, p) = e^{iy} V(x) W(p)$ avec la relation de Weyl $W(p) V(x) = e^{ixp} V(x) W(p)$, i.e.

$$W(p) V(x) W(p)^{-1} = e^{ixp} V(x) .$$

Le multiplicateur e^{ixp} met $\mathbb{R}_{(x)}$ et $\mathbb{R}_{(p)}$ en dualité et l'on peut identifier $\mathbb{R}_{(p)}$ à $\widehat{\mathbb{R}_{(x)}}$.⁴⁵

On applique alors le théorème de Stone de la section 7.5 à $V(x)$ et à l'opérateur auto-adjoint $A = i\mathbf{v}$ associé à son générateur infinitésimal \mathbf{v} . Supposons d'abord A de spectre simple. Il existe une mesure σ_A de support $\text{Spec}(A) \subset \mathbb{R}_{(r)}$ (r est la variable des valeurs spectrales de A) à valeurs dans les projecteurs de \mathcal{H} telle que $V(x) = \int_{\mathbb{R}_{(r)}} e^{ixr} d\sigma_A(r)$. La relation de commutation de Weyl entre $V(x)$ et $W(p)$ s'interprète alors comme le fait que

$$\begin{aligned} W(p) V(x) W(p)^{-1} &= e^{ixp} V(x) W(p) W(p)^{-1} = e^{ixp} V(x) \\ &= e^{ixp} \int_{\mathbb{R}_{(r)}} e^{ixr} d\sigma_A(r) = \int_{\mathbb{R}_{(r)}} e^{ix(r+p)} d\sigma_A(r) , \end{aligned}$$

autrement dit que

$$W(p) \sigma_A(B) W(p)^{-1} = \sigma_A(B + p)$$

pour les boréliens B de $\mathbb{R}_{(r)}$. Il est essentiel de remarquer, qu'à travers cette diagonalisation, l'espace $\mathbb{R}_{(x)}$ a disparu de la formule. Comme $\mathbb{R}_{(r)}$ est identifiable au dual $\widehat{\mathbb{R}_{(x)}}$ de $\mathbb{R}_{(x)}$, on peut l'identifier à $\mathbb{R}_{(p)}$ ce qui est bien compatible avec le fait que dans $r + p$ les variables r et p doivent être de même type.

Les translations par p agissent *transitivement* sur $\mathbb{R}_{(r)}$. On a donc un premier espace (un groupe) $\mathbb{R}_{(p)}$, une unirrep W de ce premier espace $\mathbb{R}_{(p)}$ dans \mathcal{H} , puis un deuxième espace linéaire $\mathbb{R}_{(r)}$ sur lequel le premier espace agit transitivement par des automorphismes $\tau_p(r)$ et même simplement transitivement car le stabilisateur des $\tau_p(r)$ est $\{0\}$ (les translations non nulles sont sans points fixes), et enfin une mesure σ sur ce second espace $\mathbb{R}_{(r)}$ à valeurs \mathcal{H} -projecteurs, ces données satisfaisant $\sigma(\mathbb{R}_{(r)}) = Id_{\mathcal{H}}$ et $W(p) \sigma(B) W(p)^{-1} = \sigma(\tau_p(B))$.

45. Ce multiplicateur e^{ixp} est, notons-le, comme le $e^{i\omega x}$ des transformées de Fourier.

Un tel système de données

$$\left\{ \begin{array}{l} G = \mathbb{R}_{(p)}, \\ W(p) : G = \mathbb{R}_{(p)} \rightarrow \mathcal{L}(\mathcal{H}), \\ M = \mathbb{R}_{(r)}, \\ \tau_p : G = \mathbb{R}_{(p)} \rightarrow \text{Aut}(M = \mathbb{R}_{(r)}), \\ \sigma : \mathfrak{B}(M = \mathbb{R}_{(r)}) \rightarrow \text{Pr}(\mathcal{H}), \\ \sigma(M = \mathbb{R}_{(r)}) = Id_{\mathcal{H}}, \\ W(p)\sigma(B)W(p)^{-1} = \sigma(\tau_p(B)) \end{array} \right.$$

est un cas particulier (contraint par les contraintes de transitivité et de nullité des stabilisateurs) de *système d'imprimitivité* au sens de Mackey.⁴⁶ Et un théorème fondamental, appelé "théorème d'imprimitivité de Mackey", dit que, dans ce cas particulier, il existe une équivalence unitaire U entre \mathcal{H} et $L^2(\mathbb{R}_{(r)}, dr)$ telle que $UW(p)U^{-1} = \tau_p$ et, pour tout borélien B de $L^2(\mathbb{R}_{(r)}, dr)$, $U\sigma(B)U^{-1}$ est tout simplement le projecteur de $L^2(\mathbb{R}_{(r)}, dr)$ associant à f sa restriction $f \upharpoonright_B$ à B .

$$\begin{array}{ccc} \mathcal{H} & \xrightarrow{U} & L^2(\mathbb{R}_{(r)}) \\ W(p) \downarrow & & \downarrow \tau_p \\ \mathcal{H} & \xrightarrow{U} & L^2(\mathbb{R}_{(r)}) \end{array}$$

D'où le théorème de Stone-von Neumann dans ce cas puisque

$$\rho_1(x, y, p) = e^{iy}V(x)W(p)$$

devient équivalente à

$$e^{i(y+xs)}u(s+p).$$

Si maintenant l'on utilise les dilatations de $\mathbb{H}_{pol} : y \mapsto \lambda y, x \mapsto \sqrt{\lambda}x, p \mapsto \sqrt{\lambda}p$ (chapitre 6, section 8), on obtient la représentation

$$e^{i(\lambda y + \sqrt{\lambda}xs)}u(s + \sqrt{\lambda}p)$$

qui, par l'automorphisme $x \mapsto \sqrt{\lambda}x, p \mapsto \frac{1}{\sqrt{\lambda}}p$, est elle-même unitairement équivalente à

$$e^{i\lambda(y+xs)}u(s+p).$$

Si A n'est pas de spectre simple, la décomposition spectrale de $V(x)$ est une intégrale

$$V(x) = \int_R e^{ixr}d\sigma_A(r)$$

sur un espace R plus compliqué. Mais un théorème, dit "théorème de multiplicité spectrale", implique que σ_A se décompose en somme directe $\sigma_A = \bigoplus_j \mu_j \sigma_{A,j}$ (μ_j entiers > 0) de $\sigma_{A,j}$ analogue au cas du spectre simple avec des $\sigma_{A,j}$ *mutuellement*

46. Répétons que nous reviendrons plus bas à la section 6.3 sur la notion d'imprimitivité.

orthogonales et uniques à équivalence près, l'équivalence $\sigma \equiv \sigma'$ signifiant que σ et σ' sont absolument continues l'une par rapport à l'autre. Comme $W(p)$ continue à agir par translation sur les boréliens B de R par

$$W(p) \sigma_A(B) W(p)^{-1} = \sigma_A(B+p) ,$$

il agit sur les $\sigma_{A,j}$ en les transformant en $\sigma'_{A,j} \equiv \sigma_{A,j}$. On utilise ce fait pour démontrer le théorème de Stone-von Neumann.

5.6.3. Généralisation aux produits semi-directs.

Georges Mackey s'est rendu compte que le théorème pouvait être généralisé à des systèmes d'imprimitivité où la PVM σ est seulement définie sur un espace M muni d'une mesure borélienne sur laquelle G opère de façon continue (ou en tout cas mesurable), du moment que la relation fondamentale

$$W(p) \sigma(B) W(p)^{-1} = \sigma(g(B))$$

est satisfaite. Si G agit sur M de façon seulement transitive (et pas "simplement" transitive) alors il n'y a qu'une G -orbite mais avec pour stabilisateur un sous-groupe non nul C de G satisfaisant $M \simeq G/C$. Par ailleurs, si G n'agit pas sur M de façon transitive il y a plusieurs G -orbites et il faut considérer l'espace des orbites G/M , mais cela peut être très compliqué car des phénomènes d'ergodicité des orbites peuvent intervenir.

Cette généralisation s'applique en particulier aux structures de produit *semi-direct*. Voyons-le pour \mathbb{H}_{pol} dont nous avons explicité la structure de produit semi-direct à la section 3.2 du chapitre 5. Le $\mathbb{R}_{(a)}^2$ des $a = (x, y, 0)$ est un sous-groupe normal abélien de \mathbb{H}_{pol} et le sous-groupe (non normal) $\mathbb{R}_{(p)}$ des $(0, 0, p)$ opère sur $\mathbb{R}_{(a)}^2$ par $p(a) = (x, y + px, 0)$ (i.e., si $a = (x, y, 0)$ est identifié à $a = (x, y)$, par $p(a) = (x, y + px)$), et le produit de \mathbb{H}_{pol} est donné par la formule de produit semi-direct

$$(a, p) \cdot (a', p') = (a + p(a'), p + p') .$$

On peut considérer

$$\begin{cases} V(a) = \rho_\lambda(a, 0) = \rho_\lambda(x, y, 0) \\ W(p) = \rho_\lambda(0, 0, p) \end{cases}$$

avec $\rho_\lambda(x, y, p) = V(a) W(p)$ et faire exactement le même développement que ci-dessus avec $V(a)$ au lieu de $V(x)$. On a bien

$$W(p) V(a) W(p)^{-1} = V(p(a)) .$$

On considère la décomposition spectrale de $V(a)$,

$$V(a) = \int_{\mathbb{R}_\omega^2} \omega(a) d\sigma(\omega)$$

où $\mathbb{R}_{(\omega)}^2 = \widehat{\mathbb{R}_{(a)}^2}$ est le dual de $\mathbb{R}_{(a)}^2$ (et donc $\omega(a)$ une forme linéaire sur $\mathbb{R}_{(a)}^2$) et où $d\sigma(\omega)$ est une mesure sur $\mathbb{R}_{(\omega)}^2$ à valeurs dans les projecteurs de \mathcal{H} . Comme précédemment, on a

$$W(p)\sigma(B)W(p)^{-1} = \sigma(p(B))$$

où $p(B)$ est l'action de p sur $\mathbb{R}_{(\omega)}^2$.

Si $\omega = (\xi^*, \eta^*)$ sont les éléments $\mathbb{R}_{(\omega)}^2$, p agit sur eux par

$$p^*(\xi^*, \eta^*) = (\xi^*, \eta^* - p\xi^*) .$$

Cette action n'est pas transitive car, pour (ξ_0^*, η_0^*) fixé, $\xi_0^* \neq 0$, l'orbite $p^*(\xi_0^*, \eta_0^*)$ de (ξ_0^*, η_0^*) est la droite (ξ_0^*, \mathbb{R}) puisque $\eta_0^* - p\xi_0^* = \eta^*$ est toujours résoluble par $p = \frac{\eta^* - \eta_0^*}{\xi_0^*}$.⁴⁷ Le stabilisateur d'une telle orbite est nul car $(\xi^*, \eta^* - p\xi^*) = (\xi^*, \eta^*)$ implique $p\xi^* = 0$ et donc $p = 0$ si $\xi^* \neq 0$.

Si l'on veut construire par cette méthode une représentation unitaire *irréductible* de \mathbb{H}_{pol} , il faut que la mesure $d\sigma(\omega)$ ait son support *dans une orbite*. Les unirreps seront donc paramétrées par le η_0^* qui paramétrise les orbites. Ce η_0^* correspond à la variable duale de y , c'est-à-dire au λ des ρ_λ ci-dessus. donnant le facteur de phase $e^{i\lambda y}$.

5.7. Lien avec la transformée de Fourier

Il existe un lien fondamental entre la représentation ρ_λ et la transformée de Fourier que nous avons définie dans le Vol I à la section 3.4.1. du chapitre 3 et reconsidérée plus haut à la section 2.1.1. Considérons en effet la transformée de Fourier $\mathcal{F}(\rho_\lambda(x, y, p)u(s))(\kappa)$. Au facteur $\frac{1}{\sqrt{2\pi}}$ près et en notant pour simplifier $\widehat{w}(\kappa) = \mathcal{F}(w(s))$ les transformées de Fourier, c'est l'intégrale⁴⁸

$$\mathcal{F}(e^{i\lambda(y+xs)}u(s+p))(\kappa) = \int_{\mathbb{R}} e^{-is\kappa} e^{i\lambda(y+xs)}u(s+p)ds .$$

En faisant le changement de variable $s+p=r$, on obtient

$$\begin{aligned} \mathcal{F}(\rho_\lambda(x, y, p)u(s))(\kappa) &= \int_{\mathbb{R}} e^{-i(r-p)\kappa} e^{i\lambda(y+x(r-p))}u(r)dr \\ &= e^{i(p\kappa+\lambda(y-xp))} \int_{\mathbb{R}} e^{-ir(\kappa-x)}u(r)dr \\ &= e^{i(p\kappa+\lambda(y-xp))}\widehat{u}(\kappa-x) . \end{aligned}$$

47. Si $\xi_0^* = 0$, tous les points de la droite $(0, \mathbb{R})$ sont des points fixes car $p^*(0, \eta_0^*) = (0, \eta_0^*)$ pour tout p . Cela correspond aux représentations triviales de \mathbb{H}_{pol} .

48. Dans le Vol I nous avons utilisé la convention où les intégrales définissant $w(s)$ et $\widehat{w}(k)$ ont toutes deux un facteur de normalisation $\frac{1}{\sqrt{2\pi}}$. Ici l'intégrale donnant $w(s)$ est sans facteur et celle définissant $\widehat{w}(k)$ est normalisée par $\frac{1}{2\pi}$.

La dernière formule ressemble à l'action de la représentation ρ_λ sur les $\widehat{u}(\kappa)$ mais à condition d'appliquer d'abord l'automorphisme α de \mathbb{V}_J défini par $p \rightsquigarrow -x$, $x \rightsquigarrow p$, $y \rightsquigarrow y - xp$. Il s'agit bien d'un automorphisme car on vérifie trivialement que

$$\begin{aligned}\alpha(v.v') &= \alpha(v) . \alpha(v') = (p + p', y + y' - x(p + p') - x'p', -(x + x')) \\ \alpha(v)^{-1} &= \alpha(v^{-1}) = (-p, -y, x) \\ \alpha^2(v) &= (-x, y, -p) .\end{aligned}$$

On obtient par conséquent

$$\mathcal{F}(\rho_\lambda(v)u)(\kappa) = \rho_\lambda(\alpha(v))\mathcal{F}(u)(\kappa) .$$

Le fait que α^2 soit la réflexion $(-x, -p)$ sur le plan (x, p) et l'identité sur le centre Z correspond à la formule d'inversion de la transformée de Fourier.

5.8. Généralisations

On peut généraliser cette théorie des représentations du groupe de Heisenberg en rajoutant une échelle (groupe de Heisenberg affine) et, lorsque la dimension de l'espace des phases (x, p) est $2n$ avec $n > 1$, le groupe des rotations $SO(n)$. On trouvera dans Kalisa-Torrésani [277] des indications à ce sujet.

5.9. Lien avec les travaux de Kirillov

En résumé, pour le groupe de Heisenberg polarisé $\mathbb{V}_J = \mathbb{H}_{pol}$ (l'espace des 1-jets), la partie non triviale du dual de Pontryagin $\widehat{\mathbb{V}}_J$ est paramétré par λ et il existe une mesure sur $\widehat{\mathbb{V}}_J$, la mesure de Plancherel $dP(\lambda) = \lambda d\lambda$, permettant les intégrations.

Remarque. C'est une situation générale dans le cas des groupes unimodulaires (i.e. dont les mesures de Haar à droite et à gauche sont identiques). Le dual \widehat{G} est paramétrable par un espace Λ muni d'une mesure de Plancherel $dP(\lambda)$. \square

On notera la relation très étroite entre ces unirreps obtenues au moyen de la structure de produit semi-direct et les orbites de la *représentation co-adjointe* de $\mathbb{H}_{pol} = \mathbb{V}_J$ décrites à la section 6 du chapitre 5 : les plans $(\mathbb{R}, \eta^*, \mathbb{R})$ pour $\eta^* \neq 0$ et chaque point du plan $(\xi^*, 0, \pi^*)$ pour $\eta^* = 0$. Ces orbites contiennent aussi la variable π^* duale de π qui correspond à p dans l'algèbre de Lie de \mathbb{H}_{pol} . Pour $\eta^* = 0$, l'unirrep $\rho_{\mu, \nu}$ correspond à $\mu = \xi^*$ et $\nu = \pi^*$. Pour $\eta_0^* \neq 0$, comme le plan entier $\eta^* = \eta_0^*$ est une orbite, on peut choisir $\varpi = (0, \eta_0^*, 0)$ comme représentant de l'orbite et paramétrer celle-ci comme $\xi^* = -p\eta_0^*$ et $\pi^* = x\eta_0^*$. Le stabilisateur de ϖ est le centre Z . Pour obtenir une *unirrep* on considère le quotient $R = \{(x, y, 0)\} \simeq \mathbb{R}^2$ de \mathbb{H}_{pol} par l'axe des p (c'est là qu'intervient la structure de produit semi-direct $\mathbb{R}_{(x,y)}^2 \rtimes \mathbb{R}_{(p)}$). Alors, ϖ définit un caractère χ sur R par

$$\chi(x, y, 0) = e^{i(\eta_0^* y)}$$

qui induit l'unirrep

$$\rho_{\eta_0^*}(x, y, p) u(s) = e^{i\eta_0^*(y+xs)} u(s+p),$$

c'est-à-dire ρ_λ pour $\lambda = \eta_0^*$.⁴⁹ Ici, la variable s appartient au même espace $\mathbb{R}_{(p)}$ que p et $s+p$ a donc un sens. Si l'on prend une fonction $\varphi_0(s)$ admissible⁵⁰ particulière pour ondelette mère, on obtient un état cohérent

$$\varphi_{(x,y,p)}(s) = \rho_{\eta_0^*}(x, y, p) \varphi_0(s)$$

et une transformée envoyant les fonctions $u(s)$ du Hilbert $\mathcal{H} = L^2(\mathbb{R}_{(s)})$ sur des fonctions appartenant à $L^2(\mathbb{V}_J)$.

En fait, le lien entre les unirreps de \mathbb{V}_J et les orbites de la représentation coadjointe est une *nécessité* impliquée par la profonde théorie de Kirillov sur les groupes nilpotents et, plus généralement, sur les groupes résolubles, utilisant la structure symplectique canonique de ces orbites que nous avons introduite à la section 6.3 du chapitre 5 et déjà mobilisée plusieurs fois. Nous l'expliciterons plus bas à la section 6.4.

5.10. Représentations de Segal-Bargmann

5.10.1. La transformée de Bargmann.

Il existe d'autres représentations du groupe de Heisenberg \mathbb{H} qui se sont révélées très utiles en physique et qu'il est intéressant d'introduire dans les études sur la vision. L'une des plus connues (qui, d'après le théorème de Stone-von Neumann, est équivalente par entrelacement avec celles de Schrödinger) est celle dite "de Segal-Bargmann" (ou "de Fock" en mécanique quantique) dans un espace de fonctions complexes *entières*, c'est-à-dire holomorphes sur \mathbb{C} tout entier (cf. Bargmann [29], [30], et aussi, par exemple Folland [187], Hall [241]).⁵¹ Elle s'obtient par prolongement analytique et introduit un état cohérent possédant des propriétés d'*analyticité*. En posant $z = p + ix$, on associe à toute fonction $f(t) \in \mathcal{H} = L^2(\mathbb{R}_{(t)})$ une fonction $(Bf)(z) \in \mathcal{B} \subset L^2(\mathbb{C})$ définie par

$$(Bf)(z) = 2^{\frac{1}{4}} \int_{\mathbb{R}} f(t) e^{2\pi tz - \pi t^2 - \frac{\pi}{2} z^2} dt = \int_{\mathbb{R}} B(z, t) f(t) dt = \langle f(t), B_{\bar{z}}(t) \rangle_{\mathcal{H}}$$

49. C'est un caractère car $\chi((x, y, 0) \cdot (x', y', 0)) = \chi(x + x', y + y', 0) = e^{i(\eta_0^*(y+y'))} = e^{i(\eta_0^*y)} e^{i(\eta_0^*y')}$.

50. C'est-à-dire telle que $0 < \int_{\mathbb{V}_J} |\langle \varphi_0, \rho_{\eta_0^*}(v) \varphi_0 \rangle|^2 dv < \infty$.

51. Irving Segal (1918-1998) et Valentine Bargmann (1908-1989).

avec $B(z, t) = 2^{\frac{1}{4}} e^{2\pi tz - \pi t^2 - \frac{\pi}{2} z^2}$.⁵² $B(z, t)$ s'appelle le "noyau de Bargmann" et Bf la "transformée de Bargmann" de f . La fonction Bf est une fonction *entière* et B est une isométrie de $\mathcal{H} = L^2(\mathbb{R}_{(t)})$ sur un sous-espace *fermé* \mathcal{B} – dit espace de Segal-Bargmann ou espace de Fock – de $L^2(\mathbb{C}, d\mu(z))$ muni de la mesure gaussienne

$$d\mu(z) = e^{-\pi|z|^2} dz = e^{-\pi z \bar{z}} dz .$$

On passe ainsi de \mathbb{R} à son complexifié \mathbb{C} .

Un théorème fondamental de Segal-Bargmann dit alors que \mathcal{B} est exactement l'espace $\mathfrak{H}(\mathbb{C}) \cap L^2(\mathbb{C}, d\mu(z))$ des fonctions *holomorphes* sur \mathbb{C} (espace $\mathfrak{H}(\mathbb{C})$) de *carré intégrable* pour $d\mu$ (espace $L^2(\mathbb{C}, d\mu(z))$). Les $B_z(t)$ (ou les $B_{\bar{z}}(t)$) de \mathcal{H} paramétrés par z constituent l'*état cohérent* associé au noyau B . Ils forment un système générateur redondant de $\mathcal{H} = L^2(\mathbb{R}_{(t)})$. Le fait qu'ils engendrent \mathcal{H} explique que B soit injective. Le fait qu'ils soient redondants explique que les transformées Bf aient des propriétés particulières, en l'occurrence soient holomorphes-entières.

Remarque. La transformée doit être $\int_{\mathbb{R}} B(z, t) f(t) dt$ afin qu'elle dépende directement des valeurs de f et soit holomorphe en z . Quand on veut écrire cette intégrale comme un produit scalaire $\langle f(t), g(t) \rangle_{\mathcal{H}} = \int_{\mathbb{R}} f(t) \bar{g}(t) dt$, il faut prendre $g = B_{\bar{z}}$. —□

On montre que l'espace \mathcal{B} des fonctions holomorphes de carré intégrable pour la mesure $d\mu(z)$ est bien un espace de Hilbert car il est *complet*. En effet on montre qu'une suite F_n de Cauchy dans \mathcal{B} est une suite de Cauchy dans l'espace des fonctions holomorphes pour la topologie de la convergence uniforme sur les compacts, qu'elle converge donc vers une fonction holomorphe F et que celle-ci appartient à \mathcal{B} .

Le noyau de Bargmann permet d'appliquer la théorie générale des états cohérents à la transformation de fonctions $f(t) \in L^2(\mathbb{R}_{(t)})$ très irrégulières en fonctions holomorphes sur \mathbb{C} .

5.10.2. Opérateurs de création et d'annihilation.

Le noyau de Bargmann semble compliqué mais il est lié de façon très naturelle aux débuts de la mécanique quantique et, plus précisément, à la théorie quantique de l'*oscillateur harmonique* qui repose sur les opérateurs dits de "création" a et d'"annihilation" a^\dagger avec la relation de commutation $[a, a^\dagger] = I$.

Nous avons présenté la quantification de l'oscillateur harmonique dans la section 8.5 du chapitre 16. L'hamiltonien classique est $H = \frac{1}{2} \left(m\omega^2 q^2 + \frac{p^2}{m} \right)$ où m est la masse et ω la fréquence propre. On peut le réduire à $H = \frac{\omega}{2} (q^2 + p^2)$ en remplaçant x par $\sqrt{m\omega} x$ et p par $\frac{p}{\sqrt{m\omega}}$. Quand on passe au formalisme quantique, on obtient⁵³ les opérateurs $\hat{q} = qI, \hat{p} = -i \frac{\partial}{\partial q}$ avec la relation d'Heisenberg $[\hat{q}, \hat{p}] = iI$. Nous en avons

52. Dans la section 8 du chapitre 16 consacrée à la quantification géométrique nous avons rencontré la variante $(Bf)(z) = \int_{\mathbb{R}} f(q) e^{\sqrt{2}qz - \frac{1}{2}q^2 - \frac{1}{2}z^2} dq$ qui redonne cette formule par le changement de variables $z \rightsquigarrow \sqrt{\pi}z$ et $q \rightsquigarrow \sqrt{2\pi}t$.

53. Avec $\hbar = 1$ pour simplifier et I l'opérateur identité.

souvent parlé, depuis la section 5.4 du chapitre 5 et dans les précédentes sections (en particulier 5.1) de ce chapitre.

Les opérateurs a, a^\dagger sont définis à partir de $\hat{q} = qI, \hat{p} = -i\frac{\partial}{\partial q}$ par les formules simples

$$\begin{aligned} a &= \frac{1}{\sqrt{2}} (\hat{q} + i\hat{p}) = \frac{1}{\sqrt{2}} \left(qI + \frac{\partial}{\partial q} \right), \\ a^\dagger &= \frac{1}{\sqrt{2}} (\hat{q} - i\hat{p}) = \frac{1}{\sqrt{2}} \left(qI - \frac{\partial}{\partial q} \right), \\ \hat{q} &= \frac{1}{\sqrt{2}} (a + a^\dagger), \\ \hat{p} &= \frac{i}{\sqrt{2}} (a - a^\dagger). \end{aligned}$$

correspondant au passage traditionnel des coordonnées réelles (q, p) d'un plan complexe aux coordonnées complexes conjuguées

$$\left(\frac{1}{\sqrt{2}} (q + ip), \frac{1}{\sqrt{2}} (q - ip) = \frac{1}{\sqrt{2}} (\overline{q + ip}) \right).$$

On vérifie immédiatement que

$$\begin{aligned} [a, a^\dagger] &= \left[\frac{1}{\sqrt{2}} (\hat{q} + i\hat{p}), \frac{1}{\sqrt{2}} (\hat{q} - i\hat{p}) \right] = \frac{1}{2} ([\hat{q}, -i\hat{p}] + [i\hat{p}, \hat{q}]) \\ &= \frac{1}{2} (-2i [\hat{q}, \hat{p}]) = \frac{1}{2} (-2iiI) = I. \end{aligned}$$

Par ailleurs, comme \hat{q} et \hat{p} sont auto-adjoints, il est évident que a et a^\dagger sont adjoints :

$$\begin{aligned} \langle a\varphi | \psi \rangle &= \left\langle \frac{1}{\sqrt{2}} (\hat{q} + i\hat{p}) \varphi \middle| \psi \right\rangle \\ &= \left\langle \frac{1}{\sqrt{2}} \hat{q} \varphi \middle| \psi \right\rangle + \left\langle \frac{1}{\sqrt{2}} i\hat{p} \varphi \middle| \psi \right\rangle \\ &= \left\langle \varphi \middle| \frac{1}{\sqrt{2}} \hat{q} \psi \right\rangle + \left\langle \varphi \middle| -\frac{1}{\sqrt{2}} i\hat{p} \psi \right\rangle \\ &= \langle \varphi | a^\dagger \psi \rangle. \end{aligned}$$

L'appellation de a et a^\dagger vient des propriétés suivantes, qui sont immédiatement déductibles de leur relation de commutation. L'hamiltonien quantique

$$\hat{H} = \frac{\hbar\omega}{2} (\hat{q}^2 + \hat{p}^2)$$

(qui quantifie en quelque sorte la norme du vecteur (q, p)) peut s'écrire sous la forme

$$\begin{aligned}\widehat{H} &= \hbar\omega \left(a^\dagger a + \frac{1}{2}I \right) = \hbar\omega \left(\left(\frac{1}{\sqrt{2}} (\widehat{q} - i\widehat{p}) \right) \left(\frac{1}{\sqrt{2}} (\widehat{q} + i\widehat{p}) \right) + \frac{1}{2}I \right) \\ &= \frac{\hbar\omega}{2} (\widehat{q}^2 + \widehat{p}^2 + i[\widehat{q}, \widehat{p}] + I) = \frac{\hbar\omega}{2} (\widehat{q}^2 + \widehat{p}^2)\end{aligned}$$

puisque $i[\widehat{q}, \widehat{p}] + I = 0$. Pour des coordonnées commutatives (q, p) , $q^2 + p^2 = (q + ip)(q - ip)$, mais pour les opérateurs a et a^\dagger la non-commutativité de \widehat{q} et \widehat{p} introduit le terme supplémentaire $\frac{1}{2}I$, ce qui signifie que l'énergie de l'état fondamental *n'est pas nulle*. Cela est une conséquence directe des relations d'incertitude.

On note traditionnellement $a^\dagger a$ par \widehat{N} et on l'interprète comme un "nombre de quanta" car ses valeurs propres sont des nombres entiers > 0 . Des calculs élémentaires montrent alors des propriétés fort intéressantes. Soit ψ_ν un état propre de \widehat{N} pour la valeur propre ν , i.e. $\widehat{N}(\psi_\nu) = \nu\psi_\nu$ (où ψ_ν est un vecteur de l'espace de Hilbert utilisé pour représenter les opérateurs). Évidemment, ψ_ν est aussi un état propre de l'hamiltonien pour la valeur propre $\hbar\omega(\nu + \frac{1}{2})$. Nécessairement $\nu \geq 0$ car

$$\|a(\psi_\nu)\|^2 = \langle a(\psi_\nu), a(\psi_\nu) \rangle = \langle \psi_\nu, a^\dagger a(\psi_\nu) \rangle = \nu \|\psi_\nu\|^2 .$$

Calculons alors l'action des opérateurs a et a^\dagger sur ψ_ν . La remarque de base est que $a^\dagger(\psi_\nu)$ est un état propre de \widehat{N} pour la valeur propre $\nu + 1$. En effet,

$$\widehat{N}(a^\dagger(\psi_\nu)) = (a^\dagger a) a^\dagger(\psi_\nu) = a^\dagger(aa^\dagger)(\psi_\nu) .$$

Mais comme $[a, a^\dagger] = I$, $aa^\dagger = I + a^\dagger a$ et donc

$$\widehat{N}(a^\dagger(\psi_\nu)) = a^\dagger(I + a^\dagger a)(\psi_\nu) = (\nu + 1)a^\dagger(\psi_\nu) .$$

En ce sens, a^\dagger "crée" un quantum supplémentaire pour un état propre à ν quanta et l'on doit donc avoir (si les valeurs propres ν sont simples) $a^\dagger(\psi_\nu) = \lambda_\nu^+ \psi_{\nu+1}$. De même $a(\psi_\nu)$ est un état propre de \widehat{N} pour la valeur propre $\nu - 1$ et en ce sens, a "annihile" un quantum pour un état propre à ν quanta avec $a(\psi_\nu) = \lambda_\nu^- \psi_{\nu-1}$. Pour ψ_0 (i.e. $\nu = 0$) on a $a(\psi_0) = 0$. Comme les nombres complexes λ_ν^+ et λ_ν^- sont de modules respectivement $\sqrt{\nu + 1}$ et $\sqrt{\nu}$ et définis à une phase près, on peut les prendre égaux à ces valeurs. D'autre part, l'application itérée de a ne devant pas permettre d'arriver à des états propres de \widehat{N} de valeur propre négative, il est nécessaire que les ν soient des *entiers* ≥ 0 car alors on arrive à ψ_0 ($\nu = 0$) avec $a(\psi_0) = 0$. L'énergie de l'état fondamental ψ_0 est $\frac{\hbar\omega}{2}$ et les états propres de \widehat{N} se calculent itérativement à partir ψ_0 par la formule simple $\psi_n = \frac{1}{\sqrt{n!}} (a^\dagger)^n(\psi_0)$.

En mécanique quantique, les relations de commutation sont interprétées comme des opérateurs sur les fonctions d'onde, i.e. au moyen de représentations dans des espaces fonctionnels appropriés. Or, très tôt, dans les années 1920, Vladimir Fock

remarque que $a = \frac{\partial}{\partial z}$ et $a^\dagger = z$ (i.e. multiplication par z) opérant sur les fonctions $F(z)$ holomorphes (l'holomorphic garantissant que $\frac{\partial F}{\partial z}$ a un sens) satisfaisaient trivialement cette relation. Cela permettait une nouvelle approche, fort originale, des fonctions d'onde. Ensuite, au début des années 1960, Valentine Bargmann se demanda pour quelle mesure sur \mathbb{C} ces opérateurs sont adjoints et trouva qu'il s'agit de la mesure gaussienne $e^{-|z|^2} dz$. C'est l'origine de la théorie (cf. par exemple [25]).

5.10.3. Dérivation de la transformée de Bargmann.

Comme souvent, il existe beaucoup de petites variantes dans les formules. Tout d'abord dans les modèles physiques (dans notre cas ceux de l'oscillateur harmonique), peuvent intervenir une masse m , une fréquence ω , la constante de Planck \hbar . Dans les modèles mathématiques, on égale souvent toutes ces constantes à 1. Ensuite il peut exister diverses constantes de normalisation que l'on peut choisir. Par exemple on peut changer z en cz , ce qui change a^\dagger en $M_z(F) = czF(z)$ et a en $D_z(F) = \frac{1}{c} \frac{\partial F}{\partial z}(z)$ sans changer le commutateur. Cherchons alors une mesure $\mu(z, \bar{z}) dz$ rendant M_z et D_z adjoints pour le produit scalaire

$$\langle F, G \rangle = \int_{\mathbb{C}} F(z) \overline{G(z)} \mu(z, \bar{z}) dz .$$

Cela signifie que $\langle D_z(F), G \rangle = \langle F, M_z(G) \rangle$, autrement dit que

$$\int_{\mathbb{C}} \frac{1}{c} \frac{\partial}{\partial z} (F(z)) \overline{G(z)} \mu(z, \bar{z}) dz = \int_{\mathbb{C}} F(z) c \bar{z} \overline{G(z)} \mu(z, \bar{z}) dz .$$

On intègre alors par parties en utilisant que

$$\frac{1}{c} \frac{\partial}{\partial z} (F \overline{G} \mu) = \frac{1}{c} F \frac{\partial}{\partial z} (\overline{G}) \mu + \frac{1}{c} \frac{\partial}{\partial z} (F) \overline{G} \mu + \frac{1}{c} F \overline{G} \frac{\partial}{\partial z} (\mu) .$$

Comme G est holomorphe, \overline{G} ne dépend pas de z et $\frac{\partial}{\partial z} (\overline{G}) = 0$. Par ailleurs, on peut supposer que $F \overline{G} \mu \xrightarrow{|z| \rightarrow \infty} 0$ assez vite, en vérifiant ensuite la consistance de l'hypothèse avec ses conséquences. Le terme de gauche s'annule alors. On en tire $\frac{1}{c} \frac{\partial}{\partial z} (F) \overline{G} \mu = -\frac{1}{c} F \overline{G} \frac{\partial}{\partial z} (\mu)$ et la relation devient

$$-\int_{\mathbb{C}} \frac{1}{c} F(z) \overline{G(z)} \frac{\partial}{\partial z} (\mu(z, \bar{z})) dz = \int_{\mathbb{C}} F(z) c \bar{z} \overline{G(z)} \mu(z, \bar{z}) dz .$$

Comme elle doit être valable pour toutes les F et G , on en tire une équation différentielle pour μ : $c \bar{z} \mu(z, \bar{z}) = -\frac{1}{c} \frac{\partial}{\partial z} (\mu(z, \bar{z}))$, soit

$$\frac{\partial}{\partial z} (\mu(z, \bar{z})) = -c^2 \bar{z} \mu(z, \bar{z}) ;$$

dont la solution est (à un facteur multiplicatif près)

$$e^{-c^2 z \bar{z}} = e^{-c^2 |z|^2} .$$

La convention que nous avons utilisée correspond à $c = \sqrt{\pi}$.

Il faut alors trouver une bonne fonction $B(z, t) = B_z(t)$ holomorphe en z permettant d'appliquer la théorie générale des états cohérents. Là encore, la forme de B est dictée par la physique. En effet les opérateurs a^\dagger (représentés par $M_z(F) = czF(z)$) et a (représentés par $D_z(F) = \frac{1}{c} \frac{\partial F}{\partial z}(z)$) sont également représentés par $a = \frac{1}{\sqrt{2}} \left(qI + \frac{\partial}{\partial q} \right)$ et $a^\dagger = \left(qI - \frac{\partial}{\partial q} \right)$ dans la représentation de Schrödinger. D'après le théorème de Stone-von Neumann, B doit donc satisfaire, suivant l'interprétation de la variable t en rapport avec z (nous avons utilisé $z = p + ix$), des équations exprimant l'équivalence des deux représentations par un entrelacement. Dans la convention que nous avons utilisée, t est scalée par $\sqrt{2\pi}$ et, étant donné que

$$\begin{cases} \frac{\partial B}{\partial t} = (2\pi z - 2\pi t) B \\ \frac{\partial B}{\partial z} = (2\pi t - \pi z) B, \end{cases}$$

B satisfait trivialement les équations

$$\begin{aligned} \sqrt{\pi} z B(z, t) = M_z(B) &= \frac{1}{\sqrt{2}} \left(\sqrt{2\pi} t B + \frac{1}{\sqrt{2\pi}} \frac{\partial B}{\partial t} \right) = A_t(B) & (M/D) \\ \frac{1}{\sqrt{\pi}} \frac{\partial B}{\partial z} = D_z(B) &= \frac{1}{\sqrt{2}} \left(\sqrt{2\pi} t B - \frac{1}{\sqrt{2\pi}} \frac{\partial B}{\partial t} \right) = A'_t(B) \end{aligned}$$

où A_t et A'_t sont des opérateurs. La première équation (M) exprime que

$$B(A_t(f(t))) = D_z(Bf(t)) .$$

En effet, cette dernière relation signifie, puisque

$$D_z(B(z, t) f(t)) = D_z(B(z, t)) f(t)$$

car $f(t)$ ne dépend pas de z , que

$$\int_{\mathbb{R}} B(z, t) A_t(f(t)) dt = \int_{\mathbb{R}} D_z(B(z, t)) f(t) dt = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \frac{\partial B}{\partial z}(z, t) f(t) dt .$$

Or, si $D_z(B) = A'_t(B)$, cette formule s'écrit

$$\int_{\mathbb{R}} B(z, t) A_t(f(t)) dt = \int_{\mathbb{R}} A'_t(B(z, t)) f(t) dt .$$

En tenant compte de la forme de A_t et A'_t , on voit que des deux côtés de la formule les intégrandes $B \frac{1}{\sqrt{2}} (\sqrt{2\pi} t f) = \frac{1}{\sqrt{2}} (\sqrt{2\pi} t B) f$ s'éliminent. Reste à vérifier que

$$\int_{\mathbb{R}} B(z, t) \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2\pi}} \frac{\partial f}{\partial t} \right) dt = - \int_{\mathbb{R}} \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2\pi}} \frac{\partial B}{\partial t} \right) f(t) dt ,$$

i.e. $\int_{\mathbb{R}} B \frac{\partial f}{\partial t} dt = - \int_{\mathbb{R}} \frac{\partial B}{\partial t} f dt$. Une intégration par parties donne

$$[Bf]_{-\infty}^{+\infty} = \left(\int_{\mathbb{R}} B \frac{\partial f}{\partial t} + \frac{\partial B}{\partial t} f \right) dt$$

et donc, comme $[Bf]_{-\infty}^{+\infty} = 0$ à cause des hypothèse d'intégrabilité sur B et les f ,

$$\int_{\mathbb{R}} B \frac{\partial f}{\partial t} dt = - \int_{\mathbb{R}} \frac{\partial B}{\partial t} f dt.$$

Dualement, la seconde équation (D) exprime que $B (A'_t (f (t))) = M_z (B f (t))$. En effet, cette dernière relation signifie que

$$\int_{\mathbb{R}} B(z, t) A'_t (f (t)) dt = \sqrt{\pi} z \int_{\mathbb{R}} B(z, t) f (t) dt = \int_{\mathbb{R}} M_z (B(z, t)) f (t) dt .$$

Si $M_z (B) = A_t (B)$, cette formule s'écrit

$$\int_{\mathbb{R}} B(z, t) A'_t (f (t)) dt = \int_{\mathbb{R}} A_t (B(z, t)) f (t) dt$$

et elle est vérifiée pour la même raison que la précédente. On obtient ainsi l'équivalence des représentations.

5.10.4. *Le noyau reproduisant et l'état cohérent.*

Calculons le produit scalaire $\langle B_z(t), B_w(t) \rangle_{\mathcal{H}}$ avec $z = p + ix$ et $w = u + iv$. On obtient :

$$\sqrt{2} \int_{\mathbb{R}} e^{2\pi tz - \pi t^2 - \frac{\pi}{2} z^2} e^{2\pi t\bar{w} - \pi t^2 - \frac{\pi}{2} \bar{w}^2} dt = \sqrt{2} e^{-\frac{\pi}{2}(z^2 + \bar{w}^2)} \int_{\mathbb{R}} e^{2\pi t(p+u+i(x-v)) - 2\pi t^2} dt .$$

Mais

$$\int_{\mathbb{R}} e^{2\pi t(p+u+i(x-v)) - 2\pi t^2} dt = \frac{1}{\sqrt{2}} e^{\frac{\pi}{2}(p+u+i(x-v))^2}$$

et un calcul immédiat montre que

$$\langle B_z(t), B_w(t) \rangle_{\mathcal{H}} = e^{\pi z \bar{w}}$$

en particulier $\|B_z(t)\|_{\mathcal{H}}^2 = e^{\pi|z|^2}$. C'est donc $K(z, w) = e^{\pi z \bar{w}}$ le noyau reproduisant de la situation. On vérifie que l'on a bien $K(z, z) = \|B_z(t)\|_{\mathcal{H}}^2$, $K(w, z) = e^{\pi w \bar{z}} = \overline{K(z, w)}$ (symétrie) et surtout

$$\int_{\mathbb{C}} K(z, w) K(w, z') d\mu(w) = \int_{\mathbb{C}} e^{\pi z \bar{w}} e^{\pi w \bar{z}'} e^{-\pi w \bar{w}} dw = K(z, z') .$$

Cette dernière identité découle du fait que $\int_{\mathbb{R}} e^{\pi(rX - r^2)} dr = e^{\frac{\pi}{4} X^2}$, X étant un nombre complexe quelconque. La vérification est alors immédiate. En séparant les

variables on obtient en effet $e^{\frac{\pi}{4}X^2}e^{\frac{\pi}{4}Y^2}$ avec

$$\begin{cases} X = p + p' + i(x - x') \\ Y = x + x' + i(p' - p) \\ X^2 + Y^2 = 4(pp' + xx' + i(xp' - x'p)) = 4z\bar{z}' . \end{cases}$$

Avec ce noyau reproduisant on obtient par conséquent

$$F(z) = \int_{\mathbb{C}} K(z, w) F(w) d\mu(w) = \int_{\mathbb{C}} e^{\pi z\bar{w}} F(w) d\mu(w)$$

ce qui impose les contraintes pour les $F(z) \in \mathcal{B}$. En particulier, d'après l'inégalité de Cauchy-Schwarz, on a $|F(z)| \leq e^{\frac{\pi|z|^2}{2}} \|F\|$.

On montre que $B : L^2(\mathbb{R}) \rightarrow \mathcal{B}$ est une isométrie d'image \mathcal{B} . On définit son inverse B^{-1} par la formule :

$$(B^{-1}F)(t) = \int_{\mathbb{C}} F(z) B(z, t) e^{-\pi|z|^2} dz = \int_{\mathbb{C}} \langle f(t), B_{\bar{z}}(t) \rangle_{\mathcal{H}} B_z(t) e^{-\pi|z|^2} dz .$$

5.10.5. Structure de l'isométrie.

Une façon commode de comprendre l'isométrie est de chercher les $h(t)$ qui sont les images inverses des z^k engendrant les séries entières. On montre qu'il s'agit d'une variante des polynômes d'Hermite (définis par les dérivées de gaussiennes, cf. Vol I, chapitre 3, section 3.3.1) calculée au moyen de la formule génératrice

$$H_k(t) = (-1)^k e^{\frac{t^2}{2}} \frac{d^k}{dt^k} \left(e^{-t^2} \right) .$$

Notre variable est ici $\sqrt{2\pi}t$. Considérons alors la transformée de Bargmann de $\lambda_k H_k(\sqrt{2\pi}t)$, λ_k étant un facteur à déterminer, c'est-à-dire l'intégrale

$$2^{\frac{1}{4}} \lambda_k (-1)^k \int_{\mathbb{R}} e^{2\pi tz - \pi t^2 - \frac{\pi}{2} z^2} e^{\pi t^2} \frac{d^k}{dt^k} \left(e^{-2\pi t^2} \right) dt .$$

Le terme en t^2 de l'exponentielle s'élimine et l'on obtient

$$2^{\frac{1}{4}} \lambda_k (-1)^k \int_{\mathbb{R}} e^{2\pi tz - \frac{\pi}{2} z^2} \frac{d^k}{dt^k} \left(e^{-2\pi t^2} \right) dt .$$

On montre alors l'identité remarquable

$$\int_{\mathbb{R}} e^{2\pi tz - \frac{\pi}{2} z^2} \frac{d^k}{dt^k} \left(e^{-2\pi t^2} \right) dt = (-1)^k \int_{\mathbb{R}} \frac{d^k}{dt^k} \left(e^{2\pi tz - \frac{\pi}{2} z^2} \right) e^{-2\pi t^2} dt . \quad (\text{IG})$$

Pour le membre de droite on obtient donc, puisque

$$\frac{d^k}{dt^k} \left(e^{2\pi tz - \frac{\pi}{2} z^2} \right) = (2\pi z)^k e^{2\pi tz - \frac{\pi}{2} z^2} ,$$

$$(-1)^k (2\pi z)^k \int_{\mathbb{R}} e^{2\pi tz - \frac{\pi}{2} z^2 - 2\pi t^2} dt.$$

Pour le membre de gauche on obtient

$$\int_{\mathbb{R}} e^{2\pi tz - \frac{\pi}{2} z^2 - 2\pi t^2} P_k(t) dt$$

où $P_k(t)$ est un polynôme de degré alternativement impair et pair qui est engendré comme les polynômes d'Hermite.⁵⁴

À partir de là, en utilisant les propriétés de ce genre d'intégrales, dites "intégrales gaussiennes", on peut démontrer l'identité remarquable (IG).⁵⁵ On peut s'amuser à faire quelques calculs à la main. Les intégrales sont des intégrales d'exponentielles du type $G_A = e^{A(t)}$ avec $A(t) = 2\pi tz - \frac{\pi}{2} z^2$ et $G_B = e^{B(t)}$ avec $B(t) = -2\pi t^2$.

- (IG) pour $k = 1$ découle d'une intégration par parties évidente.
- Pour $k = 2$, en notant $C(t) = A(t) + B(t)$, on a à droite

$$(2\pi z)^2 \int_{\mathbb{R}} e^{C(t)} dt = \frac{(2\pi z)^2}{\sqrt{2}}$$

et à gauche

$$\int_{\mathbb{R}} P_2(t) e^{C(t)} dt = \int_{\mathbb{R}} (-4\pi + (4\pi t)^2) e^{C(t)} dt.$$

Mais $\int_{\mathbb{R}} e^{C(t)} dt = \frac{1}{\sqrt{2}}$ et $\int_{\mathbb{R}} t^2 e^{C(t)} dt = \frac{1+\pi z^2}{4\sqrt{2}\pi}$, d'où le bilan

$$\frac{-4\pi}{\sqrt{2}} + (4\pi)^2 \frac{1+\pi z^2}{4\sqrt{2}\pi} = \frac{(2\pi z)^2}{\sqrt{2}}.$$

- Pour $k = 3$, on a à droite

$$-(2\pi z)^3 \int_{\mathbb{R}} e^{C(t)} dt = -\frac{(2\pi z)^3}{\sqrt{2}}$$

et à gauche

$$\int_{\mathbb{R}} e^{C(t)} P_3(t) dt = \int_{\mathbb{R}} (3 \times 16\pi^2 t - (4\pi t)^3) e^{C(t)} dt.$$

54. $P_1(t) = -4\pi t$, $P_2(t) = -4\pi + (4\pi t)^2$, $P_3(t) = 3 \times 16\pi^2 t - (4\pi t)^3$.

55. Les intégrales gaussiennes sont connues depuis longtemps. Elles sont traditionnellement définies par $I_k = \int_{\mathbb{R}} x^k e^{-x^2} dx$ et sont nulles pour k impair. Une simple intégration par parties montre qu'elles satisfont la relation de récurrence $I_{k+2} = \frac{n+1}{2} I_k$. En effet $\frac{d}{dx} (x^{k+1} e^{-x^2}) = (k+1) x^k e^{-x^2} - 2x^{k+2} e^{-x^2}$ et comme $\left[x^{k+1} e^{-x^2} \right]_{-\infty}^{+\infty} = 0$, $(k+1) I_k = 2I_{k+2}$. Comme $I_0 = \sqrt{\pi}$, on obtient la formule classique $I_{2n} = \sqrt{\pi} \frac{(2n-1)!!}{2^n}$ où $(2n-1)!! = (2n-1) \cdot (2n-3) \dots 5 \cdot 3 \cdot 1 = \frac{(2n)!}{2^n n!}$ est la double factorielle.

Mais $\int_{\mathbb{R}} te^{C(t)} dt = \frac{z}{2\sqrt{2}}$ et $\int_{\mathbb{R}} t^3 e^{C(t)} dt = \frac{z(3+\pi z^2)}{8\sqrt{2}\pi}$, d'où le bilan

$$3 \times 16\pi^2 \frac{z}{2\sqrt{2}} - (4\pi)^3 \frac{z(3+\pi z^2)}{8\sqrt{2}\pi} = -\frac{(2\pi z)^3}{\sqrt{2}}.$$

Etc.

On obtient par conséquent

$$2^{\frac{1}{4}} \lambda_k (-1)^k (-1)^k (2\pi z)^k \int_{\mathbb{R}} e^{2\pi tz - \frac{\pi}{2} z^2 - 2\pi t^2} dt.$$

Mais $2\pi tz - \frac{\pi}{2} z^2 - 2\pi t^2$ est le carré $-\left(\sqrt{\frac{\pi}{2}}z - \sqrt{2\pi}t\right)^2$ et cela implique (nous avons déjà utilisé cette propriété) que l'intégrale

$$\int_{\mathbb{R}} e^{-(\sqrt{\frac{\pi}{2}}z - \sqrt{2\pi}t)^2} dt = \frac{1}{2}$$

est une *constante* indépendante de z . On obtient donc en définitive

$$\frac{1}{2} 2^{\frac{1}{4}} \lambda_k (-1)^k (-1)^k (2\pi z)^k.$$

Si l'on veut obtenir z^k , ou mieux $(\sqrt{\pi}z)^k$, il faut prendre $\lambda_k \pi^{\frac{k}{2}} 2^{k-\frac{3}{4}} = 1$. Les z^k et les H_k constituant un système orthogonal complet respectivement pour \mathcal{B} et $L^2(\mathbb{R})$, on voit comment fonctionne l'isométrie.

5.10.6. La représentation de \mathbb{H} dans \mathcal{B} .

La transformée de Bargmann permet de transformer les $f(t) \in \mathcal{H} = L^2(\mathbb{R}_t)$ en fonctions holomorphes $F(z) = (Bf)(z) \in \mathcal{B} = B(L^2(\mathbb{R}))$. À travers la transformée de Bargmann, la représentation de Schrödinger $\rho(x, y, p)$ se transforme donc par "entrelacement" en une représentation équivalente χ de \mathbb{H} dans \mathcal{B} définie par $\chi(x, y, p)B = B\rho(x, y, p)$. Si $F(z) \in \mathcal{B}$ on a, avec $w = u + iv$ une seconde variable sur \mathbb{C} ,

$$\chi(w, t)F(z) = \chi(v, t, u)F(z) = e^{-\frac{\pi}{2}|w|^2 - \pi z\bar{w} + 2i\pi t} F(z + w).$$

On remarquera que $\chi(w, t) = e^{2i\pi t} \chi(w, 0)$.

Soit $\varphi_0(t) = 2^{\frac{1}{4}} e^{-\pi t^2}$ une ondelette mère gaussienne pour $L^2(\mathbb{R})$. Par transformée de Bargmann, elle devient la fonction constante 1 :

$$\Phi_0(z) = B\varphi_0(t) = \sqrt{2} e^{-\frac{\pi}{2} z^2} \int_{\mathbb{R}} e^{2\pi tz - 2\pi t^2} dt = 1$$

car $\int_{\mathbb{R}} e^{2\pi tz - 2\pi t^2} dt = \frac{1}{\sqrt{2}} e^{\frac{\pi}{2} z^2}$. L'action de χ donne alors

$$\chi(w, t)\Phi_0(z) = \Phi_w(z) = e^{-\frac{\pi}{2}|w|^2 - \pi z\bar{w} + 2i\pi t} = e^{-\pi z\bar{w}} e^{2i\pi t} e^{-\frac{\pi}{2}|w|^2}.$$

Le facteur $e^{2i\pi t}$ vient du centre Z , le facteur $e^{-\frac{\pi}{2}|w|^2}$ est le facteur de la mesure du signal.

Avec la transformée de Bargmann, des fonctions $f(t) \in L^2(\mathbb{R}_{(t)})$ d'une variable réelle qui peuvent être considérées comme des fonctions d'onde sur un espace de configuration $\mathbb{R}_{(t)}$ deviennent des fonctions d'onde $F(z)$ de deux variables réelles définies sur un espace des phases $\mathbb{C}_{(z)}$. Or on sait qu'en mécanique quantique les relations d'incertitude de Heisenberg interdisent une localisation trop concentrée dans l'espace des phases, l'augmentation de la localisation d'une observable conduisant à la diminution de la localisation de l'observable conjuguée. C'est là que la propriété d'*holomorphic* des $F(z) \in \mathcal{B}$ devient essentielle pour l'interprétation quantique car elle *borne inférieurement* la concentration conjointe de la localisation d'observables conjuguées.

5.10.7. Lien avec le noyau de la chaleur.

Nous parlerons plus bas à la section 9, à propos des processus de *diffusion*, de l'équation et du noyau de la chaleur en géométrie sous-riemannienne. Mais nous pouvons d'ores et déjà noter ici le lien entre le noyau de Bargmann et le noyau de la chaleur classique sur \mathbb{R} .

L'équation de la chaleur sur \mathbb{R} (déjà évoquée à la section 2.1.1) est l'équation parabolique $\frac{\partial f(x,s)}{\partial s} = \Delta f(x,s)$ que l'on veut résoudre avec une condition initiale $f(x,0) = u(x)$. Comme l'équation est linéaire, toute superposition de solutions est encore une solution et comme $u(x) = \int_{\mathbb{R}} \delta(x-y) u(y) dy$ (δ étant la distribution de Dirac), il suffit essentiellement de connaître la "solution élémentaire" ou "solution fondamentale" de l'équation, c'est-à-dire la solution de $\frac{\partial P(x,s)}{\partial s} = \Delta P(x,s)$ avec $P(x,s) \xrightarrow{s \rightarrow 0^+} \delta(x)$.⁵⁶ On trouve

$$P(x,s) = \frac{1}{\sqrt{4\pi s}} e^{-\frac{x^2}{4s}}$$

qui satisfait bien

$$\frac{\partial P(x,s)}{\partial s} = \Delta P(x,s) = \frac{(-2s + x^2)}{8\sqrt{\pi s^{\frac{5}{2}}}} e^{-\frac{x^2}{4s}},$$

$P(x,s) \xrightarrow{s \rightarrow 0^+} 0$ si $x \neq 0$, $P(0,s) \xrightarrow{s \rightarrow 0^+} \infty$ et $\int_{-\infty}^{+\infty} P(x,s) dx = 1$ pour tout $s > 0$. La solution générale est donc

$$f(x,s) = \frac{1}{\sqrt{4\pi s}} \int_{\mathbb{R}} e^{-\frac{(x-y)^2}{4s}} u(y) dy.$$

56. L'idée est intuitive. La distribution $\delta(x)$ représente la concentration au point 0 d'une quantité = 1 de substance diffusante. Par linéarité, il suffit donc de savoir comment $\delta(x)$ diffuse pour savoir comment n'importe quelle répartition $u(x)$ de substance diffusante diffuse.

La fonction

$$P(x, s) = P_s(x) = \frac{1}{\sqrt{4\pi s}} e^{-\frac{x^2}{4s}}$$

s'appelle le “noyau de la chaleur” et la solution générale s'obtient donc par la *convolution*

$$f(x, s) = u(x) * P_s(x) .$$

On peut alors remarquer que le noyau de Bargmann s'écrit

$$B(z, t) = 2^{\frac{1}{4}} e^{2\pi tz - \pi t^2 - \frac{\pi}{2} z^2} = (4\pi)^{\frac{1}{4}} \frac{P_{\frac{1}{2}}(\sqrt{\pi}z - 2\sqrt{\pi}t)}{\sqrt{P_{\frac{1}{2}}(2\sqrt{\pi}t)}}$$

et est donc de la forme $\frac{P_s(z-t)}{\sqrt{P_s(t)}}$ aux constantes de normalisation près : $z \rightsquigarrow \sqrt{\pi}z$ et $t \rightsquigarrow 2\sqrt{\pi}t$. Dans cette formule $P_s(z-t)$ est le prolongement analytique à \mathbb{C} de $P_s(t)$ défini sur \mathbb{R} , \mathbb{C} étant le complexifié de \mathbb{R} . Cela permet d'ailleurs d'introduire le paramètre supplémentaire s dans $B(z, t)$ en posant

$$B(z, t; s) = \left(\frac{2}{s}\right)^{\frac{1}{4}} e^{(2\pi tz - \pi t^2 - \frac{\pi}{2} z^2)/2s} .$$

Ce paramètre s représente une constante physique variable.

5.10.8. La généralisation de Brian Hall.

Dans son article [241] de 1994, Brian Hall a développé une vaste généralisation du formalisme de Bargmann aux groupes de Lie compacts connexes K (d'éléments notés κ), la compacité permettant de garantir tout un ensemble de techniques d'analyse (convergence d'intégrales, etc.) dans les démonstrations.

Soit $d\kappa$ la mesure de Haar de K invariante à gauche. On considère le complexifié $K_{\mathbb{C}}$ de K (de variable notée \varkappa) qui s'obtient en prenant le complexifié $\mathcal{K}_{\mathbb{C}} = \mathcal{K} \oplus i\mathcal{K}$ de l'algèbre de Lie \mathcal{K} de K .⁵⁷ Le groupe K s'injecte canoniquement comme un sous-groupe dans son complexifié $K_{\mathbb{C}}$ et $j : K \hookrightarrow K_{\mathbb{C}}$ est un morphisme réel universel pour les morphismes réels $\varphi : K \rightarrow G$ de K dans des groupes de Lie complexes \mathbb{C} -analytiques G . Cela signifie que pour tout φ , il existe un unique morphisme \mathbb{C} -analytique $\Phi : K_{\mathbb{C}} \rightarrow G$ tel que $\varphi = \Phi \circ j$. On montre que $K_{\mathbb{C}}$ existe et est unique, qu'il est connexe, que j est bien un morphisme injectif et qu'il fait de K un groupe compact⁵⁸ *maximal* de $K_{\mathbb{C}}$ (qui lui-même *n'est plus* compact) et enfin qu'il existe une conjugaison complexe anti-holomorphe $\varkappa \mapsto \bar{\varkappa}$.

On peut définir sur K le noyau de la chaleur comme sur \mathbb{R} en considérant une métrique bi-invariante sur K (dont l'existence est garantie par la compacité de K)

57. Ne pas confondre la notation du groupe compact K avec celle d'un noyau reproduisant, ni celle de l'algèbre de Lie \mathcal{K} avec la notation \mathcal{K} pour des distributions de contact.

58. L'image d'un compact par une application continue est compacte.

et le laplacien associé. On montre alors (théorème de Stein) que ce noyau existe et qu'il est une fonction $C^\infty > 0$ pour tout $s > 0$.

Soit donc ρ_s le noyau de la chaleur de K (nous gardons la notation de Hall). Une démonstration technique permet de montrer que l'on peut prolonger analytiquement ρ_s au complexifié $K_{\mathbb{C}}$. On définit alors un état cohérent par

$$B_s(\mathcal{X}, \kappa) = \frac{\rho_s(\kappa^{-1}\mathcal{X})}{\sqrt{\rho_s(\kappa)}}$$

qui généralise directement la formule classique où $\kappa^{-1}\mathcal{X} = z - t$. Puis on définit la transformée de $f(\kappa) \in L^2(K, d\kappa)$ par

$$(B_s f)(\mathcal{X}) = \int_K B_s(\mathcal{X}, \kappa) f(\kappa) d\kappa .$$

Soit μ_s la mesure associée au noyau de la chaleur sur $K_{\mathbb{C}}$. Hall montre que pour tout $s > 0$, $f \mapsto B_s f$ est une isométrie de $L^2(K, d\kappa)$ sur $\mathfrak{H}(K_{\mathbb{C}}) \cap L^2(K_{\mathbb{C}}, d\mu_s(\mathcal{X}))$ l'espace des fonctions holomorphes sur $K_{\mathbb{C}}$ de carré intégrable pour la mesure μ_s .

La formule d'inversion est alors donnée pour $f(\kappa) \in L^2(K, d\kappa)$ (en tenant compte des problèmes techniques posés par le fait que $K_{\mathbb{C}}$ n'est plus compact) par

$$f(\kappa) = \int_{K_{\mathbb{C}}} (B_s f)(\mathcal{X}) \frac{\overline{\rho_s(\kappa^{-1}\mathcal{X})}}{\sqrt{\rho_s(\kappa)}} d\mu_s(\mathcal{X}) .$$

Cette formule découle du fait que

$$D_s(\mathcal{X}, \mathcal{X}') = \int_K \frac{\rho_s(\kappa^{-1}\mathcal{X}) \overline{\rho_s(\kappa^{-1}\mathcal{X}')}}{\rho_s(\kappa)} d\kappa = \langle B_s(\mathcal{X}', \kappa), B_s(\mathcal{X}, \kappa) \rangle ,$$

qui est holomorphe en \mathcal{X} et antiholomorphe en \mathcal{X}' , est un noyau reproduisant satisfaisant

$$\int_{K_{\mathbb{C}}} D_s(\mathcal{X}, \mathcal{X}'') D_s(\mathcal{X}'', \mathcal{X}') d\mu_s(\mathcal{X}'') = D_s(\mathcal{X}, \mathcal{X}')$$

et que, si $F(\mathcal{X}) \in \mathfrak{H}(K_{\mathbb{C}}) \cap L^2(K_{\mathbb{C}}, d\mu_s(\mathcal{X}))$, alors

$$F(\mathcal{X}) = \int_{K_{\mathbb{C}}} D_s(\mathcal{X}, \mathcal{X}') F(\mathcal{X}') d\mu_s(\mathcal{X}') . \tag{59}$$

$D_s(\mathcal{X}, \mathcal{X}')$ est l'équivalent du $K(z, z') = e^{\pi z \bar{z}'}$ du cas classique.

59. La relation de noyau reproduisant n'est d'ailleurs qu'un cas particulier de cette dernière relation puisque, si $F(\mathcal{X}'') = D_s(\mathcal{X}'', \mathcal{X}')$, on a $F(\mathcal{X}'') \in \mathfrak{H}(K_{\mathbb{C}}) \cap L^2(K_{\mathbb{C}}, d\mu_s(\mathcal{X}))$ et donc $F(\mathcal{X}) = D_s(\mathcal{X}, \mathcal{X}') = \int_{K_{\mathbb{C}}} D_s(\mathcal{X}, \mathcal{X}'') F(\mathcal{X}'') d\mu_s(\mathcal{X}'') = \int_{K_{\mathbb{C}}} D_s(\mathcal{X}, \mathcal{X}'') D_s(\mathcal{X}'', \mathcal{X}') d\mu_s(\mathcal{X}'')$.

On peut également considérer le Hilbert $L^2(K, d\rho_s(\kappa))$ qui est isométrique à $L^2(K, d\kappa)$ par la transformation $f \mapsto \sqrt{\rho_s}f$. L'isométrie est évidente car

$$\int_K |f|^2 d\kappa = \int_K \left| \frac{f}{\sqrt{\rho_s(\kappa)}} \right|^2 \rho_s(\kappa) d\kappa .$$

Cela permet de définir un autre état cohérent $B_s^\rho(\mathcal{X}, \kappa) = \rho_s(\kappa^{-1}\mathcal{X})$ et une autre transformée $(B_s^\rho f)(\mathcal{X}) = \int_K B_s^\rho(\mathcal{X}, \kappa) f(\kappa) d\kappa$ qui, pour tout $s > 0$, est une isométrie de $L^2(K, d\rho_s(\kappa))$ sur $\mathfrak{H}(K_{\mathbb{C}}) \cap L^2(K_{\mathbb{C}}, d\mu_s(\mathcal{X}))$. La formule d'inversion est alors pour $f(\kappa) \in L^2(K, d\rho_s(\kappa))$

$$f(\kappa) = \int_{K_{\mathbb{C}}} (B_s^\rho f)(\mathcal{X}) \frac{\overline{\rho_s(\kappa^{-1}\mathcal{X})}}{\sqrt{\rho_s(\kappa)}} d\mu_s(\mathcal{X}) .$$

5.10.9. Le cas du cercle.

L'exemple le plus simple de la théorie de Hall est la transformée de Bargmann sur le cercle $\mathbb{S}^1 = \mathbb{S} = \{e^{i\theta}\}$ qui est un groupe compact abélien. On trouvera les calculs détaillés par exemple dans van Leeuwen [322].

L'algèbre de Lie \mathfrak{s} de \mathbb{S} est l'algèbre triviale $(\mathbb{R}, +)$ et donc sa complexifiée $\mathfrak{s}_{\mathbb{C}} = \mathfrak{s} + i\mathfrak{s}$ est l'algèbre triviale $(\mathbb{C}, +)$. Le complexifié $\mathbb{S}_{\mathbb{C}}$ de \mathbb{S} est $(\mathbb{C}^*, \times) \simeq \mathbb{S} \times (\mathbb{R}_{>0}, \times)$ (où $\mathbb{R}_{>0}$ est le groupe multiplicatif des réels > 0) et \mathbb{S} s'y trouve plongé comme le sous-groupe multiplicatif des nombres complexes de module 1, qui est bien un sous-groupe compact maximal. La structure riemannienne naturelle sur $\mathfrak{s}_{\mathbb{C}}$ est la métrique euclidienne $|\theta + ix| = \theta^2 + x^2$. On en fait par translation une métrique invariante sur $\mathbb{S}_{\mathbb{C}}$. Le laplacien sur \mathbb{S} est $\Delta_{\mathbb{S}} = \frac{d^2}{d\theta^2}$ et celui sur $\mathbb{S}_{\mathbb{C}}$ est $\Delta_{\mathbb{C}} = \frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial x^2}$.

Soit alors $f(\theta)$ une fonction sur \mathbb{S} . Sa série de Fourier est la série

$$f(\theta) = \sum_{k \in \mathbb{Z}} \widehat{f}(k) e^{ik\theta}$$

avec les coefficients

$$\widehat{f}(k) = \frac{1}{2\pi} \int_{\mathbb{S}} f(\theta) e^{-ik\theta} d\theta .$$

Il est alors facile de décrire la solution fondamentale $\rho_{\mathbb{S}}(s, \theta)$ de l'équation de la chaleur $\frac{\partial}{\partial\theta} = \Delta_{\mathbb{S}}$ sur \mathbb{S} . On utilise le fait que $\frac{d\widehat{f}}{d\theta}(k) = ik\widehat{f}(k)$ et donc

$$\widehat{\frac{d^2 f}{d\theta^2}}(k) = -k^2 \widehat{f}(k) .$$

Les coefficients de Fourier de $\rho_{\mathbb{S}}$ sont de la forme $c_k e^{-k^2 s}$ et comme il faut que $\rho_{\mathbb{S}}(s, \theta) \xrightarrow{s \rightarrow 0^+} \delta(\theta)$ (δ de Dirac), $c_k = \frac{1}{2\pi}$ et par conséquent

$$\rho_{\mathbb{S}}(s, \theta) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} e^{ik\theta} e^{-k^2 s} .$$

Par ailleurs, la solution fondamentale de l'équation de la chaleur sur \mathbb{R} est

$$\rho_{\mathbb{R}}(s, x) = \frac{1}{\sqrt{4\pi s}} e^{-\frac{x^2}{4s}}$$

(nous y reviendrons plus bas).

La solution fondamentale $\rho(s, \theta, x)$ de l'équation de la chaleur sur $\mathbb{S}_{\mathbb{C}}$ est le produit de ces deux solutions fondamentales $\rho_{\mathbb{S}}$ et $\rho_{\mathbb{R}}$ et l'on a ainsi

$$\rho(s, \theta, x) = \frac{1}{2\pi} \frac{1}{\sqrt{4\pi s}} e^{-\frac{x^2}{4s}} \sum_{k \in \mathbb{Z}} e^{ik\theta} e^{-k^2 s} .$$

On en déduit l'expression de la transformée de Bargmann $(B_s^\rho f)(\theta + ix)$ au sens de Hall qui s'obtient en convoluant $f(\theta)$ par le prolongement analytique de ρ_s :

$$(B_s^\rho f)(\theta + ix) = \sum_{k \in \mathbb{Z}} \widehat{f}(k) e^{ik(\theta+ix)} e^{-k^2 s} .$$

On vérifie alors facilement que, pour tout $s > 0$, B_s^ρ est une isométrie de $L^2(\mathbb{S}, d\rho_{\mathbb{S},s})$ dans $L^2(\mathbb{S}_{\mathbb{C}}, d\rho_{\frac{s}{2}})$. En effet, si $f(\theta) \in L^2(\mathbb{S}, d\rho_s)$, sa norme est

$$\begin{aligned} \|f(\theta)\|_{\rho_s}^2 &= \int_{\mathbb{S}} |f(\theta)|^2 \rho_{\mathbb{S}}(s, \theta) d\theta \\ &= \frac{1}{2\pi} \int_{\mathbb{S}} \left(\sum_{k, k' \in \mathbb{Z}} \overline{\widehat{f}(k)} \widehat{f}(k') e^{i(k'-k)\theta} e^{im\theta} e^{-m^2 s} \right) d\theta . \end{aligned}$$

Mais l'intégrale $\int_{\mathbb{S}} e^{i(k'-k)\theta} e^{im\theta} d\theta$ est nulle sauf pour $m = k - k'$ où elle est égale à 2π . Donc

$$\|f(\theta)\|_{\rho_s}^2 = \sum_{k, k' \in \mathbb{Z}} \overline{\widehat{f}(k)} \widehat{f}(k') e^{-(k-k')^2 s} .$$

Par ailleurs, la norme $(B_s^\rho f)$ dans $L^2(\mathbb{S}_{\mathbb{C}}, d\rho_{\frac{s}{2}})$ est

$$\begin{aligned}
\|B_s^\rho f\|_\rho^2 &= \int_{\mathbb{S}} \int_{\mathbb{R}} |(B_s^\rho f)(\theta + ix)|^2 \rho\left(\frac{s}{2}, \theta, x\right) d\theta dx \\
&= \frac{1}{2\pi} \frac{1}{\sqrt{4\pi\frac{s}{2}}} \int_{\mathbb{S}} \int_{\mathbb{R}} \left(\sum_{k, k' \in \mathbb{Z}} \overline{\widehat{f}(k)} \widehat{f}(k') e^{-ik(\theta-ix)} e^{-k^2 s} e^{ik'(\theta+ix)} e^{-k'^2 s} \right) \\
&\quad \times \left(e^{-\frac{x^2}{4\frac{s}{2}}} \sum_{m \in \mathbb{Z}} e^{im\theta} e^{-m^2 \frac{s}{2}} \right) d\theta dx \\
&= \frac{1}{2\pi} \frac{1}{\sqrt{2\pi s}} \int_{\mathbb{R}} \sum_{k, k', m \in \mathbb{Z}} \overline{\widehat{f}(k)} \widehat{f}(k') e^{-kx} e^{-k^2 s} e^{-k'x} e^{-k'^2 s} e^{-\frac{x^2}{2s}} e^{-m^2 \frac{s}{2}} \\
&\quad \times \left(\int_{\mathbb{S}} e^{-ik\theta} e^{ik'\theta} e^{im\theta} d\theta \right) dx .
\end{aligned}$$

L'intégrale sur θ donne $m = k - k'$ et introduit un facteur 2π qui s'élimine avec le $\frac{1}{2\pi}$. On se retrouve donc avec

$$\frac{1}{\sqrt{2\pi s}} \sum_{k, k', m \in \mathbb{Z}} \overline{\widehat{f}(k)} \widehat{f}(k') e^{-k^2 s} e^{-k'^2 s} e^{-(k-k')^2 \frac{s}{2}} \int_{\mathbb{R}} e^{-kx} e^{-k'x} e^{-\frac{x^2}{2s}} dx$$

c'est-à-dire une exponentielle d'exposant $-k^2 s - k'^2 s - (k - k')^2 \frac{s}{2}$ et une intégrale gaussienne qui vaut $2\sqrt{\pi\frac{s}{2}} e^{(k+k')^2 \frac{s}{2}}$. Le $2\sqrt{\pi\frac{s}{2}} = \sqrt{2\pi s}$ s'élimine et il reste une exponentielle d'exposant

$$-k^2 s - k'^2 s - (k - k')^2 \frac{s}{2} + (k + k')^2 \frac{s}{2} = -(k - k')^2 s .$$

Bref,

$$\|B_s^\rho f\|_\rho^2 = \sum_{k, k', m \in \mathbb{Z}} \overline{\widehat{f}(k)} \widehat{f}(k') e^{-(k-k')^2 s} = \|f(\theta)\|_{\rho_{\mathbb{S}}}^2 .$$

5.10.10. *La transformée de Bargmann de $SE(2)$.*

Nous allons voir à la section 8.3 comment Davide Barbieri et Giovanna Citti ont défini et utilisé la transformée de Bargmann non plus de $\mathbb{V}_J = \mathbb{H}_{pol}$ (groupe nilpotent) mais de $\mathbb{V}_S = SE(2)$ (groupe résoluble). Toutefois donnons d'abord quelques précisions à propos de l'analyse harmonique sur les groupes de Lie nilpotents plus généraux que le groupe de Heisenberg. Cela est un peu technique mais intéressant car on y voit bien fonctionner des arguments *structuraux*.

6. Analyse harmonique sur les groupes nilpotents

Nous avons vu à la section 8 du chapitre 6 le lien entre la nilpotence d'une algèbre de Lie \mathcal{G} et l'existence de dilatations. Rappelons que si $A(t)$ est un groupe

à 1-paramètre d'automorphismes de \mathcal{G} , la dérivée en $t = 0$, $D(X) = \left. \frac{d}{dt} A(t) X \right|_{t=0}$ est une dérivation de \mathcal{G} et que si α et β sont deux valeurs propres dans $\text{Spec}(D)$ d'espaces propres \mathcal{G}_α et \mathcal{G}_β alors $[\mathcal{G}_\alpha, \mathcal{G}_\beta] \subset \mathcal{G}_{\alpha+\beta}$. $A(t)$ est un groupe de dilatations si $\text{Spec}(D) \subset \mathbb{R}^+$.

6.1. Niveau 2

Supposons d'abord que G soit de niveau 2, i.e. que la suite centrale descendante de \mathcal{G} , $\mathcal{G}_{(0)} = \mathcal{G}$, $\mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}_{(0)}] \subset \mathcal{Z}$ (le centre de \mathcal{G}) et $\mathcal{G}_{(2)} = [\mathcal{G}, \mathcal{G}_{(1)}] = 0$, se termine à $\mathcal{G}_{(2)}$ (cf. chapitre 5 section 4 et chapitre 7 section 1). Si \mathcal{K} est un supplémentaire de $\mathcal{G}_{(1)}$ et si l'on écrit les W de \mathcal{G} sous la forme $W = X + Y$ avec $X \in \mathcal{K}$ et $Y \in \mathcal{G}_{(1)}$, on peut définir un $A(t)$ par $A(t)(W) = e^t X + e^{2t} Y$. $A(t)$ attribue un poids 1 aux X et un poids 2 aux Y et, par construction, les X (resp. les Y) en sont des vecteurs propres pour la valeur propre 1 (resp. 2). Il est bien un automorphisme de \mathcal{G} car d'un côté

$$\begin{aligned} A(t)[X + Y, X' + Y'] &= A(t)([X, X'] + [Y, Y'] + [X, Y'] + [Y, X']) \\ &= e^{2t}[X, X'] \end{aligned}$$

puisque $[Y, Y'], [X, Y'], [Y, X'] \in \mathcal{G}_{(2)} = 0$, alors que d'un autre côté

$$[A(t)(X + Y), A(t)(X' + Y')] = [e^t X + e^{2t} Y, e^t X' + e^{2t} Y'] = e^{2t}[X, X'] .$$

Bref, $\text{Spec}(D) = \{1, 2\}$, $\mathcal{G}_1 = \mathcal{K}$ et $\mathcal{G}_2 = \mathcal{G}_{(1)}$ et $\mathcal{G} = \mathcal{G}_1 \oplus \mathcal{G}_2$.⁶⁰

Ces groupes nilpotents de niveau 2 ont tous une structure proche de celle du groupe de Heisenberg qui en est le prototype.⁶¹ Soit en effet \mathcal{S} un sous-espace linéaire de codimension 1 de $\mathcal{G}_2 = \mathcal{G}_{(1)}$. $\mathcal{G}' = \mathcal{G}/\mathcal{S}$ est nilpotent de niveau 2 avec $\mathcal{G}'_{(1)}$ de dimension 1 et $\mathcal{K}' = \mathcal{K}$. Soit $T \in \mathcal{G}'_{(1)}$, $T \neq 0$ alors, pour $K, K' \in \mathcal{K} = \mathcal{K}'$, $[K, K'] = B(K, K')T$, $B(K, K')$ étant une forme bilinéaire antisymétrique. Soit alors \mathcal{I} l'espace des éléments I de \mathcal{K} tels que $B(I, K') = 0$ pour tout K' de \mathcal{K} et \mathcal{J} un supplémentaire de \mathcal{I} dans \mathcal{K} . Alors la forme B est non dégénérée sur \mathcal{J} car les possibilités de dégénérescence ont été éliminées avec \mathcal{I} .⁶² Cela impose que \mathcal{J} soit de dimension paire $2n$ et alors $\mathcal{J} \oplus \mathcal{G}'_{(1)}$ est isomorphe à \mathfrak{h}^n et $\mathcal{G}' \simeq \mathfrak{h}^n \oplus \mathcal{I}$. Mais comme \mathcal{I} est commutative puisque $[I, I'] = B(I, I')T = 0$ si $I, I' \in \mathcal{I}$, on a $\mathcal{I} \simeq \mathbb{R}^k$ et donc en définitive $\mathcal{G}/\mathcal{S} = \mathcal{G}' \simeq \mathfrak{h}^n \oplus \mathbb{R}^k$.

60. La confusion entre les indices 2 et (1) est inévitable car le 2 est la valeur propre 2 et le (1) est le terme 1 de la suite centrale descendante $\mathcal{G}_{(j)}$.

61. Pour qu'il n'y ait pas de confusion de notations entre l'espace de Hilbert \mathcal{H} et l'algèbre de Lie de \mathbb{H} , nous notons dans la suite cette dernière \mathfrak{h} au lieu de \mathcal{H} .

62. Si $J \neq 0 \in \mathcal{J}$ il existe nécessairement un $J' \in \mathcal{J}$ tel que $B(J, J') \neq 0$ (condition de non-dégénérescence) car si pour tout $J' \in \mathcal{J}$ on avait $B(J, J') = 0$, alors, pour tout $K = J' + I'' \in \mathcal{K}$, on aurait $B(J, K) = B(J, J' + I'') = B(J, J') + B(J, I'') = 0 + 0$ (car $I'' \in \mathcal{I}$) et donc $J \in \mathcal{I}$ ce qui est une contradiction.

En ce qui concerne les représentations, les unirreps de G nilpotent (simplement connexe) dans \mathcal{H} vont donner une représentation ρ de $\mathcal{G} = \mathcal{K} \oplus \mathcal{G}_{(1)} = \mathcal{G}_1 \oplus \mathcal{G}_2$, $\mathcal{G}_{(1)} \subset \mathcal{Z}$, dans \mathcal{H} . Comme pour Heisenberg, ρ doit agir de façon *scalaire* sur $\mathcal{G}_{(1)} = \mathcal{G}_2$ et donc soit $\rho|_{\mathcal{G}_{(1)}} = 0$ soit $\ker \rho|_{\mathcal{G}_{(1)}}$ est de codimension 1. Si $\rho|_{\mathcal{G}_{(1)}} = 0$, on peut quotienter G par le sous-groupe $G_{(1)}$ de G engendré par $\mathcal{G}_{(1)}$ et comme $G/G_{(1)}$ est abélien par définition de $\mathcal{G}_{(1)} = [\mathcal{G}, \mathcal{G}]$, l'irréductibilité de ρ impose $\dim(\mathcal{H}) = 1$ et l'on connaît déjà ces représentations. Si en revanche $\ker \rho|_{\mathcal{G}_{(1)}}$ est de codimension 1, on peut quotienter G par le sous-groupe G_K de G engendré par ce noyau, mais d'après ce que nous venons de voir, $G/G_K \simeq \mathbb{H}^n \oplus \mathbb{R}^k$ et l'on connaît les représentations par Stone-von Neumann.

6.2. Niveau > 2

Lorsque le niveau de G (connexe, simplement connexe) est > 2 , les calculs sont évidemment plus compliqués et nous renvoyons le lecteur intéressé par exemple à Taylor [524]. Les idées principales sont les suivantes. Elles sont intéressantes à esquisser car on y voit très bien opérer la richesse d'hypothèses simples sur les commutateurs. On utilise à la fois la suite centrale descendante $\mathcal{G}_{(j)}$ et le "drapeau" d'idéaux \mathcal{G}_k du théorème de structure de la section 1 du chapitre 7. Rappelons qu'il s'agit d'une suite croissante d'idéaux \mathcal{G}_j de dimension $j = 0, 1, \dots, n = \dim \mathcal{G}$ tels que $[\mathcal{G}, \mathcal{G}_j] \subset \mathcal{G}_{j-1}$. En particulier $[\mathcal{G}, \mathcal{G}_1] = \mathcal{G}_0 = \{0\}$ et donc \mathcal{G}_1 est inclus dans le centre \mathcal{Z} de \mathcal{G} et $\dim \mathcal{Z} \geq \dim \mathcal{G}_1 = 1$.

1. L'unirrep ρ de G , $\rho : G \rightarrow \mathcal{U}(\mathcal{H})$ représente \mathcal{G} par des opérateurs anti-hermitiens sur \mathcal{H} et, en particulier, représente \mathcal{Z} par des scalaires, i.e. si $Z \in \mathcal{Z}$, $\rho(Z) = i\alpha(Z)$ avec α une forme linéaire sur \mathcal{Z} . Soit \mathcal{Z}_1 le noyau $\ker(\alpha)$ de α et $G_{\mathcal{Z}_1}$ le sous-groupe de G engendré par \mathcal{Z}_1 . ρ vient d'une unirrep ρ_1 de $G/G_{\mathcal{Z}_1}$ par la composition $G \twoheadrightarrow G/G_{\mathcal{Z}_1} \rightarrow \mathcal{U}(\mathcal{H})$. On peut donc supposer $\dim \mathcal{Z} = 1$ (i.e. $\mathcal{Z} = \mathcal{G}_1$) et $\alpha \neq 0$. Soit Z un générateur de \mathcal{Z} .
2. On construit alors une sous-algèbre de \mathcal{G} isomorphe à celle de Heisenberg \mathfrak{h} . Soit $Y \in \mathcal{G}_2 - \mathcal{G}_1$. Comme $\mathcal{G}_1 = \mathcal{Z}$, il existe un $X \in \mathcal{G}$ ne commutant pas avec Y . Comme $Y \in \mathcal{G}_2$, $[X, Y] \in \mathcal{G}_1 = \mathcal{Z}$ et donc, en rescalant, on peut avoir $[X, Y] = Z$. On a évidemment $[Z, Y] = 0$. On considère alors le sous-espace \mathcal{W} des W commutant avec Y . Évidemment, $Z, Y \in \mathcal{W}$ et $X \notin \mathcal{W}$ puisque $[X, Y] = Z \neq \{0\}$. Par ailleurs \mathcal{W} est une algèbre de Lie à cause de l'identité de Jacobi.⁶³ C'est même un idéal de \mathcal{G} car si $W \in \mathcal{W}$ et $T \in \mathcal{G}$, alors $[[T, W], Y] + [[W, Y], T] + [[Y, T], W] = 0$ par Jacobi, mais $[[W, Y], T] = 0$ car $W \in \mathcal{W}$ et $[[Y, T], W] = 0$ car $[Y, T] \in \mathcal{Z}$. Par définition, on a $Y \in \mathcal{Z}_{\mathcal{W}}$. On a également $Z \in \mathcal{Z} \subset \mathcal{Z}_{\mathcal{W}}$. En fait on a $\mathcal{G} = \mathbb{R}X + \mathcal{W}$ car comme $[\mathcal{G}, \mathcal{G}_2] \subset \mathcal{G}_1$,

63. $[[W, W'], Y] + [[W', Y], W] + [[Y, W], W'] = 0$, mais $[[W', Y], W] = [[Y, W], W'] = 0$ si $W, W' \in \mathcal{W}$ et donc $[W, W'] \in \mathcal{W}$.

que \mathcal{G}_1 est de dimension 1 et que $Y \in \mathcal{G}_2$, la codimension de \mathcal{W} est ≤ 1 . Mais $X \notin \mathcal{W}$ et donc $\mathbb{R}X + \mathcal{W} = \mathcal{G}$.

3. Il est alors facile de voir que le sous-espace $\text{Span}(Z, Y, X) = \mathfrak{h}_{\mathcal{G}}$ engendré par Z, Y, X est une sous-algèbre isomorphe à \mathfrak{h} puisqu'elle satisfait les relations de commutation caractéristiques $[X, Y] = Z, [X, Z] = 0, [Y, Z] = 0$. D'après le théorème de Stone-von Neumann on connaît donc le type de représentation de ρ restreinte à $\mathfrak{h}_{\mathcal{G}}$.
4. Le problème est maintenant de comprendre la structure de ρ sur le reste de \mathcal{W} (on la connaît sur $\text{Span}(Z, Y) \subset \mathcal{W}$ et sur $\text{Span}(X)$). Pour cela on utilise la théorie des "représentations induites". D'après Stone-von Neumann, il existe une équivalence unitaire entre le Hilbert \mathcal{H} de ρ et un Hilbert $L^2(\mathbb{R}_{(s)}, \mathcal{H}^\bullet)$ transformant l'unirrep ρ en une unirrep de Schrödinger ρ' de paramètre λ à valeurs non pas dans \mathbb{C} mais dans un Hilbert \mathcal{H}^\bullet tenant compte du reste de \mathcal{W} . Soit alors $G_{\mathcal{W}}$ le sous-groupe de G exponentiant \mathcal{W} . Comme $[W, Y] = 0$ si $W \in \mathcal{W}$, si $g_W \in G_{\mathcal{W}}$ alors g_W commute avec tous les $e^{\eta Y}$. Pour les $f(s) \in L^2(\mathbb{R}_{(s)}, \mathcal{H}^\bullet)$, $\rho'(e^{\eta Y})$ agit comme la multiplication par $e^{i\lambda\eta}$ et $\rho'(e^{\xi X})$ comme la translation par ξ de $\mathbb{R}_{(s)}$. Comme $\rho'(g_W)$ commute avec toutes les multiplications, on a $\rho'(g_W) f(s) = U(g_W, s) f(s)$ avec $U(g_W, s) \in \mathcal{U}(\mathcal{H}^\bullet)$ un automorphisme unitaire de \mathcal{H}^\bullet . En utilisant le fait que \mathcal{W} est non seulement une sous-algèbre de \mathcal{G} mais également un idéal, on en tire que $U(g_W g'_W, s) = U(g_W, s) U(g'_W, s)$ et que donc, pour $s = 0$, $U_0(g_W) = U(g_W, 0)$. On obtient ainsi une unirrep de $G_{\mathcal{W}}$ dans \mathcal{H}^\bullet qui encode ce qu'il y a en plus d'une représentation de Schrödinger dans ρ .
5. C'est là qu'intervient le concept technique de représentation induite. On considère les applications mesurables $u : G \rightarrow \mathcal{H}^\bullet$ qui satisfont

$$u(gg_W) = U_0(g_W^{-1}) u(g) .$$

Comme U_0 est unitaire, la norme $\|u(g)\|_{\mathcal{H}^\bullet}$ de $u(g)$ dans \mathcal{H}^\bullet est constante sur les orbites et on peut imposer en plus aux applications u d'être de carré intégrable sur le quotient $G/G_{\mathcal{W}}$ en intégrant sur une section transverse des orbites. On obtient ainsi un nouvel espace de Hilbert \mathcal{H}_0 . La représentation de G induite par U_0 dans \mathcal{H}_0 est la représentation unitaire U_G définie pour $u(h) \in \mathcal{H}_0$ par

$$U_G(g) u(g') = u(g^{-1}g') .$$

6. On notera que, comme $\mathcal{G} = \mathcal{W} + \text{Span}(X)$, $G/G_{\mathcal{W}}$ exponentie X , i.e. est le sous groupe à 1-paramètre des $e^{\xi X}$ sur lequel ρ' agit comme la translation par ξ .

7. On considère alors l'application entre Hilbert $K : \mathcal{H}_0 \rightarrow L^2(\mathbb{R}_{(s)}, \mathcal{H}^\bullet)$ définie pour $u : G \rightarrow \mathcal{H}^\bullet$ par

$$(Ku)(\xi) = u(e^{\xi X}) .$$

Cette définition est cohérente car $u(e^{\xi X}) \in \mathcal{H}^\bullet$. Et on montre que K entrelace U_G et ρ' , autrement dit que si $u \in \mathcal{H}_0$ on a

$$(KU_G(g_W)u)(\xi) = (\rho'(g_W)Ku)(\xi) = u(g_W^{-1}e^{\xi X}) \in \mathcal{H}^\bullet .$$

6.3. Représentations induites (précisions)

Nous venons de voir un exemple de représentation induite. Donnons quelques précisions en suivant le survey classique d'Alexandre Kirillov *Introduction to the Theory of Representations and Noncommutative Harmonic Analysis* [290]. La structure générale de l'argumentation est la suivante mais sa mise en œuvre soulève nombre de détails techniques.

On a comme données initiales le groupe G , un sous-groupe fermé $H = G_W$ de G , et une représentation $\rho_H(h) = U_0(g_W)$ de H dans un espace \mathcal{H}^\bullet . On en induit une représentation $\rho_G(g)u(g') = u(g^{-1}g')$ de G dans l'espace \mathcal{H}_0 des $u : G \rightarrow \mathcal{H}^\bullet$ satisfaisant $u(gh) = \rho_H(h^{-1})u(g)$ pour tout $h \in H$.⁶⁴ Cette représentation induite est traditionnellement notée $\rho_G = \text{Ind}_H^G \rho_H$. Elle "monte" de H en indice vers G en exposant. En considérant leur graphe $\{(g \in G, u(g) \in \mathcal{H}^\bullet)\}$, on peut interpréter les $u : G \rightarrow \mathcal{H}^\bullet$ comme des sections de la fibration (triviale) $G \times \mathcal{H}^\bullet \rightarrow G$ de fibre \mathcal{H}^\bullet sur G . Si l'on regarde l'espace homogène⁶⁵ G/H des classes d'équivalence $\mathfrak{c} = gH$, u varie dans la fibre \mathcal{H}^\bullet en accord avec la représentation $\rho_H(h^{-1})$. Si $\mathcal{F}_{G/H}$ est l'espace des fonctions de G/H dans \mathbb{C} , alors l'espace \mathcal{H}_0 de $\rho_G = \text{Ind}_H^G \rho_H$ est identifiable à $\mathcal{F}_{G/H} \otimes \mathcal{H}^\bullet$, i.e. l'espace des $\Phi : \mathcal{F}_{G/H} \rightarrow \mathcal{H}^\bullet$. Pour le voir, on prend une section $s : G/H \rightarrow G$ de la projection canonique $G \rightarrow G/H$ qui sélectionne un $s(\mathfrak{c})$ dans chaque classe d'équivalence (avec évidemment $s(H) = e_G$) et qui définit un isomorphisme $G \simeq (G/H) \times H$: si $g \in G$ et $\mathfrak{c} = gH$, on prend pour $s(\mathfrak{c})$ le $h \in H$ (unique) tel que $g = s(\mathfrak{c})h$ et on pose $g \sim (s(\mathfrak{c}), h)$. On définit alors $\Phi \in \mathcal{F}_{G/H} \otimes \mathcal{H}^\bullet$ par $\Phi(\mathfrak{c}) = u(s(\mathfrak{c})) \in \mathcal{H}^\bullet$.

Regardons alors comment G opère par translations sur cette représentation. Si $\mathfrak{c} \in G/H$ est une classe d'équivalence $\mathfrak{c} = s(\mathfrak{c})H$ et si l'on considère la classe $g\mathfrak{c} = gs(\mathfrak{c})H$ translatée par g , on a

$$gs(\mathfrak{c}) = s(g\mathfrak{c})C(\mathfrak{c}, g)$$

64. Si l'on préfère, on peut prendre les u satisfaisant $u(hg) = \rho_H(h)u(g)$ et faire agir $\rho_G(g)$ par $\rho_G(g)u(g') = u(g'g)$.

65. Rappelons que ce n'est pas un groupe quotient si H n'est pas un sous-groupe normal de G .

où C est une application bien définie de $(G/H) \times G$ dans H . À cause de l'associativité du produit de G , $C(\mathfrak{c}, g)$ doit satisfaire une *condition de cocycle* analogue à celle déjà rencontrée à la section 2.1.2 à propos des transformées de Gabor. Considérons en effet $g'g$ dans l'équation précédente. On a

$$g'(gs(\mathfrak{c})) = g's(g\mathfrak{c})C(\mathfrak{c}, g) = s(g'g\mathfrak{c})C(\mathfrak{c}, g)C(g\mathfrak{c}, g')$$

et par ailleurs

$$(g'g)s(\mathfrak{c}) = s(g'g\mathfrak{c})C(\mathfrak{c}, g'g) .$$

D'où la condition de cocycle

$$C(\mathfrak{c}, g'g) = C(\mathfrak{c}, g)C(g\mathfrak{c}, g') .$$

Si $u \in \mathcal{H}_0$, qui s'identifie, nous venons de le voir, à $\Phi \in \mathcal{F}_{G/H} \otimes \mathcal{H}^\bullet$ par $\Phi(\mathfrak{c}) = u(s(\mathfrak{c}))$, alors $\rho_G = \text{Ind}_H^G \rho_H$ s'identifie à la représentation

$$\rho_G(\Phi)(\mathfrak{c}) = \rho_H(C(\mathfrak{c}, g))\Phi(g\mathfrak{c}) .$$

Pour mettre en œuvre cette argumentation générale, il faut résoudre quelques problèmes techniques. $\mathcal{F}_{G/H}$ devrait pouvoir être défini comme l'espace des fonctions de carré intégrable $L^2(G/H)$ pour une mesure quotient de la mesure de Haar de G . Ce n'est pas évident car les espaces homogènes d'orbites G/H peuvent être très compliqués et il faut souvent modifier ce Hilbert en prenant les "demi-densités" sur G/H . Il faudrait aussi pouvoir définir une "bonne" section globale transverse aux orbites $s : G/H \rightarrow G$ et des "bonnes" mesures d'intégration pour récupérer la mesure de Haar de G lorsqu'on écrit $G \simeq (G/H) \times H$. Ce n'est pas évident si G n'est pas unimodulaire et possède deux mesures de Haar, à gauche et à droite, différentes.

On peut interpréter plus abstraitement et conceptuellement $\text{Ind}_H^G \rho_H$ en termes des catégories des H -espaces et des G -espaces. En effet, la représentation ρ_H fait de \mathcal{H}^\bullet un H -espace et, si $\eta^\bullet \in \mathcal{H}^\bullet$, $\rho_H(h)(\eta^\bullet)$ peut se noter $h\eta^\bullet$. Or, dans la catégorie des H -espaces, il existe des produits directs $X \times Y$ (qui sont trivialement des H -espaces avec l'action $h(x, y) = (hx, hy)$) et des produits *fibrés* $X \times_H Y$ qui sont par définition les quotients du produit direct par l'action de H . Dans ce qui suit, il est plus commode de considérer que X est un H -espace à droite et Y un H -espace à gauche.⁶⁶ H agit alors par $h(x, y) = (xh^{-1}, hy)$ et l'équivalence $(x, y) \sim (x', y')$ signifie qu'il existe $h \in H$ tel que $x' = xh^{-1}$ et $y' = hy$ (i.e. $(x', y') = h(x, y)$).⁶⁷ En remplaçant x par xh , cela est équivalent à $(xh, y) \sim (x, hy)$. Notons $(x, y)_H$ la classe d'équivalence de (x, y) dans $X \times_H Y$.

66. Ou l'inverse, les changements à apporter dans ce qui suit étant évidents.

67. C'est bien une action de H puisque $(h'h)(x, y) = (x(h'h)^{-1}, (h'h)y) = (xh^{-1}h'^{-1}, h'hy) = h'(h(x, y))$.

Considérons alors les $u : G \rightarrow \mathcal{H}^\bullet$. Leurs graphes $\Gamma = \{(g, u(g) = \eta^\bullet)\}$ sont dans $G \times \mathcal{H}^\bullet$ et la condition $u(gh) = \rho_H(h^{-1})u(g)$, signifie que

$$(g, u(g) = \eta^\bullet) \sim (gh, h^{-1}u(g) = h^{-1}\eta^\bullet)$$

En effet $(gh, h^{-1}u(g))$ est de type (xh^{-1}, hy) avec h remplacé par h^{-1} , $x = g$, $y = u(g)$ et $(x', y') = h(x, y)$ devient l'identité $(gh, h^{-1}u(g)) = (xh, h^{-1}y) = (gh, h^{-1}u(g))$. Γ est ainsi dans le produit fibré $G \times_H \mathcal{H}^\bullet$ où G est considéré comme un H -espace à droite et \mathcal{H}^\bullet comme un H -espace à gauche. On utilise alors le fait que G est non seulement un H -espace à droite par translations à gauche mais aussi un G -espace à gauche par translations à droite et que les translations à droite et à gauche commutent. Cela permet de définir une G -action à gauche sur $G \times_H \mathcal{H}^\bullet$ par

$$g_0(g, \eta^\bullet)_H = (g_0g, \eta^\bullet)_H .$$

Cela marche bien car si $(g, \eta^\bullet) \sim (g', \eta'^\bullet)$ à travers $h \in H$, le même h établit une équivalence entre (g_0g, η^\bullet) et (g_0g', η'^\bullet) . C'est toujours le cas si X est non seulement un H -espace à droite mais également un G -espace à gauche.

On peut alors interpréter de façon conceptuelle et structurelle les représentations induites en termes de théorie des catégories. On a deux catégories, celle \mathcal{E}_G des G -espaces et celle \mathcal{E}_H des H -espaces. On a un foncteur "restriction" ⁶⁸ évident Res_H^G de \mathcal{E}_G dans \mathcal{E}_H qui consiste tout simplement à restreindre l'action de G à l'action du sous-groupe H . Le résultat structurel profond – qui remonte au *théorème de réciprocity de Frobenius* – est que ce foncteur Res_H^G admet un *adjoint à gauche* Ind_H^G qui associe à tout H -espace Y le G -espace $G \times_H Y$. Cela signifie que pour tout H -espace Y et tout G -espace X on a un isomorphisme ⁶⁹

$$\text{Hom}_{\mathcal{E}_H}(Y, \text{Res}_H^G(X)) \simeq \text{Hom}_{\mathcal{E}_G}(\text{Ind}_H^G(Y), X) .$$

Cela peut se voir assez facilement. Soit

$$\varphi_H \in \text{Hom}_{\mathcal{E}_H}(Y, \text{Res}_H^G(X)) ,$$

c'est-à-dire une application $\varphi_H : Y \rightarrow X$ telle que $\varphi_H(hy) = h\varphi_H(y)$ pour tout $h \in H$ (hy vient de l'action de H dans Y et $h\varphi_H(y)$ de l'action de H dans X). Soit également dans l'autre sens

$$\psi_G \in \text{Hom}_{\mathcal{E}_G}(\text{Ind}_H^G(Y), X) ,$$

c'est-à-dire une application $\psi_G : G \times_H Y \rightarrow X$ telle que

$$\psi_G((g'g, y)_H) = g'\psi_G((g, y)_H)$$

68. C'est-à-dire une transformation de catégories envoyant objets sur objets et morphismes sur morphismes en respectant la structure de catégorie.

69. Rappelons que dans une catégorie \mathcal{C} on note traditionnellement $\text{Hom}_{\mathcal{C}}(X, Y)$ l'ensemble des morphismes de X dans Y .

pour tout $g' \in G$, et donc

$$\psi_G((g, y)_H) = g\psi_G((e_G, y)_H)$$

(rappelons que e_G est l'élément neutre de G). L'adjonction est alors définie par

$$\begin{cases} \varphi_H(y) = \psi_G((e_G, y)_H) \\ \psi_G((g, y)_H) = g\varphi_H(y) \end{cases}$$

On notera que si l'on considère la représentation triviale I de H , alors la représentation induite est la représentation régulière de G dans l'espace des fonctions sur G constantes sur les orbites de l'action de H sur G , autrement dit sur les fonctions définies sur G/H .

Remarque. La notion d'imprimitivité que nous avons rencontrée à la section 5.6 dans la preuve du théorème de Stone-von Neumann trouve sa source dans la théorie des G -espaces. Soit $\varphi : X \rightarrow Y$ un morphisme de G -espaces. Si $y \in Y$ et si $G(y)$ est le stabilisateur de y dans Y , l'image réciproque $\varphi^{-1}(y) = X_y$ (la "fibre" de φ au-dessus de Y) est un $G(y)$ -espace puisque si $x \in X_y$ et $g \in G(y)$, $\varphi(gx) = g\varphi(x) = gy = y$ et donc $gx \in X_y$. Si X est un G -espace homogène $X \simeq G/H$ (i.e. si G agit transitivement sur X et qu'il n'y a donc qu'une seule orbite de stabilisateur H), alors $X = \text{Ind}_{G(y)}^G(X_y) = G \times_{G(y)} X_y$. Les fibres X_y lorsque Y n'est pas réduit à un point forment un système d'imprimitivité pour X . \square

6.4. De Mackey à Kirillov (méthode des orbites co-adjointes)

Nous avons noté plusieurs fois la relation très étroite des unirreps de groupes comme le groupe de Heisenberg avec les orbites de la représentation co-adjointe. Ce lien est en fait une *nécessité* impliquée par un profond théorème de Kirillov sur les groupes nilpotents et, plus généralement (comme l'a montré ensuite Bertram Kostant), sur les groupes résolubles, qui utilise d'un côté la structure symplectique canonique de ces orbites que nous avons introduite à la section 6.3 du chapitre 5 et, d'un autre côté, la notion de représentation induite. Explicitons-la de façon un peu plus générale et abstraite de façon à avoir un aperçu sur la façon dont les théories peuvent se généraliser au-delà des cas particuliers concrets élémentaires qui nous motivent dans cet ouvrage. Le lecteur intéressé pourra consulter les textes classiques de Kirillov lui-même [288], [290], [291] et le séminaire Bourbaki de 1963 d'Alain Guichardet [235] qui les présentait à partir de la version russe initiale de 1962.

L'intuition théorique initiale est que si les orbites co-adjointes sont de façon naturelle des variétés symplectiques, elles peuvent être considérées comme des espaces des phases G -invariants. Mais alors, on peut essayer de construire des systèmes quantiques, également G -invariants, qui les quantifient ("quantification géométrique"). De tels systèmes seront décrits par des unirreps de G . Alexandre Kirillov a d'abord

montré que cela était possible pour les groupes nilpotents sans conditions restrictives sur les orbites co-adjointes. Ses résultats ont ensuite été généralisés par Kostant et d'autres spécialistes aux groupes résolubles, puis, avec des conditions restrictives sur les orbites co-adjointes, à d'autres classes de groupes. La méthode de Kirillov consiste à relier la transformée de Fourier d'une orbite co-adjointe \mathcal{O} , c'est-à-dire la distribution de Dirac $\delta_{\mathcal{O}}$ de support \mathcal{O} , avec le caractère "infinitésimal" d'une unirrep (cf. section 3.4 du chapitre 7). Comme le note Kirillov dans ses *Lectures* [291] (p. xvii) :

"The idea behind the orbit method is to unite harmonic analysis with symplectic geometry."

Soit G un groupe de Lie nilpotent (supposé être connexe et simplement connexe pour assurer que l'exponentielle \exp est un difféomorphisme global). Il peut être réalisé par des matrices triangulaires supérieures non strictes dont tous les termes diagonaux sont $= 1$. Soit $\varpi \in \mathcal{G}^*$ une forme linéaire sur \mathcal{G} . Comme \mathcal{G} est une algèbre de Lie, il est naturel de regarder comment ϖ agit sur les commutateurs de \mathcal{G} : $\sigma_{\varpi}(\mathbf{t}, \mathbf{u}) = -\langle \varpi, [\mathbf{t}, \mathbf{u}] \rangle$ est une forme bilinéaire antisymétrique. On dit qu'une sous-algèbre \mathcal{S} de \mathcal{G} est "subordonnée" à ϖ si

$$\sigma_{\varpi}(\mathfrak{s}, \mathfrak{u}) = -\langle \varpi, [\mathfrak{s}, \mathfrak{u}] \rangle = 0$$

pour tous $\mathfrak{s}, \mathfrak{u} \in \mathcal{S}$, autrement dit si \mathcal{S} est *isotrope* pour la 2-forme σ_{ϖ} , ou encore si $[\mathcal{S}, \mathcal{S}] \subset \ker(\varpi)$ (ce noyau étant de codimension 1 si $\varpi \neq 0$). Comme nous l'avons vu à la section 6.3 du chapitre 5, si \mathcal{O} est l'orbite co-adjointe de ϖ , les $\mathbf{t} \in \mathcal{G}$ induisent des vecteurs tangents \mathbf{t}_{ϖ}^* de l'espace tangent $T_{\varpi}\mathcal{O}$ à \mathcal{O} en ϖ et $\sigma_{\varpi}(\mathbf{t}, \mathbf{u}) = -\langle \varpi, [\mathbf{t}, \mathbf{u}] \rangle$ devient la 2-forme symplectique de Kirillov

$$\sigma_{\mathcal{O}}(\mathbf{t}_{\varpi}^*, \mathbf{u}_{\varpi}^*) = -\langle \varpi, [\mathbf{t}, \mathbf{u}] \rangle$$

que nous avons rencontrée dans cette section 6.3. La condition d'isotropie de \mathcal{S} équivaut alors à l'isotropie de son image \mathcal{S}_{ϖ}^* dans $T_{\varpi}\mathcal{O}$.

Considérons alors la 1-forme $\mathbf{t} \rightarrow \varpi(\mathbf{t}) \in \mathbb{R}$ restreinte à \mathcal{S} . C'est une représentation de \mathcal{S} car, sur \mathcal{S} , $[\varpi(\mathfrak{s}), \varpi(\mathfrak{u})] = 0$ puisque \mathbb{R} est commutatif et par ailleurs $\varpi([\mathfrak{s}, \mathfrak{u}]) = 0$ puisque \mathcal{S} est subordonnée à ϖ . Donc

$$[\varpi(\mathfrak{s}), \varpi(\mathfrak{u})] = \varpi([\mathfrak{s}, \mathfrak{u}]) (= 0)$$

et l'on obtient ainsi une représentation réelle $1D$ (unidimensionnelle) de \mathcal{S} et donc une représentation unitaire $1D$ de $G_{\mathcal{S}} = \exp(\mathcal{S})$, à savoir

$$\chi_{\varpi, \mathcal{S}}(\exp(\mathfrak{s})) = e^{i\langle \varpi, \mathfrak{s} \rangle}.$$

On peut par conséquent considérer la représentation induite $\rho_{\varpi, \mathcal{S}} = \text{Ind}_{G_{\mathcal{S}}}^G(\chi_{\varpi, \mathcal{S}})$. Un théorème fondamental dit alors

Théorème (Kirillov). (G groupe de Lie nilpotent, connexe, simplement connexe) Toute unirrep ρ de G est induite (au sens de la section 6.3) par un $\chi_{\varpi, \mathcal{S}}$, i.e.

$\rho \simeq \rho_{\varpi, \mathcal{S}} = \text{Ind}_{G_{\mathcal{S}}}^G (\chi_{\varpi, \mathcal{S}})$, l'irréductibilité signifiant que \mathcal{S} est une sous-algèbre de \mathcal{G} subordonnée à ϖ de dimension *maximale*. Qui plus est, deux telles représentations sont isomorphes si et seulement si ϖ_1 et ϖ_2 appartiennent à la même orbite co-adjointe. \diamond

Ce théorème se démontre par induction sur la dimension n de G en utilisant de façon cruciale le fait que G soit nilpotent pour élaborer des constructions analogues à celles utilisées dans les deux sections précédentes. Rappelons que nous avons vu à la section 4 du chapitre 5 et surtout à la section 1 du chapitre 7 que, d'après le théorème d'Engel, G est nilpotent si et seulement si tous les ad_X sont nilpotents. À cause précisément de leur nilpotence, les groupes nilpotents et leurs unirreps ont un aspect "polynomial".⁷⁰ Comme nous l'avons vu plus haut dans la section 1.2.2, cela permet de définir les fonctions C^∞ à décroissance polynomiale et à décroissance rapide sur G (espace $\mathcal{C}_\downarrow^\infty(G)$).

Soit $\rho_{\mathcal{O}}$ l'unirrep correspondant à l'orbite co-adjointe \mathcal{O} et soit $f \in \mathcal{C}_\downarrow^\infty(G)$. On sait prolonger $\rho_{\mathcal{O}}$ à $\mathcal{C}_\downarrow^\infty(G)$. On montre que les opérateurs $\rho_{\mathcal{O}}(f)$ sont traçables et que la distribution $f \rightarrow \text{tr}(\rho_{\mathcal{O}}(f))$ donnant le caractère est tempérée. En effet réalisons $\rho_{\mathcal{O}}$ comme ci-dessus. Soit $U \in \mathfrak{U}(\mathcal{G})$ tel que l'opérateur différentiel à coefficients polynomiaux $\rho_{\mathcal{O}}(U)$ ait un inverse possédant une trace (un tel U existe). On peut alors écrire

$$\rho_{\mathcal{O}}(f) = \rho_{\mathcal{O}}(U)^{-1} \rho_{\mathcal{O}}(U) \rho_{\mathcal{O}}(f) = \rho_{\mathcal{O}}(U)^{-1} \rho_{\mathcal{O}}(Uf)$$

U étant considéré comme un opérateur différentiel sur G . Dans la dernière expression, $\rho_{\mathcal{O}}(U)^{-1}$ est traçable et $\rho_{\mathcal{O}}(Uf)$ est borné. Donc $\rho_{\mathcal{O}}(f)$ est traçable. On fait correspondre à cette distribution tempérée une distribution sur \mathcal{G} dont la transformée de Fourier est une mesure sur \mathcal{G}^* qui est ≥ 0 , G -invariante et portée par l'orbite \mathcal{O} . Muni de ces résultats Kirillov démontre le théorème.

Une conséquence directe de ce théorème est que le dual de Pontryagin \widehat{G} de G nilpotent (connexe, simplement connexe) est identifiable à l'espace \mathfrak{D}_G des orbites co-adjointes \mathcal{O} muni de la topologie quotient, la mesure de Plancherel $P(\lambda)$ venant de la mesure sur \mathfrak{D}_G qui vient elle-même de la mesure de Lebesgue sur \mathcal{G}^* décomposée en les mesures sur les \mathcal{O} associées aux formes symplectiques $\sigma_{\mathcal{O}}$. En fait, dans les intégrales utilisant $P(\lambda)$ on peut se restreindre aux orbites dites "en position générale" (les plans $\eta_0^* = \text{cste} \neq 0$ pour le groupe de Heisenberg) et laisser de côté les autres orbites (les points du plan $\eta_0^* = 0$ pour le groupe de Heisenberg) qui forment un ensemble de mesure nulle.

Il existe des opérations de décomposition des unirreps relativement à la structure du groupe G . Soit F un sous-groupe *fermé* connexe de G . L'inclusion des algèbres de Lie $\mathcal{F} \hookrightarrow \mathcal{G}$ définit par dualité une projection (l'opération de restriction des

70. L'analogie est que dans l'espace $\mathcal{C}^\infty(\mathbb{R})$ les polynômes sont les fonctions sur lesquelles l'opérateur différentiel $\frac{d}{dx}$ agit de façon nilpotente. Plus précisément, l'espace \mathcal{P}_k des polynômes de degré k est le noyau de $\frac{d^{k+1}}{dx^{k+1}}$ (que l'on note souvent $(\frac{d}{dx})^{k+1}$).

formes linéaires) $r : \mathcal{G}^* \rightarrow \mathcal{F}^*$. Suivant la façon dont cette projection est “située” par rapport aux orbites co-adjointes de G et F , on obtient différents résultats de décomposition.

1. La restriction de $\rho_{\mathcal{O}_G}$ à F n’est plus forcément irréductible et se décompose suivant les orbites co-adjointes \mathcal{O}_F de F dans \mathcal{F}^* qui sont contenues dans $r(\mathcal{O}_G)$.
2. Pour toute orbite \mathcal{O}_F de F dans \mathcal{F}^* , la représentation induite $\rho_{\mathcal{O}_G} = \text{Ind}_F^G(\rho_{\mathcal{O}_F})$ se décompose suivant les orbites co-adjointes \mathcal{O}_G de G dans \mathcal{G}^* qui rencontrent $r^{-1}(\mathcal{O}_F)$.

Ces résultats se démontrent par récurrence sur la dimension de G . On les applique pour décrire les spectres des $\text{Res}_H^G \rho_{\mathcal{O}_G}$ et des $\text{Ind}_H^G \rho_{\mathcal{O}_F}$: pour $\text{Res}_H^G \rho_{\mathcal{O}_G}$ on décompose $r(\mathcal{O}_G)$ en F -orbites et pour $\text{Ind}_H^G \rho_{\mathcal{O}_F}$ on décompose la “saturation” de $r^{-1}(\mathcal{O}_F)$ en G -orbites.

6.5. Transformée de Fourier non commutative

Une fois que la structure du dual \widehat{G} de G est connue avec ses unirreps ρ_λ dans les Hilbert \mathcal{H}_λ et sa mesure de Plancherel $P(\lambda)$ on peut développer la théorie des transformées de Fourier dans le cas non commutatif. Si f est une fonction sur G intégrable pour la mesure de Haar, sa transformée de Fourier est une fonction $\widehat{f}(\lambda)$ sur \widehat{G} à valeurs opérateurs définie par l’intégrale

$$\widehat{f}(\lambda) = \int_G f(g) \rho_\lambda(g) dg \in \mathcal{U}(\mathcal{H}_\lambda) ,$$

autrement dit pour un λ donné on considère l’unirrep ρ_λ dans le Hilbert \mathcal{H}_λ et on fait la moyenne sur G des opérateurs $\rho_\lambda(g) \in \mathcal{U}(\mathcal{H}_\lambda)$ en les pondérant par $f(g)$. La transformée $\widehat{f}(\lambda)$ analyse $f(g)$ en termes des unirreps de G . Dans les bons cas comme ici, on aura alors une *formule d’inversion* redonnant $f(g)$ par *synthèse* :

$$f(g) = \int_{\widehat{G}} \text{tr} \left(\widehat{f}(\lambda) \rho_\lambda(g)^* \right) dP(\lambda) .$$

$L^2(\widehat{G}, dP(\lambda))$ sera un Hilbert de fonctions φ sur \widehat{G} mais à valeurs opérateurs $\varphi(\lambda) \in \text{End}(\mathcal{H}_\lambda)$, il sera muni de la norme

$$\|\varphi\|^2 = \int_{\widehat{G}} \text{tr}(\varphi(\lambda) \varphi(\lambda)^*) dP(\lambda)$$

dérivant du produit scalaire

$$\langle \varphi, \psi \rangle = \int_{\widehat{G}} \text{tr}(\varphi(\lambda) \psi(\lambda)^*) dP(\lambda)$$

et il existera un isométrie entre $L^2(G, dg)$ et $L^2(\widehat{G}, dP(\lambda))$:

$$\|f(g)\|_{L^2(G, dg)}^2 = \|\widehat{f}(\lambda)\|_{L^2(\widehat{G}, dP(\lambda))}^2 .$$

7. Analyse harmonique sur les groupes (semi-)simples

Mais pour les groupes ni nilpotents, ni résolubles la compréhension de la structure de \widehat{G} est d'une difficulté "himalayesque". Le plus bel exemple est celui des travaux colossaux de Harish-Chandra pour comprendre la structure des unirreps (et surtout de leurs caractères) des groupes (semi-)simples comme $SL(2, \mathbb{R})$ qui, comme nous l'avons souligné à la section 4 du chapitre 5 sont à l'opposé des groupes nilpotents et résolubles. Le lecteur intéressé pourra par exemple se référer au mémoire de 1985 de Langlands sur Harish-Chandra [315] et au beau résumé de Rebecca Herb de 1991 [250]. La théorie est beaucoup trop complexe pour être ne serait-ce qu'esquissée ici. Nous nous bornerons donc à donner quelques très brèves intuitions.

Ce sont surtout les développements de la mécanique quantique qui poussèrent, comme nous l'avons vu avec le groupe de Heisenberg, à approfondir la théorie des unirreps de dimension infinie des groupes fondamentaux de la physique ($SO(3)$, $SU(2)$, Lorentz, Poincaré, etc.) dans des espaces de Hilbert. Cela commença avec Dirac et Wigner à la fin des années 1930 et ouvrit un immense domaine. Pour les groupes de Lie compacts et abéliens les unirreps ρ sont, nous l'avons vu, de dimension finie et les caractères $\chi_\rho(g) = \text{tr}(\rho(g))$ ne posent pas de problème de définition. En revanche pour les unirreps de dimension infinie les opérateurs $\rho(g)$ opérant sur un $L^2(\mathcal{H})$ ne sont pas forcément traçables. Mais, néanmoins, le caractère de ρ reste définissable comme *distribution*, ce qui signifie que l'on peut définir $\chi_\rho(f)$ comme une fonctionnelle linéaire continue sur les $f \in \mathcal{C}_c^\infty(G)$ (les fonctions test C^∞ à support compact sur G). En effet, comme nous l'avons vu, ρ est prolongeable à $\mathcal{C}_c^\infty(G)$ par

$$\rho(f) = \int_G f(g) \rho(g) dg$$

et cet opérateur sur $L^2(\mathcal{H})$ qui moyenne les $\rho(g)$ en les pondérant par une fonction C^∞ à support *compact* est *traçable*. On peut donc définir $\chi_\rho(f)$ par

$$\chi_\rho(f) = \text{tr}(\rho(f)) .$$

La mesure de Plancherel $dP(\rho)$ sur le dual de Pontryagin \widehat{G} est alors la mesure qui satisfait

$$f(e_G) = \int_{\widehat{G}} \chi_\rho(f) dP(\rho)$$

et, plus généralement,

$$f(g) = \int_{\widehat{G}} \chi_\rho(\mathcal{R}(g)f) dP(\rho)$$

(où \mathcal{R} est la représentation régulière à droite).

L'exemple standard est celui du groupe linéaire simple de dimension 3

$$G = SL(2, \mathbb{R})$$

des matrices réelles 2×2 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ de déterminant $ad - bc = 1$, objet du célèbre traité " $SL(2, \mathbb{R})$ " de Serge Lang. Il est connexe, non abélien, non compact et simple en tant que groupe de Lie et son quotient par son centre (qui est le sous-groupe à 2 éléments $\{\pm Id\}$) est le groupe spécial projectif $PSL(2, \mathbb{R})$. La simplicité se voit facilement à partir de l'algèbre de Lie $\mathcal{G} = \mathfrak{sl}(2, \mathbb{R})$, l'algèbre des matrices réelles 2×2 de trace nulle car cette algèbre est simple. En effet la base naturelle de $\mathcal{G} = \mathfrak{sl}(2, \mathbb{R})$ est constituée des matrices

$$t_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, t_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, t_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

qui satisfont les relations de commutation

$$[t_1, t_2] = 2t_2, [t_1, t_3] = -2t_3, [t_2, t_3] = t_1.$$

On notera que

$$J = -t_2 + t_3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

satisfait, comme l'unité imaginaire pure $i \in \mathbb{C}$, $J^2 = -Id$, et

$$\begin{cases} [J, t_1] = -[t_2, t_1] + [t_3, t_1] = 2(t_2 + t_3) \\ [J, t_2] = -[t_2, t_2] + [t_3, t_2] = -t_1 \\ [J, t_3] = -[t_2, t_3] + [t_3, t_3] = -t_1 \end{cases}$$

On peut montrer la simplicité de \mathcal{G} à la main. Soient (α, b, c) les coordonnées de \mathcal{G} dans la base des t_j et considérons un élément $X_0 = (\alpha_0, \beta_0, \gamma_0)$ de \mathcal{G} . Les $[X_0, \mathcal{G}]$ sont engendrés par les $[X_0, t_j]$, i.e. par $(0, -2\beta_0, 2\gamma_0)$, $(\beta_0, 0, -2\alpha_0)$, $(-\gamma_0, 2\alpha_0, 0)$.⁷¹ Donc aucune sous-algèbre de dimension 1, $\text{Span } X$, ne peut être un idéal. De même, si $\text{Span}(X, Y)$ est une sous-algèbre de dimension 2, alors $[\text{Span}(X, Y), \mathcal{G}]$ est \mathcal{G} toute entière et donc $\text{Span}(X, Y)$ ne peut pas être un idéal.

71. À la main. Si $X_0 \neq 0$, l'un de ces vecteurs ne sera pas colinéaire avec X_0 . Si par exemple $\alpha_0 \neq 0$, $(0, -2\beta_0, 2\gamma_0) = \lambda X_0$ si et seulement si $\lambda = 0$ et donc $\beta_0 = \gamma_0 = 0$, mais alors $(\beta_0, 0, -2\alpha_0) = (0, 0, -2\alpha_0)$ et $(-\gamma_0, 2\alpha_0, 0) = (0, 2\alpha_0, 0)$ ne sont pas colinéaires avec X_0 et donc $[X_0, \mathcal{G}] = \mathcal{G}$. Si $\beta_0 \neq 0$, $(\beta_0, 0, -2\alpha_0) = \lambda X_0$ si et seulement si $\lambda = 0$ et donc $\beta_0 = \alpha_0 = 0$, ce qui est contradictoire et donc $(\beta_0, 0, -2\alpha_0)$ n'est pas colinéaire avec X_0 ; de même si $\gamma_0 \neq 0$, $(-\gamma_0, 2\alpha_0, 0) = \lambda X_0$ si et seulement si $\lambda = 0$ et donc $\gamma_0 = \alpha_0 = 0$, ce qui est contradictoire et donc $(-\gamma_0, 2\alpha_0, 0)$ n'est pas colinéaire avec X_0 .

$G = SL(2, \mathbb{R})$ intervient dans de nombreux domaines des mathématiques. C'est le groupe des transformations linéaires du plan qui préserve l'aire orientée mais surtout c'est (en tant que $PSL(2, \mathbb{R})$) le groupe des isométries du plan hyperbolique, c'est-à-dire du demi-plan de Poincaré \mathbb{P} des $z = x + iy$ de partie imaginaire $\Im(z) = y > 0$ muni de la métrique hyperbolique $ds = \frac{|dz|}{y}$ de courbure constante -1 . $PSL(2, \mathbb{R})$ y agit par les homographies $g(z) = \frac{az+b}{cz+d}$. Le polynôme caractéristique de g étant $\Pi(\lambda) = \lambda^2 - \text{tr}(g)\lambda + 1$ (le 1 est le $\det(g)$), les valeurs propres sont

$$\lambda = \frac{\text{tr}(g) \pm \sqrt{\text{tr}(g)^2 - 4}}{2}$$

qui sont complexes conjuguées, égales ou réelles suivant que $|\text{tr}(g)|$ est $< 2, = 2, > 2$. Dans le premier cas g est dit "elliptique" et il est conjugué à une rotation, dans le deuxième cas g est dit "parabolique" et dans le troisième cas g est dit "hyperbolique".

$G = SL(2, \mathbb{R})$ n'a pas de représentations unitaires de dimension finie non triviale. Ses unirreps, qui sont donc de dimension infinie, ont été étudiée en profondeur dès la fin des années 1940 par Israel Gelfand, Mark Aronovitch Naimark, Valentine Bargmann et Harish-Chandra.

La base de la théorie consiste à montrer que l'on peut paramétrer les unirreps de G au-moyen des *sous-groupes abéliens maximaux* de G (ce que l'on appelle ses *sous-groupes de Cartan*). Pour $G = SL(2, \mathbb{R})$ il y a deux tels sous-groupes et ils sont de dimension 1. Le premier, non compact, est celui des matrices diagonales

$$D = \left\{ d(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \middle| a \neq 0 \right\}$$

(qui est isomorphe au groupe multiplicatif $\mathbb{R} - \{0\}$) et le second, compact, celui des rotations

$$R = \left\{ \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = e^{i\theta} \right\} .$$

Ils correspondent aux sous-algèbres abéliennes de \mathcal{G} ,

$$\mathcal{D} = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix} \right\} = \mathbb{R}t_1,$$

$$\mathcal{R} = \left\{ \begin{pmatrix} 0 & -\vartheta \\ \vartheta & 0 \end{pmatrix} \right\} = \mathbb{R}J .$$

En ce qui concerne D , on peut l'étendre au sous-groupe non commutatif de dimension 2 des matrices triangulaires supérieures

$$P = \left\{ p(a, b) = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \middle| a \neq 0 \right\} .$$

Les caractères unitaires χ de D (i.e. les éléments de \widehat{D} puisque D est abélien) sont les $\chi_0(d(a)) = |a|^{i\mu}$ et les $\chi_1(d(a)) = \text{sgn}(a)|a|^{i\mu}$ ($\text{sgn} = \text{“signe”}$, $\mu \in \mathbb{R}$). Ils se prolongent trivialement à P en posant $\chi(p(a,b)) = \chi(d(a))$. Si alors χ est un caractère unitaire de P , on lui associe la représentation induite $\rho = \text{Ind}_P^G \chi$. On obtient ainsi une série de représentations unitaires $\rho_{0,\mu}$ et $\rho_{1,\mu}$ appelée *série principale*, représentations qui sont des unirreps sauf pour $\chi = \text{sgn}(a)$ qui est la somme directe de deux unirreps. La série principale est paramétrée par le paramètre continu $\mu \in \mathbb{R}$.

En ce qui concerne le second sous-groupe de Cartan, celui R des rotations, son dual est, nous l'avons vu, \mathbb{Z} (théorie des séries de Fourier des fonctions périodiques) et l'on obtient, de façon plus compliquée que par des représentations induites, une autre série d'unirreps τ_n appelée *série discrète* et paramétrée par le paramètre discret $n \in \mathbb{Z}$.

Le résultat fondamental est alors le théorème d'Harish-Chandra de 1952 donnant la mesure de Plancherel sur $\widehat{G} = SL(2, \mathbb{R})$ (cf. Herb [250]).

Théorème (Harish-Chandra). Pour tout $f \in \mathcal{C}_c^\infty(G)$, on a

$$f(e_G) = \sum_{n \in \mathbb{Z}} |n| \chi_{\tau_n}(f) + \frac{1}{4} \int_{\mathbb{R}} \chi_{\rho_{0,\mu}}(f) \mu \tanh\left(\frac{\pi\mu}{2}\right) d\mu \\ + \frac{1}{4} \int_{\mathbb{R}} \chi_{\rho_{1,\mu}}(f) \mu \coth\left(\frac{\pi\mu}{2}\right) d\mu .$$

Ce résultat remarquable montre que la série discrète a pour mesure les masses de Dirac $|n|$ pour les τ_n et les mesures, absolument continues par rapport à la mesure de Lebesgue $d\mu$, $\mu \tanh\left(\frac{\pi\mu}{2}\right) d\mu$ et $\mu \coth\left(\frac{\pi\mu}{2}\right) d\mu$ pour la série principale des $\rho_{0,\mu}$ et $\rho_{1,\mu}$. Il y a d'autres unirreps “complémentaires” mais qui sont de mesure nulle.

8. Analyse harmonique et groupe euclidien $SE(2)$

Après ces remarques structurelles plus abstraites et plus générales, revenons à notre second cas particulier de base, celui du groupe euclidien $SE(2)$.

8.1. Unirreps et orbites co-adjointes de $SE(2)$

Pour le groupe des déplacements du plan $SE(2) = G = \mathbb{V}_S = \mathbb{R}^2 \rtimes SO(2)$ (cf. la section 7 du chapitre 5), qui n'est plus nilpotent mais reste résoluble et est un produit semi-direct avec la loi de composition $(s, r_\varphi) \circ (q, r_\alpha) = (s + r_\varphi(q), r_{\varphi+\alpha})$, on montre que les unirreps sont définies sur le Hilbert $\mathcal{H} = L^2(\mathbb{S}^1, \mathbb{C})$, \mathbb{S}^1 étant muni de la mesure de Haar-Lebesgue $d\theta$. Elles sont paramétrées par un réel $\lambda > 0$ et, pour $v = (x, y, \alpha) \in \mathbb{V}_S$, sont de la forme :

$$\begin{aligned} \rho_\lambda : \mathbb{V}_S &\rightarrow \mathcal{U}(\mathcal{H}) \\ v = (x, y, \alpha) &\mapsto \rho_\lambda(v) : \mathcal{H} \rightarrow \mathcal{H} \\ &\psi(\theta) \mapsto e^{i\lambda(x \sin(\theta) + y \cos(\theta))} \psi(\theta + \alpha) . \end{aligned}$$

On remarquera que pour les petits angles θ et α on obtient la représentation

$$e^{i\lambda(x\theta+y)} \psi(\theta + \alpha)$$

qui est de la forme $e^{i\lambda(y+xs)} u(s+p)$ des représentations de \mathbb{V}_J pour $p = \tan(\alpha) \sim \alpha$ et $s = \theta$.

Le dual $\widehat{\mathbb{V}}_S$ de \mathbb{V}_S est l'ensemble de ces unirreps avec la mesure de Plancherel $dP(\lambda) = \lambda d\lambda$

Pour comprendre l'origine de cette forme des unirreps il faut revenir à la structure des orbites de la représentation co-adjointe de G et à leur lien avec la structure de produit semi-direct expliquée à la section 3 du chapitre 11.

Remarque. Il y a un petit changement de notations, $v = (x, y, \alpha)$ étant le $v = (q, \theta)$ avec $q = (x, y)$ et le θ étant réécrit α . Nous allons voir pourquoi. \square

Soit $(\lambda_0^* \neq 0, \mu_0^*) \in \mathcal{G}^*$. Rappelons que l'orbite co-adjointe est le cylindre ayant pour base le cercle $C_{\lambda_0^*}$ du plan \mathfrak{z}^* de \mathcal{G}^* et pour génératrice l'axe $\mathbb{R}\tau^*$ des μ^* . $C_{\lambda_0^*}$ est l'orbite de l'action de $SO(2)$ et est paramétré par θ à partir de λ_0^* . C'est ce cercle paramétré par θ qui correspond au \mathbb{S}^1 du Hilbert $\mathcal{H} = L^2(\mathbb{S}^1, \mathbb{C})$ et c'est pourquoi il est naturel de traiter les $\psi(\theta) \in \mathcal{H}$ comme des fonctions de θ . D'où la nécessité de renommer l'angle de $SO(2)$ pour l'action de $SO(2)$ sur les $\psi(\theta)$. La procédure de construction des unirreps dans ce contexte est assez subtile et le lecteur en trouvera les détails par exemple dans la référence [547] déjà citée et Fresneda, Gazeau, Noguera [198]. Elle a été promue par De Bièvre [138]. À λ_0^* , avec $\lambda = |\lambda_0^*|$, on associe d'abord le caractère de $\mathbb{R}_{(q)}^2$ défini par $\chi(q) = e^{i\langle \lambda_0^*, q \rangle}$. Ensuite on fait naturellement agir $SO(2)$ par rotation de $C_{\lambda_0^*}$ paramétré par θ à partir de λ_0^* , d'où la formule

$$\rho_\lambda(x, y, \alpha)(\psi(\theta)) = e^{i\lambda(x \sin(\theta) + y \cos(\theta))} \psi(\theta + \alpha) .$$

À partir de là on peut essayer de construire des états cohérents avec une résolution de l'identité comme dans [198]. Mais cela est assez technique car les unirreps ρ_λ ne sont pas de carré intégrable. Pour construire des représentations de carré intégrable il faut passer par des quotients appropriés de G .

Comme nous l'avons vu, le cylindre $C_{\lambda_0^*} \times \mathbb{R}\tau^*$ (orbite co-adjointe de $(\lambda_0^*, 0)$) s'identifie à l'espace des phases $T^*C_{\lambda_0^*}$ de l'espace de configuration $C_{\lambda_0^*} \simeq \mathbb{S}_{(\theta)}^1$ et peut s'interpréter comme le quotient $G/H_0 \simeq \mathbb{R}_{(p)} \times \mathbb{S}_{(\beta)}^1$ où H_0 est le stabilisateur de $(\lambda_0^*, 0)$ dans \mathcal{G}^* (on a $\mathbb{R}_{(p)} = \mathbb{R}_{(q)}^2 / \mathbb{R}\lambda_0^*$). On indexe alors les états cohérents possibles par $G/H_0 \simeq \mathbb{R}_{(p)} \times \mathbb{S}_{(\beta)}^1$ muni de la mesure $d\beta dp$ associée à la 2-forme symplectique $dp \wedge d\beta$ et l'on cherche donc une ondelette mère $\varphi_0(\theta) \in \mathcal{H} = L^2(\mathbb{S}^1, \mathbb{C})$ admissible

à partir de laquelle engendrer un état cohérent $\varphi_{(p,\beta)}(\theta)$. Pour cela il faut d'abord remonter G/H_0 dans G par une *section* appropriée $\sigma : G/H_0 \rightarrow G$ de la projection $\pi : G \rightarrow G/H_0$ et définir ensuite $\varphi_{(p,\beta)}$ comme $\rho_\lambda(\sigma(p,\beta))\varphi_0(\theta)$. Ces sections vont par exemple remonter la droite $\mathbb{R}_{(p)} = \mathbb{R}_{(q)}^2/\mathbb{R}\lambda_0^*$ en une droite appropriée Δ de $\mathbb{R}_{(q)}^2$ munie de la coordonnée p et remonter $\mathbb{S}_{(\beta)}^1$ dans $SO(2) = \mathbb{S}_{(\alpha)}^1$ en déphasant $\pm\beta$, i.e. en prenant $\alpha = \pm\beta + \beta_0$.

Conformément à ce que nous avons vu, φ_0 est "admissible" si

$$\int_{G/H_0} |\langle \psi, \varphi_{(p,\beta)} \rangle_{\mathcal{H}}|^2 dp d\beta < \infty$$

pour tout $\psi \in \mathcal{H}$ ⁷² Pour cela, il suffit que

$$C_{\varphi_0} = \int_{G/H_0} |\langle \varphi_0, \varphi_{(p,\beta)} \rangle_{\mathcal{H}}|^2 dp d\beta < \infty .$$

On a alors la résolution de l'identité (en notations de Dirac)

$$\int_{G/H_0} |\varphi_{(p,\beta)}\rangle \langle \varphi_{(p,\beta)}| dp d\beta = C_{\varphi_0}$$

où les $|\varphi_{(p,\beta)}\rangle \langle \varphi_{(p,\beta)}|$ sont les projecteurs dans \mathcal{H} sur les $\varphi_{(p,\beta)}$. On obtient ainsi

$$\int_{G/H_0} \langle \psi, \varphi_{(p,\beta)} \rangle_{\mathcal{H}} \langle \varphi_{(p,\beta)}, \psi' \rangle_{\mathcal{H}} dp d\beta = C_{\varphi_0} \langle \psi, \psi' \rangle_{\mathcal{H}} .$$

Il faut trouver une bonne section σ . Un théorème de [198] explicite la procédure pour construire une droite Δ de coordonnée p dans le $\mathbb{R}_{(q)}^2$ de G de façon à ce que l'on ait $\sigma(p,\beta) = (R_\beta(\Delta), \beta)$ et que σ permette de construire un état cohérent.

Remarquons que dans un tel contexte on peut alors "quantifier" de façon naturelle, grâce à l'état cohérent, les fonction $f(p,\beta)$ sur l'espace des phases G/H_0 . Pour ce faire, on considère les projecteurs $|\varphi_{(p,\beta)}\rangle \langle \varphi_{(p,\beta)}|$ et on les intègre sur G/H_0 en les pondérant par $f(p,\beta)$ et en normalisant par la constante C_{φ_0} . On obtient ainsi un opérateur "quantique"

$$A_f = \frac{1}{C_{\varphi_0}} \int_{G/H_0} f(p,\beta) |\varphi_{(p,\beta)}\rangle \langle \varphi_{(p,\beta)}| dp d\beta .$$

Comme

$$\int_{G/H_0} |\varphi_{(p,\beta)}\rangle \langle \varphi_{(p,\beta)}| dp d\beta = C_{\varphi_0} ,$$

72. Rappelons encore une fois qu'il s'agit d'intégrales sur G/H_0 dont les intégrandes sont eux-mêmes des intégrales sur $\mathbb{S}_{(\theta)}^1 : \langle \psi, \varphi_{(p,\beta)} \rangle_{\mathcal{H}} = \int_{\mathbb{S}_{(\theta)}^1} \psi(\theta) \varphi_{(p,\beta)}(\theta)^* d\theta$.

on a bien $A_1 = Id_{\mathcal{H}}$. On peut alors étudier les propriétés de ces opérateurs et, éventuellement, leur limite semi-classique définie comme limite de

$$\tilde{f}(p, \beta) = \frac{1}{C_{\varphi_0}} \int_{G/H_0} f(p', \beta') \left| \langle \varphi_{(p', \beta')}, \varphi_{(p, \beta)} \rangle_{\mathcal{H}} \right|^2 dp' d\beta' .$$

Comme le notent les auteurs de [198] (nous gardons nos notations)

“It can be viewed as the average of the function $f(p, \beta)$ with respect to the probability distribution $(p', \beta') \mapsto \frac{1}{C_{\varphi_0}} \left| \langle \varphi_{(p', \beta')}, \varphi_{(p, \beta)} \rangle_{\mathcal{H}} \right|^2$ with respect to the measure $dp' d\beta'$.”

8.2. Les représentations de l'algèbre de Lie

L'action de la différentielle des unirreps ρ_λ sur l'algèbre de Lie des champs de vecteurs L -invariants X est

$$d\rho_\lambda : X \rightarrow d\rho_\lambda(X) := \left. \frac{d}{dt} \right|_{t=0} \rho_\lambda(e^{tX}) .$$

Pour la géométrie sous-riemannienne de $\mathbb{V}_S = SE(2)$ les plans de contact sont engendrés par X_1 et X_2 et l'on a (nous y reviendrons plus bas pour expliciter le noyau de la chaleur)

$$\begin{aligned} X_1(0) &= (1, 0, 0), \\ e^{tX_1} &= (t, 0, 0), \\ \rho_\lambda(e^{tX_1})\psi(\theta) &= e^{i\lambda t \sin(\theta)}\psi(\theta), \\ \widehat{X}_1^\lambda \psi(\theta) &= d\rho_\lambda(X_1)\psi(\theta) = \left. \frac{d}{dt} \right|_{t=0} e^{i\lambda t \sin(\theta)}\psi(\theta) = i\lambda \sin(\theta)\psi(\theta) \end{aligned}$$

et de même

$$\begin{aligned} X_2(0) &= (0, 0, 1), \\ e^{tX_2} &= (0, 0, t), \\ \rho_\lambda(e^{tX_2})\psi(\theta) &= \psi(\theta + t), \\ \widehat{X}_2^\lambda \psi(\theta) &= d\rho_\lambda(X_2)\psi(\theta) = \left. \frac{d}{dt} \right|_{t=0} \psi(\theta + t) = \frac{d\psi(\theta)}{d\theta} \end{aligned}$$

Ces plans de contact définissent une structure presque-complexe sur $\mathbb{V}_S = SE(2)$ au sens où l'on peut les identifier à \mathbb{C} en prenant X_1 et X_2 comme base. Évidemment,

\mathbb{V}_S étant de dimension 3 ne peut pas être une variété munie d'une structure complexe, mais en fait la structure de contact peut se prolonger en une structure *symplectique* qui, elle, est associée à une structure complexe dont la structure presque-complexe de \mathbb{V}_S est la restriction. Cela permet de définir une propriété de presque-holomorphicité, associée à la structure de contact, pour les fonctions $\Psi(v)$ définies sur \mathbb{V}_S . Elle signifie que $(X_2 - iX_1)\Psi(v) = 0$.

8.3. La transformée de Bargmann de $SE(2)$ (Barbieri-Citti)

Dans d'intéressants travaux [27], [28], Davide Barbieri, Giovanna Citti, Gonzalo Sanguinetti et Alessandro Sarti ont défini la transformée de Bargmann de $SE(2)$. Résumons brièvement leur démarche. Ils partent des unirreps $\rho_\lambda(v)$ ($v = (x, y, \alpha)$) de $SE(2)$ dans \mathcal{H} ,

$$\psi(\theta) \mapsto e^{i\lambda(x \sin(\theta) + y \cos(\theta))} \psi(\theta + \alpha)$$

et des

$$\begin{cases} \widehat{X}_1^\lambda \psi(\theta) = i\lambda \sin(\theta) \psi(\theta) \\ \widehat{X}_2^\lambda \psi(\theta) = \frac{d\psi(\theta)}{d\theta} \end{cases}$$

Ils utilisent ces opérateurs pour définir un état cohérent

$$\Phi_v(\theta) = \rho_\lambda(v)(\Phi_0(\theta))$$

sur $\mathcal{H} = L^2(\mathbb{S}_{(\theta)}^1, \mathbb{C})$ à partir de "l'ondelette mère" bien choisie (nous allons voir ce que cela signifie)

$$\Phi_0(\theta) = C e^{\lambda \cos(\theta)}$$

où C est un facteur de normalisation. $\Phi_0(\theta)$ satisfait trivialement

$$\left(\widehat{X}_2^\lambda - i\widehat{X}_1^\lambda\right) \Phi_0(\theta) = 0.$$

Ils définissent alors de façon naturelle la transformée de Bargmann de $\psi(\theta) \in \mathcal{H}$ comme la fonction $\Psi(v)$ (sur $\mathbb{V}_S = SE(2)$)

$$\Psi(v) = (B_\lambda \psi)(v) = \langle \Phi_v(\theta), \psi(\theta) \rangle_{\mathcal{H}}.$$

Tout le travail technique consiste ensuite à calculer l'image $B_\lambda(\mathcal{H})$ et à montrer que cette définition possède les bonnes propriétés que l'on doit attendre d'une transformée de Bargmann. Le résultat fondamental est le théorème suivant :

Théorème. B_λ est une isométrie de $\mathcal{H} = L^2(\mathbb{S}_{(\theta)}^1)$ sur l'espace $L^2(\mathbb{S}_{(\alpha)}^1, \mathfrak{H}_\lambda)$ des fonctions $\Psi(v) = \Psi(x, y, \alpha)$ telles que $\Psi_{(x,y)}(\alpha) = \Psi(x, y, \alpha) \in L^2(\mathbb{S}_{(\alpha)}^1)$ pour tout $(x, y) \in \mathbb{R}^2$ et $\Psi_\alpha(x, y) = \Psi(x, y, \alpha) \in \mathfrak{H}_\lambda$ pour tout $\alpha \in [0, 2\pi]$ où l'espace \mathfrak{H}_λ est défini ci-dessous. _____ \diamond

C'est la définition de \mathfrak{H}_λ qui est la clé du problème. Pour le définir, les auteurs considèrent des fonctions $f(x, y)$ dont la transformée de Fourier est *concentrée* sur le cercle C_λ de rayon λ dans l'espace de Fourier $\widehat{\mathbb{R}^2}$ et normalisée par $\frac{1}{\lambda}$. Soient $\kappa = (k \cos(\varphi), k \sin(\varphi))$ (notées aussi (k, φ) avec $k = |\kappa|$) des coordonnées polaires dans $\widehat{\mathbb{R}^2}$. Si $f(x, y)$ est quelconque, on peut restreindre $\widehat{f}(k, \varphi)$ à C_λ en imposant $k = \lambda$. En précisant bien les espaces fonctionnels dans lesquels on travaille, on peut introduire la distribution de Dirac $\delta(k - \lambda)$ et définir cette restriction normalisée comme la distribution

$$\widehat{f}^\lambda(k, \varphi) = \widehat{f}(k, \varphi) \frac{1}{\lambda} \delta(k - \lambda)$$

qui, en fait, ne dépend plus que de φ . On peut prendre ensuite la transformée de Fourier inverse de $\widehat{f}^\lambda(k, \varphi)$, ce qui donne une distribution $f_\lambda(x, y)$. On peut facilement calculer f_λ en utilisant le fait que la transformée de Fourier⁷³ de $\frac{1}{\lambda} \delta(k - \lambda)$ est $J_0(\lambda |(x, y)|)$ où J_0 est la fonction de Bessel⁷⁴

$$J_0(u) = \int_0^{2\pi} e^{iu \cos(\varphi)} d\varphi .$$

Comme la transformée de Fourier transforme les produits en convolution et introduit un facteur $\frac{1}{2\pi}$, on obtient

$$\begin{aligned} f_\lambda(x, y) &= \frac{1}{2\pi} (f(\bullet) * J_0(\lambda |\bullet|))(x, y) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x', y') J_0(\lambda |(x - x', y - y')|) dx' dy' . \end{aligned}$$

C'est à partir de ces \widehat{f}^λ et f_λ qui sont duaux par transformée de Fourier (i.e. $\widehat{f}^\lambda = \widehat{f}_\lambda$), que les auteurs construisent des espaces de Hilbert appropriés \mathfrak{H}^λ et \mathfrak{H}_λ . Ils partent des semi-normes $\langle \widehat{f}^\lambda, \widehat{f} \rangle$ et $\langle f_\lambda, f \rangle$ (qui ne sont que des semi-normes car $\langle \widehat{f}^\lambda, \widehat{f} \rangle = 0$ ou $\langle f_\lambda, f \rangle = 0$ n'implique pas $\widehat{f} = 0$ ou $f = 0$), quotientent les espaces fonctionnels utilisés par les éléments de semi-norme nulle et prennent les complétés. Ils obtiennent ainsi les espaces de Hilbert \mathfrak{H}^λ et \mathfrak{H}_λ qui sont duaux de Fourier et isométriques à cause de l'égalité de Parseval pour les transformées de Fourier $\langle f_\lambda, f \rangle = \langle \widehat{f}^\lambda, \widehat{f} \rangle$. Par construction \mathfrak{H}^λ est isomorphe à $L^2(\mathbb{S}^1)$.

Une fois ces espaces définis, les auteurs introduisent une condition de quasi-holomorphicité associée à la structure presque-complexe de $SE(2)$. Cette condition

73. La convention utilisée est celle d'un facteur $\frac{1}{2\pi}$ à la fois pour la transformée de Fourier et son inverse.

74. Nous avons rencontré deux fois la fonction de Bessel dans le Vol I : à la section 4.6.12 du chapitre 4 à propos des pinwheels et des champs gaussiens et à la section 5.11. du chapitre 5 à propos du modèle de Swindale pour les cartes de directions.

(CR) s'écrit

$$(X_2 - iX_1) \Psi(v) .$$

C'est l'analogie de la condition de Cauchy-Riemann définissant les fonctions holomorphes sur \mathbb{C} . Ils montrent alors

1. que, grâce au choix judicieux de $\Phi_0(\theta)$, la transformée

$$\Psi(v) = (B_\lambda \psi)(v) = \langle \Phi_v(\theta), \psi(\theta) \rangle_{\mathcal{H}}$$

satisfait CR,

2. que

$$(B_\lambda \psi)(v) = (B_\lambda \psi)(x, y, \alpha) \in \mathfrak{H}_\lambda$$

pour tout $\alpha \in [0, 2\pi]$,

3. que B_λ est une surjection de $\mathcal{H} = L^2(\mathbb{S}_\theta^1)$ sur $L^2(\mathbb{S}_\alpha^1, \mathfrak{H}_\lambda)$,

4. que B_λ est une isométrie,

5. que l'inverse de B_λ est donné par la formule

$$\psi(\theta) = \left[\int_{L^2(\mathbb{S}_\alpha^1, \mathfrak{H}_\lambda)} \Phi_v(\theta) ((B_\lambda \psi)(v)) dv \right]^\lambda .$$

La propriété (1) découle du fait que

$$\begin{aligned} (X_2 - iX_1)(B_\lambda \psi)(v) &= (X_2 - iX_1) \left(\int_{\mathbb{S}_\alpha^1} \rho_\lambda(v)(\Phi_0(\theta)) \overline{\psi(\theta)} d\theta \right) \\ &= \left(\int_{\mathbb{S}_\alpha^1} (X_2 - iX_1) \rho_\lambda(v)(\Phi_0(\theta)) \overline{\psi(\theta)} d\theta \right) \end{aligned}$$

et que $X(\rho_\lambda) = d\rho_\lambda(X) = \widehat{X}^\lambda$.⁷⁵ Comme on a par hypothèse $(\widehat{X}_2^\lambda - i\widehat{X}_1^\lambda) \Phi_0(\theta) = 0$ on a $(X_2 - iX_1)(B_\lambda \psi)(v) = 0$.

Les applications aux cartes d'orientations avec pinwheels traitées dans notre le chapitre 4 du Vol I, reposent alors sur le fait que, comme nous l'avons vu à la section 4.4.12. de ce volume, figure 4.33., les transformées de Fourier des fonctions d'autocorrélation de ces cartes sont concentrées sur des cercles. Les auteurs reconstruisent des cartes d'orientation. Ils superposent correctement des transformées de Bargmann déphasées de $\frac{\pi}{2}$ d'un bruit blanc sur \mathbb{S}^1 possédant la symétrie nécessaire pour que les orientations soient définies modulo π et que les rayons opposés d'un pinwheel correspondent à des orientations orthogonales.

75. Plus précisément, si $\widehat{X}^\lambda(\Phi_0(\theta)) = 0$, alors $X \langle \rho_\lambda(\Phi_0(\theta)), \psi(\theta) \rangle_{\mathcal{H}} = 0$ pour tout $\psi(\theta) \in \mathcal{H}$. En effet, $X \langle \rho_\lambda(\Phi_0(\theta)), \psi(\theta) \rangle_{\mathcal{H}} = \left\langle \frac{d}{dt} \Big|_{t=0} \rho_\lambda(e^{tX})(\Phi_0(\theta)), \psi(\theta) \right\rangle_{\mathcal{H}} = \left\langle \widehat{X}^\lambda(\Phi_0(\theta)), \psi(\theta) \right\rangle_{\mathcal{H}} = 0$.

9. Analyse harmonique et diffusion sous-riemannienne

Nous abordons maintenant en termes d'états cohérents de profils récepteurs les corrélats en analyse fonctionnelle de la géométrie sous-riemannienne des contours illusoires que nous avons longuement développée dans le chapitre 14. Cette étude est centrée sur les processus de *diffusion*.

Lorsqu'une métrique est définie sur un espace, la géométrie peut servir d'infrastructure pour ces processus dynamiques de transport le long de géodésiques que sont les processus de diffusion. Ces derniers sont traditionnellement décrits par l'équation de la chaleur et le noyau de la chaleur, et l'analyse harmonique (l'analyse de Fourier) en constitue un outil essentiel. Les études à ce sujet sont innombrables et ont joué un rôle absolument crucial dans les développements de l'analyse fonctionnelle.

Dans les prochaines sections, nous voudrions par conséquent présenter quelques éléments des processus de diffusion dans nos modèles \mathbb{V}_J et \mathbb{V}_S ainsi que certaines applications, en particulier en "inpainting". Comme il s'agit de processus de diffusion sous-riemannienne dans des structures de contact formalisant des architectures fonctionnelles, la neurogéométrie conduit tout naturellement à des problèmes sophistiqués d'analyse harmonique non commutative. Au début de nos recherches, les travaux de Richard Beals, Bernard Gaveau, Peter Greiner et Der-Chen Chang (cf. Beals *et al.* [35] et Chang, Greiner [105]) sur le laplacien sous-elliptique et le noyau de la chaleur associés à la structure de contact du groupe de Heisenberg nous ont beaucoup appris. Les calculs sont difficiles parce qu'il existe beaucoup de géodésiques entre deux points, avec des cut locus et des points conjugués. En géométrie riemannienne, on connaît des liens entre noyau de la chaleur et géodésiques (cf. par exemple Neel, Stroock [385] et Berceanu [46]). En géométrie sous-riemannienne la situation est nettement plus compliquée car, comme l'expliquait Bernard Gaveau dans un séminaire de l'Institut Henri Poincaré (26/10/2005) à propos de l'équation de la chaleur sur les variétés de Heisenberg, les singularités de la distance et de l'exponentielle font que la théorie classique n'est plus du tout valide.

9.1. La diffusion euclidienne

Faisons d'abord quelques brefs rappels sur le noyau de la chaleur dans le cas euclidien. On y trouve déjà toutes les idées et les résultats de base qu'il s'agit de généraliser.

9.1.1. Les solutions fondamentales.

Nous avons évoqué à la section 5.10.7 l'équation de la chaleur sur \mathbb{R} . Dans le cas classique de l'espace euclidien \mathbb{R}^3 de coordonnées $x = (x_1, x_2, x_3)$, l'équation de la chaleur⁷⁶

76. L'équation de la chaleur dépend d'un paramètre noté traditionnellement t car il s'agit en général d'un paramètre temporel. En vision et en analyse d'image, t est un paramètre

$$\frac{\partial f(x, \mathbf{r})}{\partial \mathbf{r}} = \Delta f(x, \mathbf{r})$$

se résout de la façon suivante.

La solution fondamentale (la fonction de Green) $G(x)$ de l'équation de Laplace-Poisson⁷⁷ $\Delta f(x) = \delta(x)$ (δ est la distribution de Dirac) est le potentiel newtonien

$$-\frac{1}{4\pi \|x\|}$$

(en dimension $n \geq 3$, il est proportionnel à $\frac{1}{\|x\|^{n-2}}$) et donc, l'équation étant linéaire, la solution de $\Delta f(x) = u(x)$ (c'est-à-dire, en notation symbolique, $f = \Delta^{-1}u$) est donnée par

$$f(x) = -\frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{u(y)}{\|x-y\|} dy .$$

Cela se voit très facilement au moyen de la transformée de Fourier

$$\widehat{\Delta G}(\xi) = -\|\xi\|^2 \widehat{G}(\xi) = \widehat{\delta}(\xi) = 1$$

($\|\xi\|^2$ étant, rappelons-le, le "symbole" de l'opérateur Δ) et donc

$$\widehat{G}(\xi) = -\frac{1}{\|\xi\|^2} . \quad 78$$

On en tire

$$G(x) = -\frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \frac{e^{i\xi \cdot x}}{\|\xi\|^2} d\xi$$

qui est une intégrale facile à calculer. Le point x étant donné choisissons un repère orthonormé tel que x soit sur le demi-axe $z > 0$ et prenons pour ξ les coordonnées sphériques de colatitude $\theta \in [0, \pi]$ et de longitude $\varphi \in [0, 2\pi]$. Si k est le module de ξ , la forme volume est $k^2 \sin(\theta) dk d\theta d\varphi$ et le produit scalaire $\xi \cdot x = k \|x\| \cos(\theta)$. L'intégrale est donc

$$G(x) = -\frac{1}{(2\pi)^3} \int_{\mathbb{R}^+ \times [0, \pi] \times [0, 2\pi]} \frac{e^{ik\|x\| \cos(\theta)}}{k^2} k^2 \sin(\theta) dk d\theta d\varphi .$$

d'échelle de résolution. Comme t et beaucoup d'autres notations utilisables sont déjà utilisées, nous optons pour la notation \mathbf{r} (r gothique).

77. L'équation de Laplace est l'équation $\Delta f(x) = 0$ dont les solutions sont les fonctions harmoniques, tandis que l'équation de Poisson est l'équation $\Delta f(x) = u(x)$ avec un membre de droite $\neq 0$ (en général une distribution) et des conditions aux limites. Les solutions en sont définies à une fonction harmonique près satisfaisant les conditions aux limites.

78. Le signe $-$ vient du fait que par transformée de Fourier (en dimension 1) la dérivée $\frac{\partial}{\partial x}$ devient $i\xi$ et donc $\frac{\partial^2}{\partial x^2}$ devient $-\xi^2$. D'où le $-\|\xi\|^2$ pour Δ en dimension quelconque. On préfère souvent l'opérateur $-\Delta$ car c'est lui qui intervient dans l'équation de la chaleur $\frac{\partial f(x, \mathbf{r})}{\partial \mathbf{r}} - \Delta f(x, \mathbf{r}) = 0$. La solution élémentaire change alors de signe.

L'intégration par rapport à $\varphi \in [0, 2\pi]$ introduit un facteur 2π et donc

$$G(x) = -\frac{1}{(2\pi)^2} \int_{\mathbb{R}^+} \int_{[0, \pi]} e^{ik\|x\|\cos(\theta)} \sin(\theta) dk d\theta .$$

Comme $\sin(\theta) d\theta = -d(\cos(\theta))$, l'intégrale $\int_0^\pi \cos(k\|x\|\cos(\theta)) \sin(\theta) d\theta$ donne par le changement de variable $k\|x\|\cos(\theta) = v$,

$$\int_{-k\|x\|}^{k\|x\|} \cos(v) \frac{1}{k\|x\|} dv = 2 \frac{\sin(k\|x\|)}{k\|x\|}$$

et

$$\int_0^\pi \sin(k\|x\|\cos(\theta)) \sin(\theta) d\theta = 0 ,$$

d'où la dernière intégrale

$$G(x) = -\frac{1}{2\pi^2} \int_{\mathbb{R}^+} \frac{\sin(k\|x\|)}{k\|x\|} dk .$$

Et comme il est bien connu que $\int_{\mathbb{R}^+} \frac{\sin(u)}{u} du = \frac{\pi}{2}$, on obtient bien en définitive

$$G(x) = -\frac{1}{4\pi\|x\|}$$

(ou $\frac{1}{4\pi\|x\|}$ pour $-\Delta$).

Remarque. Dans ce calcul, nous avons considéré $G(x)$ comme une fonction divergeant en 0. Mais en fait, il faut bien sûr la considérer comme une *distribution* (cf. la section 5 du chapitre 16) car elle est solution d'une équation dont le terme de droite est une distribution (la distribution de Dirac). \square

Remarque. En dimension 2 la situation est plus compliquée. La fonction de Green la plus naturelle du laplacien est

$$\frac{1}{2\pi} \log(|x|)$$

(où $|x| = \sqrt{x_1^2 + x_2^2}$ est le module), les autres s'en déduisant par addition d'une fonction harmonique. Cela peut se voir directement si l'on fait attention à quelques subtilités. En coordonnées polaires ($r = |x|, \theta$) le laplacien est

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} .$$

Une fonction de Green radiale $G(r)$ devra donc satisfaire

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial G}{\partial r} \right) = \delta(r) .$$

On constate que pour $r \neq 0$ c'est bien le cas de $G(r) = \frac{1}{2\pi} \log(r)$ avec les conditions aux limites $G(r) \xrightarrow[r \rightarrow \infty]{} \infty$ et $G(1) = 0$. En effet,

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \log(r)}{\partial r} \right) = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{1}{r} \right) = 0 .$$

Le coefficient $\frac{1}{2\pi}$ est nécessaire car si l'on applique le théorème de Stokes à $\Delta = \operatorname{div}(\operatorname{grad})$ sur un disque D de centre 0, on doit obtenir d'un côté $\int_D \delta = 1$ et de l'autre côté $\int_D \operatorname{div}(\operatorname{grad} G) = \int_{\partial D} \operatorname{grad} G = \frac{2\pi}{2\pi} = 1$. Mais pour $r = 0$ il faut faire attention et considérer $G(r) = \frac{1}{2\pi} \log(r)$ comme une distribution puisque $\delta(r)$ est une distribution. En fait on doit vérifier que, G étant considérée comme distribution, on a bien, pour toute fonction test f ,

$$\langle G, \Delta f \rangle = \langle \Delta G, f \rangle = \langle \delta, f \rangle = f(0) .$$

On notera également que $L(x) = \log(|x|)$ (considérée comme fonction paire sur \mathbb{R} divergente en 0, non intégrable mais localement intégrable en dehors de 0)⁷⁹ étant une distribution (de support singulier $\{0\}$), sa transformée de Fourier est également une distribution au sens expliqué à la section 2.1.1. On peut la calculer comme limite de fonctions en utilisant les transformées de Fourier des approximations

$$L_\varepsilon(x) = e^{-\varepsilon|x|} \log(|x|)$$

de $L(x)$ pour $\varepsilon \rightarrow 0^+$. On a

$$\widehat{L_\varepsilon}(x)(\omega) = \int_{\mathbb{R}} e^{-\varepsilon|x|} \log(|x|) e^{-i\omega x} dx .$$

Le calcul de l'intégrale donne une somme de trois termes $T_i(\omega)$

$$-\frac{2\varepsilon}{\varepsilon^2 + \omega^2} \gamma - \frac{\varepsilon}{\varepsilon^2 + \omega^2} \log(\varepsilon^2 + \omega^2) - \frac{2\omega}{\varepsilon^2 + \omega^2} \operatorname{ArcTan}\left(\frac{\omega}{\varepsilon}\right)$$

(où γ est la constante d'Euler⁸⁰). On calcule la limite distributionnelle des termes $T_i(\omega)$ en prenant des fonctions test $f(\omega)$ et en calculant $\int_{\mathbb{R}} T_i(\omega) f(\omega) d\omega$. De façon très intuitive (et pas très rigoureuse) la première limite donne, avec le changement de variable $\omega = \varepsilon\tau$ (qui est acceptable car f est une fonction test)

$$\begin{aligned} -2\gamma \int_{\mathbb{R}} \frac{\varepsilon}{\varepsilon^2 + \omega^2} f(\omega) d\omega &= -2\gamma \int_{\mathbb{R}} \frac{\varepsilon}{\varepsilon^2 + \varepsilon^2\tau^2} f(\varepsilon\tau) \varepsilon d\tau \\ &= -2\gamma \int_{\mathbb{R}} \frac{1}{1 + \tau^2} f(\varepsilon\tau) d\tau \xrightarrow[\varepsilon \rightarrow 0^+]{} -2\gamma\pi f(0) \end{aligned}$$

puisque $\int_{\mathbb{R}} \frac{1}{1+\tau^2} d\tau = \pi$. La seconde limite donne quant à elle

79. $\int_1^N \log(r) dr = 1 - N + N \log(N)$ et $\int_\varepsilon^1 \log(r) dr = -1 + \varepsilon - \varepsilon \log(\varepsilon)$.

80. $\gamma \sim 0.577216$ est la limite de la différence entre la somme $\sum_{n=1}^{n=N} \frac{1}{n}$ et $\log(N)$.

$$\begin{aligned}
 - \int_{\mathbb{R}} \frac{\varepsilon}{\varepsilon^2 + \omega^2} \log(\varepsilon^2 + \omega^2) f(\omega) d\omega &= - \int_{\mathbb{R}} \frac{\varepsilon}{\varepsilon^2 + \varepsilon^2 \tau^2} \log(\varepsilon^2 + \varepsilon^2 \tau^2) f(\varepsilon \tau) \varepsilon d\tau \\
 &= -2 \log(\varepsilon) \int_{\mathbb{R}} \frac{f(\varepsilon \tau)}{1 + \tau^2} d\tau \\
 &\quad - \int_{\mathbb{R}} \frac{\log(1 + \tau^2)}{1 + \tau^2} f(\varepsilon \tau) d\tau \\
 &\underset{\varepsilon \rightarrow 0^+}{\sim} -2\pi \log(\varepsilon) f(0) - 2\pi \log(2) f(0)
 \end{aligned}$$

puisque $\int_{\mathbb{R}} \frac{\log(1+\tau^2)}{1+\tau^2} d\tau = 2\pi \log(2)$. Enfin, pour la troisième limite, on commence par réécrire l'intégrale $\int_{\mathbb{R}} T_3(\omega) f(\omega) d\omega$ sous la forme $\int_{\mathbb{R}} = \int_{|\omega| \leq 1} + \int_{|\omega| \geq 1}$ avec

$$f(\omega) = f(0) + (f(\omega) - f(0))$$

dans $\int_{|\omega| \leq 1}$. D'où

$$\begin{aligned}
 - \int_{\mathbb{R}} \frac{2\omega}{\varepsilon^2 + \omega^2} \text{ArcTan}\left(\frac{\omega}{\varepsilon}\right) f(\omega) d\omega &= -f(0) \int_{|\omega| \leq 1} \frac{2\omega}{\varepsilon^2 + \omega^2} \text{ArcTan}\left(\frac{\omega}{\varepsilon}\right) d\omega \\
 &\quad - \int_{|\omega| \leq 1} \frac{2\omega}{\varepsilon^2 + \omega^2} \text{ArcTan}\left(\frac{\omega}{\varepsilon}\right) (f(\omega) - f(0)) d\omega \\
 &\quad - \int_{|\omega| \geq 1} \frac{2\omega}{\varepsilon^2 + \omega^2} \text{ArcTan}\left(\frac{\omega}{\varepsilon}\right) f(\omega) d\omega .
 \end{aligned}$$

La première intégrale $\int_{|\omega| \leq 1}$ peut se calculer explicitement. Le logiciel *Mathematica* donne

$$\frac{1}{6}i \left(\begin{array}{l} 12 \text{ArcCot}(\varepsilon) (\text{ArcCot}(\varepsilon) - 2i \log(1 + i\varepsilon) + 2i \log(\varepsilon) + 2i \log(2) - \pi) \\ + 12 \text{Li}_2\left(\frac{\varepsilon+i}{\varepsilon-i}\right) + \pi^2 \end{array} \right)$$

où Li_2 est le dilogarithme (Euler 1768).⁸¹ La limite pour $\varepsilon \rightarrow 0^+$ est immédiate puisque $\text{ArcCot}(0) = \frac{\pi}{2}$ et $\text{Li}_2(1) = \frac{\pi^2}{6}$. On trouve

$$\int_{|\omega| \leq 1} \underset{\varepsilon \rightarrow 0^+}{\sim} \frac{1}{6}i \left(12 \frac{\pi}{2} \left(\frac{\pi}{2} - 0 + 2i \log(\varepsilon) + 2i \log(2) - \pi \right) + 12 \frac{\pi^2}{6} + \pi^2 \right) .$$

La partie imaginaire

$$\frac{1}{6} \left(12 \frac{\pi}{2} \left(\frac{\pi}{2} - \pi \right) + 12 \frac{\pi^2}{6} + \pi^2 \right)$$

81. L'apparition de i dans la formule vient du fait que la formule exacte suppose l'intégration dans le plan complexe \mathbb{C} . Pour $z \in \mathbb{C}$, $|z| \leq 1$, $\text{Li}_2(z) = \int_1^{1-z} \frac{\log(t)}{1-t} dt$ (cf. le site *Analysis Situs* de Henri Paul de Saint Gervais [481]). On a $\left| \frac{\varepsilon+i}{\varepsilon-i} \right| = 1$ et $1 - \frac{\varepsilon+i}{\varepsilon-i} = -\frac{2i}{\varepsilon-i}$.

s'annule et la partie réelle

$$\frac{1}{6}i12\frac{\pi}{2}(2i\log(\varepsilon) + 2i\log(2)) = -\pi(2\log(\varepsilon) + 2\log(2)),$$

une fois multipliée par le facteur $-f(0)$ vient exactement compenser la deuxième limite.

Pour les deux autres intégrales de la troisième limite, on utilise des développements limités. Au voisinage de $s = \pm\infty$, $\text{ArcTan}(s) \sim \pm\frac{\pi}{2} - \frac{1}{s}$ au premier ordre et donc $\text{ArcTan}\left(\frac{\omega}{\varepsilon}\right) \underset{\varepsilon \rightarrow 0^+}{\sim} \pm\frac{\pi}{2} - \frac{\varepsilon}{\omega}$. Quand on passe à la limite on obtient, au facteur π près la distribution $V.P.\left(\frac{1}{|\omega|}\right)$ appelée "valeur principale" de $\frac{1}{|\omega|}$, d'où en définitive la transformée de Fourier de la distribution $L(x) = \log(|x|)$:

$$\widehat{L_\varepsilon(x)}(\omega) = -2\gamma\pi\delta(\omega) - \pi V.P.\left(\frac{1}{|\omega|}\right).$$

La valeur principale de la fonction impaire $\frac{1}{\omega}$ correspond à

$$\lim_{\varepsilon \rightarrow 0^+} \int_{-\infty}^{-\varepsilon} \frac{f(\omega)}{\omega} d\omega + \int_{\varepsilon}^{+\infty} \frac{f(\omega)}{\omega} d\omega$$

avec la même ε dans les intégrales (c'est de là que vient la notion de "valeur principale"). Le $(f(\omega) - f(0))$ qui intervient au lieu de $f(\omega)$ dans le calcul de la troisième limite correspond à la différence entre $V.P.\left(\frac{1}{\omega}\right)$ et $V.P.\left(\frac{1}{|\omega|}\right)$. □

Remarque. Pour la dimension 1, la fonction de Green $G(x)$ est évidente. En effet la distribution de Dirac δ est la dérivée de la fonction de Heaviside et donc (à une fonction linéaire $ax + b$ près dont la dérivée seconde = 0), $G(x) = x$ si $x \geq 0$ et $G(x) = -x$ si $x \leq 0$. $G(x) = |x|$ est une fonction continue linéaire par morceaux avec un point anguleux à l'origine. □

9.1.2. Diffusion et géodésiques.

Nous allons voir comment la diffusion est liée aux géodésiques euclidiennes de \mathbb{R}^3 (les droites). L'hamiltonien des géodésiques est (au signe près) le symbole du laplacien Δ , $H(x_j, \xi_j) = \|\xi\|^2 = \sum_{j=1}^{j=3} \xi_j^2$, les ξ_j étant les moments conjugués des x_j . Les équations de Hamilton définissant les bicaractéristiques dans $T^*\mathbb{R}^3$ (dont les projections sur \mathbb{R}^3 sont les géodésiques) sont par conséquent

$$\begin{cases} \dot{x}_j(s) = \frac{\partial H}{\partial \xi_j} = 2\xi_j \\ \dot{\xi}_j(s) = -\frac{\partial H}{\partial x_j} = 0 \end{cases}$$

Les moments ξ_j sont donc constants : $\xi_j = c_j$, et les $x_j(s)$ sont des fonctions linéaires $x_j(s) = 2c_j s + d_j$. Si les conditions aux limites sont $x(0) = 0$ et $x(\mathbf{r}) = \bar{x}$, alors les

d_j s'annulent et

$$\begin{cases} x_j(s) = \frac{\bar{x}_j}{\mathfrak{r}} s \\ \xi_j(s) = \frac{\bar{x}_j}{2\mathfrak{r}} \end{cases}$$

Le lagrangien associé à H par la transformée de Legendre $\sum_{j=1}^{j=3} \xi_j \dot{x}_j - H(x_j, \xi_j)$ est $L = \sum_{j=1}^{j=3} (\xi_j \dot{x}_j - \xi_j^2)$. Il est égal à

$$L = \sum_{j=1}^{j=3} \left(\frac{\bar{x}_j}{2\mathfrak{r}} \frac{\bar{x}_j}{\mathfrak{r}} - \left(\frac{\bar{x}_j}{2\mathfrak{r}} \right)^2 \right) = \frac{\|\bar{x}\|^2}{4\mathfrak{r}^2}$$

le long d'une telle géodésique et l'intégrale d'action est

$$S = \int_0^{\mathfrak{r}} L ds = \int_0^{\mathfrak{r}} \frac{\|\bar{x}\|^2}{4\mathfrak{r}^2} ds = \frac{\|\bar{x}\|^2}{4\mathfrak{r}} .$$

S est solution de l'équation de Hamilton-Jacobi

$$\frac{\partial S}{\partial \mathfrak{r}} + H(x, \nabla S) = 0 ,$$

c'est-à-dire $\frac{\partial S}{\partial \mathfrak{r}} = -\|\nabla S\|^2$ (les deux termes sont égaux à $-\frac{\|\bar{x}\|^2}{4\mathfrak{r}^2}$). Nous allons voir comment l'action se trouve liée au noyau de la chaleur.

9.1.3. Le noyau de la chaleur.

Venons-en donc à l'équation de la chaleur $\frac{\partial f}{\partial \mathfrak{r}}(x, \mathfrak{r}) = \Delta_x f(x, \mathfrak{r})$ avec condition initiale $f(x, 0) = u(x)$. La solution s'écrit sous forme symbolique $f(x, \mathfrak{r}) = e^{\mathfrak{r}\Delta_x} u(x)$ car symboliquement

$$\frac{\partial f(x, \mathfrak{r})}{\partial \mathfrak{r}} = \frac{\partial}{\partial \mathfrak{r}} e^{\mathfrak{r}\Delta_x} u(x) = \Delta_x e^{\mathfrak{r}\Delta_x} u(x) = \Delta_x f(x, \mathfrak{r}) .$$

Il s'agit d'expliciter l'opérateur différentiel $e^{\mathfrak{r}\Delta}$. La solution élémentaire de l'équation (i.e. de condition initiale $u(x) = P(x, 0) = \delta(x)$) est

$$P(x, \mathfrak{r}) = \frac{1}{(\sqrt{4\pi\mathfrak{r}})^3} e^{-\frac{\|x\|^2}{4\mathfrak{r}}}$$

(où l'on remarquera la gaussienne de largeur proportionnelle à \mathfrak{r} et le facteur $\mathfrak{r}^{-\frac{3}{2}}$, le 3 venant de la dimension) et donc la solution est

$$\begin{aligned} f(x, \mathfrak{r}) &= \frac{1}{(\sqrt{4\pi\mathfrak{r}})^3} \int_{\mathbb{R}^3} e^{-\frac{\|x-y\|^2}{4\mathfrak{r}}} u(y) dy = \int_{\mathbb{R}^3} \frac{1}{(4\pi\mathfrak{r})^{\frac{3}{2}}} e^{-\frac{\|x-y\|^2}{4\mathfrak{r}}} u(y) dy \\ &= \int_{\mathbb{R}^3} P_{\mathfrak{r}}(x, y) u(y) dy = \int_{\mathbb{R}^3} P_{\mathfrak{r}}(x - y) u(y) dy \\ &= (P_{\mathfrak{r}} * u)(x) , \end{aligned}$$

le noyau de la chaleur étant

$$P_{\mathfrak{r}}(x, y) = P_{\mathfrak{r}}(x - y) = \frac{1}{(4\pi\mathfrak{r})^{\frac{3}{2}}} e^{-\frac{\|x-y\|^2}{4\mathfrak{r}}}.$$

Il s'agit d'une convolution parce que le noyau $P_{\mathfrak{r}}(x, y)$ n'est fonction que de $(x - y)$.⁸² On note alors que l'exposant $\frac{\|x-y\|^2}{4\mathfrak{r}}$ intervenant dans l'exponentielle $e^{-\frac{\|x-y\|^2}{4\mathfrak{r}}}$ est l'action S . Le long de la géodésique (unique) partant de 0 et atteignant x au bout d'un temps \mathfrak{r} on a donc

$$P_{\mathfrak{r}}(x) = \frac{1}{(4\pi\mathfrak{r})^{\frac{3}{2}}} e^{-S(x, \mathfrak{r})}.$$

Le noyau de la chaleur et la solution élémentaire (fonction de Green) $G(x)$ du laplacien sont liés par la relation :

$$G(x) = \int_0^{\infty} P_{\mathfrak{r}}(x) d\mathfrak{r}.$$

En effet, comme l'intégrale $\int_0^{\infty} P_{\mathfrak{r}}(x) d\mathfrak{r}$ porte sur \mathfrak{r} et non pas sur x qui y joue le rôle d'un paramètre, on a

$$\begin{aligned} \Delta_x G(x) &= \Delta_x \int_0^{\infty} P_{\mathfrak{r}}(x) d\mathfrak{r} = \int_0^{\infty} \Delta_x P_{\mathfrak{r}}(x) d\mathfrak{r} = \int_0^{\infty} \frac{\partial P_{\mathfrak{r}}(x)}{\partial \mathfrak{r}} d\mathfrak{r} \\ &= P_{\infty}(x) - P_0(x) = 0 - \delta(x) = -\delta(x) \end{aligned}$$

et donc $G(x)$ est bien la fonction de Green de $-\Delta_x$. On vérifie d'ailleurs directement que

$$\int_0^{\infty} \frac{1}{(4\pi\mathfrak{r})^{\frac{3}{2}}} e^{-\frac{\|x-y\|^2}{4\mathfrak{r}}} d\mathfrak{r} = \frac{1}{(4\pi)^{\frac{3}{2}}} \frac{2\sqrt{\pi}}{\|x\|} = \frac{1}{4\pi\|x\|}.$$

Ces formules se généralisent en dimension $n > 3$ avec le facteur $\frac{1}{(4\pi\mathfrak{r})^{\frac{n}{2}}}$ au lieu de $\frac{1}{(4\pi\mathfrak{r})^{\frac{3}{2}}}$. Pour les dimensions inférieures il faut prendre quelques précautions.⁸³ En effet, si $\alpha > 1$,

$$\int_0^{\infty} \frac{1}{\mathfrak{r}^{\alpha}} e^{-\frac{\|x-y\|^2}{4\mathfrak{r}}} d\mathfrak{r} = 4^{\alpha-1} \frac{1}{(\|x-y\|^2)^{\alpha-1}} \Gamma(\alpha-1)$$

82. Le noyau de la chaleur $P_{\mathfrak{r}}(x, y)$ est une fonction des deux (multi)variables x et y . Ce que nous avons noté $P_{\mathfrak{r}}(x)$ est en fait $P_{\mathfrak{r}}(0, x)$.

83. Il faut tenir compte du fait qu'au voisinage de l'infini, le changement de variable $\mathfrak{r} = \frac{1}{u}$ transforme l'intégrande (à un facteur près) en

$$-u^{\frac{n}{2}} e^{-\frac{\|x-y\|^2 u}{4}} u^{-2} du = -u^{\frac{n}{2}-2} e^{-\frac{\|x-y\|^2 u}{4}} du.$$

où $\Gamma(z)$ est la fonction d'Euler $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ qui est divergente en 0. On voit que pour $\alpha \rightarrow 1^+$, c'est-à-dire $n \rightarrow 2^+$, l'intégrale susceptible de donner $G(x)$ diverge.

9.1.4. *Formule de la trace.*

Profitons de ces rappels pour signaler une formule importante. Sur un compact K de \mathbb{R}^3 avec des conditions au bord appropriées, le laplacien a un spectre *discret* de valeurs propres λ_n ($n \geq 1$), spectre associé à un système orthonormal de vecteurs propres $\psi_n(x)$ dans l'espace de Hilbert $L^2(K)$. Comme $e^{\tau\Delta}\psi_n = \lambda_n\psi_n$, $P_\tau(x, y) = \sum_{n \geq 1} e^{\tau\lambda_n}\psi_n(x)\bar{\psi}_n(y)$. On en tire immédiatement une formule pour la *trace* de l'opérateur $e^{\tau\Delta}$,

$$\text{tr } e^{\tau\Delta} = \sum_{n \geq 1} e^{\tau\lambda_n} .$$

On a en effet, en utilisant $P_\tau(x, x)$ et le fait que les ψ_n sont orthonormés,

$$\begin{aligned} \int_C P_\tau(x, x) dx &= \sum_{n \geq 1} e^{\tau\lambda_n} \int_C \psi_n(x)\bar{\psi}_n(x) dx = \sum_{n \geq 1} e^{\tau\lambda_n} \\ &= \text{tr } e^{\tau\Delta} . \end{aligned}$$

9.1.5. *Diffusion, mouvement brownien et calcul stochastique.*

Les processus de diffusion et leurs EDP comme l'équation de la chaleur peuvent être considérés comme des moyennages mésoscopiques de processus stochastiques microscopiques sous-jacents. Cela est pertinent pour la neurogéométrie car nous avons plusieurs fois expliqué que nos modèles géométriques (très proches des modèles lagrangiens et hamiltoniens de la physique classique) étaient *mésoscopiques* et possédaient une "physique statistique" sous-jacente, à savoir celle des potentiels d'action se propageant le long des connexions.⁸⁴

Nous avons déjà rencontré deux fois des processus stochastiques.

1. Dans la section 6.5. du chapitre 6 du Vol I et dans la section 3.4 du chapitre 8 de ce volume ci, nous avons rappelé comment David Mumford interprétait stochastiquement sa modélisation des contours illusoires par des *elasticæ*.⁸⁵ Il suppose que l'angle d'orientation θ est un *brownien* et donc que la courbure $\kappa = \dot{\theta}$ est un *bruit blanc*.⁸⁶ Une trajectoire microscopique réalisant un tel processus sera une marche aléatoire continue dans \mathbb{R}^2 dont les pas infinitésimaux

84. Cf. en particulier les sections 3 du chapitre 2, 3.1 du chapitre 8, 9 et en particulier 9.3 du chapitre 9.

85. Rappelons que ce modèle est différent de notre modèle géodésique sous-riemannien.

86. Lorsque le contour est régulier, la courbure est $\kappa = \dot{\theta}$, la dérivée étant prise par rapport à un temps t égal à la longueur d'arc s dans le plan euclidien (x, y) . Si θ est un brownien, il faut définir sa dérivée au sens stochastique (cf. plus bas).

seront

$$\begin{cases} dx = \cos(\theta) ds \\ dy = \sin(\theta) ds \end{cases}$$

Ces possibilités microscopiques sont inobservables. Ce qu'on observe réellement au niveau mésoscopique sont des moyennes. Si $P_s(a)$ est la probabilité, au niveau mésoscopique, que la marche aléatoire (partant par exemple de l'origine) arrive en a au temps s , $P_s(a)$ est solution d'un processus d'advection-diffusion régi par une équation de Fokker-Planck comprenant comme terme d'advection la dérivée $X_1(P)$ le long de

$$X_1 = \cos(\theta) \partial_x + \sin(\theta) \partial_y$$

et comme terme de diffusion $\frac{\sigma^2}{2} (X_2)^2(P)$ avec $X_2 = \partial_\theta$.

2. Dans les sections 4.6.10 – 12 du chapitre 4 du Vol I, nous avons par ailleurs évoqués les liens entre mésogéométrie et microphysique à propos des cartes d'orientations. Nous avons développé l'analogie de ces cartes et de leurs singularités (pinwheels) avec les champs de phases et leurs dislocations que l'on trouve en optique. Nous avons cité à ce propos les remarques de Michael Berry sur le fait que l'optique classique (qu'elle soit géométrique ou ondulatoire) est la mésogéométrie de la microphysique quantique des photons. Mathématiquement, on considère les cartes d'orientations comme des sections $Z(a)$ de la fibration $\mathbb{R}^2 \times \mathbb{S}^1 \rightarrow \mathbb{R}^2$, $(a, \theta) \rightarrow a$ qui sont des réalisations d'un champ aléatoire \mathcal{Z} de variables aléatoires \mathcal{Z}_a à valeurs dans \mathbb{P}^1 . Cela signifie que $Z(a)$ est la valeur de \mathcal{Z}_a pour une certaine réalisation du champ \mathcal{Z} . Nous avons alors expliqué certains calculs effectuables sous l'hypothèse que le champ \mathcal{Z} est *gaussien*, c'est-à-dire que les variables aléatoires \mathcal{Z}_a et toutes leurs combinaisons linéaires sont gaussiennes de moyenne m_a et de variance

$$\sigma_a^2 = \mathbb{E} [(\mathcal{Z}_a - m_a)^2]$$

(où \mathbb{E} est l'espérance de \mathcal{Z}_a). La géométrie des cartes $Z(a)$, et en particulier la statistique de leurs singularités, dépend essentiellement de la *fonction de corrélation*

$$C(a, b) = \mathbb{E} [(\mathcal{Z}_a - m_a)(\mathcal{Z}_b - m_b)] .$$

La stationnarité (invariance du champ par translation) signifie que

$$C(a, b) = C(a - b)$$

et l'isotropie signifie en plus que

$$C(a, b) = C(\|a - b\|) .$$

De profonds liens entre ces champs gaussiens et la géométrie différentielle à la Morse-Whitney-Thom sont développés dans le traité classique de Robert Adler et Jonathan Taylor [4].

Le lien entre marches aléatoires et diffusion est très facile à comprendre. L'exemple élémentaire suivant se trouve dans toutes les introductions à ce thème. Considérons sur \mathbb{R} une "particule" a partant de 0 au temps $t = 0$ et faisant des déplacements de longueur Δx en avant ou en arrière à chaque intervalle de temps Δt avec l'équiprobabilité $\frac{1}{2}$.⁸⁷ C'est un peu comme une spatialisation d'un jeu de pile ou face, le déplacement étant le gain. Il est facile de calculer la probabilité $p(m\Delta x, n\Delta t)$ de trouver la particule à la position $m\Delta x$ au moment $n\Delta t$. L'évolution temporelle est en effet évidente. Si a se trouve en $m\Delta x$ au temps $(n + 1)\Delta t$ c'est qu'elle était en $(m + 1)\Delta x$ ou $(m - 1)\Delta x$ au temps $n\Delta t$. Comme ces deux événements sont équiprobables,

$$p(m\Delta x, (n + 1)\Delta t) = \frac{1}{2}p((m + 1)\Delta x, n\Delta t) + \frac{1}{2}p((m - 1)\Delta x, n\Delta t) .$$

Traitons alors $p(m\Delta x, n\Delta t)$ comme la discrétisation d'une densité de probabilité $p(x, t)$ et discrétisons $\frac{\partial p(x, t)}{\partial t}$. On obtient en posant $\frac{(\Delta x)^2}{\Delta t} = D$

$$\begin{aligned} & \frac{p(m\Delta x, (n + 1)\Delta t) - p(m\Delta x, n\Delta t)}{\Delta t} \\ &= \frac{D(p((m + 1)\Delta x, n\Delta t) - 2p(m\Delta x, n\Delta t) + p((m - 1)\Delta x, n\Delta t))}{(\Delta x)^2} . \end{aligned}$$

Supposons alors que l'on fasse tendre Δx et Δt vers 0 en gardant le rapport $\frac{(\Delta x)^2}{\Delta t} = D$ constant, ce qui est une condition fondamentale reliant la mesure du temps à la mesure de l'espace. On obtient à la limite

$$\frac{\partial p(x, t)}{\partial t} = \frac{D}{2} \frac{\partial^2 p(x, t)}{\partial x^2} = \frac{D}{2} \Delta_x p(x, t)$$

c'est-à-dire l'équation de la chaleur pour le laplacien $\frac{D}{2}\Delta_x$. Comme par hypothèse la probabilité est 1 que a soit en 0 pour $t = 0$, le passage au continu donne $p(x, 0) = \delta(x)$ ("masse" de 1 en 0) et $p(x, t)$ est la fonction de Green de l'équation de la chaleur

$$p(x, t) = \frac{1}{\sqrt{4\pi\left(\frac{D}{2}\right)t}} e^{-\frac{x^2}{4\left(\frac{D}{2}\right)t}} .$$

Autrement dit la densité de probabilité $p(x, t)$ est une loi normale $\mathcal{N}(m, \sigma^2)$ (une gaussienne) de moyenne $m = 0$ et de variance $\sigma^2 = Dt$. La probabilité que la particule soit dans l'intervalle $[x_1, x_2]$ au temps t est

$$p(x(t) \in [x_1, x_2]) = \int_{x_1}^{x_2} p(x, t) dx .$$

87. Le symbole Δ est ici le symbole traditionnel d'incrément. Ce n'est évidemment pas le laplacien.

On vérifie que l'on a bien pour tout $t \geq 0$

$$p(x(t) \in \mathbb{R}) = \frac{1}{\sqrt{4\pi \left(\frac{D}{2}\right) t}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{4\left(\frac{D}{2}\right)t}} dx = \frac{1}{\sqrt{4\pi \left(\frac{D}{2}\right) t}} \frac{\sqrt{2\pi}}{\sqrt{\frac{1}{Dt}}} = 1$$

(la particule doit être quelque part!). Notons que la moyenne de $x^2(t)$ est

$$\frac{1}{\sqrt{4\pi \left(\frac{D}{2}\right) t}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{4\left(\frac{D}{2}\right)t}} dx = 2 \left(\frac{D}{2}\right) t = Dt .$$

Ces formules se généralisent facilement à la dimension $n > 1$, par exemple à des réseaux \mathbb{Z}^n où chaque point a pour voisins ce point déplacé de $\pm\Delta x$ parallèlement aux axes d'un repère orthonormé de \mathbb{R}^n . Le nombre de voisins d'un point n'est plus 2 mais $2n$ et la particule saute vers chaque voisin de façon équiprobable avec la probabilité $\frac{1}{2n}$. Avec la contrainte $\frac{(\Delta x)^2}{\Delta t} = D = 2n$ pour $\Delta x, \Delta t \rightarrow 0$, on obtient à la limite la généralisation :

$$\begin{cases} p(x, t) = \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{\|x\|^2}{4t}} \\ p(x, t|x_0, t_0) = \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{\|x-x_0\|^2}{4(t-t_0)}} \end{cases}$$

$p(x, t)$ étant la limite de la proportion des chemins aboutissant à x après une durée t en étant partis de $x = 0$ à $t = 0$.

On remarquera que cette limite est isotrope et invariante par translation alors que le réseau ne l'est pas. Elle reste la même si l'on change la géométrie du réseau. Derrière cette invariance se cache le théorème central limite.

On remarquera aussi que la condition que le rapport $\frac{(\Delta x)^2}{\Delta t}$ tende vers une limite bien définie pour $\Delta x, \Delta t \rightarrow 0$ signifie que, alors que la longueur de la courbe entre t_0 et t (mesurée avec l'unité Δt) est $t - t_0$, la distance parcourue $\|x - x_0\|$ est de l'ordre de $\sqrt{t - t_0}$, ce qui signifie que, à la limite, le mouvement brownien est une courbe *fractale* de dimension de Hausdorff 2.⁸⁸ D'où d'ailleurs une différence essentielle entre les dimensions $n \leq 4$ et les dimensions $n > 4$ car dans ce dernier cas des courbes browniennes qui sont comme des "surfaces" ne se coupent pas génériquement.

En termes de mouvement brownien, la fonction de Green $G(x)$ du laplacien correspond au *temps moyen* passé en un point qui est bien défini pour $n > 2$, le problème pour $n = 1, 2$ étant que ce temps moyen est alors infini.

88. Les physiciens parlent à propos de telles lois d'échelle $\|x - x_0\| \sim (t - t_0)^\nu$ d'*exposant caractéristique*. Ici $\nu = \frac{1}{2}$. Les formalismes du mouvement brownien sont des exemples simples des formalismes que l'on rencontre dans la physique statistique des phénomènes critiques et du groupe de renormalisation. Une remarquable introduction est le traité [267] de Claude Itzykson et Jean-Michel Drouffe. Le lecteur intéressé pourra aussi consulter la bibliographie de notre compilation [413] sur les phénomènes critiques.

Revenons au cas $n = 1$. Supposons pour simplifier que le coefficient de diffusion soit $D = 1$.⁸⁹ De façon plus théorique et plus formelle,⁹⁰ un mouvement brownien sur \mathbb{R} est un certain processus stochastique $W(\mathbf{r})$, appelé “processus de Wiener”, défini par un champ de variables aléatoires $\mathfrak{X}_{\mathbf{r}}$ à valeurs dans \mathbb{R} paramétrées par le “temps” $\mathbf{r} \geq 0$, une trajectoire étant une réalisation $X(\mathbf{r})$. On suppose des propriétés gaussiennes : $W(\mathbf{r})$ est un champ gaussien, ce qui implique que $W(\mathbf{r}) - W(\mathbf{s})$ soit une variable aléatoire gaussienne de moyenne nulle et de variance $\mathbf{r} - \mathbf{s}$ (i.e. qu’elle ait pour loi la loi normale $\mathcal{N}(0, \mathbf{r} - \mathbf{s})$). On suppose aussi des propriétés markoviennes (à savoir que le processus est “sans mémoire”) : si $\mathbf{r} > \mathbf{s}$, $W(\mathbf{r}) - W(\mathbf{s})$ est indépendante de $W(\mathbf{u})$ pour $\mathbf{u} < \mathbf{s}$. On dit que $W(\mathbf{r})$ est “normal” si $W(0) = 0$ au sens ou presque sûrement la position au temps initial 0 est l’origine 0 de \mathbb{R} . Dans ce cas, $W(\mathbf{r})$ est une gaussienne de moyenne nulle et de variance \mathbf{r} (comme $p(x, t)$ ci-dessus). On a aussi une propriété de “continuité” garantissant que les réalisations $x(t)$ sont presque sûrement continues.

La probabilité conditionnelle $\mathbb{P}(x, t|y)$ que la particule soit en x au temps t en étant partie de y est

$$\mathbb{P}(x, t|y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{|x-y|^2}{2t}} .$$

On obtient ainsi la fonction d’autocorrélation $C(\mathbf{s}, \mathbf{r})$ (qui est aussi l’autocovariance) du champ gaussien W pour $\mathbf{s} < \mathbf{r}$ en écrivant l’espérance du produit xy pour $W(\mathbf{s})$ suivi de $W(\mathbf{r} - \mathbf{s})$. Cela donne

$$\begin{aligned} C(\mathbf{s}, \mathbf{r}) &= \int_{\mathbb{R}} \int_{\mathbb{R}} xy \frac{1}{\sqrt{2\pi\mathbf{s}}} e^{-\frac{y^2}{2\mathbf{s}}} \frac{1}{\sqrt{2\pi(\mathbf{r}-\mathbf{s})}} e^{-\frac{(x-y)^2}{2(\mathbf{r}-\mathbf{s})}} dx dy \\ &= \mathbf{s} . \end{aligned}$$

D’après la propriété de continuité, le processus W peut être vu comme une variable aléatoire à valeurs dans l’espace $\mathcal{C}_0^0(\mathbb{R}_{(\mathbf{r})}^+, \mathbb{R}_{(x)})$ des fonctions continues $x(\mathbf{r})$ telles que $x(0) = 0$. Il permet de définir la mesure image $d\mu_W = W(d\mathbf{r})$ de $d\mathbf{r}$, mesure que l’on appelle la “mesure de Wiener” de $\mathcal{C}_0^0(\mathbb{R}_{(\mathbf{r})}^+, \mathbb{R}_{(x)})$. Ses différentes réalisations $X(\mathbf{r})$ peuvent approximer toutes les $x(\mathbf{r})$. La mesure μ_W mesure en quelque sorte la stochasticité des $x(\mathbf{r})$, les propriétés de régularité étant les plus “rares”. Si A est un sous-ensemble de $\mathcal{C}_0^0(\mathbb{R}_{(\mathbf{r})}^+, \mathbb{R}_{(x)})$,

$$\int_A d\mu_W = \mu_W(A) = \mathbb{P}(\{X(\mathbf{r})\}_{\mathbf{r} \geq 0} \in A) .$$

89. Souvent, on pose plutôt $D = 2$ de façon à avoir $\frac{D}{2} = 1$ et obtenir le laplacien Δ plutôt que le demi-laplacien $\frac{1}{2}\Delta$.

90. Le lecteur intéressé pourra se référer à l’ouvrage introductif [177] de Lawrence Evans.

On généralise facilement à la dimension $n > 1$ en considérant des processus stochastiques $W(\tau)$ à valeurs dans \mathbb{R}^n dont les projections sur les axes sont des browniens. La classe de ces processus est fermée pour les sous-processus, les isométries et les changements d'échelle.

Un tel processus définit un semi-groupe P_τ d'opérateurs sur les $f \in \mathcal{C}^0(\mathbb{R}^n)$ au moyen de la formule

$$P_\tau f(x) = \mathbb{E}[f(X(\tau)) | X(0) = x] .$$

Le générateur de $W(\tau)$ est alors la dérivée $\frac{dP_\tau}{d\tau}$ en $\tau = 0$. Dans notre exemple sur \mathbb{R} traité ci-dessus, nous voyons que c'est le laplacien $\frac{\Delta_x}{2}$ qui est le *générateur* du brownien et cela se généralise en dimension n . Le laplacien est le générateur des mouvements browniens et l'équation de la chaleur est l'équation exprimant comment un nuage de particules de densité initiale $u(x)$ diffuse parce que les particules suivent des trajectoires aléatoires d'incrémentés infinitésimaux.

9.1.6. Repères historiques du mouvement brownien.

Historiquement parlant, le mouvement brownien de particules dans un fluide a commencé à être observé au moyen de microscopes. C'est Robert Brown qui l'a imposé comme problème scientifique en 1827 en l'interprétant comme le choc des particules avec les molécules du fluide. Cette interprétation mécanique par chocs avec des molécules échappant absolument à toute observation possible a mis un certain temps à se faire jour car les premières observations portaient sur des micro-organismes dont on pouvait penser qu'ils étaient dotés d'un mouvement propre. On pouvait aussi penser que ces particules étaient entraînées par des turbulences du fluide ambiant.⁹¹

Un excellent historique mathématique est celui de Jean-Pierre Kahane [274]. Le mouvement brownien fut d'abord formalisé en 1901 par Louis Bachelier qui montra la formule $\langle \mathfrak{X}_\tau^2 \rangle = 2nD't$ qui est la formule $2\left(\frac{D}{2}\right)t$ trouvée ci-dessus en dimension 1 avec $n = 1$ et $D' = \frac{D}{2}$. Puis vint Einstein (1905-1906) avec la formule physique

$$\langle \Delta x \rangle^2 = \frac{RT}{N_A} \frac{1}{3\pi\mu a} \Delta t$$

où R est la constante des gaz parfaits, N_A le nombre d'Avogadro, μ la viscosité du fluide, a le rayon des particules. Puis vinrent Jean Perrin⁹² qui comprit en 1909 que les trajectoires browniennes sont partout sans dérivée, fractales comme on dit aujourd'hui, et Paul Langevin (1908, équations différentielles stochastiques permettant de retrouver la formule d'Einstein). Il faut aussi noter les travaux sur les probabilités

91. Un excellent exposé historique est celui d'Edward Nelson [386] de 1967. Nelson fut ensuite l'inventeur de l'*Internal Set Theory*, une version alternative de l'Analyse non standard d'Abraham Robinson.

92. Prix Nobel en 1926 pour sa mesure du nombre d'Avogadro au moyen du mouvement brownien.

d'Adriaan Fokker (1914) et Max Planck (1917) sur l'équation de Fokker-Planck⁹³, et de George Uhlenbeck et Leonard Orstein (1930) sur les processus de Gauss–Markov stationnaires. Mais c'est Norbert Wiener⁹⁴ (en 1923) et Paul Lévy⁹⁵ (également en 1923) qui établirent la théorie du brownien comme processus stochastique. Citons les travaux essentiels de Joseph Leo Doob [147] (l'un des fondateurs de la théorie des martingales) et Kiyoshi Itô [264] (le fondateur du calcul stochastique dit aussi calcul d'Itô).

9.1.7. *Passage aux cas riemanniens et sous-riemanniens.*

Nous venons de voir les liens étroits qui existent dans le cas euclidien entre laplacien, équation et noyau de la chaleur, diffusion et mouvement brownien. C'est cette théorie classique que l'on souhaiterait généraliser au cas du laplacien sous-elliptique en géométrie sous-riemannienne en explorant la façon dont se trouve préservée, pour les solutions de l'équation de la chaleur $\frac{\partial f}{\partial \tau}(x, \tau) = \Delta_x f(x, \tau)$, $f(x, 0) = u(x)$, une formule du genre

$$f(x, \tau) = \frac{1}{\alpha \tau^{\alpha(n)}} \int_M e^{-\frac{d(x,y)^2}{4\tau}} u(y) dy$$

(n = dimension de la variété M considérée), où $d(x, y)$ est la distance et où $\alpha(n)$ est une certaine fonction, .

9.2. La diffusion riemannienne

Pour généraliser la théorie euclidienne de la diffusion au cas riemannien, il faut tout d'abord définir rigoureusement la notion de laplacien. Pour ce faire, rappelons brièvement l'analyse vectorielle classique dans l'espace euclidien \mathbb{R}^3 exposée à la section 8.1 du chapitre 2.

93. C'est l'EDP régissant l'évolution des densités de probabilités dans des processus stochastiques. Des variantes sont les équations d'Andrey Kolmogorov et Marian Smoluchowski.

94. Scientifique hyper précoce et mathématicien prodige Norbert Wiener (1894-1964) obtint son PhD à 18 ans. Il fut le fondateur de la cybernétique et l'un des génies qui, comme von Neumann, furent à l'origine de la révolution qui conduira aux NBIC contemporaines (nanotechnologies, biotechnologies, technologies de l'information, sciences cognitives). Il fut un membre actif des célèbres Conférences Macy qui, de 1942 à 1950 regroupèrent (nous l'avons vu à la section 5.6 dans la note biographique sur von Neumann) une pléiade de savants aussi éminents que, outre lui-même et von Neumann, Arturo Rosenblueth, Warren McCulloch, Julian Bigelow, Gregory Bateson, Wolfgang Köhler, Heinz von Förster ou Claude Shannon. Cf. l'ouvrage de référence de Jean-Pierre Dupuy [160].

95. L'un des fondateurs de la théorie moderne des probabilités, Paul Lévy (1886-1971) fut major de l'École Normale Supérieure et vice-major de l'École Polytechnique en 1904. Après avoir soutenu sa thèse avec Jacques Hadamard, il fut professeur à l'École des Mines et à Polytechnique où son cours d'Analyse était très réputé. Il entra à l'Académie des Sciences en 1964. Il était le beau-père de Laurent Schwartz. Son traité de 1948 *Processus stochastiques et mouvement brownien* [328] est une référence fondamentale.

1. La métrique permet d'identifier les vecteurs et les covecteurs tangents. À une 1-forme ω on associe le champ de vecteurs tangents X_ω tel que pour tout autre champ Y on ait $\omega(Y) = \langle X_\omega, Y \rangle$. Duale, à tout champ de vecteurs X on associe une 1-forme ω_X par la même formule $\omega_X(Y) = \langle X, Y \rangle$. Évidemment $X_{\omega_X} = X$ et $\omega_{X_\omega} = \omega$. Si f est une fonction, df est une 1-forme et le champ X_{df} est le gradient

$$X_{df} = \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$$

de f . On a donc

$$df(Y) = \langle \nabla f, Y \rangle$$

pour tout champ Y .

2. On peut aussi associer naturellement à un champ X une 2-forme $\sigma = d\omega_X$. En dimension $n = 3$,

$$\begin{aligned} \sigma &= d\omega_X \\ &= (\partial_y X_z - \partial_z X_y) dy \wedge dz + (\partial_z X_x - \partial_x X_z) dz \wedge dx + (\partial_x X_y - \partial_y X_x) dx \wedge dy \end{aligned}$$

correspond au "pseudo-vecteur" rotationnel

$$\text{rot}(X) = \nabla \wedge X = (\partial_y X_z - \partial_z X_y, \partial_z X_x - \partial_x X_z, \partial_x X_y - \partial_y X_x)$$

(où \wedge est le produit vectoriel de vecteurs). De façon générale, on associe à tout champ $A = (A_x, A_y, A_z)$ la 2-forme

$$\sigma_A = A_x dy \wedge dz + A_y dz \wedge dx + A_z dx \wedge dy .$$

On a donc par définition $d\omega_X = \sigma_{\text{rot}(X)}$. Notons aussi l'égalité des 2-formes

$$\omega_X \wedge \omega_Y = \sigma_{X \wedge Y} .$$

3. On peut également considérer la 3-forme $d\sigma_A$ et, comme toutes les 3-formes sont des multiples de la forme *volume canonique* $\mu_v = dx \wedge dy \wedge dz$, on doit avoir $d\sigma_A = c_A \mu_v$ où c_A est un facteur de proportionnalité. Effectivement, on a

$$d\sigma_A = (\partial_x A_x + \partial_y A_y + \partial_z A_z) \mu_v .$$

Le facteur de proportionnalité est, par définition, la divergence de A . Si X et A sont deux champs, la 3-forme

$$\sigma_A \wedge \omega_X = \langle A, X \rangle \mu_v .$$

Pour généraliser ces constructions au cas riemannien (M, g) , il faut remarquer deux choses.

(i) La définition du gradient d'une fonction par la formule $df(Y) = \langle \nabla f, Y \rangle$ gardera un sens car on a le produit scalaire $g(X, Y)$.

(ii) La divergence d'un champ X est définissable au niveau de la seule structure différentiable dès que l'on se donne une forme volume μ_v . En effet, elle exprime la

façon dont la forme volume se déforme sous l'action de X . La dérivée de Lie $\mathcal{L}_X \mu_v$ restant une n -forme, elle est proportionnelle à μ_v et l'on a, par définition,

$$\mathcal{L}_X \mu_v = (\operatorname{div}_{\mu_v}(X)) \mu_v .$$

D'après la formule d'homotopie de Cartan (chapitre 2, section 7.2.12) on a

$$\mathcal{L}_X \mu_v = i_X d\mu_v + d(i_X \mu_v) = d(i_X \mu_v)$$

(puisque $d\mu_v$ étant une $(n+1)$ -forme est nécessairement identiquement nulle), autrement dit

$$(\operatorname{div}_{\mu_v}(X)) \mu_v = d(i_X \mu_v)$$

et donc $(\operatorname{div}_{\mu_v}(X)) \mu_v$ est une forme *exacte*. Cela implique d'ailleurs, d'après le théorème de Stokes, que l'intégrale de $\operatorname{div}_{\mu_v}(X)$ est toujours *nulle* si X est à support compact. En coordonnées locales, μ_v est toujours proportionnelle à $\bigwedge_{i=1}^{i=n} dx_i$,

$$\mu_v = V(x) \bigwedge_{i=1}^{i=n} dx_i$$

avec un facteur $V(x)$ partout $\neq 0$. Si X est de composantes X^i dans la base des $\{\partial_{x_i}\}$, on trouve

$$\operatorname{div}_{\mu_v}(X) = \frac{dV(X)}{V} + \sum_i \partial_{x_i} X^i .^{96}$$

Si $g_{\mu\nu}$ est le tenseur métrique et si $g^{\mu\nu}$ est son inverse, le gradient est donné par la formule $\sum_{\mu} (\nabla f)^{\mu} \partial_{\mu}$ avec les composantes données en coordonnées holonomes par

$$(\nabla f)^{\mu} = \sum_{\tau} g^{\mu\tau} \partial_{\tau} f .$$

En effet, on a bien

$$df(Y) = g(\nabla f, Y)$$

puisque

$$\begin{cases} df(Y) = \sum_{\nu} (\partial_{\nu} f) Y^{\nu} \\ g(\nabla f, Y) = \sum_{\mu, \nu} g_{\mu\nu} (\nabla f)^{\mu} Y^{\nu} \\ \sum_{\mu} g_{\mu\nu} g^{\mu\tau} = \delta_{\nu}^{\tau} \end{cases}$$

Donc

96. Le second terme de droite redonne la formule classique. La divergence se réduit à ce terme dans ce cas car $V \equiv 1$ et donc $dV = 0$.

$$\begin{aligned} g(\nabla f, Y) &= \sum_{\mu, \nu} g_{\mu\nu} (\nabla f)^\mu Y^\nu = \sum_{\mu, \nu} g_{\mu\nu} \left(\sum_{\tau} g^{\mu\tau} \partial_\tau f \right) Y^\nu \\ &= \sum_{\nu, \tau} \delta_\nu^\tau (\partial_\tau f) Y^\nu = \sum_{\nu} (\partial_\nu f) Y^\nu = df(Y) . \end{aligned}$$

Cette formule pour le gradient permet de dériver la formule du laplacien riemannien – dit “opérateur de Laplace-Beltrami”⁹⁷ – en utilisant la divergence par rapport à la forme volume canonique qui est donnée en coordonnées holonomes par

$$\mu_g = \sqrt{\det(g)} \bigwedge_{i=1}^{i=n} dx^i .$$

En appliquant la formule

$$\begin{aligned} \operatorname{div}_{\mu_g}(X) &= \frac{d(\sqrt{\det(g)})}{\sqrt{\det(g)}}(X) + \sum_{\mu} \partial_\mu X^\mu \\ &= \frac{1}{\sqrt{\det(g)}} \left(\sum_{\mu} \partial_\mu (\sqrt{\det(g)}) X^\mu \right) + \frac{1}{\sqrt{\det(g)}} \left(\sum_{\mu} \sqrt{\det(g)} \partial_\mu X^\mu \right) \\ &= \frac{1}{\sqrt{\det(g)}} \left(\sum_{\mu} \partial_\mu (\sqrt{\det(g)} X^\mu) \right) \end{aligned}$$

au champ ∇f de composantes $(\nabla f)^\mu = \sum_{\tau} g^{\mu\tau} \partial_\tau f$, on trouve

$$\begin{aligned} \Delta f &= \frac{1}{\sqrt{\det(g)}} \left(\sum_{\mu} \partial_\mu \left(\sqrt{\det(g)} \sum_{\tau} g^{\mu\tau} \partial_\tau f \right) \right) \\ &= \frac{1}{\sqrt{\det(g)}} \left(\sum_{\mu, \tau} \partial_\mu (\sqrt{\det(g)} g^{\mu\tau} \partial_\tau f) \right) . \end{aligned}$$

On notera que, d’après le théorème de Stokes, l’opérateur de Laplace-Beltrami est formellement auto-adjoint pour le produit scalaire global $\langle f, h \rangle = \int_M (fh) d\mu_g$ des fonctions C^∞ lorsque celles-ci sont à support compact sur M . Pour le voir il suffit d’appliquer la formule

$$\operatorname{div}_{\mu_g}(fX) = f \operatorname{div}_{\mu_g}(X) + df(X) = f \operatorname{div}_{\mu_g}(X) + g(\nabla f, X)$$

97. On peut en fait définir plusieurs laplaciens reliés entre eux par des formules explicites dues à Roland Weitzenböck que nous avons déjà rencontré à la section 6.5 du chapitre 6 à propos du parallélisme absolu dans les géométries de Cartan-Einstein et des connexions à courbure nulle mais à torsion non nulle.

à $X = \nabla h$ et d'intégrer sur M . Comme f est à support compact,

$$\int_M \operatorname{div}_{\mu_g} (f \nabla h) d\mu_g = 0$$

et donc

$$\begin{aligned} \int_M f \operatorname{div}_{\mu_g} (\nabla h) d\mu_g &= \int_M f \operatorname{div}_{\mu_g} (\nabla h) d\mu_g = \int_M f \Delta h d\mu_g = \langle f, \Delta h \rangle \\ &= - \int_M g (\nabla f, \nabla h) d\mu_g . \end{aligned}$$

Comme la dernière intégrale est symétrique en f et h puisque g est symétrique, on a bien l'auto-adjonction

$$\langle f, \Delta h \rangle = \langle \Delta f, h \rangle .$$

Sur la base de ces formules, la diffusion riemannienne est la généralisation directe de la diffusion euclidienne. La première déforme la seconde de façon bien contrôlable en utilisant les exponentielles $e^{-\frac{d(x,y)^2}{4\tau}}$ au lieu des exponentielles $e^{-\frac{\|x-y\|^2}{4\tau}}$. Pour l'analyse des liens entre le noyau de la chaleur et les géodésiques nous renvoyons entre autres à Neel et Stroock [385] et Berceanu [46] déjà cités. Par exemple, un théorème de Li et Yau [331] dit que pour $x, y \in M$ variété riemannienne connexe munie de la distance d et complète pour d (avec quelques conditions appropriées sur la courbure de Ricci), le noyau de la chaleur $P_\tau(x, y)$ exprimant la diffusion d'une masse de Dirac en x est encadré par des exponentielles du type ci-dessus avec des changements d'échelle de τ et un facteur lié au volume V de la boule riemannienne de rayon $\sqrt{\tau}$: il existe deux constantes $C, D > 0$ telles que

$$\frac{C}{V(B(x, \sqrt{\tau}))} e^{-\frac{d(x,y)^2}{4\tau}} \leq P_\tau(x, y) \leq \frac{D}{V(B(x, \sqrt{\tau}))} e^{-\frac{d(x,y)^2}{4\tau}} .$$

Sans changement d'échelle sur τ , i.e. avec uniquement des exponentielles $e^{-\frac{d(x,y)^2}{4\tau}}$, on a des encadrements un peu plus compliqués comme celui d'un théorème d'Adam Sikora dans [503] disant que si $\sup_{x \in M} (P_\tau(x, y)) \leq C\tau^{-\frac{d}{2}}$ si $\tau \leq 1$ et $\leq C\tau^{-\frac{D}{2}}$ si $\tau > 1$ avec $d \leq D$ alors il existe une constante $C_{d,D}$ telle que

$$P_\tau(x, y) \leq CC_{d,D} \min \left\{ \tau^{-\frac{d}{2}} \left(1 + \frac{d(x,y)}{\sqrt{\tau}} \right)^{d-1}, \tau^{-\frac{D}{2}} \left(1 + \frac{d(x,y)}{\sqrt{\tau}} \right)^{D-1} \right\} e^{-\frac{d(x,y)^2}{4\tau}}$$

avec le min remplacé par le max si $d > D$.

Mais c'est la généralisation sous-riemannienne de la diffusion, plus compliquée, qui nous intéresse au premier chef. Nous allons y consacrer le reste de ce chapitre.

9.3. Le volume intrinsèque en diffusion sous-riemannienne

Pour pouvoir étudier la généralisation des processus de diffusion au cas sous-riemannien, il faut d'abord pouvoir définir une forme volume "canonique" qui puisse être considérée comme adaptée à la distribution \mathfrak{D} sur laquelle est définie la métrique. Cela est nécessaire pour pouvoir définir le laplacien sous-riemannien Δ_{SR} à partir de la formule de base

$$\Delta_{SR}f = \operatorname{div}_\mu(\operatorname{grad}_{SR}(f))$$

puisque la notion de divergence dépend du volume : $\mathcal{L}_X\mu = (\operatorname{div}_\mu(X))\mu$ où μ est une forme volume sur la variété M considérée. Plusieurs travaux ont été consacrés à la recherche d'une bonne définition (Mitchell, Gromov). La plus couramment utilisée est celle de Popp (au début des années 2000).

La forme volume de Popp utilise la graduation

$$gr(\mathfrak{D}) = \bigoplus_{j=0}^{j=r} \mathfrak{D}^j / \mathfrak{D}^{j-1}$$

de la distribution \mathfrak{D} que nous avons définie à la section 1.3 du chapitre 15 sur la typologie des distributions.⁹⁸ On montre (c'est le résultat clé) que l'espace des n -formes volume $\bigwedge^n(T_xM)$ et $\bigwedge^n(gr_x(\mathfrak{D}))$ sur les deux espaces T_xM et $gr_x(\mathfrak{D})$ sont *canoniquement isomorphes*. Or il existe sur les quotients D_x^j/D_x^{j-1} des fibres une forme volume *canonique* μ_j induite par leur structure euclidienne. En effet, comme l'expliquent Andrei Agrachev, Davide Barilari et Ugo Boscain dans [5] on peut considérer les projections

$$\pi_j : \otimes^j D_x \rightarrow D_x^j / D_x^{j-1}$$

données par

$$\pi_j(X_1 \otimes \dots \otimes X_j) = [X_1 \dots [X_{j-1}, X_j]] \text{ modulo } D_x^{j-1}.$$

Elles sont surjectives et héritent du produit intérieur induit sur $\otimes^j D_x$ par celui donné sur \mathfrak{D} . D'où un produit intérieur "somme directe" sur le gradué et donc une forme volume. La forme volume de Popp en x est alors définie comme l'image de $\bigwedge_{j=1}^{j=r} \mu_j$ par l'isomorphisme canonique. Et ces formes volume locales *se recollent* correctement. Le lecteur intéressé trouvera de nombreuses précisions techniques dans [5].

Dans ce qui suit, nous allons présenter la diffusion sous-riemannienne dans nos modèles et nous n'aurons pas besoin de trop entrer dans ces subtilités.

98. Le premier terme de la somme directe est $\mathfrak{D}^0 = \mathfrak{D}$.

10. Diffusion sous riemannienne dans \mathbb{V}_J et \mathbb{V}_S

10.1. Le cas Heisenberg

Nous voudrions maintenant appliquer cette analyse des processus de diffusion (noyau de la chaleur) et d'advection-diffusion (Fokker-Planck) à l'architecture fonctionnelle de $V1$, implémentant la géométrie sous-riemannienne du groupe de Heisenberg (pour \mathbb{V}_J) et de $SE(2)$ (pour \mathbb{V}_S).

Nous allons commencer par le groupe de Heisenberg pour lequel Richard Beals, Bernard Gaveau, Peter Greiner, Der Chen Chang, ont commencé à construire le noyau à partir d'un papier fondateur de 1977. Comme le notait B. Gaveau dans un séminaire à l'IHP (26/10/2005), le problème est rendu difficile par l'existence de *singularités* au voisinage de *chaque* point.

10.1.1. Noyau de Heisenberg (Gaveau, Beals, Greiner).

Traitions d'abord le cas du groupe de Heisenberg \mathbb{H} non polarisé. Comme nous voulons garder nos notations standard (x, y, p) du chapitre 5 pour le \mathbb{H}_{pol} polarisé modélisant \mathbb{V}_J , nous utilisons dans cette section pour les points q de \mathbb{H} les coordonnées $q = (x_1, x_2, t) = (z, t)$, avec (x_1, x_2) les variables symplectiques conjuguées et $z = x_1 + ix_2$ la coordonnée complexe associée. Nous reprenons les notations de la section 2.2 du chapitre 14 où nous avons traité des géodésiques sous-riemanniennes de \mathbb{H} . Nous avons vu au chapitre 5 que le groupe de Heisenberg polarisé \mathbb{H}_{pol} s'identifie à \mathbb{H} non polarisé par le changement de variable $x = x_1$, $y = t + 2x_1x_2$, $p = 4x_2$ et si $q = (x_1, x_2, t) = (z, t)$, la loi de groupe de \mathbb{H} s'écrit :

$$q \cdot q' = (x_1, x_2, t) \cdot (x'_1, x'_2, t') = (x_1 + x'_1, x_2 + x'_2, t + t' + 2(x_2x'_1 - x_1x'_2)) ,$$

soit

$$(z, t) \cdot (z', t') = (z + z', t + t' + 2\Omega(z, z'))$$

où $\Omega(z, z')$ est la 2-forme symplectique standard sur le plan des z .⁹⁹ En effet, si l'on note \circ le produit dans \mathbb{H}_{pol} et si F est le changement de variable, il faut vérifier que $F(q \cdot q') = F(q) \circ F(q')$ pour $q = (x_1, x_2, t)$, $q' = (x'_1, x'_2, t') \in \mathbb{H}$. Mais

$$\begin{aligned} F(q \cdot q') &= (x_1 + x'_1, 4x_2 + 4x'_2, t + t' + 2(x_2x'_1 - x_1x'_2) + 2(x_1 + x'_1)(x_2 + x'_2)) \\ &= (x_1 + x'_1, 4(x_2 + x'_2), t + t' + 2x_1x_2 + 2x'_1x'_2 + 4x_2x'_1) \\ &= (x_1, 4x_2, t + 2x_1x_2) \circ (x'_1, 4x'_2, t' + 2x'_1x'_2) \\ &= F(q) \circ F(q') . \end{aligned}$$

99. Avec le changement de variables utilisé, le facteur de $(x_2x'_1 - x_1x'_2)$ est 2 au lieu du plus classique $\frac{1}{2}$. Nous avons parlé de cette variante à la section 5.1 du chapitre 5. Cela simplifie un peu les formules plus bas.

Dans l'espace de base \mathbb{C} , cette loi est simplement la translation $z + z'$, alors que dans la fibre \mathbb{R} elle est la translation augmentée du couplage

$$2(x_2x'_1 - x_1x'_2) = \langle \Lambda z, z' \rangle$$

où Λ est la multiplication par $-2i$. On choisit donc la fibration $(x_1, x_2, t) \rightarrow (x_1, x_2)$ pour laquelle la distribution de contact \mathcal{K} est une *connexion* dont les plans de contact K_q sont les plans "horizontaux" (cf. chapitre 6, section 2).

La structure de contact de \mathbb{H} translate le plan de contact à l'origine K_0 , c'est-à-dire le plan \mathbb{C} engendré par $X_1 = (1, 0, 0)$, $X_2 = (0, 1, 0)$. Les vecteurs tangents X_1 et X_2 engendrant les plans de contact K_q sont donc :

$$\begin{cases} X_1 = (1, 0, 2x_2) = \frac{\partial}{\partial x_1} + 2x_2 \frac{\partial}{\partial t} \\ X_2 = (0, 1, -2x_1) = \frac{\partial}{\partial x_2} - 2x_1 \frac{\partial}{\partial t} \end{cases}$$

Si $T = (0, 0, 1)$ est le vecteur unité "vertical", la structure d'algèbre de Lie nilpotente de $\mathfrak{h} = T_0\mathbb{H}$ est donnée par

$$[X_1, X_2] = -4T, \quad [X_1, T] = 0, \quad [X_2, T] = 0,$$

tous les commutateurs de niveau 2 étant donc nuls.¹⁰⁰ La forme de contact est

$$\omega = -2x_2 dx_1 + 2x_1 dx_2 + dt.$$

On a

$$d\omega = 4dx_1 \wedge dx_2 = 4\Omega$$

et

$$\omega \wedge d\omega = 4dt \wedge dx_1 \wedge dx_2$$

est une forme volume

Si $f : \mathbb{H} \rightarrow \mathbb{R}$ est une fonction différentiable, son gradient relativement à la connexion \mathcal{K} est donné par (cf. chapitre 5, section 2) :

$$\begin{aligned} \nabla_{\mathcal{K}}(f) &= X_1(f)X_1 + X_2(f)X_2 \\ &= \left(\frac{\partial f}{\partial x_1} + 2x_2 \frac{\partial f}{\partial t} \right) (1, 0, 2x_2) + \left(\frac{\partial f}{\partial x_2} - 2x_1 \frac{\partial f}{\partial t} \right) (0, 1, -2x_1) \\ &= \left(\frac{\partial f}{\partial x_1} + 2x_2 \frac{\partial f}{\partial t}, \frac{\partial f}{\partial x_2} - 2x_1 \frac{\partial f}{\partial t}, 2x_2 \frac{\partial f}{\partial x_1} - 2x_1 \frac{\partial f}{\partial x_2} + 4(x_1^2 + x_2^2) \frac{\partial f}{\partial t} \right). \end{aligned}$$

Par ailleurs, si $X = u_1X_1 + u_2X_2$ est un champ de vecteurs tangents à \mathcal{K} , sa divergence relativement à \mathcal{K} est donnée par :

$$\operatorname{div}_{\mathcal{K}} X = X_1(u_1) + X_2(u_2).$$

100. Rappelons que nous notons \mathfrak{h} et non pas \mathcal{H} l'algèbre de Lie de \mathbb{H} pour éviter la confusion avec les Hilbert \mathcal{H} .

Si $X = \nabla_{\mathcal{K}}(f)$ alors

$$\begin{aligned} \operatorname{div}_{\mathcal{K}}(\nabla_{\mathcal{K}}(f)) &= X_1(X_1(f)) + X_2(X_2(f)) \\ &= \Delta_{\mathcal{K}} = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + 4x_2 \frac{\partial^2}{\partial x_1 \partial t} - 4x_1 \frac{\partial^2}{\partial x_2 \partial t} + 4(x_1^2 + x_2^2) \frac{\partial^2}{\partial t^2} \\ &= \Delta_{\mathcal{K}}(f) \text{ (laplacien de Horn) .} \end{aligned}$$

Une fois ces rappels effectués, revenons à la diffusion dans \mathbb{H} . Notons d’abord que, d’après un théorème de Folland de 1973, la solution fondamentale de ce laplacien sous-riemannien $\Delta_{\mathcal{K}}$ de \mathbb{H} est

$$C \frac{1}{\sqrt{|z|^4 + 16t^2}} ,$$

formule que l’on comparera au cas euclidien dans \mathbb{R}^3 qui est, comme nous l’avons vu plus haut, $C \frac{1}{\|x\|}$. Le fait que l’on obtienne $|z|^4 + 16t^2$ à la puissance $-\frac{1}{2}$ est subtil. Pour \mathbb{H}^n on obtient $\|z\|^4 + 16t^2$ à la puissance $\frac{1}{4}(2 - 2(n + 1)) = -\frac{n}{2}$, $2(n + 1)$ étant la dimension homogène de \mathbb{H}^n . Le terme $|z|^4 + 16t^2$ s’explique à partir de l’équation de la chaleur.

La diffusion dans \mathbb{H} est infiniment plus complexe que dans le cas riemannien car il existe, répétons-le encore une fois, des singularités de la distance, des points de cut locus dans le voisinage de tout point et des points conjugués. L’exponentielle est mal définie et il existe plusieurs géodésiques entre deux points quelconques qui, toutes, contribuent au noyau de la chaleur.

Les travaux fondamentaux inauguraux sont ceux de Bernard Gaveau [206], [207] et de Der Chen Chang et Peter Greiner [105] que nous avons déjà longuement évoqués dans les sections 2.1 et 2.2 du chapitre 14 à propos des géodésiques. Dès 1977, Bernard Gaveau a défriché le problème. Nous avons vu que la diffusion classique résulte d’un mouvement brownien. Gaveau a d’abord montré que la diffusion associée à l’équation de la chaleur sur \mathbb{H} peut s’écrire comme un processus stochastique de type mouvement brownien *sous-riemannien* $(\mathfrak{Z}(\mathbf{r}), \mathfrak{I}(\mathbf{r}))$ où $\mathfrak{Z}(\mathbf{r})$ est un mouvement brownien à valeurs complexes $\mathfrak{Z}(\mathbf{r}) = \mathfrak{X}_1(\mathbf{r}) + i\mathfrak{X}_2(\mathbf{r})$ (avec $\mathfrak{X}_i(\mathbf{r}) = \int_0^{\mathbf{r}} d\mathfrak{X}_i$) et où $\mathfrak{I}(\mathbf{r}) = 2 \int_0^{\mathbf{r}} (\mathfrak{X}_2 d\mathfrak{X}_1 - \mathfrak{X}_1 d\mathfrak{X}_2)$. En effet l’hamiltonien

$$H(x_1, x_2, t, \xi_1^*, \xi_2^*, \theta^*) = \frac{1}{2} [(\xi_1^* + 2x_2\theta^*)^2 + (\xi_2^* - 2x_1\theta^*)^2]$$

s’écrit

$$\frac{1}{2} (\xi_1^*, \xi_2^*, \theta^*) g \begin{pmatrix} \xi_1^* \\ \xi_2^* \\ \theta^* \end{pmatrix} ,$$

g étant la matrice

$$g = \begin{pmatrix} 1 & 0 & 2x_2 \\ 0 & 1 & -2x_1 \\ 2x_2 & -2x_1 & 4|z|^2 \end{pmatrix}.$$

La projection de $\Delta_{\mathcal{K}}$ sur la base $\mathbb{C}_{(z)}$ de coordonnée complexe $z = x_1 + ix_2$ est la diffusion standard. C'est donc un mouvement brownien $\mathfrak{Z}(\mathbf{r})$. Pour relever ce dernier dans \mathbb{H} il faut donc calculer dt . En effet, comme l'explique B. Gaveau ([206], p. 99)

“la matrice σ donnant en chaque point les covariances des accroissements infinitésimaux de la diffusion”

est la matrice

$$\sigma = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2x_2 \\ 0 & 1 & -2x_1 \end{pmatrix}$$

et

$$g = \sigma^T \sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2x_2 & -2x_1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2x_2 \\ 0 & 1 & -2x_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2x_2 \\ 0 & 1 & -2x_1 \\ 2x_2 & -2x_1 & 4(x_1^2 + x_2^2) \end{pmatrix}.$$

Les accroissements de la diffusion sont alors donnés par

$$(dx_1, dx_2, dt) = (dx_1, dx_2)^T \sigma = (dx_1, dx_2, dt = 2(x_2 dx_1 - x_1 dx_2))$$

et la diffusion est donc le mouvement brownien $\mathfrak{Z}(\mathbf{r}) = \mathfrak{X}_1(\mathbf{r}) + i\mathfrak{X}_2(\mathbf{r})$ auquel s'ajoute $\mathfrak{T}(\mathbf{r}) = 2 \int_0^{\mathbf{r}} (\mathfrak{X}_2 d\mathfrak{X}_1 - \mathfrak{X}_1 d\mathfrak{X}_2)$.

B. Gaveau a montré ensuite le théorème suivant pour le laplacien hypo-elliptique

$$\Delta_{\mathcal{K}} = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + 4x_2 \frac{\partial^2}{\partial x_1 \partial t} - 4x_1 \frac{\partial^2}{\partial x_2 \partial t} + 4(x_1^2 + x_2^2) \frac{\partial^2}{\partial t^2}.$$

Théorème (Gaveau, 1977). Soit $P(z, t, \mathbf{r})$ la solution fondamentale de l'équation de la chaleur sous-elliptique $\frac{\partial f(z, t, \mathbf{r})}{\partial \mathbf{r}} = \frac{1}{2} \Delta_{\mathcal{K}} f(z, t, \mathbf{r})$ calculée au point (z, t) au temps \mathbf{r} .¹⁰¹ On a

101. Rappelons que la variable temporelle est ici notée \mathbf{r} parce que la variable t est déjà utilisée comme coordonnée de \mathbb{H} . Comme beaucoup d'auteurs, Bernard Gaveau utilise $\frac{1}{2} \Delta_{\mathcal{K}}$ et non pas $\Delta_{\mathcal{K}}$ pour l'équation de la chaleur. Cela change un petit peu les formules par rapport à $\Delta_{\mathcal{K}}$.

$$\begin{aligned}
 P(z, t, \mathfrak{r}) &= P_{\mathfrak{r}}(z, t) = \frac{1}{(2\pi\mathfrak{r})^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} e^{\left(\frac{i\tau t}{\mathfrak{r}} - \left(\frac{|z|^2}{2\mathfrak{r}}\right) \frac{2\tau}{\tanh(2\tau)}\right)} d\tau \\
 &= \frac{1}{(2\pi\mathfrak{r})^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} e^{-\left(\frac{|z|^2}{2\mathfrak{r}}\right) \frac{2\tau}{\tanh(2\tau)}} \cos\left(\frac{\tau t}{\mathfrak{r}}\right) d\tau \\
 &\hspace{15em} (\text{à cause de la parité des fonctions}) \\
 &= \frac{1}{(2\pi\mathfrak{r})^2} \int_{\mathbb{R}} V(\tau) e^{(-\frac{\Sigma(z,t,\tau)}{\mathfrak{r}})} d\tau,
 \end{aligned}$$

formule où $V(\tau) = \frac{2\tau}{\sinh(2\tau)}$ est un volume sur la variété bicaractéristique de $\Delta_{\mathcal{K}}$ et $\Sigma(z, t, \tau) = -i\tau t + |z|^2 \frac{\tau}{\tanh(2\tau)}$ est une action *complexe*. \diamond

Remarque. Nous allons voir dans la prochaine section comment, à la même époque que Bernard Gaveau, Andrzej Hulanicki avait trouvé une formule analogue pour le groupe d’Heisenberg *polarisé*. \square

La connaissance de $P_{\mathfrak{r}}(z, t)$ permet de calculer facilement la fonction de Green $G_{\mathbb{H}}(z, t)$ du laplacien hypo-elliptique de \mathbb{H} . On a en effet, en intervertissant l’ordre des deux intégrales,

$$\begin{aligned}
 G_{\mathbb{H}}(z, t) &= \int_0^\infty P_{\mathfrak{r}}(z, t) d\mathfrak{r} \\
 &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \left(\int_0^\infty \frac{1}{\mathfrak{r}^2} e^{-\left(\frac{|z|^2}{2\mathfrak{r}}\right) \frac{2\tau}{\tanh(2\tau)}} \cos\left(\frac{\tau t}{\mathfrak{r}}\right) d\mathfrak{r} \right) d\tau.
 \end{aligned}$$

Mais

$$\int_0^\infty \frac{1}{\mathfrak{r}^2} e^{\frac{a}{\mathfrak{r}}} \cos\left(\frac{b}{\mathfrak{r}}\right) d\mathfrak{r} = -\frac{a}{a^2 + b^2}$$

et il faut donc calculer, puisque $a = -\frac{|z|^2 2\tau}{2 \tanh(2\tau)}$ et $b = \tau t$,

$$\frac{1}{(2\pi)^2} \int_{\mathbb{R}} \frac{2|z|^2 \cosh(2\tau)}{|z|^4 \cosh^2(2\tau) + t^2 \sinh^2(2\tau)} d\tau = \frac{1}{(2\pi)^2} \frac{1}{t} \int_{\mathbb{R}} F(u) du$$

avec $u = 2\tau$, $A = \frac{|z|^2}{t}$ et

$$F(u) = \frac{A \cosh(u)}{A^2 \cosh^2(u) + \sinh^2(u)}.$$

Ce genre d’intégrale se calcule traditionnellement en utilisant la puissante méthode d’analyse complexe dite des *résidus*.¹⁰² On prolonge $F(u)$ au plan complexe en une fonction *méromorphe* qui est *holomorphe* en dehors de pôles isolés u_k pour lesquels

102. Historiquement, la théorie s’est édifiée entre Cauchy et Riemann.

le dénominateur $|z|^4 \cosh^2(2\tau) + t^2 \sinh^2(2\tau)$ s'annule et l'on utilise les propriétés d'intégration remarquables de ces fonctions le long de circuits dans le plan complexe des u . Ces pôles sont faciles à calculer, ce sont les solutions de $\tanh^2(u) = -\frac{|z|^4}{t^2} = -A^2$, autrement dit de

$$\tanh(u) = \pm i \frac{|z|^2}{t} = \pm iA.$$

Comme $\tanh(u) = \frac{e^{2u}-1}{e^{2u}+1}$, les solutions sont

$$e^{2u} = \frac{1 \pm i \frac{|z|^2}{t}}{1 \mp i \frac{|z|^2}{t}} = \frac{t \pm i |z|^2}{t \mp i |z|^2} = \frac{1 \pm iA}{1 \mp iA}.$$

Elles sont définies modulo $i\pi$ et si $t + i|z|^2 = Me^{i\Phi}$, $\Phi \in [0, \pi]$, en coordonnées polaires, on a $\frac{t \pm i|z|^2}{t \mp i|z|^2} = e^{\pm 2i\Phi}$ et par conséquent des pôles

$$u_k^\pm = \pm i\Phi + ik\pi (k \in \mathbb{Z})$$

définis modulo $i\pi$. Au voisinage d'un pôle u_k^\pm , le résidu de $F(u)$ est le coefficient de $\frac{1}{u}$ dans le développement de Laurent de $F(u)$. On montre (théorème de Cauchy) qu'il est donné par $\frac{1}{2\pi i} \oint F(u) du$ où \oint est prise le long d'un petit circuit entourant une fois le pôle (et seulement lui) dans le sens direct. Cela se déduit immédiatement de ce qui se passe pour les u^n au voisinage de 0. Si $u = \rho e^{i\theta}$ en coordonnées polaires et si l'on intègre le long d'un petit cercle de rayon ρ_0 ,

$$\oint u^n du = (\rho_0)^n \int_0^{2\pi} e^{in\theta} \rho_0 (ie^{i\theta}) d\theta = (\rho_0)^{n+1} \int_0^{2\pi} ie^{i(n+1)\theta} d\theta.$$

Cette intégrale est trivialement nulle si $(n+1) \neq 0$, i.e. si $n \neq -1$. En revanche, pour $n = -1$ elle vaut $2\pi i$ et donc, par définition, le résidu de $\frac{1}{u}$ au pôle 0 est 1. On voit que lorsqu'on intègre le long des circuits, seuls les résidus aux pôles vont contribuer à l'intégration.

Ici les résidus aux pôles u_k^+ sont donnés par la formule

$$\begin{aligned} \frac{A}{2(1+A^2)} \frac{1}{\sinh(u_k^+)} &= -i \frac{A}{2(1+A^2)} \frac{\sqrt{1+A^2}}{A} \\ &= -i \frac{1}{2\sqrt{1+A^2}} \end{aligned}$$

car

$$\tanh^2(u_k^\pm) = -A^2 = \frac{\sinh^2(u_k^\pm)}{1 + \sinh^2(u_k^\pm)}$$

implique

$$\sinh^2(u_k^\pm) = -\frac{A^2}{1+A^2}.$$

Les résidus aux pôles $u_k^- = -u_k^+$ sont donc

$$i \frac{1}{2\sqrt{1+A^2}} .$$

Le théorème des résidus dit que, dans une intégrale le long d'un circuit Γ , chaque pôle intervient dans l'intégrale par $2\pi i$ fois le résidu du pôle, avec en facteur l'index de Γ , c'est-à-dire le nombre algébrique de tours que fait le circuit autour du pôle. Prenons alors pour circuit Γ l'intervalle $[-R, R]$ suivi du demi-cercle supérieur de rayon R et faisons tendre R vers ∞ . Comme le module $|F(u)|$ de $F(u)$ décroît rapidement vers 0 lorsque $R \rightarrow \infty$, l'intégrale le long du demi-cercle tend vers 0 et donc notre intégrale est la somme des résidus des pôles inclus à l'intérieur de Γ . Mais comme les u_k^\pm sont définis modulo $i\pi$ et que $k \in \mathbb{Z}$, on voit que tous les résidus s'annulent par paires sauf celui de u_0^+ . On obtient donc la valeur

$$\int_{\mathbb{R}} F(u) du = (2\pi i) i \frac{1}{2\sqrt{1+A^2}} = \frac{\pi}{\sqrt{1+A^2}} = \frac{\pi t}{\sqrt{|z|^4+t^2}}$$

et, en multipliant par le facteur $\frac{1}{(2\pi)^2 t}$, on obtient la fonction de Green

$$G_{\mathbb{H}}(z, t) = \frac{1}{4\pi\sqrt{|z|^4+t^2}}$$

du groupe de Heisenberg \mathbb{H} .

On comparera cette formule à celle du cas euclidien qui est en $\frac{1}{4\pi\|x\|}$. Le fait que l'on ait $\frac{1}{\sqrt{(|z|^2)^2+t^2}}$ au lieu de $\frac{1}{\sqrt{|z|^2+t^2}} = \frac{1}{\|x\|}$ manifeste de façon simple et frappante l'écart entre la géométrie euclidienne isotrope de \mathbb{R}^3 et la géométrie sous-riemannienne anisotrope de \mathbb{H} où la base et la fibre ont des statuts différents.

À partir de $P_{\mathfrak{r}}(z, t) = P_{\mathfrak{r}}(q)$, le noyau de la chaleur est défini

$$K_{\mathfrak{r}}(q, q') = P_{\mathfrak{r}}(q \cdot q'^{-1}) = P_{\mathfrak{r}}(q - q')$$

et la solution $\varphi(q, \mathfrak{r})$ de la diffusion dans \mathbb{H} à partir de l'état initial $\varphi(q, 0) \in L^2(\mathbb{H})$ est donnée par

$$\varphi(q, \mathfrak{r}) = \int_{\mathbb{H}} K_{\mathfrak{r}}(q, q') \varphi(q', 0) dq' .$$

Si $\varphi(q, 0) = \delta(q)$, $\varphi(q, \mathfrak{r}) = K_{\mathfrak{r}}(q, 0) = P_{\mathfrak{r}}(q)$ et donc $P_{\mathfrak{r}}(q)$ représente bien la diffusion sous-riemannienne d'une masse de Dirac à l'origine 0 de \mathbb{H} .

La formule remarquable de Gaveau se démontre en appliquant un théorème fondamental de Paul Lévy sur le mouvement brownien (et $\mathfrak{Z}(\mathfrak{r})$ est un mouvement brownien sur $\mathbb{C}_{(z)}$) à la transformée de Fourier $\widehat{P}_{\mathfrak{r}}(z, t)$ de $P_{\mathfrak{r}}(z, t)$ ($\zeta^* = (\xi_1^*, \xi_2^*)$) :

$$\widehat{P}_{\mathfrak{r}}(\zeta^*, \theta^*) = \int_{\mathbb{R}^3} e^{i(\langle \zeta^*, z \rangle + \theta^* t)} P_{\mathfrak{r}}(z, t) dz dt .$$

Pour calculer \widehat{P} , Gaveau l'écrit sous forme d'espérance \mathcal{E} pour $(z(\mathbf{r}), t(\mathbf{r}))$ par rapport à la mesure

$$P_{\mathbf{r}}(z, t) dz dt : \widehat{P}_{\mathbf{r}}(\zeta^*, \theta^*) = \mathcal{E} \left(e^{i(\langle \zeta^*, z(\mathbf{r}) \rangle + \theta^* t(\mathbf{r}))} \right) = \mathcal{E} \left(e^{i\langle \zeta^*, z(\mathbf{r}) \rangle} \mathcal{E}^{z(\mathbf{r})} \left(e^{i\theta^* t(\mathbf{r})} \right) \right)$$

avec $\mathcal{E}^{z(\mathbf{r})}$ l'espérance conditionnelle pour l'extrémité $z(\mathbf{r}) = z$ fixée. Mais

$$\mathcal{E}^{z(\mathbf{r})=z} \left(e^{i\theta^* t(\mathbf{r})} \right) = \mathcal{E}^z \left(e^{2i\theta^* \int_0^{\tau} (x_2 dx_1 - x_1 dx_2)} \right)$$

et le résultat de Lévy dit que cette espérance est égale à

$$\frac{2\mathbf{r}\theta^*}{\sinh(2\mathbf{r}\theta^*)} e^{\frac{|z|^2}{2\mathbf{r}}(1-2\mathbf{r}\theta^* \coth(2\mathbf{r}\theta^*))} .$$

On en tire alors

$$\widehat{P}_{\mathbf{r}}(\zeta^*, \theta^*) = \frac{1}{\cosh(2\mathbf{r}\theta^*)} e^{-\frac{1}{2}|\zeta^*|^2 \frac{\tanh(2\mathbf{r}\theta^*)}{2\theta^*}}$$

puis $P_{\mathbf{r}}(z, t)$ par transformée de Fourier inverse.¹⁰³

L'action complexe Σ intervient pour la raison suivante. Nous avons vu à la section 2.2.3 du chapitre 14 que la valeur de l'hamiltonien H le long d'une géodésique allant de l'origine $(0, 0, 0)$ à un point terminal $(z = z(\tau), t = t(\tau))$ est donnée par

$$H_0 = \frac{2\theta^{*2}}{\sin^2(2\tau\theta^*)} |z|^2 .$$

Nous avons vu également que le lagrangien est donné par

$$L = \langle \xi^*, \dot{z} \rangle + \theta^* \dot{t} - H = \left\langle \xi^*, \frac{\partial H}{\partial \xi^*} \right\rangle + \theta^* \frac{\partial H}{\partial \theta^*} - H .$$

Mais la forme spéciale de H permet de calculer facilement L . En effet, $\frac{\partial H}{\partial \xi^*} = \zeta^*$ (rappelons que $\zeta^* = \xi^* + \theta^* \Lambda(z)$ et que $2H$ est la norme $\langle \zeta^*, \zeta^* \rangle$ de ζ^*) et $\frac{\partial H}{\partial \theta^*} = \langle \Lambda(z), \zeta^* \rangle$ et donc

$$\begin{aligned} \left\langle \xi^*, \frac{\partial H}{\partial \xi^*} \right\rangle + \theta^* \frac{\partial H}{\partial \theta^*} &= \langle \xi^*, \zeta^* \rangle + \theta^* \langle \Lambda(z), \zeta^* \rangle \\ &= \langle \xi^* + \theta^* \Lambda(z), \zeta^* \rangle = \langle \zeta^*, \zeta^* \rangle = 2H , \end{aligned}$$

c'est-à-dire $L = 2H$. L'action S est donc

$$S = \int_0^{\tau} H_0 d\tau = \frac{2\tau\theta^{*2}}{\sin^2(2\tau\theta^*)} |z|^2 = \theta^* (t - t_0) + \theta^* \cot(2\tau\theta^*) |z|^2 .$$

103. Une démonstration plus récente (2020) et originale (et très calculatoire) dans le contexte sous-riemannien se trouve dans Agrachev-Barilari-Boscain [5].

L'intégrale donnant $P(z, t, \mathfrak{r})$ est une "intégrale oscillante" difficile à évaluer.¹⁰⁴ On cherche à connaître en priorité sa valeur *asymptotique* pour $\mathfrak{r} \rightarrow 0$, c'est-à-dire son "approximation semi-classique". Pour avoir une idée des calculs, écrivons d'abord l'action complexe Σ sous la forme $\Sigma(z, t, \tau) = \tau\Gamma(z, t, \tau)$ avec

$$\Gamma(z, t, \tau) = -it + |z|^2 \coth(2\tau) .$$

On voit que pour $\theta^* = -i$ on a

$$\Gamma(z, t, \tau) = S(z, t, \tau, -i) - it_0 .$$

En tant qu'action, Γ satisfait l'équation d'Hamilton-Jacobi et il est facile de voir que celle-ci s'écrit

$$\frac{\partial \Gamma}{\partial \tau} + \frac{1}{\tau} (\Gamma + it_0) = 0 .$$

Supposons $z \neq 0$ (c'est-à-dire qu'on n'est pas dans le centre Z de \mathbb{H}) et calculons alors les points critiques de Σ par rapport à τ . Comme on a

$$\frac{\partial \Sigma}{\partial \tau} = \Gamma(z, t, \tau) + \tau \frac{\partial \Gamma}{\partial \tau} ,$$

on voit que $\frac{\partial \Sigma}{\partial \tau} = -it_0$. Les points critiques de l'action complexe correspondent donc à $t_0 = 0$ considéré comme une condition sur τ . Pour ces valeurs critiques τ_c on a l'égalité $\Gamma(z, t, \tau_c) = S(z, t, \tau_c, -i)$.

10.1.2. *Asymptotique pour $\mathfrak{r} \rightarrow 0^+$.*

Revenons à la solution fondamentale

$$P_{\mathfrak{r}}(z, t) = \frac{1}{(2\pi\mathfrak{r})^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \exp\left(\frac{i\tau t}{\mathfrak{r}} - \left(\frac{|z|^2}{2\mathfrak{r}}\right) \frac{2\tau}{\tanh(2\tau)}\right) d\tau .$$

Il est intéressant d'en étudier certaines valeurs.

On remarquera d'abord que par rapport au cas euclidien où $P_{\mathfrak{r}}(x) = \frac{1}{(2\sqrt{\pi\mathfrak{r}})^3} e^{-\frac{\|x\|^2}{4\mathfrak{r}}}$ est une gaussienne avec un facteur en $\mathfrak{r}^{-\frac{3}{2}}$ et donc $P_{\mathfrak{r}}(x=0) = \frac{1}{(2\sqrt{\pi\mathfrak{r}})^3}$, on a ici un facteur en $\mathfrak{r}^{-2} = \mathfrak{r}^{-\frac{4}{2}}$ avec

$$P_{\mathfrak{r}}(z=0, t=0) = \frac{1}{(2\pi\mathfrak{r})^2} \frac{\pi^2}{4} = \frac{1}{16\mathfrak{r}^2}$$

(avec pour \mathfrak{r} la divergence de la distribution de Dirac δ) puisque $\int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} d\tau = \frac{\pi^2}{4}$. Cela est dû au fait que, comme nous l'avons vu plusieurs fois,¹⁰⁵ la dimension de Hausdorff h de \mathbb{H} est 4 et non pas 3. Bien que de dimension 3 comme espace vectoriel

104. Cf. chapitre 16, section 9.2. Rappelons aussi [418] et, surtout, sa bibliographie.

105. Cf. par exemple les sections 7 du chapitre 6 et 8 du chapitre 15.

et espace topologique, \mathbb{H} muni de sa distance sous-riemannienne $d_{\mathcal{K}}$ est métriquement de dimension

$$h = 1 \times \dim(\text{Span}\{X_1, X_2\}) + 2 \times \dim(\text{Span}\{[X_1, X_2]\}) = 1.2 + 2.1 = 4 .$$

Il s'agit là d'un résultat général : en géométrie sous-riemannienne, où la condition d'Hörmander implique que le laplacien est sous-elliptique, la diffusion possède la forme

$$f(x, \mathfrak{r}) = \frac{1}{\alpha \mathfrak{r}^{h/2}} \int_M e^{-\frac{d(x-y)^2}{4\mathfrak{r}}} u(y) dy$$

où h est la dimension de Hausdorff de l'espace M .

Pour $t = 0$, on est dans le plan de base de coordonnée $z = x_1 + ix_2$ et l'on doit estimer

$$P_{\mathfrak{r}}(z, 0) = \frac{1}{(2\pi\mathfrak{r})^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \exp\left(-\left(\frac{|z|^2}{2\mathfrak{r}}\right) \frac{2\tau}{\tanh(2\tau)}\right) d\tau$$

pour $\mathfrak{r} \rightarrow 0^+$. On peut pour cela appliquer la *méthode de Laplace*¹⁰⁶ qui dit que lorsque $\mathfrak{r} \rightarrow 0^+$ et que la phase $-\left(\frac{|z|^2}{2\mathfrak{r}}\right) \frac{2\tau}{\tanh(2\tau)}$ de l'exponentielle (en la variable τ sur laquelle on intègre) a son facteur en $\frac{1}{\mathfrak{r}}$ qui $\rightarrow \infty$, alors l'intégrale se concentre sur les points critiques (relativement à τ) de la phase.

Posons $u = 2\tau$ ce qui introduit un facteur $\frac{1}{2}$ devant l'intégrale. On doit évaluer l'intégrale

$$\int_{\mathbb{R}} \frac{u}{\sinh(u)} \exp\left(-\frac{1}{\mathfrak{r}} \left(\frac{|z|^2}{2}\right) \frac{u}{\tanh(u)}\right) du = \int_{\mathbb{R}} f(u) \exp\left(\frac{1}{\mathfrak{r}} \psi(u)\right) du ,$$

avec les fonctions paires

$$\begin{cases} \psi(u) = -\left(\frac{|z|^2}{2}\right) \frac{u}{\tanh(u)} = -\left(\frac{|z|^2}{2}\right) \varphi(u) \\ \varphi(u) = \frac{u}{\tanh(u)} \end{cases}$$

Le théorème de Laplace permet d'estimer une intégrale de la forme

$$I(\lambda) = \int_{\mathbb{R}} f(u) \exp(\lambda\psi(u)) du$$

lorsque $\psi(u)$ est C^2 et possède un point critique unique non dégénéré (un maximum strict) en u_0 (i.e. $\psi'(u_0) = 0$ et $\psi''(u_0) > 0$), que $f(u_0) \neq 0$ et que l'intégrande

106. Une généralisation de la méthode de Laplace aux phases imaginaires pures (intégrales oscillantes) est la méthode de la phase stationnaire que nous avons déjà évoquée dans les sections 9 et 6 des chapitres 9 et 16. Il y a aussi la méthode dite "du col".

considéré est bien dans L^1 (ce qui est le cas ici). La formule est

$$I(\lambda) \underset{\lambda \rightarrow \infty}{\sim} \frac{\sqrt{2\pi}}{\sqrt{\psi''(u_0)}} \frac{1}{\sqrt{\lambda}} f(u_0) \exp(\lambda\psi(u_0)) .$$

Appliquons cette formule à notre cas où $\lambda = \frac{1}{\mathfrak{r}}$. On a $\psi(u) = -\left(\frac{|z|^2}{2}\right) \varphi(u)$. Comme $\tanh(u)$ tend très vite vers 1 avec le signe de u , $\varphi(u)$ tend très vite vers $|u|$. Comme $\frac{u}{\tanh(u)} \underset{u \rightarrow 0}{\sim} 1 + \frac{u^2}{3}$ au 2^e ordre, on a $\varphi(0) = 1$ et donc $\psi(0) = -\left(\frac{|z|^2}{2}\right)$. On a

$$\varphi'(u) = \left(\frac{u}{\tanh(u)}\right)' = \operatorname{cotanh}(u) - \frac{u}{\sinh^2(u)} \underset{u \rightarrow 0}{\sim} \frac{2u}{3} - \frac{4u^3}{45} \text{ au } 3^{\text{e}} \text{ ordre}$$

$$\begin{aligned} \varphi''(u) &= \left(\frac{u}{\tanh(u)}\right)'' = -\frac{2}{\sinh^2(u)} (1 - 2u \operatorname{cotanh}(u)) \\ &= -\frac{2}{\sinh^3(u)} (\sinh(u) - u \cosh(u)) \underset{u \rightarrow 0}{\sim} \frac{2}{3} - \frac{4u^2}{15} \text{ au } 2^{\text{e}} \text{ ordre} \end{aligned}$$

et donc 0 est un point critique (un minimum) non dégénéré de $\varphi(u)$ et donc de $\psi(u)$ avec $\psi''(u) = -\frac{2}{3} \left(\frac{|z|^2}{2}\right) = -\frac{|z|^2}{3}$. La formule de Laplace donne

$$\frac{\sqrt{2\pi}}{\sqrt{\psi''(u_0)}} \frac{1}{\sqrt{\lambda}} f(u_0) \exp(\lambda\psi(u_0))$$

et donc

$$P_{\mathfrak{r}}(z, 0) \underset{\mathfrak{r} \rightarrow 0^+}{\sim} \frac{1}{(2\pi\mathfrak{r})^2} \frac{1}{2} \frac{\sqrt{2\pi}}{\sqrt{\frac{|z|^2}{3}}} \sqrt{\mathfrak{r}} e^{-\left(\frac{|z|^2}{2\mathfrak{r}}\right)} = \frac{1}{2} \frac{\sqrt{3}}{(2\pi\mathfrak{r})^{\frac{3}{2}}} \frac{1}{|z|} e^{-\left(\frac{|z|^2}{2\mathfrak{r}}\right)} .$$

Cette approximation déduite de la formule de Laplace n'est pas valide pour $z = 0$ car alors $\psi(u)$ est identiquement nulle.

Par ailleurs, pour $z = 0$, on est dans la fibre des t où tous les points sont des points de la caustique de 0 et l'on trouve

$$\begin{aligned} P_{\mathfrak{r}}(0, t) &= \frac{1}{(2\pi\mathfrak{r})^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \exp\left(\frac{i\tau t}{\mathfrak{r}}\right) d\tau = \frac{1}{(2\pi\mathfrak{r})^2} \int_{\mathbb{R}} \frac{2\tau}{\sinh(2\tau)} \cos\left(\frac{\tau t}{\mathfrak{r}}\right) d\tau \\ &= \frac{1}{(2\pi\mathfrak{r})^2} \frac{\pi^2}{4 \cosh^2\left(\frac{\pi t}{4\mathfrak{r}}\right)} = \frac{1}{(2\mathfrak{r})^2} \frac{1}{4 \cosh^2\left(\frac{\pi t}{4\mathfrak{r}}\right)} . \end{aligned}$$

Il s'agit là encore d'une intégrale oscillante. On intègre en utilisant des contours appropriés dans le plan complexe en tenant compte des pôles de $\frac{2\tau}{\sinh(2\tau)}$ pour

$$2\tau = ik\pi (k > 0) ,$$

pôles où le résidu est

$$(-1)^k i k \pi \exp\left(-\frac{k\pi t}{2\mathfrak{r}}\right).$$

D'où, en sommant sur les résidus et en utilisant la formule $\frac{a}{(1+a)^2} = a - 2a^2 + 3a^3 - \dots$ avec $a = \exp\left(-\frac{k\pi t}{2\mathfrak{r}}\right)$, l'expression

$$P_{\mathfrak{r}}(0, t) \underset{\mathfrak{r} \rightarrow 0^+}{\sim} \frac{1}{(2\mathfrak{r})^2} \frac{\exp\left(-\frac{\pi t}{2\mathfrak{r}}\right)}{\left(1 + \exp\left(-\frac{\pi t}{2\mathfrak{r}}\right)\right)^2} = \frac{1}{(2\mathfrak{r})^2} \frac{1}{4 \cosh^2\left(\frac{\pi t}{4\mathfrak{r}}\right)}$$

qui redonne bien $\frac{1}{16\mathfrak{r}^2}$ pour $t \sim 0$.

Dans le cas général, les calculs sont nettement plus complexes (cf. Gaveau [206]) et l'on doit, encore une fois, utiliser le principe de la phase stationnaire. Gaveau trouve

$$P_{\mathfrak{r}}(z, t) \sim \exp\left(-\frac{S(z, t, \mathfrak{r})}{2}\right) \left(\frac{\Phi(t)}{\mathfrak{r}^{\frac{3}{2}}} + O(\mathfrak{r}^{-1})\right)$$

avec, en posant $T = T\left(\frac{t}{|z|^2}\right)$ où $T(\varphi)$ est la fonction $\frac{\varphi}{\sin^2(\varphi)} - \cot(\varphi)$,

$$\Phi(t) = \frac{1}{(2\pi)^{\frac{3}{2}} |z|} \left(\frac{\sin(T)}{\sin(T) - T \cos(T)}\right)^{\frac{1}{2}}.$$

Nous avons vu à la sous-section 2.1.3 de la section 2 du chapitre 14 que $T(\varphi)$ est la fonction qui contrôle la multiplicité des géodésiques du groupe de Heisenberg \mathbb{H} . Nous voyons donc bien comment la diffusion suit les géodésiques.

Le fait que $\log(P_{\mathfrak{r}}(z, t)) \underset{\mathfrak{r} \rightarrow 0^+}{\sim} -\frac{S(z, t, \mathfrak{r})}{2}$ où S est l'action le long de la géodésique minimale de 0 à (z, t) est fondamental.

10.1.3. Généralisations.

La formule de Beals-Gaveau-Greiner pour \mathbb{H} peut être généralisée. De nombreux travaux y ont été consacrés. Wojciech Cygan [130] l'a étendu à tous les groupes de Carnot de niveau 2 de type Heisenberg.¹⁰⁷ Le lecteur intéressé pourra par exemple se référer à l'article de Nicola Garofalo et Giulio Tralli [204] qui vont beaucoup plus loin dans la généralisation.

10.1.4. Diffusion, points conjugués et cut locus.

La diffusion étant guidée par les géodésiques il faut s'attendre à ce que l'existence de points singuliers comme les points du cut locus où des géodésiques de même longueur sont en compétition (perte d'optimalité globale par conflit) et les points conjugués sur l'enveloppe des géodésiques (perte d'optimalité locale par bifurcation) transforme les formules générales. Nous avons vu aux sections 9.1 et 9.2 que le noyau de la chaleur $P_{\mathfrak{r}}(q, q')$ possède dans les cas euclidiens un facteur en $\frac{1}{\mathfrak{r}^\alpha}$ où α dépend

107. Voir par exemple Kaplan [282] pour la théorie générale de ces groupes.

de la dimension $(\frac{n}{2}$ pour $n \geq 3$) suivi d'une exponentielle $e^{-\frac{d(q,q')^2}{4\tau}}$ où $d(q, q')$ est la distance entre q et q' .¹⁰⁸

Considérons la diffusion sur la sphère \mathbb{S}^2 partant d'un δ de Dirac au point q (pôle Nord). Les géodésiques de q à q' sont bien définies sauf si q' est le point antipodal $c(q)$ (pôle Sud) qui est le cut locus de q (et aussi son lieu conjugué). Nous les avons explicitées dans la sous-section 6.2 de la section 6 du chapitre 12. Toutes les géodésiques partant de q convergent vers $c(q)$ et l'on s'imagine intuitivement que la chaleur diffusant de q va se reconcentrer en $c(q)$. On montre effectivement que

$$P_{\tau}(q, q') \sim \frac{1}{4\pi\tau} e^{-\frac{d(q,q')^2}{4\tau}}$$

si $q' \neq c(q)$ mais que en revanche

$$P_{\tau}(q, c(q)) \sim \frac{C}{\tau^{\frac{3}{2}}} e^{-\frac{d(q,c(q))^2}{4\tau}}$$

si $q' = c(q)$.¹⁰⁹

Pour les ellipsoïdes, nous avons également explicité les géodésiques dans la section 6 du chapitre 12 et nous avons rappelé les travaux qui, de Jacobi (ellipsoïdes de révolution) jusqu'à Itoh et Kiyohara en 2004 (ellipsoïdes triaxiaux), ont montré que, génériquement, le lieu conjugué d'un point est une astroïde à quatre cusps et le cut locus un segment joignant deux cusps opposés. Davide Barilari et Jacek Jendrej ont prouvé en 2013 dans [32] que, pour les ellipsoïdes de révolution, si q est sur l'équateur et si q' est une extrémité cusp du cut locus de q (q' est donc aussi un point conjugué de q), on a en temps petit (i.e. pour $\tau \rightarrow 0^+$) la valeur asymptotique

$$P_{\tau}(q, q') \sim \frac{C}{\tau^{\frac{5}{4}}} e^{-\frac{d(q,q')^2}{4\tau}} .$$

En fait, à partir de la fin des années 1960 (Srinivasa Varadhan 1967, S. Molchanov 1975, Robert Azencott 1980, Bernard Gaveau 1977, Jean-Michel Bismut 1984, Rémi Léandre 1987), on s'est mis à étudier de façon détaillée les asymptotiques en temps petit des diffusions riemanniennes et sous-riemanniennes. Une des motivations était qu'il existe un lien étroit entre ces asymptotiques et les propriétés *géométriques* de la variété considérée. Une bonne synthèse des travaux de l'époque se trouve dans le recueil [22] de 1981 édité par Robert Azencott. Un résultat général assez satisfaisant est le théorème de Ben Arous [37] de 1988 disant que lorsque $q \neq q'$ sont reliés par une unique géodésique minimisante "normale" (projection d'une bicaractéristique)

108. Le noyau de la chaleur $P_{\tau}(q, q')$ dépend de deux points. Ci-dessus $P_{\tau}(z, t) = P_{\tau}(q)$ abrégéait $P_{\tau}(0, q)$.

109. Dans les paragraphes suivants le dénominateur $\frac{1}{4\tau}$ dans l'exponentielle correspond au choix de Δ pour l'équation de la chaleur. Le choix de $\frac{1}{2}\Delta$ conduit à $\frac{1}{2\tau}$.

le long de laquelle ils ne sont pas conjugués, on a

$$P_{\mathfrak{r}}(q, q') \sim \frac{C}{\mathfrak{r}^{\frac{n}{2}}} e^{-\frac{d(q, q')^2}{4\mathfrak{r}}}$$

(n étant la dimension de la variété considérée). En fait, on a même un développement limité en

$$\frac{1}{\mathfrak{r}^{\frac{n}{2}}} e^{-\frac{d(q, q')^2}{4\mathfrak{r}}} \left(\sum_{k \geq 0} c_k(q, q') \mathfrak{r}^k \right)$$

assurant la dérivabilité et l'uniformité sur les compacts. Les coefficients $c_k(q, q')$ satisfont des équations de transport analogues à celles que nous avons vues pour les caustiques en optique. Les méthodes utilisées pour démontrer ce résultat sont du type de celles que nous avons déjà évoquées plus haut et mobilisent beaucoup de résultats d'analyse fonctionnelle sophistiquée.

Mais la situation se complique évidemment lorsqu'on se situe en des points du cut locus et des points conjugués. Pour les surfaces riemanniennes complètes quelconques on montre (Davide Barilari, Ugo Boscain, Robert Neel [31], 2012) que, pour les points conjugués le long d'au moins une géodésique minimale, on a la valeur asymptotique pour $\mathfrak{r} \rightarrow 0^+$

$$\frac{C_1}{\mathfrak{r}^{\frac{5}{2}}} e^{-\frac{d(q, c(q))^2}{4\mathfrak{r}}} \leq P_{\mathfrak{r}}(q, q') \leq \frac{C_2}{\mathfrak{r}^{\frac{3}{4}}} e^{-\frac{d(q, q')^2}{4\mathfrak{r}}} .$$

En géométrie sous-riemannienne un point q peut appartenir à son cut-locus. Dans ce cas Gérard Ben Arous a montré en 1989 [38] que sur la diagonale (où $d(q, q) = 0$ et donc $e^{-\frac{d(q, q)^2}{4\mathfrak{r}}} = 1$), on a

$$P_{\mathfrak{r}}(q, q) \sim \frac{C}{\mathfrak{r}^{\frac{h}{2}}}$$

où h est la dimension *homogène* de la variété sous-riemannienne.

Le théorème de Barilari, Boscain, Neel [31] de 2012 juste évoqué dit que si $q' \neq q$ est cut-conjugué de q le long d'au moins une géodésique minimale, alors en temps petit

$$\frac{C_1}{\mathfrak{r}^{\frac{n}{2} + \frac{1}{4}}} e^{-\frac{d(q, q')^2}{4\mathfrak{r}}} \leq P_{\mathfrak{r}}(q, q') \leq \frac{C_2}{\mathfrak{r}^{n - \frac{1}{4}}} e^{-\frac{d(q, q')^2}{4\mathfrak{r}}} .$$

Pour Heisenberg et les points du cut locus de l'origine (l'axe des t) on a (en renormalisant avec Δ notre formule qui était avec $\frac{1}{2}\Delta$)

$$P_{\mathfrak{r}}(0, t) \underset{\mathfrak{r} \rightarrow 0^+}{\sim} \frac{C}{\mathfrak{r}^2} e^{-\frac{t}{4\mathfrak{r}}} = \frac{C}{\mathfrak{r}^2} e^{-\frac{d(0, (z, t))^2}{4\mathfrak{r}}}$$

et l'on a bien $\frac{3}{2} + \frac{1}{4} \leq 2 \leq 3 - \frac{1}{4}$.

10.1.5. *Noyau de Heisenberg polarisé (Hulanicki).*

En 1976, l'année précédant la publication de Bernard Gaveau, Andrzej Hulanicki avait trouvé la même formule, mais pour Heisenberg polarisé et avec une méthode différente. Il est intéressant de voir deux façons différentes de calculer la même formule.

On revient donc aux variables $v = (x, y, p) \in \mathbb{V}_J$ et $(\xi^*, \eta^*, \pi^*) \in \mathcal{V}_J^*$. Le contexte est celui d'une conception physique du groupe de Heisenberg polarisé $\mathbb{H}_{pol} = \mathbb{V}_J$. Hulanicki considère x et p soumis à des mouvements browniens indépendants $\mathfrak{X}(\mathbf{r})$ et $\mathfrak{P}(\mathbf{r})$, l'un concernant une "position" x et l'autre une "force" p et il considère y comme une "énergie" donnée par $\mathfrak{Y}(\mathbf{r}) = \int_0^{\mathbf{r}} \mathfrak{P}(\mathbf{r}) d\mathfrak{X}(\mathbf{r})$.¹¹⁰ Un incrément élémentaire $d\mathfrak{X}(\mathbf{r})$ du brownien sur la position $\mathfrak{X}(\mathbf{r})$ à force $\mathfrak{P}(\mathbf{r})$ constante donnera un incrément de l'énergie $\mathfrak{P}(\mathbf{r}) d\mathfrak{X}(\mathbf{r})$ alors que, au contraire, un incrément élémentaire $d\mathfrak{P}(\mathbf{r})$ du brownien sur la force $\mathfrak{P}(\mathbf{r})$ à position $\mathfrak{X}(\mathbf{r})$ constante donnera un incrément nul de l'énergie. C'est ainsi que la loi de multiplication asymétrique de \mathbb{H}_{pol} intervient de façon naturelle.

Il s'agit de calculer explicitement le noyau $P_{\mathbf{r}}(v) = P_{\mathbf{r}}(x, y, p)$ de l'équation de la chaleur $\frac{\partial}{\partial \mathbf{r}} = \frac{1}{2}\Delta_{\mathcal{K}}$ et sa transformée de Fourier $\widehat{P}_{\mathbf{r}}(\xi^*, \eta^*, \pi^*)$ pour le Laplacien hypo-elliptique

$$\Delta_{\mathcal{K}} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial p^2} + 2p \frac{\partial^2}{\partial x \partial y} + p^2 \frac{\partial^2}{\partial y^2} .$$

Pour les variables (x, p) on a un mouvement brownien euclidien et le noyau de la chaleur est

$$\Pi_{\mathbf{r}}(x, p) = \frac{1}{2\pi\mathbf{r}} e^{-\frac{x^2+p^2}{2\mathbf{r}}} . \quad 111$$

On a

$$\Pi_{\mathbf{r}}(x, p) = \int_{\mathbb{R}} P_{\mathbf{r}}(x, y, p) dy .$$

Pour le calcul, Hulanicki utilise les unirreps de \mathbb{H}_{pol} ,

$$\rho_{\lambda}(x, y, p) u(s) = e^{i\lambda(y+xs)} u(s+p)$$

110. Comme nous l'avons vu à la section 2.3 du chapitre 3, cette interprétation "physique" est différente de (mais formellement équivalente à) l'interprétation "géométrique" de \mathbb{V}_J comme espace de jets où l'on a $dy = p dx$ et où un incrément de la position x à p constant produit un incrément $p dx$ de y alors qu'un incrément de la tangente p à x constant ne change pas y . Elle correspond à la complémentarité des connexions excitatrices inter-hypercolonnes et des connexions inhibitrices intra-hypercolonnes de $V1$.

111. Par rapport aux formules du cas euclidien déjà évoquées, l'exponentielle $e^{-\frac{x^2+p^2}{2\mathbf{r}}}$ plutôt que $e^{-\frac{x^2+p^2}{4\mathbf{r}}}$ vient du fait que l'auteur utilise, comme Gaveau et beaucoup d'autres spécialistes, $\frac{1}{2}\Delta_{\mathcal{K}}$ plutôt que $\Delta_{\mathcal{K}}$.

et leurs relations aux polynômes d'Hermite. Il note que les $\rho_\lambda(v)$ peuvent se prolonger en des $\rho_\lambda(P_\tau(v))$ au moyen de la formule

$$\rho_\lambda(P_\tau(v)) u(s) = \int_{\mathbb{R}^3} P_\tau(x, y, p) e^{i\lambda(y+xs)} u(s+p) dx dy dp$$

et qu'il s'agit là d'une transformée de Fourier partielle. En effet, on écrit

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^2} P_\tau(x, y, p) e^{i\lambda(y+xs)} dx dy \right) u(s+p) dp = \int_{\mathbb{R}} P_\tau(-\widehat{\lambda}s, -\widehat{\lambda}, p) u(s+p) dp$$

où un $\widehat{\lambda}$ à une place de variable z signifie qu'on a pris la transformée de Fourier par rapport à la variable correspondante (intégration par rapport à z avec le facteur $e^{-i\lambda z}$). En appliquant cela aux ondes planes $u(s) = e^{i\pi^* s}$, on trouve

$$\begin{aligned} \rho_\lambda(P_\tau(v)) e^{i\pi^* s} &= \int_{\mathbb{R}^3} P_\tau(x, y, p) e^{i\lambda(y+xs)} e^{i\pi^*(s+p)} dx dy dp \\ &= e^{i\pi^* s} \widehat{P}_\tau(-\lambda s, -\lambda, -\pi^*) . \end{aligned}$$

En utilisant alors des développements sur la base orthonormée des polynômes d'Hermite qui est adaptée à l'opérateur laplacien (nous y reviendrons dans la prochaine section), Hulanicki trouve la version polarisée de la formule de Gaveau :

$$\widehat{P}_\tau(\xi^*, \eta^*, \pi^*) = \frac{1}{\cosh(2\tau\eta^*)^{\frac{1}{2}}} \exp\left(\frac{-\frac{1}{2}(|\xi^*|^2 + |\eta^*|^2) \sinh(2\tau\eta^*) + i\xi^* \eta^* \sinh^2(\tau\eta^*)}{2\eta^* \cosh(2\tau\eta^*)}\right).$$

Lorsque $\eta^* \rightarrow 0$, on trouve, puisque $\cosh(\varepsilon) \sim 1$ et $\sinh(\varepsilon) \sim \varepsilon$ au premier ordre pour ε infiniment petit,

$$\begin{aligned} \widehat{P}_\tau(\xi^*, \varepsilon, \pi^*) &\sim \frac{1}{1} \exp\left(\frac{-\frac{1}{2}(|\xi^*|^2 + |\eta^*|^2) 2\tau\varepsilon + i\xi^* \eta^* (\tau\varepsilon)^2}{2\varepsilon}\right) \\ &\sim e^{-\frac{1}{2}(|\xi^*|^2 + |\eta^*|^2)\tau} \end{aligned}$$

qui est bien la transformée de Fourier de $\frac{1}{2\pi\tau} e^{-\frac{x^2+p^2}{2\tau}}$.

10.2. La transformée de Fourier non commutative pour \mathbb{V}_J

Donnons une présentation plus abstraite et générale des processus de diffusion de \mathbb{V}_J et \mathbb{V}_S en termes d'analyse harmonique non commutative du laplacien hypoelliptique associé à leur géométrie sous-riemannienne. Nous suivrons la façon dont, à partir des résultats de Gaveau et Hulanicki, Andrei Agrachev, Jean-Paul Gauthier, Ugo Boscain, et leur doctorant Francesco Rossi ont prouvé en 2009 [6] un théorème général pour les groupes de Lie de dimension 3 unimodulaires (i.e. ceux dont les mesures de Haar invariantes à droite et à gauche sont identiques) munis d'une géométrie sous-riemannienne invariante. Ils utilisent la transformée de Fourier

généralisée non commutative sur le dual \widehat{G} du groupe G (l'ensemble des unirreps de G) et ils calculent le noyau de la chaleur associé au laplacien hypo-elliptique $\Delta_{\mathcal{K}} = X_1^2 + X_2^2$ en termes de fonctions propres et de valeurs propres (i.e. en termes de spectres).¹¹²

Pour le groupe de Heisenberg polarisé \mathbb{V}_J nous avons vu à la section 5 que d'après le théorème de Stone-von Neumann les unirreps dans le groupe $\mathcal{U}(\mathcal{H})$ des automorphismes unitaires du Hilbert $\mathcal{H} = L^2(\mathbb{R}, \mathbb{C})$ sont paramétrées par un λ et données, dans la version (R4) du théorème, par

$$\begin{aligned} \rho_\lambda : \mathbb{V}_J &\rightarrow \mathcal{U}(\mathcal{H}) \\ v = (x, y, p) &\mapsto \rho_\lambda(v) : \mathcal{H} \rightarrow \mathcal{H} \\ &u(s) \mapsto e^{i\lambda(y+xs)}u(s+p) \end{aligned}$$

si elles sont non triviales sur le centre Z (l'axe y) et sinon par les "caractères"

$$\rho_{\mu,\nu}(x, y, p) u(s) = e^{i(\mu x + \nu p)}u(s) .$$

La mesure de Plancherel qui permet les intégrations sur $\widehat{\mathbb{V}}_J$ est $dP(\lambda) = \lambda d\lambda$. Pour calculer la transformée de Fourier du laplacien sous-riemannien $\Delta_{\mathcal{K}}$ on regarde l'action de la différentielle des unirreps sur les champs de vecteurs X invariants à gauche sur \mathbb{V}_J . Par définition,

$$d\rho_\lambda : X \rightarrow d\rho_\lambda(X) := \left. \frac{d}{dt} \right|_{t=0} \rho_\lambda(e^{tX})$$

et l'on obtient ainsi la transformée de Fourier $\widehat{X}_i^\lambda = d\rho_\lambda(X_i)$. On a

$$\begin{aligned} X_1(0) &= (1, 0, 0) , \\ e^{tX_1} &= (t, 0, 0) , \\ \rho_\lambda(e^{tX_1})u(s) &= e^{i\lambda ts}u(s) , \\ \widehat{X}_1^\lambda u(s) &= d\rho_\lambda(X_1)u(s) = \left. \frac{d}{dt} \right|_{t=0} \rho_\lambda(e^{tX_1})u(s) \\ &= \left. \frac{d}{dt} \right|_{t=0} e^{i\lambda ts}u(s) = i\lambda s u(s) , \end{aligned}$$

ainsi que

112. Rappelons qu'Hulanicki et Gaveau utilisaient $\frac{1}{2}\Delta_{\mathcal{K}}$ pour l'équation de la chaleur et non pas $\Delta_{\mathcal{K}}$ comme ici.

$$\begin{aligned}
X_2(0) &= (0, 0, 1) , \\
e^{tX_2} &= (0, 0, t) , \\
\rho_\lambda(e^{tX_2}) u(s) &= u(s+t) , \\
\widehat{X}_2^\lambda u(s) &= d\rho_\lambda(X_2) u(s) = \left. \frac{d}{dt} \right|_{t=0} \rho_\lambda(e^{tX_2}) u(s) \\
&= \left. \frac{d}{dt} \right|_{t=0} u(s+t) = \frac{du(s)}{ds} .
\end{aligned}$$

La transformée de Fourier du laplacien sous-riemannien est par conséquent la somme hilbertienne (l'intégrale sur λ avec la mesure de Plancherel $dP(\lambda) = \lambda d\lambda$) des $\widehat{\Delta}_{\mathcal{K}}^\lambda$ avec

$$\widehat{\Delta}_{\mathcal{K}}^\lambda u(s) = \left((\widehat{X}_1^\lambda)^2 + (\widehat{X}_2^\lambda)^2 \right) u(s) = \frac{d^2 u(s)}{ds^2} - \lambda^2 s^2 u(s) .$$

Cette équation n'est rien d'autre que celle de l'*oscillateur harmonique*. Nous l'avons étudiée plus haut section 5.10.2 et aussi dans la section 8.6.4 du chapitre 16 à propos de la quantification géométrique.

Dans une formulation plus générale, on peut dire que le noyau de la chaleur est donné par l'intégrale

$$P(v, t) = \int_{\widehat{\mathbb{V}}_J} \text{tr} \left(e^{t\widehat{\Delta}_{\mathcal{K}}^\lambda} \rho_\lambda(v) \right) dP(\lambda) , \quad t \geq 0 .$$

et que si les $\widehat{\Delta}_{\mathcal{K}}^\lambda$ ont un spectre *discret* et un ensemble complet de fonctions propres normalisées $\{u_n^\lambda\}$ alors

$$P(v, t) = \int_{\widehat{\mathbb{V}}_J} \left(\sum_n e^{\alpha_n^\lambda t} \langle u_n^\lambda, \rho_\lambda(v) (u_n^\lambda) \rangle \right) dP(\lambda) , \quad t \geq 0 .$$

C'est le cas ici. Les fonctions propres u_n^λ de l'oscillateur harmonique ont pour valeurs propres les $\alpha_n^\lambda = -\frac{2n+1}{\lambda}$ et sont essentiellement les fonctions d'Hermite scalées par λ :

$$u_n^\lambda(s) = (2^n n! \sqrt{\pi})^{-\frac{1}{2}} \lambda^{\frac{1}{4}} e^{-\lambda \frac{s^2}{2}} H_n(\sqrt{\lambda} s)$$

H_n étant le n -ième polynôme d'Hermite. On obtient ainsi le noyau dit *noyau de Mehler*.¹¹³

113. Ferdinand Mehler (1835-1895) introduisit la formule en 1866.

Avant que d'en venir aux formules analogues pour $\mathbb{V}_S = SE(2)$ faisons une petite remarque.

Remarque. Le problème est que le groupe de Heisenberg, de par sa nilpotence et ses automorphismes de dilatation non isotrope $(x, t) \rightarrow (\delta x, \delta^2 t)$ est un mauvais modèle pour les variétés de Heisenberg. Comme nous l'avons vu au chapitre 14, section 8.3, il n'en constitue que la structure "tangente". Or le groupe $SE(2)$ des déplacements du plan n'est pas nilpotent et, aussi étonnant que cela puisse paraître étant donné le caractère ultra classique de ce groupe, le problème de trouver une formule *explicite* pour la diffusion et le noyau de chaleur n'est pas encore résolu! \square

10.3. La transformée de Fourier non commutative pour $\mathbb{V}_S = SE(2)$

Dans [6], Andrei Agrachev, Ugo Boscain, Jean-Paul Gauthier and Francesco Rossi ont également donné la version *spectrale* du noyau de la chaleur pour $\mathbb{V}_S = SE(2)$ et le laplacien hypo-elliptique $\Delta_{\mathcal{K}} = X_1^2 + X_2^2$. Cela permet d'approximer les processus de diffusion, même lorsqu'on ne connaît pas de formule explicite du noyau de la chaleur (ce qui est précisément le cas de $SE(2)$). La diffusion sous-riemannienne sur \mathbb{V}_S reste très fortement anisotrope puisqu'elle se réduit à une diffusion angulaire de θ et à une diffusion spatiale restreinte à la direction X_1 . Elle satisfait la loi gestaltiste de "bonne continuation" et sa différence avec la diffusion euclidienne classique est frappante comme le montrent les images d'inpainting de la section suivante 11.

Le dual $\widehat{\mathbb{V}}_S$ du groupe \mathbb{V}_S est l'espace des unirreps de \mathbb{V}_S dans l'espace de Hilbert $\mathcal{H} = L^2(\mathbb{S}^1, \mathbb{C})$ et les unirreps génériques sont paramétrées par λ :

$$\begin{aligned} \rho_\lambda : \mathbb{V}_S &\rightarrow \mathcal{U}(\mathcal{H}) \\ v = (x, y, \alpha) &\mapsto \rho_\lambda(v) : \mathcal{H} \rightarrow \mathcal{H} \\ &\psi(\theta) \mapsto e^{i\lambda(x \sin(\theta) + y \cos(\theta))} \psi(\theta + \alpha) . \end{aligned}$$

La mesure de Plancherel sur $\widehat{\mathbb{V}}_S$ est encore $dP(\lambda) = \lambda d\lambda$ et pour calculer la transformée de Fourier nous devons regarder l'action de la différentielle des unirreps sur les champs de vecteurs invariants à gauche X . Comme précédemment,

$$d\rho_\lambda : X \rightarrow d\rho_\lambda(X) := \left. \frac{d}{dt} \right|_{t=0} \rho_\lambda(e^{tX})$$

et l'on a $\widehat{X}_i^\lambda = d\rho_\lambda(X_i)$ avec

$$\begin{aligned}
X_1(0) &= (1, 0, 0) , \\
e^{tX_1} &= (t, 0, 0) , \\
\rho_\lambda(e^{tX_1})\psi(\theta) &= e^{i\lambda t \sin(\theta)}\psi(\theta) , \\
\widehat{X}_1^\lambda \psi(\theta) &= d\rho_\lambda(X_1)\psi(\theta) = \left. \frac{d}{dt} \right|_{t=0} \rho_\lambda(e^{tX_1})\psi(\theta) \\
&= \left. \frac{d}{dt} \right|_{t=0} e^{i\lambda t \sin(\theta)}\psi(\theta) = i\lambda \sin(\theta)\psi(\theta) ,
\end{aligned}$$

ainsi que

$$\begin{aligned}
X_2(0) &= (0, 0, 1) , \\
e^{tX_2} &= (0, 0, t) , \\
\rho_\lambda(e^{tX_2})\psi(\theta) &= \psi(\theta + t) , \\
\widehat{X}_2^\lambda \psi(\theta) &= d\rho_\lambda(X_2)\psi(\theta) = \left. \frac{d}{dt} \right|_{t=0} \rho_\lambda(e^{tX_2})\psi(\theta) \\
&= \left. \frac{d}{dt} \right|_{t=0} \psi(\theta + t) = \frac{d\psi(\theta)}{d\theta} .
\end{aligned}$$

La transformée de Fourier du laplacien sous-elliptique est la somme hilbertienne (l'intégrale en λ avec la mesure de Plancherel $dP(\lambda) = \lambda d\lambda$) des $\widehat{\Delta}_{\mathcal{K}}^\lambda$

$$\widehat{\Delta}_{\mathcal{K}}^\lambda \psi(\theta) = \left(\left(\widehat{X}_1^\lambda \right)^2 + \left(\widehat{X}_2^\lambda \right)^2 \right) \psi(\theta) = \frac{d^2\psi(\theta)}{d\theta^2} - \lambda^2 \sin^2(\theta)\psi(\theta) .$$

On reconnaît là l'équation de Mathieu.¹¹⁴ Pour les petits angles on retrouve l'équation

$$\widehat{\Delta}_{\mathcal{K}}^\lambda \psi(\theta) = \frac{d^2\psi(\theta)}{d\theta^2} - \lambda^2 \theta^2 \psi(\theta)$$

et le noyau de Mehler $\widehat{\Delta}_{\mathcal{K}}^\lambda u(s) = \frac{d^2 u(s)}{ds^2} - \lambda^2 s^2 u(s)$.

Le noyau de la chaleur est maintenant

$$P(v, t) = \int_{\widehat{\mathbb{V}}_S} \text{Tr} \left(e^{t\widehat{\Delta}_{\mathcal{K}}^\lambda} \rho_\lambda(v) \right) dP(\lambda) , \quad t \geq 0 .$$

114. Émile Mathieu (1835-1890) a introduit son équation en 1865.

Comme précédemment, les $\widehat{\Delta}_{\mathcal{K}}^\lambda$ ont un spectre discret et un ensemble complet d'états propres normalisés $\{\psi_n^\lambda\}$ avec des valeurs propres $\{\alpha_n^\lambda\}$ et donc

$$P(v, t) = \int_{\widehat{\mathbb{V}}_S} \left(\sum_n e^{\alpha_n^\lambda t} \langle \psi_n^\lambda, \rho_\lambda(v) (\psi_n^\lambda) \rangle \right) dP(\lambda), \quad t \geq 0.$$

Les fonctions propres 2π -périodiques de l'équation de Mathieu satisfont

$$\frac{d^2\psi(\theta)}{d\theta^2} - \lambda^2 \sin^2(\theta) \psi(\theta) = E\psi(\theta)$$

et, puisque $\sin^2(\theta) = \frac{1}{2}(1 - \cos(2\theta))$,

$$\begin{aligned} \frac{d^2\psi(\theta)}{d\theta^2} - \frac{\lambda^2}{2}\psi(\theta) - E\psi(\theta) + \frac{\lambda^2}{2}\cos(2\theta)\psi(\theta) &= 0 \\ \frac{d^2\psi(\theta)}{d\theta^2} + (a - 2r\cos(2\theta))\psi(\theta) &= 0, \end{aligned}$$

avec $a = -\left(\frac{\lambda^2}{2} + E\right)$ et $r = -\frac{\lambda^2}{4}$. Les fonctions propres 2π -périodiques normalisées sont connues. Elles sont paires et impaires et traditionnellement notées $ce_n(\theta, r)$ et $se_n(\theta, r)$, les $a_n(r)$ et $b_n(r)$ associés étant appelés les valeurs *caractéristiques*.

10.4. Confluence entre \mathbb{V}_J et \mathbb{V}_S

Pour conclure cet aperçu des processus de diffusion dans nos deux modèles \mathbb{V}_J et \mathbb{V}_S , notons que l'on peut *interpoler* entre les deux modèles, ce qui correspond à une *confluence* de singularités entre les deux équations associées. Mohammed Brahim Zahaf et Dominique Manchon [572] ont construit une telle interpolation donnée par une famille de modèles \mathbb{V}^α et ont étudié la confluence des équations différentielles correspondantes dans l'espace de Fourier. Par des changements de variables appropriés, toutes ces équations peuvent être réduites à la forme

$$P_0(t) y''(t) + P_1(t) y'(t) + P_2(t) y(t) = 0$$

pour des polynômes bien choisis $P_0(t)$, $P_1(t)$ et $P_2(t)$. Les singularités de telles EDO du deuxième ordre (considérées comme des équations d'une variable complexe t) sont les zéros t_j de $P_0(t)$. Elles sont appelées singularités "régulières" lorsque $(t - t_j) \frac{P_1(t)}{P_0(t)}$ et $(t - t_j)^2 \frac{P_2(t)}{P_0(t)}$ sont analytiques dans un voisinage de t_j .

Le modèle \mathbb{V}^α peut être résumé par la table suivante :

$$\begin{aligned}
X_1^\alpha &= \cos(\theta) \partial_x + \frac{1}{\alpha} \sin(\alpha\theta) \partial_y , \\
X_2^\alpha &= \partial_\theta , \\
X_3^\alpha &= -\alpha \sin(\alpha\theta) \partial_x + \cos(\theta) \partial_y , \\
[X_1^\alpha, X_2^\alpha] &= -X_3^\alpha , \\
[X_2^\alpha, X_3^\alpha] &= \alpha^2 X_1^\alpha , \\
[X_1^\alpha, X_3^\alpha] &= 0 , \\
\mathbb{V}^\alpha &= SE_\alpha(2) \text{ avec } \mathbb{S}_\alpha^1 = \frac{\mathbb{R}}{2\pi\alpha^{-1}\mathbb{Z}} , \\
X_1^\alpha(\psi(\theta)) &= i\lambda\alpha^{-1} \sin(\alpha\theta) \psi(\theta) , \\
X_2^\alpha(\psi(\theta)) &= \psi'(\theta) , \\
\hat{\Delta}^\lambda : \psi''(\theta) - \frac{\lambda^2}{\alpha^2} \sin^2(\alpha\theta) \psi(\theta) , \\
\psi''(\theta) + \left(\mu - \frac{\lambda^2}{\alpha^2} \sin^2(\alpha\theta) \right) \psi(\theta) &= 0 .
\end{aligned}$$

Le changement de variable $\frac{\sin^2(\alpha\theta)}{\alpha^2} \rightarrow t$ donne l'équation $t(1 - \alpha^2 t) y''(t) + \frac{1}{2}(1 - 2\alpha^2 t) y'(t) + \frac{1}{4}(\mu - \lambda^2 t) y(t) = 0$,
3 singularités : 0, α^{-2} sont régulières, ∞ est irrégulière.

Pour $\alpha = 1$, \mathbb{V}^1 donne le modèle \mathbb{V}_S :

$$\begin{aligned}
X_1 &= \cos(\theta) \partial_x + \sin(\theta) \partial_y , \\
X_2 &= \partial_\theta , \\
X_3 &= -\sin(\theta) \partial_x + \cos(\theta) \partial_y , \\
[X_1, X_2] &= -X_3 , \\
[X_2, X_3] &= X_1 , \\
[X_1, X_3] &= 0 , \\
\mathbb{V}^1 &= \mathbb{V}_S = SE(2) \text{ avec } \mathbb{S}^1 = \frac{\mathbb{R}}{2\pi\mathbb{Z}} , \\
X_1(\psi(\theta)) &= i\lambda \sin(\theta) \psi(\theta) , \\
X_2(\psi(\theta)) &= \psi'(\theta) , \\
\hat{\Delta}^\lambda : \psi''(\theta) - \lambda^2 \sin^2(\theta) \psi(\theta) , \\
\psi''(\theta) + (\mu - \lambda^2) \sin^2(\theta) \psi(\theta) &= 0 .
\end{aligned}$$

Le changement de variable $\sin^2(\theta) \rightarrow t$ donne l'équation $t(1 - t) y''(t) + \frac{1}{2}(1 - 2t) y'(t) + \frac{1}{4}(\mu - \lambda^2 t) y(t) = 0$,
3 singularités : 0, 1 sont régulières, ∞ est irrégulière.

Pour $\alpha \rightarrow 0$ et θ petit (noté p), \mathbb{V}^0 donne le modèle \mathbb{V}_J :

$$\begin{aligned}
X_1^0 &= \partial_x + p\partial_y , \\
X_2^0 &= \partial_p , \\
X_3^0 &= \partial_y , \\
[X_1^0, X_2^0] &= -X_3^0 , \\
[X_2^\alpha, X_3^\alpha] &= 0 , \\
[X_1^\alpha, X_3^\alpha] &= 0 , \\
\mathbb{V}^0 &= \mathbb{V}_J \text{ avec } S_0^1 = \mathbb{R} , \\
X_1^0(y(s)) &= i\lambda sy(s) , \\
X_2^0(y(s)) &= y'(s) , \\
\hat{\Delta}^\lambda : y''(s) - \lambda^2 s^2 y(s) , \\
y''(s) + (\mu - \lambda^2) s^2 y(s) &= 0 .
\end{aligned}$$

Le changement de variable $s^2 \rightarrow t$ donne l'équation

$$ty''(t) + \frac{1}{2}y'(t) + \frac{1}{4}(\mu - \lambda^2 t)y(t) = 0 ,$$

2 singularités : 0 est régulière, $\alpha^{-2} = \infty$ est irrégulière.

11. Inpainting par diffusion sous-riemannienne : le modèle semi-discret

11.1. Un exemple d'inpainting

Disposant désormais d'une part de la géométrie des géodésiques sous-riemanniennes de $SE(2)$ et d'autre part de rudiments sur la diffusion sous-riemannienne nous pouvons expliquer plus en détail comment ces méthodes neurogéométriques peuvent conduire, dans une perspective *neuromimétique*, à des modèles d'*inpainting*, c'est-à-dire de reconstruction d'images lacunaires, particulièrement efficaces.

La figure 1, due à Jean-Paul Gauthier, part d'une image initiale qui est celle d'un œil masqué par une grille blanche et lui applique la diffusion jusqu'à ce que la grille ait disparu. En dépit d'une diffusion très importante la géométrie de l'image demeure excellente. La figure 2 montre à quel point une diffusion classique normale aurait détruit l'image.

Jean-Paul Gauthier a raffiné ces techniques en améliorant leur coût computationnel au moyen de méthodes *semi-discrètes* que nous allons brièvement esquisser. Les figures 3, 4, représentent à gauche l'image très lacunaire de son visage et, à droite, l'image reconstruite. La reconstruction est spectaculaire et, en fait, très supérieure aux capacités du système visuel. Cela est dû au fait que, dans ces algorithmes améliorés, on connaît la position des lacunes de l'image et cela permet d'itérer des diffusions sous-riemanniennes pendant un temps court et, à chaque étape, savoir quelles parties des lacunes peuvent être considérées comme remplies. Le modèle neurogéométrique conduit ainsi à des algorithmes "neuro-mimétiques" d'analyse d'images et de vision artificielle particulièrement puissants.

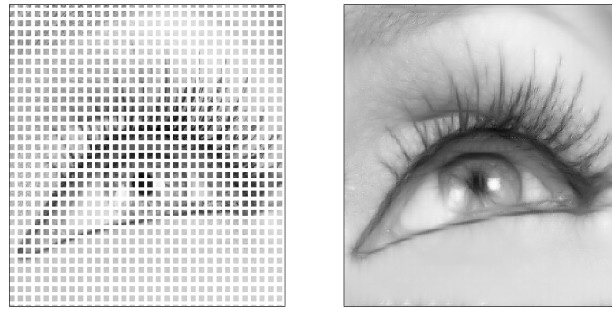


FIGURE 1. Diffusion dans \mathbb{V}_S selon Jean-Paul Gauthier. L'image initiale est celle d'un œil masqué par une grille blanche. On applique la diffusion jusqu'à ce que la grille ait disparu. En dépit d'une diffusion très importante la géométrie de l'image demeure excellente.

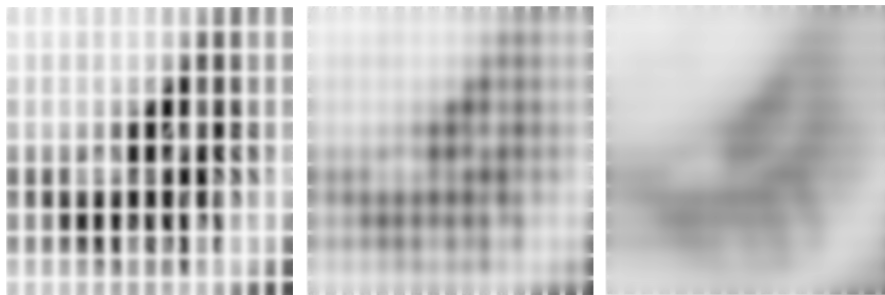


FIGURE 2. Pour faire disparaître la grille de la figure 1 avec une diffusion isotrope classique (Gaussian blurring) il faut complètement brouiller l'image.

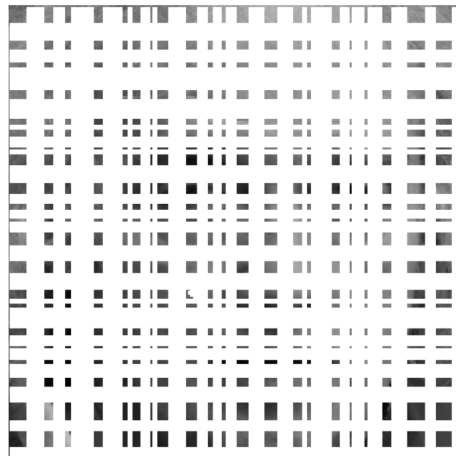


FIGURE 3. Image très lacunaire du visage de Jean-Paul Gauthier



FIGURE 4. La reconstruction par diffusion sous-riemannienne de l'image corrompue.

Nous allons pour conclure ce chapitre présenter brièvement cette version semi-discrète de notre modèle développée par Jean-Paul Gauthier avec Dario Prandi dans [455]. Une partie de leurs résultats s'appliquent à des groupes topologiques localement compacts mais nous nous restreignons, car cela est ici bien suffisant, aux groupes de Lie de dimension finie connexes.

11.2. Le modèle théorique de reconstruction

On part donc d'une image lacunaire $f_W(x, y)$ qui devrait être définie sur l'espace rétinien $R \simeq \mathbb{R}^2$ mais qui est définie en fait seulement sur un domaine W , en général très morcelé, de R . Il est essentiel que W fasse partie de la donnée initiale puisqu'il s'agit de compléter f_W en la prolongeant de W à \mathbb{R}^2 tout entier. On suppose que, même si l'image $f_W(x, y)$ comprend beaucoup de singularités (bords, coins, croisements, etc.), elle est quand même C^∞ . Cette hypothèse est légitime puisque, comme nous l'avons longuement expliqué dans le Vol I, $f_W(x, y)$ est définie, à cause de la taille, même minime, des champs récepteurs définissant une certaine échelle, par une convolution du signal avec une gaussienne.

La méthode, développée à partir de nos travaux avec Giovanna Citti et Alessandro Sarti [121], [484] et d'autres travaux comme ceux de Remco Duits et Em Erik Franken [157], consiste alors

1. à relever $f_W(x, y)$ dans le fibré $\mathbb{R}^2 \times \mathbb{S}^1$ au-dessus de W en relevant ses lignes de niveau ; on obtient alors des courbes $f_W(x, y, \theta)$ composant une surface S_W ;
2. à effectuer la diffusion sous-riemannienne hypo-elliptique permettant de compléter le domaine de définition S_W ;

3. à construire, par une certaine opération de projection sur la base R , une image $f(x, y)$ sur R qui complète $f_W(x, y)$.

L'idée est donc de compléter les lignes de niveau de f_W (qui sont en général géométriquement très compliquées) de façon géodésique. On reprend les équations du système différentiel dont les relevées legendriennes sont les solutions du système :

$$\begin{cases} \dot{x} = u \cos(\theta) \\ \dot{y} = u \sin(\theta) \\ \dot{\theta} = v \end{cases}$$

(en remarquant que l'on peut tenir compte du modèle $\mathbb{R}^2 \times \mathbb{P}^1$ en acceptant que u , qui est la norme du vecteur (\dot{x}, \dot{y}) dans le modèle $\mathbb{R}^2 \times \mathbb{S}^1$, puisse être négatif) et on minimise l'intégrale de $(u^2 + v^2)$ le long des trajectoires.

Évidemment, un problème majeur par rapport à celui des géodésiques que nous avons traité à la section 4 du chapitre 14, est que, étant donnée une ligne de niveau sortant de W , on ne sait pas à quelle autre ligne de niveau elle doit être connectée en revenant dans W . D'où la nécessité de passer à un modèle stochastique en tenant compte de toutes les possibilités de connexions et en choisissant la plus vraisemblable. On transforme donc les équations des relevées legendriennes en équations stochastiques faisant intervenir deux mouvements browniens (processus de Wiener) indépendants du_t et dv_t .

$$\begin{cases} dx_t = \cos(\theta_t) du_t \\ dy_t = \sin(\theta_t) du_t \\ d\theta_t = dv_t \end{cases}$$

Cela conduit à la diffusion sous-riemannienne sur $SE(2)$, l'équation de la chaleur avec le laplacien hypo-elliptique, et comme le note Jean-Paul Gauthier

“By the Feynman-Kac formula, integrating [the] equation with the corrupted image as the initial condition, one expects to reconstruct the most probable missing level curves (among admissible).” ([455], p. 6)

11.3. Produits semi-directs semi-discrets

Mais, comme nous l'avons vu, l'équation de la chaleur sur $SE(2)$ est très difficile à intégrer car les fonctions de Mathieu sont difficiles à implémenter et la convergence du développement en série du noyau de la chaleur qu'elles permettent est lente. Il est donc nécessaire de trouver des algorithmes numériques plus rapides permettant d'en intégrer une version approximative satisfaisante. C'est ce qu'a proposé Jean-Paul Gauthier en discrétisant $SO(2) = \mathbb{S}^1$ dans le produit semi-direct

$\mathbb{V}_S = \mathbb{R}^2 \rtimes SO(2)$.¹¹⁵ Cette simplification est empiriquement justifiée car la détection des orientations dans $V1$ est discrète et elle est algorithmiquement très efficace.

Gauthier divise donc \mathbb{S}^1 en N secteurs de $\frac{2\pi}{N}$, ce qui donne le sous groupe cyclique d'ordre N de \mathbb{S}^1 ,

$$\mathbb{S}_N^1 = \left\{ \vartheta_k = e^{i\frac{2\pi k}{N}} \right\}_{k=0, \dots, N-1} ,$$

qui est isomorphe à $\mathbb{Z}_N = \frac{\mathbb{Z}}{N\mathbb{Z}}$, et considère le groupe semi-discret qu'est le produit semi-direct

$$SE(2, N) = \mathbb{R}^2 \rtimes \mathbb{S}_N^1 = H \rtimes K$$

où $H = \mathbb{R}^2$ est un groupe de Lie abélien, K un groupe abélien fini de cardinal N agissant par automorphismes $k.x$ sur H (ici les rotations $e^{i\frac{2\pi k}{N}}$), la mesure de Haar dx de H étant K -invariante et $k \in K, k \neq 1$, agissant librement (i.e. sans point fixe) sur $H - \{0\}$.¹¹⁶ Nous allons voir que l'avantage de $SE(2, N)$ sur $SE(2)$ est d'être un *groupe de Moore* c'est-à-dire un groupe localement compact dont toutes les unirreps sont de dimension *finie* bien qu'il soit non commutatif et non compact. Cela simplifie notablement les calculs. Le laplacien peut être décomposé ("désintégré") en une intégrale d'opérateurs de type Mathieu mais de dimension *finie* et le noyau de la chaleur peut être explicité de façon pas trop compliquée.

Soient $x, y, \dots \in H$ les éléments de H et $\chi, \zeta, \dots \in \hat{H}$ les caractères de H . K agit sur les χ par $k.\chi(x) = \chi(k^{-1}.x)$. Regardons les représentations régulières de H, K et G . Elles sont unitaires. Celle de H sur $\mathcal{U}(L^2(H))$ (\mathcal{U} = groupe unitaire) donne les translations

$$\tau_x f(y) = f(y - x)$$

sur H et celle de K sur $\mathcal{U}(L^2(K)) = \mathcal{U}(\mathbb{C}^N)$ donne les shifts

$$\sigma_k \varphi(m) = \varphi(k^{-1}m)$$

sur K . Quant à celle $\Lambda(g)$ de G , elle donne

$$g.\psi(h) = \psi(g^{-1}h)$$

sur $\mathcal{U}(L^2(G))$. Mais, comme nous l'avons vu à la section 4.3.1 du chapitre 17, la structure de produit semi-direct induit aussi la représentation *quasi-régulière* de G sur $\mathcal{U}(L^2(H))$ donnée, avec la notation Λ_{\rtimes} , par

$$\Lambda_{\rtimes}(g = (x, k)) f(y) = f(k^{-1}.(y - x)) .$$

115. Nous avons vu plus haut dans la section 2.2.2 de ce chapitre la discrétisation de la base \mathbb{R}^2 conduisant aux trames de Gabor de Liontou-Marcolli.

116. La notation $H \rtimes K$ a pour but de souligner les propriétés de $SE(2, N)$ qui ne dépendent que des propriétés mentionnées de H et K .

De même que Λ , Λ_{\times} n'est pas du tout irréductible puisque, pour tout fermé K -invariant F de \widehat{H} , le sous-espace fermé de $L^2(H)$ constitué des $f(x)$ dont la transformée de Fourier $\widehat{f}(\chi)$ a son support dans F est Λ_{\times} -invariant. En prenant la transformée de Fourier sur H de Λ_{\times} , on obtient une représentation duale $\widehat{\Lambda}_{\times}$ sur \widehat{H} donnée par

$$\widehat{\Lambda}_{\times}(x, k) \widehat{f}(\chi) = \overline{\chi}(x) \widehat{f}(k^{-1} \cdot \chi) .$$

Grâce aux méthodes exposées dans le chapitre 17 (théorème de Mackey, etc.) on obtient les unirreps de G .

11.4. Résultats de finitude

Les algorithmes du modèle semi-discret de Jean-Paul Gauthier reposent sur une propriété particulière de $SE(2, N)$ qui est d'être "maximalement presque-périodique" (MAP : "maximally almost periodic"). On caractérise cette propriété en utilisant la notion de "compactification de Bohr" ¹¹⁷ $G^{\mathfrak{B}}$ de G qui est l'objet universel pour les applications continues de G dans des groupes compacts séparés. ¹¹⁸ Si G est lui-même compact, $G^{\mathfrak{B}} = G$ et si G est abélien, $G^{\mathfrak{B}}$ s'obtient en prenant le groupe dual \widehat{G} de G , en le munissant de la topologie discrète et en reprenant le dual, c'est-à-dire l'espace de *tous* les caractères, pas seulement les caractères continus, de \widehat{G} . On a un morphisme canonique (pas forcément injectif) \mathfrak{b} de G dans $G^{\mathfrak{B}}$ ($g \in G$ est identifiable au caractère $g(\chi) = \chi(g)$ de \widehat{G}) dont le dual est l'identité et qui envoie G sur un sous-groupe *dense* de $G^{\mathfrak{B}}$. On montre que \mathfrak{b} induit une *bijection* entre les représentations unitaires continues de dimension *finie* de G et de $G^{\mathfrak{B}}$. Le noyau de \mathfrak{b} est constitué des éléments g de G qui ne peuvent pas être séparés de e_G (l'élément neutre) par des représentations unitaires de dimension finie. On dit alors que G est MAP si et seulement si \mathfrak{b} est *injective*. $SE(2)$ n'est pas MAP alors que $SE(2, N)$ l'est. $SE(2, N)$ est MAP et un groupe de Moore non compact.

Des théorèmes disent que G est MAP si et seulement si ses représentations continues unitaires de dimension finie sont assez nombreuses pour *séparer les points*, ¹¹⁹ et aussi (Freudenthal-Weil) si et seulement si G (supposé localement compact et connexe) est un produit *direct* $\mathbb{R}^n \times K$ d'un \mathbb{R}^n et d'un groupe compact K . Cela exclut $SE(2)$ qui est un produit semi-direct.

Les liftings d'images $f(x, y)$ sur $W \subset R \simeq \mathbb{R}^2$ vont être alors des fonctions sur $G = SE(2, N)$. Parmi ces fonctions il y a l'espace $\mathfrak{B}(G)$ des fonctions, dites *presque-périodiques* (PP) au sens d'Harald Bohr (1927), qui sont les pull back par $\mathfrak{b} : G \rightarrow G^{\mathfrak{B}}$ des fonctions continues sur $G^{\mathfrak{B}}$ (i.e. si f_B est une fonction continue

117. Harald Bohr, le frère de Niels.

118. Rappelons qu'un espace topologique est séparé si deux points différents ont toujours deux voisinages disjoints (chapitre 2, section 4.2).

119. Nous avons déjà rencontré cette notion à la section 3.1 du chapitre 17.

$f_B : G^{\mathfrak{B}} \rightarrow \mathbb{C}$, on prend $f = f_B \circ \mathfrak{b}$.¹²⁰ Mais $SE(2, N)$ étant MAP, on peut utiliser un théorème disant que $\mathfrak{B}(G)$ est *dense* dans l'espace $C(G)$ muni de la topologie de la convergence uniforme sur les compacts et que l'on peut donc approximer les f sur G par des fonctions PP, ce qui simplifie et accélère considérablement le traitement numérique.

De façon plus précise, les fonctions PP sur $SE(2, N)$ $f(x, y, \theta)$ sont les fonctions $f(x, y)$ PP sur \mathbb{R}^2 à valeurs dans \mathbb{C}^N de composantes $f_j(x, y) = f(x, y, \vartheta_j)$, $j = 1, \dots, N$, qui sont des limites uniformes de polynômes trigonométriques en (x, y) . On peut donc les approximer par de tels polynômes

$$f_j(x, y) = \sum_{k=0}^{k=M} c_{j, \zeta_k} e^{i\langle \zeta_k, z \rangle}$$

et l'on peut en plus forcer les ζ_k à appartenir à un (grand) sous-ensemble fini F de \mathbb{R}^2 . Avec toutes ces approximations, la diffusion sous-elliptique se ramène à un très grand système d'équations différentielles ordinaires sur \mathbb{C}^N . Les auteurs ont travaillé sur un système de 1.996.080 EDOs!

Le processus stochastique sous-jacent n'est plus, comme pour $SE(2)$, de la forme $dx_t = \cos(\theta_t) du_t$, $dy_t = \sin(\theta_t) du_t$, $d\theta_t = dv_t$ mais des mouvements browniens (processus de Wiener) indépendants couplés avec un processus de Markov (en fait de Poisson) de sauts ("jump process") pour θ_t : la probabilité de k sauts entre $t = 0$ et t est $\frac{(\alpha t)^k}{k!} e^{-\alpha t}$ avec α un paramètre > 0 (cf. [70]). Il en découle immédiatement que

$$P(\theta_t = \vartheta_{j\pm 1} | \theta_0 = \vartheta_j) = \frac{1}{2} (\alpha t) e^{-\alpha t}$$

(le $\frac{1}{2}$ vient du fait qu'il y a 2 possibilités ; rappelons que $\vartheta_j = e^{i\frac{2\pi j}{N}}$) et

$$P(\theta_t = \vartheta_{j\pm 2} | \theta_0 = \vartheta_j) = \frac{1}{4} \left(\frac{1}{2} (\alpha t)^2 e^{-\alpha t} \right).$$

Pour calculer le générateur infinitésimal de ce processus de Poisson, on calcule les limites de ces probabilités pour $t \rightarrow 0$, soit

$$p_{j,k} = \lim_{t \rightarrow 0} \frac{1}{t} P(\theta_t = \vartheta_k | \theta_0 = \vartheta_j).$$

On trouve $p_{j,j\pm 1} = \frac{1}{2}\alpha$, $p_{j,j\pm m} = 0$ si $m > 1$ et $p_{j,j} = -\alpha$ (et donc $\sum_k p_{j,k} = 0$). Quand on regarde alors le laplacien hypo-elliptique de $SE(2, N)$ il est la somme (i) de l'opérateur $(\cos(\theta) \partial_x + \sin(\theta) \partial_y)^2$ sur \mathbb{R}^2 , mais avec θ contraint d'être l'un des ϑ_j , et

120. Dans le cas des $f : \mathbb{R} \rightarrow \mathbb{C}$, les f PP sont bornées et uniformément continues, forment une algèbre et ont de bonnes propriétés de leurs transformées de Fourier au sens où elles constituent la fermeture du sous-espace engendré par les $e^{i\omega x}$. Elles sont donc uniformément approximables par des polynômes trigonométriques.

(ii) de la discrétisation de l'opérateur ∂_θ^2 : pour une fonction $f(x, y, \theta)$, si l'on note $f_j(x, y) = f(x, y, \vartheta_j)$, le ∂_θ^2 devient $\alpha f_{j-1}(x, y) - 2f_j(x, y) + f_{j+1}(x, y)$.

Les auteurs de [70] et [455] donnent alors la formule *explicite* du noyau de la chaleur pour $SE(2, N)$ en insistant sur le fait qu'il est rare de disposer de telles formules pour les groupes non compacts. Nous avons vu une telle formule explicite pour le groupe de Heisenberg qui est nilpotent, mais pas pour le groupe $SE(2)$. D'où l'intérêt algorithmique de $SE(2, N)$.

Sur quelques prolongements de la neurogéométrie

Les modèles neurogéométriques dont nous avons détaillé les bases mathématiques dans ce volume ont été développés et approfondis par un certain nombre d’auteurs. Pour ne prendre qu’un exemple, Paul Bressloff et Jack Cowan ont proposé dans [79] un modèle ”sphérique” des hypercolonnes de V1 destiné à rendre compte de la sensibilité des neurones visuels à la fréquence spatiale. Nous l’avons commenté dans le Vol.I à la section 4.9.3. du chapitre 4 et l’avons comparé à un modèle dû à Daniel Bennequin ; Jérôme Ribot *et al.* et fondé sur la notion de dipôle. Dans [15] Dmitri Alekseevsky et Andrea Spiro proposent une synthèse de nos modèles et du modèle sphérique en utilisant des éléments de géométrie conforme dont ils sont spécialistes.

Mais il ya aussi de nombreux aspects de la neurogéométrie que nous ne traitons pas dans ce volume. Et nous voudrions quand même en évoquer quelques-uns avant de conclure.

Nous voudrions d’abord revenir sur deux phénomènes particulièrement frappants, celui de la “squelettisation” des formes et celui de la “vision entoptique” qui, encore plus que les contours illusoires, ont en commun le fait que la géométrie du percept se trouve *construite* à partir de données sensorielles *qui ne la contiennent pas*, même pas à l’état très lacunaire. Leur origine se situe donc ailleurs, dans la structure interne même de la machine corticale et, comme l’affirment Dirk Jancke *et al.* [271], ce sont des structures subjectives globales qui

“reveal fundamental principles of cortical processing”.

1. Le cut locus des contours 2D

1.1. La réalité neuronale du cut locus

Le premier phénomène, déjà évoqué dans l’Introduction du Vol I, est celui de la reconstruction par le système visuel du *cut locus* (de “l’axe de symétrie généralisé” ou du “squelette”) d’un contour C . Des expériences permettent de prouver la réalité neuronale de la construction de ce squelette interne dont aucune trace n’existe au niveau de l’input sensoriel (qui se réduit à la seule donnée du contour C).

Dans les figures 1 et 2 (déjà anciennes et déjà insérées dans l’*Introduction* du Vol I), dues au disciple de David Mumford Tai Sing Lee (Lee [321]), on considère la

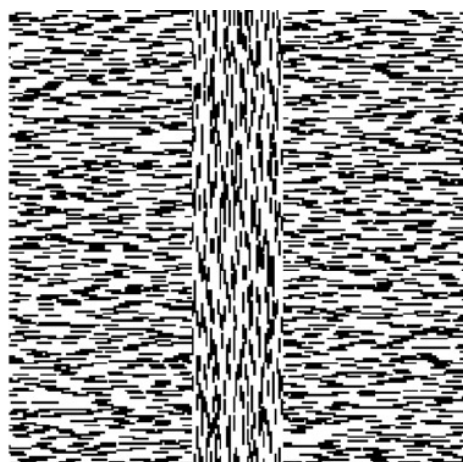


FIGURE 1. Un stimulus de l'expérience de Tai Sing Lee. (D'après Lee [321])

réponse d'une population de neurones simples de $V1$, dont l'orientation préférentielle est verticale, à des textures dont le bord est défini par une opposition d'orientations.

La figure 1 montre d'abord un stimulus de l'expérience. Il s'agit d'une bande verticale dont la texture est composée de petits segments verticaux se détachant sur un fond dont la texture est composée de petits segments horizontaux. Le cut locus de la bande (on ne tient pas compte du cadre du stimulus) est l'axe de symétrie vertical de la bande centrale.

La figure 2 montre la réponse.

(a) Au départ, entre 40 et 100ms, la réponse précoce ne concerne que l'orientation *locale* du stimulus : le neurone répond tant que son champ récepteur est dans la région homogène "verticale" du stimulus (son orientation préférentielle).

(b) Ensuite, après 100ms, la réponse concerne la structure perceptuelle *globale* du stimulus : le neurone répond au bord global de la région et au changement brutal d'orientation de la texture, autrement dit à des discontinuités qui brisent l'homogénéité de la texture.

(c) Puis, à partir de 200ms, le neurone répond à la fois au bord et au cut locus et le cut locus se trouve progressivement construit.¹

On remarquera que la réponse à l'intérieur de la zone verticale décroît progressivement alors que celle concentrée aux bords de la bande verticale centrale reste forte. Ce phénomène manifeste le fait que les bords deviennent "actifs". Les neurones détecteurs du cut locus (virtuel et *absent* du stimulus) sont progressivement activés,

1. Il en allait exactement de même, avec les mêmes doubles temps de réponse, dans les expériences sur les contours illusoire modaux à la Kanizsa exposées à la section 2.1 du chapitre 8

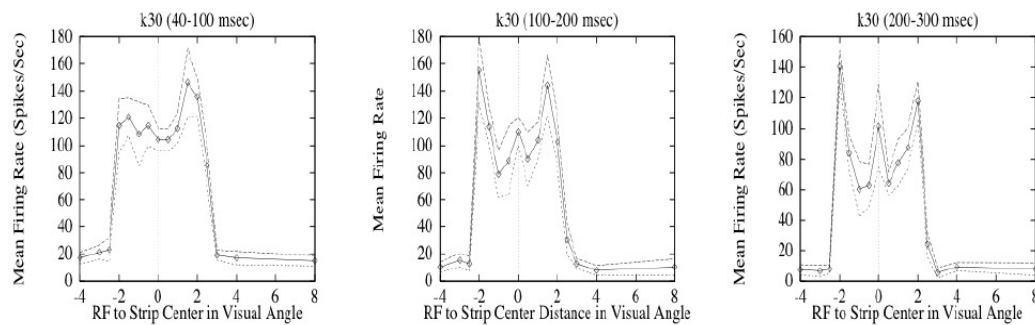


FIGURE 2. Enregistrement de la construction du cut locus. (D'après Lee [321]).

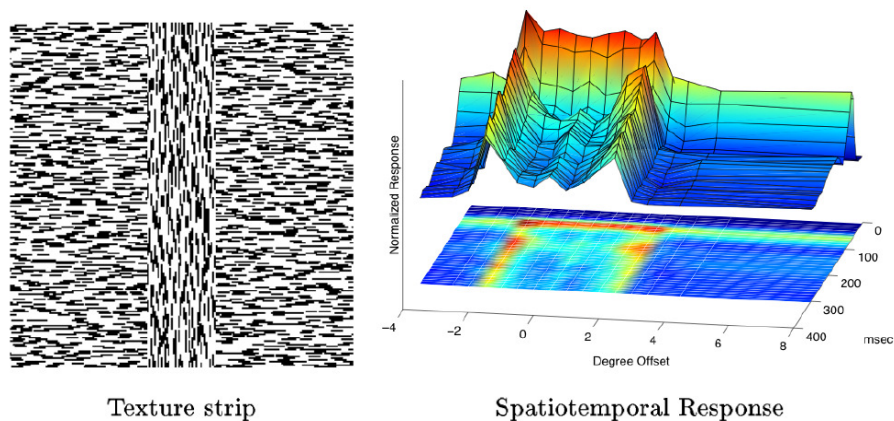


FIGURE 3. La dynamique de la construction du cut locus. (D'après Lee [321]).

certes au début moins que ceux activant les bords qui sont, eux, bien présents dans le stimulus, mais néanmoins de façon presque aussi marquée après 300ms.

Quant à la figure 3, elle montre la dynamique de la construction du cut locus.

Donnons quelques précisions sur cette expérience.

(i) Il s'agit d'enregistrements de quelques centaines de neurones de *V1* chez des singes macaques rhésus éveillés.

(ii) Pour chaque neurone on connaît son champ récepteur (position, taille) et l'orientation préférentielle qu'il détecte (ici verticale). Le "firing rate" des trains de spikes mesure l'intensité de la réponse.

(iii) Le protocole expérimental est, comme toujours dans ce genre d'expérience, assez technique. Nous ne le détaillons pas ici. Disons simplement qu'il y a un point de fixation sur l'écran et que l'on translate la forme par rapport au profil récepteur. Pour certaines positions, le profil récepteur est sur le contour (réel) ou sur le cut

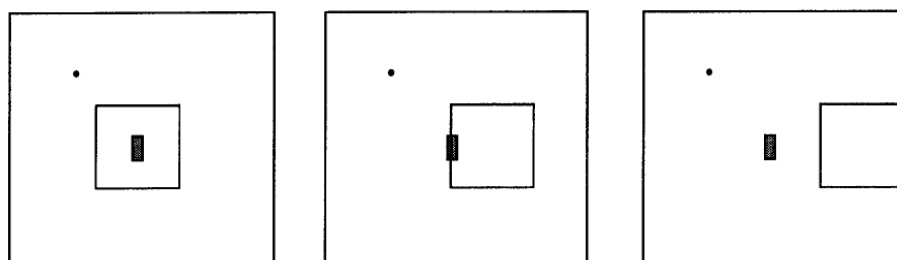


FIGURE 4. Protocole de l'expérience de T.S. Lee. Il y a un point de fixation sur l'écran et l'on translate une forme (ici un carré) par rapport au profil récepteur (la petite barre verticale épaisse). Pour certaines positions le profil récepteur est sur le contour (réel) ou sur le cut locus (virtuel) et l'on constate que la réponse est forte dans les deux cas.

locus (virtuel) et l'on constate que la réponse est forte dans les deux cas. L'exemple d'un carré est représenté figure 4.

(iv) Le fait que les textures de petites barres parallèles soient de bons stimuli pour des expériences de ségrégation de régions homogènes a été établi par Victor Lamme (un spécialiste de la segmentation des images pixelisées de scènes composées d'objets avec des relations figure/fond).²

(v) Le traitement de l'image rétinienne par les aires visuelles primaires n'est pas feedforward. Il y a de nombreux feedbacks sur $V1$ des traitements post- $V1$ ($V2$, $V3$, $V4$, MT)³. Ils sont essentiels pour le cut locus (et les contours illusoires),

“the later part of $V1$ neuron's response reflecting higher order perceptual computations”.

- (vi) Il y a deux processus dynamiques qui se reflètent par feedback dans $V1$.
- Celui de la segmentation par construction de contours globaux (discontinuités qualitatives) intégrant des détections locales.
 - Celui de la propagation des contours conduisant au lieu singulier qu'est le medial axis.

“signals sharpen spatially over time at the boundaries (...), and spread spatially over time from the boundaries (...).”

Des recherches en imagerie non invasive (fMRI) ont aussi été développées. Citons par exemple l'article de 2013 “Cortical Representation of Medial Axis Structure” [326] de Mark Lescroart et Irving Biederman.

2. Nous renvoyons à la section 4.3 du chapitre 2 pour l'opposition figure/fond et ses liens avec l'opposition fermé/ouvert en topologie.

3. Cf. les sections 5.8. à 5.10. du Vol I.

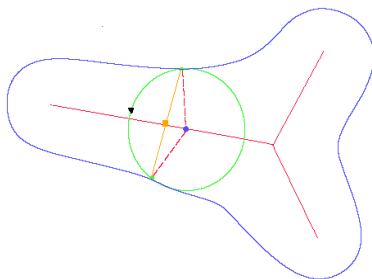


FIGURE 5. Le cut locus d'un contour C comme lieu singulier d'une propagation "optique" : c'est le lieu des centres des cercles bitangents à C . (D'après Kimia [287])

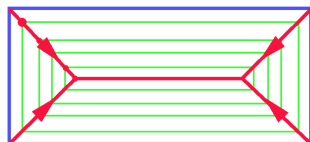


FIGURE 6. Un contour C (bleu) engendre une surface dont les lignes de gradient (rouge) sont les rayons caractéristiques de l'équation de propagation. Les fronts d'onde successifs (vert) sont les courbes de niveau d'une surface dont le cut locus est la ligne de crête. (D'après Kimia [287])

1.2. Squelettisation des formes

Précisons maintenant la géométrie mise en jeu par des expériences comme celle de Lee. Le cut locus d'un contour (fermé) régulier $C = \partial E$ bord d'une forme E (domaine connexe de \mathbb{R}^2) est le lieu singulier K de la propagation de C dans E par l'équation eikonale de l'optique géométrique (modèle de Huygens). Chaque point de C devient un centre de propagation d'ondes et C se propage ainsi parallèlement à lui-même sous la forme de fronts d'ondes dont les points évoluent le long de rayons orthogonaux à ces fronts. Les singularités de la propagation, à savoir le lieu des centres des cercles bitangents à C , définissent le cut locus comme "squelette", de la forme E (cf. figure 5, déjà vue dans notre *Introduction*).⁴

Les fronts d'onde peuvent être identifiés aux courbes de niveau d'une surface dont le cut locus constitue une sorte de "ligne de crête" (cf. figure 6).

La figure 7 donne un exemple de squelettisation d'une forme un peu plus compliquée, un cerf de l'art des steppes.

Dans ces processus "optiques" de propagation appelés aussi modèles de "grass-fire", deux sortes de singularités peuvent apparaître :

4. En fait, pour être précis, il faut dire qu'il existe plusieurs variantes des points du cut locus : celle de Blum (centres des cercles bitangents à C maximaux), celle de Brady (milieux des cordes joignant des points de bitangence), celle de Leyton [330] (milieux des arcs de C joignant des points de bitangence).

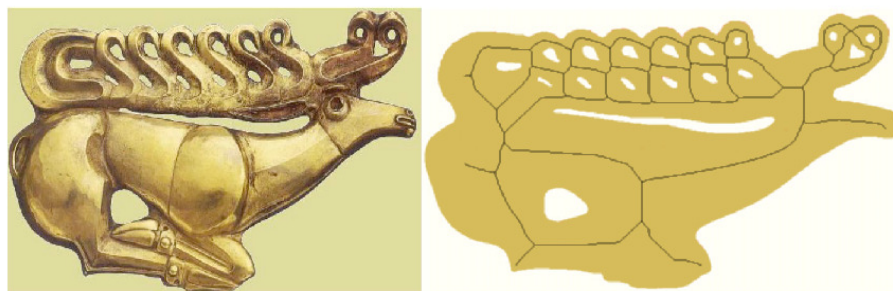


FIGURE 7. La squelettisation d'une forme un peu plus compliquée (cerf de l'art des steppes).

- (i) les caustiques, c'est-à-dire les enveloppes des rayons ;
- (ii) les cut loci, c'est-à-dire essentiellement les lieux des points qui sont atteints au même moment par deux rayons venant de deux points différents du bord $C = \partial E$ de la région E .

1.2.1. D'Harry Blum à René Thom.

Depuis les travaux pionniers du grand psychologue Harry Blum [59], le rôle fondamental du cut locus a été souligné par plusieurs grands géomètres et spécialistes de la vision comme René Thom, David Marr, David Mumford, Steve Zucker ou James Damon (voir [417] et Kimia [287]). Il est donc essentiel d'avoir pu constater qu'il est bien neurophysiologiquement réel et qu'il est bien une conséquence de l'architecture fonctionnelle des aires visuelles.

Dans son article pionnier de 1973 *Biological Shape in Visual Science* [59], Harry Blum introduisait le cut locus pour des raisons précises. Il commençait par la question :

“How do organisms describe and characterize other organisms' shapes ?”

et essayait de trouver la géométrie “biologique” implicite et interne sous-jacente aux relations entre perception visuelle et action motrice. Cette géométrie biologiquement immanente n'est pas la géométrie euclidienne du monde externe et il faut par conséquent l'explicitier car

“without a proper shape mathematics for biology, we are in the position that physics would have been in trying to develop mechanics without Euclidean geometry.”

Blum cherchait en fait à établir une relation *sensori-motrice* entre la forme des objets et le mouvement moteur de leur prise. Pour cela, il lui fallait une schématisation informationnellement très économique et compressée de la forme. Il est parti du fait de base que

“shapes are normally described by their boundaries”

et développa l'idée clé que les contours étaient “actifs” et déclenchaient un processus de propagation de contours à la Huygens qu'il appelait un “grassfire model”. La squelettisation apparaissait ainsi comme un bon candidat pour un codage très compact de la forme.

C'est ce qui a particulièrement intéressé René Thom qui avait lui-même promu depuis longtemps l'idée principielle que les singularités d'applications définies sur des formes *encodaient* la géométrie de ces formes. Dans son article “Perception et Préhension” [534],⁵ il rappelle sa découverte de Harry Blum au début des années 1970 et comment il en discuta avec la psychologue Liliane Lurçat. Il reprend la définition classique du cut locus.

“Dès que l'objet E est globalement perçu, tout se passe comme si chaque point de C se mettait à émettre des ondes circulaires vers l'intérieur de C .”⁶
(p. 165, nous utilisons nos notations)

En dehors du cut locus K un point de E (i.e. interne à C) a une seule normale à C de longueur minimale. K est le lieu des points $q \in E$ à partir desquels deux normales à C de longueur minimale sont en compétition. C'est un ensemble de Maxwell. Sur K la distance $d(q)$ de q à C est bien définie et peut avoir elle-même des points critiques. Comme le note Thom,

“le ‘codage’ de la forme E est réalisé par la donnée du cut-locus K (de ses singularités) et de la fonction $d(q)$ (et de ses points singuliers).” (p. 166)

Ce “codage” est très important parce qu'il participe aux couplages sensori-moteurs qui manifestent que la perception visuelle et la motricité partagent le *même* espace, l'apprentissage conduisant, comme dit Thom

“à une identification de l'espace tactilo-kinesthésique de la main avec l'espace visuel.” (p. 170)

Une telle identification est “miraculeuse” si l'on songe à la différence absolue des principes de construction de ces deux espaces. Leur isomorphisme est hautement non trivial.

En résumé, pour Thom :

“La théorie de Harry Blum affirme donc que la perception visuelle comporte un mécanisme automatique dans le champ visuel permettant de localiser, à l'issue d'un processus de type ‘onde de choc’, les prises ‘possibles’ de l'objet.” (p. 169)

5. Article inédit écrit en 1971-1972 et publié en 1990 dans *Apologie du Logos*.

6. La condition “Dès que l'objet E est globalement perçu” est essentielle. Elle signifie que le contour de E devient *actif*.

Ou encore,

“La forme d’un objet se définit par l’ensemble des stratégies nécessaires pour saisir cet objet” (p. 169)

et il y a une

“structuration du champ visuel par le champ moteur de la main.”

Les propositions théoriques de Blum ont été confirmées expérimentalement dès les années 70 en psychophysique par Psozka qui a présenté plusieurs contours à des sujets en leur demandant de placer de façon réflexe un point à l’intérieur “in the first place that come to mind”. Les points s’accumulaient de façon spectaculaire sur le cut locus.

1.2.2. *Après David Marr.*

Ceci dit, dans son article “Perception et préhension” [534], René Thom regrettait que

“les psychologues professionnels semblent tenir cette théorie (d’Harry Blum) en médiocre estime.” (p. 162)

Mais ce qui était vrai en 1972 n’était plus vrai dix ans après à cause

- (i) d’une profonde transformation de la psychologie traditionnelle sous l’impulsion des sciences cognitives et
- (ii) des développements rapides des algorithmes de vision artificielle.

Un repère de ce tournant a été la publication en 1982 (hélas à titre posthume) de l’ouvrage de référence *Vision* [354] de David Marr (l’un des fondateurs de la théorie computationnelle de la vision). Les travaux pionniers de Harry Blum et le cut locus y jouent un rôle fondamental.

Depuis, les recherches ont bien avancé et c’est ainsi que Benjamin Kimia a pu affirmer dans sa synthèse de 2003 “On the Role of Medial Geometry in Human Vision” [287]

“the medial axis became a central concept in mathematical morphology”
(p. 7)

1.2.3. *Cut locus et fonction distance.*

Dans son article de 1972 “Sur le cut-locus d’une variété plongée” [530] René Thom remercie Harry Blum

“dont la théorie sur la reconnaissance visuelle des formes a stimulé mon intérêt pour le cut-locus”

et précise la définition géométrique du cut locus. On regarde les points critiques dégénérés et les points de Maxwell de la fonction distance au bord

$$V_q(c) = \|q - c\|^2, \quad q \in E, \quad c \in C.^7$$

- (i) Le fait que c soit point critique (minimum) de $V_q(c)$ signifie que q est sur la normale à C issue de c .
- (ii) Le fait qu'il existe un c qui est un point critique (minimum) dégénéré de $V_q(c)$ signifie que q est le centre de courbure de C en c et qu'il est sur l'enveloppe des normales (l'évolute) de C .
- (iii) Le fait que q soit sur le cut locus K signifie que $V_q(c)$ possède deux minima absolus de même hauteur (points de Maxwell).

1.2.4. Implémentation neuronale.

Il est facile d'implémenter le cut locus dans un réseau de neurones artificiels standard dont l'activité des neurones est solution des équations de Wilson-Cowan-Hopfield. La figure 8 montre un exemple (celui du rectangle) remontant à 1991 et dû au jeune polytechnicien Hugh Bellemare. Le réseau comporte cinq couches :

1. La première couche est une "rétine" recevant l'input.
- 2-3. Les seconde et troisième couches calculent les composantes X et Y des rayons caractéristiques.
4. La quatrième couche calcule les points singuliers de la propagation.
5. Enfin la cinquième couche calcule le cut locus en utilisant comme critère géométrique les discontinuités de la divergence du champ (X, Y) .

La figure 9 montre quant à elle l'activation du réseau à partir du contour *actif* initial.

Les figures 10 et 11 montrent le cut locus interne K_{int} et externe K_{ext} de contours possédant des concavités.

1.2.5. Squelettisation et constituance mérologique.

Le cut locus K d'une forme E est une structure particulièrement intéressante et puissante pour l'analyse de sa "constituance mérologique" (sa décomposition tout/parties).

- (i) C'est un lieu *singulier* qui permet de reconstruire la forme globale E à partir de la fonction rayon (c'est-à-dire le rayon $r(q)$ du disque maximal centré en $q \in K$).
- (ii) C'est un objet *dynamique* construit à partir de la propagation des fronts, donc dans la direction des $r(q)$ croissants.

7. On prend $\|q - c\|^2$ plutôt que $\|q - c\|$ pour que $V_q(c)$ reste différentiable si $q \in C$. En effet, dans ce cas, $\|q - c\|$ a pour son minimum absolu 0 ($c = q$) qui est localement un point anguleux.

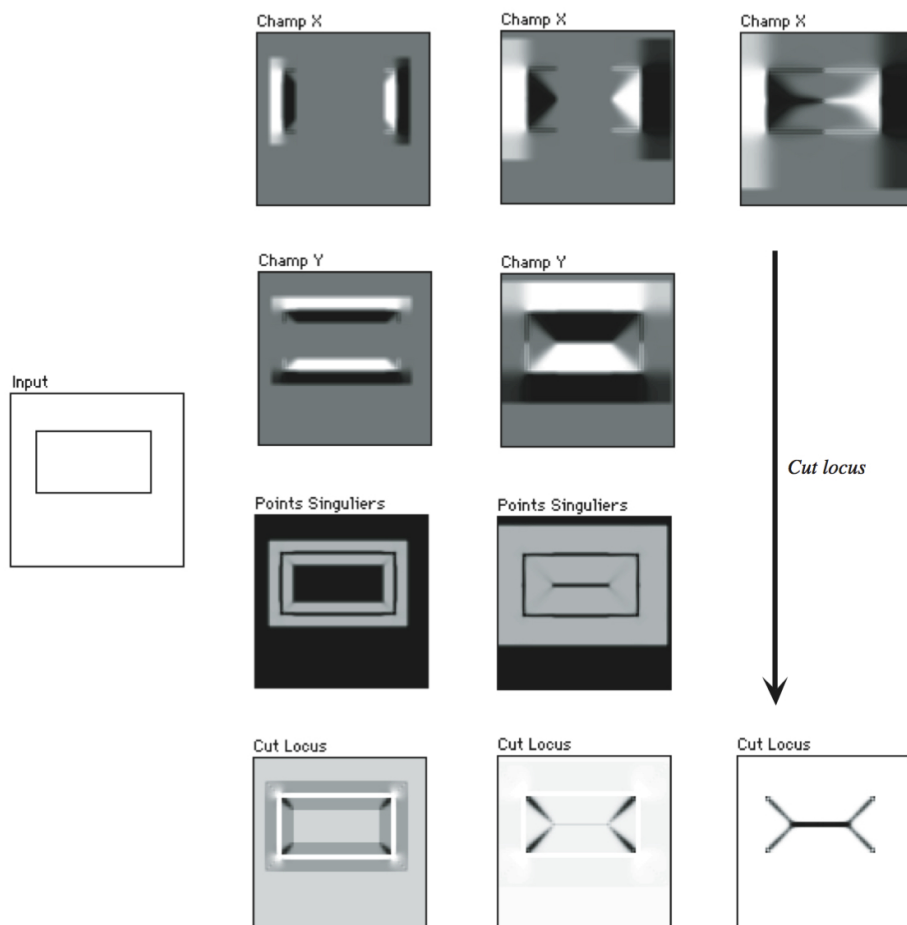


FIGURE 8. Implémentation dans un réseau de neurones de la dynamique de propagation d'un contour. La figure montre l'input E (rectangle), les composantes X, Y des lignes de gradient de la surface engendrée par le contour C de E , les singularités de la propagation et enfin le cut locus qui est l'axe de symétrie de la forme E . (Hugh Bellemare, 1991)

- (iii) Ses propriétés *topologiques* – et en particulier ses singularités : points triples et points d'arrêt – sont des indices fondamentaux des propriétés géométriques de E , comme par exemple sa convexité.

C'est pourquoi le cut locus K d'une forme E constitue une information fondamentale pour reconstruire la forme au moyen d'un processus temporel et dynamique virtuel de *morphogenèse*. Le cut locus d'un cercle métrique est un point et d'une certaine manière le cut locus mesure la non-trivialité de la forme. K est au contour C de E ce que le centre est au cercle bordant un disque. Il fournit une décomposition canonique de la forme, une sorte de système de coordonnées *intrinsèque* qui généralise

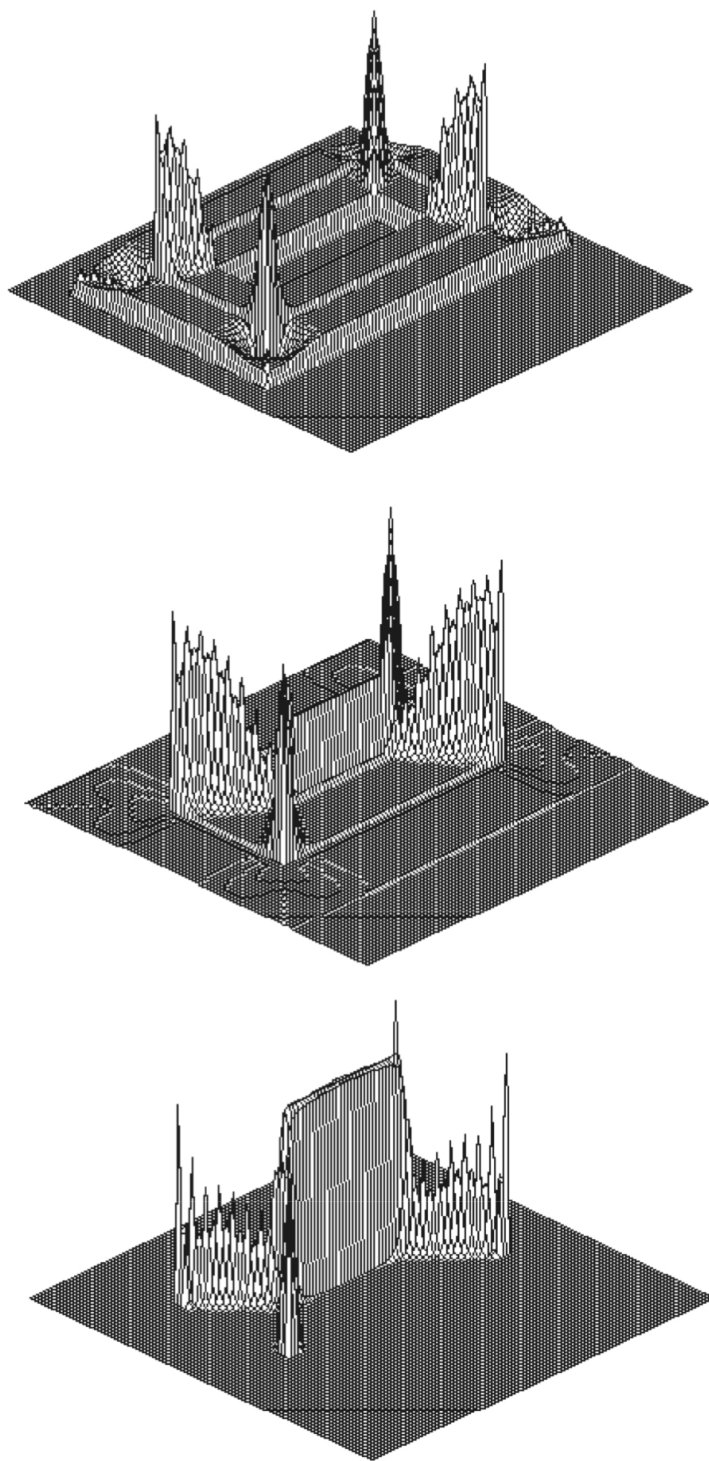


FIGURE 9. L'activité du réseau de la figure 8.

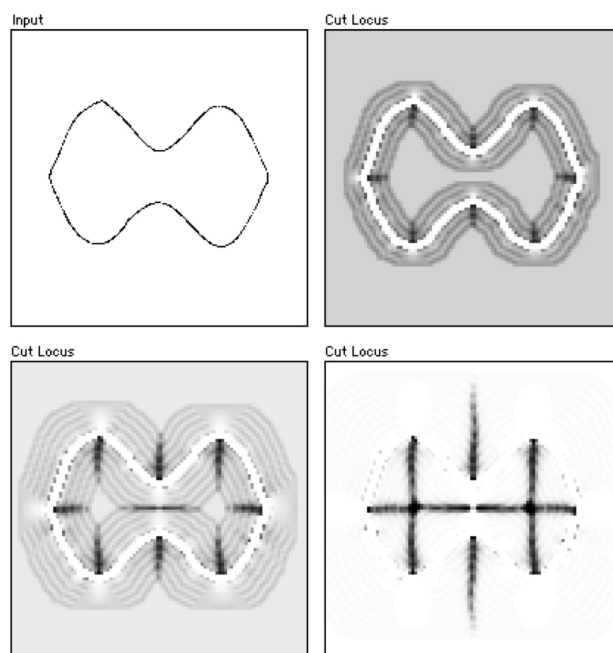


FIGURE 10. Le cut locus interne K_{int} et externe K_{ext} d'un contour possédant des concavités.

les coordonnées polaires et permet de développer une “grammaire structurale” des formes.

Cela est particulièrement important pour le cut locus *externe* K_{ext} d'une configuration A de domaines E_i : K_{ext} partitionne – catégorise, stratifie – l'espace ambiant autour des E_i en “zones d'influence” R_i associées aux E_i . En théorie morphologique⁸ il s'appelle le SKIZ (“skeleton by influence zones”) de la configuration A . Comme le montre la figure 12, la géométrie de K_{ext} encode qualitativement les positions relatives des E_i .

1.2.6. *Cut locus multi-échelle.*

L'une des principales critiques adressée à l'usage du cut locus K pour l'analyse méréologique des formes E est qu'il est très sensible au bruit : toute petite aspérité du bord engendre une nouvelle branche de K . Mais si l'on régularise la forme, on peut éliminer ce défaut et élaguer les branches parasites (“pruning”).

En fait, il est très intéressant d'utiliser une version *multi-échelle* du cut locus, par exemple en utilisant les algorithmes de “curve shortening”, de “flow by curvature” ou de “heat flow on isometric immersions” (Michael Gage, Richard Hamilton, Matthew Grayson, Stan Osher, James Sethian, Lawrence Evans, Joel Spruck). D'après un

8. Cf. les algorithmes développés par l'École dite “de Fontainebleau”, entre autres Georges Matheron et Jean Serra. Cf. [495].

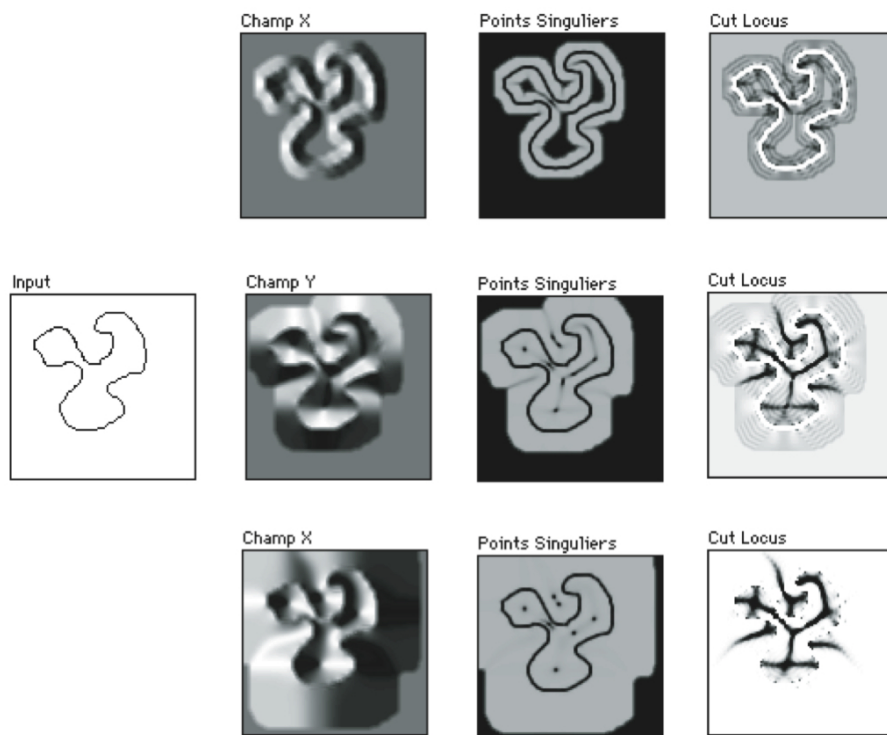


FIGURE 11. Le cut locus interne K_{int} et externe K_{ext} d'un contour un peu plus complexe.

théorème de Grayson pour la dimension 2,⁹ ces algorithmes de diffusion anisotrope convexifient le contour C en diffusant la courbure le long du bord et C converge métriquement vers un cercle dont le cut locus est trivial (réduit à un point). Cela signifie que le K de la forme E se contracte sur un point via une séquence de bifurcations (fusions successives de points d'arrêt et de points triples éliminant successivement les branches de K). Si l'on *inverse* le processus en faisant de l'anti-diffusion, on déploie progressivement un point dans le cut locus, déploiement qui explique comment la forme E peut être construite à partir de la forme triviale du cercle (cf. figure 13).

2. Patterns de vision entoptique

Nous voudrions aussi évoquer tout un ensemble de travaux fondamentaux obtenus en implémentant la neurogéométrie des architectures fonctionnelles dans des systèmes d'équations différentielles "physiques" de réseaux de neurones à la Wilson-Cowan-Hopfield. La géométrie force la physique des états d'activité à se stabiliser

9. Ce théorème est faux en dimension > 2 .

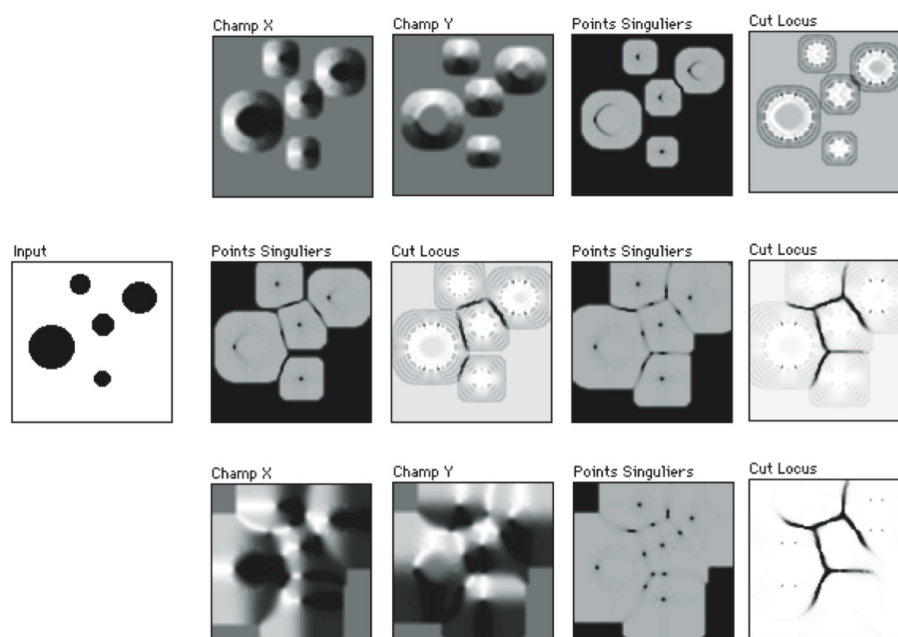


FIGURE 12. Le SKIZ (“skeleton by influence zones”) d’une configuration A de domaines E_i . Sa géométrie encode qualitativement les positions relatives des E_i .

sur des patterns bien définis qui, en quelque sorte, la traduisent. Cela est intéressant parce que, jusqu’ici, nous avons modélisé la connectivité neuronale (géométrie) et les profils récepteurs (analyse) mais nous n’avons pas véritablement étudié les *dynamiques d’activité* possibles du réseau neuronal. Dans cette section nous allons dire un mot des états d’activité stables de ces dynamiques (les percepts) ainsi que de leurs bifurcations possibles en utilisant les équations d’évolution standard des réseaux de neurones.

2.1. Les premières données de Heinrich Klüver

Nous allons appliquer cette stratégie à un exemple spectaculaire, celui des hallucinations visuelles (allant bien au-delà des contours illusoires) relevant de ce qu’on appelle la “vision entoptique”. L’exemple est particulièrement intéressant puisque, en l’absence de stimulus visuel, ces perceptions géométriques purement *endogènes* procèdent d’une activation spontanée des architectures fonctionnelles corticales. Ces modèles remarquables sont dus à Paul Bressloff, Jack Cowan et Martin Golubitsky sur la base de travaux pionniers de Bard Ermentrout et Jack Cowan et nous les avons déjà évoqués dans notre Introduction du Vol I (section 2.11.3.) et explicités

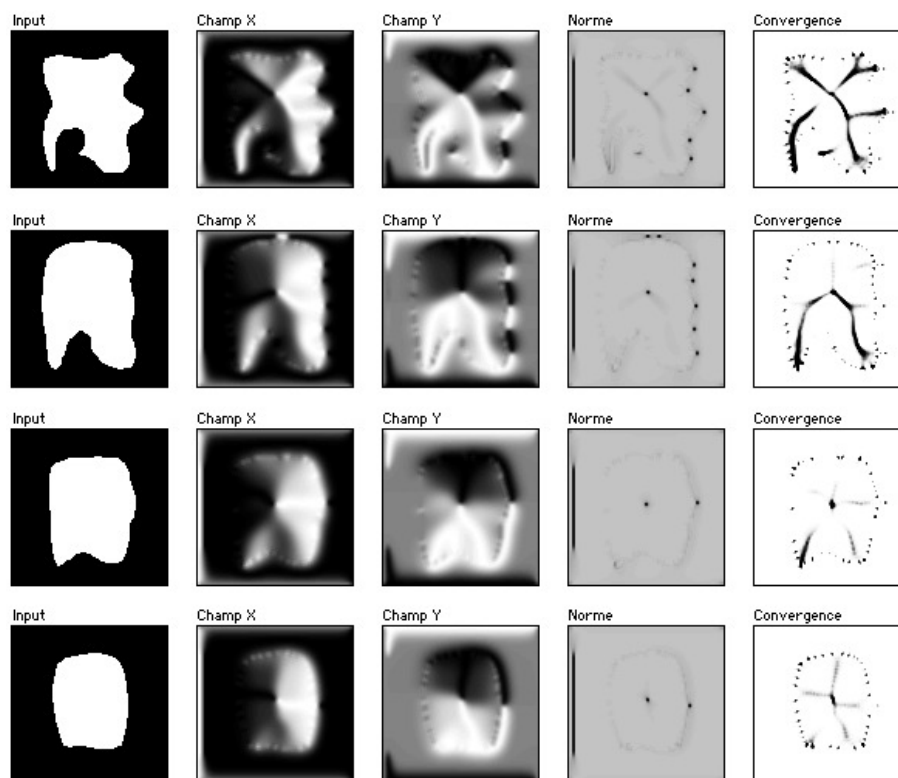


FIGURE 13. Un cut locus multi-échelle. Quatre étapes sont représentées. La convexification de la forme initiale E fait que les branches du cut locus K disparaissent progressivement dans un ordre séquentiel bien précis reflétant la hiérarchie méréologique des composants de E . Les cinq couches du réseau de neurones implémentant l’algorithme sont expliquées plus haut à la figure 8.

dans notre précédent *Neurogéométrie de la vision* de 2008 [435]. Ils montrent comment la perception peut engendrer des patterns géométriques *indépendants* de tout stimulus sensoriel.

La vision entoptique (Tyler 1978, [548]) concerne plus précisément l’émergence de patterns géométriques de phosphènes lors d’endormissements ou de réveils, après de fortes pressions sur les globes oculaires (stimulation mécanique), des stimulations électro-magnétiques (stimulations magnétiques transcrâniennes, stimulations électriques au moyen de microélectrodes), des expositions à de violents flashes intermittents, de vraies migraines, l’absorption de substances comme la mescaline, le LSD, la psilocybine, la kétamine ou des alcaloïdes comme le peyotl (stimulations neuropharmacologiques), ou des “near death experiences”, toutes expériences au cours desquelles le cortex visuel est suractivé. Ils sont liés à une excitabilité anormale des neurones et sont en compétition avec l’activité issue des inputs visuels (cf. Rauschecker [462]).

Il existe de nombreuses descriptions phénoménologiques de telles expériences (par exemple Oster, [396]). Les sujets observent de façon spontanée des patterns typiques en forme de tunnel, d'entonnoir, de spirale, de nid d'abeille, de toile d'araignée, etc. Comme l'explique Yves Frégnac [195] :

“Such visual imagery is dynamic and the illusory contours usually explode from the center of gaze to the periphery, appearing initially in black and white before bright colors take over, and eventually pulsate and rotate in time as the experience progresses.”

Certains patterns illusoires purement géométriques de ce type ont été classés il y a déjà longtemps (1928) par le grand neurophysiologiste Heinrich Klüver (1897 – 1979) qui en a fourni plusieurs descriptions cliniques. Klüver était un étudiant de Max Wertheimer et introduisit la Gestalttheorie aux États-Unis où il devint l'un des fondateurs de la neurobiologie. Ses expériences sur la mescaline, un alcaloïde hallucinogène dérivé du cactus “*Lophophora Williamsii*”, sont remarquables. Citons sa biographie académique [383] due en 1998 à Frederik Nahm et Karl Pribram

“Klüver’s interest in mescal ‘buttons’ or peyote (...) can be traced back to his earlier publications on eidetic visual phenomena, for mescal visions were thought to resemble visual eidetic imagery. (...) He experienced recurring visual forms such as those used by Miro in his painting and suggested that their existence might be of some interest to anthropologists studying visions and symbolic art of various tribes. Klüver always recognized the importance of his data for other fields, and he confidently pointed out that psychoactive compounds were an important tool in the study of visual abilities such as color and space phenomena, dreams, illusions, and hallucinations.”

Le livre de référence de Klüver est Klüver [296] (cf. aussi son papier inaugural [295]).

Il existe aujourd’hui de nombreux programmes (en particulier au National Institute of Mental Health américain) visant à tester la réactivité des neurorécepteurs à différentes classes de substances de façon à explorer le substrat chimique de l’esprit. On localise chaque substance dans un espace abstrait comportant une dimension par récepteur et on interprète les phénomènes observés comme des bifurcations d’attracteurs des dynamiques neurales. Par exemple le “Psychoactive Drug Screening Program” (NIMH-PDSP) a testé 19 substances (2C-B, 2C-B-fly, DOB, DOI, DOM, 2C-E, 2C-T-2, ALEPH-2, Mescaline, MEM, MDA, MDMA, DMT, 5-MeO-DMT, 5-MeO-MIPT, DIPT, 5-MeO-DIPT, DPT, Psilocine¹⁰) et 3 contrôles (lisuride, 6-fluoro-DMT, 4C-T-2) pour une centaine de récepteurs. Comme l’explique Thomas Ray [463] de l’Université d’Oklahoma :

10. Des enzymes réduisent la psilocybine à la psilocine qui est neuralemement active.

“The project aims to understand the mechanisms underlying the qualitative diversity of actions of psychedelics, by locating each drug in an abstract ‘receptor space’, a coordinate system with one axis for each receptor. Drugs shift the balance of activity of the brain away from the origin, by a vector representing the profile of binding affinities at different receptors. Drugs perturb the system through increasing or decreasing transmission or transmitter levels, or up or down regulating receptor populations. ”

“In a brain-centered reference frame, the origin is based on absolute levels of activity at each receptor population. The state of the brain is constantly on the move. We can think of it as a complex dynamical system, in which the trajectory follows high-dimensional orbits, and switches among many ‘attractors’. [...] In this dynamic reference frame, drugs will create a perturbation along the binding vector, thereby pushing the system into a new attractor.”

2.2. Architecture fonctionnelle et réseaux de neurones

Dans une série d’articles reprenant un premier travail de Bard Ermentrout et Jack Cowan de 1979 [170], Paul Bressloff, Jack Cowan, Martin Golubitsky, Peter Thomas et Matthew Wiener [78] ont construit en 2001, en se fondant sur l’architecture fonctionnelle de $V1$, des modèles géométriques précis des patterns visuels qui apparaissent spontanément dans certaines expériences de vision entoptique (voir aussi [77] et [79]). Leur méthode consiste

- (i) à partir des équations de Wilson-Cowan-Hopfield pour les réseaux de neurones et à passer à la limite continue pour obtenir une EDP déterminant l’activité $A(v, t)$ de $V1$ (v étant un élément de $V1$ et t le temps) ;
- (ii) à exprimer dans les poids synaptiques de l’EDP l’architecture fonctionnelle de $V1$ (l’organisation en pinwheels et le système des connexions cortico-corticales horizontales qui implémentent la structure de contact du plan rétinien R) ;
- (iii) à tenir compte du fait que cette architecture fonctionnelle est invariante sous l’action du groupe $E(2)$ des déplacements du plan rétinien R ;
- (iv) à montrer que la solution homogène initiale $A(v, t) \equiv 0$ devient *instable* et *bifurque* spontanément lorsque le paramètre d’excitabilité μ de $V1$ dépasse une certaine valeur critique μ_c ;
- (v) à linéariser l’EDP au voisinage de la bifurcation ;
- (vi) à calculer les modes propres (“eigenforms” ou “planforms”) en faisant l’hypothèse qu’il s’agit de solutions doublement périodiques sur R et en tirant les conséquences du fait que les bifurcations sont dominées par la brisure de la $E(2)$ -symétrie ;
- (vii) à faire une analyse de stabilité pour pouvoir sélectionner les eigenforms stables ;

(viii) à engendrer des patterns visuels en appliquant à ces dernières l'application réciproque de la projection rétinotopique $\chi : R \rightarrow V1$ (qui est une application conforme de type logarithme complexe).

Les calculs sont explicites et l'accord avec les données expérimentales extrêmement frappant. Il s'agit de la première application neurocomputationnelle des techniques de théorie des bifurcations par brisure de symétrie, techniques qui sont universellement présentes en physique et en chimie.

Les auteurs travaillent (comme nous l'avons fait dans nos modèles) dans la fibration

$$\pi : \mathbb{V} = R \times \mathbb{S}^1 \simeq \mathbb{R}^2 \times \mathbb{S}^1 \rightarrow R \simeq \mathbb{R}^2$$

avec des coordonnées $v = (a, \theta)$ ($a = (x, y)$). Les "neurones" (il s'agit en fait de petites populations "méso" de vrais neurones) paramétrés par les éléments (a, θ) de \mathbb{V} ont une activité $A(a, \theta, t)$ et sont connectés par des connexions de poids synaptiques $W \langle a, \theta | a', \theta' \rangle$. L'état d'activité nulle $A \equiv 0$ décrit l'état de base de l'activité cérébrale et, comme en physique, ce "ground state" peut être très compliqué. Ce que décrit l'activité A est donc en fait la déviation relativement à l'état de base, c'est-à-dire les états excités. Comme le souligne Franz Vollenweider (University Hospital of Psychiatry, Zurich) :

"There is converging evidence from brain imaging, behavioural and electrophysiological studies indicating that the psychedelic effects of these drugs arise, at least in part, from their common capacity to disrupt fronto-striato-thalamic gating or filtering of external and internal information flow to the cortex."

Comme en physique, ces modèles comprennent deux parties : une infrastructure géométrique et un champ d'activation gouverné par une EDP.

Les auteurs partent de l'EDP de Wilson-Cowan-Hopfield gouvernant l'évolution de $A(a, \theta, t)$:

$$\frac{\partial A(a, \theta, t)}{\partial t} = -\alpha A(a, \theta, t) + \frac{\mu}{\pi} \int_0^\pi \int_{\mathbb{R}^2} W \langle a, \theta | a', \theta' \rangle \sigma(A(a', \theta', t)) da' d\theta' \quad (1) \\ + H(a, \theta, t)$$

où $W \langle a, \theta | a', \theta' \rangle$ est le poids de la connexion entre le neurone $v = (a, \theta)$ et le neurone $v' = (a', \theta')$, α un paramètre de decay, μ un paramètre d'excitabilité de $V1$ qui exprime la façon dont une substance agit sur l'activité cérébrale et $H(a, \theta, t)$ l'input rétinien. La croissance de μ traduit une augmentation de l'excitabilité de $V1$ due à l'action de la substance sur les noyaux (locus coeruleus, raphé) qui produisent des neurotransmetteurs comme la sérotonine ou la noradrénaline.¹¹

11. Il existe d'autres mécanismes pouvant expliquer l'induction de phosphènes, par exemple des mécanismes inhibant le courant d'hyperpolarisation I_h dans les cellules rétinienues (cf. entre autres Cervetto *et al.* [101]).

Le point clé du travail des auteurs est que *l'architecture fonctionnelle* (simplifiée) de V1 peut être *encodée* dans les poids synaptiques $W \langle a, \theta | a', \theta' \rangle$.

1. Les connexions “verticales” locales à l’intérieur d’une hypercolonne fournissent un terme :

$$W \langle a, \theta | a', \theta' \rangle = W_{loc}(\theta - \theta') \delta(a - a')$$

où $W_{loc}(\varphi)$ est une fonction paire donnée (caractéristique du système) ne dépendant que de l’orientation et où le Dirac $\delta(a - a')$ force $a = a'$.

2. Les connexions “horizontales” latérales entre différentes hypercolonnes fournissent un terme :

$$W \langle a, \theta | a', \theta' \rangle = W_{lat}(a - a', \theta) \delta(\theta - \theta')$$

où le Dirac $\delta(\theta - \theta')$ force $\theta = \theta'$ et exprime le fait que ces connexions cortico-corticales connectent des paires (a, θ) et (a', θ') avec une orientation commune $\theta = \theta'$.

3. Qui plus est, la co-axialité $\theta = \theta' =$ “orientation du segment aa' ” s’exprime par le fait que

$$W_{lat}(a - a', \theta) = W_{lat}(s) \delta(a - a' - se_\theta) = \widehat{W}(r_{-\theta}(a - a'))$$

où e_θ est le vecteur unité dans la direction θ et où $W_{lat}(s)$ est une fonction donnée (caractéristique du système) ne dépendant que de la distance. On notera $\widehat{W}(r_{-\theta}(a - a'))$ la fonction $W_{lat}(s) \delta(a - a' - se_\theta)$.

4. D’où l’expression générale des poids synaptiques :

$$W \langle a, \theta | a', \theta' \rangle = W_{loc}(\theta - \theta') \delta(a - a') + \beta \widehat{W}(r_{-\theta}(a - a')) \delta(\theta - \theta') \quad (2)$$

où β est un coefficient donné (caractéristique du système) mesurant la force relative des connexions verticales et horizontales.

5. L’EDP est donc (pour un input $H(a, \varphi, t) = 0$) :

$$\begin{aligned} \frac{\partial A(a, \theta, t)}{\partial t} = & -\alpha A(a, \theta, t) \\ & + \mu \left[\int_0^\pi W_{loc}(\theta - \theta') \sigma(A(a, \theta', t)) \frac{d\theta'}{\pi} \right] \\ & + \mu \left[\beta \int_{\mathbb{R}^2} W_{lat}(a - a', \theta) \sigma(A(a', \theta, t)) da' \right]. \end{aligned} \quad (3)$$

On voit donc que les auteurs utilisent la même architecture fonctionnelle que celle de nos modèles *géométriques* mais en l’implémentant explicitement dans les poids synaptiques de connexions neuronales *physiques*.

Un modèle plus réaliste consiste à introduire un peu de diffusion et à remplacer les Dirac par des gaussiennes ou des laplaciens de gaussiennes (qui confinent mieux

que les gaussiennes) mais nous nous en tiendrons ici aux Dirac qui formalisent bien l'idéalisation qui intéresse les auteurs.

Le point essentiel pour les auteurs est que l'architecture fonctionnelle de $V1$ (les poids synaptiques) est $E(2)$ -invariante et fournit une représentation de $E(2)$ plus riche que la représentation standard sur la base \mathbb{R}^2 . Cette action "shift-twist" du groupe euclidien $E(2)$ sur $\mathbb{V} = \mathbb{R}^2 \times \mathbb{S}^1$ est, répétons-le, donnée pour les translations τ de \mathbb{R}^2 , les rotations r_ψ d'angle ψ et la réflexion κ , par

$$\begin{cases} \tau(a, \theta) &= (a + \tau, \theta) \\ \psi(a, \theta) &= (r_\psi a, \theta + \psi) \\ \kappa(a, \theta) &= (\kappa a, -\theta) \end{cases}$$

et, comme y insistent les auteurs (Bressloff *et al.* [78], p. 306),

"The particular form of the action of rotations (...) reflects a crucial feature of the lateral connections, namely that they tend to be oriented along the direction of their cell's preference."

Nous retrouvons donc le parallélisme et la co-axialité caractéristiques de la structure de contact de $V1$.

Si $A(a, \theta)$ est une fonction $A : \mathbb{R}^2 \times \mathbb{S}^1 \rightarrow \mathbb{R}$ et $\gamma \in E(2)$, la transformée de A par γ est par définition

$$\gamma A(a, \theta) = A(\gamma^{-1}(a, \theta))$$

et, pour les poids synaptiques,

$$\gamma W \langle a, \theta | a', \theta' \rangle = W \langle \gamma^{-1}(a, \theta) | \gamma^{-1}(a', \theta') \rangle .$$

Il est trivial de vérifier que, eu égard à leur forme très spécifique (2), les poids synaptiques sont $E(2)$ -invariants.

1. L'invariance est triviale pour les translations $\tau \in R$ puisque W ne dépend que de $a - a'$ et que $(a + \tau) - (a' + \tau) = a - a'$.
2. Pour les rotations on a :

$$\begin{aligned} & W \langle r_{-\psi} a, \theta - \psi | r_{-\psi} a', \theta' - \psi \rangle \\ &= W_{loc} ((\theta - \psi) - (\theta' - \psi)) \delta(r_{-\psi} a - r_{-\psi} a') \\ &\quad + \beta \widehat{W}(r_{-(\theta - \psi)}(r_{-\psi} a - r_{-\psi} a')) \delta((\theta - \psi) - (\theta' - \psi)) \\ &= W_{loc} (\theta - \theta') \delta(a - a') + \beta \widehat{W}(r_{-\theta}(a - a')) \delta(\theta - \theta') \\ &= W \langle a, \theta | a', \theta' \rangle . \end{aligned}$$

3. Pour la réflexion κ on a :

$$\begin{aligned} W \langle \kappa a, -\theta | \kappa a', -\theta' \rangle &= W_{loc}(-\theta + \theta') \delta(\kappa a - \kappa a') + \beta \widehat{W}(r_\varphi(\kappa a - \kappa a')) \delta(-\theta + \theta') \\ &= W_{loc}(\theta - \theta') \delta(a - a') + \beta \widehat{W}(r_{-\varphi}(a - a')) \delta(\theta - \theta') \\ &= W \langle a, \theta | a', \theta' \rangle \end{aligned}$$

car W_{loc} est pair, $r_\psi \kappa = \kappa r_{-\psi}$ et $\widehat{W}(\kappa a) = \widehat{W}(a)$.

Par conséquent, l'EDP (1) $\frac{\partial A(a, \theta, t)}{\partial t} = F(A(a, \theta, t))$ (on suppose que l'input H est nul) est $E(2)$ -équivariante au sens où $\gamma F(A) = F(\gamma A)$. En effet, soit $\gamma \in E(2)$. γ opère sur l'EDP (1) par

$$\begin{aligned} \frac{\partial A(\gamma^{-1}(a, \theta), t)}{\partial t} &= -\alpha A(\gamma^{-1}(a, \theta), t) \\ &\quad + \frac{\mu}{\pi} \int_0^\pi \int_{\mathbb{R}^2} W \langle \gamma^{-1}(a, \theta) | a', \theta' \rangle \sigma(A(a', \theta', t)) da' d\theta' . \end{aligned}$$

Soit (a'', θ'') le point défini par $(a', \theta') = \gamma^{-1}(a'', \theta'')$, c'est-à-dire $(a'', \theta'') = \gamma(a', \theta')$. Comme W est γ -invariant

$$W \langle \gamma^{-1}(a, \theta) | a', \theta' \rangle = W \langle a, \theta | \gamma(a', \theta') \rangle = W \langle a, \theta | a'', \theta'' \rangle$$

et $da' d\theta' = da'' d\theta''$, on obtient

$$\begin{aligned} \frac{\partial A(\gamma^{-1}(a, \theta), t)}{\partial t} &= -\alpha A(\gamma^{-1}(a, \theta), t) + \\ &\quad \frac{\mu}{\pi} \int_0^\pi \int_{\mathbb{R}^2} W \langle \gamma^{-1}(a, \theta) | \gamma^{-1}(a'', \theta'') \rangle \sigma(A(\gamma^{-1}(a'', \theta''), t)) d(\gamma^{-1}(a'', \theta'')) . \end{aligned}$$

Mais

$$\frac{\partial A(\gamma^{-1}(a, \theta), t)}{\partial t} = \gamma \frac{\partial A(a, \theta, t)}{\partial t} = \gamma F(A)$$

et d'autre part

$$\begin{aligned} &-\alpha A(\gamma^{-1}(a, \theta), t) \\ &+ \frac{\mu}{\pi} \int_0^\pi \int_{\mathbb{R}^2} W \langle \gamma^{-1}(a, \theta) | \gamma^{-1}(a'', \theta'') \rangle \sigma(A(\gamma^{-1}(a'', \theta''), t)) d(\gamma^{-1}(a'', \theta'')) \\ &= F(\gamma A) \end{aligned}$$

d'où l'équivariance.

Comme y insistent les auteurs (Bressloff *et al.* [78], p. 307),

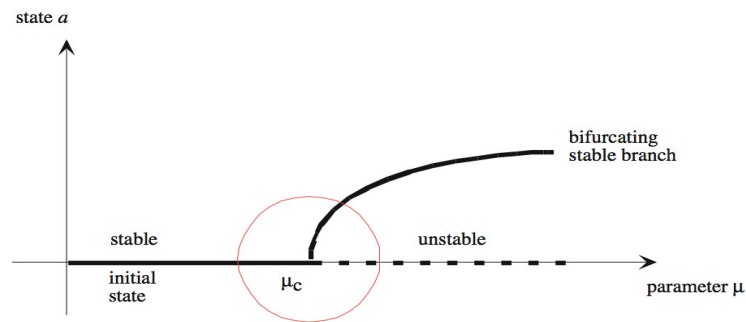


FIGURE 14. Lorsque le paramètre μ augmente et traverse la valeur critique μ_c , l'état de base homogène $A \equiv 0$ jusque-là stable (trait gras) devient instable (traits pointillés) et bifurque vers un autre état stable.

“The equivariance of the operator F with respect to the action of $E(2)$ has major implications for the nature of solutions bifurcating from the homogeneous resting state.”

En particulier, comme nous allons le voir, elle implique que lors d'une bifurcation, sont sélectionnées des représentations scalaires ou pseudo-scalaires de $E(2)$.

2.3. Analyse spectrale et bifurcations

On peut alors expliquer au moyen de l'analyse spectrale de l'EDP (1) pourquoi et comment des patterns géométriques peuvent émerger *spontanément* dans de tels réseaux neuronaux.

On suppose qu'il n'existe pas d'input externe ($H = 0$). Pour $\mu = 0$, l'état de base $A \equiv 0$ est trivialement l'état d'équilibre du réseau et il est stable (on ne voit rien). Rappelons que $A \equiv 0$ est un “ground state” qui peut être très compliqué (activité endogène de $V1$, bruit spontané, etc.) et que le modèle décrit en fait l'écart à cet état de base. L'analyse spectrale de l'EDP (1) montre que l'état de base initial $A \equiv 0$ peut devenir instable et *bifurquer* pour des valeurs critiques μ_c du paramètre μ (cf. figure 14).

Il faut insister sur le fait que lorsqu'un état d'activation bifurque, il décroche de l'input qui l'avait causé initialement et se met à refléter les propriétés structurales de l'EDP qui le pilote et comme, par construction, celle-ci est équivariante sous l'action du groupe $E(2)$ encodé dans l'architecture fonctionnelle, on comprend pourquoi les nouveaux états d'activation issus des bifurcations peuvent traduire en fait la structure de l'architecture fonctionnelle.

L'analyse spectrale de l'EDP et l'analyse des bifurcations sont assez techniques. Elles utilisent les méthodes classiques déjà évoquées :

1. Linéarisation de l'EDP au voisinage de la solution $A \equiv 0$ et de la valeur critique μ_c .
2. Analyse spectrale de l'équation linéarisée.
3. Calcul de ses vecteurs propres (eigenmodes).
4. Après avoir linéarisé l'EDP autour de la solution triviale $A \equiv 0$, on regarde des solutions de la forme :

$$A(a, \theta, t) = e^{\lambda t} A(a, \theta), \quad \lambda \in \mathbb{R} \quad (4)$$

où une exponentielle temporelle $e^{\lambda t}$ module un pattern spatial $A(a, \theta)$. Cela est justifié par le fait que, quand $\mu = 0$, les solutions de l'EDP linéaire

$$\frac{\partial A(a, \theta, t)}{\partial t} = -\alpha A(a, \theta, t)$$

sont

$$A(a, \theta, t) = e^{-\alpha t} A(a, \theta) .$$

5. Les solutions $A(a, \theta, t) = e^{\lambda t} A(a, \theta)$ ne peuvent être stationnaires que si $\lambda = 0$. Sinon, elles décroissent ($\lambda < 0$) ou divergent ($\lambda > 0$) exponentiellement. Dans ce dernier cas, la saturation imposée par la fonction de gain non linéaire σ les stabilise.
6. En substituant la solution (4) dans l'EDP (1), on obtient une équation pour les valeurs propres λ qui est de la forme :

$$\lambda A(a, \theta) = -\alpha A(a, \theta) + \sigma'(0)\mu \left(\begin{array}{l} \int_0^\pi W_{loc}(\theta - \theta') A(a, \theta') \frac{d\theta'}{\pi} \\ + \beta \int_{\mathbb{R}^2} W_{lat}(a - a', \theta) A(a', \theta) da' \end{array} \right) \quad (5)$$

où β est une constante mesurant la force relative des connexions verticales et horizontales.

7. En utilisant les séries de Fourier de A , W_{loc} et W_{lat} pour la variable périodique θ et la transformée de Fourier pour la bivariable spatiale a et en identifiant les coefficients des termes de même type des deux côtés de l'équation, on obtient des *relations de dispersion* de la forme $\lambda = -\alpha + \mu F(\dots)$ où F est une fonction (compliquée) des coefficients de Fourier. Ce qui fournit les équations $\lambda = 0$.
8. Pour $\mu = 0$, $\lambda = -\alpha$. Mais quand μ croît, λ s'annulera une première fois pour une certaine fonction F et une certaine valeur critique μ_c de μ .
9. La bifurcation activera alors les termes correspondants dans les séries et transformées de Fourier, ce qui donnera les eigenmodes.
10. Comme c'est la valeur propre $\lambda = 0$ qui correspond aux bifurcations de l'état de base, on travaille par conséquent dans *le noyau du linéarisé*.

11. Le point clé est que les symétries du problème imposent des contraintes extrêmement fortes. La symétrie de translation impliquera que les fonctions propres du linéarisé devront avoir en facteur des ondes planes $e^{ik \cdot a}$. De même, la symétrie de rotation impliquera que si un vecteur d'onde k devient critique à la bifurcation, alors il en ira de même de tous les vecteurs d'onde de même module, et pour la même valeur propre, l'espace propre correspondant étant par conséquent de dimension infinie. Il s'agit de la partie la plus technique de la déduction. Il est important de souligner que si les eigenmodes sont bien mathématiquement déduits de l'équation (1) ils jouissent néanmoins d'une certaine autonomie car ils dépendent de façon essentielle de brisures de symétrie *universelles*. En ce sens, une dérivabilité mathématique n'a rien à voir avec un réductionnisme physicaliste. Nous reviendrons sur ce point épistémologique crucial à la section 1 de notre Conclusion.¹²
12. Comme $E(2)$ opère sur le fibré \mathbb{V} et pas simplement sur la base \mathbb{R}^2 , il n'y a pas que la représentation classique, dite scalaire, sur \mathbb{R}^2 et sur les fonctions de \mathbb{R}^2 qui interviendra, mais aussi l'action dite "pseudo-scalaire". Il s'agit d'un phénomène vraiment remarquable car, comme le soulignent avec force les auteurs (Bressloff *et al.* [77], p. 5) :

"The visual cortex models (...) provide the first example of pseudo-scalar bifurcations that are relevant in applications."

Précisons un peu les choses. On suppose donc que la partie spatiale de la solution est une onde plane de vecteur d'onde $k = (\rho \cos(\psi), \rho \sin(\psi))$ modulée par un facteur de phase $u(\theta - \psi)$ ayant un angle ψ comme origine (dans la formule c.c. signifie "complexe conjugué") :

$$A(a, \theta) = u(\theta - \psi)e^{ik \cdot a} + c.c.$$

Les contraintes imposées aux fonctions de phase $u(\theta)$ se traduisent par des relations de dispersion sur les coefficients des séries de Fourier (les coefficients de celles des $u(\theta)$ sont inconnus, ceux de celles des $W_{loc}(\theta - \theta')$ et $W_{lat}(a - a', \theta)$ sont connus).

Les auteurs introduisent en plus une contrainte de *discrétisation* en travaillant sur des réseaux \mathcal{L} de la couche corticale et en cherchant des solutions doublement périodiques par rapport à ces réseaux. Puis ils tiennent compte de l'application rétinotopique qui projette la rétine sur la couche corticale de V1. À travers cette application, les droites dans la couche corticale correspondent à des spirales au niveau de la rétine, les verticales donnant des cercles et les horizontales donnant des rayons. Les réseaux \mathcal{L} n'ont donc rien à faire avec des réseaux dans la base \mathbb{R}^2

12. La situation est tout à fait analogue à celle que l'on rencontre en physique statistique où les comportements critiques des systèmes peuvent être classés, au moyen du groupe de renormalisation, en classes universelles indépendantes de la structure physique fine spécifique du substrat.

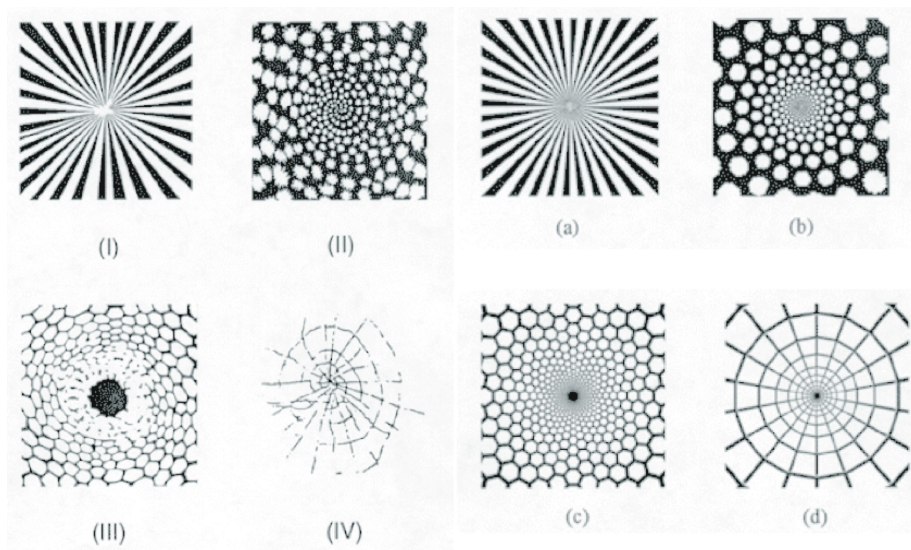


FIGURE 15. À gauche (*I, II, III, IV*) les hallucinations visuelles décrites par Klüver. À droite (*a, b, c, d*), leurs modèles neurogéométriques d'après Bressloff, Cowan et Golubitsky.

et leurs directions ne sont pas des directions dans le champ visuel. Les translations verticales font tourner l'image rétinienne associée et les translations horizontales la font changer d'échelle par une homothétie radiale. Les auteurs considèrent trois réseaux, carré, hexagonal et rhombique.

Les solutions contraintes par cette double périodicité sont appelées *planforms*. Elles sont bien connues car, comme le soulignent les auteurs,

“There is a common approach to all lattice bifurcation problems.”

Les calculs par rupture de symétrie sont néanmoins compliqués. Ils débouchent sur des formes explicites des $u(\theta)$ et montrent comment s'effectuent les bifurcations. Ils utilisent de façon essentielle l'“equivariant branching lemma” de Golubitsky et conduisent à des résultats étonnants. La figure 15 montre, à gauche (*I, II, III, IV*) les hallucinations visuelles décrites par Klüver ainsi que, à droite (*a, b, c, d*), leurs modèles neurogéométriques.

Quant à la table, elle montre des images rétiniennes virtuelles plus précises de *planforms*

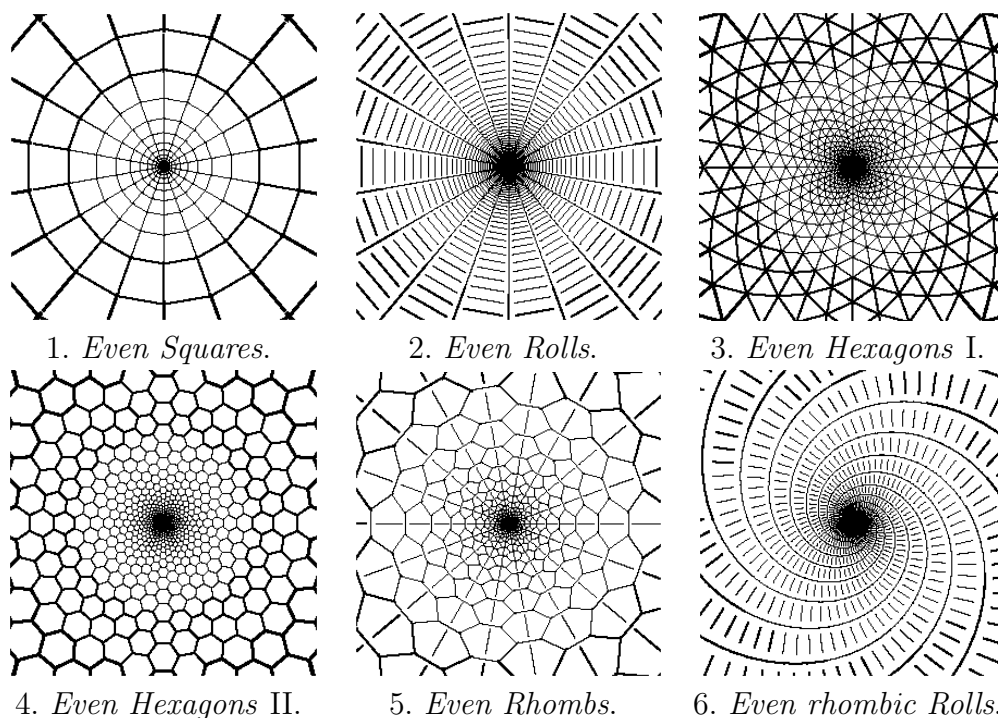


Table. Images rétiniennes virtuelles de planforms paires
(d'après Bressloff, Cowan, Golubitsky *et al.* [78]).

Remarque. Dans un article fort intéressant Olivier Faugeras et Pascal Chossat ont repris le concept de “planform” dans un contexte de géométrie *hyperbolique*. Ils enrichissent les modèles de la façon dont les hypercolonnes de V1 extraient les dérivées (premières) de l'intensité du signal $I(x, y)$ en introduisant un “tenseur de structure”. Ils expliquent ensuite que les tenseurs de structure vivent dans un espace dont la géométrie naturelle est hyperbolique. Les groupes et les sous-groupes d'isométries hyperboliques interviennent donc naturellement et ils reprennent dans ce nouveau cadre la théorie des bifurcations engendrant les planforms par rupture de symétrie. \square

3. Stéréopsie, kinesthésie, motricité

3.1. La neurogéométrie de la stéréopsie

Une des fonctions essentielles de la détection des contours $2D$ est la reconstruction des objets $3D$ à partir de leurs esquisses $2D$ et en particulier de leurs contours apparents. Nous lui avons consacré de nombreuses études et nous l'avons mise en relation avec la théorie des esquisses (*Abschattungen*) de Husserl. Le problème est difficile car il consiste à résoudre le *problème inverse* de celui de la projection d'un

objet $3D$ sur un plan $2D$ et, pour le résoudre, il faut mobiliser de nombreux outils de théorie des singularités. Le lecteur trouvera des précisions dans nos études [420], [423], [429], [433].

Ce passage $2D \rightarrow 3D$ est intimement lié à la *stéréopsie*. Or nos modèles de $V1$ sont des modèles de vision monoculaire. Pourtant nous avons signalé à la section 2.1 du chapitre 8 que les contours illusoires pouvaient être rendus plus nets en utilisant la stéréopsie (cf. figure 9). On peut même montrer que, pour un triangle de Kanizsa, des contours illusoires courbes légèrement différents dans les deux images vont, lorsqu'on fusionne celles-ci, apparaître comme des contours illusoires parfaitement droits d'un triangle $3D$ convexe ou concave.

Dans le Vol I section 4.10.3. du chapitre 4 nous avons donné quelques indications sur la dominance oculaire, la rivalité binoculaire et la stéréopsie.

Pour la vision *binoculaire* le modèle neurogéométrique des relevées legendriennes des courbes planes doit être notablement complexifié. Dans un article fort intéressant de 2023 [60], “Good continuation in $3D$: the neurogeometry of stereo vision” [60], Maria Virginia Bolelli, Giovanna Citti, Alessandro Sarti et Steven Zucker ont fait le lien entre la structure binoculaire de la stéréopsie et les relevées legendriennes des courbes gauches $\gamma(s)$ de \mathbb{R}^3 dans le fibré de dimension cinq $\mathbb{R}^3 \times \mathbb{S}^2$ dont nous avons brièvement parlé à la section 8 du chapitre 5.¹³ Il s'agit de comprendre comment deux modèles monoculaires \mathbb{V}_S permettent de reconstruire ces courbes gauches de \mathbb{R}^3 et leurs trièdres de Frenet comme courbes intégrales d'une distribution \mathcal{K} de sous-espaces de dimension 3 de $\mathbb{R}^3 \times \mathbb{S}^2$ qui satisfont la condition de Hörmander.

Le formalisme se développe élégamment une fois qu'on a bien mis en place la construction et les variables idoines. Les auteurs travaillent dans \mathbb{R}^3 avec des coordonnées fixes (x, y, z) et les vecteurs unitaires tangents aux courbes

$$\gamma(s) = (\gamma_x(s), \gamma_y(s), \gamma_z(s))$$

(paramétrées par leur longueur d'arc) sont des vecteurs de \mathbb{S}^2 paramétrés par deux angles (θ, φ) :

$$\dot{\gamma}(s) = (\dot{\gamma}_x(s), \dot{\gamma}_y(s), \dot{\gamma}_z(s)) = (\cos(\theta) \sin(\varphi), \sin(\theta) \sin(\varphi), \cos(\varphi))$$

On suppose que les rayons lumineux viennent d'en haut, l'axe des z étant orienté vers le bas. On simplifie les modèles monoculaires en se donnant deux centres optiques $c_L = (-c, 0, 0)$ et $c_R = (c, 0, 0)$ et, en dessous, à une distance focale verticale de valeur $z = f$ deux petits disques rétiniens horizontaux \mathbb{D}_L et \mathbb{D}_R centrés sur les points $O_L = (-c, 0, f)$ et $O_R = (c, 0, f)$ et de coordonnées respectives (x_L, y_L) et (x_R, y_R) . Si $\gamma(s)$ est une courbe gauche on la projette à travers les centres optiques c_L et c_R sur les rétines \mathbb{D}_L et \mathbb{D}_R et l'on obtient deux courbes γ_L et γ_R (on suppose évidemment que ces projections ne débordent pas des rétines). Si $P = (x, y, z)$ est

13. On suppose que s est la longueur d'arc et donc que les vecteurs tangents sont unitaires.

un point de γ qui se projette en P_L et P_R les coordonnées sont

$$x_L = \frac{f(x+c)}{z}, \quad y_L = \frac{fy}{z}, \quad x_R = \frac{f(x-c)}{z}, \quad y_R = \frac{fy}{z} = y_L.$$

On considère alors les relevées legendriennes de γ_L et γ_R dans leurs fibrés respectifs $\mathbb{D}_L \times \mathbb{S}_L^1$ et $\mathbb{D}_R \times \mathbb{S}_R^1$.

En terme de composantes, la correspondance entre la situation 3D et la situation binoculaire $2D_{L/R}$ est donc

$$\begin{aligned} \gamma_L &= \left(\frac{f(\gamma_x + c)}{\gamma_z}, \frac{f\gamma_y}{\gamma_z} \right), \quad \gamma_R = \left(\frac{f(\gamma_x - c)}{\gamma_z}, \frac{f\gamma_y}{\gamma_z} \right), \\ \dot{\gamma}_L &= \left(\frac{f(\gamma_z \dot{\gamma}_x - (\gamma_x + c) \dot{\gamma}_z)}{\gamma_z^2}, \frac{f(\gamma_z \dot{\gamma}_y - \gamma_y \dot{\gamma}_z)}{\gamma_z^2} \right), \\ \dot{\gamma}_R &= \left(\frac{f(\gamma_z \dot{\gamma}_x - (\gamma_x - c) \dot{\gamma}_z)}{\gamma_z^2}, \frac{f(\gamma_z \dot{\gamma}_y - \gamma_y \dot{\gamma}_z)}{\gamma_z^2} \right), \\ \tan(\theta_L) &= \frac{\gamma_z \dot{\gamma}_y - \gamma_y \dot{\gamma}_z}{\gamma_z \dot{\gamma}_x - (\gamma_x + c) \dot{\gamma}_z} = \frac{\gamma_z \sin(\theta) \sin(\varphi) - \gamma_y \cos(\varphi)}{\gamma_z \cos(\theta) \sin(\varphi) - (\gamma_x + c) \cos(\varphi)}, \\ \tan(\theta_R) &= \frac{\gamma_z \dot{\gamma}_y - \gamma_y \dot{\gamma}_z}{\gamma_z \dot{\gamma}_x - (\gamma_x - c) \dot{\gamma}_z} = \frac{\gamma_z \sin(\theta) \sin(\varphi) - \gamma_y \cos(\varphi)}{\gamma_z \cos(\theta) \sin(\varphi) - (\gamma_x - c) \cos(\varphi)}. \end{aligned}$$

et les auteurs explicitent en détail comment on peut inverser ces relations et reconstruire γ et son trièdre de Frenet à partir des couples monoculaires (γ_L, θ_L) et (γ_R, θ_R) .

3.2. Le contrôle kinesthésique du regard

Les modèles traités dans cet ouvrage sont restreints à $V1$ et $V2$ alors qu'il faudrait prendre en compte les liens de $V1 - V2$ avec les autres aires visuelles et avec le système de contrôle kinesthésique du regard (globes oculaires, tête, corps). Là aussi il existe des architectures fonctionnelles fascinantes à modéliser ainsi que des *changements de coordonnées* neuralemement implémentés. L'espace visuel est visuo-moteur.

Pour en tenir compte, et pour montrer à quel point tout reste encore à faire en neurogéométrie, il suffit de se référer un instant à l'ouvrage d'Alain Berthoz [53] *Le sens du mouvement* que nous avons évoqué dans la section 5 de notre Introduction. Répétons qu'Alain Berthoz est l'un des neurophysiologistes qui s'est le plus intéressé aux résonances remarquables entre les neurosciences cognitives contemporaines et les réflexions philosophiques de Husserl, Poincaré et Merleau-Ponty. Nous renvoyons de nouveau le lecteur intéressé à son ouvrage [55] rédigé avec Jean-Luc Petit : *Phénoménologie et physiologie de l'action*.

Il faut insister sur le fait que Husserl a été l'un des premiers philosophes à avoir mis en avant trois thèses fondamentales¹⁴ :

1. Contrairement à ce qu'impose l'évidence de l'attitude naturelle, la perception n'est pas la réception passive de structures objectives externes mais bien au contraire une construction active. L'objectivité du monde externe (espace 3D approximativement euclidien, objets 3D pouvant se déplacer par rotations et translations, etc.) est une objectivité construite (constituée) par des opérations mentales (noèses) très multiples et très complexes, très différenciées et très intégrées.
2. La perception ne peut pas exister sans la kinesthésie.
3. Un caractère essentiel de la perception est son caractère proactif (“protentionnel”) et anticipateur permettant à la conscience de boucler ses perceptions sur ses actions.

De telles thèses sont en accord profond avec les conceptions actuelles et en particulier les théories dites “motrices” de la perception.

Dès le début du *Sens du mouvement*, Alain Berthoz insiste sur le fait que le cerveau est une machine proactive qui a pour une de ses fonctions principales d'anticiper les conséquences de ses actions à partir de modèles internes du monde et du corps (p. 7). Cette anticipation prédictive à partir de modèles internes constamment comparés à la situation réelle actuelle a été sélectionnée par l'évolution car elle est vitale pour la survie des proies et des prédateurs. Et Alain Berthoz enchaîne sur le fait que le sens du mouvement est la kinesthésie

“qui résulte de la coopération de plusieurs capteurs et exige que le cerveau reconstruise le mouvement du corps et de l'environnement de façon cohérente.”
(p. 11)

Alain Berthoz fait lui-même, de façon récurrente, le lien entre son approche proactive et d'autres approches :

- (i) Husserl et Merleau-Ponty ;

14. Depuis très longtemps (les années 1970) nous avons consacré de longs développements à la modélisation géométrique et dynamique des descriptions eidétiques de la perception chez Husserl. La naturalisation de ces descriptions au moyen des outils morphodynamiques élaborés par René Thom est depuis toujours un axe majeur de nos recherches et de nos enseignements. Un premier texte synthétique est notre intervention [415] au Colloque de Cerisy que nous avons organisé en 1982 en hommage à René Thom. Une synthèse plus complète est “Morphological eidetics for Phenomenology of Perception” [429] dans l'ouvrage *Naturalizing Phenomenology* que nous avons édité en 1999 avec Jean-Michel Roy, Francisco Varela et Bernard Pachoud. Sur la formalisation par la théorie des topoi de la théorie husserlienne des jugements perceptifs dans *Erfahrung und Urteil*, on pourra se référer à [438].

- (ii) Poincaré pour qui localiser un objet consiste à se représenter les mouvements pour l'atteindre (ce qui posait pour Poincaré un problème théorique car la kinesthésie n'est pas d'essence géométrique) ;
- (iii) Gibson (le fondateur de "l'écologisme " perceptif) pour qui
 "la transduction des grandeurs physiques (lumière, son, pression, etc.) répond à des questions préexistantes que le système nerveux pose au monde," (p. 50)
 cela étant dû au fait que les espèces ne s'intéressent qu'à certains aspects, pour elles pertinents, de la réalité, aspects que Gibson appelle des "affordances" (p. 67).
- (iv) Nous pourrions y ajouter ce que nous avons vu dans la section 1 à propos du cut locus d'un contour chez Harry Blum et René Thom.

La différence fondamentale entre phénoménologie et neurosciences est que la première s'arrête à la description eidétique des évidences perceptives alors que les secondes cherchent l'explication neuronale causale sous-jacente. Le cas est spectaculaire pour les kinesthéses oculaires dont nous avons parlé en Ouverture section 3.2.

Rappelons d'abord les trajets des deux voies rétino-géniculo-corticales essentielles, respectivement magnocellulaire et parvocellulaire, déjà bien séparées dans les couches du corps genouillé latéral (*CGL*). La voie magnocellulaire arrive à l'aire *MT - V5* (médio-temporale) traitant spécifiquement le mouvement rétinien et le transmettant à l'aire *MST* (médio-temporale supérieure) qui le combine avec les signaux extra-rétiens (mouvements des yeux et signaux vestibulaires) alors que la voie parvocellulaire arrive aux aires visuelles *V2, V3, V4* traitant spécifiquement la forme et la couleur.

Le rapport au corps propre passe quant à lui par la proprioception. Les fibres musculaires sont doublées par des fuseaux neuromusculaires qui sont des capteurs sensoriels mesurant leur allongement. Il existe également d'autres récepteurs qui mesurent la vitesse d'élongation. Les fuseaux neuromusculaires sont eux-mêmes contrôlés par des motoneurons qui sont à leur tour contrôlés par l'intention motrice, ce qui permet au cerveau de simuler le mouvement (p. 32-35).

Il existe aussi sur les tendons des récepteurs de Golgi qui mesurent les forces (c'est-à-dire des accélérations) et même des variations de forces (des "secousses"). Le cerveau est ainsi à même de combiner des variables d'élongation, de vitesse d'étirement et de force en des variables qu'Alain Berthoz appelle "composites" en se référant au roboticien Jean-Jacques Slotine (p. 37). Le calcul de ces variables composites mixant des grandeurs à plusieurs de leurs dérivées permet d'actualiser des synergies musculaires qui présupposent des mécanismes neuronaux d'intégration (au sens mathématique) permettant de passer des secousses aux accélérations-forces, des accélérations aux vitesses et des vitesses aux positions (p. 54).

Le contrôle kinesthésique des saccades oculaires est particulièrement fascinant. Pour en saisir toute la complexité, il faut d'abord rappeler brièvement quelques données neurophysiologiques supplémentaires.

Quand on fixe un objet, le système vestibulaire stabilise l'image rétinienne par des réflexes qui "relient les capteurs vestibulaires aux muscles de l'œil" (p. 51),¹⁵ éliminent le glissement rétinien et permettent donc aux objets de rester fixes pendant un déplacement de la tête en produisant des mouvements des yeux exactement inverses à ceux de la tête (p. 52).

Les mouvements de la tête sont codés par les capteurs inertiels de l'oreille interne dans un référentiel euclidien $3D$ fixe. Au contraire, la rétine code les images et leur glissement dans un repère $2D$ mobile (p. 68). Le passage de l'un à l'autre est donc tout à fait non trivial. En ce qui concerne le premier codage, celui effectué par "la centrale gravito-inertielle" du système vestibulaire, il comprend les canaux semi-circulaires qui détectent les accélérations angulaires et les otolithes qui détectent les accélérations linéaires et les inclinaisons de la tête (accélération de la pesanteur) (p. 39). Le groupe des déplacements de \mathbb{R}^3 qui est le produit semi-direct du groupe des rotations par le groupe des translations se trouve ainsi implémenté, mais d'une façon "motrice" totalement *sui generis* et non classique.¹⁶

Dans la poursuite oculaire d'un objet-cible en mouvement, le traitement des mouvements effectué dans les aires MT et MTS est envoyé dans le champ oculomoteur frontal (COF) du cortex frontal à partir duquel les signaux de commande motrice de poursuite de la cible sont envoyés vers les noyaux du pont commandant la contraction des muscles des yeux, les noyaux dorsolatéral ($NDLP$) et dorsomédian ($NDMP$) (p. 71).

Évolutivement beaucoup plus ancien que la poursuite oculaire, le réflexe optocinétique permet quant à lui de suivre les mouvements de l'environnement à travers le "système optique accessoire" où les neurones répondent de façon préférentielle à des mouvements visuels dans les plans des canaux semi-circulaires :

- (i) les noyaux du tractus optique (NTO) pour l'horizontal ;
- (ii) le noyau terminal médian (NTM) pour l'un des plans verticaux ;
- (iii) le noyau terminal dorsal (NTD) pour l'autre plan vertical.

Quant aux saccades oculaires et à leurs différents contrôles, Alain Berthoz décrit avec précision la cascade hiérarchique de contrôles inhibiteurs permettant cette étonnante performance (p. 209-228).

15. Le système vestibulaire assure également la stabilité de l'équilibre du corps et corrige de façon réflexe les déséquilibres posturaux (p. 51).

16. Comme la cochlée, les canaux semi-circulaires sont remplis d'endolymphe qui, par inertie lors d'un mouvement, agissent sur des cils sensoriels qui eux-mêmes agissent sur les nerfs vestibulaires (p. 40). Les otolithes sont des cristaux flottant dans l'endolymphe et dont les déplacements lors d'accélérations linéaires activent les nerfs vestibulaires (p. 46).

1. Une saccade est produite par des motoneurones oculaires (Mn). Son extrême rapidité (20ms, vitesse angulaire jusqu'à $800^\circ/s$) est due à leur décharge phasique forte.
2. Elle est déclenchée avec une anticipation de 10-15ms par des neurones prémoteurs appelés neurones excitateurs phasiques (NEP) (ségrégués suivant qu'il s'agit de mouvements horizontaux ou verticaux).
3. Les NEP se projettent aussi sur des interneurones inhibiteurs (situés dans les noyaux vestibulaires) qui bloquent le réflexe vestibulo-oculaire pendant la saccade (ce qui assure l'invariance perceptive).
4. Pour maintenir la position de l'œil dans l'orbite, il faut aussi une décharge tonique. Il s'agit d'un processus d'intégration se projetant lui aussi sur les motoneurones Mn .
5. Ce circuit neuronal prémoteur est contrôlé par un premier système inhibiteur situé dans la partie médiane de la formation réticulée. Des neurones inhibiteurs "pauseurs" (P) y inhibent les NEP et empêchent la saccade. Les intentions motrices top-down ne peuvent donc s'actualiser que lorsque cette inhibition est levée, l'activation possédant le statut d'une inhibition d'inhibition.
6. L'inhibition (P) est elle-même inhibée à partir de deux voies : celle du colliculus supérieur (CS) et celle du champ oculomoteur frontal (COF) dans le cortex préfrontal. Le colliculus supérieur est une structure fondamentale de la coopération visuo-motrice (et auditive). Il est relié à de très nombreuses aires cérébrales. Ses couches superficielles comprennent des cartes sensorielles (rétinotopique, somatotopique et audiotopique) et ses couches intermédiaires et profondes des cartes motrices (p. 89). C'est en grande partie le CS qui explique que la topie spatiale (l'isomorphisme entre la distribution spatiale des capteurs périphériques et l'organisation spatiale des neurones de l'aire considérée) soit en fait multisensorielle. Le CS reçoit directement (sans passer par le CGL et les aires visuelles) une carte rétinienne et ses neurones tecto-réticulo-spinaux ($NTRS$) se projettent sur le tronc cérébral et la moelle, de façon inhibitrice sur (P) (qui inhibe les NEP) et aussi directement (de façon excitatrice) sur les NEP et les neurones réticulo-spinaux (NRS) contrôlant les couplages entre cou, tête et yeux. Cela constitue un second système inhibiteur.
7. Mais le CS est lui-même, au niveau de ses couches intermédiaires, contrôlé de façon inhibitrice par les ganglions de la base ($SNpr$: substance noire). L'ensemble de cette structure de contrôle possède une architecture fonctionnelle précise. Elle est "fibrée" (au sens mathématique) au-dessus de l'espace des directions de l'œil (qui est un domaine du plan projectif \mathbb{P}^2). C'est ce que les neurophysiologistes appellent la sélection spatiale. Les neurones désinhibés

dans le *CS* correspondent à une direction précise du regard qui doit être atteinte par la saccade.

8. La *SNpr* est elle-même sous le contrôle inhibiteur du noyau caudé du Striatum (*Str*). C'est le troisième niveau inhibiteur du contrôle. Il comprend lui-même deux voies : une voie directe $Str \rightarrow SNpr$ et une voie passant par le Globus pallidus *Gpe* et le noyau sub-thalamique *NST*.
9. Il existe également un autre mécanisme inhibiteur régulateur des saccades et dépendant lui du cervelet

“qui joue un rôle essentiel dans la régulation des commandes motrices qui modulent l'amplitude, la vitesse et la durée des saccades” (p. 219).

10. Il existe enfin le contrôle cortical des saccades. En plus de la voie directe rétine $\rightarrow CS$, la voie visuelle standard rétino-géniculo-corticale intervient de façon top-down sur les saccades. Cela est nécessaire à ce que celles-ci réalisent une exploration de l'espace cognitivement motivée. Le *CGL* est déjà sensible aux saccades. Autrement dit,

“la commande motrice modifie l'entrée sensorielle [et] l'action modifie la perception à sa source” (p. 221).

11. La voie $V1 \rightarrow V3$ aboutit dans le cortex pariétal de la voie dorsale (*LIPL*, *LIPV*, *7a*, *7b*) qui est essentiel

“pour la représentation de l'espace et les relations entre le corps et le monde extérieur” (p. 223).

1. Il reconstitue par exemple la position d'un objet-cible dans l'espace en additionnant les positions relatives objet-œil, œil-tête et tête-corps.
12. Les centres du cortex pariétal se projettent dans le cortex pré-frontal (*CPF*) sur le champ oculomoteur supplémentaire (*COS*) et sur le *COF* qui assure la coordination des séquences des saccades et se projette lui-même sur le *CS* et le tronc cérébral.
13. Le *CPF* participe aux aspects les plus cognitifs des saccades : organisation en séquences, saccades avec délai, stratégies d'exploration, mémorisation de saccades, anticipation, prédiction, etc.
14. Citons enfin le rôle, pour l'anticipation et la prédiction, de l'aire motrice supplémentaire (*AMS*) située dans la partie médiane du lobe frontal en avant des aires motrices. Elle participe à des actions globales, à des synergies motrices, des séquencements de gestes et des mouvements endogènes produits sans stimulus externe.

3.3. La neurogéométrie de la motricité

Du côté de la motricité proprement dite, par exemple atteindre une cible avec la main, ou pointer du doigt une cible donnée, ou encore tracer manuellement une courbe (c'est tout le problème de l'écriture) de très nombreux résultats expérimentaux sont disponibles.¹⁷ Par exemple, du côté psychophysique beaucoup de mesures quantitatives détaillées ont été effectuées sur la dynamique du tracé des courbes planes $(x(s), y(s))$ (s = longueur d'arc). La loi temporelle du mouvement est $s(t)$. Depuis longtemps (fin XIX^e siècle) on a noté des relations entre la courbure $C(s)$ et la vitesse tangentielle $V(s) = \frac{ds}{dt}$. Elle manifeste la façon dont le système moteur interagit avec les forces pilotant la dynamique. La plus connue des lois découvertes est la loi de Viviani dite "des deux tiers" dont la forme la plus simple (pour une certaine classe de courbes tracées par des adultes sans pathologies) est

$$V(\theta) = kC(s)^{\frac{2}{3}}$$

où $V(\theta) = C(s)V(s)$ est la vitesse angulaire $\frac{d\theta}{dt}$ et k une constante.¹⁸ On a relié ce genre de relation de puissance entre vitesse et courbure à des modèles variationnels minimisant un coût lié au carré du taux de changement de l'accélération (le "jerk").¹⁹

Et du côté des neurosciences, on a découvert et étudié, depuis, entre autres, les travaux d'Apostolos Georgopoulos (cf. [210]) dans les années 1980, une organisation *columnaire* du cortex moteur primaire. La modélisation neurogéométrique en est commencée.

Citons par exemple le travail [359] et [360] de Caterina Mazzetti, Alessandro Sarti et Giovanna Citti sur la façon dont le bras peut atteindre une cible. Il s'agit d'un modèle sous-riemannien de l'aire $M1$. Pour le tracé manuscrit de courbes planes, les variables sont le temps t , la position (x, y) du crayon tenu par la main, l'orientation du mouvement θ , la vitesse v et l'accélération a du mouvement le long de l'orientation. Les auteurs considèrent alors les formes de Pfaff associées

$$\begin{cases} \varpi_\theta = -\sin(\theta) dx + \cos(\theta) dy = 0 \text{ pour } \theta \\ \varpi_v = \cos(\theta) dx + \sin(\theta) dy = -v dt = 0 \text{ pour } v \\ \varpi_a = dv - a dt = 0 \text{ pour } a. \end{cases}$$

La structure sous-riemannienne correspondante est définie par une distribution \mathfrak{D} de dimension 3 dans un espace M de dimension 6 de base (x, y, t) et de fibre (θ, v, a) .

17. Nous avons très brièvement évoqué la motricité d'un tracé dans la section 7 du chapitre 10 à propos de la modélisation des contours illusoire par des géodésiques "legendriennes".

18. $C(s) = \frac{d\theta}{ds}$ et donc $V(\theta) = \frac{d\theta}{dt} = \frac{d\theta}{ds} \frac{ds}{dt} = C(s)V(s)$.

19. Pour un exemple des recherches sur ces problèmes de multi-géométrie de la motricité, cf. l'article [44] de Daniel Bennequin, Ronit Fuchs, Alain Berthoz et Tamar Flash.

\mathfrak{D} est engendrée par les 3 champs de vecteurs tangents linéairement indépendants

$$\begin{cases} X_1 = v \cos(\theta) \partial_x + v \sin(\theta) \partial_y + a \partial_v + \partial_t \\ X_2 = \partial_\theta \\ X_3 = \partial_a \end{cases}$$

qui satisfont la condition de Hörmander puisque

$$\begin{cases} [X_1, X_2] = v \sin(\theta) \partial_x - v \cos(\theta) \partial_y = X_4 \\ [X_3, X_1] = \partial_v = X_5 \\ [X_5, X_1] = \cos(\theta) \partial_x + \sin(\theta) \partial_y = X_6 \end{cases}$$

les 6 champs X_i étant linéairement indépendants et engendrant tout le fibré tangent TM . On peut alors tester l'hypothèse de mouvements géodésiques. Le lecteur intéressé trouvera dans les articles de nombreuses données expérimentales et une bibliographie détaillée.

4. Extensions nécessaires des modèles

Nos modélisations se sont toujours situées à un niveau de granularité mésoscopique assez grossier et ont toujours considéré des architectures fonctionnelles indépendantes de la structure microscopique fine des potentiels d'action voyageant le long des connexions neuronales. Or ces hypothèses sont trop simplificatrices. Elles permettent de proposer des modèles dont le contexte théorique est certes déjà mathématiquement complexe mais elles ne correspondent pourtant seulement qu'à une toute petite partie de la réalité neurobiologique. Une foule de problèmes restent en suspens et une foule de développements techniques sont envisageables et parfois d'ores et déjà disponibles au niveau expérimental. Nous voudrions en citer quelques-uns.

4.1. Stimuli artificiels-simples VS stimuli naturels-complexes

Nous avons utilisé des modèles de fibration avec "engrafted secondary variables" au sens de Hubel et cela signifiait que, pour calculer les champs et profils récepteurs des neurones visuels, nous considérons des stimuli artificiels descriptibles par peu de paramètres (par exemple une orientation). Mais aujourd'hui des laboratoires de plus en plus nombreux travaillent sur des stimuli *naturels* de beaucoup plus grande complexité. Cela est nécessaire car, comme y insiste Michel Imbert dans son *Traité du Cerveau*,

"Pour comprendre quelles sont les caractéristiques des stimuli qui sont encodés dans la décharge d'un neurone sensoriel, (...) il ne suffit plus d'utiliser des stimuli simplifiés. (...) Étant donné la nature très fortement non linéaire des réponses neurales, il faut coupler cette présentation de stimuli *naturels*

à des méthodes puissantes d'*analyse des données* recueillies, qui tiennent compte du fait que ces réponses sont bruitées et fortement non stationnaires.” ([263], p. 230)

Le lecteur intéressé trouvera de nombreux résultats expérimentaux dans le Vol I (par exemple section 3.6.), en particulier certains dus à l'équipe UNIC d'Yves Frégnac (section 5.3.). Un problème est alors de savoir comment réduire l'énorme dimensionnalité des stimuli naturels même locaux. Un autre problème est d'enrichir les profils récepteurs utilisés et de complexifier les états cohérents associés. Mais nous avons vu que le coût mathématique de la dialectique “squelette géométrique / chair fonctionnelle” est considérable.

4.2. Multi-fibrations et interactions entre cartes

Notre modélisation neurogéométrique s'est essentiellement développée dans une structure de contact tridimensionnelle correspondant aux hypercolonnes d'orientation avec parfois l'introduction de dimensions supplémentaires lorsqu'il s'agissait de tenir compte de la courbure (passage aux 2-jets et à la structure d'Engel) ou d'introduire un facteur d'échelle (symplectisation de la structure de contact), ou encore des repères mobiles (orientations + normales détectées par $V2$). Mais nous avons vu dans le Vol I (section 4.10.) qu'il existe de nombreux autres paramètres traités par les couches corticales bidimensionnelles : la phase, la fréquence spatiale, la dominance oculaire. Comment ces cartes bidimensionnelles cohabitent et s'organisent entre elles ? Comme le demandait Nicholas Swindale [519]

“How many maps are there in visual cortex?”.

Nous avons introduit un “principe de transversalité” posant que les lignes de niveau des cartes sont maximalelement transversales là où leur gradient est fort. La traduction neurogéométrique de ce principe reste à étudier.

4.3. Du niveau mésoscopique au niveau microphysique

Notre niveau de granularité n'est pas le niveau véritablement “microscopique” des neurones individuels et de leurs synapses mais un niveau “mésoscopique” que l'on peut caractériser en disant que les éléments de contact (a, p) y paramétrisent les colonnes corticales. Or si une colonne est bien une entité “micro” relativement aux structures globales de la perception, elle est en revanche tout à fait “macro” relativement aux arbres dendritiques buissonnants et au nombre énorme de synapses des neurones individuels. D'ailleurs, la simulation détaillée de l'état dynamique d'une colonne de quelques 10.000 neurones appartenant à quelques 200 types différents à partir des équations différentielles de leurs potentiels de membrane est l'un des défis computationnels actuels. On peut citer l'exemple du projet “Blue Brain” du Brain

Mind Institute de l'École Polytechnique Fédérale de Lausanne, projet financé par IBM qui utilise pour ce faire son nouveau superordinateur BlueGene. C'est dire qu'il existe une neurophysique fine incroyablement complexe sous-jacente à notre neurogéométrie mésoscopique.

4.4. Évo/Dévo

Même une fois effectuées ces extensions, il resterait à les insérer dans une théorie génétique et évolutionniste du développement. Nous avons évoqué dans le Vol I (chapitre 5, section 5.12.) la morphogenèse de la voie rétino-géniculo-corticale et de l'architecture fonctionnelle du cortex visuel ainsi que son contrôle génétique. Le guidage axonal ("axonal guidance") qui permet à ces connectivités très spéciales de se mettre place est fascinant. Résumons à titre de rappel cette section 5.12. du Vol I.

Des mécanismes moléculaires spécifiques interviennent avec des protéines membranaires ou diffusibles comme les sémaphorines, les nétrines, les éphrines et les Slits, accompagnées de morphogènes, de facteurs de croissance, de CAMS ("cell-adhesion molecules") et d'immunoglobulines, qui stimulent ou inhibent et orientent spatialement leur croissance. Un grand nombre de mécanismes sont nécessaires : le guidage des axones et des dendrites, leur branchement, la reconnaissance des cibles, la synaptogenèse, la dégénérescence (apoptose) des neurones et de leurs axones, la régénération des axones, le pruning (élagage) des arbres dendritiques, etc.

Au niveau génétique, le développement, la différenciation et la régulation sont contrôlés par un réseau de facteurs de transcription. Un progrès considérable a été la découverte à la fin des années 1970, en particulier par Edward Lewis²⁰ des gènes "homéotiques". Ces homéogènes sont des gènes de régulation déterminant les sites où certaines structures anatomiques vont se développer.

Le développement des voies optiques avec la formation du chiasme optique et les projections sur le colliculus supérieur et le corps genouillé latéral est particulièrement remarquable. Il est rendu possible par la présence le long du tractus optique de molécules de guidage sensibles à la position topographique. À l'extrémité des axones se trouvent des cônes de croissance qui, à travers des mécanismes de signalisation de l'*information positionnelle spatiale* portés par des protéines attractives ou répulsives, vont permettre aux axones de rejoindre leurs cibles et de se regrouper en faisceaux.

L'un des plus fascinants processus morphogénétiques concerne le contrôle du positionnement spatial dans les cartes rétino-topiques. Il exige un ciblage extrêmement précis et le guidage des axones n'y suffit pas. Il y faut un contrôle génétique plus précis et un raffinement de ce contrôle par l'activité due aux interactions entre les axones devenus voisins par guidage. L'activité spontanée des réseaux neuronaux

20. 1918-2004, Nobel 1995.

se révèle cruciale car elle engendre des ondes qui corrèlent les arbres dendritiques voisins.

Un mécanisme essentiel est celui de la “chimiotaxie axonale” au moyen de *gradients* de molécules chimio-attractives et chimio-répulsives comme les éphrines avec leurs récepteurs Eph.²¹ La nature moléculaire exacte de ces gradients n’a commencé à être comprise que dans les années 1990.

La génétique du développement est un monde en soi. Pour une introduction à leur problématique le lecteur pourra consulter *Machine-esprit* d’Alain Prochiantz [456].

Le développement du système visuel, sa phylogénèse et son ontogénèse et sa stabilisation par apprentissage soulève un passionnant problème pour la modélisation neurogéométrique. Il s’agit en effet de comprendre comment une connectivité au départ peu structurée peut évoluer et engendrer des structures comme des fibrations, des distributions de contact et des géométries sous-riemanniennes.

21. L’acronyme “Eph” vient de “Erythropoietin-producing human hepatocellular carcinoma cell line” et “ephrin” vient de “Eph family receptor interacting proteins”.

Conclusion

Le parcours de cet ouvrage a été bien long. Partis dans le Vol I de données neuro-physiologiques précises sur les champs et profils récepteurs des neurones visuels et sur les architectures fonctionnelles par lesquelles ils se trouvent connectés, nous avons exposé dans ce volume II tout un ensemble de modèles physico-mathématiques. Sous le titre de “neurogéométrie”, notre but était de promouvoir un certain type de structures géométriques en neurosciences, structures permettant d’expliquer comment des traits locaux peuvent être *intégrés* en formes globales.

Nous espérons avoir convaincu le lecteur que de la géométrie subtile se trouve neuralemement implémentée : géométrie de contact, géométrie symplectique, géométrie sous-riemannienne, groupes de Carnot, variétés de Heisenberg, tout cela en relation avec les états cohérents obtenus en faisant opérer le groupe euclidien $SE(2)$ sur une ondelette mère. En utilisant ces formalismes simplement pour $V1$, nous avons déjà pu rendre compte de plusieurs phénomènes perceptifs remarquables et énigmatiques.

Tout reste encore à explorer et élaborer dans ce vaste continent encore essentiellement non cartographié de la neurogéométrie et il faut se convaincre que les exigences de construction des modèles y sont considérables. Nous avons plusieurs fois expliqué que la neurogéométrie en est à un stade assez analogue à celui de la mécanique au début du XVIII^e siècle. Une fois établie l’idée que les phénomènes mécaniques étaient régis par des équations différentielles, on arriva à mathématiser correctement, mais déjà avec de grandes difficultés, des phénomènes simples que, comme le disait Jacques Bernoulli en 1694 (nous reprenons la magnifique citation du début du chapitre 9),

“la nature produit elle-même sans aucun artifice, d’un mouvement rapide et pour ainsi dire instantané”.

Il en va de même en neurogéométrie. Les outils de base sont ceux de la géométrie différentielle à la Lie-Cartan, des systèmes de Pfaff, des structures de contact, des structures symplectiques, des métriques sous-riemanniennes et des représentations de groupes. Ils permettent par exemple de modéliser des contours illusoires que

la perception naturelle produit elle-même sans aucun artifice, d’un mouvement rapide et pour ainsi dire instantané.

En conclusion de cette longue exploration nous voudrions revenir à des problèmes plus épistémologiques et en tirer quelques conséquences, selon nous critiques, pour la *philosophie de l'espace*. Nous souhaitons présenter quelques réflexions plus proprement *fondationnelles* à propos des données expérimentales neurobiologiques et des différentes modélisations physico-mathématiques exposées jusqu'ici. Notre hypothèse est que leur "autoréflexion" philosophique conduit à repenser profondément la nature de la perception spatiale et, à travers elle, une partie importante de plusieurs traditions philosophiques dans la mesure où le problème de la perception spatiale est central dans l'histoire de la pensée.

Nous déploierons l'autoréflexion sur deux niveaux.

1. Nous allons d'abord montrer comment les modèles neurogéométriques de la perception visuelle permettent d'éclairer le débat philosophique sur la conscience perceptive et la fameuse hypothèse d'un "explanatory gap" qui disjoindrait irréductiblement et irréversiblement les explications scientifiques des vécus subjectifs.
2. Nous concluons enfin par une incursion dans le domaine de la philosophie transcendantale de l'espace sensible.

La motivation d'un "finale" transcendantal est facile à comprendre. Notre exploration a abouti à une correspondance étroite entre, d'un côté, des descriptions conceptuelles abstraites ("eidétiques") de la phénoménologie de la perception et, d'un autre côté, des modèles géométriques et fonctionnels mettant en jeu des idéalités mathématiques abstraites. À partir de là, se pose tout naturellement la question de la correspondance entre ces deux types d'idéalités. Mais l'eidétique phénoménologique de la perception conduit à la reprise du problème kantien de l'esthétique transcendantale. En ce sens, la neurogéométrie peut être considérée comme la *naturalisation*, dans le cadre d'un "matérialisme neuronal", de l'idéalité transcendantale de l'espace au sens de l'esthétique transcendantale kantienne.

Ces remarques philosophiques conclusives s'inscrivent dans un programme de *naturalisation de la phénoménologie transcendantale* que nous poursuivons depuis de très longues années.¹ Au cours des années 1990 nous avons organisé avec Jean-Michel Roy, Francesco Varela et Bernard Pachoud un séminaire sur ce tournant naturaliste rendu possible, d'un côté, par les neurosciences cognitives et, d'un autre côté, par les modèles de la géométrie perceptive. Cela a abouti à la publication en

1. Citons par exemple nos études dans *Logos et Théorie des Catastrophes* [415] (le Colloque de Cerisy de 1982 en hommage à René Thom), dans le numéro spécial d'*Intellectica* "Philosophie et Sciences cognitives" dirigé par Jean-Michel Salanskis en 1993 [420], dans l'hommage à Gilles-Gaston Granger dirigé par Joëlle Proust en 1994 [423], ou dans le numéro spécial "Sciences cognitives et Phénoménologie" des *Archives de Philosophie* en 1995 [425].

1999 du volume *Naturalizing Phenomenology* dont la longue Introduction [471] est en grande partie due à Jean-Michel Roy.²

1. Conscience visuelle et neurogéométrie

1.1. Réductionnisme VS conscience phénoménale

Un certain monisme naturaliste étant apparemment la seule philosophie des sciences acceptable, il est tentant d'opter a priori pour une conception naturaliste de la conscience perceptive. Toutefois, adopter le point de vue d'un réductionnisme neurophysique causal semble extrêmement difficile à cause de ce que l'on appelle l'"explanatory gap" entre

- (i) la biophysique des réseaux de neurones,
- (ii) les traits qualitatifs de l'expérience consciente en première personne, ce que l'on appelle la "conscience phénoménale".

Rappelons brièvement l'opposition désormais classique introduite par Ned Block [58] entre la conscience d'accès (attention orientée objet, intentionnalité, contenus représentationnels) et conscience phénoménale (expérience en première personne, évidence subjective, qualia). C'est cette dernière qui fait problème. Elle semble en effet conduire inéluctablement à l'hypothèse dualiste d'une solution de continuité infranchissable entre une neurophysiologie objective et une conscience phénoménale subjective qui réfuterait tout physicalisme et constituerait ce que des philosophes comme David Chalmers [103], [104] ont appelé le "hard problem" de la conscience. En général, ce dernier est focalisé sur les *qualia*, c'est-à-dire sur le vécu subjectif de qualités sensibles comme la couleur.³ Mais il concerne selon nous de façon encore plus fondamentale le problème de l'espace.

Trois grands points de vue s'affrontent sur les relations entre neurophysiologie et expérience en première personne :

- (i) un réductionnisme matérialiste éliminativiste (monisme naturaliste) selon lequel le niveau qualitatif et intentionnel de la conscience n'est qu'un épiphénomène permettant de parler de façon commode de processus neurophysiologiques trop complexes ;

2. Jean-Michel Roy est l'auteur d'un ouvrage de référence [470] sur le "schisme" entre la phénoménologie et la philosophie analytique qui a joué un rôle considérable dans les débats épistémologiques du XX^e siècle.

3. Pour simplifier de façon outrancière, la thèse est que toutes les connaissances scientifiques possibles sur une couleur comme le rouge ne pourront jamais expliquer mon vécu subjectif *hic et nunc* de cette couleur rouge ci. Mais pour le matérialisme neuronal ce "hard problem" n'en est pas un car on ne voit pas bien comment l'activation des aires corticales impliquées dans des facultés conceptuelles abstraites de haut niveau pourraient bien activer de façon "normale" les aires visuelles primaires responsables du "vécu" chromatique.

- (ii) un dualisme ontologique à la Descartes ;
- (iii) un émergentisme selon lequel la conscience perceptive émerge des dynamiques neurophysiologiques sous-jacentes et n'est pas épiphénoménale dans la mesure où elle possède un niveau de réalité *sui generis*.

Le débat entre les monismes naturalistes (i) et les dualismes “cartésiens” (ii) est vif. C'est un exemple typique de couple thèse / antithèse au sens de la dialectique transcendantale kantienne. Les réductionnistes affirment que les phénomènes, structures et processus complexes de haut niveau peuvent être réduits, en ce qui concerne leur explication scientifique, à des phénomènes, structures et processus sous-jacents de bas niveau. L'exemple paradigmatique reste celui de la réduction de la thermodynamique macroscopique à des mouvements moléculaires microscopiques (température = énergie cinétique moyenne par degré de liberté, etc.). Donnons quelques précisions à leur sujet.

1. Le réductionnisme peut être une thèse scientifique particulière concernant une théorie scientifique spécifique : c'est précisément le cas avec la réduction de la thermodynamique macro à une mécanique statistique micro. Mais il peut être également une affirmation métaphysique générale à propos de la nature ultime de la réalité. Les monismes matérialistes affirment la réductibilité universelle non seulement causale mais également ontologique de toute réalité à la matière et l'énergie. Dans cet ouvrage, nous nous sommes restreints au réductionnisme scientifique et nous n'avons abordé aucun problème ontologique au sens fort du terme (celui d'une réalité indépendante de notre connaissance) car le principe de restrictions aux observables interdit aux sciences tout contenu ontologique de ce type.

2. Le réductionnisme scientifique peut lui-même être “objectif” ou “méthodologique”. Il est “objectif” quand il concerne des explications en termes d'objets primitifs (atomes, neurones, etc.) et “méthodologique” quand il concerne des compréhensions ontologiquement déflationnistes (rasoir d'Ockham). Un important débat en sciences cognitives porte sur l'éliminabilité des concepts “mentaux” et leur réductibilité à des concepts “neuronaux” (cf. par exemple la controverse entre Daniel Dennett et David Chalmers que nous évoquons plus bas). Dans cet ouvrage, nous n'avons envisagé que des formes “objectives” de réductionnisme.

3. Au sens étroit, scientifique et objectif, qui a été ici le nôtre, le réductionnisme n'est pas un éliminativisme et concerne les systèmes complexes possédant au moins deux niveaux de réalité : un niveau micro sous-jacent où interagissent un nombre considérable d'unités élémentaires et un niveau macro où émergent des macrostructures auto-organisées. La médiation entre les deux niveaux s'effectue à travers un niveau fonctionnel intermédiaire (méso). Dans une telle perspective, le réductionnisme n'est ni ontologique, ni conceptuel et reste inséparable de concepts

anti-éliminativistes tels ceux d’“émergence”, de “supervenience” ou de “fonctionnalisme”. Le fonctionnalisme signifie que les macro-structures possédant un rôle fonctionnel ne peuvent certes exister que si elles sont matériellement implémentées dans un substrat matériel sous-jacent, mais qu’elles sont dans le même temps, précisément parce que dotées de signification fonctionnelle, en grande partie indépendantes des propriétés physiques fines du substrat. L’exemple paradigmatique est celui de l’opposition software / hardware en informatique (cf. les travaux de philosophes fonctionnalistes comme le premier Hilary Putnam, ou Jerry Fodor et Zenon Pylyshyn, et tant d’autres), mais le problème est beaucoup plus général.

4. Dans les systèmes complexes possédant différents niveaux de réalité à plusieurs échelles, il existe des comportements collectifs régis par des lois qui ne sont pas les lois du niveau micro sous-jacent. C’est le cas dans d’innombrables domaines : phénomènes critiques, percolation, criticalité auto-organisée, processus de réaction-diffusion, structures dissipatives, turbulence, automates cellulaires, réseaux de neurones, colonies de fourmis, essaims, marchés financiers, etc. Suivant la conception que l’on se fait des lois, on peut interpréter ce *factum rationis* de différentes façons :

- (i) À partir d’un empirisme éliminativiste et épiphénoménaliste : les lois ne sont que des régularités empiriques statistiques et ne possèdent aucun contenu objectif (et a fortiori ontologique) ; les structures émergentes sont purement épiphénoménales et peuvent être scientifiquement éliminées “salva veritate”.
- (ii) À partir d’un réalisme holistique (c’est la position converse) : les lois sont réelles au sens ontologique ; les niveaux émergents possèdent une réalité ontologique et par conséquent ne peuvent pas être réduits.
- (iii) À partir d’un réductionnisme causal complété d’un émergentisme objectif (c’est la solution dépassant l’antinomie précédente) : les lois sont objectives (ce qui ne veut pas dire ontologiques), c’est-à-dire à la fois empiriquement fondées et mathématiquement formalisables ; les niveaux émergents ne possèdent pas de contenu ontologique propre mais sont néanmoins beaucoup plus que de simples régularités statistiques empiriques dégagées inductivement. Le niveau macro est certes *causalement* réductible à des interactions complexes au niveau micro sous-jacent, mais il possède néanmoins une *autonomie* observationnelle et théorique.

5. Dans la perspective (iii), la principale difficulté est, par définition, de concilier la réduction causale avec l’autonomie théorique. On ne peut pas la résoudre si l’on ne comprend pas le rôle fondamental que jouent les mathématiques. En effet, la légitimation formelle de la réduction causale d’un niveau de réalité est que l’on puisse en déduire mathématiquement les propriétés à partir d’un niveau sous-jacent, ce que l’on peut appeler une “synthèse computationnelle”. Mais la déductibilité est

une propriété syntaxique et n'implique en tant que telle aucune dérivabilité purement conceptuelle.⁴ Par conséquent, le fait que des structures et des propriétés du niveau macro puissent être mathématiquement déduites des processus micro sous-jacents ne signifie pas pour autant que leur conceptualisation puisse être *conceptuellement* réduite au contenu du niveau micro. La réduction causale formalisée par une déduction mathématique n'est pas une réduction conceptuelle. C'est pourquoi les structures émergentes peuvent appartenir à des niveaux théoriques autonomes de réalité.

1.2. La conscience visuelle comme émergence

Le problème de l'émergence de la conscience visuelle est, comme le dit Max Velmans [552], que :

“non conscious processes within the brain produce consciously experienced events which may be subjectively located and extended in the phenomenal space beyond the brain.”

L'une de ses principales difficultés est qu'il intrique deux sous-problèmes complètement différents :

1. l'émergence de macro-structures hors de mécanismes physiologiques dans un substrat neuronal micro N (*Neuro*),
2. l'accord ou matching entre certaines macro-structures émergentes M (*Macro*) et certaines expériences phénoménales E (*Exp*).

Pour ne prendre qu'un exemple, Alva Noë et Evan Thompson ont publié en 2004 dans le *Journal of Consciousness Studies* un “target paper” “Are there Neural Correlates of Consciousness?” [391] contre le dualisme de David Chalmers, en utilisant la définition suivante des “corrélats neuronaux de la conscience” (NCC) :

“A NCC for a content is a minimal neural representational system N such that the representation of a content in N is sufficient, under conditions C , for the representation of that content in consciousness.” (p. 31)

Ils y affirment que (RF = “receptive field”) :

“RF-content is too thin to sustain a match with perceptual experience.” (p. 90)

“It is difficult to see how a structural coherence could be built up (...) out of RF-contents atoms.” (p. 14)

4. Nous entendons par dérivabilité “conceptuelle” une dérivabilité se situant au niveau du “sens” des concepts. L'analyse conceptuelle d'un niveau de réalité (ici le niveau “macro”) est très différente de la synthèse computationnelle de ce niveau au moyen d'algorithmes mathématiques. Nous avons même soutenu ailleurs la thèse qu'il s'agit de processus inverses l'un de l'autre.

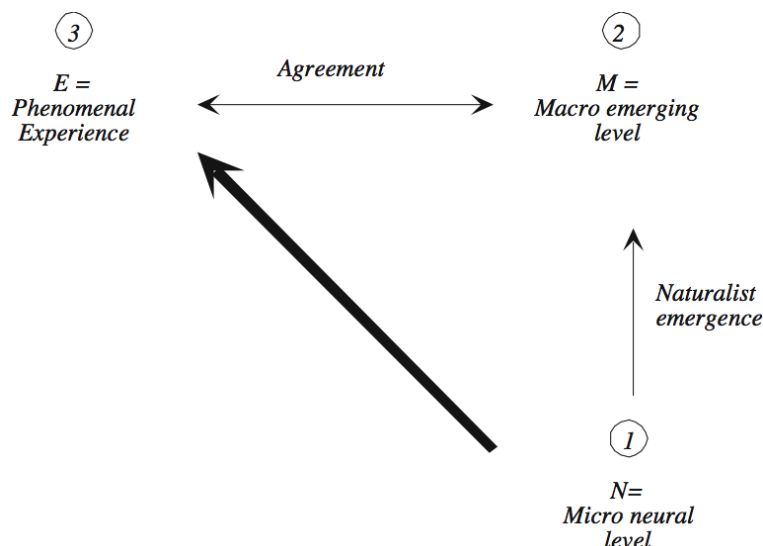


FIGURE 1. La factorisation, à travers un niveau macro émergent M , de la relation d'émergence de l'expérience phénoménale E hors du micro-niveau neuronal N .

Mais précisons un peu l'argument d'un gap $Neuro \rightarrow Exp$. Si l'on cherche à corrélérer directement E avec N , on rencontre de grands obstacles car l'émergence s'intrique avec l'opposition "objectivité à la troisième personne" / "subjectivité à la première personne". Une réponse émergentiste au problème des NCC ne peut être philosophiquement vraiment clarifiée que si une émergence éventuelle $N \rightarrow E$ peut d'abord être "factorisée" (comme diraient les mathématiciens) à travers un niveau macro naturel M ne mettant pas en jeu des vécus subjectifs (cf. figure 1).

La factorisation désintrique en effet le problème.

- (i) L'émergence "verticale" $1 \rightarrow 2$ est strictement naturaliste. Le problème qu'elle pose n'est pas un problème de naturalisation des vécus subjectifs mais le problème de changement de niveau : *déduire* 2 de 1, dériver le "macro" du "micro".
- (ii) L'accord "horizontal" $3 \leftrightarrow 2$ pose un problème de matching de contenus représentationnels de même niveau, mais pas de problème d'émergence.

Dans notre cas de la perception visuelle de formes, le macro-niveau naturaliste 2 est morphologique, et par conséquent géométrique. Et c'est précisément son caractère géométrique (qui n'est plus biophysique et peut s'accorder avec un vécu perceptif) qui apporte une solution et révèle l'obstruction méthodologique conduisant à l'hypothèse d'un "explanatory gap" :

- (i) le micro-niveau neurophysique 1 est assez bien connu grâce aux résultats de la neurophysiologie ;

- (ii) le niveau phénoménal 3 est assez bien décrit grâce aux descriptions eidétiques de la phénoménologie ;
- (iii) en revanche, et c'est toute la difficulté, le macro-niveau morphologique 2 demeure essentiellement inconnu tant qu'il n'a pas fait l'objet d'une démonstration mathématique : il ne peut pas être conceptuellement dérivé du niveau 1 mais seulement reconstruit mathématiquement par synthèse computationnelle ; il dépend donc entièrement de théorèmes dont il est impossible d'anticiper la démonstration un jour ou l'autre.

On voit par conséquent le dilemme : en l'absence de 2 la question d'une émergence possible $1 \rightarrow 3$ devient insoluble, mais en présence de 2 sa solution dépend *hic et nunc* des progrès contingents des mathématiques. C'est pourquoi les thèses philosophiques dogmatiquement dualistes affirmant que "il ne sera jamais possible de réduire ceci à cela" sont certes compréhensibles mais ne sont pas pour autant recevables car il est impossible d'anticiper les futurs progrès des mathématiques. Elles reposent sur une méconnaissance profonde du fait, qu'en science, l'analyse conceptuelle joue un rôle second (ce qui ne veut pas dire secondaire) par rapport à la synthèse computationnelle.

L'"explanatory gap" est par conséquent *relatif* dans la mesure où il dépend entièrement de l'histoire scientifique du niveau 2. Si 2 fait défaut alors, comme il n'y a pas de matching direct de contenu entre 1 et 3, le gap paraît exister. Mais plus la déduction $1 \rightarrow 2$ se renforce par synthèse computationnelle, plus il se réduit. Et il n'y a aucune raison de principe pour rejeter l'idée qu'un niveau 2 suffisamment complet ne puisse un jour le combler.

1.3. Retour sur la vision entoptique

L'exemple de la vision entoptique exposé à la section 2 du chapitre 18 fournit un exemple frappant d'émergence de type $1 \rightarrow 2$ et nous avons vu à quel point la dérivation mathématique en est non triviale. Pour spécifier le statut philosophique de l'expérience vécue qui lui sert de base, nous nous inspirerons de la liste des caractères possibles d'un vécu de conscience dressée par Robert Van Gulick [550] dans son item "Consciousness" de la *Stanford Encyclopedia of Philosophy*.

1. C'est une expérience subjective, qualitative et phénoménale, qui ne relève pas de la conscience d'accès.⁵
2. Elle n'est ni conceptuelle ni linguistique mais communicable par des comptes-rendus graphiques.
3. Elle n'est pas consciente au sens réflexif : il ne s'agit pas de savoir si elle est accompagnée par le cogito comme état méta-mental.

5. Cf. plus haut l'opposition "conscience d'accès"/"conscience phénoménale" chez Ned Block.

4. Elle ne possède pas de contenu intentionnel et sémantique.
5. Elle n'est pas un événement dans un flux de conscience compliqué "narrativement" organisé.
6. Elle n'est pas impliquée dans des boucles sensori-motrices.

En fait, il s'agit d'un exemple (assez exceptionnel) d'expérience qualitative ne possédant qu'un pur contenu phénoménal et il est donc particulièrement bien adapté à la preuve d'une possibilité d'accord entre un contenu phénoménal et un contenu neuro-dynamique.

Nous avons vu à la section 2 du chapitre 18 que les modes propres de l'activité neuronale de $V1$ modélisant les planforms de Klüver étaient mathématiquement déductibles des équations de Hopfield, mais qu'ils jouissaient en même temps d'une certaine autonomie dans la mesure où ils dériveraient essentiellement de brisures de symétrie relativement au groupe $E(2)$ implémenté par l'architecture fonctionnelle. Il s'agit là, nous ne saurions trop le souligner, du point épistémologique central. Les contraintes géométriques imposées au modèle physique ne sont pas elles-mêmes de nature physique. Elles sont *idéales* (formelles) et exemplifient un type important, souvent négligé dans le débat philosophique, de "fonctionnalisme". Si donc un accord peut s'établir entre elles et des vécus perceptifs, ces derniers peuvent eux-mêmes être considérés comme émergents. Entre le micro-physique 1 et l'expérience vécue 3 il existe le macro-niveau morphologique et géométrique 2 qui est *à la fois* physique *et* vécu et réalise la factorisation $1 \rightarrow 2 \rightarrow 3$.

1.3.1. *L'accord* $2 \leftrightarrow 3$.

En revenant à la figure 15 de Klüver/Bressloff de la section 2 du chapitre 18, nous constatons que l'accord entre le macro-niveau géométrique (morphologique) émergent M (niveau 2) et l'expérience phénoménale E (niveau 3) est extrêmement fort, beaucoup plus fort qu'une simple corrélation. C'est même la forme la plus forte possible de matching de contenus puisque, à la limite, c'est même un *isomorphisme*.

Mais nous devons être ici très prudents : les contenus en jeu ne sont pas des contenus conceptuels mais des contenus géométriques non conceptuels. L'isomorphisme dont il s'agit est un isomorphisme *géométrique* entre deux contenus géométriques, celui de l'espace intuitif de l'expérience et celui d'un espace mathématique (cf. figure 2).

Le jeu ne se joue donc pas entre une expérience phénoménale holistique macro et une neurophysique locale micro mais entre une expérience phénoménale, une neurophysique locale *et* une géométrie globale émergente. L'isomorphisme est un matching entre deux morphologies qui sont des contenus géométriques. Il n'est pas problématique en tant que tel. Ce qui est problématique est le matching entre *les deux types d'espaces* où ces contenus morphologiques sont plongés et formatés.

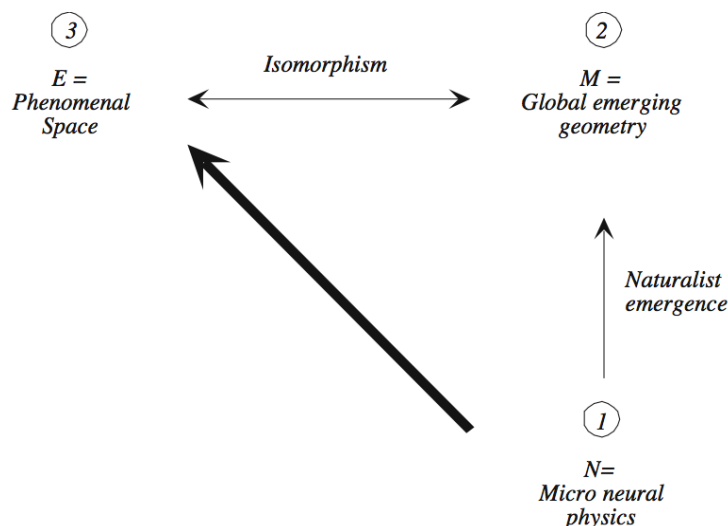


FIGURE 2. Le matching $2 \leftrightarrow 3$ est un isomorphisme.

Nous rencontrons ici le véritable “hard problem”, celui qui court de l’esthétique transcendantale de Kant à Poincaré et Husserl : quel est le lien entre l’“intuition pure” de l’espace en tant que “forme de l’intuition” phénoménale (un format des percepts) et l’espace en tant qu’“intuition formelle” géométrique objective⁶.

C’est en définitive celui de *l’idéarité transcendantale* de l’espace, dont la méconnaissance est la source de la croyance à un “explanatory gap”. L’espace intuitif est un format multisensoriel commun constitutif de la conscience phénoménale. Il est la forme de la phénoménalité elle-même. Interne et subjectif, fondé dans la temporalité interne de l’expérience subjective en première personne, il est en même temps la forme universelle de l’objectivité externe à la troisième personne.

Si l’on accepte l’identité de ces deux types d’espaces respectivement objectif et intuitif, alors le matching $2 \leftrightarrow 3$ devient un cas très fort d’identité de contenus non conceptuels : modes propres et planforms \leftrightarrow formes entoptiques. Si au contraire l’on n’accepte pas cette identité, alors le matching devient un isomorphisme entre morphologies plongées dans des espaces ambiants différents, respectivement géométrique et phénoménal.

6. Nous reprenons la différence essentielle introduite par Kant entre “forme de l’intuition” et “intuition formelle”. “L’espace représenté comme *objet* (ainsi qu’il en est réellement besoin dans la géométrie), contient plus que la simple forme de l’intuition ; il contient le *rassemblement* en une représentation *intuitive* du divers, donné selon la forme de la sensibilité, de telle sorte que la *forme de l’intuition* donne simplement du divers, tandis que l’*intuition formelle* donne l’unité de la représentation.” (*Critique de la Raison pure*, §26, Pléiade, I, 873, *Kants gesammelte Schriften*, III, 125)

1.3.2. *L'émergence 1 → 2 comme corrélat neuronal.*

Toujours dans le cas de cet exemple, le modèle mathématique d'Ermentrout, Cowan, Golubitsky et Bressloff explique complètement, au moyen d'une longue déduction mathématique, le niveau morphologique émergent *Macro 2* à partir de la dynamique du substrat neuronal sous-jacent *Neuro 1*. Dans ce cas, il n'existe donc pas d'"explanatory gap" et, comme nous venons de l'argumenter, le problème se trouve déplacé vers celui de l'idéalité transcendantale de l'espace. Nous sommes donc en présence d'un type non réductionniste de physicalisme : *M* "appartient" à *N* dans la mesure où les propriétés naturelles de *N* sont des conditions nécessaires et suffisantes pour son instanciation, mais *M* n'est pas pour autant conceptuellement inclus "dans" *N* puisqu'il en est conceptuellement autonome.

Dans cet exemple, on voit que $N = V1$ (niveau 1) fournit un NCC pour l'expérience visuelle *E* (niveau 3). Il fournit ce que l'on appelle un "bridge locus" pour *M* (niveau 2), et à travers *M*, pour *E* (niveau 3). Bien sûr, il est bien connu que *V1* ne se corrèle pas bien avec des structures perceptuelles de plus haut niveau possédant un contenu intentionnel et sémantique (reconnaissance d'objets, etc.). D'autres aires (par exemple l'aire inféro-temporale IT) se corrèlent mieux. Mais cela ne soulève pas de question philosophique car, dans notre cas, *V1* se corrèle fort bien avec les morphologies que l'on veut expliquer. *V1* est bien un NCC, et qui plus est un NCC *minimal* car on a besoin de toute son architecture fonctionnelle pour expliquer les phénomènes considérés.

Pour bien argumenter, il faut définir correctement le contenu représentationnel de $N = V1$. La limite des critiques dualistes est qu'elles limitent ce contenu représentationnel à celui, très local, des champs récepteurs (RF) des neurones visuels. Mais, comme tout cet ouvrage a essayé de le montrer, celui-ci n'épuise pas, loin de là, celui de *V1*. Le véritable contenu représentationnel de $N = V1$ est défini par *l'ensemble* de son architecture fonctionnelle globale. Tant que l'on ne met pas ce fait au centre de la réflexion épistémologique, aucune argumentation ne saurait être valide. Ce n'est qu'à travers ses connexions rétino-topiques "verticales" et ses connexions cortico-corticales "horizontales" que *V1* acquiert un contenu géométrique global émergent :

Contenu représentationnel neuronal local	≡	Champs récepteurs, orientations préférentielles
Architectures fonctionnelles	≡	1. Rétinotopie, 2. connexions horizontales
Contenu représentationnel géométrique global	≡	Modes propres (planforms)

Si l'on traite les modes propres mathématiquement déduits comme une partie du contenu représentationnel du système neuronal $N = V1$, alors l'exemple montre que

des NCC existent effectivement. Si en revanche l'on n'accepte pas que les contenus soient déductivement transitifs, c'est-à-dire qu'un contenu C_2 mathématiquement déductible d'un contenu C_1 soit considéré comme inclus dans C_1 , alors on ouvre un "explanatory gap". On voit ainsi que la réponse que l'on est en droit d'apporter à la question des NCC dépend à la fois de progrès mathématiques inanticipables et d'options épistémologiques très générales. Appelons "contenu représentationnel déduit" (RC_2) une structure mathématiquement déduite du contenu représentationnel de base RC_1 d'un système neuronal N . La question est de savoir si l'on accepte ou non RC_2 comme une partie de RC_1 . Notre thèse est que, même si RC_2 ne peut pas être inféré *conceptuellement* de RC_1 , il doit quand même être considéré comme "appartenant" à RC_1 s'il peut en être mathématiquement déduit. Dans ce cas, il est beaucoup plus probable de trouver des NCC. Mais, dans la mesure où les démonstrations dépendent du progrès contingent des mathématiques, l'existence de tels NCC ne peut pas être un fait empirique mais seulement une inférence théorique possible.

1.4. Émergentisme dynamique

Précisons encore un peu le statut philosophique de l'exemple.

1. Il ne relève pas d'un réductionnisme éliminativiste et d'un épiphénoménalisme : les modes propres modélisant les planforms de Klüver existent en tant qu'idéalités géométriques indépendamment de leur implémentation neurophysiologique fine ; leur condition nécessaire et suffisante est l'architecture fonctionnelle.
2. Il n'est pas dualiste : la corrélation *Neuro* (1) \rightarrow *Macro* (2) ne relie pas des domaines ontologiquement hétérogènes. Les morphologies émergentes sont des idéalités géométriques et ne possèdent par conséquent aucun contenu ontologique propre.
3. Il n'est donc pas du type d'un émergentisme ontologique où les structures émergentes seraient non seulement formellement autonomes mais aussi causalement irréductibles. Ici, ces dernières sont dans le même temps formellement autonomes *et* causalement réductibles.
4. Il ne relève pas d'un fonctionnalisme classique où le haut niveau macro M (2) serait simplement codé au sens d'une machine de Turing universelle dans le bas niveau micro N (1). Ici, les structures 2 émergent à travers un processus morphodynamique de bifurcation et d'un phénomène critique de brisure de symétrie qui "décroche" 2 de 1 et l'autonomise.
5. On peut le qualifier de "fonctionnalisme dynamique" : la géométrie globale M (2) ne peut être déduite ni du comportement de neurones individuels isolés ni du comportement collectif de réseaux organisés différemment ; mais elle

est néanmoins “law-like” et peut être nomologiquement déduite de la dynamique contrôlée par l’architecture fonctionnelle spécifique du système neuronal $N(1) = V1$ considéré.

Peut-être qu’une bonne formulation de ce statut philosophique de notre exemple serait d’en faire une variante de ce que Thomas Metzinger appelle la “double-aspect theory” :

“Scientifically describing $[N]$ and phenomenally experiencing $[E]$ are just two different ways of accessing one and the same underlying reality.” ([367], p. 4)

Ici la réalité sous-jacente commune est la *cohérence globale* d’une dynamique neuronale. Elle peut donner lieu d’un côté à une réalité géométrique idéale émergente (contenu représentationnel géométrique) et d’un autre côté à une expérience phénoménale (contenu représentationnel phénoménal). En ce sens, l’émergentisme dynamique que nous invoquons peut être considéré comme une théorie naturaliste du processus même de phénoménalisation qui, à travers des dynamiques neuronales, transforme les données sensorielles en concepts. Bref, la conscience phénoménale doit elle-même être conçue comme un phénomène naturel.

Si nous revenons à notre factorisation $1 \rightarrow 2 \rightarrow 3$ nous pouvons conclure que :

- (i) il n’existe pas de “hard problem” philosophique pour l’émergence $1 \rightarrow 2$ d’une géométrie globale hors d’une physique locale ; en revanche il existe bien un “hard problem” mathématique pour démontrer que 2 est dérivable de 1 ;
- (ii) il n’existe pas de “hard problem” pour le “content matching” $2 \leftrightarrow 3$ (isomorphisme des morphologies) ;
- (iii) en revanche, il existe bien un “hard problem” philosophique concernant la relation entre l’espace géométrique et l’espace phénoménal ;
- (iv) cette dernière relation n’est pas un matching de contenus conceptuels, mais un matching entre des intuitions pures immanentes et des idéalités mathématiques servant de “background structure” à l’objectivité externe.

2. Esthétique transcendantale et neurogéométrie

Nous allons maintenant développer les réflexions entamées dans notre petit “interlude philosophique” de la section 5 du chapitre 2.

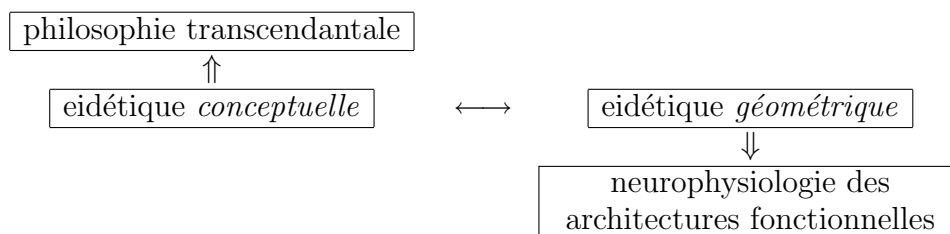
2.1. Les deux eidétiques et leurs fondations

Nous avons donné plusieurs exemples de l’accord profond qui existe entre, d’un côté, les descriptions phénoménologiques exprimées en termes de vécus immanents et, d’un autre côté, les structures géométriques et morphologiques de l’apparaître.

Nous avons ainsi établi un parallèle entre l'eidétique *conceptuelle* d'une phénoménologie descriptive et l'eidétique *géométrique* des modèles morphologiques :



Mais il faut aller beaucoup plus loin et s'interroger sur les fondements mêmes de ces deux eidétiques. Du côté de l'eidétique géométrique, "aller plus loin" signifie accéder expérimentalement et mathématiquement à son implémentation neuronale et c'est précisément le but de la neurogéométrie. Mais du côté de l'eidétique conceptuelle, "aller plus loin" signifie tout autre chose que d'aller chercher des causes profondes à l'intérieur d'une boîte noire cérébrale, d'ailleurs considérée comme inaccessible par les philosophes anti-psychologistes. Il s'agit d'une auto-réflexion conceptuelle qui renvoie à l'ancienne et puissante tradition philosophique de la recherche de *conditions de possibilité* des structures, règles et lois mises à jour par la description scientifique. Cette tradition est celle de la philosophie transcendantale et c'est pourquoi il existe une solidarité architectonique entre neurogéométrie et philosophie transcendantale :



2.2. Synthétique a priori et neurosciences

Si nous appliquons ce schéma général à l'espace, alors nous voyons apparaître un *lien architectonique entre neurogéométrie et esthétique transcendantale*. Certes, l'esthétique kantienne semble fort loin de nous, d'autant plus qu'elle a été fortement critiquée. Mais elle garde pourtant une pertinence remarquable dans ce nouveau contexte et se trouve en quelque sorte revivifiée et même "vindicated".

Nous avons vu dans le Vol I que l'origine neuronale de l'espace sensible se trouve dans la synchronisation temporelle fine (le "binding") des activités d'un double système de connexions, celui de la "rétinotopie" de la voie rétino-géniculo-corticale et celui de l'architecture fonctionnelle des aires primaires du cortex visuel, en particulier l'aire V1. L'espace apparaît ainsi comme un *format* pour le traitement des informations sensorielles véhiculées par les fibres du nerf optique.

Or il suffit de comprendre que l'espace est un format neurophysiologique défini par l'architecture fonctionnelle du système visuel pour valider immédiatement la

plupart des thèses kantienne sur l'espace sensible. Ce formatage est en effet par définition une forme ontogénétique a priori des contenus sensoriels et n'appartient donc pas à ces contenus eux-mêmes puisque ces derniers ne sont que des inputs qui, ainsi formatés, structurés et traités, se transforment en percepts. Le fait, qu'en tant que format sensoriel, l'espace soit un résultat de l'évolution biologique ne remet pas en cause son statut a priori car les a posteriori de la phylogenèse sont des a priori de l'ontogenèse.

Le formatage spatial est "synthétique" et non conceptuel (ante-prédictif et pré-judicatif) dans la mesure où les aires visuelles occipitales ne sont pas les aires temporales du langage et leur architecture fonctionnelle détermine un format non prédictif. On est vraiment encore très proche du Kant qui affirmait :⁷

"Il n'y a ainsi pour mon intuition qu'une seule façon possible de précéder la réalité effective de l'objet et de se produire comme connaissance a priori, c'est de ne contenir rien d'autre que la forme de la sensibilité, qui dans le sujet que je suis précède toutes les impressions effectives par lesquelles je suis affecté par des objets" (*Prolégomènes*, P II 50, Ak IV 282).

La thèse de l'analyticité de l'espace doit donc être profondément repensée si l'on admet que toute structure analytique présuppose un format propositionnel. En effet, il est neurophysiologiquement faux que la perception s'identifie à des *jugements* perceptifs. Son format n'est pas prédictif et judicatif mais intuitif et géométrique en un sens synthétique.

2.3. Neurogéométrie et exposition métaphysique

Ce qui relevait chez Kant de l'"exposition métaphysique" des formes de l'intuition dans l'Esthétique transcendantale relève selon nous aujourd'hui de la neurogéométrie qui formate les sensations, c'est-à-dire de la méso-géométrie des architectures fonctionnelles cérébrales. Une telle "naturalisation" change évidemment le statut de l'a priori en le faisant passer du transcendantal au neuronal, mais beaucoup moins qu'on pourrait le croire. En effet, les architectures fonctionnelles appartiennent au patrimoine génétique de notre espèce et les structures héritées de la phylogenèse sont, répétons-le, des a priori ontogénétiques pour les sujets. Certes Kant était anti-innéiste mais de façon subtile. Dès la *Dissertation* de 1770, il explique bien sûr que les catégories ne sont pas innées mais acquises, mais elles sont acquises réflexivement en tant que

"concepts abstraits des lois inhérentes à l'esprit" (*Dissertation*, P I 642, Ak II 395)

7. Dans cette section, nous indiquerons les références aux textes de Kant par les volumes *P I* ou *P II* de l'édition de la Pléiade et par le volume de l'édition *Ak* des œuvres complètes à l'Académie de Berlin : *Kants gesammelte Schriften*, Preussische Akademie der Wissenschaften.

et on ne voit pas comment l'inhérence pourrait ne pas être d'une certaine façon innée.

Quant aux intuitions pures, en faire des représentations innées relèverait certes d'une philosophie "paresseuse". Elles sont acquises, mais au sens de

"abstraites de l'action même par laquelle l'esprit coordonne, selon des lois permanentes, ses sensations." (*Dissertation*, P I 658, Ak II 406)

Or, cette coordination est elle-même innée et fonctionne comme un fondement de l'acquisition. D'ailleurs, dans sa réponse à Eberhard de 1790, Kant parle "d'acquisition originaire" des intuitions pures et insiste sur le fait que

"Il faut cependant qu'il y ait pour ce faire un fondement dans le sujet, fondement par lequel il est possible que les représentations en question naissent ainsi et pas autrement, et qu'en outre elles puissent être rapportées à des objets qui ne sont pas encore donnés; ce fondement, du moins, est inné." (P II 1351, Ak VIII 221-222)

Autrement dit, le fondement de la possibilité d'une intuition spatiale est inné et c'est en lui que s'enracine la condition subjective a priori d'être affecté par des objets.

Mais c'est exactement cela la fonction des architectures fonctionnelles : être la condition de possibilité des formats qui formatent les données sensorielles. Par un passage de la philosophie à la science, l'Esthétique transcendantale se trouve ainsi relayée par une problématique neurophysiologique précise qui, en retour, en confirme le bien fondé.

2.4. Géométrie et réflexion : "décompiler" le calcul neuronal

Le fait qu'il existe des "formes de l'intuition" qui formatent les sensations peut par conséquent être désormais considéré comme une donnée scientifique. Non seulement le synthétique a priori a un sens, mais les jugements synthétiques a priori géométriques peuvent être interprétés comme la "réflexion" de structures neurophysiologiques fondamentales. On ne saurait donc arguer – et ce point est vraiment crucial – du fait que toute géométrie se développe nécessairement, en tant que théorie mathématique, comme un système d'énoncés, pour conclure que les formes de l'intuition sont "analytiques" au sens logiciste. En effet, la géométrie intuitive que nous éprouvons constamment dans la perception visuelle provient de l'architecture fonctionnelle des aires visuelles.

Nous avons vu les exemples du transport parallèle (reconnaître que deux orientations en deux points différents du champ visuel sont les mêmes), du principe gestaltiste de "bonne continuation" (le système visuel a tendance à prolonger co-axialement les segments orientés), des mécanismes d'intégration des bords (des détections locales de bords d'objets sont intégrées en bords globaux), des contours

illusoires comme géodésiques, des patterns entoptiques comme modes propres d'activité neuronale. Toute cette neurogéométrie fonctionnelle de base qui conduit progressivement à la géométrie euclidienne est une conséquence des architectures fonctionnelles visuelles. Les contenus sensoriels correspondent à des inputs neuronaux qui se propagent dans ces réseaux, mais c'est le "design" de ces derniers qui les formate et fonctionne comme une forme a priori de la sensibilité. La géométrie classique de l'espace "externe" est l'axiomatisation de certaines constructions résultant de cette neurogéométrie "interne" et, en tant que telle, elle fait passer des formes de l'intuition à ce que Kant appelait des "intuitions formelles".⁸

Exprimées en termes de formatage, les thèses du §15 de la *Dissertation* et l'Esthétique transcendantale de la *Critique de la Raison pure* deviennent évidentes. L'espace ne peut pas être abstrait des sensations (l'abstraction relationnelle leibnizienne est neurophysiologiquement problématique), il n'est pas un réceptacle d'objets, il est une représentation singulière, il est une intuition pure et il n'est pas réel mais idéal en tant que forme coordonnant (liant) le divers de la sensation. C'est en effet l'architecture fonctionnelle qui explique par son formatage les phénomènes de binding de données sensorielles locales en Gestalts perceptives globales. Cela correspond exactement à la synthèse kantienne du divers de la sensation, synthèse phénoménologiquement conditionnée par les formes de l'intuition et mathématiquement déterminée par les intuitions formelles.

La géométrie formalise et axiomatise cette situation exactement comme en informatique la logique formelle formalise et axiomatise des calculs machine (correspondance de Curry-Howard).⁹ Elle "décompile" – c'est ce que signifie ici "réflexion" – les algorithmes neuronaux du système visuel et, dans la mesure où ces algorithmes sont matériellement réalisés dans le hardware neuronal, elle ne peut pas être conceptuelle et analytique dans son contenu bien qu'elle soit composée de jugements mathématiques. L'intuition pure n'est pas conceptuelle et l'Esthétique transcendantale n'est pas résorbable dans l'Analytique transcendantale à travers une extension de la logique formelle.

8. Cf. la célèbre note au §26 de la Dédution transcendantale de la *Critique de la Raison pure* (P I 873, Ak III 125), note que nous avons déjà citée plus haut.

9. Comme nous l'avons déjà expliqué dans la section 5 du chapitre 2, la correspondance de Curry-Howard établit une correspondance entre des programmes de bas niveau (λ -calcul typé) proches du langage machine et des preuves logiques de haut niveau qui s'y trouvent compilées et qui les typent, et cela de façon à ce que les démonstrations logiques correspondent au fait que les calculs de bas niveau soient exécutés correctement sur la machine (β -réduction).

2.5. Intuitif VS Conceptuel

La réhabilitation qu'apportent les neurosciences cognitives contemporaines aux thèses kantienne se renforce encore lorsqu'on prend en compte les relations qu'entretiennent les aires visuelles primaires occipitales avec d'autres aires cérébrales. En effet, nous avons vu au chapitre 3, section 3.1. du Vol I, qu'il existe deux grandes "voies" corticales, la première, pariétale, allant vers les aires du mouvement et de la motricité et concernant la localisation des objets et leurs relations spatiales, la seconde, ventrale, allant vers les aires temporales du langage et concernant l'identification des objets (reconnaissance de forme) et leurs propriétés prises en charge par leurs descriptions linguistiques. La première voie est appelée celle du "Where" par les neurophysiologistes et la seconde celle du "What". L'opposition "Where VS What" correspond tout à fait à celle entre intuitif et conceptuel chez Kant.

Il est donc, insistons-y une dernière fois, fort problématique de poser que la perception est réductible aux jugements perceptifs et que les contenus perceptifs sont conceptuels et de format propositionnel. Cette thèse ne semble pas être compatible avec les résultats des neurosciences, même si elle est encore très largement dominante.

2.6. Géométrie et physiologie

Il existe une longue et puissante tradition de discussions entre géométrie et physiologie. À propos du "Raumproblem" (chapitre 6, section 3.2.2) nous sommes revenus sur le débat entre Helmholtz et Riemann. Rappelons que dans sa conférence de 1868 "Über die Tatsachen, welche der Geometrie zu Grunde liegen" [249] écrite dans la foulée de son monumental traité d'*Optique physiologique* de 1867, Helmholtz répondait au grand texte de Riemann de 1854 "Über die Hypothesen, welche der Geometrie zu Grunde liegen" [464]. Il cherchait à retrouver la géométrie de l'espace sensible à partir du fait empirique de l'existence de corps rigides mobiles. Dans un premier temps, il crut que cela caractérisait la géométrie euclidienne si l'espace était infini (ce qui élimine des espaces elliptiques comme la sphère) mais Beltrami lui fit remarquer que cette propriété était partagée par les espaces hyperboliques de courbure constante < 0 . Quoi qu'il en soit, de Helmholtz à Poincaré, le lien entre physiologie et géométrie s'est souvent focalisé sur les propriétés mécaniques de l'espace physique.

En ce qui concerne la neurophysiologie, des travaux tout aussi remarquables, allant de Johannes Müller à Helmholtz en passant par la théorie des signes locaux (*Localzeichen*) d'Hermann Lotze conduisirent à l'idée que les fibres nerveuses du système visuel sont associées à des positions rétinienne précises, que chacune est adaptée à des signaux spécifiques, et que l'espace sensible est le résultat des interactions entre les sensations ainsi localisées. On trouve là les premières intuitions

des champs et profils récepteurs, de la rétinitopie et des interactions entre champs récepteurs à travers une architecture fonctionnelle. Mais aucune géométrie explicite correspondante n'est envisagée.

Inspirés par les théories physiques de leur époque, les pères fondateurs de la théorie de la Gestalt et de la phénoménologie pensaient qu'il fallait développer une sorte de "théorie des champs" qui permettrait d'expliquer les propriétés intégrales très particulières de la perception visuelle. L'intuition était remarquable mais les données expérimentales neuronales totalement inexistantes à l'époque et il était impossible de préciser les choses de façon détournée au moyen d'expériences psychologiques ou psychophysiques. Celles-ci ne pouvaient que fournir de nouveaux matériaux pour une théorie future. Avec le développement foudroyant des méthodes d'enregistrement et d'imagerie ces dernières décennies, la situation s'est totalement transformée. Nous nous sommes fondés sur ces résultats nouveaux et fins de neurophysiologie pour élaborer nos modèles et, en fin de compte, nous en arrivons bien à une *théorie des champs*. Les descriptions eidétiques de la phénoménologie, les expériences de la Gestalt, les résultats psychophysiques quantitatifs, tout converge, au niveau des mécanismes neuronaux sous-jacents où s'implémentent les algorithmes en cause, vers un concept de champ neuronal.

Cela permet de mieux comprendre notre proximité et, tout à la fois, notre distance avec la phénoménologie husserlienne. Convaincu qu'il ne pouvait pas, pour des raisons de principe, exister de géométrie interne "immanente" des vécus perceptifs,¹⁰ Husserl a conçu les structures génétiques constituantes des morphologies perceptives comme des structures *de sens* et a donc conçu son eidétique descriptive et sa noématique comme relevant d'une théorie de la signification alternative à une introuvable eidétique géométrique descriptive des vécus (la géométrie était selon lui le seul exemple d'eidétique existant avant la phénoménologie mais ne pouvait concerner que les "choses" du monde externe). Nous reprenons entièrement à notre compte toute la phénoménologie husserlienne de la perception et l'idée que la donation des objets perçus par des actes intentionnels objectivants est le résultat d'un processus dynamique de constitution qui, comme le rappellent Alain Berthoz et Jean-Luc Petit ([55], p. 82) :

"nous rendent possible de nous arracher à notre intériorité psychique et de rejoindre un objet connu à travers l'expérience."

Mais, partout, nous remplaçons les structures de sens par des structures géométriques, celles – inanticipables par Husserl – d'une géométrie "immanente" neuralemment implémentée.

Nous pourrions ainsi conclure que la neurogéométrie se propose de faire avec la neurophysiologie ce que des physiologistes comme Helmholtz et des géomètres

10. Sur la discussion husserlienne de la géométrie dans les *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*, cf. nos études [420] et [423].

comme Poincaré avaient fait avec la géométrie euclidienne et les mouvements de corps rigides. C'est la géométrie au sens de Sophus Lie et d'Élie Cartan qui se révèle être la géométrie la mieux adaptée au problème.

Or cette géométrie à la Lie-Cartan est aussi omniprésente dans les théories physiques modernes. Il n'est donc pas étonnant que nous trouvions, au niveau du formalisme, des analogies profondes – même si au départ elles furent pour nous tout à fait imprévues – entre la neurogéométrie et la théorie des champs où les champs de matière sont interprétés comme des sections de fibrés sur l'espace-temps externe E , les coordonnées des fibres F comme des degrés internes de liberté (et donc les groupes de symétrie des fibres comme des symétries internes) et les champs de jauge comme des connexions sur ces fibrés.

Dans nos modèles neurogéométriques, l'espace de base n'est évidemment pas l'espace-temps E mais le champ rétinien R , les fibres F ne sont évidemment pas des espaces de paramètres internes comme la charge électrique mais des espaces de traits géométriques comme l'orientation, mais il n'empêche que la "physique" du système (l'activité de $V1$) se trouve géométrisée par une connexion, à savoir le champ des plans de contact de $R \times F$. L'architecture fonctionnelle est un système d'interactions entre des propriétés locales (orientation, phase, échelle, dominance oculaire) variant avec la localisation, et la géométrie des percepts (contours illusoires, etc.) émerge de ces interactions décrites par la neurogéométrie. Autrement dit, la neurogéométrie peut se concevoir comme la théorie neuronale des champs sous-jacente à la géométrie "naïve" de la perception visuelle.

*
* * *

Nous avons commencé le Vol I en soulignant que des neurophysiologistes et des psychologues cognitivistes aussi éminents que John O'Keefe et Stanislas Dehaene avaient réhabilité Kant. Rappelons quelques citations.

En juin 2014, John O'Keefe (prix Nobel la même année pour sa découverte des "place cells" en 1971) a donné une conférence intitulée "Immanuel Kant : Pioneer neuroscientist" [393] où il affirma

"In his *Critique of Pure Reason*, Kant argued that our concept of space was not derived from sensations arising from our interaction with the physical world but instead represented the a priori basis for our perception of the world in the first place. Extensive work in modern neuroscience has provided strong evidence in support of this position."

En 2012, Stanislas Dehaene a proposé avec Elizabeth Brannon [139] un projet de recherche "kantien" "Space, time, and number : a Kantian research program" pour le Flagship "Future and Emerging Technologies" du *Human Brain Project*, l'un des

plus importants programmes européens de neurosciences. Il y affirma dans un projet associé de revue

“In his *Critique of Pure Reason*, Immanuel Kant famously argued that they (the representations of space, time and number) provide “a priori intuitions” that precede and structure how humans experience the environment. (...) The articles in this special issue all support this point of view : from grid cells to number neurons, the richness and variety of mechanisms by which animals and humans, including infants, can represent the dimensions of space, time and number is bewildering and suggests evolutionary processes and neural mechanisms by which Kantian intuitions might universally arise. (...) If Immanuel Kant were born today, he would probably be a cognitive neuroscientist !”

À propos du schématisme transcendantal Kant disait dans la *Critique de la raison pure* qu’il est

“un art caché dans les profondeurs de l’âme humaine, et dont nous aurons de la peine à arracher à la nature les secrets du fonctionnement pour les mettre à découvert sous les yeux.” (A 141, B 180-181, AK III, 136)¹¹

Le schématisme transcendantal reste encore une énigme, mais nous pensons que “l’art caché dans les profondeurs” *neuronales* “de l’âme humaine” qu’est la géométrie perceptive commence à être “mis à découvert sous les yeux” grâce à l’imagerie *in vivo* (c’est le contenu de notre volume 1) et mathématisé grâce à la neurogéométrie (c’est le contenu de ce volume 2).

11. “Dieser Schematismus unseres Verstandes, in Ansehung der Erscheinungen und ihrer bloßen Form, ist eine verborgene Kunst in den Tiefen der menschlichen Seele, deren wahre Handgriffe wir der Natur schwerlich jemals abraten, und sie unverdeckt vor Augen legen werden.”

Bibliographie

- [1] Abel, M.N.H., *Recherches sur les fonctions elliptiques*. Disponible sur le *Site du Prix Abel*, www.abelprisin.no.
- [2] Abraham, R., Marsden, J. , *Foundations of Mechanics*, Benjamin/Cummings, Reading, 1978.
- [3] Adams, J.F., Vector fields on spheres, *Annals of Mathematics*, 75/3 (1962) 603-632.
- [4] Adler, R., Taylor, J., *Random Fields and Geometry*, Springer, New York, 2007.
- [5] Agrachev, A.A., Barilari, D., Boscain, U., *A Comprehensive Introduction to Sub-Riemannian Geometry*, Cambridge University Press, 2020.
- [6] Agrachev, A.A., Boscain, U., Gauthier, J-P., Rossi, F., The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups, *Journal of Functional Analysis*, 256/8 (2009) 2621-2655.
- [7] Agrachev, A.A., Charlot, G., Gauthier, J-P., Zakalyukin, V.M., On sub-Riemannian caustics and wave fronts for contact distributions in the three-space, *J. of Dynamical and Control Systems*, 6/3 (2000) 365-395.
- [8] Agrachev, A.A., El Alaoui, El-H.C., Gauthier, J-P., Kupka, I., Generic singularities of Sub-Riemannian metrics on \mathbb{R}^3 , *CRAS (Comptes Rendus de l' Académie des Sciences de Paris)*, t. 322, série I (1996) 377-384.
- [9] Agrachev, A.A., Gamkrelidze, R.V., The Pontryagin Maximum Principle 50 years later, www.maik.ru/abstract/prsteks/6/prsteks0004_abstract.pdf.
- [10] Agrachev, A.A., Sachkov, Y.L., *Control Theory from the Geometric Viewpoint*, SISSA 77/2002/M, Trieste. *Encyclopedia of Mathematical Sciences*, Springer-Verlag, New York, Heidelberg, Berlin, 2004.
- [11] Airy, G.B., On the Intensity of Light in the neighbourhood of a Caustic, *Transactions of the Cambridge Philosophical Society*, 6 (1838) 379-402.
- [12] Akivis, M.A., Rosenfeld, B.A., *Élie Cartan (1869-1951)*, (transl. V.V. Goldberg), *Translations of Mathematical Monographs*, volume 123, American Mathematical Society, Providence, 1993.
- [13] Alekseev, A., Meinrenken, E., On the Kashiwara-Vergne conjecture, *Inventiones Mathematicae*, 164 (2006) 615-634.
- [14] Alekseevsky, D.V., Michor, P.W., Differential Geometry of Cartan Connections, *Publicationes Mathematicae Debrecen*, 47/3-4 (1995) 349-375. arXiv :math/9412232v1
- [15] Alekseevsky, D.V., Spiro, A.F., *Conformal models for hypercolumns and the visual V1 cortex*, arXiv :2202.10157v2
- [16] Annaratone, S., Les premières démonstrations de la formule intégrale de Fourier, *Revue d'histoire des mathématiques*, 3 (1997), 99-136.

- [17] Andler, M., Théorie des représentations de $GL(2, \mathbb{R})$, *Les représentations linéaires et le grand théorème de Fermat*, Éditions de l'École Polytechnique, Palaiseau, 2009, 19-83.
- [18] Ardentov, A., Sachkov, Y., Extremal trajectories in a nilpotent sub-Riemannian problem on the Engel group, *Sbornik : Mathematics*, 202/11 (2011) 1593–1615.
- [19] Arnold, V.I., Remarks on the Stationary Phase Method and Coxeter Numbers, *Russian Mathematical Surveys*, 28 (1973) 19-48.
- [20] Arnold, V.I., *Méthodes mathématiques de la Mécanique classique*, Editions Mir, Moscou, 1976.
- [21] Atick, J., Could Information theory provide an ecological theory of sensory processing?, *Network*, 3 (1992) 213-251.
- [22] Azencott, R., (éd.), Géodésiques et diffusions en temps petit, *Astérisque*, 84-85, Société Mathématique de France, 1981.
- [23] Bachman, D., *A Geometric Approach to Differential Forms*, Birkhäuser, Basel, 2012.
- [24] Barbaresco, F., Gay-Balmaz, F., Lie Group Cohomology and (Multi)Symplectic Integrators : New Geometric Tools for Lie Group Machine Learning Based on Souriau Geometric Statistical Mechanics, *Entropy* 22/5 (2020) 498, 1-52.
- [25] Barbieri, S., *The quantum mechanical Segal-Bargmann transform using Jordan algebras*, Thesis, Universiteit Gent, 2014.
- [26] Barbieri, D., *Reconstructing group wavelet transform from feature maps with a reproducing kernel iteration*, arXiv :2110.00600, 2021.
- [27] Barbieri, D., Citti, G., Sanguinetti, G., Sarti, A., *Coherent states of the Euclidean group and activation regions of primary visual cortex*, arXiv :1111.0669v1, 2011.
- [28] Barbieri, D., Citti, G., Sanguinetti, G., Sarti, A., An uncertainty principle underlying the functional architecture of V1, *Journal of Physiology Paris* 106/5-6 (2012) 183-193.
- [29] Bargmann, V., On a Hilbert space of analytic functions and an associated integral transform, *Communications on Pure and Applied Mathematics*, Part I, 14/3 (1961) 187-214.
- [30] Bargmann, V., On a Hilbert space of analytic functions and an associated integral transform, *Communications on Pure and Applied Mathematics*, Part II. 20/1 (1967) 1-101.
- [31] Barilari, D., Boscain, U., Neel, R. W., Small-time heat kernel asymptotics at the sub-Riemannian cut locus, *Journal of Differential Geometry*, 92 (2012), 373–416.
- [32] Barilari, D., Jendrej, J., *Small time heat kernel asymptotics at the cut locus on surfaces of revolution*, arXiv :1211.1811, 2013.
- [33] Barilari, D., Rizzi, L., *On Jacobi fields and a canonical connection in sub-Riemannian geometry*, arXiv :1506.01827v3, 2017.
- [34] Beals, R., Greiner, P.C., *Calculus on Heisenberg manifolds*, Annals of Mathematics Studies, vol. 119, Princeton University Press, Princeton, NJ, 1988.
- [35] Beals, R., Gaveau, B., Greiner, P.C., Hamilton-Jacobi theory and the heat kernel on Heisenberg groups, *Journal de Mathématiques Pures et Appliquées*, 79/7 (2000) 633-689.
- [36] Bellaïche, A., The tangent space in sub-Riemannian geometry, *Sub-Riemannian Geometry*, (A. Bellaïche, J. Risler, eds.), *Progress in Mathematics*, 144, Birkhäuser, Basel, 1996, 4-78.
- [37] Ben Arous, G., Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus, *Annales Scientifiques de l'École Normale Supérieure*, 21/3 (1988) 307-331.

- [38] Ben Arous, G., Développement asymptotique du noyau de la chaleur hypoelliptique sur la diagonale, *Annales de l'Institut Fourier*, 39/1 (1989) 73-99.
- [39] Benci, V., Di Nasso, M., Alpha Theory : Mathematics with Infinite and Infinitesimal Numbers, 2012, <http://people.dm.unipi.it/dinasso/papers/14.pdf>
- [40] Bennequin, D., *Entrelacements et équations de Pfaff*, Thèse d'État, Université de Paris VII, sous la direction d'Alain Chenciner, 1982.
- [41] Bennequin, D., Entrelacements et équations de Pfaff, *IIIe Rencontre de Géométrie du Schnepfenried* (1982), *Astérisque* 107-108 (1983) 87-161.
- [42] Bennequin, D., Caustique mystique, *Séminaire Bourbaki*, n° 634, *Astérisque*, 133-134 (1986) 19-56.
- [43] Bennequin, D., Topologie symplectique, convexité holomorphe et structures de contact (d'après Y. Eliashberg, D. Mc Duff *et al.*), *Séminaire Bourbaki*, n° 725, *Astérisque*, 189-190 (1990) 285-323.
- [44] Bennequin, D., Fuchs, R., Berthoz, A., Flash, T., Movement Timing and Invariance Arise from Several Geometries, *PLoS Comput Biol*, 5(7) : e1000426.
- [45] Ben-Shahar, O., Huggins, P.S., Izo, T., Zucker, S., Cortical connections and early visual function : intra- and inter-columnar processing, *Neurogeometry and Visual Perception*, (J. Petitot, J. Lorenceau, eds.), *Journal of Physiology - Paris*, 97 (2003) 191-208.
- [46] Berceanu, S., Coherent states and geodesics : cut locus and conjugate locus, *Journal of Geometry and Physics*, 21 (1997) 149-168.
- [47] Berestycki, H., Nirenberg, L., Travelling fronts in cylinders, *Annales de l'Institut Henri Poincaré* (Analyse non linéaire), 9 (1992) 497-572.
- [48] Berkovitz, L. D., *Optimal Control Theory*, Applied Mathematical Sciences, Springer-Verlag, New York, Heidelberg, Berlin, 1974.
- [49] Bernier, D., Taylor, K.F., Wavelets from Square-Integrable Representations, *SIAM Journal on Mathematical Analysis*, 27/2 (1996) 594-608.
- [50] Bernouilli, Ja., *Der Briefwechsel von Jacob Bernouilli* (André Weil, ed.), Birkhäuser, Basel, 1993 (avec une contribution de C. Truesdell sur la Mécanique et l'Élasticité dans la correspondance avec Leibniz).
- [51] Bernouilli, Ja. et Jo., (1694), *Die Streitschriften von Jacob und Johann Bernouilli*, (H.H. Goldstine, ed.), Birkhäuser, Basel, 1991.
- [52] Berry, M.V., Upstill, C., Catastrophe Optics : Morphologies of Caustics and their Diffraction Patterns, *Progress in Optics*, (E. Wolf, ed.), North-Holland, Amsterdam, 1980, 258-345.
- [53] Berthoz, A., *Le sens du mouvement*, Odile Jacob, Paris, 1997.
- [54] Berthoz, A., *La simplicité*, Odile Jacob, Paris, 2009.
- [55] Berthoz, A., Petit, J.-L., *Phénoménologie et physiologie de l'action*, Odile Jacob, Paris, 2006.
- [56] Bianchi, L., *Memorie di Matematica e di Fisica della Societa Italiana delle Scienze*, Serie III, Tomo XI, 267-352, 1898. On the three-dimensional spaces which admit a continuous group of motions, *General Relativity and Gravitation*, 33/12 (2001) 2157-2170 (Editor's note) & 2171-2253.
- [57] Blau, M., *Symplectic Geometry and Geometric Quantization*, Lecture Notes, Albert Einstein Center for Fundamental Physics, Bern University, 1992.

- [58] Block, N., On a confusion about a function of consciousness, *The Nature of Consciousness*, (N. Block, O. Flanagan, G. Guzeldere, eds.), MIT Press, Cambridge MA, 1997.
- [59] Blum, H., Biological Shape and Visual Science, *Journal of Theoretical Biology*, 38 (1973) 205-287.
- [60] Bolelli, M.V., Citti, C., Sarti, A., Zucker, S.W., *Good continuation in 3D : the neurogeometry of stereo vision*, arXiv :2301.04542, 2023.
- [61] Bonfiglioli, A., Lanconelli, E., Uguzzoni, F., *Stratified Lie Groups and Potential Theory for their sub-Laplacians*, Springer, Berlin, 2007.
- [62] Bonhöffer, T., Grinvald, A., Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns, *Nature*, 353 (1991) 429-431.
- [63] Bonnard, B., Caillaud, J.-B., Sinclair, R., Tanaka, M., Conjugate and cut loci of a two-sphere of revolution with application to optimal control, *Annales de l'Institut Henri Poincaré*, 26 (2009) 1081-1098.
- [64] Bonnet, P.-O., *Astronomie sphérique* (cours de 1887, J. Blondin, Guillet, A., édés), Georges Carré, Paris, 1889.
- [65] Borel, A., Hermann Weyl and Lie groups, *Hermann Weyl 1885-1985 : Centenary Lectures* (K. Chandrasekharan, ed.) Springer, Berlin-Heidelberg, 1986, 53-82.
- [66] Born, M., *Untersuchungen über die Stabilität der elastischen Linie in Ebene und Raum, unter verschiedenen Grenzbedingungen* (*Stability of elastic lines in the plane and the space*), Thesis, Göttingen, 1906.
- [67] Boscain, U., Rossi, F., Invariant Carnot-Carathéodory metrics on S^3 , $SO(3)$, $SL(2)$ and Lens Spaces, *Preprint SISSA 58/2007/M*, Trieste, 2007.
- [68] Boscain, U., Piccoli, B. *An Introduction to Optimal Control*, <http://www.cmapx.polytechnique.fr/~boscain/AUTOMATICS/introduction-to-optimal-control.pdf>
- [69] Boscain, U., Duits, R., Rossi, F., Sachkov, Y., *Curve cusplless reconstruction via sub-Riemannian geometry*, arXiv : 1203.3089v4, 2013.
- [70] Boscain, U., Chertovskih, R., Gauthier, J-P., Remizov, A., Hypoelliptic diffusion and human vision : a semi-discrete new twist in the Petitot theory, *SIAM Journal on Imaging Sciences*, 7/2 (2014) 669-695.
- [71] Bosking, W.H., Zhang, Y., Schoenfield, B., Fitzpatrick, D., Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex, *Journal of Neuroscience* 17/6 (1997) 2112-2127.
- [72] Bourgeois, F., A survey of contact homology, *New Perspectives and Challenges in Symplectic Field Theory*, CRM Proceedings & Lecture Notes, American Mathematical Society, 49 (2009) 45-72.
- [73] Bourgeois, F., Lectures on Symplectic and Contact Homology, *Proceedings of Gökova Geometry-Topology Conference 2010*. <https://gokovagt.org/proceedings/2010/>
- [74] Bourguignon, J.-P., *Calcul variationnel*, Cours de l'École Polytechnique, Paris, 1993.
- [75] Brady, M., Grimson, W.E.L., Langridge, D., Shape encoding and subjective contours, *Proceedings of the AAAI*, Stanford University, 1980, 15-17.
- [76] Bressan, P., Mingolla, E., Spillmann, L., Watanabe, T., Neon color spreading : a review, *Perception*, 26 (1997) 1353-1366.

- [77] Bressloff, P., Cowan, J., Golubitsky, M., Thomas, P., Scalar and pseudoscalar bifurcations motivated by pattern formation on the visual cortex, *Nonlinearity*, 14 (2001) 739-775.
- [78] Bressloff, P., Cowan, J., Golubitsky, M., Thomas, P., Wiener, M., Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex, *Philosophical Transactions of the Royal Society of London B*, 356 (2001) 299-330.
- [79] Bressloff, P., Cowan, J., The functional geometry of local and horizontal connections in a model of V1, *Neurogeometry and Visual Perception* (J. Petitot, J. Lorenceau, eds.), *Journal of Physiology-Paris*, 97, 2-3 (2003) 221-236.
- [80] Brillouin, L., Remarques sur la mécanique ondulatoire, *Journal de Physique et le Radium*, 7 (1926) 353-368.
- [81] Bruce, I., *Some Mathematical Works of the 17th & 18th Centuries, including Newton's Principia, Euler's Mechanica, Introductio in Analysin, etc., translated mainly from Latin into English*, www.17centurymaths.com.
- [82] Bryant, R.L., Griffiths, P.A., Reduction for constrained variational problems and $\int \frac{\kappa^2}{2} ds$, *American Journal of Mathematics*, 108 (1986) 525-570.
- [83] Buliga, M., *Sub-Riemannian Geometry and Lie Groups*, arXiv :math/0210189v3, 2002.
- [84] Buser, P., Imbert, M., *Vision*, Hermann, Paris, 1987.
- [85] Byrd, P.F., Friedman, M.D., *Handbook of Elliptic Integrals for Engineers and Physicists*, Springer, Berlin, Heidelberg, 1954.
- [86] Calin, O., Mangione, V., Variational calculus on sub-Riemannian manifolds, *Balkan Journal of Geometry and Its Applications*, 8/1 (2003) 21-32.
- [87] Calkin, M.G., *Lagrangian and Hamiltonian Mechanics*, World Scientific, Singapour, 1996.
- [88] Cantù, P., Grassmann's Concept Structuralism, *The Prehistory of Mathematical Structuralism* (Reck, E.H., Schiemer, G. eds), Oxford University Press, 2020, 21-58.
- [89] Carlson, B.C., *Special Functions of Applied Mathematics*, Academic Press, New York, 1977.
- [90] Cartan, É., Sur certaines expressions différentielles et le problème de Pfaff, *Annales Scientifiques de l'École Normale Supérieure*, 16 (1899) 239-332.
- [91] Cartan, É., L'intégration des systèmes d'équations aux différentielles totales, *Annales Scientifiques de l'École Normale Supérieure*, 18 (1901) 241-311.
- [92] Cartan, É., Les systèmes de Pfaff à cinq variables et les équations aux dérivées partielles du second ordre, *Annales Scientifiques de l'École Normale Supérieure*, 27 (1910) 109-192.
- [93] Cartan, É., Sur l'équivalence absolue de certains systèmes d'équations différentielles et sur certaines familles de courbes, *Bulletin de la Société Mathématique de France*, XLII (1914) 1-36.
- [94] Cartan, É., Sur une généralisation de la notion de courbure de Riemann et les espaces à torsion, *Comptes Rendus de l'Académie des Sciences de Paris*, 174 (1922) 593-595.
- [95] Cartan, É., Sur les variétés à connexion affine, et la théorie de la relativité généralisée, *Annales Scientifiques de l'École Normale Supérieure*, 40 (1923) 325-412, 41 (1924) 1-25, 42 (1925) 17-88.
- [96] Cartan, É., Le rôle de la théorie des groupes de Lie dans l'évolution de la géométrie moderne, *Comptes rendus du Congrès international des mathématiciens d'Oslo*, 92-103, 1936.

- [97] Cartan, É., *La théorie des groupes finis et continus et la géométrie différentielle traitées par la méthode du repère mobile*, Gauthier-Villars, Paris, 1937.
- [98] Cartan, É., Notice sur les travaux scientifiques, *Selecta* (1939) 219-272. Repris dans *Œuvres Complètes*, Partie I, Volume I, 15-98.
- [99] Cartan, É., *Œuvres complètes*, Paris, Gauthier-Villars, (publiées avec le concours du CNRS), 1952-1955.
- [100] Cassou-Noguès, P., *De l'expérience mathématique. Essai sur la philosophie des sciences de Jean Cavaillès*, Vrin, Paris, 2001.
- [101] Cervetto, L., Demontis, G.C., Gargini, C., Cellular mechanisms underlying the pharmacological induction of phosphenes, *British Journal of Pharmacology*, 150 (2007) 383-390.
- [102] *1830-1930 : A Century of Geometry. Epistemology, History, Mathematics* (L. Boi, D. Flament, J.-M. Salanskis, eds), Springer-Verlag, Berlin, 1992.
- [103] Chalmers, D., *The Conscious Mind : in Search of a Fundamental Theory*, Oxford University Press, 1997.
- [104] Chalmers, D., What is a Neural Correlate of Consciousness ?, *Neural Correlates of Consciousness : Empirical and Conceptual Questions*, (T. Metzinger, ed.), MIT Press, Cambridge MA, 2000.
- [105] Chang, D.-C., Greiner, P., Harmonic Analysis and Subriemannian Geometry on Heisenberg Groups, *Bulletin of the Institute of Mathematics, Academia Sinica*, 30/3 (2002) 153-190.
- [106] Chaperon, M., Commentaires au texte [528] de René Thom "Sur la théorie des enveloppes", *Œuvres mathématiques*, II, 145-154, Société Mathématique de France, Paris, 2019.
- [107] Chaperon, M., Commentaires au texte [531] de René Thom "Sur les équations différentielles multiformes et leurs intégrales singulières", *Œuvres Mathématiques*, III, 182-188, Société Mathématique de France, Paris, 2021.
- [108] Châtelet, G., La capture de l'extension comme dialectique géométrique : dimension et puissance selon l'*Ausdehnungslehre* de Grassmann (1844), *1830- 1930 : A Century of Geometry. Epistemology, History and Mathematics*, (L. Boi, D. Flament, J.-M. Salanskis, eds), Springer, New York, 1992, 222-244.
- [109] Cheaito, M., Mormul, P., Rank-2 distributions satisfying Goursat condition : all their local models in dimension 7 and 8, *ESAIM : Control, Optimisation and Calculus of Variations*, 4 (1999) 137-158.
- [110] Chenciner, A., Travaux de Thom et Mather sur la Stabilité Topologique, *Séminaire Bourbaki* 424 , 1973.
- [111] Chenciner, A., Singularités des Fonctions Différentiables, *Encyclopædia Universalis*, Paris, 1980.
- [112] Chenciner, A., Poincaré and the Three-Body Problem, *Séminaire Poincaré (Bourbaphy)* XVI, 2012, 45-133.
- [113] Chenciner, A., Montgomery, R., A remarkable periodic solution of the three-body problem in the case of equal masses, *Annals of Mathematics*, 152 (2000) 881-901.
- [114] Chern, S.S., Chevalley, C., Élie Cartan and his mathematical work, *Bulletin of the American Mathematical Society*, 58 (1952) 219-272.
- [115] Chern, S.S., Hamilton, R.S., On Riemannian metrics adapted to three-dimensional manifolds, *Lecture Notes in Mathematics*, 1111, Springer, Berlin, New York, 1985, 279-308.

- [116] Chevalley, C., Weil, A., Hermann Weyl (1885-1955), *L'Enseignement Mathématique*, 2/3 (1957) 157-187.
- [117] Cho, J.K., Geometry of CR-manifolds of contact type, *Proceedings of the eighth International Workshop on Differential Geometry*, 8 (2004) 137–155.
- [118] Chossat, P., Faugeras, O., *Hyperbolic planforms in relation to visual edges and textures perception*, arXiv :0907.0963v3.
- [119] Christensen, O., Frames, Riesz bases, and discrete Gabor/wavelet expansions, *Bulletin of the American Mathematical Society*, 38/3 (2001) 273-291.
- [120] Christoffel, E.B., Über die Transformation der homogenen Differentialausdrücke zweiten Grades, *Journal de Crelle*, LXX (1869) 46–70.
- [121] Citti, G., Sarti, A., A cortical based model of perceptual completion in the roto-translation space, *Journal of Mathematical Imaging and Vision*, 24/3 (2006) 307-326.
- [122] Clairaut, A.C., *Théorie de la Figure de la Terre, Tirée de l'Hydrostatique*, David Fils, Paris, 1743.
- [123] Clebsch, A., Über das Pfaffsche Problem, *Journal de Crelle*, LX (1862) 193-251, LXI (1863) 146-179.
- [124] Cohen-Tannoudji, G., Spiro, M., *La Matière-Espace-Temps*, Fayard, Paris, 1986.
- [125] Connes, A., *Noncommutative Geometry*, Academic Press, Cambridge (MA), 1994.
- [126] Connes, A., Gravity Coupled with Matter and the Foundation of Noncommutative Geometry, *Communications in Mathematical Physics*, 182 (1996) 155-176. ArXiv :hep-th/9603053.
- [127] Coquereaux, R., Espaces fibrés et connexions. Une introduction aux géométries classiques et quantiques de la physique théorique, <http://www.cpt.univ-mrs.fr/~coque/EspacesFibresCoquereaux.pdf>
- [128] CST, *Contact and Symplectic Topology*, (V. Colin, A. Stipsicz, eds), Springer, 2014.
- [129] Curto, C., Itskov, V., Veliz-Cuba, A., Youngs, N.E., *The neural ring : an algebraic tool for analyzing the intrinsic structure of neural codes*, arXiv :1212.4201v2, 2013.
- [130] Cygan, J., Heat kernels for class 2 nilpotent groups, *Studia Mathematica*, 64 (1979) 227-238.
- [131] d'Alembert, J. Le Rond, 1754-1765, Articles “Différentiel” et “Limite”, *Encyclopédie*, vol. IV, 985-989, vol. IX, 542.
- [132] Dangelmayr, G., Güttinger, W., Veit, W., Semi-classical Path Integrals in Terms of Catastrophes, *Structural Stability in Physics*, (W. Güttinger, H. Eikemeier eds.), Springer, Berlin, Heidelberg, New-York, 1979, 104-121.
- [133] Darboux, G., Sur le problème de Pfaff, *Bulletin des Sciences Mathématiques et Astronomiques*, VI (1882) 14-36 et 49-68.
- [134] Dat, J.-F., Groupes et Algèbres de Lie, <https://webusers.imj-prg.fr/~jean-francois.dat/enseignement/GAL/GALM2.pdf>
- [135] d'Espagnat, B., Zwirn, H., (éds), *Le monde quantique. Les débats philosophiques de la physique quantique*, Éditions Matériologiques, Paris, 2014.
- [136] DeAngelis, G.C, Ghose, G.M, Ohzawa, I, Freeman, R.D, Functional Micro-Organization of Primary Visual Cortex : Receptive Field Analysis of Nearby Neurons, *Journal of Neuroscience*, 19/9 (1999) 4046–4064.

- [137] DeAngelis, G.C., Ozhawa, I., Freeman, R.D., Receptive-field dynamics in the central visual pathways, *Trends in Neuroscience* 18/10 (1995) 451-458.
- [138] De Bièvre, S., Coherent states over symplectic homogeneous spaces, *Journal of Mathematical Physics*, 30/7 (1989) 1401-1407.
- [139] Dehaene, S., Brannon, E.M., Space, time, and number : a Kantian research program, *Trends in Cognitive Sciences*, Special Issue : *Space, Time and Number*, 14/12 (2010) 517-519.
- [140] Dehornoy, P., *La théorie des ensembles*, Calvage & Mounet, Paris, 2017.
- [141] Desolneux, A., Moisan, L., Morel, J.-M., Computational gestalts and perception thresholds, *Neurogeometry and Visual Perception* (J. Petitot, J. Lorenceau, eds.), *Journal of Physiology-Paris*, 97/2-3 (2003) 311-324.
- [142] Desolneux, A., Moisan, L., Morel, J.-M., *From Gestalt Theory to Image Analysis : A Probabilistic Approach*, Interdisciplinary Applied Mathematics, vol.34, Springer-Verlag, Heidelberg, New York, 2008.
- [143] Dhombres, J., La méthode fonctionnelle chez J.F. Pfaff : une filiation leibnizienne, *Sciences et Techniques en Perspective*, Université de Nantes, 26 (1993) 97-147.
- [144] Dieudonné, J., *Cours de Géométrie algébrique*, Presses Universitaires de France, Paris, 1974.
- [145] DLMF : NIST, *Digital Library of Mathematical Functions*, <https://dlmf.nist.gov>.
- [146] Doddoli, A., Montgomery, R., *Geodesics in jet-space*, arXiv :2109.13835, 2022.
- [147] Doob, J.L., The Brownian movement and stochastic equations, *Annals of Mathematics*, 43 (1942) 351-369.
- [148] Douady, A., Systèmes dynamiques holomorphes, *Séminaire Bourbaki*, n° 599, 39-63, *Astérisque*, 105-106, 1983.
- [149] Douady, A., Nœuds et structures de contact en dimension 3 (d'après Daniel Bennequin), *Séminaire Bourbaki* n° 604, *Astérisque*, 105-106 (1983) 129-148.
- [150] Doubrov, B.M., Komrakov, B.P., Contact Lie algebras of vector fields on the plane, *Geometry and Topology*, 3 (1999) 1-20.
- [151] Dresp, B., Bonnet, C., Psychophysical evidence for low level processing of illusory contours and surfaces in the Kanizsa square, *Vision Research*, 31 (1991) 1813-1817.
- [152] Dresp, B., Bonnet, C., Psychophysical measures of illusory form perception : further evidence for local mechanisms, *Vision Research*, 33/5-6 (1993) 759-766.
- [153] Dresp, B., Bonnet, C., Subthreshold summation with illusory contours, *Vision Research*, 35 (1995) 1071-1078.
- [154] Dresp, B., Grossberg, S., Contour integration across polarities and spatial gaps : from local contrast filtering to global grouping, *Vision Research*, 37/7 (1997) 913-924.
- [155] Duistermaat, J.J., Oscillatory integrals, Lagrange immersions and unfolding of singularities, *Communications on Pure and Applied Mathematics*, 27/2, (1974) 207-281.
- [156] Duits, R., van Almsick, M., The explicit solutions of linear left-invariant second order stochastic evolution equations on the 2D Euclidean motion group, *Quarterly of Applied Mathematics*, 66 (2008) 27-67.
- [157] Duits, R., Franken, E., Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores, I et II, *Quarterly of Applied Mathematics*, 68/2 (2010) 255-292 et 293-331.

- [158] Duits, R., Boscain, U., Rossi, F., Sachkov, Y., Association Fields via Cuspless Sub-Riemannian Geodesics in $SE(2)$, *Journal of Mathematical Imaging and Vision*, 49 (2014) 384–417.
- [159] Duits, R., Ghosh, A., Dela Haije, T., Sachkov, Y., Cuspless Sub-Riemannian Geodesics within the Euclidean Motion Group $SE(d)$, *Neuromathematics of Vision*, (G. Citti, A. Sarti, eds.), Springer, Berlin, New York, 2014, 173–215.
- [160] Dupuy, J.-P., *Aux origines des sciences cognitives*, La Découverte, Paris, 1994. *Mechanization of the Mind : on the Origins of Cognitive Sciences*, Princeton University Press, Princeton, 2000.
- [161] Dušek, Z., Survey on homogeneous geodesics, *Note di Matematica*, 1 (2008), suppl. n. 1, 147–168.
- [162] Eckes, C., Weyl et la géométrisation de la physique, *Revue d'Histoire des Mathématiques*, 20/1 (2014) 95–170.
- [163] Einstein, A., Auf die Riemann-Metrik und den Fern-Parallelismus gegründete einheitliche Feldtheorie, *Mathematische Annalen*, 102 (1930) 685–697.
- [164] Einstein, A., Théorie unitaire du champ physique, *Annales de l'Institut Henri Poincaré*, 1/1 (1930) 1–24.
- [165] Eliashberg, Y., Classification of overtwisted contact structures on 3-manifolds, *Inventiones Mathematicae*, 98 (1989) 623–637.
- [166] Eliashberg, Y., Contact 3-manifolds twenty years since J. Martinet's work, *Annales de l'Institut Fourier*, 42 (1992), 165–192.
- [167] Eliashberg, Y., Givental, A., Hofer, H., Introduction to Symplectic Field Theory, *Geometric and Functional Analysis*, Special Volume, Part II, Birkhäuser, Basel, 2000, 560–673. ArXiv :math/0010059, 2000.
- [168] Engel, F., Zur Invariantentheorie der Systeme von Pfaff'schen Gleichungen, *Leipzig Berichte Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften Mathematisch-Physikalische*, Klasse XLI (1889) 157–176.
- [169] Engel, F., Lie, S., *Theorie der Transformationsgruppen*, Erster-Zweiter-Dritter und Letzter Abschnitt, Teubner, Leipzig, Berlin, 1888, 1890, 1890. Reprinted by Chelsea Publishing Co, New York, 1970.
- [170] Ermentrout, G.B., Cowan, J.D., A Mathematical Theory of Visual Hallucinations, *Biological Cybernetics*, 34 (1979) 137–150.
- [171] Ermentrout, G.B., Kopell, N., Frequency plateaus in a chain of weakly coupled oscillators, *SIAM J. Math. Anal.*, 15, 2 (1984) 215–237.
- [172] Etnyre, J.B., *Introductory Lectures on Contact Geometry*, arXiv :math/0111118, 2001.
- [173] Euler, L., *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici lattissimo sensu accepti*, 1744. Référence E65 sur le site *Euler Archive* [176] et *Opera Omnia* : Series 1, Volume 24. Trad. All. A. Aycock et A. Diener.
- [174] Euler, L., De curvis elastica, *Additamentum 1* du *Methodus* [173], 1744. Trad. Angl. [394].
- [175] Euler, L., *Mathematische Werke*, Leipzig, Berlin, 1912.
- [176] Euler Archive : eulerarchive.maa.org.

- [177] Evans, L.C., *An Introduction to Stochastic Differential Equations*, American Mathematical Society, Providence, 2013.
- [178] Fagnano, G.C. (comte de), *Produzioni matematiche*, Tome II, Pesaro, 1750. In *Opere matematiche*, Società Editrice Dante Alighieri, Milan-Rome-Naples, 1911.
- [179] Fatou, P., Séries trigonométriques et séries de Taylor, *Acta Mathematica*, 30 (1906) 335-400.
- [180] Faulhuber, M., Shafkulovska, I., *Gabor frame bound optimization*, arXiv :2204.02917, 2022.
- [181] Fellmann, E., *Leonhard Euler*, Rowohlt Taschenbuch, 1995.
- [182] Field, D.J., Hayes, A., Hess, R.F., Contour integration by the human visual system : evidence for a local “association field”, *Vision Research*, 33/2 (1993) 173-193.
- [183] Finkel, L.H., Edelman, G.M., Integration of distributed cortical systems by reentry : a computer simulation of interactive functionally segregated visual areas, *Journal of Neuroscience*, 9/9 (1989) 3188-3208.
- [184] Fitzpatrick, D., The functional organization of local circuits in visual cortex : insights from the study of tree shrew striate cortex, *Cerebral Cortex*, 6 (1996) 329-341.
- [185] Flament, D., H. G. Grassmann et l’introduction d’une nouvelle discipline mathématique : L’*Ausdehnungslehre*, *Philosophia Scientiæ*, 5 (2005) 81-141.
- [186] Florack, L.M.J., *The Syntactical Structure of Scalar Images*, PhD, University of Utrecht, 1993.
- [187] Folland, G.B., *Harmonic Analysis in Phase Space*, Princeton University Press, Princeton, 1989.
- [188] Forsyth, A.R., *Theory of Differential Equations, Part I : Exact equations and Pfaff’s problem*, Cambridge University Press, 1890.
- [189] Fourier, J-B.J., *Théorie analytique de la chaleur*, Firmin Didot, Paris, 1922 (fac-similé chez Jacques Gabay, accessible sur Gallica).
- [190] Fowler-Wright, A., *The Classification of Three-dimensional Lie Algebras*, Thesis, University of Warwick, 2014.
- [191] Fraser, C., J. L. Lagrange’s Early Contributions to the Principles and Methods of Mechanics, *Archive for History of Exact Science*, 28/3 (1983) 197-241.
- [192] Fraser, C., J. L. Lagrange’s Changing Approach to the Foundations of the Calculus of Variations, *Archive for History of Exact Science*, 32/2 (1985) 151-191.
- [193] Fraser, C., The Calculus as Algebraic Analysis : Some Observations on Mathematical Analysis in the 18th Century, *Archive for History of Exact Science*, 39/4 (1989), 317-335.
- [194] Frégnac, Y., Shulz, D., Activity-dependent regulation of receptive field properties of cat area 17 by supervised Hebbian learning, *Journal of Neurobiology*, 41/1 (1999) 69-82.
- [195] Frégnac, Y., Neurogeometry and entoptic visions of the functional architecture of the brain, *Neurogeometry and Visual Perception*, (J. Petitot, J. Lorenceau, eds.), *Journal of Physiology - Paris*, 97/2-3 (2003) 87-92.
- [196] Frégnac, Y., Big data and the industrialization of neuroscience : A safe roadmap for understanding the brain ?, *Science*, 358 (2017) 470-477.
- [197] Frégnac, Y., How Blue is the Sky ?, *eNeuro*, <https://doi.org/10.1523/ENEURO.0130-21.2021>
- [198] Fresneda, R., Gazeau, J.-P., Noguera, D., Quantum localisation on the circle, arXiv :1708.03693v1, 2017.

- [199] Friedman, M., *Kant's Construction of Nature*, Cambridge University Press, 2013.
- [200] Frobenius, F.G., Über das Pfaff'sche Problem, *Journal de Crelle*, LXXXII (1877) 230-315.
- [201] Frobenius, F.G., Über homogene totale Differentialgleichungen, *Journal de Crelle*, LXXXVI (1879) 1-19.
- [202] Fulton, W., Harris, J., *Representation Theory. A First Course*, Springer, New York, 2004.
- [203] Gamkrelidze, R., *Principles of Optimal Control Theory* (édition russe Tbilisi University Press), Plenum Press, New York, 1978.
- [204] Gårding, L., Hyperbolic Equations in the Twentieth Century, *Matériaux pour l'histoire des Mathématiques au XX^e siècle*, Actes du colloque à la mémoire de Jean Dieudonné (Nice 1996), Société Mathématique de France, Paris, 1998, 41-72.
- [205] Gauss, C.F., Compte rendu de Pfaff [447], *Göttingische gelehrte Anzeigen* (1815) 1035-1088. *Gesammelte Werke*, t.III, 231-241, Königlichen Gesellschaft der Wissenschaften zu Göttingen.
- [206] Gaveau, B., Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents, *Acta Mathematica*, 139 (1977) 96-153.
- [207] Gaveau, B., Équation de la chaleur sur les variétés d'Heisenberg, *Séminaire Institut Henri Poincaré*, 26/10/2005.
- [208] Gelfand, I. M., Fomin, S. V., *Calculus of Variations*, Dover Publications, Inc., New York, 1963.
- [209] Georges, S., Serié, P., Frégnac, Y., Lorenceau, J., Orientation-dependent modulation of apparent speed : psychophysical evidence, *Vision Research*, 42 (2002) 2557-2572.
- [210] Georgopoulos, A.,P., Columnar organization of the motor cortex : direction of movement, *Recent Advances on the Modular Organization of the Cortex*, (M.F. Casanova, I. Opris, eds.), Springer, 2015, 123-141.
- [211] Giaro, A., Kumpera, A., Ruiz, C., Sur la lecture correcte d'un résultat d'Élie Cartan, *CRAS (Comptes Rendus de l'Académie des Sciences de Paris)*, 287 (1978) 241-244.
- [212] Gibson, J.J., *The Ecological Approach to Visual Perception*, Houghton-Mifflin, Boston, 1979.
- [213] Giorello, G., *Lo Spettro e il Libertino*, Mondadori, Milan, 1985.
- [214] Giroux, É., Convexité en topologie de contact, *Commentarii Mathematici Helvetici*, 66 (1991) 637-677.
- [215] Godement, R., *Topologie Algébrique et théorie des faisceaux*, Hermann, Paris, 1960
- [216] Goldstine, H.H., *A History of the Calculus of Variations from the 17th through the 19th Century*, Springer, New York, 1980.
- [217] Goursat, É., Sur certains systèmes d'équations aux différentielles totales et sur une généralisation du problème de Pfaff, *Annales de la Faculté des Sciences de l'Université de Toulouse*, 7 (1915) 1-58.
- [218] Goursat, É., *Leçons sur le problème de Pfaff*, Hermann, Paris, 1922.
- [219] Grassmann, H.G., *Die lineale Ausdehnungslehre*, Leipzig, Wiegand, 1844. Traduction anglaise L. Kannenberg, *A new branch of mathematics*, Open Court, Chicago, 1995.
- [220] Grassmann, H.G., *Die Ausdehnungslehre. Vollständig und in strenger Form bearbeitet*, Berlin, Enslin, 1862. Traduction anglaise L. Kannenberg, *Extension Theory*, American Mathematical Society, 2000.

- [221] Gray, J.W., Some global properties of contact structures, *Annals of Mathematics*, 69/2 (1959) 421-450.
- [222] Gregory, R.L., *The Intelligent Eye*, Weidenfeld & Nicolson, London and McGraw-Hill, New York, 1970.
- [223] Grill-Spector, K., Kourtzi, Z., Kanwisher, N., The lateral occipital complex and its role in object recognition, *Vision Research*, 41 (2001) 1409–1422.
- [224] Gröchenig, K., *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston, 2001.
- [225] Gröchenig, K., *New function spaces associated to representations of nilpotent Lie groups and generalized time-frequency analysis*, arXiv :2007.04615, 2020.
- [226] Gröchenig, K., Lyubarskii, Y., Gabor frames with Hermite fonctions, *CRAS (Comptes Rendus de l'Académie des sciences de Paris)*, Série I, 344 (2007) 157-162.
- [227] Gromov, M., Pseudo-holomorphic Curves in Symplectic Manifolds, *Inventiones Mathematicae*, 82 (1985) 307-347.
- [228] Gromov, M., Carnot-Carathéodory spaces seen from within, *Subriemannian Geometry* (A. Bellaïche, J. Risler, eds.), Progress in Mathematics 144, Birkhäuser, Basel, 1996, 79-323.
- [229] Gromov, M., *Metric Structures for Riemannian and Non Riemannian Spaces*, Progress in Mathematics 152, Birkhäuser, Boston, Basel, 1999.
- [230] Grossberg, S., Mingolla, E., Neural dynamics of form perception : boundary completion, illusory figures and neon color spreading, *Psychological Review*, 92 (1985) 173-211.
- [231] Grossberg, S., Mingolla, E., Neural dynamics of perceptual grouping : textures, boundaries and emergent segmentation, *Perception and Psychophysics*, 38/2 (1985) 141-171.
- [232] Grossberg, S., Mingolla, E., The role of illusory contours in visual segmentation, *The Perception of Illusory Contours* (S. Petry, G.E. Meyer, eds.), Springer, Berlin, New York, 1987, 116-125.
- [233] Grünbaum, A., *Philosophical Problems of Space and Time*, Dordrecht-Boston, Reidel, 1973.
- [234] Guckenheimer, J., Caustics and Non-Degenerate Hamiltonians, *Topology*, 13 (1974) 127-133.
- [235] Guichardet, A., Représentations des groupes de Lie nilpotents (d'après Kirillov), *Séminaire Bourbaki* n°249, 1963.
- [236] Guillemin, V., *Differential Forms*, World Scientific Publishing, Singapour, 2019.
- [237] Guillemin, V., Sternberg, S., *Symplectic Techniques in Physics*, Cambridge University Press, Cambridge, 1984.
- [238] Guillemin, V., Sternberg, S., *Semi-classical analysis*, International Press of Boston, 2010.
- [239] Haffner, H., Des calculs dans l'atelier de fabrication d'Élie Cartan, 2017. <http://images.math.cnrs.fr/Des-calculs-dans-l-atelier-de-fabrication-d-Élie-Cartan>
- [240] Haiko, M.A., Mingolla, E., Somers, D.C., Multiple mechanisms of illusory contour perception, *Journal of Vision*, 8/11 (2008) 1–17.
- [241] Hall, B.C. The Segal-Bargmann “coherent state” transform for compact Lie groups, *Journal of Functional Analysis*, 122 (1994) 103-151.
- [242] Hawkins, T., From General Relativity to Group Representations. The Background to Weyl Papers of 1925-26, *Matériaux pour l'histoire des Mathématiques au XX^e siècle*, Actes du colloque à la mémoire de Jean Dieudonné (Nice 1996), Société Mathématique de France, Paris, 1998, 69-100.

- [243] Hawkins, T., *The Mathematics of Frobenius in Context. A Journey Through 18th to 20th Century Mathematics*, Springer, New York, 2013.
- [244] Hehl, F.W., Obukhov, Y., *Élie Cartan's torsion in geometry and in field theory, an essay*. arXiv :0711.1535, 2007.
- [245] Heinzmann, G., Petitot, J., The functional role of structures in Bourbaki, *The Prehistory of Mathematical Structuralism*, (E. Reck, G. Schiemer, eds), Oxford University Press, 2020, 187-214.
- [246] Heisenberg, W., Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen, *Zeitschrift für Physik*, 33 (1925) 879-893.
- [247] Heitger, F., Rosenthaler, L., Von Der Heydt, R., Peterhans, E., Kübler, O., Simulation of neural contour mechanisms : from simple to end-stopped cells, *Vision Research* 32/5 (1992) 963-981.
- [248] Heitger, F., Von Der Heydt, R., A computational model of neural contour processing : figure-ground segregation and illusory contours, *Proceedings of the 4th IEEE International Conference on Computer Vision*, IEEE, 1993, 32-40.
- [249] Helmholtz, H. von, Über die Tatsachen, die der Geometrie zum Grunde liegen, *Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen*, 9, 1868. *Wissenschaftliche Abhandlungen*, Vol. II, Johann Ambrosius Barth, Leipzig, 1883, 618-639. Eng. transl. in *Epistemological Writings*, Reidel, Dordrecht, 1977, 39-58.
- [250] Herb, R.A., Harish-Chandra and his work, *Bulletin of the American Mathematical Society (New Series)*, 25/1 (1991) 1-17.
- [251] Hilbert, D., Cohn-Vossen, S., *Anschauliche Geometrie*, Springer, Berlin, 1932. Engl. transl. *Geometry and the Imagination*, AMS Chelsea Pub, Providence, R.I, 1952.
- [252] Hladky, R.K., Pauls, S.D., Minimal surfaces in the roto-translation group with applications to a neuro-biological image completion model, *Journal of Mathematical Imaging and Vision*, 36 (2010) 1-27, arXiv :math/0509636v1.
- [253] Hoffman, W.C., Some reasons why algebraic topology is important in neuropsychology : perceptual and cognitive systems as fibrations, *International Journal of Man-Machine Studies*, 22 (1985) 613-650.
- [254] Hoffman, W.C., The visual cortex is a contact bundle, *Applied Mathematics and Computation*, 32 (1989) 137-167.
- [255] Hogan, J.A., Lakey, J.D., *Time-Frequency and Time-Scale Methods : Adaptive Decompositions, Uncertainty Principles, and Sampling*, Birkhäuser, Basel, 2005.
- [256] Hörmander, L., *The Analysis of Linear Partial Differential Operators* (4 volumes), Springer, Berlin, New York, 1983-1985.
- [257] Horn, B.K.P., The curves of least energy, *ACM Transactions on Mathematical Software*, 9/4 (1983) 441-460.
- [258] Houzel, C., Histoire de la théorie des faisceaux, *Matériaux pour l'histoire des mathématiques au XX^e siècle. À la mémoire de Jean Dieudonné (1996)*, Séminaires et Congrès 3, Société Mathématique de France, Paris, 1998, 101-119.
- [259] Hubel, D.H., *Eye, Brain and Vision*, Scientific American Library, W.H. Freeman & Co, New York, 1988.

- [260] Husserl, E., (1907), *Ding und Raum, Vorlesungen 1907*, Husserliana XVI, Martinus Nijhoff, La Hague, 1973. Éd. française *Chose et Espace* (trad. J-F. Lavigne), Presses Universitaires de France, Paris, 1989.
- [261] Husserl, E., (1913), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*, Husserliana III-IV, Martinus Nijhoff, La Hague, 1950. Éd. française *Idées Directrices pour une Phénoménologie* (trad. P. Ricoeur), Gallimard, Paris, 1950.
- [262] Husserl, E., *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, Claassen & Goverts, Hamburg, 1954.
- [263] Imbert, M., *Traité du Cerveau*, Odile Jacob, Paris, 2006.
- [264] Itô, K., Differential equations determining a Markov process, *Selected Papers*, Springer, New York, 1987, 42-75.
- [265] Itoh, J., Kiyohara, K., The Cut Loci and the Conjugate Loci on Ellipsoids, *Manuscripta Mathematica*, 114 (2004) 247-264.
- [266] Itzykson, C., Zuber, J.B., *Quantum Field Theory*, McGraw-Hill, Singapour, 1985.
- [267] Itzykson, C., Zuber, J.B., *Théorie statistique des champs*, InterÉditions/Éditions du CNRS, Paris, 1989.
- [268] Jacobi, C.G.J., Über die Pfaffsche Methode, eine gewöhnliche lineäre Differentialgleichung zwischen $2n$ Variablen durch ein System von n Gleichungen zu integrieren, *Journal de Crelle*, II (1827) 347-357. *Gesammelte Werke*, IV, 17-29, Preussische Akademie der Wissenschaften, Reimer, 1886.
- [269] Jacobi, C.G.J., Note von der geodätischen Linie auf einem Ellipsoid und den verschiedenen Anwendungen einer merkwürdigen analytischen Substitution, *Journal de Crelle*, XIX (1839) 309-313. Trad. française : De la ligne géodésique sur un ellipsoïde et des différents usages d'une transformation analytique remarquable, *Journal de Mathématiques Pures et Appliquées*, 6 (1841) 267-272.
- [270] Jacobi, C.G.J., *Vorlesungen über Dynamik* (1842-1843), édité par A. Clebsch, Georg Reimer, Berlin, 1866.
- [271] Jancke, D., Chavane, F., Naaman, S., Grinvald, A., Imaging cortical correlates of illusion in early visual cortex, *Nature*, 428 (2004) 423-426.
- [272] Janssen, A.J.E.M., Strohmer, Th., Hyperbolic secants yield Gabor frames, *Applied and Computational Harmonic Analysis*, 12/2 (2002) 259-267. arXiv :math/0301134.
- [273] Joshi, M. S., *Lectures on Pseudo-differential Operators*, 1999, arXiv :math/9906155v1.
- [274] Kahane, J.-P., Le mouvement brownien, *Matériaux pour l'histoire des mathématiques au XX^e siècle. À la mémoire de Jean Dieudonné (1996)*, Séminaires et Congrès 3, Société Mathématique de France, Paris, 1998, 123-155.
- [275] Kahane, J.-P., Le retour de Fourier, *Académie des sciences*, Août 2005.
- [276] Kajzer, V. V., Conjugate points of left-invariant metrics on Lie groups, *Journal of Soviet Mathematics*, 34 (1990), 32-44.
- [277] Kalisa, C., Torrèsani, B., N -Dimensional Affine Weyl-Heisenberg Wavelets, *Annales de l'Institut Henri Poincaré (A) Physique Théorique*, 59/2 (1993) 201-236.
- [278] Kang, K., Shelley, M., Sompolinsky, H., Mexican hats and pinwheels in visual cortex, *PNAS (Proceedings of the National Academy of Sciences USA)*, 100/5 (2003) 2848-2853.

- [279] Kanizsa, G., *Grammatica del Vedere*, Il Mulino, Bologna, 1980. Trad. française d'A. Chamblolle *La Grammaire du Voir* (Préface de J.-M. Morel), Diderot Editeur, Paris, 1998.
- [280] Kant, E., (1786), *Metaphysische Anfangsgründe der Naturwissenschaft*, Kants gesammelte Schriften, Band IV, Preussische Akademie der Wissenschaften, Berlin, Georg Reimer, 1911. *Premiers Principes métaphysiques de la Science de la Nature*, (trad. J. Gibelin), Vrin, Paris, 1971.
- [281] Kant, E., *Kants gesammelte Schriften*, Preussische Akademie der Wissenschaften, Georg Reimer, Berlin, 1911. *Œuvres philosophiques* (F. Alquié, éd.), Bibliothèque de la Pléiade, Gallimard, Paris, 1980-1986.
- [282] Kaplan, A., *Harmonic analysis on groups of Heisenberg type*, Lecture Notes in Mathematics, vol.992, Springer-Verlag, Berlin, 1983, 416-435.
- [283] Karney, C., Geodesics on an ellipsoid, *Wikipedia*.
- [284] Keller, B., *Introduction aux groupes et algèbres de Lie*, <https://webusers.imj-prg.fr/~bernhard.keller/lie/>
- [285] Keller, J.B., Corrected Bohr-Sommerfeld Quantum Conditions for Nonseparable Systems, *Annals of Physics*, 4 (1958) 180-188.
- [286] Kellman, P.J., Shipley, T.F., A theory of visual interpolation in object perception, *Cognitive Psychology*, 23 (1991) 141-221.
- [287] Kimia, B.B., On the role of medial geometry in human vision, *Neurogeometry and Visual Perception* (J. Petitot, J. Lorenceau, eds.), *Journal of Physiology-Paris*, 97/2-3 (2003) 155-190.
- [288] Kirillov, A. A., *Elements of the theory of representations*, Grundlehren der Mathematischen Wissenschaften, 220, Springer-Verlag, Berlin, New York, 1972.
- [289] Kirillov, A. A., Geometric Quantization, *Dynamical Systems IV* (V.I. Arnold, S.P. Novikov, eds), 2ème édition, Springer, Berlin, Heidelberg, New York, 2001. (Édition russe, 1985).
- [290] Kirillov, A.A., Introduction to the Theory of Representations and Noncommutative Harmonic Analysis, *Representation Theory and Noncommutative Harmonic Analysis I*, (Kirillov A.A., ed.), *Encyclopaedia of Mathematical Sciences*, vol 22. Springer, Berlin, Heidelberg, 1994, 1-156.
- [291] Kirillov, A. A., *Lectures on the Orbit Method*, Graduate Studies in Mathematics 64, American Mathematical Society, Providence, Rhode Island, 2004.
- [292] Kisil, V., Integral representations and coherent states, *Bulletin of the Belgian Mathematical Society*, 2 (1995) 529-540.
- [293] Kisil, V., *Wavelets in Applied and Pure Mathematics*, <http://maths.leeds.ac.uk/~kisilv/courses/wavelets.html>.
- [294] Klein, F., L'œuvre géométrique de Sophus Lie, *Nouvelles annales de mathématiques*, 3e série, 15 (1896) 1-20 (accessible sur *Numdam*). Traduit par M.L. Langel d'une conférence au *Evans-ton Colloquium*, Chicago, 1893, parue en 1894 chez Mac Millan, Londres, New York.
- [295] Klüver, H., Mescal vision and eidetic vision, *American Journal of Psychology*, 37 (1926) 502-515.
- [296] Klüver, H., *Mescal and Mechanisms of Hallucinations*, University of Chicago Press, Chicago, 1966.

- [297] Koenderink, J.J., Simultaneous order in nervous nets from a functional standpoint, *Biological Cybernetics*, 50 (1984) 35-41.
- [298] Koenderink, J.J., Operational significance of receptive field assemblies, *Biological Cybernetics*, 58 (1988) 163-171.
- [299] Koenderink, J.J., Van Doorn, A.J., Representation of local geometry in the visual system, *Biological Cybernetics*, 55 (1987) 367-375.
- [300] Kohn, J.J., Nirenberg, L., On the algebra of pseudodifferential operators, *Communications in Pure and Applied Mathematics*, 18 (1965) 269-305.
- [301] Kohn, R.V., A Few of Louis Nirenberg's Many Contributions to the Theory of Partial Differential Equations, *The Abel Prize 2013-2017*, (H. Holden, R. Piene, eds), Springer, Cham, 2019.
- [302] Kok, P., de Lange, F.P., Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex, *Current Biology*, 24 (2014) 1531-1535.
- [303] Kolmogorov, A., Une série de Fourier-Lebesgue divergente partout, *CRAS (Comptes-Rendus de l'Académie des Sciences de Paris)*, 183 (1926) 1327-1328.
- [304] Kolmogorov, A.N., Yushkevich, A.P. (eds), *Mathematics of the 19th Century*, vol. 3, *Constructive Function Theory (According to Chebyshev). Ordinary Differential Equations. Calculus of Variations. Theory of Finite Differences* (Izdatel'svo "Nauka", Moskva, 1987), English translation by R. Cooke, Birkhäuser, Basel, 1998.
- [305] Kostant, B., *Quantization and Unitary Representations*, Lecture Notes in Mathematics, Vo.1 170, Springer, Berlin.
- [306] Kowalski, O., Szenthe, J., On the existence of homogeneous geodesics in homogeneous Riemannian manifolds, *Geometriae Dedicata*, 81 (2000), 209-214. (Erratum : *Geom. Dedicata* 84 (2001), 331-332.)
- [307] Kramers, H.A., Wellenmechanik und halbzahlige Quantisierung, *Zeitschrift für Physik*, 39 (1926) 828-840.
- [308] Kumpera, A., Ruiz, C., Sur l'équivalence locale des systèmes de Pfaff en drapeau, *Monge-Ampère Equations and Related Topics* (F. Gherardelli, ed.), Istituto Nazionale di Alta Matematica, Rome, 1982, 201-248.
- [309] Lablée, O., *Une vue panoramique sur l'analyse semi-classique*, 2008. Hal-00398069, Arxiv 0906.4535. 2008.
- [310] Lagrange, J.-L., Une nouvelle méthode de Calcul Intégral pour les différentielles affectées d'un radical carré sous lequel la variable ne passe pas le quatrième degré, *Mémoires de l'Académie royale des Sciences de Turin*, II, 1784-1785.
- [311] Lagrange, J.-L., *Mécanique analytique* (Mécanique analitique), chez la Veuve Desaint, Libraire, Paris, 1788. Réédité dans *Œuvres de Lagrange* (J.-A. Serret, G. Darboux, eds), 1867-1892, vol. 11 et 12. Fac simulé de l'édition originale chez Jacques Gabay, Paris, 1989.
- [312] Lagrange, J.-L. (colloque) *La mécanique analytique de Lagrange et son héritage*, Collège de France, Paris, septembre 1988, *Atti della Accademia delle Scienze di Torino*, 124 (1990).
- [313] Lamme, V.A.F., Super, H., Spekreijse, H., Feedforward, horizontal and feedback processing in the visual cortex, *Current Opinion in Neurobiology*, 8 (1998) 529-535.
- [314] Lang, S., *SL₂(ℝ)*, Addison-Wesley, Reading, 1975.

- [315] Langlands, R.P., *Harish-Chandra*, Biographical Memoirs of Fellows of the Royal Society, 31 (1985) 197-225.
- [316] Laudенbach, F., Orbites périodiques et courbes pseudo-holomorphes. Application à la conjecture de Weinstein en dimension 3, *Séminaire Bourbaki* n°786, *Astérisque*, 227 (1995) 309-333.
- [317] LC, *Le Labyrinth du Continu*, (J.-M. Salanskis, H. Sinaceur, édés), Actes du Colloque de Cerisy, Springer-Verlag France, 1992.
- [318] Lê, F., *Theory and applications of invariants in Alfred Clebsch's papers*, <http://math.univ-lyon1.fr/~fle/>
- [319] Lee, E. B., Markus, L., *Foundations of Optimal Control Theory*, John Wiley and Sons, New York, 1967.
- [320] Lee, T.S., Mumford, D., Romero, R., Lamme, V.A.F., The role of primary visual cortex in higher level vision, *Vision Research*, 38 (1998) 2429-2454.
- [321] Lee, T.S., Computations in the early visual cortex, *Neurogeometry and Visual Perception* (J. Petitot, J. Lorenceau, eds.), *Journal of Physiology - Paris*, 97, 2-3 (2003) 121-139.
- [322] Leeuwen, S.J. van, *The Segal-Bargmann transform and its generalizations*, Master Thesis, University of Utrecht, 2009.
- [323] Legendre, A.-M., *Traité des fonctions elliptiques et des intégrales Eulériennes*, trois volumes, Buzard-Courcier, 1825-1828. Accessible sur *Gallica* et Hachette/BNF, 2014.
- [324] Leibniz, G.W., Mémoire de M. Leibniz touchant son sentiment sur le calcul différentiel (1701), *Mathematische Schriften*, V, Olms, Berlin, 1962.
- [325] Leray, J., *Solutions asymptotiques et groupe symplectique*, Séminaire Jean Leray, 3 (1973-1974), 1-25.
- [326] Lescroart, M., Biederman, I., Cortical Representation of Medial Axis Structure, *Cerebral Cortex*, 23/3 (2013) 629-637.
- [327] Levien, R., *The elastica : a mathematical history*, Technical Report UCB/EECS-2008-103, Berkeley, 2008.
- [328] Lévy, P., *Processus stochastiques et mouvement brownien*, Gauthier-Villars, Paris, 1948.
- [329] Lewis, A.D., The Maximum Principle of Pontryagin in control and in optimal control, 2006. <https://mast.queensu.ca/~andrew/teaching/pdf/maximum-principle.pdf>
- [330] Leyton, M., *Symmetry, Causality, Mind*, MIT Press, Cambridge MA, 1992.
- [331] Li, P., Yau, S.T., On the parabolic kernel of the Schrödinger operator, *Acta Mathematica*, 156/3-4 (1986) 153-201.
- [332] Libermann, P., Élie Cartan (1869-1951). Travaux mathématiques, *Centre Universitaire de Luxembourg*, 8 (1996) 115-158.
- [333] Lie, S., Sur les complexes et en particulier les complexes de droites et de sphères et de leur application à la théorie des équations aux dérivées partielles, *Mathematische Annalen*, V (1872) 145-256.
- [334] Lie, S., Theorie des Pfaff'schen Problems, *Archiv for Mathematik og Naturvidenskab*, II (1877) 338-379, Kristiania.
- [335] Lie, S., Theorie der Transformationsgruppen I, *Mathematische Annalen*, 16 (1880) 441-528. Engl. Transl. M. Ackermann, R. Hermann, *Sophus Lie's 1880 Transformation Group paper*, Math Sci Press, Brookline, MA, 1975.

- [336] Liontou, V., Marcolli, M., *Gabor Frames from Contact Geometry in Models of the Primary Visual Cortex*, arXiv :2111.02307, 2022.
- [337] Longo, G., *Le cauchemar de Prométhée. Les sciences et leurs limites*, Presses Universitaires de France, Paris, 2023.
- [338] Longo, G., Petitot, J., Teissier, B., *Géométrie et Cognition* (manifesto), ENS, 1999. <http://jeanpetitot.com/ArticlesPDF/Geo-Co-ENS.pdf>
- [339] LTC, *Logos et Théorie des Catastrophes. Colloque de Cerisy en l'honneur de R. Thom* (J. Petitot, éd.), Éditions Patiño, Genève, 1989.
- [340] Lund, J., Fitzpatrick, D., Humphrey, A.L., The striate visual cortex of the tree shrew, *Cerebral Cortex* (E. G. Jones, A. Peters, eds.), Plenum, New York, 1985, 157-205.
- [341] Lutz, R., Structures de contact sur les fibrés principaux en cercles de dimension 3, *Annales de l'Institut Fourier*, 27 (1977) 1-15.
- [342] Mackey, G. W., A theorem of Stone and von Neumann, *Duke Mathematical Journal*, 16 (1949), 313-326.
- [343] Mackey, G. W., *The Theory of Unitary Group Representations*, The University of Chicago Press, 1976.
- [344] Maffei, M., Fiorentini, A., The unresponsive regions of visual cortical receptive fields, *Vision Research*, 16 (1976) 1131-1139.
- [345] Maier-Hein, K., The challenge of mapping the human connectome based on diffusion tractography, *Nature Communications*, 8 :1349 (2017) 1-13.
- [346] Maldonado, P.E., Gödecke, I., Gray, C.M., Bonhöffer, T., Orientation Selectivity in Pinwheel Centers in Cat Striate Cortex, *Science*, 276 (1997) 1551-1555.
- [347] Mallat, S., Multifrequency Channel Decompositions of Images and Wavelet Models, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37, 12 (1989) 2091-2110.
- [348] Mallat, S., *A Wavelet Tour of Signal Processing*, Academic Press, New York, 1998.
- [349] Manin, Y.I., Marcolli, M., *Homotopy theoretic and categorical models of neural information networks*, arXiv :2006.15136v1, 2020.
- [350] Marcolli, M., Tsao, D., *Geometry of Neurosciences*, California Institute of Technology, 2017. <http://www.its.caltech.edu/~matilde/GeomNeuroClass.html>
- [351] Margulis, G.A., Mostow, G.D., Some remarks on the definition of tangent cones in a Carnot-Carathéodory space, *Journal d'Analyse Mathématique*, 80 (2000), 299-317.
- [352] Marinosci, R.A., Homogeneous geodesics in a three-dimensional Lie group, *Commentationes Mathematicae Universitatis Carolinae*, 43/2 (2002) 261-270.
- [353] Marle, C.M., *Quantification géométrique : théorie et exemples*, Séminaire Paul Krée, tome 2 (1975-1976), exposé n°2.
- [354] Marr, D., *Vision*, W.H. Freeman, San Francisco, 1982.
- [355] Martinet, J., Formes de contact sur les variétés de dimension 3, Proceedings of the *Liverpool Singularities Symposium II, Lectures Notes in Mathematics 209*, Springer, Berlin, 1971, 142-163.
- [356] Maslov, V.P., *TEORIJA VOS MOUTCHENII ACYMPTOTICHESTIE METODI (Perturbation Theory and Asymptotic Methods)*, Éditions de l'Université de Moscou, 1964. Trad. fr. J. Lascoux, R. Sénéor, *Théorie des perturbations et méthodes asymptotiques*, Dunod, Paris, 1972.

- [357] Massot, P., *Le théorème de Bennequin*, <http://images.math.cnrs.fr/Le-theoreme-de-Bennequin-II.html>.
- [358] Massot, P., *Topological methods in 3-dimensional contact geometry*, arXiv :1303-1063, 2013.
- [359] Mazzetti, C., Sarti, A., Citti, G., *Functional architecture of M1 cells encoding movement direction*, arXiv :2208.03352, 2022.
- [360] Mazzetti, C., Sarti, A., Citti, G., *A model of reaching via subriemannian geodesics in Engel-type group*, arXiv :2301.05765, 2023.
- [361] Mendola, J.D., Dale, A.M., Fischl, B., Liu, A.K., Tootell, R.B.H., The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging, *The Journal of Neuroscience*, 19/19 (1999) 8560–8572.
- [362] Merker, J., *Le problème de l'espace. Sophus Lie, Friedrich Engel et le problème de Riemann-Helmholtz*, Hermann, Paris, 2010.
- [363] Merker, J. & de Marçay, F., *Séries de Fourier*, <https://www.imo.universite-paris-saclay.fr/~joel.merker/>
- [364] Merker, J. & de Marçay, F., *Espaces de Hilbert*, <https://www.imo.universite-paris-saclay.fr/~joel.merker/>
- [365] Merker, J. & de Marçay, F., *Faisceaux*, <https://www.imo.universite-paris-saclay.fr/~joel.merker/>
- [366] Métivier, G., Équations aux dérivées partielles sur les groupes nilpotents, *Séminaire Bourbaki* n° 583, novembre 1981, *Astérisque*, 92-93 (1982) 75-99.
- [367] Metzinger, T., (ed.), *Neural Correlates of Consciousness : Empirical and Conceptual Questions*, MIT Press, Cambridge MA, 2000.
- [368] Milnor, J., Curvatures of Left-invariant Metrics on Lie Groups, *Advances in Mathematics*, 21/3 (1976) 293-329.
- [369] Misner, C.W., Thorne, K.S., Wheeler, J.A., *Gravitation*, San Francisco, Freeman, 1973.
- [370] Mitchell, J., On Carnot-Carathéodory metrics, *Journal of Differential Geometry*, 21 (1985) 35–45.
- [371] Mitchinson, G., Crick, F., Long axons within the striate cortex : their distribution, orientation, and patterns of connections, *PNAS (Proceedings of the National Academy of Sciences USA)*, 79 (1982) 3661-3665.
- [372] Moiseev, I., *Geometry of strokes*, SISSA, Trieste, 2006.
- [373] Moiseev, I., Sachkov, Y.L., Maxwell strata in sub-Riemannian problem on the group of motions of a plane, *ESAIM : Control, Optimisation and Calculus of Variations*, 16 (2010) 380-399. arXiv :0807.4731v1, 2008.
- [374] Monge, G., Supplément. Où l'on fait voir que les Équations aux différences ordinaires, pour lesquelles les conditions d'intégrabilité ne sont pas satisfaites, sont susceptibles d'une véritable intégration, et que c'est de cette intégration que dépend celle des équations aux différences partielles élevées, *Académie Royale des Sciences*, 1784, 502-576.
- [375] Montgomery, R., *A Tour of Subriemannian Geometries, Their Geodesics and Applications*, Mathematical Surveys and Monographs 91, American Mathematical Society, 2002.
- [376] Montgomery, R., Zhitomirskii, M., Geometric approach to Goursat flags, *Annales de l'Institut Henri Poincaré, Analyse non linéaire*, 184 (2001) 459-493.

- [377] Montgomery, R., Zhitomirskii, M., *Points and curves in the Monster tower*, Memoirs of the American Mathematical Society, 956, 2010.
- [378] Morin, B., Champs de vecteurs sur les sphères, *Séminaire Bourbaki*, n° 233 (1962) 163-189.
- [379] Mormul, P., Goursat flags : classification of codimension-one singularities, *Journal of Dynamical and Control Systems*, 6/3 (2000) 311-330.
- [380] Mäkkiläinen, R., Bednar, J.A., Choe, Y., Sirosh, J., *Self-organizing neural network model of the primary visual cortex*, Proceedings of the Fifth International Conference on Neural Information Processing (ICONIP'98), Ohmsha, 1998.
- [381] Mumford, D., Elastica and computer vision, *Algebraic Geometry and Applications* (C. Bajaj, ed.), Springer Verlag, Heidelberg, 1992, 491-506.
- [382] Nabonnand, P., Contribution à l'histoire de la théorie des géodésiques au XIX^e siècle, *Revue d'histoire des mathématiques*, 1 (1995) 159-200.
- [383] Nahm, F.K.D., Pribram, K.H., Heinrich Klüver, *Biographical Memoirs*, vol. 73, 1998, 289-305, National Academies Press, Washington D.C. nap.nationalacademies.org/read/9650/chapter/16.
- [384] Natani, L., Über totale und partielle Differentialgleichungen, *Journal de Crelle*, LVIII (1861) 301-328.
- [385] Neel, R., Stroock, D., Analysis of the Cut Locus via the Heat Kernel, *Surveys in Differential Geometry*, 9 (2004) 337-349.
- [386] Nelson, E., Dynamical theories of Brownian motion, Princeton University Press, 1967.
- [387] Neuman, E., On Gauss lemniscate functions and lemniscatic mean, *Mathematica Pannonica*, 18/1 (2007) 77-94.
- [388] von Neumann, J., *Mathematische Grundlagen der Quantenmechanik*, Springer-Verlag, Berlin, 1932. *Mathematical Foundations of Quantum Mechanics*, (Engl Trans. R.T. Beyer), Princeton University Press, 1955.
- [389] Nieder, A., Seeing more than meets the eye : processing of illusory contours in animals, *Journal of Comparative Physiology A*, 188 (2002) 249-260.
- [390] Nitzberg, M., Mumford, D., Shiota, T., *Filtering, segmentation and depth*, Lecture Notes in Computer Science, 662, Springer Verlag, Berlin, New York, 1993.
- [391] Noë, A., Thompson, E., Are there Neural Correlates of Consciousness ?, *Journal of Consciousness Studies*, 11/1 (2004) 3-28.
- [392] Nurowski, P., Tafel, J., Symmetries of Cauchy-Riemann Spaces, *Letters in Mathematical Physics*, 15 (1988) 31-38.
- [393] O'Keefe, J., Immanuel Kant : Pioneer neuroscientist, Public Lecture, *Royal Institution of Great Britain*, London, June 2, 2014.
- [394] Oldfather, W.A., Ellis, C.A., Brown, D.M., Leonhard Euler's Elastic Curves, *Isis*, 20/1 (1933) 72-160.
- [395] Ollivier, Y., *Cohomologie de Čech des faisceaux*, www.yann-ollivier.org, 2001.
- [396] Oster, G., Phosphenes, *Scientific American*, 222, 2 (1970) 83-87.
- [397] Pansu, P., Métriques de Carnot-Carathéodory et quasi-isométries des espaces symétriques de rang un, *Annals of Mathematics*, 129/1 (1989) 1-60.
- [398] Panza, M., *Newton et les origines de l'analyse : 1664-1666*, Blanchard, Paris, 2005.

- [399] Parent, P., Zucker, S.W., Trace inference, curvature consistency, and curve detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence II*, 8 (1989) 823-839.
- [400] Pasillas-Lépine, W., Respondek, W., On the geometry of Goursat structures; *ESAIM : Control, Optimisation and Calculus of Variations*, 6 (2001) 119-181.
- [401] Pauls, S., Minimal surfaces in the Heisenberg group, *Geometriae Dedicata* 104 (2004) 201–231, Springer.
- [402] Pavliotis, G., A., *Stochastic Processes and Applications : Diffusion Processes, the Fokker-Planck and Langevin Equations*, Springer, Berlin, Heidelberg, 2014.
- [403] Pearcey, T., The structure of an electromagnetic field in the neighbourhood of a cusp of a caustic, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 37/268 (1946) 311–317.
- [404] Peiffer, J., Jacob Bernouilli, maître et rival de son frère Johann, www.jehps.net, juin 2008.
- [405] Pénin, J.-C., *Les fonctions elliptiques : une histoire*, fonctionelliptique.pagesperso-orange.fr, 2007.
- [406] Perrin, D., *Résolution par radicaux*, Université d'Orsay, www.imo.universite-paris-saclay.fr/~perrin/TER/radicaux.pdf, 2006.
- [407] Perrin, N., *Introduction to Lie algebras*, <https://lmv.math.cnrs.fr/wp-content/uploads/2019/09/lie-alg-2.pdf>
- [408] Peterhans, E., Von Der Heydt, R., Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps, *Journal of Neuroscience*, 9/5 (1989) 1749-1763.
- [409] Peterhans, E., Von Der Heydt, R., Subjective contours : bridging the gap between psychophysics and psychology, *Trends in Neuroscience*, 14/3 (1991) 112-119.
- [410] Petitot, J., Infinitesimale, *Enciclopedia Einaudi*, VII, 443-521, Einaudi, Turin, 1979.
- [411] Petitot, J., Locale/Globale, *Enciclopedia Einaudi*, VIII, 429-490, Einaudi, Turin, 1979.
- [412] Petitot, J., *Eléments de théorie des singularités*, 1982, http://jeanpetitot.com/ArticlesPDF/Petitot_Sing.pdf.
- [413] Petitot, J., *Introduction aux phénomènes critiques*, 1982, http://jeanpetitot.com/ArticlesPDF/Petitot_CritPh.pdf.
- [414] Petitot, J., Refaire le Timée. Introduction à la philosophie mathématique d'Albert Lautman. *Revue d'Histoire des Sciences*, XL/1 (1987) 79-115. Version complète dans hal-01130394.
- [415] Petitot, J., Structuralisme et Phénoménologie : la théorie des catastrophes et la part maudite de la Raison, *Logos et Théorie des Catastrophes*, Colloque de Cerisy (1982, J. Petitot, ed.), Éditions Patiño, Genève, 1988. 345-376.
- [416] Petitot, J., Rappels sur l'Analyse non standard, *La Mathématique non standard* (J.-M. Salanskis, H. Barreau, eds.), Éditions du CNRS, Paris, 1989, 187-209.
- [417] Petitot, J., Modèles morphodynamiques pour la grammaire cognitive et la sémiotique modale, *RSSI Canadian Semiotic Association*, 9/1-2-3 (1989) 17-51.
- [418] Petitot, J., *Physique du Sens*, Éditions du CNRS, Paris, 1992.
- [419] Petitot, J., Actuality of Transcendental Aesthetics for Modern Physics, *1830-1930 : A Century of Geometry*, (L. Boi, D. Flament, J.-M. Salanskis, eds), Springer, Berlin, New York, 1992, 273-304.

- [420] Petitot, J., Phénoménologie naturalisée et Morphodynamique : la fonction cognitive du synthétique a priori, *Philosophie et Sciences cognitives* (J.-M. Salanskis, éd.), *Intellectica*, 17 (1993) 79-126.
- [421] Petitot, J., *The unity of mathematics as a method of discovery : Wiles' example*, XIXth International Congress of History of Science, Zaragoza, 1993 and 7th French Philosophy of Mathematics Workshop, University of Paris-Diderot, 2015. http://jeanpetitot.com/ArticlesPDF/Petitot_TSW_Wiles.pdf.
- [422] Petitot, J., La sémiophysique : de la physique qualitative aux sciences cognitives, *Passion des Formes, à René Thom* (M. Porte, éd.), École Normale Supérieure, Éditions Fontenay-Saint Cloud, 1994, 499-545.
- [423] Petitot, J., Phénoménologie computationnelle et objectivité morphologique, *La connaissance philosophique. Essais sur l'œuvre de Gilles-Gaston Granger* (J. Proust, È. Schwartz, eds.), Presses Universitaires de France, Paris, 1994, 213-248.
- [424] Petitot, J., Phenomenology of Perception, Qualitative Physics and Sheaf Mereology, *Philosophy and the Cognitive Sciences*, Proceedings of the 16th International Wittgenstein Symposium (R. Casati, B. Smith, G. White, eds), Hölder-Pichler-Tempsky, Vienna, 387-408.
- [425] Petitot, J., La réorientation naturaliste de la phénoménologie, *Archives de Philosophie, Sciences cognitives et Phénoménologie*, 58/4 (1995) 631-658.
- [426] Petitot, J., Morphodynamics and Attractor Syntax. Dynamical and morphological models for constituency in visual perception and cognitive grammar, *Mind as Motion* (T. van Gelder, R. Port, eds.), MIT Press, Cambridge MA, 1995, 227-281.
- [427] Petitot, J., Objectivité faible et Philosophie transcendante, *Physique et Réalité. Débat avec B. d'Espagnat*, (M. Bitbol, S. Laugier, eds.), Diderot Editeur, Paris, 1997, 201-236. http://jeanpetitot.com/ArticlesPDF/Petitot_DEspagnat.pdf
- [428] Petitot, J. (avec une collaboration de Y. Tondut), Vers une Neuro-géométrie. Fibrations corticales, structures de contact et contours subjectifs modaux, *Mathématiques, Informatique et Sciences Humaines*, 145 (1999) 5-101, École des Hautes Études en Sciences Sociales, Paris.
- [429] Petitot J., Morphological Eidetics for Phenomenology of Perception, *Naturalizing Phenomenology : Issues in Contemporary Phenomenology and Cognitive Science* (J. Petitot, F. J. Varela, J.-M. Roy, B. Pachoud, eds.), Stanford University Press, Stanford, 1999, 330-371. Trad. française *Naturaliser la phénoménologie. Essais sur la phénoménologie contemporaine et les sciences cognitives*, CNRS Éditions, Paris, 2002, 427-484.
- [430] Petitot J., Les infinitésimales comme éléments nilpotents : actualité du débat Nieuwentijt / Leibniz, *L'actualité de Leibniz : les deux labyrinthes* (D. Berlioz, F. Nef, eds), *Studia Leibnitiana Supplementa*, 34 (1999) 567-575, Franz Steiner, Stuttgart.
- [431] Petitot, J., The Neurogeometry of Pinwheels as a Sub-Riemannian Contact Structure, *Neurogeometry and Visual Perception* (J. Petitot, J. Lorenceau, eds.), *Journal of Physiology-Paris*, 97/2-3 (2003) 265-309.
- [432] Petitot, J., Neurogéométrie et phénoménologie de la perception, *Philosophie de la Perception* (J. Bouveresse, J.-J. Rosat, eds.), Collège de France & Odile Jacob, Paris, 2003, 53-76.
- [433] Petitot, J., Géométrie et Vision dans *Ding und Raum* de Husserl, *Des lois de la pensée aux constructivismes* (M.-J. Durand-Richard, éd.), *Intellectica*, 39, (2004) 139-167.
- [434] Petitot, J., Neurogéométrie des architectures fonctionnelles de la vision, *Journée annuelle de la SMF, Mathématiques et Vision*, 2006, 69-128.

- [435] Petitot, J., *Neurogéométrie de la Vision. Modèles mathématiques et physiques des architectures fonctionnelles*, Les Éditions de l'École Polytechnique, Palaiseau, 2008.
- [436] Petitot, J., Noncommutative Geometry and Transcendental Physics, *Constituting Objectivity. Transcendental Perspectives on Modern Physics* (M. Bitbol, P. Kerszberg, J. Petitot eds), Springer, Berlin, New York, 2009, 415-455.
- [437] Petitot, J., A Transcendental View on the Continuum : Woodin's Conditional Platonism, *Le continu mathématique. Nouvelles conceptions, nouveaux enjeux* (M. De Glas ed.), *Intellectica*, 51/1, 93-133. arXiv :1502.05898, 2015.
- [438] Petitot, J., 'Le hiatus entre le logique et le morphologique'. Prédication et perception, *Semiosis and Catastrophes. René Thom's Semiotic Heritage* (W. Wildgen, P.A. Brandt eds), Peter Lang, Bern, 141-166.
- [439] Petitot, J., La simplicité de la notion géométrique de jet, *Leçons "Complexité-Simplicité" du Collège de France*, 2012. OpenEditions Book, 2014.
- [440] Petitot, J., Landmarks for Neurogeometry, *Neuromathematics of Vision*, (G. Citti, A. Sarti, eds.), Springer, Berlin, New York, 2014, 1-85.
- [441] Petitot, J., *Noncommutative geometry and physics* (a compilation of Alain Connes' noncommutative geometry in physics), arXiv :1505.00132 [math-ph]
- [442] Petitot, J., *Elements of Neurogeometry. Functional Architectures of Vision*, Lecture Notes in Morphogenesis, Springer, Berlin, New York, 2017.
- [443] Petitot, J., *Axiomatics as a strategy for complex proofs : the case of Riemann Hypothesis*, http://jeanpetitot.com/ArticlesPDF/Petitot_AIPS_2017.pdf, 2017.
- [444] Petitot, J., Calcul différentiel neuronal et architectures fonctionnelles, *Les Neurosciences au sein des Sciences de la Cognition entre Neuroenthousiasme et Neuroscepticisme*, (C. Monier, A. Sarti, eds), *Intellectica*, 69/1-2 (2018) 303-346.
- [445] Petry, S., Gannon, R. Time, motion, and objectedness in illusory contours, *Petry, S., Meyer, G.E. (eds.)* [446], 193-200.
- [446] Petry, S., Meyer, G.E. (eds.), *The Perception of Illusory Contours*, Springer, Berlin, New York, 1987.
- [447] Pfaff, J.F., Methodus generalis, aequationes differentiarum partialium, nec non aequationes differentiales vulgares, utrasque primi ordinis inter quocumque variables, complete integrandi, *Abhandlungen der Königlichten Akademie der Wissenschaften zu Berlin*, 1814-1815, 76-136. Trad. allemande par Gerhard Kowalewski, *Ostwald's Klassiker der exakten Wissenschaften*, n° 129, Leipzig, 1902.
- [448] Pinna, B., Brelstaff, G., Spillmann, L., Surface color from boundaries : a new "intercolor" illusion, *Vision Research*, 41 (2001) 2669-2676.
- [449] Poincaré, H., *Les méthodes nouvelles de la Mécanique céleste*, III, Gauthier-Villars, Paris, 1899.
- [450] Poincaré, H., *La Science et l'Hypothèse*, Flammarion, Paris, 1902. Réédition avec une préface de J. Vuillemin, 1968.
- [451] Poincaré, H., Sur les lignes géodésiques des surfaces convexes, *Transactions of the American Mathematical Society*, 17 (1905) 237-274.
- [452] Ponge, R., *Heisenberg calculus and spectral theory of hypoelliptic operators on Heisenberg manifolds*, arXiv :math/0509300, 2005.

- [453] Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., Mishchenko, E. F., *The Mathematical Theory of Optimal Processes*, Classics of Soviet Mathematics, John Wiley & Sons, 1962, Gordon and Breach, 1986.
- [454] Porteous, I.R., The Normal Singularities of Surfaces in \mathbb{R}^3 , *Proceedings of the A.M.S. Summer Institute on Singularities*, Arcata, Californie, 1981.
- [455] Prandi, D., Gauthier, J.-P., *A Semidiscrete Version of the Petitot Model as a Plausible Model for Anthropomorphic Image Reconstruction and Pattern Recognition*, arxiv :1704.03069v1, 2017.
- [456] Prochiantz, A., *Machine-esprit*, Odile Jacob, Paris, 2001.
- [457] Quigg, C., *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, Benjamin-Cummings, Reading, 1983.
- [458] Ramis, J.-P., Les derniers travaux de Jean Martinet, *Annales de l'Institut Fourier*, 42/1-2 (1992) 15-47.
- [459] Ramis, J.-P., Warusfel, A., et al., *Mathématiques - Tout-en-un pour la Licence*, Dunod, Paris.
- [460] Ramsden, B.M., Hung, C.P., Roe, A., Real and illusory contour processing in area V1 of the primate : a cortical balancing act, *Cerebral Cortex*, 11 (2001) 648–665.
- [461] Rangan, A.V., Cai, D., McLaughlin, D.W., Modeling the spatiotemporal cortical activity associated with the line-motion illusion in primary visual cortex, *PNAS (Proceedings of the National Academy of Sciences USA)*, 102/52 (2005) 18793-18800.
- [462] Rauschecker, A.M., Bestman, S., Walsh, V., Thilo, K.V., Phosphene threshold as function of contrast of external visual stimuli, *Experimental Brain Research*, 157/1 (2004) 124-127.
- [463] Ray, T., The Chemical Architecture of the Human Mind : Probing Receptor Space with Psychedelics, <https://bandura.sbs.arizona.edu>, *Toward a Science of Consciousness VI*, Tucson, 2004.
- [464] Riemann, B., (1854), Über die Hypothesen, welche der Geometrie zu Grunde liegen, *Bernhard Riemanns Gesammelte Mathematische Werke* (R. Dedekind, H. Weber, eds.), Druck und Verlag von B. G. Teubner, Leipzig, 1892.
- [465] Risken, H., *The Fokker–Planck Equation. Methods of Solutions and Applications*, Springer, Berlin, Heidelberg, 1996.
- [466] Robinson, A., *Non standard Analysis*, North-Holland, Amsterdam.
- [467] Rockland, K.S., Lund, J.S., Widespread periodic intrinsic connections in the tree shrew visual cortex, *Science*, 215 (1982) 532-534.
- [468] Rockland, K.S., Lund, J.S., Intrinsic laminar lattice connections in the primate visual cortex, *Journal of Comparative Neurology*, 216/3 (1993) 303-318.
- [469] Roe, A.W., Modular Complexity of Area V2 in the Macaque Monkey, in *The Primate Visual System*, (J.H. Kaas, C.E. Collins, eds), Chap. 5, CRC Press LLC, Boca Raton, FL, 2003, 109-138.
- [470] Roy, J.-M., *Rhin et Danube. Essais sur le schisme analytico-phénoménologique*, Vrin, Paris, 2010.
- [471] Roy, J.-M.; Petitot J., Pachoud, B., Varela, F., Beyond the gap, *Naturalizing Phenomenology : Issues in Contemporary Phenomenology and Cognitive Science*, Stanford University Press, Stanford, 1999, 1-80. Trad. française Comblent le déficit, *Naturaliser la phénoménologie. Essais sur la phénoménologie contemporaine et les sciences cognitives*, CNRS Éditions, Paris, 2002, 1-100.

- [472] Rosenberg, J., *A Selective History of the Stone–von Neumann Theorem*, Contemporary Mathematics, 365, American Mathematical Society, 2004.
- [473] Rudin, W., *Functional Analysis*, McGraw-Hill, New York, 1973/1991.
- [474] Ryckman, T., *The Reign of Relativity : Philosophy in Physics 1915–1925*, Oxford University Press, New York, 2005.
- [475] Ryckman, T., Hermann Weyl and ‘First Philosophy’ : Constituting Gauge Invariance, *Constituting Objectivity, Transcendental Perspectives on Modern Physics* (M. Bitbol, P. Kerszberg, J. Petitot, eds), Springer, Berlin, New York, 2009, 279-298.
- [476] Sachkov, Y., Maxwell strata in Euler’s elastic problem, *arXiv :0705.0614v1* (2007) puis *Journal of Dynamical and Control Systems*, 14/2 (2008) 169-234.
- [477] Sachkov, Y., Conjugate points in the Euler elastic problem, *Journal of Dynamical and Control Systems*, 14/3 (2008) 409-439.
- [478] Sachkov, Y., Cut locus and optimal synthesis in the sub-Riemannian problem on the group of motions of a plane, *ESAIM : Control, Optimisation and Calculus of Variations*, 2009, www.emath.fr/cocv
- [479] Sachkov, Yu. L., Cut time and optimal synthesis in sub-Riemannian problem on the group of motions of a plane, *Quaderni di Matematica*, Università di Milano-Bicocca, 3/2009, arXiv :0903.0727v1.
- [480] Sachkov, Y., *Exponential mapping in Euler’s elastic problem*, arXiv :1303.1746v1, 2013.
- [481] de Saint Gervais, H.P., *Analysis Situs*, <http://analysis-situs.math.cnrs.fr>.
- [482] Salanskis, J.-M., *Le constructivisme non standard*, Presses Universitaires du Septentrion, Lille, 1999.
- [483] Sanguinetti, G., *Invariant models of vision between phenomenology, image statistics and neurosciences*, Thesis, Montevideo University, 2011.
- [484] Sanguinetti, G., Citti, G., Sarti, A., Image completion using a diffusion driven mean curvature flow in a sub-Riemannian space, *VISIGRAPP 2008 (International Joint Conference on Computer Vision and Computer Graphics Theory and Applications)*, 46-53.
- [485] Sanguinetti, G., Citti, G., Sarti, A., A model of natural image edge co-occurrence in the roto-translation group, *Journal of Vision*, 10/14 (2010) 1-16.
- [486] Sarti, A., Citti, G., Petitot, J., On the Symplectic Structure of the Primary Visual Cortex, *Biological Cybernetics*, 98/1 (2008) 33-48.
- [487] Sauer, T., Field equations in teleparallel spacetime : Einstein’s *Fernparallelismus* approach towards unified field theory, arXiv :physics/0405142, 2004.
- [488] Scheibe, E., Hermann Weyl and the nature of space-time, *Exakte Wissenschaften und ihre philosophische Grundlegung* (W. Deppert, Kurt Hübner, eds), Peter Lang, Frankfurt am Main, 1988, 61-82.
- [489] Schlosshauer, M., *Decoherence and the Quantum-to-Classical Transition*, Springer-Verlag, Heidelberg, Berlin, 2009 (4th ed.).
- [490] Scholz, E., Hermann Weyl’s Contribution to Geometry, 1917–1923, *The Intersection of History and Mathematics* (S. Chikara, S. Mitsuo, J.W. Dauben, eds.), Birkhäuser, Berlin, 1994, 203-230.

- [491] Scholz, E., Hermann Weyl's "purely infinitesimal geometry", *Proceedings of the International Congress of Mathematics, Zürich, 1994*, (S.D. Chatterji, ed.), Birkhäuser, Zürich, 1995, 1512-1603.
- [492] Schummers, J., Marino, J., Sur, M., Synaptic integration by V1 neurons depends on location within the orientation map, *Neuron*, 36 (2002) 969-978.
- [493] Schwartz, L., *Théorie des distributions*, Hermann, Paris, 2 volumes, 1950-1951.
- [494] Seip, K., Wallstén, R., Density theorems for sampling and interpolation in the Bargmann-Fock space, I, II, *Journal für die reine und angewandte Mathematik*, 429 (1992) 91-106, 107-114.
- [495] Serra, J., *Image Analysis and Mathematical Morphology*, Academic Press, New York, 1982.
- [496] Sharpe, R., *Differential Geometry*, Springer, New-York, 1997.
- [497] Sheth, B.R., Sharma, J., Rao, C., Sur, M., Orientation maps of subjective contours in visual cortex, *Nature*, 274 (1996) 2110-2115.
- [498] Shipley, T.F., Kellman, P.J., Perception of partly occluded objects and illusory figures : evidence for an identity hypothesis, *Journal of Experimental Psychology*, 18/1 (1992) 106-120.
- [499] Shipley, T.F., Kellman, P.J., Strength of visual interpolation depends on the ratio of physically specified to total edge length, *Perception and Psychophysics*, 52 (1992) 97-106.
- [500] Shmuel, A., Grinvald, A., Coexistence of linear zones and pinwheels within orientation maps in cat visual cortex, *PNAS (Proceedings of the National Academy of Sciences USA)*, 97/10 (2000) 5568-5573.
- [501] Shubin, M., *Pseudodifferentsialnye operatory i spektralnaya teoria*, Nauka, Moscou, 1978. Eng. transl. S.I. Anderson, *Pseudodifferential Operators and Spectral Theory*, Springer, Berlin, 1987 & 2001.
- [502] Sigman, M., Cecchi, G.A., Gilbert, C.D., Magnasco, M.O., On a common circle : Natural scenes and Gestalt rules, *PNAS (Proceedings of the National Academy of Sciences USA)*, 98/4 (2001) 1935-1940.
- [503] Sikora, A., Sharp pointwise estimates on heat kernels, *The Quarterly Journal of Mathematics*, 47/3 (1996) 371-382.
- [504] Singer, D.A., *Lectures on Elastic Curves and Rods*, Case Western Reserve University, <https://case.edu/artsci/math/singer/publish/elaslecs.pdf>.
- [505] Smadja, I., *La lemniscate de Fagnano et la multiplication complexe*, halshs-00456361, 2010.
- [506] Snippe, H.P., Koenderink, J.J., Discrimination thresholds for channel-coded systems, *Biological Cybernetics*, 66 (1992) 543-551.
- [507] Soriano, M., Spillmann, L., Bach, M., The Abutting Grating Illusion, *Vision Research*, 36/1 (1996) 109-116.
- [508] Sotomayor, J., Gutierrez, C., Structurally Stable Configurations of Lines of Principal Curvature, *Bifurcations, théorie ergodique et applications, Astérisque*, 98-99 (1982) 195-215.
- [509] Souriau, J.M., *Géométrie symplectique et physique mathématique*, Colloques Internationaux du CNRS, 237, Paris, 1975.
- [510] Spillmann, L., Dresch, B., Phenomena of illusory form : can we bridge the gap between levels of explanation?, *Perception*, 24 (1995) 1333-1364.

- [511] Spivak, M., *A Comprehensive Treatise on Differential Geometry*, Publish or Perish, Boston MA, 1974.
- [512] Sporns, O., Tononi, G., Kötter, R., The Human Connectome : A Structural Description of the Human Brain, *PLOS*, 1/4 (2005) e42, 245-251.
- [513] Stanton, N., Review of Beals-Greiner [34], *Bulletin of the American Mathematical Society*, 21/1 (1989) 177-179.
- [514] Stone, M.H., Linear transformations in Hilbert Space, III. Operational methods and group theory, *PNAS (Proceedings of the National Academy of Sciences USA)*, 16 (1930) 172-175.
- [515] Stone, M.H., *Linear Transformations in Hilbert Space and their Applications to Analysis*, American Mathematical Society, New York, 1932.
- [516] Strichartz, R.S., Sub-Riemannian Geometry, *Journal of Differential Geometry*, 24 (1986) 221-263.
- [517] Sussmann, H. J., Willems, J. C., 300 years of optimal control : From the brachystochrone to the maximum principle, *IEEE Control Systems Magazine*, 1997.
- [518] Swanson, L.R., The predictive processing paradigm has roots in Kant, *Frontiers in Systems Neuroscience*, 10 (2016) 1-13.
- [519] Swindale, N.V., How many maps are there in visual cortex ?, *Cerebral Cortex*, 7 (2000) 633-643.
- [520] Szczeciniarz, J.-J., The Mysterious Strength of the Galois Theory, *Proceedings of the Evariste Galois Bicentenary Meeting*, (F.M. Bertato, J.C. Cifuentes, J.-J. Szczeciniarz, eds), Hermann, Paris, 2014, 53-102.
- [521] Szczeciniarz, J.-J., Réflexions sur Riemann, philosophe mathématicien, *Les mathématiques et l'expérience, Ce qu'en ont dit les philosophes et les mathématiciens*, Herrmann, Paris, 2015, 195-221.
- [522] Tallon-Baudry, C., Bertrand, O., Oscillatory gamma activity in humans and its role in object representation, *Trends in Cognitive Science*, 3/4 (1999) 151-162.
- [523] Tanaka, N., On differential systems, graded Lie algebras and pseudo-groups, *Journal of Mathematics of Kyoto University*, 10/1 (1970) 1-82.
- [524] Taylor, M.E., *Noncommutative Harmonic Analysis*, American Mathematical Society, 1986.
- [525] Teissier, B., Géométrie et Cognition : l'exemple du continu, *Ouvrir la logique au monde. Philosophie et Mathématique de l'interaction*, (J.-B. Joinet, S. Tronçon, eds), Actes du Colloque de Cerisy de septembre 2006 *Logique et interaction ; vers une géométrie de la cognition*, Hermann, Paris, 2009.
- [526] Teissier, B., Trotman, D., Commentaire sur l'article "Ensembles et morphismes stratifiés" [529] de René Thom, *Œuvres mathématiques*, I1, 440-445, Société Mathématique de France, Paris, 2019.
- [527] Thom, R., Les Singularités des applications différentiables, *Annales de l'Institut Fourier*, 6 (1956) 43-87. *Œuvres mathématiques*, 1, 337-382, Société Mathématique de France, Paris, 2017.
- [528] Thom, R., Sur la théorie des enveloppes, *Journal de Mathématiques Pures et Appliquées*, 41 (1962) 177-192. *Œuvres mathématiques*, I1, 130-144, Société Mathématique de France, Paris, 2019.

- [529] Thom, R., Ensembles et morphismes stratifiés, *Bulletin of the American Mathematical Society*, 75/2 (1968) 240-284. *Œuvres mathématiques*, II, 395-439, Société Mathématique de France, Paris, 2019.
- [530] Thom, R., Sur le cut-locus d'une variété plongée, *Journal of Differential Geometry*, 6/4 (1972) 577-586. *Œuvres Mathématiques*, III, 125-133, Société Mathématique de France, Paris, 2021.
- [531] Thom, R., Sur les équations différentielles multiformes et leurs intégrales singulières, *Boletim da Sociedade Brasileira de Matemática*, 3/1 (1972) 1-11. *Œuvres Mathématiques*, III, 171-181, Société Mathématique de France, Paris, 2021.
- [532] Thom, R., Sur la propagation des fronts d'ondes et l'optique géométrique, *Proceedings of the C. Carathéodory International Symposium, Athens, 1973*, (A. Panayotopoulos, ed.), Greek Mathematical Society, 1974, 580-585.
- [533] Thom, R., *Modèles mathématiques de la morphogenèse*, Bourgeois, Paris, 1980. Ouvrage commenté par Alain Chenciner dans *Œuvres Mathématiques*, III, 116-124, Société Mathématique de France, Paris, 2021.
- [534] Thom, R., *Apologie du Logos*, Hachette, Paris, 1990.
- [535] Thurston, W.P., Winkelnkemper, H.E., On the existence of contact forms, *Proceedings of the American Mathematical Society*, 52 (1975) 345-347.
- [536] Toet, A., Blom, J., Koenderink, J.J., The construction of a simultaneous functional order in nervous systems, *Biological Cybernetics*, 57 (1987) 115-125.
- [537] Torossian, C., La conjecture de Kashiwara-Vergne (d'après Alekseev, Meinrenken), *Séminaire Bourbaki*, n° 980, 2007. *Astérisque* 317 (2008) 441-466.
- [538] Torrèsani, B., Time-frequency representation : wavelet packets and optimal decomposition, *Annales de l'Institut Henri Poincaré*, 56 (1992) 215-234.
- [539] Torrèsani, B., Some Remarks on Wavelet Representations and Geometric Aspects, *Wavelets : Theory, Algorithms and Applications*, 1993, Taormina, Italy. hal-01739594.
- [540] *Troisième Rencontre de Géométrie du Schnepfenried* (10-15 mai 1982), 2 volumes, *Astérisque* 107-108 et 109-110, 1983.
- [541] Trélat, E., *Contrôle optimal : théorie & applications*, Vuibert, Paris, 2005.
- [542] Trélat, E., Analyse spectrale des Laplaciens sous-Riemanniens, mesure de Weyl, Journée Sous-Riemannienne, 13/10/2016, Institut Fourier, <https://hal.archives-ouvertes.fr/medihal-01383669/>
- [543] Truesdell, C., The Rational Mechanics of Flexible or Elastic Bodies, Introduction aux Vol. X et XI des *Opera Omnia* de L. Euler, Birkhäuser, Basel, 1960.
- [544] Truesdell, C., The influence of elasticity on analysis : the classical heritage, *Bulletin of the American Mathematical Society*, 9/8 (1983) 293-310.
- [545] Ts'o, D.Y., Gilbert, C.D., Wiesel T.N., Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis, *Journal of Neuroscience*, 6/4 (1986) 1160-1170.
- [546] Turiel, A., Nadal, J.-P., Parga, N., Orientational minimal redundancy wavelets : from edge detection to perception, *Vision Research*, 43, 9 (2003) 1061-1079.
- [547] Twareque Ali, S., Antoine, J.-P., Gazeau, J.-P., Mueller, U.A., Coherent states and their generalizations : a mathematical overview, *Reviews in Mathematical Physics*, 07/07 (1995) 1013-1104.

- [548] Tyler, C.W., Some new entoptic phenomena, *Vision Research*, 18 (1978) 1633-1639.
- [549] Ullman, S., Filling in the gaps : the shape of subjective contours and a model for their generation, *Biological Cybernetics*, 25 (1976) 1-6.
- [550] Van Gulick, R., Consciousness, *Stanford Encyclopedia of Philosophy*, 2004. <http://plato.stanford.edu/entries/consciousness/>
- [551] Varadarajan, V. S., George Mackey and His Work on Representation Theory and Foundations of Physics, *Group representations, ergodic theory, and mathematical physics : a tribute to George W. Mackey, Contemporary Mathematics*, 449, American Mathematical Society, Providence, RI, 2008, 417–446.
- [552] Velmans, M., *Understanding Consciousness*, Routledge, New York, 2000.
- [553] Vershik, A.M., Gershkovich, V.Y., Non Holonomic Dynamical Systems, Geometry of Distributions and Variational Problems, *Encyclopædia of Mathematical Sciences*, 16 (1994) 1-81, Springer, Berlin.
- [554] Viviani, P., Schneider, R., A developmental study of the relation between geometry and kinematics in drawing movements, *Journal of Experimental Psychology : Human Perception and Performance*, 17 (1991) 198-218.
- [555] Vogel, T., *Existence of Engel structures*, Thesis, University of Munich, 2004. arXiv :math/0411217
- [556] Von Der Heydt, R., Peterhans, E., Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity, *Journal of Neuroscience*, 9/5 (1989) 1731-1748.
- [557] Vuillemin, J., *Physique et Métaphysique kantienne*, Presses Universitaires de France, Paris, 1955.
- [558] Waters, T., *The Conjugate Locus on the Triaxial Ellipsoid*, Wolfram Demonstrations Project, 2014. <https://demonstrations.wolfram.com/TheConjugateLocusOnTheTriaxialEllipsoid/>
- [559] Weber, E. von, Zur Invariantentheorie der Systeme Pfaff'scher Gleichungen, *Leipzig Berichte Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften Mathematisch-Physikalische*, Klasse L (1898) 207-229.
- [560] Weinstein, A., 1977, *Lectures on Symplectic Manifolds*, CBMS Regional Conference Series in Mathematics, 29, American Mathematical Society, Providence, 1977.
- [561] Weyl, H., *Die Idee der Riemannschen Fläche*, B.G. Teubner, Leipzig, Berlin, 1913.
- [562] Weyl, H., *Das Kontinuum : Kritische Untersuchungen über die Grundlagen der Analysis*, Verlag von Veit & Comp., Leipzig, 1918.
- [563] Weyl, H., *Space-Time-Matter*, Dover, New-York, 1922.
- [564] Weyl, H., Theorie der Darstellung kontinuierlicher hal-beinfacher Gruppen durch lineare Transformationen, I, II, III, *Mathematische Zeitschrift*, 23 (1925) 271-309, 24 (1926) 328-376, 24 (1926) 377-395.
- [565] Weyl, H., *Gruppentheorie und Quantenmechanik*, Hirzel, Leipzig, 1928 et 1931. *The Theory of Groups & Quantum Mechanics* (tr. by H.P. Robertson), Dover Publications, New York, 1950.
- [566] Weyl, H., Review of Cartan [97], *Bulletin of the American Mathematical Society*, 44 (1938) 598-601.
- [567] Whittaker, E.T., *A Course of Modern Analysis*, Cambridge University Press, Cambridge, 1902 (4^{eme} édition, 1927 avec Watson, G.N.).

- [568] Williams, L.R., Jacobs, D.W., Stochastic completion fields : a neural model of illusory contour shape and salience, *Proceedings of IEEE International Conference on Computer Vision*, 5 (1995) 408-415. *Neural Computation*, 9/4 (1997) 837-858.
- [569] Woodhouse, N.M.J., *Geometric Quantization*, Clarendon Press, Oxford, 1980/1992.
- [570] Wu, A., *Bayesian latent structure discovery for large-scale neural recordings*, Thesis, Princeton University, 2019.
- [571] Yen, S.C., Finkel, L.H., Extraction of perceptually salient contours by striate cortical networks, *Vision Research*, 38, 5 (1998) 719-741.
- [572] Zahaf M.B., Manchon, D., Confluence of singularities of differential equations : A Lie algebras contraction approach, hal-00292676, 2008.
- [573] Zhitomirskii, M., Singularities of vector distributions, *Geometry, Analysis and Dynamics on sub-Riemannian Manifolds*, vol. II, (D. Barilari, U. Boscain, M. Sigalotti, eds), European Mathematical Society, 2016.
- [574] Zuber, J.-B., *Introduction à la théorie des groupes et de leurs représentations*, <https://cel.archives-ouvertes.fr/cel-00092968>.
- [575] Zygmund, A., *Trigonometric Series* (second edition), Cambridge University Press, Cambridge UK, 1959 (first edition, Warsaw, 1935).