



HAL
open science

Balades aléatoires

Jonas Kahn

► **To cite this version:**

| Jonas Kahn. Balades aléatoires. Mathématiques [math]. Université de Lille 1, 2016. tel-01500300

HAL Id: tel-01500300

<https://hal.science/tel-01500300>

Submitted on 3 Apr 2017

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.



Document de synthèse
présenté par
Jonas KAHN
en vue de l'obtention de
l'Habilitation à Diriger des Recherches
Spécialité : Mathématiques

Balades aléatoires

soutenue le 16 décembre 2016

Composition du jury :

Cristina BUTUCEA	Professeur (Université Paris-Est Marne-la-Vallée), <i>Rapporteur</i>
Sergueï DACHIAN	Professeur (Université Lille 1)
David DEREUDRE	Professeur (Université de Lille 1)
Wilfrid KENDALL	Professeur (University of Warwick), <i>Rapporteur</i>
Gabriele STEIDL	Professeur (Technische Universität Kaiserslautern), <i>Rapporteur</i>

Remerciements

Je remercie d'abord le CNRS pour les conditions de travail qu'il m'accorde.

Mes collègues de Lille m'auront chaudement accueilli après ma thèse, et le groupe de travail de géométrie stochastique m'a, en particulier, beaucoup apporté. Et merci à tous pour toutes les tablettes de chocolat, et merci pour ta chaleur, ma collègue de bureau !

À Lille toujours, les biologistes de l'IRI auront dialogué avec moi pendant plusieurs années. Je me sens encore coupable de ne pas avoir fourni les méthodes que j'avais promises.

À Toulouse, où je suis désormais, je dois énormément à Pierre Weiss, dont la capacité à me recentrer ne s'est jamais démentie.

Un grand merci enfin à mes rapporteurs, qui ont accepté malgré mon retard de lire ce maladroit manuscrit.

Avant-propos

Je ne parle dans ce manuscrit que de mes travaux postérieurs à la thèse, sans mentionner les derniers articles de statistique quantique qui ont été publiés après mes soutenances.

Le mémoire d'habilitation est souvent l'occasion d'essayer de tisser une histoire, de relier les éléments de travail des années précédentes, en leur donnant une cohérence, en les inscrivant dans un seul tracé, fût-il parsemé de surprises et de détours, en un mot dans un plan. Il n'y en a pas.

Mon travail après ma thèse est le fruit du hasard des rencontres. Cela peut être un simple problème évoqué dans un séminaire, que je pense résoudre. Ou cela peut être une collaboration suivie avec quelqu'un dont les questions m'intéressent.

Il y a aussi des regrets au cours de ces années, des projets inachevés, par paresse ou par perfectionnisme. Depuis une étude des effets de l'internat sur le succès aux concours, à des idées autour de la carte brownienne. Le seul évoqué dans ces notes est le projet d'amélioration de données FLIM-FRET, où je n'ai jamais implémenté le programme que j'avais promis. . . J'aurai appris de ces années à rechercher davantage de collaborations pour partager les enjeux avec d'autres.

La première partie du manuscrit évoque deux points de mon travail de «mercenaire», où je résous un problème entendu une fois et ne revient pas sur le thème : l'un est sur le calcul quantique avec ancilla, l'autre sur des comparaisons de chaînes de Markov,

La seconde partie rassemble des questions liées à l'imagerie en science, et plus généralement à l'étude de données spatialement organisées. Le caractère en est donc essentiellement statistique. Sont évoquées des données de microscopie par fluorescence, et des questions trouvant leur origine dans l'IRM. Les premières m'ont mené à corriger les taux connus pour l'estimation des mélanges finis. L'IRM a motivé des questions d'échantillonnage compressé sous contraintes, de projection sur des espaces de mesure, et d'algorithme de clustering ultra-rapides, qui évitent de percoler.

La troisième partie parle de géométrie stochastique, entendue au sens le plus large. J'y range donc l'étude de tout objet créé de manière aléatoire dans ses aspects géométriques, et notamment métriques. Le premier travail traite de l'existence d'une mesure de Gibbs pour des modèles de mosaïques en T. Le second montre que le processus de droites de Poisson impropre a suffisamment de symétries et de comportement hyperbolique pour être un SIRS_N au sens d'Aldous, en toute dimension.

Je commence ce mémoire par un bref résumé, en une phrase, de chaque résultat important, avec l'article correspondant.

Table des matières

Remerciements	4
Avant-propos	6
Résultats	9
Présentation	13
1 Mercanariat	13
1.1 Ancilla	13
1.1.1 Notions de calcul quantique	13
1.1.2 Calcul quantique avec ancilla	15
1.1.3 Intrication et fidélité des portes quantiques inexactes	15
1.2 Inégalités de comparaison entre chaînes de Markov	17
1.2.1 Mesures de vitesse de convergence	17
1.2.2 Propriétés élémentaires des inégalités de comparaison	19
1.2.3 Comparaisons de vitesse générales	20
1.2.4 Ajouter des pas ne ralentit pas une chaîne de Markov	20
1.2.5 Chaînes de naissance et de mort	21
2 Images en science	23
2.1 Mesures d'interactions protéine-protéine par FLIM-FRET	24
2.2 Vitesses d'estimation optimales dans les modèles de mélange	26
2.2.1 Mélanges	26
2.2.2 Vitesse minimax	27
2.2.3 Vitesse point par point et interprétation	29
2.3 IRM fonctionnelle	30
2.3.1 Perspectives	34
2.4 Acquisition comprimée sous contraintes physiques	34
2.4.1 Motivation initiale : acquisition en IRM	34
2.4.2 Voyageur de commerce comme échantillonneur de densité variable	36

2.4.3	Projection sur les mesures	37
2.4.4	Perspectives	42
3	Géométrie stochastique et espaces métriques aléatoires	44
3.1	Mosaïques en T	45
3.2	Processus de droites de Poisson impropre	48
3.2.1	SIRSN et définitions	48
3.2.2	Propriétés	50
3.2.3	Liens	54
3.2.4	Perspectives	54

Résultats

Résultat 1

La fidélité d'une porte quantique non-idéale dans le modèle de calcul quantique avec ancilla est bornée supérieurement par une fonction strictement décroissante de l'intrication des qubits concernés avec le reste du registre.

Result 1

The fidelity of an inaccurate quantum gate in ancilla-driven quantum computation is upper bounded by a decreasing function of the entanglement between the qubits that are acted on and the remainder of the register.

MORIMAE et J. KAHN (2010)

Résultat 2

Si K et L sont deux noyaux de Markov réversibles stochastiquement monotones sur un espace partiellement ordonné, avec la même distribution d'équilibre, et satisfont une *inégalité de comparaison* – une nouvelle relation d'ordre partiel – alors la chaîne de Markov associée à K mélange plus vite que celle associée à L à *tout instant* en variation totale, L^2 et séparation, entre autres, pour de bonnes conditions initiales.

Result 2

If K and L are two reversible stochastically monotone Markov kernels on a partially ordered space, and they satisfy a *comparison inequality* – a new partial order – then the Markov chain with kernel K mixes faster than the Markov chain with kernel L *at all times* in total variation and L^2 distances, and separation, among others.

J. FILL et J. KAHN (2013, Corollaire 3.3)

Résultat 3

Parmi les chaînes de vie et de mort sur $\{0, \dots, n\}$ qui convergent vers la distribution uniforme depuis l'état initial 0, celle uniforme – une chance sur deux d'aller à gauche ou à droite, ou de rester sur place aux extrémités – majorise à *tout instant* toutes les autres. Elle mélange donc plus vite dans de nombreux sens.

La seule exception classique est au sens du temps de mélange de Lovász-Winkler quand n est impair, auquel cas la chaîne la plus rapide est aussi identifiée.

Result 3

The uniform birth-and-death chain – one chance in two to go right and left, or stay at endpoints – majorizes *at all times* any other birth-and-death chain on $\{0, \dots, n\}$. Hence it mixes faster in many senses.

The only classical exception are Lovász-Winkler mixing times when n is even, in which case the fastest chain is also given.

J. FILL et J. KAHN (2013, Théorèmes 4.3 and 6.5)

Résultat 4

Ajouter un pas à une chaîne de Markov ne ralentit pas le mélange, dans un certain nombre de cas particuliers.

Result 4

Extra updates do not delay mixing for several Markov chains.

J. FILL et J. KAHN (2013, Section 8)

Résultat 5

Le clustering récursif par plus proches voisins est extrêmement rapide et permet une réduction de dimension qui conserve le signal intéressant en IRM fonctionnelle.

Result 5

Recursive nearest neighbour clustering is extremely fast and reduces dimension while preserving the relevant signal in functional MRI.

HOYOS-IDROBO et al. (2016)

Résultat 6

Le test du maximum de vraisemblance est relativement efficace pour déterminer le nombre d'espèces en un pixel d'une image FLIM-FRET. La distance de transport permet de déterminer la similitude entre différents pixels.

Result 6

Maximum likelihood ratio test is comparatively efficient for finding the number of species in a pixel of a FLIM-FRET image. The transportation distance allows to measure similarity between pixels.

HEINRICH, Jonas KAHN et al. (2011) and HEINRICH, PISFIL et al. (2014)

Résultat 7

La vitesse minimax d'estimation d'une loi de mélange à m composantes au plus, localement autour d'un mélange à m_0 composantes, est en $n^{-1/(4(m-m_0)+2)}$ sous des conditions de régularité et d'identifiabilité. Donc la vitesse globale est en $n^{-1/(4m-2)}$.

Result 7

The optimal local minimax rate of estimation of a finite mixture with at most m components around a mixture with m_0 components is $n^{-1/(4(m-m_0)+2)}$, under sufficient regularity and identifiability conditions. Hence the global minimax rate of estimation of a finite mixture with at most m components is $n^{-1/(4m-2)}$.

HEINRICH et Jonas KAHN (2015, Théorèmes 3.2 et 3.3)

Résultat 8

Il existe des estimateurs qui convergent non uniformément à vitesse $n^{-1/2}$ vers toutes les lois de mélange fini.

Result 8

There are estimators that converge non-uniformly at rate $n^{-1/2}$ to all finite mixing distributions.

HEINRICH et Jonas KAHN (2015, Théorème 3.5)

Résultat 9

Un voyageur de commerce reliant des points tirés proportionnellement à $\pi^{1-1/d}$ approche une densité π par des courbes continues. Il est aussi possible d'approcher une probabilité π par des tirages sous contraintes par un algorithme général de projection sur des mesures. Une application parmi d'autres est l'acquisition compressée sous contraintes, en visant une densité adaptée à la paire de bases acquisition/compression.

Result 9

A travelling salesman connecting points sampled according to $\pi^{1-1/d}$ converges to a density π with continuous curves. A more general projection algorithm may be used to approach a probability π by sampling under very general constraints. As a typical application, compressed sensing under acquisition constraints is considered, pairing acquisition/compression bases to determine the target π .

CHAUFFERT, CIUCIU, Jonas KAHN et P. WEISS (2014), CHAUFFERT, CIUCIU, Jonas KAHN et P. WEISS (2016) et BOYER et al. (2016)

Résultat 10

On ne peut construire que $o(a^k k^k)$ mosaïques en T différentes sur k droites données, pour tout a . Aussi les modifications Gibbsiennes de la mosaïque en T complètement aléatoires existent si l'énergie est bornée inférieurement par le nombre de segments.

Result 10

There are at most $o(a^k k^k)$ different T -tessellations on k given lines, for any a . Hence Gibbsian modifications of the CRTT (Completely Random T -Tessellation) exist if the energy is bounded from below by the number of lines.

Jonas KAHN (2014)

Résultat 11

Les géodésiques du processus de droites de Poisson impropre génère un SIRS_N en toute dimension.

Result 11

The geodesics of the improper Poisson line process are a SIRS_N (scale-invariant random spatial network) in any dimension.

Jonas KAHN (2015)

Présentation

Chapitre 1

Mercanariat

1.1 Ancilla

1.1.1 Notions de calcul quantique

L'informatique quantique (NIELSEN et CHUANG, 2010) consiste à effectuer des calculs en utilisant directement des systèmes et phénomènes quantiques. Dans certains cas, elle permet des calculs considérablement plus rapides que tout ordinateur classique.

L'une des idées initiales était la simulation efficace de systèmes quantiques (FEYNMAN, 1982), vraisemblablement exponentiellement lente sur des ordinateurs classiques. DEUTSCH et JOZSA (1992) et SIMON (1997) ont mis au point des algorithmes permettant de résoudre en temps polynomial un certain problème avec oracle, nécessitant un temps exponentiel dans le cas classique. Mais l'intérêt pour les ordinateurs quantiques a véritablement explosé avec l'algorithme de SHOR (1994), permettant la factorisation d'entiers et le log discret en temps polynomial, et l'algorithme de GROVER (1996), permettant une recherche annuelle inverse en temps racine carrée.

Il existe un certain nombre de modèles de calcul quantique, c'est-à-dire d'ensembles de ressources (l'équivalent des bits dans le monde classique) et d'opérations permises (portes logiques en classique), à partir desquels le calcul doit être effectué. Les réalisations physiques de ces modèles ont donc «seulement» à implémenter chacune de ces ressources et opérations. Tous ces modèles sont équivalents et universels, dans le sens où un algorithme écrit dans un modèle peut être traduit à coût polynomial près dans un autre modèle, par exemple la machine de Turing quantique (DEUTSCH, 1985).

Le premier modèle est celui du circuit de portes quantiques, où les qubits sont stockés dans un registre, et où des portes quantiques peuvent être appliquées directement à chaque qubit ou paire de qubits. Expérimentalement, il peut être difficile de maintenir l'intrication entre les qubits du registre au cours de ces opérations.

Un autre modèle plus récent est celui du calcul quantique à sens unique («one-way quantum computation» (RAUSSENDORF et BRIEGEL, 2001)), où les qubits

sont initialement intriqués tous ensemble dans un certain état («cluster state»), et où le calcul s'effectue simplement en mesurant les qubits individuellement. L'intrication permet de propager des effets aux autres qubits. Ainsi, la préparation de la ressource est clairement séparée du calcul lui-même.

Le modèle de calcul quantique avec ancilla (ANDERS et al., 2010) peut être vu comme intermédiaire entre les deux. Les qubits restent stockés dans un registre, mais il n'y a pas de porte appliquée directement à des paires de qubits du registre. Au lieu de cela, chaque qubit peut être intriqué via une porte bien choisie avec une ancilla, et les opérations effectuées sur cette ancilla propagent leur effet au qubit du registre.

Un peu plus formellement, à un objet quantique est associé un espace de Hilbert complexe \mathbb{H} , et un état quantique pur – les seuls que nous utiliserons – est un élément de norme 1 dans cet espace de Hilbert. Dans le cas d'un qubit, l'espace de Hilbert est \mathbb{C}^2 . On dénote par $|0\rangle$ et $|1\rangle$ deux vecteurs formant une base orthonormale quelconque mais désormais fixée de \mathbb{C}^2 . On l'appelle la *base de calcul*. Une autre base utile est définie par :

$$\begin{aligned} |+\rangle &= \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle), \\ |-\rangle &= \frac{1}{\sqrt{2}} (|0\rangle - |1\rangle) \end{aligned}$$

Un état quantique peut soit suivre soit une évolution unitaire, c'est-à-dire $|\psi\rangle \mapsto U|\psi\rangle$ pour U un opérateur unitaire sur \mathbb{H} , soit être projeté suite à une mesure sur une base orthonormale de \mathbb{H} , c'est-à-dire que si $\{|\phi_i\rangle\}_{1 \leq i \leq d(\mathbb{H})}$ est une telle base, l'état quantique $|\psi\rangle$ devient $|\phi_i\rangle$ avec probabilité $\|\langle\psi|\phi_i\rangle\|^2$.

Nous utiliserons les deux évolutions unitaires suivantes :

- la porte de Hadamard \hat{H} , agissant sur un qubit via $\hat{H}|0\rangle = |+\rangle$ et $\hat{H}|1\rangle = |-\rangle$.
- la porte Z contrôlée (CZ), agissant sur une paire de qubits de la manière suivante : $\widehat{CZ} = |00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 10| - |11\rangle\langle 11|$.

D'une manière générale, nous pourrions préciser sur quel qubit une opération est effectuée via des indices : par exemple \hat{H}_A signifie la porte de Hadamard appliquée à l'ancilla A .

L'intrication entre deux ou plusieurs objets quantiques signifie qu'en général un système de plusieurs objets quantiques même sans interactions doit être considéré comme un seul grand objet. Il n'est pas suffisant de connaître l'état de chacun des petits objets pour connaître l'état du système. Cela provient du fait que l'espace de Hilbert associé au grand objet est le produit tensoriel de ceux des petits objets. Et beaucoup des vecteurs de norme 1 ne peuvent pas être écrits comme des produits directs d'états des petits objets. Un exemple typique d'état intriqué pour deux qubits est

$$\frac{1}{\sqrt{2}} (|0\rangle|0\rangle + |1\rangle|1\rangle). \quad (1.1)$$

Pour évaluer l'intrication, nous faisons usage de la matrice densité réduite, définie par $\rho_A = \text{Tr}_B (|\phi\rangle_{A\otimes B} \langle\phi|_{A\otimes B})$, où $|\phi\rangle_{A\otimes B}$ est l'état quantique des deux sous-systèmes potentiellement intriqués A et B , et où Tr_B est la trace partielle sur B . C'est une matrice positive de trace 1. S'il n'y a pas d'intrication, c'est-à-dire si $|\phi\rangle_{A\otimes B} = |\phi\rangle_A \otimes |\phi\rangle_B$, alors $\rho_A = |\phi\rangle_A \langle\phi|_A$. En cas d'intrication maximale, comme pour l'état (1.1), on obtient $\rho_A = \frac{1}{\dim A} 1_A$.

Une mesure d'intrication pour un qubit avec un autre système et comprise entre 0 et 1, est donc la suivante :

$$S = 2 (1 - \text{Tr}(\rho_A^2)). \quad (1.2)$$

1.1.2 Calcul quantique avec ancilla

Le modèle de calcul quantique avec ancilla est donné par :

- un registre de N qubits.
- une ancilla d'un qubit
- la possibilité d'appliquer à n'importe quel qubit du registre une porte de Hadamard.
- la possibilité d'intriquer n'importe quel qubit R du registre avec l'ancilla A via l'opération $\hat{E} = \hat{H}_A \hat{H}_R \widehat{CZ}_{AR}$.
- la possibilité de mesurer l'ancilla dans n'importe quelle direction ou, de manière équivalente, de lui faire subir n'importe quelle rotation et de la mesurer dans la base de calcul.

En effet, l'implémentation de toutes les transformations unitaires sur un qubit, plus une opération d'intrication sur n'importe quelle paire de qubit, permet d'obtenir un ordinateur quantique universel (ANDERS et al., 2010). L'ordinateur quantique avec ancilla permet de réaliser ces deux opérations, comme illustré à la Figure 1.1.

1.1.3 Intrication et fidélité des portes quantiques inexactes

Bien entendu, en pratique, les rotations, mesures et portes E ne sont jamais implémentées de manière complètement exacte. Comment les erreurs sur les portes se propagent-elles aux états ?

Avec Tomoyuki Morimae, nous nous sommes intéressés au cas où les mesures (effectuées en (c) et (g) sur la Figure 1.1) sont imprécises : l'axe de projection dévie de ε , c'est-à-dire que la réponse 1 est obtenue pour $\cos(\varepsilon/2) |1\rangle + \sin(\varepsilon/2) |0\rangle$ si on mesure dans la base de calcul.

Notons F la fidélité moyenne de la porte, définie par $F = \mathbb{E} [|\langle\phi|\psi\rangle|^2]$, où $|\phi\rangle$ est l'état attendu dans le registre et $|\psi\rangle$ l'état (aléatoire) en fait obtenu. Une porte idéale a une fidélité de 1, et la fidélité est positive.

Suivant l'état initial du registre, cette fidélité moyenne peut varier, pour un ε donné. Dans des cas chanceux, elle peut même rester de 1. Mais il n'y a pas de

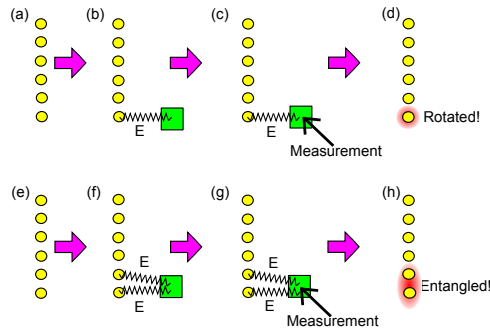


FIGURE 1.1 – Les cercles jaunes représentent les qubits du registre. En haut : rotation d’un qubit. (b) L’ancilla (carré vert) préparée dans l’état $|+\rangle$ est couplée au qubit à tourner via l’interaction E . (c) Après l’interaction, l’ancilla est mesurée dans la direction dans laquelle on veut tourner le qubit. (d) La rétro-action fait tourner le qubit du registre de la manière désirée, quitte à devoir changer la base de calcul en fonction du résultat de la mesure.

En bas : intrication de deux qubits. (f) L’ancilla préparée dans l’état $|+\rangle$ est couplée à chacun des deux qubits individuellement via l’interaction E , la même que celle utilisée pour la rotation. (g) L’ancilla est mesurée dans la base de calcul. (h) La rétroaction cause l’intrication des deux qubits du registre, l’équivalent d’un \widehat{CZ} et de rotations individuelles des qubits.

cas chanceux si les qubits du registre sur lesquels on agit sont intriqués au reste du registre :

Théorème 1.1.1. (MORIMAE et J. KAHN, 2010) Avec les notations ci-dessus, une mesure d’imprécision ε , et S l’intrication (1.2) entre l’un des qubits sur lequel on agit et le reste du registre, la fidélité de la rotation comme celle de l’intrication sont bornées ainsi :

$$F \leq 1 - S \sin^2 \frac{\varepsilon}{2}.$$

Comme le calcul quantique exige cette intrication entre tous les qubits du registre pour être efficace, cela prouve que les erreurs dans les opérateurs de mesure ont de toute façon des conséquences. Toutefois, celles-ci peuvent être corrigées en pratique par des codes correcteurs d’erreur quantiques.

Notons que nous avons des résultats tout à fait pour la seule autre façon d’implémenter le calcul quantique avec ancilla et une seule opération, à savoir des portes $\widehat{CZ} + \widehat{SWAP}$ au lieu de \widehat{CZ} . Les mesures d’intrication sont un peu plus compliquées dans ce cas-là, cependant, et forcent à considérer l’intrication des paires de qubit avec le reste du registre.

1.2 Inégalités de comparaison entre chaînes de Markov

1.2.1 Mesures de vitesse de convergence

Les chaînes de Markov sont des marches aléatoires sur un espace fini où chaque pas ne dépend que du présent. Leur souplesse leur permettent d'être fréquemment utilisées comme modèles en science, leur facilité d'analyse fournit une théorie riche et leurs propriétés pratiques en font un bon outil en informatique.

Elles sont définies à partir d'un espace fini d'états \mathcal{X} , et de noyaux de Markov K sur \mathcal{X} , aussi appelés matrices de transition. Un noyau de Markov est une matrice indexée par \mathcal{X}^2 , dont toutes les entrées sont positives, et telle que $\sum_{y \in \mathcal{X}} K(x, y) = 1$ pour tout $x \in \mathcal{X}$. L'entrée $K(x, y)$ est la probabilité que la chaîne en x à un temps donné soit en y un pas plus tard.

Sauf précision contraire, nous nous intéresserons à des chaînes de Markov homogènes, de noyau K , où le noyau ne change pas avec le temps. Ainsi, si σ est la distribution de départ sur \mathcal{X} , la distribution au bout de t pas est σK^t . Dans quelques cas toutefois, nous regarderons des chaînes de Markov inhomogènes, où la transition au pas de temps t est donnée par un noyau K_t . On notera toujours x_t pour la valeur atteinte par la chaîne de Markov au temps t .

Une manière de se représenter un noyau de Markov est comme un graphe dont \mathcal{X} est l'ensemble des sommets et l'arête dirigée de x à y est munie d'un poids $K(x, y)$ et absente si $K(x, y) = 0$. Si le graphe est connexe et apériodique, alors le noyau de Markov admet une unique probabilité stationnaire π , c'est-à-dire $\pi K = \pi$. De plus la chaîne de Markov correspondante converge vers π depuis n'importe quelle distribution initiale.

Le noyau de Markov peut être vu comme agissant à droite sur les fonctions sur \mathcal{X} dans $L^2(\pi)$. L'avantage d'inclure π dans le produit scalaire est que K^* devient le renversement du temps pour la chaîne de Markov. En particulier, si son noyau est auto-adjoint, une chaîne de Markov est *réversible*, c'est-à-dire qu'elle est invariante par renversement du temps. Localement, $\pi(x)K(x, y) = \pi(y)K(y, x)$.

La question naturelle est donc d'évaluer la vitesse de cette convergence.

L'approche traditionnelle est basée sur les valeurs propres de K . Le vecteur $(1, \dots, 1)$ est un vecteur propre à droite pour la valeur propre 1. Par Perron-Frobenius, toutes les autres valeurs propres sont de module au plus un. Dans le cas apériodique, elles sont de module strictement inférieur à un. Notons $1 = \beta_1 > |\beta_2| \geq |\beta_3| \geq \dots \geq |\beta_{|\mathcal{X}|}|$ les valeurs propres et V_i les vecteurs propres correspondants. Une simple décomposition de la distribution initiale σ sur les vecteurs propres de K donne $\sigma K^t = \pi + \sum_{i=2}^{|\mathcal{X}|} V_i c_i(\sigma) \beta_i^t$. Du coup, pour toute distance raisonnable, comme la distance L^1 , on obtient $d(\sigma K^t, \pi) = O(\beta_2^t)$. La valeur $1 - \beta_2$ est appelée le trou spectral, et plus il est grand, plus la chaîne de Markov converge vite, asymptotiquement.

Le problème essentiel de cette méthode est son caractère asymptotique. L'une des motivations essentielles pour connaître les vitesses de convergence est l'échantillonnage (HASTINGS, 1970) d'une variable aléatoire par des méthodes de Monte

Carlo (MCMC) : on suit la chaîne de Markov suffisamment longtemps, et quand on s'arrête, la valeur est à peu près échantillonnée suivant la distribution stationnaire π . Ou bien, on veut évaluer l'espérance πf d'une fonction f par rapport à π , et on prend $\frac{1}{T} \sum_{t=1}^T f(x_t)$ pour T assez grand. Mais que veut dire «assez grand»? Ce résultat asymptotique ne le dit pas. Il est possible de le renforcer sous une forme non asymptotique, mais la taille de l'espace $|\mathcal{X}|$ apparaît en facteur, et la borne est en général très pessimiste. Il est possible d'obtenir des estimées plus précises, mais elles nécessitent la pleine connaissance du spectre et des vecteurs propres (DIACONIS, 1988, par exemple).

Un exemple net du pessimisme de cette méthode est le phénomène de cutoff, mis en évidence par DIACONIS (1996). Assez fréquemment, notamment pour les chaînes de Markov avec de fortes symétries, la distance en variation totale entre la distribution au bout de t pas et la distribution stationnaire va rester très proche de 1 jusqu'à un certain T puis décroître exponentiellement à partir de ce T et approcher 0 dès $T + o(T)$. En d'autres termes, la chaîne de Markov atteint le régime asymptotique au temps T , et ne mélange pas du tout avant. La question pertinente en pratique est donc la valeur de T , plutôt que le taux de l'exponentielle. C'est le sens du résultat célèbre qui dit qu'il faut effectuer sept riffle-shuffle pour qu'un paquet de 52 cartes soit mélangé.

JOULIN, OLLIVIER et al. (2010) ont notamment donné des résultats en temps fini sous des conditions de courbure de Ricci discrète positive. Ces conditions ont l'avantage de pouvoir être vérifiées localement, donc facilement. En revanche, cela constitue une forte contrainte sur la chaîne de Markov.

Pour évaluer la vitesse de mélange de manière fine, il nous faut une divergence adaptée. La plus courante est la distance en variation totale $d_{TV}(\sigma, \pi) = \sup_{A \in \mathcal{B}(\mathcal{X})} \pi(A) - \sigma(A)$. Deux autres choix usuels sont $L^2(\pi)$, très pratique pour l'approche spectrale, et la séparation $\sup 1 - \frac{\sigma_i}{\pi_i}$, utile pour son lien avec les temps stationnaires forts ((DIACONIS et J. A. FILL, 1990)).

Dans le cas où π est uniforme, toutes ces distances et plusieurs autres peuvent être traitées simultanément, via la majorisation (MARSHALL et OLKIN, 1979). Si v et w sont deux séries de N nombres ayant la même somme, on dit que v *majorise* w si pour tout $k \leq N$, la somme des k plus grandes entrées de v est plus grande que la somme des k plus grandes entrées de w . On peut écrire toutes les divergences à π qui vont suivre comme des fonctions ϕ Schur-convexes de σ , c'est-à-dire telles que $\phi(\rho_1) \geq \phi(\rho_2)$ si ρ_1 majorise ρ_2 : toutes les distances L^p pour $p \geq 1$, la séparation, la distance de Hellinger, la divergence de Kullback $K(\sigma, \pi)$ et la divergence de Kullback $K(\pi, \sigma)$.

LOVÁSZ et WINKLER (1995) ont imaginé une mesure très différente de la vitesse de mélange, inspirée par l'utilisation en échantillonnage : un temps d'arrêt. Spécifiquement, on considère tous les temps d'arrêt T tels que X_T soit de loi π . On peut alors montrer que l'espérance d'un tel temps d'arrêt est minimale si et seulement s'il existe un état d'arrêt x , c'est-à-dire que si $X_t = x$, alors $T = t$. L'espérance $\mathbb{E}[T] = T_{mix}$ est le temps de mélange de Lovász-Winkler. On peut noter qu'il existe plusieurs constructions, et qu'il existe des constructions pratiques per-

mettant d'approcher π de manière efficace sans connaître complètement le graphe.

J. FILL et J. KAHN (2013) ont mis au point une méthode permettant d'affirmer qu'une chaîne de Markov mélange plus vite qu'une autre au sens où, à *tout temps fini*, la distance de π à la distribution de la chaîne rapide et plus faible qu'à celle de la chaîne lente. Certaines propriétés de monotonie sont nécessaires.

1.2.2 Propriétés élémentaires des inégalités de comparaison

À partir de maintenant, l'espace \mathcal{X} est muni d'un ordre partiel. Tous les produits scalaires sont pris par rapport à π . De plus, dans les sections qui suivent, « Y mélange plus vite que Z en d » a le sens très fort que pour tout temps, Y est plus proche de π que Z pour la divergence d .

Soient \mathcal{K} , \mathcal{M} et \mathcal{F} les ensembles respectivement (i) des noyaux de Markov sur \mathcal{X} de distribution stationnaire π , (ii) des fonctions positives monotones sur \mathcal{X} , (iii) des noyaux de \mathcal{K} stochastiquement monotones, c'est-à-dire tels que pour tout $f \in \mathcal{M}$, on a encore $Kf \in \mathcal{M}$. De manière équivalente, si $x \leq y$, alors la distribution de probabilité au bout d'un pas partant de y domine stochastiquement celle partant de x : $K(x, D) \geq K(y, D)$ pour toute section commençante D , c'est-à-dire que si $z \leq w$ et $w \in D$, alors $z \in D$.

On définit les *inégalités de comparaison* comme la relation d'ordre partiel sur \mathcal{K} qui dit que $K \preceq L$ si $\langle Kf|g \rangle \leq \langle Lf|g \rangle$ pour tout $f, g \in \mathcal{M}$. Les inégalités de comparaison sont très stables :

Proposition 1.2.1 (J. FILL et J. KAHN, 2013, Propositions 2.3 et 2.9).

1. Si $K \preceq L$, alors $K^* \preceq L^*$.
2. Si $K_n \preceq L_n$ pour tout n et $K_n \rightarrow K$ et $L_n \rightarrow L$, alors $K \preceq L$.
3. Si $K_0 \preceq L_0$ et $K_1 \preceq L_1$, alors pour tout $0 \leq \lambda \leq 1$:

$$\lambda K_0 + (1 - \lambda)K_1 \preceq \lambda L_0 + (1 - \lambda)L_1.$$

4. Si \mathcal{X} est partitionné en \mathcal{X}_0 et \mathcal{X}_1 avec les distributions stationnaires et ordres partiels induits, et si $K_i \preceq L_i$ sur \mathcal{X}_i pour $i = 0, 1$, alors avec le noyau K (resp. Y) somme directe de K_0 et K_1 (resp. L_0 et L_1), on a $K \preceq L$.
5. Si K_1, \dots, K_t et L_1, \dots, L_t sont tous dans \mathcal{F} , ainsi que leurs adjoints, et si $K_i \preceq L_i$ pour tout $1 \leq i \leq t$, alors $K_1 \dots K_t \preceq L_1 \dots L_t$ et ces deux noyaux sont dans \mathcal{F} .

J'ai laissé le dernier point ici pour montrer que l'on n'a pas besoin, *stricto sensu*, de la réversibilité pour que la notion soit utile. Dans un certain nombre de cas, on a besoin que ce soit le renversement en temps de la chaîne de Markov qui soit monotone, et pas nécessairement la chaîne de Markov elle-même. Cependant, le cas le plus simple reste celui des chaînes homogènes réversibles, et nous formulons les théorèmes suivants dans ce cadre.

1.2.3 Comparaisons de vitesse générales

Les inégalités de comparaison entraînent facilement que si Y est de noyau K et Z de noyau L , tous deux réversibles dans \mathcal{F} , avec $K \preceq L$ et une distribution de départ commune $\hat{\pi}$ telle que $\hat{\pi}/\pi$ soit décroissante, alors Y domine Z . En effet l'indicatrice d'une section commençante D est une fonction décroissante, et donc

$$\mathbb{P}[Y_t \in D] = \langle K^t \mathbf{1}_D | \frac{\hat{\pi}}{\pi} \rangle \leq \langle L^t \mathbf{1}_D | \frac{\hat{\pi}}{\pi} \rangle = \mathbb{P}[Z_t \in D].$$

On remarque également que dans ce cadre, $\mathbb{P}[Y_t = \cdot] / \pi(\cdot)$ est une fonction décroissante pour tout t . Avec la domination, ceci implique $\mathbb{P}[Z_t = \cdot] / \pi(\cdot)$ majorise $\mathbb{P}[Y_t = \cdot] / \pi(\cdot)$. En conséquence :

Théorème 1.2.2 (J. FILL et J. KAHN, 2013, Corollaires 3.3 et 3.7).

Si K et L sont réversibles, dans \mathcal{F} , et vérifient $K \preceq L$, et si $\hat{\pi}/\pi$ est décroissante, alors la chaîne de Markov Y de noyau K mélange plus vite que la chaîne Z de noyau L , en variation totale, en séparation et en L^2 . Si de plus, π est uniforme, alors Y mélange plus vite que Z dans tous les L^p pour $p \geq 1$, en distance de Hellinger et en divergence de Kullback.

Une astuce permet, dans le cas L^2 , de demander seulement que $K^2 \preceq L^2$ et soient stochastiquement monotones, pas nécessairement K et L .

1.2.4 Ajouter des pas ne ralentit pas une chaîne de Markov

(PERES et WINKLER, 2013) ont demandé si ajouter des pas à une chaîne de Markov pouvait ralentir sa convergence. Intuitivement, la réponse est non, et ils ont pu le prouver dans le cas des systèmes de spin monotones pour la variation totale. HOLROYD (2011) a trouvé des contre-exemples.

Ajouter des pas revient à comparer à une autre une chaîne de Markov inhomogènes où certains K_t ont été remplacés par des I , le noyau identité, qui correspond à ne pas bouger.

La proposition 1.2.1 met en évidence que les inégalités de comparaison se comportent bien avec des chaînes inhomogènes. Notons que le noyau identité I , correspondant à la marche qui ne bouge pas, est toujours dans \mathcal{F} . Il suffit donc de prouver que $I \preceq K_t$ et que tous les noyaux concernés sont dans \mathcal{F} et réversibles. Dans ce cas, pour de bonnes distributions initiales, le théorème 1.2.2 implique immédiatement que les pas supplémentaires n'ont pu que rapprocher en distance variation totale, L^2 , et en séparation.

Ce programme est assez facilement mis en œuvre dans les cas suivants, dont le dernier généralise les résultats de PERES et WINKLER (2013) :

Proposition 1.2.3 (J. FILL et J. KAHN, 2013, Théorèmes 8.3, 8.5 et 8.6).

Ajouter des pas à la marche aléatoire ne peut qu'accélérer la convergence si la distribution initiale $\hat{\pi}$ vérifie que $\hat{\pi}/\pi$ est décroissante et soit :

- L'espace \mathcal{X} est totalement ordonné et K_t et π est uniforme.
- L'espace \mathcal{X} est l'ensemble des permutations muni de l'ordre de Bruhat, et les pas sont choisis parmi les K_i : on range les cartes en position i et $i + 1$ avec probabilité p , et on les range en ordre inverse sinon.
- L'espace \mathcal{X} est un ensemble de configuration de spins : sur chaque sommet d'un graphe (V, E) se trouve un élément de S , qui est partiellement ordonné. L'ordre sur \mathcal{X} est le produit des ordres sur chaque sommet. De plus, il faut que la distribution d'équilibre π soit monotone au sens où la loi d'un sommet conditionnellement à l'état de tous les autres sites soit monotone en tous les autres sites. On suppose enfin que ces lois conditionnelles π_v vérifient $\langle f|g \rangle_{\pi_v} \geq \langle f|1 \rangle_{\pi_v} \langle g|1 \rangle_{\pi_v}$ pour tout $f, g \in \mathcal{M}$. Enfin, les pas sont de la forme : on choisit un v et on choisit son état suivant π_v .

1.2.5 Chaînes de naissance et de mort

Une chaîne de naissance et de mort est une chaîne de Markov sur $\mathcal{X} = \{0, 1, \dots, n\}$ dont les transitions se font entre entiers consécutifs, ou sur place. Toutes les chaînes de ce paragraphe partent de zéro. Toutes ces chaînes sont réversibles.

La probabilité stationnaire π est uniforme si et seulement si K est symétrique. Dans ce cas, on peut vérifier que K^2 est monotone, même si K ne l'était pas.

Nous appelons *chaîne uniforme* la chaîne de Markov symétrique dont les transitions U sont $U(i, i + 1) = 1/2$ et $U(0, 0) = U(n, n) = 1/2$. Un court calcul montre que $U^2 \preceq K^2$. Ainsi, pour tout temps pair, la distribution σ_t de la chaîne uniforme est majorisée par celle de n'importe quelle chaîne de vie et de mort.

Ce résultat peut être généralisé à tous les temps par des méthodes *ad hoc*. On obtient :

Théorème 1.2.4 (J. FILL et J. KAHN, 2013, Théorème 4.3). *Si X est une chaîne de naissance et de mort symétrique, et π_t sa distribution après t pas partant de 0 majorise σ_t la distribution de la chaîne uniforme, pour tout temps t .*

Ceci implique que la chaîne uniforme mélange plus vite que toute autre chaîne pour toutes les divergences couvertes par la majorisation.

C'est également la chaîne avec le plus faible temps de mélange de Lovász-Winkler, si n est pair. Mais pas si n est impair. Contre-exemple : si $n = 1$, un temps d'arrêt de Lovász-Winkler pour l'uniforme est juste d'attendre un pas, espérance 1. La chaîne symétrique avec $K(0, 1) = 1$ permet, elle, de s'arrêter avec une chance sur deux au temps 0, et une chance sur deux au temps 1. On obtient bien X_T de loi uniforme, et $\mathbb{E}[T] = \frac{1}{2}$.

Plus généralement, un temps d'arrêt est de Lovász-Winkler si X_T est uniforme et s'il existe un état d'arrêt. Un T qui remplit ces conditions pour une chaîne de vie et de mort est simplement de tirer j selon π , et d'arrêter la chaîne quand elle atteint j . En effet, n n'est atteinte qu'après tous les autres états, donc la chaîne s'arrête

toujours quand $x_t = n$. Donc

$$T_{mix} = \sum_{i=0}^n \pi_i T_i,$$

où T_i est le premier temps de passage en i partant de zéro. Quelques manipulations algébriques dans le cas uniforme permettent d'arriver à

$$T_{mix} = \sum_{k=0}^{n-1} \frac{(k+1)(n-k)}{p_k},$$

où $K(i, i+1) = p_i$. Partant de là, il est possible d'arriver à :

Théorème 1.2.5. *La chaîne de vie et de mort symétrique sur $\mathcal{X} = \{0, \dots, n\}$ au départ de 0 qui minimise le temps de mélange de Lovász-Winkler T_{mix} est :*

- la chaîne uniforme si n est pair.
- si n est impair, alors :

$$p_k = \begin{cases} 1 - \theta_n & \text{si } k \text{ est pair} \\ \theta_n & \text{si } k \text{ est impair} \end{cases} \quad (k = 0, \dots, n-1), \quad (1.3)$$

avec pour tout m :

$$\theta_{m-1} := \frac{1}{6} \left[\sqrt{(m^2 + 2)(m^2 - 4)} - (m^2 - 4) \right]. \quad (1.4)$$

Quitte parfois à se restreindre aux noyaux monotones, une partie de ces résultats se généralise au cas où la distribution stationnaire n'est pas celle uniforme.

Chapitre 2

Images en science

Le monde des sciences naturelles compte de nombreuses données organisées spatialement, éventuellement sur une grille régulière. Il ne s'agit toutefois pas toujours nécessairement d'images *stricto sensu*. Soit parce que l'organisation spatiale n'est pas celle des données, mais celle de la cible, comme en IRM où on observe essentiellement la transformée de Fourier de l'image d'intérêt. Soit parce que la donnée en un point spatial (pixel ou voxel) n'est pas réduite à une intensité. Un tel pixel peut soit contenir des informations directement intéressantes, soit contenir des données restant à analyser. Un exemple de données intéressantes pourrait être la présence ou non de molécules à cet endroit, pour plusieurs types simultanément. Un exemple à analyser serait une série de temps d'arrivée dont le ou les temps caractéristiques sont l'information cible, comme en imagerie par fluorescence FLIM-FRET. Un exemple intermédiaire pourrait être une série temporelle d'un voxel du cerveau en IRM fonctionnelle.

Dans tous les cas, l'analyse de ces données, et l'optimisation de leur acquisition, posent des problèmes statistiques originaux. Ils peuvent faire appel tout à la fois à des techniques classiques d'analyse d'images (le traitement du signal à des fins non artistiques est une branche des statistiques) et à des techniques de statistiques paramétriques, ou non-paramétriques plus générales.

Ce qui suit est un florilège de thématiques que j'ai rencontrées en essayant de me rapprocher d'expérimentateurs et d'applications, tant les problèmes originaux que les questions spécifiquement mathématiques qui ont pu être soulevées.

La section 2.1 décrit la microscopie par temps de vie de fluorescence et transfert d'énergie à résonance (FLIM-FRET), et les données qu'elle génère. Nous mentionnons quelques méthodes d'analyse et quelques suggestions pour les analyser. En particulier, pixel par pixel, nous sommes confrontés à un problème d'estimation des paramètres d'une densité de mélange. La section 2.2 est donc consacrée aux vitesses d'estimation optimales dans les mélanges, venant corriger les taux faux de la littérature.

La section 2.4 est consacrée à l'acquisition compressée en IRM, avec ses contraintes physiques, et suggère différentes applications des algorithmes développés dans ce

but.

La section 2.3 évoque rapidement des algorithmes de clustering ultra-rapides aux propriétés adaptées à l'IRM fonctionnelle.

2.1 Mesures d'interactions protéine-protéine par FLIM-FRET

La microscopie par fluorescence prend véritablement son essor quand CHALFIE et al. (1994) parviennent à inclure le code de la GFP (green fluorescent protein) dans l'ADN d'une protéine quelconque d'une cellule. Cela permet d'ajouter à la protéine fabriquée par la cellule une petite partie fluorescente, dont on espère qu'elle n'influera pas sur sa fonction. Il est ensuite possible de voir où ces protéines se trouvent dans la cellule vivante.

De nombreuses améliorations et astuces permettent d'accéder à des données éventuellement subtiles (LAKOWICZ, 2013), comme les coefficients de diffusion d'une protéine à une échelle donnée. Nous allons nous intéresser au transfert d'énergie par résonance de type FÖRSTER (1948), ou FRET.

Il s'agit d'un transfert d'énergie entre un fluorophore donneur dont la fréquence d'émission correspond à la fréquence de réception d'un fluorophore accepteur. Ce transfert n'a lieu que si les fluorophores sont très proches, à moins de 10 nanomètres l'un de l'autre. Ainsi, si deux protéines sont marquées respectivement par le donneur et l'accepteur, et qu'un tel transfert est détecté en un point de la cellule, ces deux protéines sont très proches. On espère que cela met en évidence une interaction entre ces deux protéines.

L'une des manières les plus fiables de détecter ce FRET est d'observer les temps de vie de fluorescence (Fluorescence Lifetime Imaging Microscopy), ce qui peut se faire par comptage de photons uniques (DUNCAN et al., 2004). Plus précisément, une molécule fluorescente excitée peut se désexciter de plusieurs manières, notamment en émettant un photon : c'est la fluorescence. Toutes les manières de se désexciter sont soit strictement quantiques, et donc ont une demi-vie, soit liées à un environnement suffisamment aléatoire pour qu'une telle approximation soit possible. Ainsi, si un photon de fluorescence est détecté, celui-ci l'est après un temps aléatoire exponentiel de paramètre naturel λ . Le FRET ajoute un mode de désexcitation. Donc si λ_D est le paramètre sans FRET, et λ_F le paramètre du FRET seul, le paramètre en présence de FRET (à une distance donnée) est $\lambda_{DA} = \lambda_D + \lambda_F$.

En termes peut-être plus intuitifs de temps de vie, $\tau_{DA} = \frac{1}{\tau_D^{-1} + \tau_F^{-1}}$. Le temps de vie de la fluorescence du donneur diminue en présence de FRET. Il est de l'ordre de 10^{-9} secondes. Expérimentalement, une impulsion laser est envoyée à une fréquence d'excitation du donneur, et le microscope détecte l'émission du donneur, photon par photon, avec les temps d'arrivée à une précision de l'ordre de 10^{-11} secondes (WAHARTE ET AL., 2006). Mathématiquement, on a donc des échantillons tirés suivant la loi du temps de vie du donneur.

Les observations se font sur un microscope confocal : un tout petit volume est observé, puis on change le point d'observation. Ainsi, on obtient une suite de temps d'arrivée en chaque pixel d'une image. Les volumes sont toutefois assez grands pour qu'il n'y ait jamais que des molécules en interaction, toutes seules. Spécifiquement, l'intensité en un temps t après chaque impulsion laser est donnée par

$$I(t) = I_0 + \sum_i I\pi_i \exp\left(-\frac{t}{\tau_i}\right), \quad (2.1)$$

où I_0 correspond à diverses sources de bruit constant et est connu, I est l'intensité totale du donneur juste après l'impulsion, proportionnelle au nombre de molécules du donneur dans le volume, et π_i est la proportion des molécules du donneur dans la situation i : typiquement, i est soit «éloigné de l'accepteur», soit «proche de l'accepteur». Mais il peut parfois y avoir d'autres influences. Les τ_i sont les temps de vie correspondants.

La dépendance des temps de vie en l'environnement fait que les τ_i ne sont généralement pas supposés connus. La première question d'intérêt est de savoir si en un pixel, les temps d'arrivée sont générés par une mono-exponentielle (probablement pas d'interaction) ou par une multi-exponentielle. HEINRICH, Jonas KAHN et al. (2011) met en évidence qu'un simple test du maximum de vraisemblance permet de répondre à cette question avec dix fois moins d'observations qu'un test du chi-deux.

D'une manière plus générale, déterminer les proportions et temps de vie en chaque point est l'objectif. Il se trouve que, sans fixer d'*a priori* sur le problème, comme de connaître le temps de vie sans accepteur, ce problème d'estimation de paramètres dans un modèle de mélange est extrêmement difficile. La détermination des vitesses optimales est l'objet de la prochaine section. Mentionnons simplement que si les temps caractéristiques d'une bi-exponentielle sont dix fois plus proches, il faut 10^6 fois plus de données pour avoir la même précision relative. . .

Les méthodes traditionnelles d'estimation de biologistes sont certainement perfectibles (ajustement de courbe L^2 . . .). Plus fondamentalement, cependant, nous avons affaire à une «image». Il doit donc être possible de débruiter l'estimation des différents paramètres comme on le fait pour des images générales. Voire en mutualisant directement les données des différents pixels. À partir du moment où l'on trouve une bonne mesure de similitude entre différents pixels, ou petits carrés de pixels, il est possible d'employer des moyennes non locales entre petits carrés, par exemple. Cela permettrait d'augmenter largement les données effectives en chaque pixel.

Dans ce sens, HEINRICH, PISFIL et al. (2014) montrent que la distance de transport entre les probabilités empiriques entre chaque pixel sont un indicateur de similarité relativement robuste.

D'autres directions d'amélioration viennent de l'impact possible de l'analyse mathématique sur les méthodes expérimentales, comme illustré par REBAFKA (2009).

Celle-ci montre comment étudier les temps d'arrivée malgré le problème d'empilement. Le problème d'empilement est le suivant : si deux photons de fluorescence sont émis en réponse à la même impulsion laser, le second n'est pas détecté. Du coup la distribution des temps d'arrivée change, mais de manière connue. Auparavant, les expérimentateurs s'assuraient d'avoir de faibles intensités pour que ce cas arrive rarement, et utilisaient les formules à base d'exponentielles. Les travaux de REBAFKA (2009) permettent une acquisition plus rapide en autorisant de plus fortes intensités. Dans le même ordre d'idée, une description physique précise du photoblanchiment, qui interrompt durablement la fluorescence des protéines exposées à trop de lumière trop longtemps, permettrait sans doute une étude davantage prolongée.

2.2 Vitesses d'estimation optimales dans les modèles de mélange

2.2.1 Mélanges

Motivé par l'estimation des temps de vie de fluorescence (2.1), j'ai regardé la littérature sur l'estimation des paramètres de modèles de mélange. Il se trouve que les vitesses qui y sont données sont erronées. Voici les corrections.

Un *mélange* est une loi de probabilité de la forme

$$P(\cdot, G) = \int P(\cdot, \theta) dG(\theta), \quad (2.2)$$

où G est la *loi de mélange*, une probabilité sur l'espace Θ des paramètres θ , et où les $F(\cdot, \theta)$ sont des lois de probabilité sur le même espace pour tout θ .

Les mélanges sont principalement utilisés dans trois cas. Premièrement, en classification, où il s'agit d'étiqueter chaque donnée comme appartenant à un groupe (MCLACHLAN et PEEL, 2000), une approche consiste à supposer que chaque groupe génère des données suivant une loi $F(\cdot, \theta)$ pour différents θ . Du coup, l'ensemble des données a la distribution du mélange.

Deuxièmement, et c'est peut-être le cas le plus courant, les mélanges sont utilisés simplement pour leur souplesse pour représenter des données hétérogènes. Dans ce cas, le but du statisticien est d'approcher $P(\cdot, G)$, la mesure suivie par les données (GENOVESE et WASSERMAN, 2000, par exemple).

Troisièmement, et c'est le cas qui nous intéresse, les données sont produites par un processus qui s'interprète directement comme un mélange, et le but est de trouver les paramètres du mélange eux-mêmes : nous cherchons G et pas P .

Ce troisième problème est nettement plus dur que le second. Identifier P se fait à vitesse paramétrique $n^{-1/2}$ dans les cas habituels, multiplié par une puissance de $\ln(n)$ si G n'est pas à support fini. Par contre, un exemple d'identification de G est la déconvolution, où G est la vraie fonction, et $P(\cdot, \theta)$ est la loi du bruit de convolution déplacé de θ . Dans ce cas, les vitesses minimax sont logarithmiques.

Nous nous placerons désormais dans le cas de mélanges finis, donc G est une somme de pics de Dirac. Nous notons \mathcal{G}_m les lois de mélange à exactement m composantes distinctes, et $\mathcal{G}_{\leq m}$ l'ensemble des lois de mélange à au plus m composantes. Sauf précision contraire, nous supposons que nous savons que le vrai mélange a au plus m composantes. Par ailleurs, nous supposons que P est une probabilité sur \mathbb{R} , et nous noterons $f(\cdot, \theta)$ les densités correspondantes, et $F(\cdot, \theta)$ leurs fonctions de répartition. De plus nous supposons que Θ est un compact de \mathbb{R} .

Ce qui suit peut sans doute se généraliser facilement à des probabilités sur d'autres espaces que \mathbb{R} , ou pour Θ un paramètre à plusieurs dimensions. La non-compacité imposerait sans doute d'autres conditions d'identifiabilité, par contre.

La distance adaptée sur les lois de mélange est la distance de transport (L^1), qui élimine tout problème d'identifiabilité :

$$W(G_1, G_2) = \sup_{|f|_{Lip} \leq 1} \int_{\Theta} f(\theta) d(G_1 - G_2)(\theta), \quad (2.3)$$

où $\|\cdot\|_{Lip}$ représente la semi-norme de Lipschitz. Nous noterons $\mathcal{W}_G(\varepsilon)$ les boules de rayon ε autour de G , en distance de transport.

Les vitesses connues étaient $n^{-1/4}$, ce qui est étrange quand on compare à la déconvolution, certes un mélange infini. Mais d'habitude, quand les taux sont constants pour tout nombre fini de paramètres, ce sont les mêmes pour un nombre infini de paramètres. Nous allons résoudre ce paradoxe.

2.2.2 Vitesse minimax

Donnons une intuition de ce qui se passe. Les observations sont la fonction de répartition empirique F_n . Par l'inégalité de Dvoretzky-Kiefer-Wolfowitz, la distance à la vraie fonction de répartition F décroît en racine de n :

$$\|F_n - F\|_{\infty} \approx n^{-1/2}. \quad (2.4)$$

Nous considérons l'estimateur \hat{G}_n du minimum de distance de DEELY et KRUSE (1968) :

$$\|F(\cdot, \hat{G}_n) - F_n\|_{\infty} = \inf_{G \in \mathcal{G}_{\leq m}} \|F(\cdot, G) - F_n\|_{\infty}. \quad (2.5)$$

Par l'inégalité triangulaire et la remarque (2.4), le mélange $F(\cdot, \hat{G}_n)$ est à distance $n^{-1/2}$ du vrai mélange F . Donc, si nous avons un contrôle sur la distance de transport $W(G_1, G_2)$ en fonction des distances entre mélanges $\|F(G_1) - F(G_2)\|_{\infty}$, nous obtenons du même coup la vitesse de convergence de l'estimateur de G .

Supposons maintenant que nous savons que G est à proximité d'un certain G_0 ,

de telle sorte que G est de la forme :

$$G = \sum_{i=1}^{m_0} \sum_{j=1}^{J_i} \pi_j \delta_{\theta_i + \varepsilon_n h_j},$$

$$\sum_{j=1}^{J_i} \pi_j = \pi_i + O(\varepsilon_n).$$

Alors un développement limité en θ donne

$$F(\cdot, G) = \sum_{k=0}^K \varepsilon_n^k \sum_{i=1}^{m_0} \left(\sum_{j=1}^{J_i} \frac{\pi_j h_j^k}{k!} \right) F^{(k)}(\cdot, \theta_i) + O(\varepsilon_n^{K+1}), \quad (2.6)$$

où les $F^{(k)}$ sont des dérivées par rapport à θ .

Nous voyons donc que l'on peut s'attendre à une distance entre $F(\cdot, G_1)$ et $F(\cdot, G_2)$ de l'ordre de ε_n^{K+1} si nous pouvons nous assurer des égalités de moments $\sum_{j=1}^{J_{i,1}} \frac{\pi_{j,1} h_{j,1}^k}{k!} = \sum_{j=1}^{J_{i,2}} \frac{\pi_{j,2} h_{j,2}^k}{k!}$ pour tout $i \leq m_0$ et tout $k \leq K$.

Nous disposons donc de $2J_i$ paramètres pour la composante i , à savoir les π_j et h_j pour $j \in [1, J_i]$, et nous souhaitons disposer d'au moins deux solutions différentes à $(K+1)$ équations, à savoir $\sum \pi_j h_j^k = c_k$ pour $k \in [0, K]$. Ceci va être possible si $K+2 \leq 2J_i$. Comme il faut au moins une composante auprès de chaque composante θ_i de G_0 , nous avons au plus $J_i = (m - m_0 + 1)$ composantes auprès de θ_i . Donc la distance entre les mélanges de loi G_1 et G_2 près de G_0 est au moins d'ordre $\varepsilon_n^{-(2(m-m_0)+1)}$.

De plus, comme les h_j sont à l'échelle ε_n , on peut s'attendre à ce que la distance entre G_1 et G_2 soit ε_n . Ainsi, en réutilisant la distance (2.4), nous obtenons une vitesse d'estimation en $n^{-\frac{1}{4(m-m_0)+2}}$.

Pour rendre le raisonnement ci-dessus rigoureux, il faut s'assurer que :

- Le problème soit suffisamment lisse en θ pour prendre les dérivées.
- Une condition d'identifiabilité k -forte est satisfaite pour k assez grand, à savoir : pour toute famille finie de θ_j distincts, l'égalité

$$\left\| \sum_{p=0}^k \sum_j \alpha_{p,j} F^{(p)}(\cdot, \theta_j) \right\| = 0$$

implique $\alpha_{p,j} = 0$ pour tout p et j . En effet, nous l'utilisons pour dire que l'égalité (ou presque) des moments est nécessaire pour l'égalité (ou presque) des distributions dans le développement (2.6).

- Si $W(G_1, G_2) \ll \varepsilon_n$, alors $\|F(\cdot, G_1) - F(\cdot, G_2)\| \geq \delta W(G_1, G_2)^{2(m-m_0)+1}$ reste vrai pour δ qui ne dépend que de G_0 . Cela revient à gérer les cas où les G_i convergent vers G_0 asymptotiquement suivant les mêmes lignes. Il

est nécessaire pour cela d'être plus soigneux dans les développements limités, en les effectuant par sous-ensembles de coefficients organisés selon une structure d'arbre. Ça marche.

- Aucun estimateur ne peut faire mieux, à constante près. Cela nécessite certaines hypothèses de régularité, essentiellement le fait que des espérances du type $\mathbb{E}_{\theta_1} \left| \frac{f^{(p)}(\cdot, \theta_2)}{f(\cdot, \theta_3)} \right|^q$ sont finies.

Nous obtenons finalement, en utilisant la compacité de Θ pour le passage du local au global :

Théorème 2.2.1 (HEINRICH et Jonas KAHN, 2015, Théorèmes 3.2 et 3.3). *Sous des hypothèses de régularité et d'identifiabilité suffisantes (et explicites), la vitesse d'estimation minimax autour de G_0 à m_0 composantes dans l'espace $\mathcal{G}_{\leq m}$ des mélanges à au plus $m \geq m_0$ composantes est en $n^{-\frac{1}{4(m-m_0)+2}}$. C'est-à-dire que si on note \hat{G}_n toute séquence d'estimateurs et $\varepsilon_n = n^{-\frac{1}{4(m-m_0)+2} + \kappa}$ pour un $\kappa > 0$:*

$$\infty > \liminf_{n \rightarrow \infty} \inf_{\hat{G}_n} \sup_{G_1 \in \mathcal{G}_m \cap \mathcal{W}_{G_0}(\varepsilon_n)} n^{1/(4(m-m_0)+2)} \mathbb{E}_{G_1} \left[W(G_1, \hat{G}_n) \right] > 0.$$

Sous les mêmes conditions, la vitesse d'estimation minimax sur \mathcal{G}_m globalement est en $n^{-1/(4m-2)}$.

Plusieurs constats sont possibles à l'énoncé du théorème. Le premier est que la situation empire avec le nombre de composantes possibles. Il n'est donc pas surprenant que les taux pour l'estimation des lois de mélanges infinis ne soient pas paramétriques.

Ensuite, c'est le doute sur le nombre de composantes qui est gênant : le pire cas est quand le mélange ressemble beaucoup à un mélange à une seule composante. Plusieurs composantes peuvent ressembler fortement à plusieurs composantes très différentes en égalisant leurs moments. Notons que le problème n'est pas ce qui se passe si le vrai mélange a peu de composantes, mais ce qui se passe à proximité, c'est-à-dire si le vrai mélange est suffisamment proche, à n donné, d'un mélange à peu de composantes.

Quand le nombre de composantes est connu, c'est-à-dire que $m = m_0$, nous obtenons une vitesse en $n^{-1/2}$. Cela n'est pas très surprenant puisque nous sommes alors dans une situation paramétrique classique, au moins localement. Mais cela motive l'étude de la vitesse de convergence point par point.

2.2.3 Vitesse point par point et interprétation

Sous des conditions d'identifiabilité, il existe des estimateurs du nombre de composantes tel que, pour tout G , l'estimateur est exact avec une probabilité $1 - \varepsilon_n$, avec $\varepsilon_n \ll n^{-1/2}$. En fait, le risque asymptotique est même une exponentielle amortie. Par exemple, si $G_{n,m}$ est l'estimateur du minimum de distance dans $\mathcal{G}_{\leq m}$, alors

$$\hat{m} = \hat{m}_n = \inf \left\{ m \geq 1 : \|F(\cdot, \hat{G}_{n,m}) - F_n\|_\infty \leq n^{-1/2+\kappa} \right\}, \quad (2.7)$$

pour $\frac{1}{2} > \kappa > 0$, convient.

Or, si on connaît le nombre de composantes, l'estimation peut se faire en $n^{-1/2}$, puisque nous sommes dans le cas paramétrique. Nous en déduisons :

Théorème 2.2.2 (HEINRICH et Jonas KAHN, 2015, Théorème 3.5). *Sous des conditions suffisantes de régularité (moindres que pour le théorème 2.2.1), il existe des estimateurs tels que pour toute loi de mélange fini $G \in \mathcal{G}_{<\infty}$, l'estimateur converge vers G à vitesse $n^{-1/2}$:*

$$\mathbb{E}_G \left[W(\widehat{G}_n, G) \right] = C(G)n^{-1/2}, \quad (2.8)$$

où $C(G)$ ne dépend que de G .

À première vue, les théorèmes 2.2.1 et 2.2.2 peuvent sembler se contredire. La différence entre une vitesse d'estimation minimax et une vitesse d'estimation point par point pour tous les points est assez rare en statistique pour prendre la peine de mettre en lumière ce qu'elle signifie.

La différence entre les deux concepts est l'uniformité. Tout point peut être approché à vitesse $n^{-1/2}$, mais le moment où ce régime asymptotique est atteint dépend du point. Donc le pire cas parmi tous les points, ou même tous les points d'une petite boule, n'est pas forcément en $n^{-1/2}$. Une autre manière de voir les choses dans le cas présent est que la constance $C(G)$ dans le risque (2.8) explose quand G s'approche de certains G_0 , spécifiquement ceux qui ont strictement moins de composantes.

En pratique, cela signifie que si certaines composantes d'un mélange sont très proches, il faut un nombre phénoménal d'observations pour bien estimer le mélange. Il est plus facile de déterminer le nombre de composantes, et les moments d'ordre inférieur. Mais si par exemple il faut N observations pour distinguer G_1 de G_2 avec probabilité 0.9, et que ces deux mélanges ont trois composantes bien séparées, il faudra $10^{10}N$ pour séparer les lois de mélange où les θ_i ont été divisés par 10. Irréalizable en général.

Ainsi, si un expérimentateur a le choix, via le choix d'un marqueur par exemple, il convient de séparer autant que possible les composantes du mélange, même quitte à avoir beaucoup moins d'échantillons.

2.3 IRM fonctionnelle

Le titre de cette section est peut-être un peu trompeur. Il fait référence à mes dialogues avec Gaël Varoquaux, et j'évoquerai une autre idée en fin de section. L'essentiel de celle-ci, néanmoins, portera sur le clustering par plus proches voisins. Le problème qu'il vise à résoudre est la réduction de dimension rapide pour des données suffisamment continues spatialement. Il peut certainement être appliqué dans d'autres domaines.

En IRM fonctionnelle, on observe le cerveau en haute résolution, de l'ordre de 10^5 ou 10^6 voxels à l'heure actuelle (ZALESKY et al., 2014). L'activité du cerveau en chaque voxel est mesurée dans le temps, avec plusieurs modes d'analyse possible. Le plus simple est de remplacer chaque voxel par une activité moyenne standardisée durant le temps d'observation.

On obtient ainsi une image pouvant faire 10^6 voxels. Voire un grand nombre d'images, associées à différents sujets ou différentes activités. Le but final pourra être de déterminer quelles zones sont actives pour une activité donnée, ou de différencier des sujets en fonction de leur état de santé, par exemple. Dans tous les cas, ces images sont donc utilisées en entrée d'autres algorithmes, comme une analyse en composantes indépendantes ou un algorithme de classification.

Ces algorithmes souffrent d'une grande lenteur et exigence mémoire pour des entrées d'une dimension pareille. Il est donc utile de réduire cette dimension. Pour être clair, nous parlons de réduire à 10^4 dimensions, gardant ainsi l'essentiel du signal, pas de résumer en trois paramètres. Mais les algorithmes de réduction de dimension doivent eux-mêmes être rapide et suffisamment fidèles.

Une approche traditionnelle *a priori* efficace est celle des projections aléatoires : une projection de l'espace d'état de dimension n sur un espace aléatoire de dimension k avec $k \ll n$. La projection est très rapide, et il y a garantie de préservation approximative de la norme L^2 via le lemme de JOHNSON et LINDENSTRAUSS (1984). Toutefois les projections aléatoires présentent l'inconvénient de ne pas tenir compte de la structure des images : celles-ci sont en général continues par morceaux, et utiliser ce phénomène doit permettre une plus grande efficacité. Par ailleurs, un axe de projection n'a pas de signification intuitive, comme elle est «répartie» sur toute l'image.

THIRION et al. (2015) ont donc recherché une méthode de clustering des voxels. Les voxels sont rassemblés en «super-voxels», vus comme une manière de compresser les données en perdant peu d'information, tout en pouvant opérer directement sur les données compressées. L'heuristique suivante permet d'avoir une idée de quelles sont les bonnes propriétés d'un tel algorithme.

On observe $Y = X + \varepsilon$, où Y et X sont des fonctions de Ω dans \mathbb{R} où Ω est typiquement un sous-ensemble de \mathbb{Z}^3 . Nous notons X et Y comme des vecteurs de \mathbb{R}^n , avec $X_\omega = X(\omega)$. La réalité sous-jacente X est supposée L -Lipschitz, par exemple pour la distance de graphe. Le bruit ε est supposé gaussien de variance σ^2 , indépendant pour chaque ω . Supposons que le clustering Φ se fasse sur une partition (C_1, \dots, C_k) de Ω indépendante des données. Nous notons $i(\omega)$ le i tel que $\omega \in C_i$, et $\bar{X}_i = \frac{1}{|C_i|} \sum_{\omega \in C_i} X_\omega$.

Alors $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ et $(\Phi Y)_i = \sqrt{|C_i|} \bar{Y}_i$. Donc ΦY est un vecteur gaussien de variance σ^2 et de moyenne $\sqrt{|C|} \bar{X}$, coordonnée par coordonnée. En notant Φ^\dagger

le pseudo-inverse de Φ , on peut ramener la projection sur l'espace d'origine :

$$\begin{aligned}\Phi^\dagger \Phi Y_\omega &= \bar{Y}_{i(\omega)} \\ &= \bar{X}_{i(\omega)} + \frac{1}{|C_i|} \sum_{\chi \in C_i(\omega)} \varepsilon_\chi.\end{aligned}\quad (2.9)$$

Pour quantifier la fidélité, nous comparons $\|X\|_2^2$ et $\|\Phi Y\|_2^2$:

$$\begin{aligned}\|\Phi Y\|_2^2 &= \sum |C_i| \bar{X}_i^2 + \varepsilon_i^2 \\ &= \sum_\omega X_\omega^2 - \left(X_\omega^2 - \bar{X}_{i(\omega)}^2 \right) + F \\ &= \|X\|_2^2 + F - \sum_i |C_i| \text{Var}_i(X)\end{aligned}\quad (2.10)$$

où F suit une loi $\chi^2(k)$, et $\text{Var}_i(X)$ est la variance intra-cluster de X , soit $\text{Var}_i(X) = \frac{1}{|C_i|} \sum_{\omega \in C_i} (X(\omega) - \bar{X}_{i(\omega)})^2$.

On observe donc :

- L'expression (2.9) met en évidence les super-voxels, et l'interprétation en moyenne locale.
- En conséquence, le clustering a un effet de débruitage. Il risque d'être légèrement moindre dans les algorithmes proposés comme ceux-ci dépendent des données.
- Pour beaucoup de méthodes à noyau (RAHIMI et RECHT, 2007), seules les distances entre sujets comptent. C'est-à-dire que si Y^α et Y^β sont les observations correspondants aux sujets α et β , il convient seulement de conserver $\|Y^\alpha - Y^\beta\|$. En conséquence de quoi ce sont les différences entre Y qui doivent être L -Lipschitz, ce qui est *a priori* plus simple.
- Quand on considère plusieurs individus Y^α , chacun est projeté sur les mêmes clusters. On peut envisager d'utiliser un estimateur de James-Stein sur tous les ΦY_i^α pour améliorer le débruitage.
- Un bon algorithme va avoir une bonne fidélité (2.10), donc le terme de variance intra-cluster doit être petit.
- Comme cette variance peut être bornée supérieurement par $L^2 \text{Diam}(C_i)^2$, et que $\text{Diam}(C_i) \leq |C_i|$, on s'attend à ce qu'un bon algorithme ait de petits clusters.
- De même, si les clusters sont des boules en dimension d , alors $\text{Diam}(C_i)$ est d'ordre $|C_i|^{1/d}$. Un bon algorithme devrait donc générer des clusters assez compacts.
- Un bon algorithme devrait relier des voxels similaires : cela correspond à faire baisser L à l'intérieur du cluster.

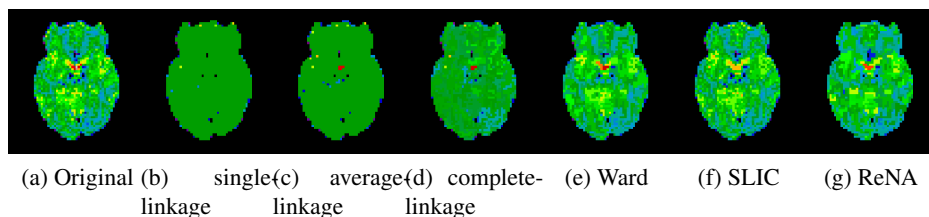


FIGURE 2.1 – **Approximation d'une image IRM par différents algorithmes de clustering**

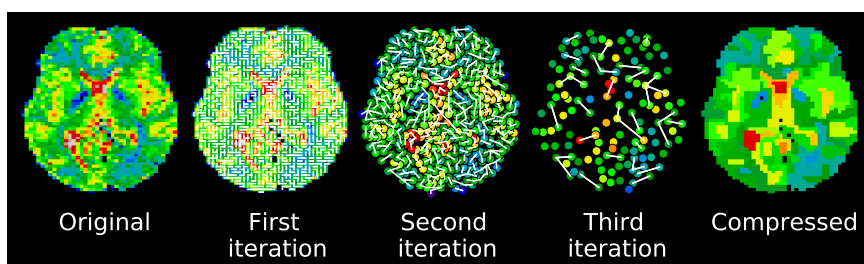


FIGURE 2.2 – **Illustration du principe de ReNA** : Les lignes blanches sont les arêtes du graphe. L'algorithme traite chaque voxel de l'image de départ comme un cluster. Puis, à chaque itération, les clusters les plus proches sont réunis, produisant un graphe réduit, jusqu'à obtention du nombre désiré de clusters.

La figure 2.1 illustre le résultat de différents algorithmes de clustering ; il est facile de voir que l'apparition de composantes géantes est extrêmement nuisible. Il s'agira donc d'éviter la percolation.

Le clustering de WARD (1963) vise directement à minimiser la variance intra-cluster des observations, il n'est donc pas étonnant qu'il donne de bons résultats en pratique. Cependant, même bien optimisé, sa complexité est $O(n^2)$, et il est trop lent pour les applications visées.

Le single-linkage clustering (GOWER et ROSS, 1969) se calcule extrêmement rapidement. Mais il souffre de percolation : une composante géante apparaît, et les résultats sont décevants dans notre contexte.

D'où la proposition d'un clustering récursif par plus proche voisin (ReNA, Recursive Nearest Neighbour Algorithm), illustré à la figure 2.2. À chaque itération, chaque cluster est relié à un autre cluster, parmi ceux qui ont une arête en commun sur le graphe de l'image, quand on voit les voxels comme des valeurs associées à des points de \mathbb{Z}^d . Cette contrainte assure la cohérence spatiale. Le cluster choisi parmi ces voisins spatiaux est celui qui est le plus proche au sens des données, c'est-à-dire qui minimise $|\bar{X}_i - \bar{X}_j|$.

Le calcul est linéaire en le nombre de voxels, itérations comprises.

En pratique, ce graphe ne percole pas, et les clusters sont tous de taille assez

comparable. L'algorithme semble donc répondre aux exigences intuitives.

Expérimentalement, la fidélité est bonne. Comme prévu, le calcul d'analyses ultérieures est grandement accéléré. Le débruitage mentionné rend même ceux testés plus efficaces en sortie que sur des données brutes.

Les autres méthodes étudiées par HOYOS-IDROBO et al. (2016) sont ou bien beaucoup plus lentes, ou bien perdent trop de données, au moins pour certains usages. Un cas frappant est celui des projections aléatoires pour une analyse en composantes indépendantes. Sur cette dernière les projections aléatoires sont catastrophiques, comme ces composantes sont localisées spatialement.

2.3.1 Perspectives

J'ai en projet d'établir théoriquement la non-percolation du système. Le résultat le plus proche que je connaisse est celui de TENG et YAO (2007), sur la percolation du graphe des k plus proches voisins dans un processus ponctuel de Poisson.

Un autre usage classique des données d'IRM fonctionnelle est la recherche de connectivité fonctionnelle : trouver des zones du cerveau qui travaillent simultanément. En ce cas, plutôt que de résumer les séries temporelles des voxels à une seule valeur, l'expérimentateur est intéressé par la matrice de covariance entre des régions d'intérêt, sur lesquelles les voxels sont intégrés. Une condition, comme l'état de repos chez un sujet sain, est alors modélisée par une loi de probabilité sur cette matrice de covariance. Le but est d'obtenir un diagnostic en attribuant la matrice de covariance d'un sujet à un modèle. Il convient donc d'avoir des modèles efficaces.

Traditionnellement, les neurobiologistes utilisent des modèles de la forme $\Sigma = \Sigma^* + d\Sigma$, où $d\Sigma$ est une variable aléatoire (FAIR et al., 2007, par exemple). Or la matrice de corrélation est forcément positive, et ne le sera pas dans ce modèle. Parmi mes projets avec Gaël Varoquaux, j'aimerais voir ce que l'on gagne à modéliser explicitement Σ comme une matrice positive, en définissant l'aléa par un vecteur tangent à Σ^* sur cette variété, afin de rester avec un modèle assez simple.

2.4 Acquisition comprimée sous contraintes physiques

2.4.1 Motivation initiale : acquisition en IRM

La plupart des signaux sont structurés, au sens où ils peuvent être représentés parcimonieusement dans une base (par exemple) connue à l'avance. C'est-à-dire que les quelques plus gros coefficients du signal dans la base portent presque toute l'information du signal, presque toute la norme L^2 , par exemple. Cependant, on ne sait pas à l'avance quels coefficients sont gros.

CANDÈS, ROMBERG et TAO (2006) ont mis en évidence la possibilité de reconstituer un tel signal parcimonieux en effectuant peu de mesures. Ces mesures, des applications linéaires, doivent être effectuées dans des directions qui n'ont

«rien à voir» avec le signal, comme des projections aléatoires. C'est le début de la théorie de l'acquisition comprimée (*compressed sensing*). LUSTIG, DONOHO et PAULY (2007) l'ont rapidement appliqué à l'IRM.

Cependant, la théorie suppose que chaque mesure peut être choisie indépendamment. Dans les applications réelles, les contraintes physiques laissent rarement cette liberté. Souvent, les mesures doivent être effectuées dans une base donnée. De plus, et c'est l'aspect le plus différent, les mesures successives doivent être suffisamment lisses : si les éléments de la base de mesure sont indexés par une image, les mesures choisies suivent une courbe continue ou plus, ou sa discrétisation.

Pour clarifier les idées, fixons les notations et exprimons un théorème typique. On part d'une matrice orthogonale $\mathbf{A}_0 = [\mathbf{a}_1 | \dots | \mathbf{a}_n]^*$, où les \mathbf{a}_i sont les vecteurs d'une base. De ceci, on extrait la matrice de mesure aléatoire $\mathbf{A} = [\mathbf{a}_{J_1} | \dots | \mathbf{a}_{J_m}]^*$, où les J_i sont des variables i.i.d. dans $[1, n]$. Ce qui est observé est $\mathbf{y} = \mathbf{A}\mathbf{x}$ et le but est de retrouver \mathbf{x} . Comme le nombre de coordonnées mesurées m est (très) inférieur à la dimension de l'espace, il y a beaucoup de solutions à cette équation linéaire. L'estimateur utilisé est :

$$\bar{\mathbf{x}} \in \underset{\mathbf{A}\mathbf{x}=\mathbf{y}}{\text{Argmin}} \|\mathbf{x}\|_1. \quad (2.11)$$

Dans les cas où il y a du bruit ou si la parcimonie attendue n'est pas parfaite, la condition d'égalité est relâchée, comme dans un estimateur LASSO, mais cela ne change rien à l'idée de base.

Théorème 2.4.1 (CHAUFFERT, CIUCIU, Jonas KAHN et P. WEISS, 2014 ; BIGOT, BOYER et P. WEISS, 2016, Théorème 3.1). *Supposons que \mathbf{x} est s -parcimonieux, i.e. qu'il ne contient que s composantes non nulles parmi n . Posons $\pi_k = \frac{\|\mathbf{a}_k\|_\infty^2}{\sum_{i=1}^n \|\mathbf{a}_i\|_\infty^2}$. Si le nombre de mesures m satisfait :*

$$m \geq Cs \left(\sum_{i=1}^n \|\mathbf{a}_i\|_\infty^2 \right) \log \left(\frac{n}{\epsilon} \right),$$

où $C > 0$ est une constante universelle, alors $\bar{\mathbf{x}} = \mathbf{x}$ avec probabilité $1 - \epsilon$.

Dans le cas de l'IRM, le signal est censé être parcimonieux dans une base d'ondelettes, dont la matrice de changement de bases est Ψ , et les mesures sont effectuées dans la base de Fourier, de matrice \mathbf{F} . Ainsi $\mathbf{A}_0 = \mathbf{F}^* \Psi$. Dans ce cas, on peut montrer que $\|\mathbf{a}_i\|_\infty^2 \propto \log(n)$. Il suffit donc $O(s \log(n)^2)$ coefficients de Fourier pour pouvoir reconstruire exactement une image s -parcimonieuse.

Notons que la norme L^1 est ici une relaxation convexe de la norme L^0 , et permet un calcul rapide du minimiseur.

Traditionnellement, ces théorèmes sont plutôt exprimés en choisissant uniformément les coordonnées à mesurer, et en exprimant le nombre d'échantillons nécessaires via la cohérence $n \max_{1 \leq k \leq n} \|\mathbf{a}_k\|^2$. Toutefois, la cohérence est grande entre les bases de Fourier et d'ondelettes, de l'ordre de n . La méthode consistant

à tirer plus souvent les coordonnées cohérentes est bien connue, et ce théorème la formalise.

Notre but est donc de tirer selon la densité cible π . Comme nous l'avons dit, physiquement, l'IRM mesure des transformées de Fourier de l'image, à strictement parler sur un sous-volume de l'espace et pas forcément à des fréquences entières. Et les mesures successives sont effectuées le long d'une courbe à la vitesse et à l'accélération limitées. Nous ne pouvons donc pas mesurer des points indépendamment. Un premier pas vers une acquisition plus réaliste est de l'effectuer le long d'une courbe continue.

2.4.2 Voyageur de commerce comme échantillonneur de densité variable

Une manière naturelle de générer une courbe continue quand on sait tirer des points au hasard est de les relier. De plus intuitivement, chaque nouveau point apporte de l'information, on veut donc en relier un maximum en un temps donné, c'est-à-dire pour une longueur de courbe donnée. D'où l'idée d'effectuer un voyageur de commerce entre des points tirés selon une certaine densité.

Notons $X_N = \{x_i\}_{1 \leq i \leq N}$ un N -échantillon i.i.d. de loi $\tilde{\pi}$. Notons $\gamma_N : [0, 1] \rightarrow \Omega$ la paramétrisation à vitesse constante du voyageur de commerce entre ces N points, où Ω est un compact convexe de \mathbb{R}^d . Notons $\Pi_N = (\gamma_N)_* \lambda_{[0,1]}$ la mesure image sur Ω de Lebesgue sur $[0, 1]$. Intuitivement, le poids d'un volume est proportionnel à la longueur de son intersection avec le voyageur de commerce. Dire que nous souhaitons approcher une densité cible π revient à dire que nous souhaitons que Π_N converge vers π .

Comment choisir la distribution $\tilde{\pi}$ des points X_N ? Un premier réflexe serait de les tirer suivant π , mais ce serait faire erreur. En effet, il faut tenir compte de la distance entre les points.

Pour obtenir une intuition, regardons ce qui se passe dans un petit cube. Le nombre de points dans le petit cube N_c est proportionnel à $\tilde{\pi}$, approximativement constant dans le petit cube. La distance typique entre deux points est proportionnelle à $N_c^{-1/d}$. Le voyageur de commerce a tendance à relier des points proches, donc les points du cube entre eux; par ailleurs il est invariant d'échelle. Donc la longueur attendue du voyageur de commerce dans le cube sera proportionnelle à $N_c N_c^{-1/d} \propto \tilde{\pi}^{1-1/d}$.

Cet argument heuristique doit pouvoir s'appliquer à beaucoup de constructions de courbes à partir de points (algorithme gourmand, par exemple). Pour le voyageur de commerce, il est possible d'employer certaines propriétés de sous-additivité d'un processus associé et l'estimation asymptotique de la longueur du voyageur de commerce par BEARDWOOD, HALTON et HAMMERSLEY (1959) pour prouver :

Théorème 2.4.2 (CHAUFFERT, CIUCIU, Jonas KAHN et P. A. WEISS, 2013, Théorème 3.1). *Soit π une mesure d'un compact Ω de \mathbb{R}^d à densité par rapport à Lebesgue. Soit $\tilde{\pi} = \frac{\pi^{(d-1)/d}}{\int_{\Omega} \pi^{(d-1)/d}(x) dx}$. Alors, $\tilde{\pi}^{\otimes n}$ -presque sûrement vis-à-vis de la*

suite de points $\{x_i\}_{i \in \mathbb{N}}$, la distribution du voyageur de commerce converge faiblement vers π :

$$\Pi_N \xrightarrow[N \rightarrow \infty]{} \pi \quad \tilde{\pi}^{\otimes n} p.s.$$

Notons la nécessité qu'il s'agisse d'une densité. Cependant, si la mesure cible a des atomes, on peut simplement rester sur place durant le temps correspondant.

Les simulations numériques montrent de bonnes reconstructions. Cependant, toutes les contraintes de l'IRM ne sont pas encore respectées : nous nous autorisons une accélération infinie. Une première approche est de projeter la courbe du voyageur de commerce sur l'ensemble des courbes respectant les contraintes (CHAUFFERT, P. WEISS et al., 2016). Numériquement, ce problème se résout efficacement, mais nous perdons la garantie de convergence vers la densité cible.

Une autre approche est de projeter directement la densité cible sur l'espace des contraintes, ce que je vais détailler maintenant.

2.4.3 Projection sur les mesures

Nous pouvons élargir le cadre. Les contraintes de l'IRM font que l'ensemble des mesures image de la courbe d'acquisition forment un sous-ensemble \mathcal{M}_N des mesures de Radon sur l'espace. Et nous essayons d'approcher π avec un élément de cet ensemble.

Plus généralement, comment projeter une probabilité π sur un sous-ensemble \mathcal{M}_N des mesures de Radon sur $\Omega \subset \mathbb{R}^d$? CHAUFFERT, CIUCIU, Jonas KAHN et P. WEISS (2016) définissent la projection comme une solution de ce problème variationnel, qui permet à la fois certains résultats théoriques et une implémentation numérique :

$$\mu_N^* \hat{=} \inf_{\mu \in \mathcal{M}_N} \|h \star (\mu - \pi)\|_2^2, \quad (2.12)$$

où h est un noyau dans $L^2(\Omega)$, avec $\Omega \in \mathbb{R}^d$, et \star est la convolution.

Un bon noyau h doit permettre de définir une norme $\mathcal{N}_h(\mu) = \|h \star \mu\|_2^2$ sur l'espace \mathcal{M} des mesures signées sur Ω bornées en variation totale. Ainsi, si $\pi \in \mathcal{M}_N$, alors π est l'unique solution du problème de minimisation.

Si h est continue, auquel cas $h \star \mu \in L^2(\Omega)$, la condition de norme est équivalente au fait que h a tous ses coefficients de Fourier non nuls. Dans ce cas, la norme \mathcal{N}_h métrise la topologie faible sur les boules de \mathcal{M} , et en particulier sur l'ensemble des mesures de probabilité. Du coup, si \mathcal{M}_N est faiblement fermée (elle est bornée en variation totale puisque ce sont des probabilités), alors le problème (2.12) admet au moins une solution. En particulier, si \mathcal{M}_N est un ensemble de mesures image $\{p_*\gamma : p \in \mathcal{P}\}$ avec \mathcal{P} un ensemble de paramétrisations $p : X \rightarrow \Omega$, et que \mathcal{P} est compact pour la topologie de la convergence simple, alors le problème (2.12) admet au moins une solution. Ce sera le cas pour tous les exemples qui vont suivre.

L'aspect pratique de ce choix de projection se réalise en écrivant, si Ω est le tore :

$$\begin{aligned} \langle h \star (\mu - \pi) | h \star (\mu - \pi) \rangle_2^2 &= \langle H \star (\mu - \pi) | \mu - \pi \rangle \\ &= \langle H\mu | \mu \rangle - 2\langle H\mu | \pi \rangle + \langle H\pi | \pi \rangle, \end{aligned}$$

avec $\hat{H}(\xi) = |\hat{h}|(\xi)^2$ pour tout $\xi \in \mathbb{Z}^d$. En particulier, si \mathcal{M}_N ne contient que des mesures de N points de même poids p_i , minimiser (2.12) est équivalent à minimiser l'équation d'attraction-répulsion suivante :

$$\min_{p \in \mathcal{M}_N} \frac{1}{2N^2} \sum_{i=1}^N \sum_{j=1}^N H(p_i - p_j) - \frac{1}{N} \sum_{i=1}^N \int_{\Omega} H(x - p_i) dx. \quad (2.13)$$

Si H a un minimum en zéro, le premier terme représente une répulsion entre les points, et le second une attraction vers π . TEUBER et al. (2011) ont proposé cette équation avec $H(x) = \|x\|_2$ et n'importe quelle position des points pour effectuer du demi-ton.

Trouver un minimum exact d'une telle équation est en général un problème difficile. En prenant Ω égal à la sphère \mathbb{S}^2 et $H(x) = \frac{1}{\|x\|_2}$ et π l'uniforme, on récupère le problème de THOMSON (1904), consistant à trouver la position d'énergie minimale de N électrons sur la sphère, et encore non résolu.

Toutefois, si trouver un minimum global est difficile, l'équation d'attraction-répulsion (2.13) est continue en les p_i si H est gentil, y compris si on discrétise l'intégrale, donnant une fonction $J(p)$. Ceci permet de trouver un point critique par une descente de gradient projetée dans des cas très généraux.

Théorème 2.4.3 (CHAUFFERT, CIUCIU, Jonas KAHN et P. WEISS, 2016, Corollaire 2). *Si H est une fonction définissable sur une structure o -minimale dotée d'un gradient continu L -Lipschitz, et si C est un ensemble fermé définissable sur une structure o -minimale, alors la suite suivante converge vers un point critique de la fonctionnelle d'attraction-répulsion $J(p)$:*

$$p^{(k+1)} \in P_C \left(p^{(k)} - \gamma \nabla J(p^{(k)}) \right)$$

avec $0 < \gamma < \frac{N}{3L}$.

Ce théorème est un cas particulier d'un théorème très général d'ATTOUCH, BOLTE et SVAITER (2013). Sans détailler ce que sont les fonctions et ensembles définissables sur une structure o -minimale (COSTE, 2000), disons simplement que cela inclut tous les polynômes de fonctions élémentaires, leurs compositions, leurs ensembles de niveau...

Idéalement, à quoi devrait ressembler les ensembles de mesure \mathcal{M}_N pour avoir de bonnes propriétés ?

Si nous souhaitons approcher une mesure π par une suite de mesures dans des \mathcal{M}_N successifs (et emboîtés), il faut et suffit que $\bigcup_N \mathcal{M}_N$ soit faiblement dense

dans les probabilités sur Ω . La vitesse de convergence peut être estimée via le contrôle par la distance de transport si h est L -Lipschitz, c'est-à-dire $\mathcal{N}_h(\mu - \pi) \leq LW_1(\mu, \pi)$.

En particulier, si les \mathcal{M}_N sont les ensembles de N pics de Dirac, on trouve une convergence en $LN^{-1/d}$ dans le pire des cas.

Plus généralement, pour appliquer l'algorithme de descente de gradient projetée, il convient d'approcher \mathcal{M}_N par un ensemble \mathcal{A}_n de mesures discrètes. C'est toujours possible, au sens de la distance de Hausdorff. Mais en pratique, \mathcal{A}_n n'est pas forcément simple à exprimer, et il n'est pas forcément simple de projeter dessus. Dans les exemples qui suivront, elle l'est. De plus \mathcal{A}_n joue le rôle de C dans le théorème 2.4.3, et doit donc être définissable sur une structure o -minimale.

L'algorithme de descente de gradient projetée sur \mathcal{A}_n tend alors un point critique du problème initial (2.12), et peut tourner jusqu'à environ $n = 10^5$ points sur des machines actuelles, en quelques jours. Les résultats sont beaucoup plus rapides pour $n = 10^3$ points.

Pour illustrer sa grande souplesse, voici quelques exemples d'ensembles de projection \mathcal{M}_N et d'applications, certaines non encore traitées.

- L'ensemble des sommes de N pics de Dirac $\mathcal{M}_N = \{\frac{1}{N} \sum \delta_{p_i} : p_i \in \Omega\}$.
- Les courbes soumises à des contraintes de vitesse, d'accélération, ou de dérivées supérieures, pas forcément isotropiquement, qui peuvent être représentées par des boules de Sobolev :

$$\mathcal{M}_N = \left\{ \frac{1}{N} p_* \lambda_{[0, N]} : p \in (W^{m, q}([0, N]))^d, p([0, N]) \subset \Omega, \forall 1 \leq j \leq m, \|p^{(j)}\|_q \leq \alpha_j \right\},$$

où λ est la mesure de Lebesgue et $q \in [1, \infty]$ et les α_j sont des réels positifs. Dans ce cas, en posant $\Delta t = \frac{N}{n}$, et l'opérateur de dérivation discrète $(Ds)_i = (s_i - s_{i-1})/\Delta t$, avec les conditions au bord idoines :

$$\mathcal{A}_n = \left\{ s \in \mathbb{R}^{n \cdot d}, \forall 1 \leq i \leq n, s_i \in \Omega \text{ et } \forall 1 \leq j \leq m, \|D^j s\|_q \Delta t \leq \alpha_j \right\}.$$

- Les ensembles de k bâtonnets parcourus à des vitesses constantes, chacun dans le même temps $\frac{N}{k}$:

$$\mathcal{M}_N = \left\{ \frac{1}{N} \sum_{b=1}^k (l_b)_* \lambda_{[0, \frac{N}{k}]} : l_b : [0, \frac{N}{k}] \rightarrow \Omega, l_b(\frac{N}{k}t) = tx_b^1 + (1-t)x_b^2 \right\}.$$

Dans ce cas, en discrétisant chaque segment par $m = \frac{n}{N}$ points, on utilise :

$$\mathcal{A}_n = \left\{ q \in \mathbb{R}^{n \cdot d} : \forall 1 \leq i \leq n, s_i \in \Omega \text{ et } q_j = q_i + \frac{j-i-1}{m-2} (q_{i+m-1} - q_i) \right. \\ \left. \forall i \in \left\{ 1, m+1, 2m+1, \dots, \frac{N}{k}m+1 \right\} \text{ et } i \leq j \leq i+m-1 \right\}.$$

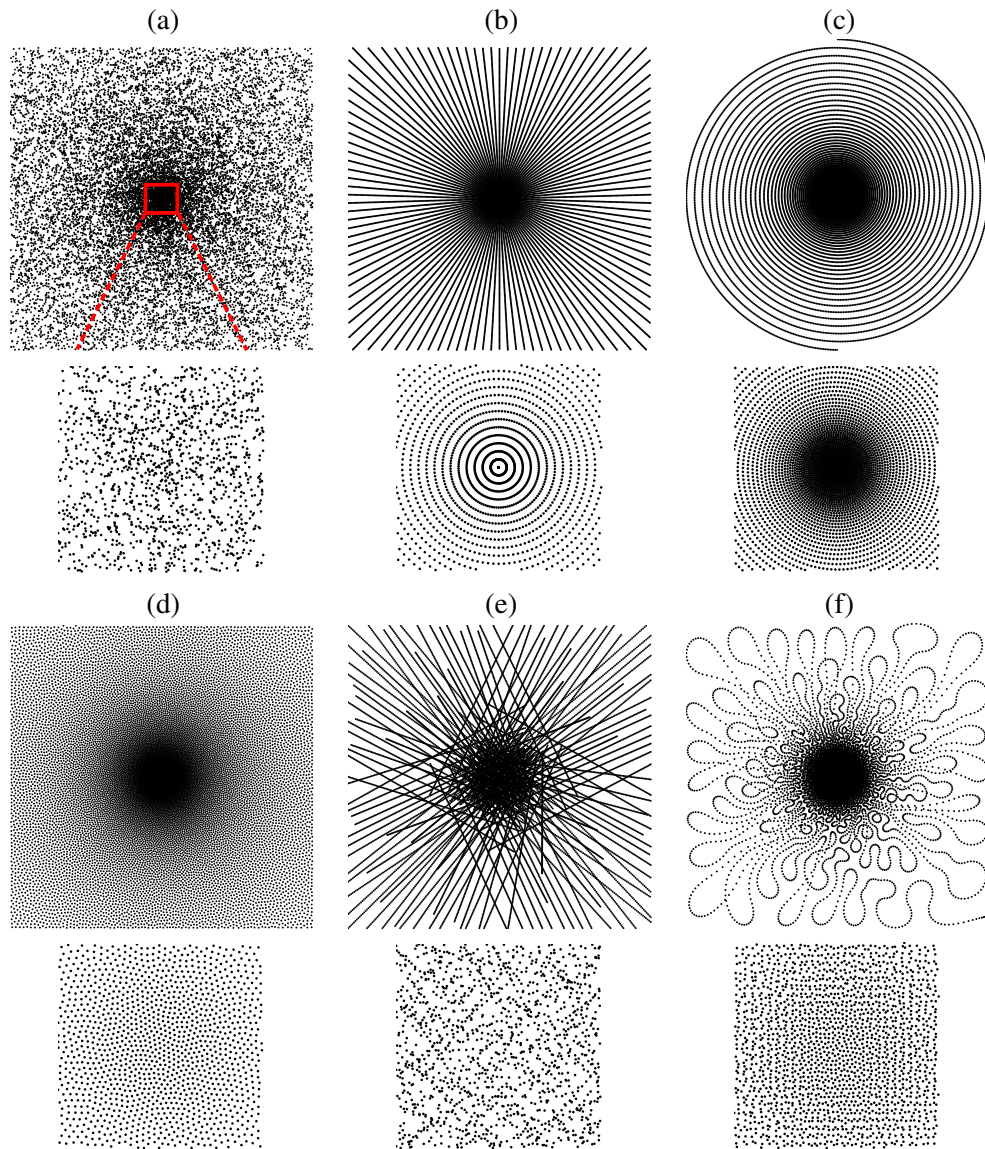


FIGURE 2.3 – Schémas d'échantillonnage classiques (a-c) et schémas d'échantillonnage obtenus par projection de mesure (d-f). Haut : (a) : tirage indépendant ; (b) : lignes radiales ; (c) : spirale. Seconde ligne : zoom au centre. Troisième ligne : (d) : points isolés ; (e) : lignes isolées de longueur fixée ; (f) : courbe admissible en IRM. Bas : zoom au centre. La Figure 2.4 montre les reconstructions correspondantes.

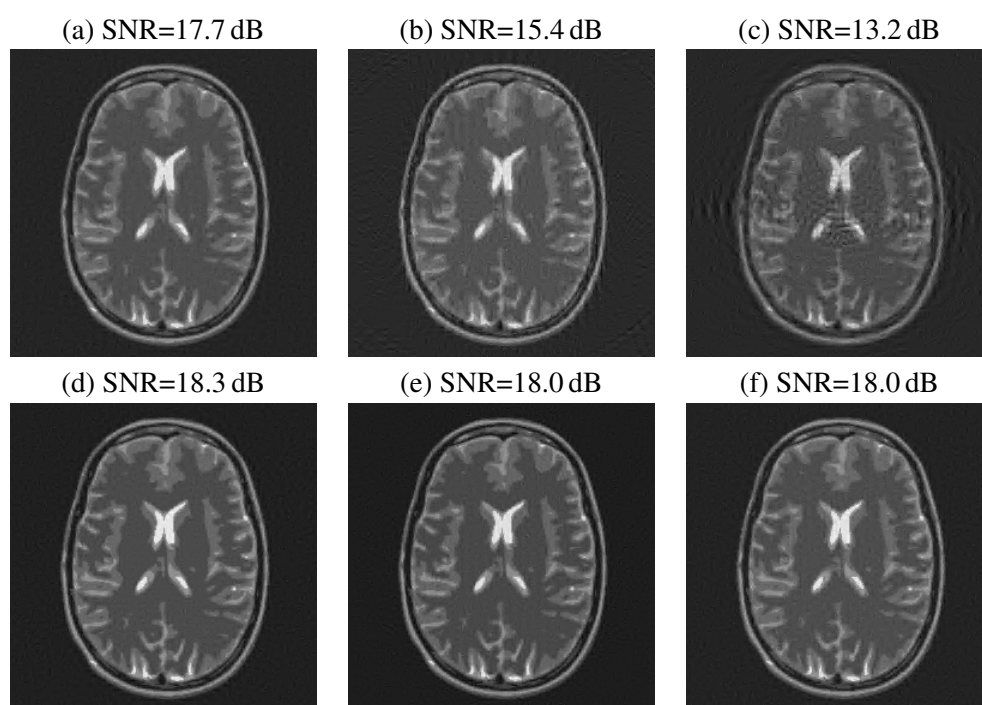


FIGURE 2.4 – Reconstructions pour les schémas d'échantillonnage de la Figure 2.3 sur une image test classique.

- Il est bien sûr possible d’adapter l’équation (2.13) en associant des poids à chacun des p_i .

Pour ce qui est des applications :

- La motivation initiale de l’IRM est couverte à la fois par les courbes dans les boules de Sobolev, et les bâtonnets, faciles à implémenter en pratique. Les figures 2.3 et 2.4 donnent une idée de l’efficacité. Des figures pour de l’imagerie à plus haute résolution, plus favorables encore, sont présentées par BOYER et al. (2016)
- Tout simplement l’aspect artistique. Outre le demi-ton utilisé par les imprimantes, tracer une image en niveaux de gris d’un seul trait ouvre des possibilités, comme la Jeune Fille à la Perle en Figure 2.5.
- Il doit être possible d’adapter l’algorithme à la déconvolution de pics isolés. CANDÈS et FERNANDEZ-GRANDA (2014) propose un algorithme de reconstruction exacte pour une fonction constituée de pics suffisamment distants, après convolution. Toutefois l’algorithme devient lent dès que l’espace sous-jacent est de dimension plus grande que 1. On peut alterner la descente de gradient dans l’équation (2.13) avec une projection sur l’espace des pics séparés pour obtenir des solutions, quelle que soit la dimension d .

2.4.4 Perspectives

Je compte poursuivre ma collaboration avec Pierre Weiss et ceux qui l’entourent. Un nouveau doctorant est d’ailleurs lancé sur ces sujets.

Outre les applications déjà évoquées, et d’autres comme la détection de singularité ; outre d’autre part des problèmes purement numériques pour accélérer les algorithmes, quelques problèmes purement mathématiques se présentent.

À la frontière, et assez simplement, adapter la théorie et les algorithmes à une projection en distance de transport, très naturelle.

De manière nettement plus ambitieuse, trouver des conditions sous lesquelles la distance de transport entre la densité cible π et l’échantillonnage offre des garanties sur la reconstruction en acquisition compressée. De manière évidente, dans le cas général, cela ne tient pas, puisqu’une grille régulière correspondra à du down-sampling.

Par ailleurs, tous les points critiques auxquels on arrive en pratique sont similaires, au sens de leur distance à la cible, et ce quel que soit le point de départ de l’algorithme. Comprendre ce phénomène semble à la fois peu évident et très significatif : quand le nombre de points d’échantillonnage augmente, est-ce que tous les points critiques ont une valeur proche de celle du minimum ? Ou tout du moins les bassins d’attraction de tels points occupent-ils tout l’espace des phases ?



FIGURE 2.5 – *La Jeune Fille à la Perle* de VERMEER (1665). Projection sur une boule de Sobolev $W^{1,\infty}$, avec $N = 150000$ points. Résultat au bout de 10000 itérations

Chapitre 3

Géométrie stochastique et espaces métriques aléatoires

La géométrie stochastique s'intéresse à des objets aléatoires générés dans des espaces doués d'une géométrie, et souvent dans leurs aspects géométriques. Cela pourra être l'étude de la forme d'une cellule de Voronoï pour la mosaïque générée par un processus de Poisson dans \mathbb{R}^d , ou se demander si le processus de Poisson marqué par un cercle de rayon 1 percole, par exemple. Savoir comment va se passer la transmission entre des capteurs d'un réseau ou caractériser une roche par ses pores sont deux exemples de problèmes pratiques qui se formalisent dans le cadre de la géométrie stochastique. STOYAN, KENDALL et MECKE (1996) offrent une bonne introduction au sujet.

Certains des objets de la géométrie stochastique créent leur propre géométrie : imaginons un graphe entre des points qui sont reliés s'ils sont assez proches. Le graphe fournit son propre espace métrique sur ses sommets. D'une manière générale, la géométrie stochastique permet de générer facilement des espaces métriques aléatoires. Ce point de vue permet d'englober dans le champ de la géométrie stochastique les espaces métriques aléatoires qui sont générés de manière topologique et combinatoire, comme la carte brownienne, que nous évoquerons seulement pour quelques brèves comparaisons.

La section 3.1 présente un dénombrement des mosaïques en T sur un nombre de droites fixé, ce qui implique l'existence de certaines mesures de Gibbs sur ces mosaïques. Ceci permet une certaine souplesse dans la modélisation de paysages agricoles.

Dans la section 3.2, nous présentons la notion de SIRS (scale-invariant random spatial network) introduite par ALDOUS (2014) pour modéliser les réseaux routiers. Le processus de droites de Poisson impropre est un SIRS. Par ailleurs, il génère un espace métrique aléatoire, d'un intérêt indépendant.

3.1 Mosaïques en T

Les mosaïques en T sont les mosaïques d'une partie (compacte convexe pour simplifier) du plan dont tous les sommets sont de degré trois et avec un angle plat, en forme de T. Elles ressemblent notamment au découpage des terres agricoles, et KIËU et al. (2013) ont développé un modèle dans ce sens.

Les premiers modèles de mosaïques en T sont des cas particuliers de champs de Markov polygonaux d'ARAK, CLIFFORD et SURGAILIS (1993), un modèle très général de graphe planaire (géométrique) aléatoire, dont les belles propriétés permettent de l'échantillonner exactement. Toutefois KIËU et al. (2013) souhaitaient davantage de souplesse, et s'autoriser des mesures de Gibbs sur un modèle sous-jacent adapté.

Ce modèle est la mosaïque en T complètement aléatoire (completely random T-tessellation) qui est intuitivement définie de la manière suivante : ajouter un segment donné ne dépend pas du reste de la mosaïque, tant que nous gardons une mosaïque en T. Un tout petit plus précisément, si T est une mosaïque en T, et si quand on ajoute le segment s à T , obtenant $T + \{s\}$, on a encore une mosaïque en T, alors le ratio de leurs densités de probabilité $\frac{p(T+\{s\})}{p(T)}$ ne dépend pas de T . La définition s'écrit bien en termes de noyau de Papangelou, et rappelle celle d'un processus de Poisson.

Du coup, la densité de probabilité d'une mosaïque est donnée par la densité de probabilité des droites qui portent les segments de la mosaïque, si celles-ci ont été choisies suivant un processus de Poisson. Nous détaillerons en section 3.2.1 ce qu'est un processus de droites de Poisson pour ceux qui veulent un énoncé plus précis que «jeter des droites complètement au hasard». Rappelons simplement que le nombre de droites qui intersectent un volume fini, comme ici, suit une loi de Poisson de paramètre λ , son intensité.

Mais il n'est pas évident que cette mesure μ sur les mosaïques soit finie, et donc puisse être une probabilité. En effet, à une configuration de droites de Poisson donnée est associé un poids égal au nombre de mosaïques en T que l'on peut former en mettant exactement un segment (une union maximale d'arêtes alignées et contiguës) sur chacune de ces droites. Il faut que ce poids ne soit pas trop grand en moyenne. Nous allons contrôler le pire des cas, ce qui s'avère suffisant.

Plus précisément, si pour k droites quelconques, on peut toujours construire au plus $\mathcal{N}(k) = o(a^k k^k)$ mosaïques en T sur ces droites, pour tout $a > 0$, alors la variation totale de la mesure non normalisée que nous avons décrite est

$$\begin{aligned} |\mu| &\leq \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \mathcal{N}(k) \\ &= \sum_{k=0}^{\infty} o((a\lambda e)^k) < \infty. \end{aligned}$$

Il est possible d'obtenir la borne suivante par des moyens entièrement combinatoires :

Théorème 3.1.1 (Jonas KAHN, 2014, Théorème 2.1). *Le nombre de mosaïques en T différentes qui peuvent être construites sur k droites est borné ainsi :*

$$\mathcal{N}(k) \leq C^k \left(\frac{k}{(\ln k)^{1-\varepsilon}} \right)^{k-k/(\ln k)},$$

où C ne dépend que de ε .

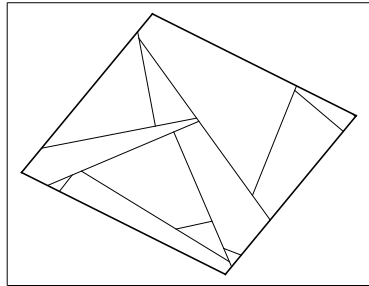
Idée de preuve. Le principe est, pour chaque mosaïque en T , d'associer à chaque droite un certain nombre d'informations qui permettent ensemble de reconstruire la mosaïque. Le nombre de mosaïques est donc borné par le nombre de jeux d'information différents.

Donnons une version simplifiée. Choisissons une direction qui n'est parallèle à aucune droite ni aux bords du domaine, que nous appellerons l'axe du temps. On dit qu'un segment de la mosaïque naît à son point de temps minimal et meurt ou est tué à son point de temps maximal. Il suffit maintenant d'associer à chaque droite la date de naissance de son segment, et le nombre de meurtres effectué par son segment, c'est-à-dire d'autres segments qui sont tués quand ils l'intersectent. La reconstruction est expliquée et illustrée à la figure 3.1. Maintenant, un segment ne peut naître que sur le bord ou un autre segment, donc en intersectant une autre droite : cela fait k possibilités par segment, soit k^k au total. De plus, chaque segment ne meurt qu'une fois, et ne peut être tué que par une droite ou le bord. Cela revient à répartir k objets dans $(k+1)$ boîtes, soit $\binom{2k+1}{k} \leq 4^k$ possibilités. D'où une borne $\mathcal{N}(k) \leq (4k)^k$ sur le nombre de mosaïques en T sur k droites.

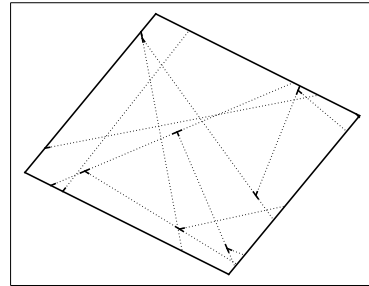
Cette borne est insuffisante. Pour obtenir la borne du théorème, il faut oublier une partie bien choisie des dates de naissance, ajouter d'autres informations similaires au nombre de meurtres, et effectuer une reconstruction plus compliquée, avec des aller-retours le long de l'axe du temps. \square

Cette borne combinatoire est essentiellement optimale : en prenant une grille avec $k/(\ln k)$ droites verticales et $k - k/(\ln k)$ droites horizontales, on se convainc facilement que l'on peut construire $\left(\frac{k}{\ln k}\right)^{k-k/(\ln k)}$ mosaïques en T différentes. Il suffit de prendre les segments verticaux de longueur maximale, $k/(\ln k)$ positions possibles pour chaque segment horizontal.

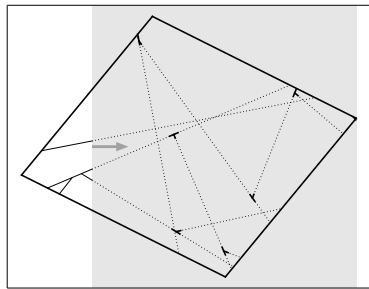
Cependant, une analyse qui tiendrait compte de la géométrie et de la position typique de droites aléatoires devrait permettre de montrer que le nombre moyen de mosaïques sur k droites est beaucoup plus faibles. En effet, nous pouvons envisager l'heuristique suivante : Supposons que $(k-1)$ segments soient déjà fixés. Combien de manières a-t-on de placer le dernier segment ? Une pour chaque intervalle entre deux segments que sa droite croise, soit une plus le nombre n de segments que sa droite croise. Du coup, la longueur de ce segment sera à peu près $1/n$, en normalisant à un pour la longueur maximale. Or, la probabilité que sa droite croise un segment d'une droite donnée est proportionnel à la longueur de celui-ci. Soit typiquement $1/n$ pour chaque droite. Donc le nombre de segments croisés par la



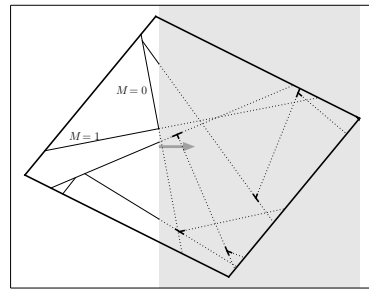
(a)



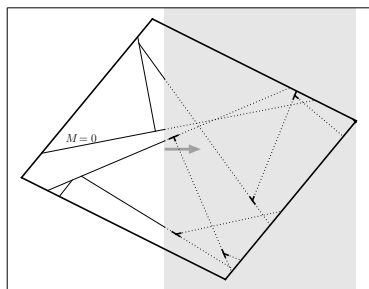
(b)



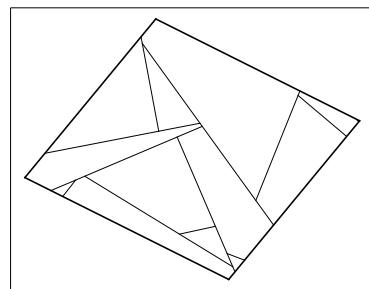
(c)



(d)



(e)



(f)

FIGURE 3.1 – (a) est la mosaïque à reconstruire. Nous connaissons initialement les droites, les dates de naissance (b) et les nombres de meurtre. Nous nous déplaçons le long de l'axe du temps en prolongeant les segments vivants (c). Quand deux segments se croisent (d), nous tuons celui à qui il reste 0 meurtres, et baissons de 1 le nombre de meurtres de l'autre (e). À la fin du parcours, on retrouve la mosaïque (f).

droite est à peu près k/n , donc $n = \sqrt{k}$. Il y a \sqrt{k} possibilités pour placer le dernier segment, donc on s'attend à \sqrt{k}^k mosaïques différentes.

Toutefois, raisonner rigoureusement selon ces lignes semble considérablement plus difficile...

3.2 Processus de droites de Poisson impropre

3.2.1 SIRS et définitions

ALDOUS (2014) a introduit la notion de SIRS (scale-invariant random spatial network) comme une abstraction des réseaux routiers et des cartes telles qu'on les utilise sur Internet. D'où la définition :

1. Un utilisateur veut aller d'un point A à un point B : un SIRS est donc la donnée d'un itinéraire $\mathcal{R}(x_1, x_2)$ entre chaque paire de points x_1, x_2 de \mathbb{R}^d , tel que $\mathcal{R}(x_1, x_2) = \mathcal{R}(x_2, x_1)$. L'itinéraire est aléatoire et presque sûrement unique. Un itinéraire est un chemin de longueur finie.
2. Quand on se déplace sur la carte, ou qu'on la tourne, l'aspect général ne change pas. De même lorsque l'on zoome, et que les routes plus petites apparaissent. Un SIRS doit donc être statistiquement équivariant par translation, rotation et changement d'échelle : en notant $\mathcal{N}(x_1, \dots, x_k)$ pour le réseau formé des itinéraires entre les paires de x_i pour tout ensemble fini fixé de point $x_1, \dots, x_k \in \mathbb{R}^d$, et \mathfrak{R} pour une similarité euclidienne, les réseaux $\mathcal{N}(\mathfrak{R}x_1, \dots, \mathfrak{R}x_k)$ et $\mathfrak{R}\mathcal{N}(x_1, \dots, x_k)$ ont la même distribution.
3. Un bon réseau routier permet de lier deux points sans trop de détours : si D_1 est la longueur de l'itinéraire entre deux points à distance 1, alors $\mathbb{E}[D_1] < \infty$.
4. Un bon réseau routier ne demande pas de construire trop de routes à une échelle donnée. Et on passe par les grandes routes, à part près du départ et de l'arrivée : si les $\{\Xi_n\}_{n \in \mathbb{N}}$ sont des processus ponctuels de Poisson d'intensité n (voir définition ci-dessous) imbriqués, c'est-à-dire tels que $\Xi_n \subset \Xi_{n+1}$, et si $\Xi = \bigcup_n \Xi_n$, alors le réseau à longue distance \mathcal{F} suivant a une intensité (longueur moyenne par unité de surface) $p(1)$ finie :

$$\mathcal{F} = \bigcup_{x_1, x_2 \in \Xi} (\mathcal{R}(x_1, x_2) \setminus (B(x_1, 1) \cup B(x_2, 1))). \quad (3.1)$$

Pour clarifier la dernière condition, l'utilisation des processus de Poisson est pour avoir à gérer un nombre dénombrable d'itinéraires. Moralement, nous demandons que tous les itinéraires entre toutes les paires de point de l'espace, auxquels on a retiré la «sortie de parking» au départ et à l'arrivée, utilisent tous uniquement les mêmes (auto)routes.

Ces quelques propriétés suffisent à fortement contraindre le modèle. Par exemple, ALDOUS (2014) a montré qu'il existait des chemins simplement infinis dont tout

sous-ensemble faisait partie d'un itinéraire, mais pas de chemin doublement infini (infini au deux bouts). En revanche, il n'est pas facile de construire des exemples explicites de SIRS. ALDOUS (2014) a créé un modèle binaire hiérarchique dans \mathbb{R}^2 , mais il est un peu artificiel dans la mesure où l'invariance est ajoutée en symétrisant à la fin, si bien que les réalisations ont une forte dépendance à longue portée. C'était le seul connu.

À la lecture de la définition, on peut envisager une manière générale de créer des candidats à être un SIRS : on part d'une bonne métrique aléatoire sur \mathbb{R}^d , et les itinéraires sont les géodésiques pour cette métrique. Pour peu que la construction soit invariante, on obtient automatiquement la propriété 2. La propriété 4 présente un caractère hyperbolique (les géodésiques se rapprochent). Or, les espaces métriques aléatoires sont en général hyperboliques (GROMOV, 2003, par exemple).

ALDOUS (2014) a proposé deux métriques aléatoires pareilles. Nous allons nous intéresser à l'une d'elles : le processus de droites de Poisson impropre.

Intuitivement, il s'agit de jeter des droites uniformément au hasard dans \mathbb{R}^d , et de les doter d'une limite de vitesse aléatoire bien choisie. Il y a de plus en plus de droites «lentes» (limite de vitesse faible), si bien que ces droites sont denses dans \mathbb{R}^d . La métrique est le temps nécessaire pour joindre deux points en respectant les limites de vitesse. Curieusement, il est possible de créer des chemins continus d'un point à un autre même en dimension 3 ou plus presque partout portés par les droites, bien que les droites ne se croisent pas.

Plus rigoureusement, commençons par rappeler ce qu'est un processus ponctuel de Poisson sur un espace mesuré (\mathbb{X}, μ) : il s'agit d'un ensemble aléatoire de points sur \mathbb{X} tels que, en notant $N(B)$ est le nombre de points dans B :

- Si les B_i sont mesurables et disjoints, alors les $N(B_i)$ sont indépendants.
- $N(B_i)$ est une loi de Poisson de paramètre $\mu(B_i)$.

Il est par ailleurs possible de paramétriser l'ensemble des droites de \mathbb{R}^d par \mathcal{H} , de telle sorte que la mesure de Lebesgue dans l'espace des paramètres d'un ensemble mesurable quelconque de droites soit invariante par les isométries de \mathbb{R}^d . En deux dimensions par exemple, cet espace de paramètres est $\mathcal{H} = [0, \pi[\times \mathbb{R}$, où le premier paramètre est l'angle de la droites avec les abscisses, et le second paramètre la distance algébrique de la droite à l'origine d'un repère donné.

Nous pouvons donc caractériser une droite l munie d'une vitesse v comme un élément (l, v) de $\mathcal{H} \times \mathbb{R}^+$. Le processus de droites de Poisson impropre Π est l'image du processus ponctuel de Poisson sur $\mathcal{H} \times \mathbb{R}^+$ pour la mesure de densité $(\gamma - 1)v^{-\gamma}$ par rapport à Lebesgue. Nous exigeons $\gamma > d$.

Sous ces conditions, il existe presque sûrement entre chaque paire de points a et b de l'espace \mathbb{R}^d des chemins localement Lipschitz $\xi : [0, T] \rightarrow \mathbb{R}^d$ tels que :

- $\xi(0) = a$ et $\xi(T) = b$.
- pour presque tout $t \in [0, T]$, ou bien $\xi'(t) = 0$, ou bien ξ suit une droite de Π , c'est-à-dire qu'il existe $v \geq |\xi'(t)|$ tel que $(\xi(t) + \xi'(t)\mathbb{R}, v) \in \Pi$.

Nous pouvons donc considérer la métrique sur \mathbb{R}^d donnée par l'infimum des T pour de tels chemins reliant a et b . Nous l'appelons la Π -distance, ou le temps T_{ab} qu'il faut pour aller de a et b , en s'inspirant de l'analogie routière.

De la même manière, nous parlerons de Π -chemin, Π -boule, Π -diamètre, etc.

Nous avons donc défini un espace métrique aléatoire. La loi de puissance sur les vitesses nous assure qu'il est invariant d'échelle, à constante multiplicative près. Ainsi ses géodésiques restent les mêmes en loi quand on zoome.

Notons encore \mathcal{L}_ξ le support d'un Π -chemin ξ , c'est-à-dire l'ensemble des droites de Π sur lesquelles ξ parcourt une distance euclidienne non nulle : $\mathcal{L}(\xi) = \{l \in \Pi : m_1(l \cap \xi(0, T)) > 0\}$ où m_1 est la mesure de Hausdorff à une dimension, et $\xi : [0, T] \rightarrow \mathbb{R}^d$.

Dans ce qui va suivre, nous noterons $B(x, r)$ les boules de centre x et de rayon r .

3.2.2 Propriétés

KENDALL (2014) a obtenu des résultats importants sur ce processus : on obtient bien un espace métrique géodésique. De plus, en dimension 2, les géodésiques sont presque partout uniques, localement de longueur finie, et le sous-réseau reliant les points d'un processus de Poisson d'intensité 1 a une longueur d'espérance finie sur les compacts (une forme faible de la propriété 4 du SIRS).

Pour mieux comprendre ce processus, commençons par comparer le diamètre des boules pour la métrique euclidienne et la métrique aléatoire :

Théorème 3.2.1 (Jonas KAHN, 2015, forme simplifiée et preuve des Théorèmes 3.1 et 5.1). *Il existe T_1 tel que pour tout $\frac{1}{2} > \varepsilon > 0$, pour tout $x \in \mathbb{R}^d$, pour tout rayon r , avec probabilité au moins $1 - \varepsilon$:*

$$\begin{aligned} T_{x,r} &\hat{=} \sup_{y,z \in B(x,r)} T_{y,z} \\ &\leq T_r \left(\ln \frac{1}{\varepsilon} \right)^{\frac{1}{\gamma-1}} \quad \text{avec} \\ T_r &= r^{\frac{\gamma-d}{\gamma-1}} T_1. \end{aligned}$$

En particulier, le Π -diamètre de la boule euclidienne de rayon r a plus qu'un moment exponentiel. Pour tout $\delta < T_r^{\frac{1}{\gamma-1}}$:

$$\mathbb{E} \left[\exp \left(\delta T_{x,r}^{\gamma-1} \right) \right] < \infty.$$

L'exposant $\gamma - 1$ dans ce moment ne peut pas être amélioré.

D'autre part, le diamètre euclidien d'une Π -boule de rayon r est d'espérance finie.

La propriété 3 d'un SIRS_N découle facilement de la dernière remarque.

De plus, les deux contrôles permettent de montrer que l'espace métrique aléatoire généré par Π est homéomorphe à \mathbb{R}^d .

On remarque que les boules euclidiennes sont très bien contrôlées en Π -distance, tandis que les Π -boules le sont beaucoup moins. En fait, puisque la probabilité pour qu'une droite de vitesse v passe près du point x est en $v^{-(\gamma-1)}$, le moment polynomial maximum de ce diamètre euclidien ne peut pas être meilleur que $\gamma - 1$. Cela peut probablement être vu comme lié à un aspect fractal du processus.

L'unicité presque sûre des géodésiques en dimension 2 a été obtenue par KENDALL (2014). Notons que le «presque sûr» n'est pas pour tous les points simultanément : le cut-locus n'est pas vide.

En dimension au moins 3 nous obtenons cette unicité via la notion technique de *plein de directions*, qui formalise l'idée que des droites passant près d'un point ne sont pas concentrées dans peu de directions :

Définition 3.2.2 (Jonas KAHN, 2015, Définitions 4.3 et 4.4). *Pour un ensemble de droites $\mathcal{L} = \{l_j\}_{j \in \mathcal{J}}$ et un sous-ensemble X de \mathbb{R}^d , un \mathcal{L} -tour dans X est une courbe continue f dans X telle que pour tout $j \in \mathcal{J}$, il existe un t_j telle que $f(t_j) \in l_j$. Le tour est fini si f est rectifiable, infini sinon.*

Un Π -chemin ξ a plein de directions près d'un point x si, pour tout $\varepsilon > 0$, tous les \mathcal{L}_ε -tours dans $\mathbb{R}^d \setminus B(x, \varepsilon)$ sont infinis.

L'intérêt de cette notion est que, si ξ a plein de directions près de x , tout chemin fermé qui est un \mathcal{L}_ε -tour fini contient x . De plus :

Lemme 3.2.3 (Jonas KAHN, 2015, Lemme 4.6). *Soit $d \geq 3$. Soit l une droite fixée indépendante de Π . Presque sûrement, pour tout $x \in l$, pour tout $y \notin l$, les géodésiques g_{xy} ont toutes plein de direction près de x .*

Ce lemme est assez technique comme il s'applique simultanément à un nombre non dénombrable de points alors qu'il n'est pas vrai pour tout l'espace.

Comme les droites de Π viennent d'un processus de Poisson, chacune de ses droites l est indépendante de $\Pi \setminus \{l\}$, de même loi que Π . En remarquant que presque sûrement les géodésiques entre x et y doivent toutes employer les mêmes droites, le Lemme 3.2.3 permet de déduire que les extrémités de chaque segment de géodésique sont contenues dans chacune de ces géodésiques, et donc que ce sont les mêmes. Ceci donne la propriété 1 d'un SIRS_N.

La dernière propriété d'un SIRS_N, la propriété 4, découle d'un joli principe des tiroirs.

L'équation (3.1) nous invite à contrôler la longueur totale de l'intersection de toutes les géodésiques entre paires de points de Ξ , privées de boules de rayon 1 autour de leurs extrémités, et la boule $B(0, \frac{1}{3})$, par exemple. Cette longueur est $\ell = m_1(\mathcal{F} \cap B(0, \frac{1}{3}))$, où m_1 est la mesure de Hausdorff. Notons \bar{g}_{xy} pour cette trace associée à la géodésique entre x et y .

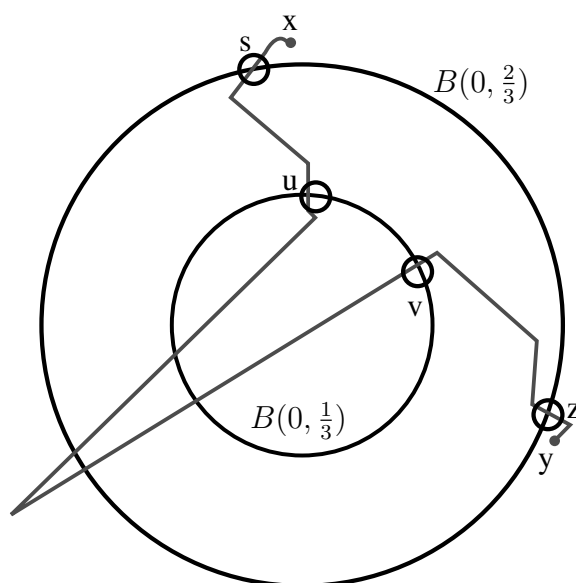


FIGURE 3.2 – La géodésique de x à y atteint $B(0, \frac{2}{3})$ pour la première fois en s et la quitte pour la dernière fois z . Elle atteint $B(0, \frac{1}{3})$ pour la première fois en u et la quitte pour la dernière fois en v .

Si x ou y est dans la boule $B(0, \frac{2}{3})$, l'intersection \bar{g}_{xy} est vide. Donc les géodésiques g_{xy} qui contribuent doivent être de la forme suivante, illustrée à la figure 3.2 :

- x et y sont en dehors de $B(0, \frac{2}{3})$.
- Elles frappent une première fois $B(0, \frac{2}{3})$ en un point s de sa sphère.
- Elles frappent une première fois $B(0, \frac{1}{3})$ en un point u de sa sphère.
- Elles frappent une dernière fois $B(0, \frac{1}{3})$ en un point v de sa sphère.
- Elles frappent une dernière fois $B(0, \frac{2}{3})$ en un point z de sa sphère.

La géodésique g_{xy} doit donc traverser la couronne entre $B(0, \frac{2}{3})$ et $B(0, \frac{1}{3})$ à deux reprises, entre s et u , et entre v et z . Chacune de ces traversées a une longueur euclidienne d'au moins $\frac{1}{3}$.

D'autre part, le Théorème 3.2.1 nous donne un bon contrôle sur la Π -distance T entre s et z . Nous avons aussi un bon contrôle sur le nombre de droites plus rapides

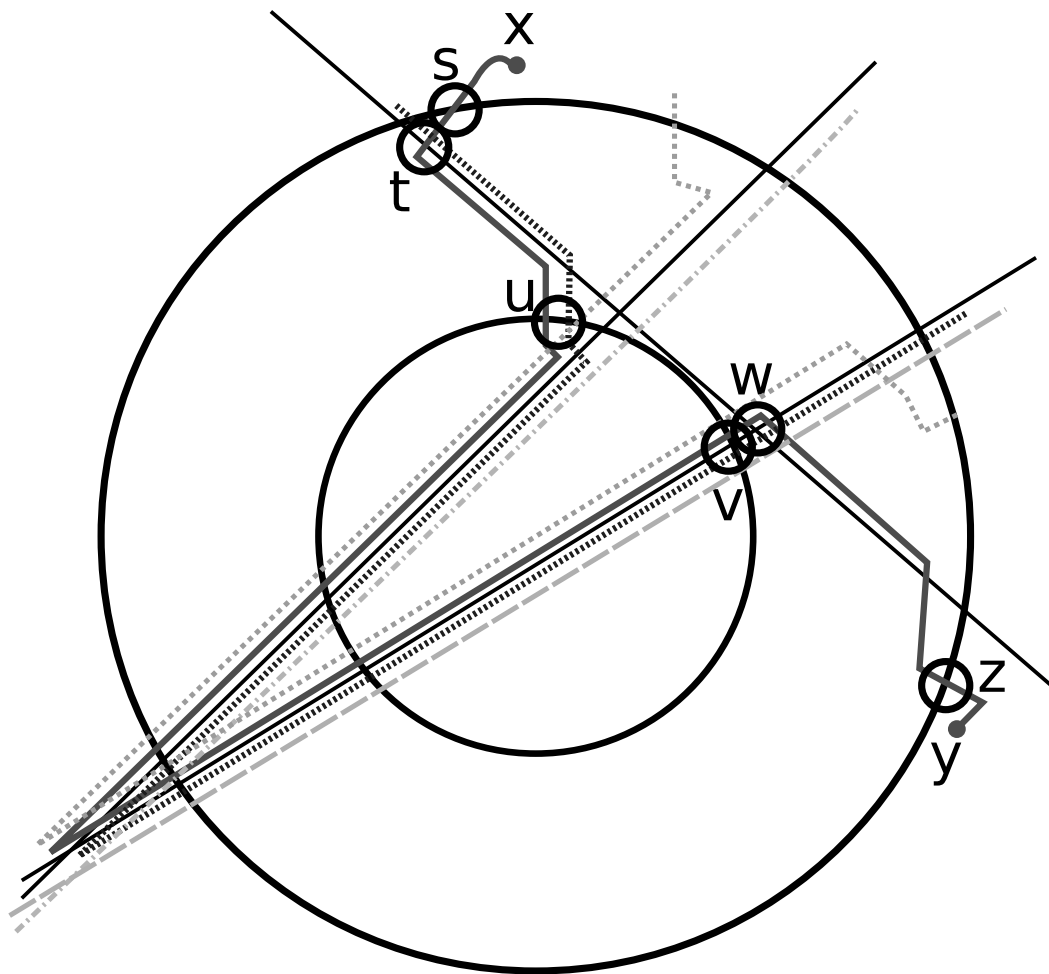


FIGURE 3.3 – Les lignes parallèles sont confondues. Elles sont séparées sur la figure pour la rendre lisible. Les trois lignes noires fines sont les lignes rapides. Les lignes en tirets ou pointillés sont une famille de géodésiques $\{g^i\}$. La ligne pleine qui connecte x et y est la géodésique g_{xy} . La géodésique g_{xy} a un point commun t avec la géodésique en pointillés noirs sur une des droites rapides lors de la première traversée, entre s et u , et un autre en w lors de la dernière traversée, entre v et z . Elles sont donc confondues entre u et v . Toute autre géodésique qui contribue à ℓ intersecte une géodésique de $\{g^i\}$ de la même manière.

que $\frac{1}{6T}$ qui frappent la boule $B(0, \frac{2}{3})$. Notons \mathcal{V} l'intersection de ces droites avec $B(0, \frac{2}{3})$, qui est donc un ensemble de dimension 1. Alors l'intersection de g_{xy} avec des droites plus lentes que $\frac{1}{6T}$ est de longueur $\frac{1}{6}$ au maximum. Donc g_{xy} passe une longueur au moins $\frac{1}{6}$ sur des droites rapides entre s et u , et à nouveau entre v et z . Autrement dit, l'ensemble des paires de points $(t, w) \in \mathcal{V}^2$ avec t sur g_{xy} entre s et u et w sur g_{xy} entre v et z , a mesure au moins $\frac{1}{36}$.

Comme nous contrôlons la mesure de \mathcal{V}^2 , nous pouvons trouver un ensemble fini de géodésiques $\{g^i\}$ telle que toute géodésique g_{xy} intersecte une g^i entre s et u , et entre w et z . Par unicité des géodésiques, elle est donc confondue avec g^i entre u et v . Donc $\bar{g}_{xy} = \bar{g}^i$ et $\mathcal{F} \cap B(0, \frac{1}{3})$ est composé d'un nombre fini de traces. Les calculs précis donnent un moment de forme exponentielle :

Théorème 3.2.4 (Jonas KAHN, 2015, Théorème 6.1). *Soit $\gamma > d \geq 2$. Soit $\ell = m_1(\mathcal{F} \cap B(0, \frac{1}{3}))$ la longueur du réseau longue distance dans $B(0, \frac{1}{3})$. Pour tout $\varepsilon < \varepsilon_{max}$, avec probabilité $1 - \varepsilon$, cette longueur est inférieure à $C(\ln(C_1/\varepsilon))^2$, où les constantes ε_{max} , C et C_1 ne dépendent que de γ et d . D'où le moment, pour tout $\delta < \sqrt{C}$,*

$$\mathbb{E} \left[\exp \left(\delta \sqrt{\ell} \right) \right] < \infty. \quad (3.2)$$

En particulier, ℓ est de moyenne finie, et le processus de droites de Poisson impropre est un SIRS.

3.2.3 Liens

Comme le processus de droites de Poisson impropre est un espace métrique aléatoire, il peut être instructif de le comparer à la carte brownienne (LE GALL, 2014, par exemple)).

La carte brownienne est une métrique aléatoire sur la sphère \mathbb{S}^2 . Elle conserve la topologie de la sphère. Mais elle est métriquement de dimension 4. Elle a un comportement très hyperbolique : l'ensemble des points de ses géodésiques privées de leurs points extrémaux est de dimension 1. Par ailleurs, le cut-locus partant d'un point est un arbre de dimension 2.

Le processus de Poisson impropre est une métrique aléatoire sur \mathbb{R}^d . Il conserve la topologie de \mathbb{R}^d . Métriquement, il est de dimension $(d\gamma - d)/(\gamma - d)$, qui est plus grand que d . Notons que pour $d = 2$ et $\gamma = 3$, nous retrouvons les dimensions de la carte brownienne. Par ailleurs, si on peut prouver que toute géodésique peut être approchée par des géodésiques entre points du processus Ξ , l'ensemble des points de ses géodésiques privées de leurs points extrémaux est aussi de dimension 1. Je n'ai pas d'idées sur le cut-locus.

3.2.4 Perspectives

J'aimerais vérifier si le plan brownien, le cône tangent de Gromov-Hausdorff à la carte brownienne, soit un SIRS. Plus spécifiquement, comme la définition de

SIRSN implique une attache précise au plan \mathbb{R}^2 , une limite correctement choisie de triangulations aléatoires placées par l'empilement de cercle. Les invariances sont évidentes.

Par ailleurs, on peut s'interroger sur ce qui se passe si on fait tomber des éléments autres que des droites, avec des tailles mises à l'échelle. Obtient-on encore un SIRSN ?

D'une manière générale, y a-t-il un objet universel ?

Dans ce cas, que se passe-t-il si les éléments tombent sur un autre espace que l'espace euclidien, comme la sphère ? Comment se comporte l'espace métrique aléatoire si on renonce à l'invariance d'échelle, et qu'on change la fréquence des droites rapides ?

Bibliographie

- ALDOUS, D. J. (2014). “Scale-invariant random spatial networks”. In : *Electronic Journal of Probability* 19.15, p. 1–41. URL : [arxiv:1204.0817](https://arxiv.org/abs/1204.0817).
- ANDERS, Janet et al. (2010). “Ancilla-driven universal quantum computation”. In : *Physical Review A* 82.2, p. 020301.
- ARAK, T., P. CLIFFORD et D. SURGAILIS (1993). “Point-based polygonal models for random graphs”. In : *Advances in Applied Probability* 25, p. 348–372.
- ATTOUCH, Hedy, Jérôme BOLTE et Benar Fux SVAITER (2013). “Convergence of descent methods for semi-algebraic and tame problems : proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods”. In : *Mathematical Programming* 137.1-2, p. 91–129.
- BEARDWOOD, Jillian, John H HALTON et John Michael HAMMERSLEY (1959). “The shortest path through many points”. In : *Proc. Cambridge Philos. Soc.* T. 55. 4. Cambridge Univ Press, p. 299–327.
- BIGOT, Jérémie, Claire BOYER et Pierre WEISS (2016). “An analysis of block sampling strategies in compressed sensing”. In : *IEEE Transactions on Information Theory* 62.4, p. 2125–2139.
- BOYER, Claire et al. (2016). “On the generation of sampling schemes for Magnetic Resonance Imaging”. In : *Accepté à SIAM Imaging Science*.
- CANDÈS, Emmanuel J et Carlos FERNANDEZ-GRANDA (2014). “Towards a Mathematical Theory of Super-resolution”. In : *Communications on Pure and Applied Mathematics* 67.6, p. 906–956.
- CANDÈS, Emmanuel J, Justin ROMBERG et Terence TAO (2006). “Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information”. In : *Information Theory, IEEE Transactions on* 52.2, p. 489–509.
- CHALFIE, Martin et al. (1994). “Green fluorescent protein as a marker for gene expression”. In : *Science* 263.5148, p. 802–805.
- CHAUFFERT, Nicolas, Philippe CIUCIU, Jonas KAHN et Pierre WEISS (2014). “Variable density sampling with continuous trajectories. Application to MRI.” In : *SIAM Journal of Imaging Sciences* 7.4, p. 1962–1992. URL : <https://hal.inria.fr/hal-00908486>.
- (2016). “A projection algorithm on measure sets”. In : *Constructive Approximation*, p. 1–29.

- CHAUFFERT, Nicolas, Philippe CIUCIU, Jonas KAHN et Pierre Armand WEISS (2013). “Travelling salesman-based variable density sampling”. In : *10th international conference on Sampling Theory and Applications (SampTA 2013)*. Bremen, Germany, p. 509–512.
- CHAUFFERT, Nicolas, Pierre WEISS et al. (2016). “Gradient waveform design for variable density sampling in Magnetic Resonance Imaging”. In : *IEEE Transactions on Medical Imaging* 35.9, p. 2026–2039.
- COSTE, Michel (2000). *An introduction to o-minimal geometry*. Istituti editoriali e poligrafici internazionali Pisa.
- DEELY, J. J. et R. L. KRUSE (1968). “Construction of Sequences Estimating the Mixing Distribution”. In : *The Annals of Mathematical Statistics* 39.1, p. 286–288.
- DEUTSCH, David (1985). “Quantum theory, the Church-Turing principle and the universal quantum computer”. In : *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*. T. 400. 1818. The Royal Society, p. 97–117.
- DEUTSCH, David et Richard JOZSA (1992). “Rapid solution of problems by quantum computation”. In : *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*. T. 439. 1907. The Royal Society, p. 553–558.
- DIACONIS, Persi (1988). “Group representations in probability and statistics”. In : *Lecture Notes-Monograph Series* 11, p. i–192.
- (1996). “The cutoff phenomenon in finite Markov chains”. In : *Proceedings of the National Academy of Sciences* 93.4, p. 1659–1664.
- DIACONIS, Persi et James Allen FILL (1990). “Strong stationary times via a new form of duality”. In : *The Annals of Probability*, p. 1483–1522.
- DUNCAN, RR et al. (2004). “Multi-dimensional time-correlated single photon counting (TCSPC) fluorescence lifetime imaging microscopy (FLIM) to detect FRET in cells”. In : *Journal of microscopy* 215.1, p. 1–12.
- FAIR, Damien A et al. (2007). “Development of distinct control networks through segregation and integration”. In : *Proceedings of the National Academy of Sciences* 104.33, p. 13507–13512.
- FEYNMAN, Richard P (1982). “Simulating physics with computers”. In : *International journal of theoretical physics* 21.6/7, p. 467–488.
- FILL, J. et J. KAHN (2013). “Comparison Inequalities and Fastest-Mixing Markov Chains”. In : *Annals of Applied Probability* 23.5, p. 1778–1816. URL : arxi : 1109.6075.
- FÖRSTER, Th (1948). “Zwischenmolekulare energiewanderung und fluoreszenz”. In : *Annalen der physik* 437.1-2, p. 55–75.
- GENOVESE, Christopher R. et Larry WASSERMAN (2000). “Rates of convergence for the Gaussian mixture sieve”. In : *Ann. Statist.* 28.4, p. 1105–1127.
- GOWER, John C et GJS ROSS (1969). “Minimum spanning trees and single linkage cluster analysis”. In : *Applied statistics*, p. 54–64.

- GROMOV, Mikhail (2003). “Random walk in random groups”. In : *Geometric and Functional Analysis* 13.1, p. 73–146.
- GROVER, Lov K (1996). “A fast quantum mechanical algorithm for database search”. In : *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. ACM, p. 212–219.
- HASTINGS, W Keith (1970). “Monte Carlo sampling methods using Markov chains and their applications”. In : *Biometrika* 57.1, p. 97–109.
- HEINRICH, Philippe et Jonas KAHN (2015). “Optimal rates for finite mixture estimation”. In : *arXiv preprint arXiv :1507.04313 ; submitted to Annals of Statistics*.
- HEINRICH, Philippe, Jonas KAHN et al. (2011). *Remarks on the statistical study of protein-protein interaction in living cells*. URL : [arXiv:1105.5738](https://arxiv.org/abs/1105.5738).
- HEINRICH, Philippe, Mariano Gonzalez PISFIL et al. (2014). “Implementation of Transportation Distance for Analyzing FLIM and FRET Experiments”. In : *Bulletin of mathematical biology* 76.10, p. 2596–2626.
- HOLROYD, Alexander E (2011). “Some circumstances where extra updates can delay mixing”. In : *Journal of Statistical Physics* 145.6, p. 1649–1652.
- HOYOS-IDROBO, Andrés et al. (2016). “Recursive nearest agglomeration (ReNA) : fast clustering for approximation of structured signals”. In : *IEEE Transactions on Pattern Analysis and Machine Intelligence*. URL : <https://hal.inria.fr/hal-01366651>.
- JOHNSON, William B et Joram LINDENSTRAUSS (1984). “Extensions of Lipschitz mappings into a Hilbert space”. In : *Contemporary mathematics* 26.189-206, p. 1.
- JOULIN, Aldéric, Yann OLLIVIER et al. (2010). “Curvature, concentration and error estimates for Markov chain Monte Carlo”. In : *The Annals of Probability* 38.6, p. 2418–2442.
- KAHN, Jonas (2014). “How many T-tessellations on k lines ? Existence of associated Gibbs measures on bounded convex domains”. In : *Random Structures & Algorithms*, n/a–n/a. ISSN : 1098-2418. DOI : 10.1002/rsa.20557. URL : <http://dx.doi.org/10.1002/rsa.20557>.
- (2015). “Improper poisson line process as SIRS in any dimension”. In : *arXiv preprint arXiv :1503.03976 ; submitted to Annals of Probability*.
- KENDALL, W. S. (2014). “From random lines to metric spaces”. In : *Annals of Probability, to appear*. URL : [arxiv:1403.1156v1](https://arxiv.org/abs/1403.1156v1).
- KIËU, K. et al. (2013). “A completely random T-tessellation model and Gibbsian extensions”. In : *ArXiv e-prints*. arXiv : 1302.1809 [math.ST].
- LAKOWICZ, Joseph R (2013). *Principles of fluorescence spectroscopy*. Springer Science & Business Media.
- LE GALL, J.-F. (2014). “Random geometry on the sphere”. In : *ArXiv e-prints*. arXiv : 1403.7943 [math.PR].
- LOVÁSZ, László et Peter WINKLER (1995). “Mixing of Random Walks and Other Diffusions on a Graph”. In : *Survey in Combinatorics*. T. 218. Lecture Note Series. Cambridge University Press, p. 119–154.

- LUSTIG, Michael, David DONOHO et John M PAULY (2007). “Sparse MRI : The application of compressed sensing for rapid MR imaging”. In : *Magnetic resonance in medicine* 58.6, p. 1182–1195.
- MARSHALL, Albert W. et Ingram OLKIN (1979). *Inequalities : theory of majorization and its applications*. New York : Academic Press Inc. [Harcourt Brace Jovanovich Publishers], p. xx+569. ISBN : 0-12-473750-1.
- MCLACHLAN, G. et D. PEEL (2000). *Finite mixture models*. Wiley Series in Probability and Statistics : Applied Probability and Statistics. Wiley-Interscience, New York.
- MORIMAE, T. et J. KAHN (2010). “Entanglement-fidelity relations for inaccurate ancilla-driven quantum computation”. In : *Phys. Rev. A* 82, p. 052314. URL : [arXiv:1011.3806](https://arxiv.org/abs/1011.3806).
- NIELSEN, Michael A et Isaac L CHUANG (2010). *Quantum computation and quantum information*. Cambridge university press.
- PERES, Yuval et Peter WINKLER (2013). “Can extra updates delay mixing ?” In : *Communications in Mathematical Physics* 323.3, p. 1007–1016.
- RAHIMI, Ali et Benjamin RECHT (2007). “Random features for large-scale kernel machines”. In : *Advances in neural information processing systems*, p. 1177–1184.
- RAUSSENDORF, Robert et Hans J BRIEGEL (2001). “A one-way quantum computer”. In : *Physical Review Letters* 86.22, p. 5188.
- REBAFKA, Tabea (2009). “ESTIMATION DANS LE MODÈLE D’EMPILEMENT AVEC APPLICATION AUX MESURES DE LA FLUORESCENCE RÉSO-LUE EN TEMPS”. Thèse de doct. Télécom ParisTech.
- SHOR, Peter W (1994). “Algorithms for quantum computation : Discrete logarithms and factoring”. In : *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on*. IEEE, p. 124–134.
- SIMON, Daniel R (1997). “On the power of quantum computation”. In : *SIAM journal on computing* 26.5, p. 1474–1483.
- STOYAN, D., W. S. KENDALL et Joseph MECKE (1996). *Stochastic Geometry and Its Applications*. WILEY.
- TENG, Shang-Hua et Frances F YAO (2007). “K-nearest-neighbor clustering and percolation theory”. In : *Algorithmica* 49.3, p. 192–211.
- TEUBER, Tanja et al. (2011). “Dithering by differences of convex functions”. In : *SIAM Journal on Imaging Sciences* 4.1, p. 79–108.
- THIRION, B. et al. (2015). “Fast clustering for scalable statistical analysis on structured images”. In : *ICML Workshop on Statistics, Machine Learning and Neuroscience (Stamlins 2015)*.
- THOMSON, Joseph John (1904). “XXIV. On the structure of the atom : an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle ; with application of the results to the theory of atomic structure”. In : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 7.39, p. 237–265.

- WAHARTE ET AL., François (2006). “Setup and Characterization of a Multiphoton FLIM Instrument for Protein-Protein Interaction Measurement in Living Cells”. In : *Cytometry Part A* 69A, p. 299–306.
- WARD, Joe H (1963). “Hierarchical grouping to optimize an objective function”. In : *Journal of the American statistical association* 58.301, p. 236–244.
- ZALESKY, Andrew et al. (2014). “Time-resolved resting-state brain networks”. In : *Proceedings of the National Academy of Sciences* 111.28, p. 10341–10346.