



HAL
open science

Riemannian and stratified geometries on covariance and correlation matrices

Yann Thanwerdas

► **To cite this version:**

Yann Thanwerdas. Riemannian and stratified geometries on covariance and correlation matrices. Differential Geometry [math.DG]. Université Côte d'Azur, 2022. English. NNT : 2022COAZ4024 . tel-03698752v2

HAL Id: tel-03698752

<https://hal.science/tel-03698752v2>

Submitted on 21 Oct 2022

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.

THÈSE DE DOCTORAT

Géométries riemanniennes et stratifiées des matrices de covariance et de corrélation

Yann THANWERDAS

Inria, Epione, ERC G-Statistics

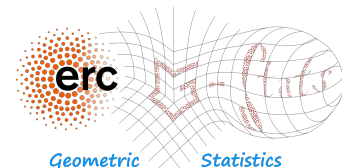
Présentée en vue de l'obtention du grade de docteur en Automatique, traitement du signal et des images
d'Université Côte d'Azur, dirigée par Xavier Penneec, soutenue le 24 mai 2022

Devant le jury composé de :

Directeur	Xavier PENNEC	Directeur de recherches	Inria, Epione, Sophia Antipolis
Rapporteurs	Pierre-Antoine ABSIL	Professeur	Université de Louvain
	Marc ARNAUDON	Professeur des universités	Université de Bordeaux
Examineurs	Rajendra BHATIA	Professeur	Ashoka University, Sonipat
	Minh HA QUANG	Professeur associé	RIKEN Center for AIP, Tokyo
	Huiling LE	Professeur émérite	University of Nottingham
	Frank NIELSEN	Directeur de recherches	Sony CSL, Tokyo
	Rodolphe SEPULCHRE	Professeur	Université de Cambridge

Inria

Epione
e-patient / e-medicine



Géométries riemanniennes et stratifiées des matrices de covariance et de corrélation

Dans de nombreuses applications, les données sont des matrices de covariance ou de corrélation entre plusieurs signaux (EEG, MEG, fMRI), grandeurs physiques (cellules, gènes) ou instants (autocorrélation). L'ensemble des matrices de covariance est un cône convexe qui est un espace stratifié non euclidien : il a un bord qui est lui-même un espace stratifié de dimension inférieure. Ses strates sont les variétés de matrices de covariance de rang fixé et la strate principale des matrices Symétriques Définies Positives (SPD) est dense dans l'espace total. L'ensemble des matrices de corrélations admet une structure similaire.

Les concepts géométriques comme les géodésiques, le transport parallèle ou la moyenne de Fréchet permettent de généraliser les opérations classiques (interpolation, extrapolation, recalage) et statistiques (moyenne, analyse en composantes principales, classification, régression) à ces espaces non linéaires. Cependant, ces généralisations reposent sur le choix d'une géométrie supposée connue à l'avance, c'est-à-dire d'un opérateur de base tel qu'une distance, une connexion affine, une métrique riemannienne, une divergence. En général il n'existe pas une unique géométrie adaptée à une application mais plutôt une famille de géométries à explorer pour faire ce choix.

D'abord, la géométrie doit correspondre au problème. Par exemple, si les matrices de covariance doivent être inversibles, les matrices dégénérées doivent être rejetées à l'infini. Ensuite, elle doit satisfaire aux invariances naturelles du problème par des groupes de transformations : si multiplier chaque variable par un facteur indépendant n'a pas d'influence, alors il faut une métrique invariante par le groupe des matrices diagonales strictement positives, par exemple une métrique produit qui découple les échelles et les corrélations. Enfin, de bonnes propriétés numériques (formes closes, algorithmes efficaces) sont essentielles pour utiliser cette géométrie en pratique.

Dans ma thèse, j'étudie des géométries sur les matrices de covariance et de corrélation suivant ces principes. En particulier, je fournis les opérations géométriques associées qui sont les briques élémentaires pour calculer avec ces matrices.

Sur les matrices SPD, je m'inspire de la caractérisation des métriques affine-invariantes pour caractériser les métriques continues invariantes par $O(n)$ au moyen de trois fonctions multivariées continues. Je construis ainsi une classification de métriques : les contraintes imposées sur ces fonctions définissent des classes emboîtées vérifiant des propriétés de stabilité. En particulier, je réinterprète la classe des "kernel metrics", j'introduis la famille des métriques "mixed-Euclidean" dont je calcule la courbure, et je résume et complète les connaissances sur les métriques classiques (log-euclidien, Bures-Wasserstein, BKM, power-Euclidean).

Sur les matrices de corrélation de rang plein, je calcule les opérations riemanniennes de la métrique quotient-affine et je montre que, malgré sa construction intéressante et son invariance par permutations, sa courbure est non majorée et de signe non constant, ce qui rend sa géométrie très complexe en pratique. Pour pallier ce défaut majeur, j'introduis des métriques Hadamard ou même log-euclidiennes ainsi que leurs opérations géométriques. Pour retrouver l'invariance par permutations perdue, je définis deux nouvelles métriques log-euclidiennes invariantes par permutations, l'une d'elle étant invariante par une involution naturelle de l'espace. Je fournis aussi un algorithme efficace pour calculer les opérations géométriques associées, qui s'appuie sur le "scaling" de matrices SPD.

Enfin, j'étudie la structure riemannienne stratifiée de la distance de Bures-Wasserstein sur les matrices de covariance. Je calcule le domaine de définition des géodésiques et le domaine d'injection dans chaque strate, puis je caractérise les courbes minimisant la longueur entre toutes les strates.

Mots clés: Géométrie riemannienne, Matrices de covariance, Matrices de corrélation, Familles de métriques, Géodésiques, Espaces stratifiés.

Riemannian and stratified geometries of covariance and correlation matrices

In many applications, the data can be represented by covariance matrices or correlation matrices between several signals (EEG, MEG, fMRI), physical quantities (cells, genes), or within a time window (autocorrelation). The set of covariance matrices forms a convex cone that is not a Euclidean space but a stratified space: it has a boundary which is itself a stratified space of lower dimension. The strata are the manifolds of covariance matrices of fixed rank and the main stratum of Symmetric Positive Definite (SPD) matrices is dense in the total space. The set of correlation matrices can be described similarly.

Geometric concepts such as geodesics, parallel transport, Fréchet mean were proposed for generalizing classical computations (interpolation, extrapolation, registration) and statistical analyses (mean, principal component analysis, classification, regression) to these non-linear spaces. However, these generalizations rely on the choice of a geometry, that is a basic operator such as a distance, an affine connection, a Riemannian metric, a divergence, which is assumed to be known beforehand. But in practice there is often not a unique natural geometry that suits the application. Thus, one should explore more general families of geometries that exploit the data properties.

First, the geometry must match the problem. For instance, degenerate matrices must be rejected to infinity whenever covariance matrices must be non-degenerate. Second, we should identify the invariance of the data under natural group transformations: if scaling each variable independently has no impact, then one needs a metric invariant under the positive diagonal group, for instance a product metric that decouples scales and correlations. Third, good numerical properties (closed-form formulae, efficient algorithms) are essential to use the geometry in practice.

In my thesis, I study geometries on covariance and correlation matrices following these principles. In particular, I provide the associated geometric operations which are the building blocks for computing with such matrices.

On SPD matrices, by analogy with the characterization of affine-invariant metrics, I characterize the continuous metrics invariant by $O(n)$ by means of three multivariate continuous functions. Thus, I build a classification of metrics: the constraints imposed on these functions define nested classes satisfying stability properties. In particular, I reinterpret the class of kernel metrics, I introduce the family of mixed-Euclidean metrics for which I compute the curvature, and I survey and complete the knowledge on the classical metrics (log-Euclidean, Bures-Wasserstein, BKM, power-Euclidean).

On full-rank correlation matrices, I compute the Riemannian operations of the quotient-affine metric. Despite its appealing construction and its invariance under permutations, I show that its curvature is of non-constant sign and unbounded from above, which makes this geometry practically very complex. I introduce computationally more convenient Hadamard or even log-Euclidean metrics, along with their geometric operations. To recover the lost invariance under permutations, I define two new permutation-invariant log-Euclidean metrics, one of them being invariant under a natural involution on full-rank correlation matrices. I also provide an efficient algorithm to compute the associated geometric operations based on the scaling of SPD matrices.

Finally, I study the stratified Riemannian structure of the Bures-Wasserstein distance on covariance matrices. I compute the domain of definition of geodesics and the injectivity domain within each stratum and I characterize the length-minimizing curves between all the strata.

Keywords: Riemannian geometry, Covariance matrices, Correlation matrices, Families of metrics, Geodesics, Stratified spaces.

Remerciements

Je voudrais adresser toute ma reconnaissance à mon directeur de thèse Xavier Pennec pour m'avoir accompagné et guidé pendant ces cinq années depuis mon premier stage jusqu'à l'aboutissement de cette thèse. Je me suis tout de suite senti en confiance tant scientifiquement qu'humainement, j'ai apprécié sa direction souple, c'est à dire le parfait équilibre entre d'un côté la présence infaillible, le suivi attentif et soucieux, la clarté des grandes lignes, la vision, et de l'autre la liberté laissée à l'expression personnelle jusqu'à l'humble effacement pour laisser l'oiseau voler de ses propres ailes. Ces qualités induisent un climat de confiance qui permet l'accomplissement personnel et le dépassement de soi. J'ai eu la chance de bénéficier de ce même climat de confiance avec mes parents, et à plusieurs reprises dans mon parcours, notamment à Centrale. Je suis heureux de l'avoir retrouvé ici auprès de Xavier et je l'en remercie infiniment. J'ai aussi apprécié les efforts constants pour nous emmener à de nombreux workshops et conférences, notamment Nicolas et moi au début, lorsque cela était possible, et plus généralement le fait de nous avoir généreusement présentés aux communautés de statistiques géométriques et cousines. La constitution de l'équipe G-Statistics au sein d'Epione et l'organisation des hackathons geomstats furent une façon de restaurer les interactions perdues à cause de la crise sanitaire. Je mesure bien la valeur de ces riches échanges qui, outre l'intérêt crucial de contrecarrer la dangereuse distanciation « sociale », m'ont beaucoup apporté dans mes recherches. Un immense merci pour tout cela.

I warmly thank Pierre-Antoine Absil and Marc Arnaudon for having accepted to review this manuscript and for their very careful and detailed reports and observations. I would like to thank Rodolphe Sepulchre who accepted to be the president of the jury and Rajendra Bhatia, Minh Ha Quang, Huiling Le and Frank Nielsen for being part of the jury. I am especially grateful to Frank for his strong support throughout the thesis, the many references he pointed to me and for being part of my PhD committee with Marco Lorenzi, whom I also thank a lot for his support and accurate suggestions. I would like to thank Minh for his invitation to spend some time in Japan, which unfortunately did not occur, and for sending me Rajendra's paper on the Bures-Wasserstein metric, without which I may have never had a look at it. In the same spirit, I thank Estelle Massart for showing me Takatsu's and her work on this Bures-Wasserstein metric, and Peter Michor for helping me understand the geometry of orbit spaces.

Je remercie Nicholas, grâce à qui tout a commencé, ainsi que Xavier, Hervé, Maxime, Marco et Irene de faire vivre l'équipe Epione de façon aussi dynamique et passionnée. Je mesure la chance que nous avons d'en faire partie, de pouvoir interagir avec tant de collègues et aussi, simplement, de pouvoir se faire des amis parmi eux. Je remercie tout particulièrement Isabelle pour son aide précieuse sur les missions, les conférences, les paiements, les contrats, les journaux, les réservations, la soutenance et tout ce que l'on ne voit pas mais qui est si important.

Au sein de la petite entreprise de Xavier, G-Statistics, qui ne connaît pas la crise, j'aimerais avant tout remercier Nicolas qui fut mon compagnon de route, qui m'a parlé Gallier et Postnikov tandis que je lui répondais Paulin et O'Neill – entre autres Sharpe, Lee, Kobayashi et Nomizu –, qui a eu la patience de m'initier à geomstats, PyCharm, et même au cluster, et avec lequel j'ai eu plaisir à découvrir quelques opéras et instruments de musique. Je remercie Nina pour avoir initié l'aventure geomstats et pour nous y avoir vite associés.

Je remercie Sofia pour nos échanges passionnés sur la géométrie riemannienne et les espaces barycentriques, Bastien pour ses explications patientes sur la théorie des représentations et la topologie, Dimby pour nos discussions sur les connexions, ses questions profondes qui furent la source de grand tracasseries cependant fructueuses et ses conseils sur le monde académique, Anna pour nos discussions sur les graphes, les arbres, les forêts et autres espaces stratifiés. Je tiens à les remercier ainsi qu'Elodie, Morten, James, Josquin, Luís et Tom pour nos échanges stimulants et pour l'ambiance très amicale de nos rendez-vous. Je remercie aussi les participants aux hackathons geomstats, particulièrement Saiteja, Alice, Yann, Thomas, Hadi, Jonas, Yann, Sylvain, Daniel, Niklas et Stefan.

Je remercie mes collègues d'Epione qui furent de formidables camarades, notamment pour décompresser au MK, à Green ou sur la plage. Je remercie particulièrement Buntheng pour son aide technique avant et pendant la soutenance, ainsi que Paul, Hind et Dimitri pour leur précieuse aide sur la présentation. Je remercie chaleureusement mes collègues de l'ère Asclepios qui m'ont donné envie de revenir, notamment Julian, Raphaël, Shuman, Wen, Nicoco, Qiao et Pawel, puis Luigi, Clément, Jaume, Tania, Benoît, Zihao, Gaëtan, Florent et Santi avec qui j'ai passé de très agréables moments.

Je voudrais remercier le comité de la Bourse d'Excellence 2018 d'Université Côte d'Azur de me l'avoir accordée ainsi que leur confiance. Je remercie particulièrement Hanane Berouag, Françoise Renaudat, Anne Linares et Emmanuel Scarsi de la Maison des Etudes Doctorales ainsi que Claire Migliaccio et Régine Saelens de l'Ecole Doctorale STIC pour leur aide tout au long de ma thèse.

Je suis très reconnaissant à Philippe Maisonobe, Ann Lemahieu, Isabelle Delorme, Erwan Aubry, Christian Pauly et François Delarue d'avoir rendu possible mon intervention comme vacataire au sein du département de mathématiques d'UCA. Ces deux TD ont été une respiration au sein de mon travail de thèse, j'ai beaucoup apprécié mes visites à Valrose et les échanges avec les étudiants et étudiantes pour qui j'ai une pensée. Je remercie particulièrement Ann et Isabelle pour leur aide immense sans laquelle j'aurais été perdu, ainsi que François pour son écoute, son expérience et ses précieux conseils.

Je souhaite remercier John, Erick et Michel pour leur oreille attentive, leurs conseils avisés et l'amitié dont ils m'honorent. Je profite de cette occasion pour saluer et remercier tous mes professeurs.

Je remercie tous mes amis, avec qui j'ai vécu d'intenses moments et dont la présence m'est très précieuse. Je remercie Clément et Selma pour nos débats endiablés et nos fous rires à la table de ce très cher Miguel, que je remercie aussi chaleureusement. Je remercie Pauline pour sa délicatesse et ses messages toujours attentionnés, Loriane pour son écoute réconfortante et sa compréhension intuitive, Marine pour sa mémoire de tout et son énergie communicative, Marine pour sa débrouillardise fascinante et inspirante, Benjamin pour sa résilience modèle et son humour, Florent pour sa liberté et son authenticité, Guillaume pour notre complicité émotionnelle, Paul pour son originalité et celle de ses réflexions, Jérôme pour sa parole réfléchie et sa folie touche-à-tout, Santi pour sa drôlerie spirituelle et son sourire perpétuel, Jaume pour sa profonde gentillesse et sa paix intérieure, Guillaume pour sa chaleur accueillante, Juanjo pour sa sincérité apaisante et Koji pour sa générosité inclusive, ainsi que pour leur écoute bienveillante à tous les trois. Je remercie Radomir qui me suit de son affection immodérée et Thierry qui me guide de sa profonde sérénité.

Je remercie ma famille d'être toujours là et de croire en moi, mes grands-parents pour

leur amour inconditionnel, ma Tati-marraine qui veille toujours sur moi, Laurent pour notre complicité qui supporte la distance, Jean-Michel, Marie-Cécile, Tonton Jacky, Claude, Rémi, Juliet, Joël, Taoufik, Khadija, Marie-Louise, Jeannine pour leurs pensées toujours tendres et affectueuses. J'ai moi-même une pensée émue pour Jean qui restera une grande source d'inspiration, et une pensée espiègle pour Leila et Yannis. Je remercie mes parents pour leur soutien indéfectible et leur présence réconfortante. Merci d'avoir toujours encouragé ma curiosité, mon indépendance, et de m'avoir transmis tant de vous, en particulier le goût des mathématiques, le dépassement de soi et le respect de chacun. Je remercie Sara pour son soutien quotidien, sa tendresse, sa profonde gentillesse, sa fraîcheur innocente, son humour décalé, son honnêteté et son sens de l'honneur, sa curiosité et sa culture. Volim te.

Funding

This thesis was funded by the French government, through the UCAJEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01. The contract extensions, travels, equipment, participation to conferences and publication fees were funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant G-Statistics agreement No 786854).

Contents

I	Introduction	13
1	Introduction	14
1.1	Which geometric structure for non-linear data?	14
1.1.1	Riemannian manifolds	14
1.1.2	Metric spaces	16
1.1.3	Stratified spaces	17
1.1.4	Riemannian orbit spaces	18
1.1.5	How to choose the geometry	19
1.2	Spaces of covariance and correlation matrices	22
1.2.1	Two stratified spaces	22
1.2.2	The many geometries of SPD and PSD matrices	23
1.2.3	The few geometries of correlation matrices	25
1.3	Comprehensive presentation of the contributions	25
1.3.1	Purposes and methodology	25
1.3.2	Classification of Riemannian metrics on SPD matrices	27
1.3.3	New families of Hadamard metrics on full-rank correlation matrices	29
1.3.4	Bures-Wasserstein geodesics in the orbit space of covariance matrices	31
1.4	Organization of the thesis	32
1.4.1	Overview	32
1.4.2	Summary of contributions by chapter	33
II	Mathematical preliminaries	36
2	Regularity of eigenvalues and eigenvectors	37
2.1	Introduction	37
2.2	No continuous map of eigenvectors on $\text{Sym}^{\neq}(n)$	40
2.3	Local Lipschitz maps of eigenvectors in $\text{Sym}^{\neq}(n)$	41
2.3.1	Majoration of the distance between eigenvector matrices in $\text{Sym}^{\neq}(n)$	41
2.3.2	Existence of local Lipschitz maps of eigenvectors in $\text{Sym}^{\neq}(n)$	42
2.4	Smoothness of eigenvalues and local eigenvectors on $\text{Sym}^{\neq}(n)$	43
2.5	Majoration of the distance between eigenvector matrices in $\text{Sym}(n)$	44
2.6	Conclusion	46

3	Characterization of invariant inner products	47
3.1	Introduction	47
3.2	Characterization of invariant inner products of a Euclidean space using representation theory	49
3.2.1	Preliminaries	49
3.2.2	Step 1: from invariant inner products to equivariant automorphisms	50
3.2.3	Steps 2 and 3: find all equivariant automorphisms and all invariant inner products	51
3.3	Examples	54
3.3.1	Invariant inner products on \mathbb{R}^n	55
3.3.2	$O(n)$ -invariant inner products on $\text{Mat}(n)$	56
3.3.3	$O(I)$ -invariant inner products on $\text{Mat}(n)$	58
3.3.4	$\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$	61
3.4	Conclusion	64

III Covariance matrices of full rank: the open cone of symmetric positive definite matrices **66**

4	$O(n)$-invariant Riemannian metrics	67
4.1	Introduction	67
4.1.1	Results and organization of the chapter	69
4.1.2	Notations and conventions	70
4.2	Preliminary concepts and results	73
4.2.1	Extending maps defined on diagonal matrices	73
4.2.2	Univariate maps	74
4.2.3	$O(n)$ -invariant inner products on symmetric matrices	74
4.3	Main $O(n)$ -invariant metrics on SPD matrices with new formulae	76
4.3.1	$O(n)$ -invariant Euclidean metrics	76
4.3.2	$O(n)$ -invariant log-Euclidean metrics	76
4.3.3	Affine-invariant metrics	77
4.3.4	Bures-Wasserstein metric	79
4.3.5	Bogoliubov-Kubo-Mori metric	81
4.4	The interpolating class of kernel metrics: new observations	83
4.4.1	The general class of kernel metrics	83
4.4.2	New observations on kernel metrics	84
4.5	Characterization of $O(n)$ -invariant metrics	87
4.5.1	Theorem and corollaries	87
4.5.2	Proof of the theorem	88
4.5.3	Reinterpretation of kernel metrics	90
4.5.4	Key results on $O(n)$ -invariant metrics	90
4.5.5	Bivariate separable metrics	92
4.6	Conclusion	93

5	Geometry of Mixed-Euclidean metrics	95
5.1	Introduction	95
5.2	Notations and preliminary concepts	97
5.2.1	Notations	97
5.2.2	Univariate maps	97
5.2.3	Classes of $O(n)$ -invariant metrics	98
5.3	Deformed metrics	100
5.3.1	Use of a deformation in the literature	100
5.3.2	Principle of deformed metrics	101
5.3.3	The new family of deformed-Wasserstein metrics	102
5.4	Balanced metrics	103
5.4.1	Information geometry and dually-flat manifolds	103
5.4.2	Principle of balanced bilinear forms	105
5.5	The new family of mixed-Euclidean metrics	106
5.5.1	Definition	106
5.5.2	Information geometry of Mixed-Euclidean metrics	106
5.5.3	Riemannian geometry of Mixed-Euclidean metrics	108
5.6	Conclusion	111
IV	Correlation matrices of full rank: the open ellipsope	112
6	Geometry of Quotient-affine metrics	113
6.1	Introduction	113
6.2	Quotient-affine metrics	114
6.2.1	The quotient manifold of full-rank correlation matrices	114
6.2.2	The affine-invariant metrics and the quotient-affine metrics	115
6.3	Fundamental Riemannian operations	116
6.3.1	Vertical and horizontal distributions and projections	116
6.3.2	Horizontal lift and metric	116
6.3.3	Geodesics	117
6.3.4	Levi-Civita connection and sectional curvature	117
6.4	Illustration in dimension 2	118
6.5	Conclusion	119
7	Theoretically and computationally convenient Cholesky-based geometries	121
7.1	Introduction	121
7.1.1	Overview of the results	123
7.1.2	Concepts and notations	124
7.2	Quotient-affine metrics	127
7.2.1	Quotient metrics	127
7.2.2	Definition of quotient-affine metrics	127
7.2.3	Complement: bounds of curvature	128
7.3	Generalization of quotient-affine metrics	129
7.3.1	Congruence actions of matrix Lie groups on SPD matrices	129

7.3.2	Lie-Cholesky metrics	130
7.3.3	Quotient-Lie-Cholesky metrics	132
7.4	New geometric structures with unique mean	133
7.4.1	Symmetric space: poly-hyperbolic-Cholesky metrics	133
7.4.2	Vector space: Euclidean-Cholesky metrics	135
7.4.3	Vector space: log-Euclidean-Cholesky metrics	135
7.4.4	Nilpotent Lie group structure	136
7.4.5	Explicit geodesics in dimension 2	137
7.5	Conclusion	138
8	Permutation-invariant Log-Euclidean metrics	139
8.1	Introduction	139
8.1.1	Results and organization of the chapter	141
8.1.2	Notations	142
8.2	Permutation-invariant log-Euclidean metrics via the off-log diffeomorphism	145
8.2.1	The off-log bijection	146
8.2.2	Permutation-invariant pullback metrics via the off-log diffeomorphism	147
8.2.3	Numerical computation of the off-log inverse map	148
8.3	The log-scaling bijection	149
8.3.1	Is the conjecture true with $\mathcal{V}^* = \text{Hol}(n)$?	149
8.3.2	Why $\mathcal{V}^* = \text{Row}_0(n)$ seems to be a better choice	150
8.3.3	The conjecture is true with $\mathcal{V}^* = \text{Row}_0(n)$	151
8.3.4	Properties of the log-scaling bijection	153
8.4	Permutation-invariant log-Euclidean metrics via the log-scaling bijection	153
8.4.1	The cor-exp bijection is a diffeomorphism	154
8.4.2	Permutation-invariant inner products on $\text{Row}_0(n)$	154
8.4.3	Pullback metrics via the log-scaling diffeomorphism	154
8.5	Numerical computation of the log-scaling	155
8.5.1	The optimization problem	156
8.5.2	Optimizing an α -strongly convex β -smooth map on a closed convex set	156
8.5.3	Approximate the projection	158
8.5.4	Approximated projection gradient descent	160
8.6	Conclusion	164
V	Stratified spaces of covariance and correlation matrices	165
9	Bures-Wasserstein stratified geometry of covariance matrices	166
9.1	Introduction	166
9.2	Preliminary concepts	170
9.2.1	Geodesics	171
9.2.2	Quotient spaces	174
9.3	Bures-Wasserstein geometry of covariance matrices	176
9.3.1	The quotient geometry of covariance matrices	177
9.3.2	The Bures-Wasserstein distance	177

9.3.3	Topology, metric and smooth structure of the strata	178
9.4	Geodesics of the Bures-Wasserstein metric on $\text{Sym}^+(n)$	179
9.5	Geodesics of the Bures-Wasserstein metric on $\text{Sym}^+(n, k)$	181
9.6	Minimizing geodesics of the Bures-Wasserstein distance on $\text{Cov}(n)$	186
9.6.1	Characterization of minimizing geodesics	186
9.6.2	Number of minimizing geodesics	186
9.7	Conclusion	188
VI	Conclusion	190
10	Conclusion and perspectives	191
10.1	Summary of contributions and beyond	191
10.2	Future works	193
10.2.1	Other geometries for the rank stratification	193
10.2.2	Geometries of matrices of any size and any rank	194
10.2.3	Eigenvalue multiplicity and block-wise transformations	195
10.2.4	Applications	195
10.3	Publications	196
VII	Appendix	198
11	Proofs	199
11.1	Proofs of Chapter 2	199
11.2	Proofs of Chapter 3	200
11.3	Proofs of Chapter 4	213
11.4	Proofs of Chapter 5	219
11.5	Proofs of Chapter 6	226
11.6	Proofs of Chapter 7	230
11.7	Proofs of Chapter 8	240
11.8	Proofs of Chapter 9	250
	Bibliography	262

Part I

Introduction

Chapter 1

Introduction

1.1 Which geometric structure for non-linear data?

1.1.1 Riemannian manifolds

Data science consists in extracting knowledge from the data collected in a broad variety of contexts. The data may take multiple forms such as signals (brain activity, radar), images (MRI, scanner), videos, shapes (organs, bones, proteins, cells), networks, etc. These data are generally modelled by mathematical objects such as structured matrices, trees, graphs, meshes, shapes, curves, diffeomorphisms. To analyze the data, one needs to compute with this kind of objects. However, these mathematical objects often live in non-linear spaces where the classical statistical operations are not defined. Indeed, without a vector space structure, it seems impossible to compute means, to perform principal component analyses, to do interpolations or extrapolations for example.

Therefore, a natural idea that became popular in data science over the last 20 years is to consider that the data live in a Riemannian manifold, that is a locally Euclidean space. It allows to generalize a lot of Euclidean operations, in particular statistical operations. Firstly, thanks to the differential structure, one can differentiate curves, functions and define the gradient. If the manifold is oriented, which is often the case in practice, there is a canonical volume form which defines the Riemannian measure. Thus, absolutely continuous measures and especially densities of probability measures can be defined. Secondly, the Riemannian metric defines a canonical affine connection called the Levi-Civita connection. Thus, the Riemannian manifold inherits the notions of geodesics (or self-parallel curves), exponential map, parallel transport and Riemann curvature. Other notions of curvature may be defined such as the scalar curvature, the Ricci curvature and especially the sectional curvature which characterizes the Riemann curvature. Thirdly, the Riemannian metric defines a notion of length and an intrinsic distance called the Riemannian distance on each connected component of the manifold. The locally minimizing geodesics of the Riemannian distance coincide with the geodesics of the Levi-Civita connection. Moreover, the distance allows to define (globally) minimizing geodesics and Riemannian logarithms. In addition, the Hopf-Rinow theorem ensures that the metric space is complete if and only if the Riemannian manifold is geodesically complete. All this is summarized on the left part of Table 1.1.

This very rich set of geometric tools led to generalize many statistical concepts and algorithms from Euclidean spaces to Riemannian manifolds. In a metric space (\mathcal{M}, d) , a

Fréchet mean [Fréchet, 1948] of the points $x_1, \dots, x_k \in \mathcal{M}$ is a minimizer of the continuous map $f : x \in \mathcal{M} \mapsto \sum_{i=1}^k a_i d(x, x_i)^p$ with $p = 2$ and $a_1 = \dots = a_k = 1$. More generally, this defines the notion of Riemannian L^p center of mass for any $p \geq 1$ and the notion of Riemannian barycenter for any coefficients a_1, \dots, a_k . Thus, these notions are defined in Riemannian manifolds via the Riemannian distance. Moreover, precise conditions were derived to ensure their local existence, uniqueness and convexity [Karcher, 1977, Buser and Karcher, 1981, Kendall, 1990, Le, 2004, Yang, 2010, Afsari, 2011, Arnaudon and Miclo, 2014] and the consistence and robustness of empirical estimators [Bhattacharya and Patrangenaru, 2003, Arnaudon et al., 2013]. In addition, several gradient descent algorithms were proposed to compute them [Penec, 2006, Arnaudon et al., 2012, Arnaudon et al., 2013, Afsari et al., 2013, Bonnabel, 2013]. More generally, many optimization algorithms (Riemannian Newton methods, line-search methods, trust-region methods) based on the concept of retraction were developed on matrix manifolds to tackle classical problems such as the eigenvalue problem [Absil et al., 2009, Boumal et al., 2014]. Other main statistical operations were defined such as the covariance of a sample, the Mahalanobis distance, the normal law or the χ^2 law [Penec, 2006]. A central limit theorem was established [Bhattacharya and Patrangenaru, 2005] and several generalizations of Principal Component Analysis (PCA) were imagined: tangent PCA, Principal Geodesic Analysis [Fletcher and Joshi, 2004, Sommer et al., 2014], Geodesic PCA [Huckemann et al., 2010]. Approximated parallel transport such as Schild's ladder [Lorenzi et al., 2011] and pole ladder [Lorenzi and Penec, 2013, Guigui and Penec, 2021a] were also proposed for registration purposes. This is summarized on the right part of Table 1.1.

Geometric operations	Statistical operations, results, algorithms
<ul style="list-style-type: none"> • <u>Manifold</u> 	Fréchet mean, p -mean, barycenters [Fréchet, 1948]
Coordinate charts	From [Karcher, 1977] to [Arnaudon and Miclo, 2014]
Tangent space, tangent vector	Gradient descent algorithms [Penec, 2006, Bonnabel, 2013]
Derivative, speed vector	[Arnaudon et al., 2013, Afsari et al., 2013]
<ul style="list-style-type: none"> • <u>Riemannian metric</u> 	Optimization algorithms [Absil et al., 2009, Boumal et al., 2014]
Gradient	Covariance matrix, Mahalanobis distance,
(If oriented) Volume form	Normal law, χ^2 law [Penec, 2006]
Riemannian measure	(If compact) Uniform law
Probability measures	Central limit theorem
<ul style="list-style-type: none"> • <u>Levi-Civita connection</u> 	[Bhattacharya and Patrangenaru, 2005]
Geodesics, exponential map	Tangent PCA
Parallel transport	Principal Geodesic Analysis [Fletcher and Joshi, 2004]
Curvature	[Sommer et al., 2014]
+ sectional/Ricci/scalar	Geodesic PCA [Huckemann et al., 2010]
<ul style="list-style-type: none"> • <u>Riemannian distance</u> 	Geodesic regression [Fletcher, 2013]
Length, length space	Schild's ladder [Lorenzi et al., 2011]
Minimizing geodesics	Pole ladder [Lorenzi and Penec, 2013]
Logarithm map	[Guigui and Penec, 2021a]

Table 1.1: Main geometric and statistical operations in Riemannian manifolds.

In other words, the framework of Riemannian manifolds is very rich. It offers many geometric operations that were used to build efficient algorithms and adapt the usual statistical

operations to non-linear spaces. However, it is not always possible to equip a space with a differential structure so the hypothesis of working in a Riemannian manifold may be a bit restrictive. A more general framework is given by metric spaces.

1.1.2 Metric spaces

Spaces of trees and graphs are typical examples of spaces that are too complex to admit a manifold structure. We can cite the BHV space of phylogenetic trees [Billera et al., 2001], the QED space of unlabeled trees [Feragen et al., 2010] or more recently the Graph space [Calissano et al., 2020] and the Wald space of forests [Garba et al., 2021]. The three first spaces are complete length spaces and even geodesic spaces. A length space is a metric space such that the distance between any two points coincides with the infimum of lengths of curves between them. A geodesic space is a metric space such that between any two points, there exists a (globally minimizing) geodesic. Beware that in the context of metric spaces, a geodesic often designates a globally minimizing geodesic whereas in Riemannian manifolds, it rather stands for a locally minimizing geodesic, or equivalently a self-parallel curve. The Hopf-Rinow theorem ensures that any complete and locally compact length space is a geodesic space. A notion of curvature due to Aleksandrov also exists in metric spaces [Aleksandrov et al., 1986, Alexander et al., 2019, Feragen and Nye, 2020]. It consists in comparing the areas of triangles with the model spaces of constant curvature: the sphere, the Euclidean space and the hyperbolic space. It defines upper or lower bounds for the curvature of the metric space.

The notions of minimizing geodesics, Fréchet means and its variants are defined in a metric space (\mathcal{M}, d) . The consistency of the Fréchet mean and median empirical estimators still holds [Bhattacharya and Patrangenaru, 2003, Arnaudon et al., 2013]. More generally, a distance defines a measure of dissimilarity between points, which can be taken as a loss function. Thus, one can define least square models [Feragen and Nye, 2020] and therefore geodesic regression or principal geodesics. However, the main difference with Riemannian manifolds is the scarce quantity of available tools to compute these statistical operations. The tools for metric spaces are more specific to a given space, such as the computation of geodesics in the BHV space [Kupczok et al., 2008, Owen and Provan, 2011], the phylogenetic PCA [Nye, 2011, Nye et al., 2017] or the computation of Fréchet means for persistence diagrams [Turner et al., 2014] for example.

Fortunately, there are subclasses of metric spaces where we retrieve generic results and methods. For instance, there exist many references on non-positively curved spaces [Bridson and Haefliger, 1999], results on their probability measures [Sturm, 2003] and algorithms to compute medians and means in Hadamard spaces [Bacák, 2014]. Hadamard spaces are complete non-positively curved geodesic spaces, they are generalizations of Hadamard manifolds and satisfy analogous properties. For example, there exists a unique geodesic segment between any two points. In contrast, it seems that few tools exist on non-negatively curved spaces. Another example of notable metric spaces is given by quotients of metric spaces by isometric group actions. When the total space is a Riemannian manifold, Geodesic PCA is still defined [Huckemann et al., 2010], potentially with generalized geodesics. When the group is finite like for the Graph space, there exist algorithms of type “Align All and Compute” to compute the Fréchet mean or the geodesic components [Calissano et al., 2020]. They

converge in finite time since the group is finite.

Thus, the framework of metric spaces is much more limited than the one of Riemannian manifolds. However, there exist subclasses of convenient spaces and examples such as spaces of trees that were extensively studied. Most of examples of metric spaces that are not Riemannian manifolds actually convey more structure than a simple distance: they are often stratified spaces.

1.1.3 Stratified spaces

Stratified spaces were defined and studied by Hassler Whitney, René Thom and John Mather [Thom, 1969, Mather, 1970]. A \mathcal{C}^p stratification of a closed set \mathcal{X} in a manifold \mathcal{M} is a sequence of nested spaces $\mathcal{X} = \mathcal{X}_k \supset \mathcal{X}_{k-1} \supset \dots \supset \mathcal{X}_0 \supseteq \mathcal{X}_{-1} := \emptyset$ such that for all $i \in \{0, \dots, k\}$, the set $\mathcal{M}_i = \mathcal{X}_i \setminus \mathcal{X}_{i-1}$ is a \mathcal{C}^p manifold and the map $i \in \{0, \dots, k\} \mapsto \dim \mathcal{M}_i$ is increasing. (As in [Trotman, 2020], we can always impose $\dim \mathcal{M}_i = i$ by adding copies of \mathcal{X}_i 's and by allowing that \mathcal{M}_i might be empty.) The connected components of the \mathcal{M}_i 's are called the strata. Therefore \mathcal{M}_i can be split into strata \mathcal{S}_i^j for $1 \leq j \leq \alpha_i$ where α_i is the number of connected components of $\mathcal{M}_i = \bigsqcup_{1 \leq j \leq \alpha_i} \mathcal{S}_i^j$. Thus the space \mathcal{X} is split into connected manifolds, $\mathcal{X} = \bigsqcup_{i=0}^k \bigsqcup_{j=1}^{\alpha_i} \mathcal{S}_i^j$. We say that the stratum \mathcal{S} is adjacent to the stratum \mathcal{T} and we write $\mathcal{T} < \mathcal{S}$ when $\mathcal{T} \subset \bar{\mathcal{S}} \setminus \mathcal{S}$. This defines the partial order of adjacency on the set of strata. We can represent the adjacency by the Hasse diagram of this partial order.

We say that a pair of strata $(\mathcal{S}, \mathcal{T})$ satisfies Whitney's condition (b) at $t \in \mathcal{T}$ when for all sequences $(s_n) \in \mathcal{S}^{\mathbb{N}}$ and $(t_n) \in \mathcal{T}^{\mathbb{N}}$ tending to t , if in a chart of \mathcal{M} around t , the sequence of tangent spaces $(T_{s_n} \mathcal{S})$ tends to a space τ in the Grassmannian $\text{Gr}(\dim \mathcal{M}, \dim \mathcal{S})$ and the sequence of lines $(s_n t_n)$ tends to a line λ in the Grassmannian $\text{Gr}(\dim \mathcal{M}, 1)$, then $\lambda \subseteq \tau$. A Whitney stratification is a \mathcal{C}^1 stratification such that the number of strata is locally finite (each point has a neighborhood that meets a finite number of strata) and such that for all strata $\mathcal{T} < \mathcal{S}$, the pair $(\mathcal{S}, \mathcal{T})$ satisfies the Whitney's condition (b) at all $t \in \mathcal{T}$. Therefore, one can show that $\mathcal{T} < \mathcal{S}$ implies $\dim \mathcal{T} < \dim \mathcal{S}$ [Trotman, 2020], so if $\mathcal{T} \subseteq \mathcal{M}_i$ and $\mathcal{S} \subseteq \mathcal{M}_{i'}$, then $i < i'$. In other words, the map $\mathcal{S} \subseteq \mathcal{M}_i \mapsto i \in \{0, \dots, k\}$ is increasing. Let us illustrate the definition on examples.

A manifold with boundary \mathcal{M} (that is locally diffeomorphic to an open set of a half-space $\mathbb{R}^{n-1} \times [0, +\infty)$) is a stratified space with $\mathcal{X}_0 = \partial \mathcal{M}$ (the boundary) and $\mathcal{X}_1 = \mathcal{M}$ (the manifold without boundary). For example, the closed ball of \mathbb{R}^n for the Euclidean norm $\bar{\mathcal{B}}_2^n(0, 1) = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}$ is composed of the sphere $\mathcal{M}_0 = \mathcal{S}_2^n(0, 1)$ and the open ball $\mathcal{M}_1 = \mathcal{B}_2^n(0, 1)$. The dimensions are $\dim(\mathcal{M}_0) = n - 1$ and $\dim(\mathcal{M}_1) = n$. The closed ball of \mathbb{R}^3 for the infinite norm $\bar{\mathcal{B}}_\infty^3(0, 1) = \{x \in \mathbb{R}^3 \mid \|x\|_\infty = \max |x_i| \leq 1\}$ is a cube where the strata are the 8 vertices \mathcal{M}_0 , the 12 edges \mathcal{M}_1 , the 6 faces \mathcal{M}_2 and the open cube \mathcal{M}_3 . The dimensions are $\dim(\mathcal{M}_i) = i$. The Hasse diagram of adjacency of the closed unit square of \mathbb{R}^2 is represented on Figure 1.1. Note that when the strata are connected manifolds, the diagram is a trivial line $\mathcal{M}_0 \rightarrow \dots \rightarrow \mathcal{M}_k$. Other popular examples are the spider and the open book exposed in [Feragen and Nye, 2020]. The spider is a set of half-lines connected at their finite extremity. The open book is the Cartesian product of \mathbb{R} and the spider, that is a set of half-planes sharing their boundary line. Their Hasse diagrams of adjacency are represented on Figure 1.2.

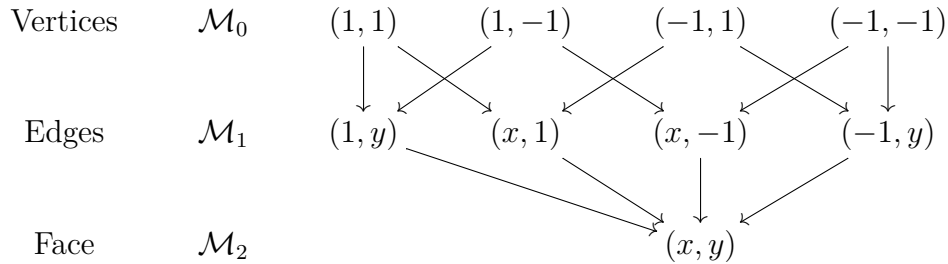


Figure 1.1: Hasse diagram of adjacency of the canonical stratification of the closed unit square $\bar{\mathcal{B}}_\infty^2(0, 1) = \{(x, y) \in [-1; 1]^2 \mid \max(|x|, |y|) \leq 1\}$. This is an upper semi-lattice. On the diagram, the free variables x and y stand for all the interval $(-1; 1)$.

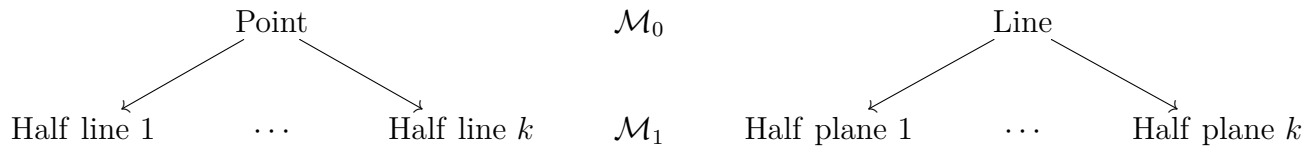


Figure 1.2: Hasse diagrams of adjacency of the spider (left) and the open book (right). These are lower semi-lattices.

Metric stratified spaces are stratified spaces endowed with a distance inducing the same topology. The BHV space [Billera et al., 2001] and orthant spaces [Barden and Le, 2018], the QED space [Feragen et al., 2010], the Graph space [Calissano et al., 2020], the Wald space [Garba et al., 2021] are important examples of metric stratified spaces used in applications. They offer more structure than simple metric spaces. For example, the BHV space is a collection of Euclidean orthants in which the topology of the tree is preserved. They are glued together so that the intersection between two orthants corresponds to the shrinking of an edge in the tree. Therefore, minimizing geodesics are piecewise Euclidean [Billera et al., 2001, Nye, 2011, Feragen and Nye, 2020]. A logarithm, generalizing the Riemannian logarithm, is defined and the Fréchet mean and its limiting distribution can be precisely characterized [Barden and Le, 2018]. We would like to go beyond spaces of trees and graphs where methods seem to be mainly ad hoc. Thus, we promote an interesting subclass of metric stratified spaces with more structure which is the class of Riemannian orbit spaces [Alekseevsky et al., 2001].

1.1.4 Riemannian orbit spaces

In this thesis, we choose to work with a particular type of metric stratified spaces that we call Riemannian orbit spaces. A Riemannian orbit space is the quotient of a Riemannian manifold by a proper and isometric Lie group action [Alekseevsky et al., 2001, Huckemann et al., 2010]. (Note that the condition of a *proper* group action is classical to get a Hausdorff separated quotient space.) When the action is free in addition, the quotient space has a unique structure of smooth manifold such that the canonical projection is a submersion [Lee, 2012]. It can be turned into a Riemannian submersion by descending the Riemannian metric. The theory of Riemannian submersions apply so the geodesics and the curvature can

be deduced from the ones of the total space [O’Neill, 1966]. When the action is not free, the isotropy group is not reduced to identity: the points can be grouped by conjugation class of isotropy group, this creates strata in the total space and in the quotient space, called orbit space [Alekseevsky et al., 2001, Michor, 2008]. The canonical projection becomes a stratified Riemannian submersion in a sense. Moreover, the Riemannian distance of the total space descends to a quotient distance on the orbit space so we get a metric stratified space. The strata are proved to be Riemannian manifolds [Alekseevsky et al., 2001, Michor, 2008].

Riemannian orbit spaces can be seen as the stratified generalization of quotients of Riemannian manifolds by a free, proper, isometric Lie group action. Some results on complete Riemannian quotient manifolds still hold in complete Riemannian orbit spaces. For example, any minimizing geodesic segment of the orbit space is the projection of a horizontal geodesic segment, called a horizontal lift [Alekseevsky et al., 2001]. More precisely, the length of a horizontal curve coincides with the length of its projection as long as it does not meet an orbit of more singular type [Michor, 2008]. Moreover, all the horizontal lifts of a geodesic differ by the action of an element in the Lie group [Alekseevsky et al., 2001]. These nice properties allow to characterize the geodesics in a complete Riemannian orbit space. Hence, they are a particularly convenient type of metric stratified spaces.

1.1.5 How to choose the geometry

After detailing these possible choices of geometric structures, how can we choose between them? Moreover, even within one of these classes of structures, does there exist a canonical geometric structure to use? Otherwise, how can we choose it? In Euclidean spaces, the notions of straight line or mean do not depend on the choice of the inner product because they all share the same canonical affine connection. In non-linear spaces, all these notions strongly depend on the choice of the geometric structure, should it be a Riemannian metric or a distance. We identify three principles to choose the geometric structure for non-linear data.

1.1.5.1 First principle: adequacy of the model to the data

The first principle is the adequacy of the model to the data. Do they live in a manifold or are there singularities to take into account in a more complex model such as a stratified space? For example, in the set of covariance matrices, rank-deficient matrices may correspond to a non-physical reality. For instance, in diffusion tensor imaging or electro-encephalography, these matrices are often considered as degenerate and there exist tricks to make them positive definite such as adding εI_n to the dataset. In this case, the model should be the manifold of symmetric positive definite (SPD) matrices with a Riemannian metric that rejects rank-deficient matrices to infinity. However, in other contexts, rank-deficient matrices should be kept at finite distance because they correspond to a physical reality. It occurs when covariance matrices are of the type $\Sigma = XX^\top$ with $X \in \mathbb{R}^{n \times k}$ with $n > k$, i.e. empirical covariance matrices between many features with few data points. So the treatment of singularities depends on the application. Another example of singularity on SPD matrices is given by matrices with two equal eigenvalues. For instance, in diffusion tensor imaging (DTI), they may indicate the crossing of two fibers. Therefore, considering this space as stratified by

eigenvalue multiplicity is more suited to this kind of problem [Groisser et al., 2017].

Another important element is the notion of invariance under the action of a group of transformations or under a set of symmetries of the space. Being invariant under a transformation means that if it is applied jointly to all the dataset, then the results of geometric operations such as the Fréchet mean will be transformed accordingly. Therefore, if the inverse transformation is applied at the end, the result will be the same as the one without any transformation so the statistical analysis will be unchanged. Let us take the example of permuting the nodes in an undirected graph with vertices $\{A, B, C\}$ for three types of electronic device which exchange information. The attribute of an edge is the quantity of information exchanged between the two types of device, allowing self loops. The adjacency matrix of the graphs is built by associating A, B, C to 1, 2, 3 respectively. A distance on adjacency matrices with non-negative coefficients $M \in \text{Sym}^{\geq 0}(3)$ is invariant under permutations if and only if the coefficients M_{11}, M_{22}, M_{33} (resp. M_{12}, M_{23}, M_{31}) are treated equally. For example, $d(M, M') = \sum_i |M_{ii} - M'_{ii}| + 3 \sum_{i \neq j} |M_{ij} - M'_{ij}|$ is permutation-invariant but $d'(M, M') = \sum_{i,j} i |M_{ij} - M'_{ij}|$ is not. Moreover, the choice of associating A to 1, B to 2 and C to 3 is arbitrary so we probably want the statistical analysis to be independent from this choice. That is exactly what a permutation-invariant distance allows to do. If the same permutation is applied to the whole dataset, then the Fréchet mean will be affected by the same permutation so we will be able to retrieve the information on A, B, C respectively since we know how they are associated to 1, 2, 3. Thus, a permutation-invariant distance preserves the difference between A, B and C, despite what one could think.

Invariance is not independence. Being independent from a transformation is much stronger: it means that if a set of different transformations is applied independently to each data within the dataset, then the result of the geometric operations will not even be affected. It means that the geometry has to be defined on the quotient of the space by this group of transformations. Here, the distance does not distinguish anymore between A, B and C. Actually, the graph itself is defined up to permutation, the vertices have no label here. For example, the Graph space is defined as the quotient of graphs by the permutation of nodes [Calissano et al., 2020].

In addition to singularities, invariance or independence under natural transformations of the space, another type of hypothesis consists in decoupling variables and turning the space into a Cartesian product endowed with a product metric or a product distance. This can come from the observation that two variables seem to be unrelated. This can also be a refinement of the assumption of independence under one of the two variables. Indeed, the respective weights in the product metric can give more or less importance to each variable.

In general, these assumptions lead to families of admissible structures rather than one unique structure. Thus, we need other principles to discriminate between them.

1.1.5.2 Second principle: good theoretical properties

We saw that for metric spaces, some hypotheses such as completeness, non-positive curvature or quotient structure, lead to important theoretical results of existence and uniqueness. We also argued that Riemannian manifolds convey many more geometric tools and results which allow to generically perform some statistical operations. Let us specify some particular types of Riemannian manifolds where there are even more tools and results to use.

When a manifold \mathcal{M} is embedded in a Riemannian manifold (\mathcal{M}^0, g^0) , the ambient Riemannian metric g^0 induces a Riemannian metric g on the submanifold. Gauss-Codazzi equations provide the Levi-Civita connection and the curvature in (\mathcal{M}, g) in function of those in the embedding space (\mathcal{M}^0, g^0) .

Another typical situation is when a submersion $\pi : \mathcal{M}^0 \rightarrow \mathcal{M}$ is given. The submersion π and the metric g^0 allow to define a vertical distribution $\mathcal{V} = \ker d\pi$, a horizontal distribution $\mathcal{H} = \mathcal{V}^\perp$, a horizontal lift $\#_x = (d_x\pi)|_{\mathcal{H}_x}^{-1} : T_{\pi(x)}\mathcal{M} \rightarrow \mathcal{H}_x$ at each point $x \in \mathcal{M}^0$, and finally an induced metric $g_{\pi(x)} = g_x^0 \circ (\#_x, \#_x)$ on \mathcal{M} . The map $\pi : (\mathcal{M}^0, g^0) \rightarrow (\mathcal{M}, g)$ is called a Riemannian submersion and O'Neill's equations provide the Levi-Civita connection, the curvature and the exponential map in (\mathcal{M}, g) in function of those in the total space (\mathcal{M}^0, g^0) [O'Neill, 1966]. More precisely, the geodesics in \mathcal{M} are the projections of horizontal geodesics in \mathcal{M}^0 so the exponential map writes $\text{Exp}_{\pi(x)} = \pi \circ \text{Exp}_x^0 \circ \#_x$. In particular, if (\mathcal{M}^0, g^0) is geodesically complete, then (\mathcal{M}, g) is geodesically complete.

Many particular cases of Riemannian submersions can be met such as Riemannian quotient manifolds, Riemannian homogeneous manifolds, naturally reductive homogeneous manifolds or Riemannian symmetric spaces. Each of these classes is included in the previous one and brings more properties to the space. They are summarized along with the embedded spaces in Table 1.2.

Riemannian manifold	Theoretical properties
Embedded	Levi-Civita connection, curvature (if known in embedding space)
Submersed/quotient	+ exponential map (if known in the total space)
Homogeneous	+ completeness, transitive isometric group action
Naturally reductive	+ geodesics at identity are one-parameter subgroups
Symmetric	+ $\nabla R = 0$, parallel transport is a composition of two symmetries
Euclidean	+ everything in closed form

Table 1.2: Theoretical properties of Riemannian manifolds implying practical properties.

Apart from Lie group actions that are compatible with the Riemannian metric, Riemannian manifolds may also carry other compatible structures such as a pair of dual connections or a divergence, or it may be seen as the parameter space of a statistical model. This is well studied in information geometry [Amari and Nagaoka, 2000, Amari, 2016, Nielsen, 2022]. Two connections ∇, ∇^* are dual with respect to the metric g if they satisfy $X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z)$ for all vector fields X, Y, Z . A divergence is a distance-like function $D : \mathcal{M}^2 \rightarrow [0, +\infty)$ satisfying separation and which induces by differentiation a dualistic structure (g, ∇, ∇^*) , that is a Riemannian metric g with dual connections ∇, ∇^* [Amari and Nagaoka, 2000].

A statistical model, or family of probability densities, induces by differentiation a Riemannian metric called the Fisher-Rao metric g . Then, the natural gradient is defined as the gradient associated to the Fisher-Rao metric. A statistical model also induces a pair of dual connections $(\nabla^{(1)}, \nabla^{(-1)})$, or more generally the one-parameter family of Amari-Chentsov α -connections $(\nabla^{(\alpha)})_{\alpha \in \mathbb{R}}$ [Amari and Nagaoka, 2000]. The dualistic structure $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ can also be viewed as emanating from a divergence. If the statistical model is an exponential family or a mixture family, then $\nabla^{(1)}$ and $\nabla^{(-1)}$ are flat [Amari and Nagaoka, 2000].

Information-geometric tools are used in many algorithms for parameter estimation [Amari, 1995], computing centroids [Nielsen and Nock, 2009], clustering [Liu et al., 2012] or function optimization [Ollivier et al., 2017], to cite a few.

To summarize, the Riemannian geometry can be enriched with other compatible geometric structures that provide more tools to compute on the manifold.

1.1.5.3 Third principle: good practical properties

Our third principle is the ability of the geometry to be implemented efficiently, for example with closed-form formulae or fast algorithms to compute the geometric operations such as the Fréchet mean, the geodesics or the parallel transport. As we argued earlier, this ability is often a consequence of a rich geometry. As shown in Table 1.2, the simpler is the richer. Note that it does not mean that the simplest geometry is the best suited for the application at hand. Application constraints and practical constraints might be in conflict so one might need to find a trade-off between them. This supports the idea of defining families of Riemannian metrics encompassing different geometries to be able to go continuously from one to the other.

This principle also pushes us to give a particular attention to numerical questions. The python package `geomstats` [Miolane et al., 2020a, Miolane et al., 2020b] aims at offering generic implementations for certain classes of manifolds such as Riemannian quotient spaces, Lie groups, or more generally Riemannian manifolds and manifolds equipped with an affine connection. For example, the exponential map of an affine connection is computed by integrating the Hamiltonian flow. These generic methods can be overridden by specific methods, for example when one has closed-form formulae or more efficient algorithms. This encourages us to push the theoretical computations as far as possible because we often get formulae that can be computed faster and with more precision than the generic implementations. The geometric operations being the building blocks of more complex algorithms, it is crucial to simplify their implementation when it is possible.

1.2 Spaces of covariance and correlation matrices

We now focus on the spaces studied in this thesis: the spaces of covariance and correlation matrices.

1.2.1 Two stratified spaces

In many applications, data are modelled by covariance matrices or correlation matrices for example between signals, physical quantities or within a time window (auto-correlation). The application domains are electro-encephalography (EEG) [Barachant et al., 2012, Barachant et al., 2013], magneto-encephalography (MEG), functional Magnetic Resonance Imaging (fMRI) [Marrelec et al., 2006, Varoquaux et al., 2010], diffusion tensor imaging (DTI) [Pennec et al., 2006, Lenglet et al., 2006, Fletcher and Joshi, 2007, Moakher, 2005, Batchelor et al., 2005], trees and graphs [Journée et al., 2010, Severn et al., 2019, Garba et al., 2021], signal processing, radar [Barbaresco, 2013], computer vision, genomics [de la Fuente et al.,

2004, Peng et al., 2009], etc. Covariance matrices are symmetric positive semi-definite matrices. They don't form a manifold but a stratified space whose strata are the manifolds of symmetric positive semi-definite matrices of fixed rank. Since these manifolds are connected, the Hasse diagram of adjacency is trivial. The principal stratum is the manifold of Symmetric Positive Definite (SPD) matrices, it is dense in the set of covariance matrices. Covariance matrices form a convex cone in the vector space of symmetric matrices.

Correlation matrices are covariance matrices with unit diagonal. The correlation coefficient between two variables is the quotient of their covariance by the product of their standard deviations: $\text{Cor}(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Cov}(X_i, X_i)}\sqrt{\text{Cov}(X_j, X_j)}}$. Thus correlation matrices $C = \text{Cor}(X)$ are built from covariance matrices $\Sigma = \text{Cov}(X)$ with non-zero variances by dividing each entry by the product of square-roots of the diagonal elements on its row and its column: $C_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}}\sqrt{\Sigma_{jj}}}$ or $C = \text{Diag}(\Sigma)^{-1/2} \Sigma \text{Diag}(\Sigma)^{-1/2}$. Correlation matrices form a convex space called the ellipsope. They admit a similar stratified structure where the strata are also the manifolds of correlation matrices of fixed rank. In terms of strata, a first difference with covariance matrices is that the null matrix is not a correlation matrix. A second difference is that the correlation matrices of rank one form a discrete manifold corresponding to the corners of the ellipsope [Tropp, 2018].

Back to the difference explained earlier between invariance and independence, let us consider the action of the Lie group of positive diagonal matrices on SPD matrices. We can decompose the space of SPD matrices into the Cartesian product of positive diagonal matrices and full-rank correlation matrices. Therefore, the action of the positive diagonal group only acts on the diagonal component. If an experiment with covariance matrices as data is very weakly sensitive to the scales of variables, it means that the correct model to use should be correlation matrices instead of covariance matrices.

To study the geometry of these stratified spaces, it seems natural to proceed step by step by studying firstly the principal stratum, secondly the other strata, thirdly the whole space. Since the principal stratum is dense in the whole space, it is worth spending some time studying it because it is a good approximation of the whole space. In other words, random generic covariance matrices are almost surely SPD. However, covariance matrices may not be generic in many practical problems. For example, when empirical covariance matrices are computed with less data than the number of features, they are mandatorily degenerate and the previous statement fails. That is precisely why it is also important to have theoretical tools on the boundary.

1.2.2 The many geometries of SPD and PSD matrices

Many Riemannian metrics were proposed on the manifold of SPD matrices, emanating from different contexts from algebra to information geometry or quantum information geometry: affine-invariant [Siegel, 1943, Skovgaard, 1984, Pennec et al., 2006, Lenglet et al., 2006, Fletcher and Joshi, 2007, Moakher, 2005], log-Euclidean [Arsigny et al., 2006, Fillard et al., 2007, Hà Quang et al., 2014], Bures-Wasserstein [Dowson and Landau, 1982, Olkin and Pukelsheim, 1982, Takatsu, 2010, Takatsu, 2011, Malagò et al., 2018, Bhatia et al., 2019], Bogoliubov-Kubo-Mori [Petz and Toth, 1993, Michor et al., 2000], power-Euclidean [Hiai and Petz, 2009, Dryden et al., 2010], alpha-Procrustes [Hà Quang, 2022], power-affine

[Thanwerdas and Pennec, 2019b], Cholesky [Wang et al., 2004], log-Euclidean-Cholesky [Li et al., 2017] log-Cholesky [Pinheiro and Bates, 1996, Lin, 2019], etc. We generically call them the noted metrics. Let us follow the general method we sketched to choose a Riemannian metric. On the manifold of SPD matrices, the singularities are the degenerate matrices, those which have at least one null eigenvalue and which are located on the boundary. Therefore, depending on the problem, we may either want these degenerate matrices to be at finite or infinite distance of SPD matrices.

One natural transformation of the space of SPD matrices is given by the invertible affine transformation of the feature vector $X \mapsto AX + X_0$ with $A \in GL(n)$. It is well known that the new covariance matrix is congruent to the original one via the linear part A of the transformation: $\text{Cov}(AX + X_0) = A \text{Cov}(X) A^\top$. Therefore, choosing an affine-invariant metric means that the geometry (and the associated geometric operations) will not change if we apply jointly an affine transformation to the *whole set* of data. For example, in brain-computer interfaces, we may assume that between two different sessions, the electrode cap is not positioned at the same location and that this induces a linear transformation in the signals acquired at first order. In contrast, within a session, we may assume that the transformation remains the same for all the experiments. Then, under this hypothesis, an affine-invariant metric seems to be a good choice to analyze experiments within a session but there should be an alignment step or a change of reference across different sessions [Barachant et al., 2013].

Furthermore, all the subgroups of the general linear group define a notion of invariance that may be of interest or not depending on the application: orthogonal transformations (orthogonal group), independent positive scalings of variables (positive diagonal group), permutations between variables (permutation group) or global positive scaling (positive real group) for instance. All the metrics cited above are invariant under orthogonal transformations, except those named after Cholesky. Their quantity suggests to propose other theoretical properties to classify and distinguish them.

On singular PSD matrices of fixed rank, it seems that four Riemannian metrics were proposed [Bonnabel and Sepulchre, 2010, Vandereycken et al., 2009, Vandereycken et al., 2013, Massart and Absil, 2020]. These spaces are a bit more difficult to handle since they are not open sets of vector spaces. On PSD matrices, the invariance under affine transformations or even under the positive diagonal group is impossible to get for any distance [Bonnabel and Sepulchre, 2010]. However, it might be quite difficult to glue these Riemannian manifolds together to form a global interesting geometry on covariance matrices of any rank. Therefore, it appears that it is important to think to a geometry on the whole space *before* studying it on each stratum separately.

The only distance defined on the whole space of covariance matrices that we are aware of is the Bures-Wasserstein distance [Dowson and Landau, 1982, Olkin and Pukelsheim, 1982]. The stratified topology was described in [Takatsu, 2011] as well as the tangent cone (which replaces the notion of tangent space at singular points). It was also shown to be an Aleksandrov space of non-negative curvature. There seems to be no other work on stratified geometries of covariance matrices so far.

1.2.3 The few geometries of correlation matrices

In contrast, very few Riemannian metrics or more general geometric structures were proposed in the literature on the manifold of correlation matrices. The oldest geometry seems to be the quotient by the orthogonal group of a product of n spheres \mathbb{S}^{n-1} of dimension $n - 1$. It was used to simulate correlation matrices in [Rebonato and Jaeckel, 2001]. The quotient topology was later described in [Kercheval, 2008]. However, a complete description of the geometry does not exist neither on strata nor on the whole space of correlation matrices so far. Cholesky parametrizations were used to compute the nearest low-rank correlation matrix [Grubišić and Pietersz, 2007].

On full-rank correlation matrices, we can cite the Hilbert projective geometry, which is a distance on any open bounded convex subset of \mathbb{R}^n . It was applied to the open elliptope of full-rank correlation matrices for clustering [Nielsen and Sun, 2019]. A Riemannian metric was proposed by taking the quotient of the affine-invariant metric on SPD matrices by the action of positive diagonal matrices [David and Gu, 2019, David, 2019]. More recently, a Euclidean bijective parametrization of the open elliptope was proposed via the matrix exponential [Archakov and Hansen, 2021].

The same remark as for SPD matrices can be made about singularities. If one wants to avoid singular matrices, then the open elliptope should be considered without its boundary. Since it is an open set of an affine space, the simplest Riemannian geometry should be diffeomorphic to a Euclidean geometry. Therefore, one should find natural diffeomorphisms between the open elliptope and Euclidean spaces. The recent bijective parametrization of [Archakov and Hansen, 2021] could be a nice tool to do so.

About the natural group transformations of the space, the situation is very different than the one of SPD matrices. Indeed, the affine transformation of the feature vector does not induce a group action on correlation matrices. Correlation matrices were often treated as covariance matrices in the applications. However, their structure is completely different. Indeed, among the groups cited above for covariance matrices, only the permutation group stabilizes the manifold of full-rank correlation matrices. Following the discussion on invariance and independence, it seems natural to use permutation-invariant Riemannian metrics when the features of the feature vector should be treated equally, for example for electrodes, brain regions, cells or genes.

One could also require the weaker invariance by groups of features, that is by block permutation matrices with fixed sizes of blocks. However, the invariance under permutation might not be relevant for features such as times in auto-correlation matrices because contrarily to the previous cases, the order between times is probably a key element of the analysis.

1.3 Comprehensive presentation of the contributions

1.3.1 Purposes and methodology

Our purposes follow the analysis we made of theoretical and practical tools available on covariance and correlation matrices. On SPD matrices, there is a need of classification to understand the common points and the differences between all the Riemannian metrics proposed in the literature. In particular, the invariance under orthogonal transformations

seems to be a key classification criterion to study. In other words, the identified need is essentially theoretical since there exist already many tools to compute with SPD matrices.

On the contrary, there is an obvious lack of intrinsic and efficient tools even to interpolate between two full-rank correlation matrices. Our goal is to define several types of geometries with different properties so that people have the choice in function of their application. Furthermore, proposing families of metrics rather than isolated metrics is a leitmotiv to offer more flexibility and allow optimizing the parameters of the family under other constraints for example. In practice, the goal is to get simple formulae of geodesics or parallel transport to perform the basic geometric and statistical operations on the manifold of full-rank correlation matrices.

About the stratified spaces of covariance and correlation matrices, we aim at proposing a first example of stratified structure with a Riemannian metric on each stratum and a global distance between any two points of the space. This structural description should be enriched by some practical tools such as the geodesics between any two points.

Our methodology is the following. From the literature on applications using covariance and correlation matrices, we extract some assumptions that may be relevant in different contexts such as invariances, symmetries, singularities, independent variables. Then, we abstract these considerations and enrich them with other theoretical constraints on the algebraic or topological structure of the space, on the groups acting on it, on the relevant functions that were used in other domains for example. After that, we try to define new structures by manipulating all these concepts together. A crucial element of our methodology is to push the computations as far as they can be. We noticed that in several works, calculations were left with implicit formulae that are difficult to grasp and implement. Moreover from a computational point of view, a closed-form formula often offers faster and more precise results than a composition of generic functions that may propagate numerical errors. Following these guidelines, we got several results from theorems of classification, characterization or stability to practical formulae (geodesics, curvature, parallel transport, etc.) or algorithms. The package `geomstats` offers a place to gather all these tools on manifolds. Our methodology is schematized on Figure 1.3.

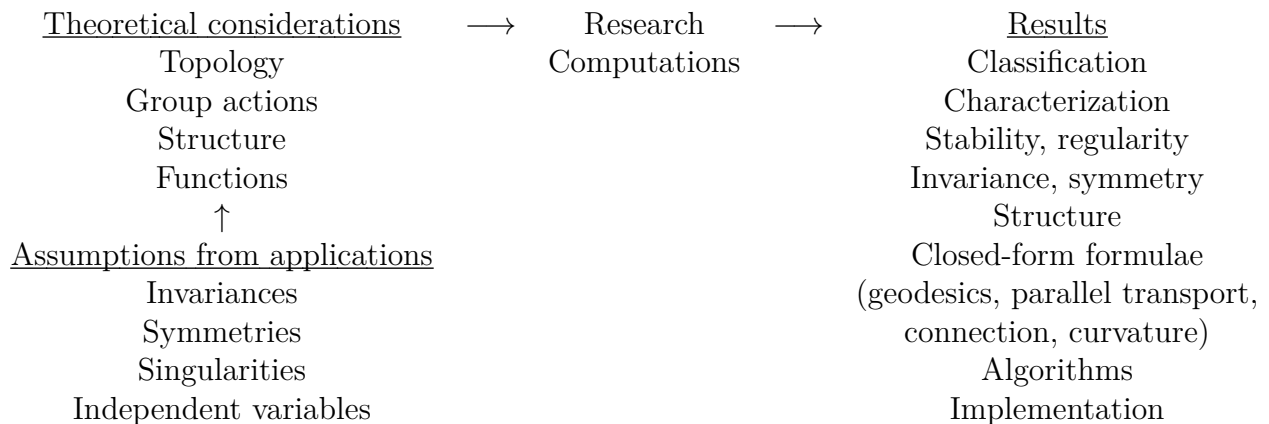


Figure 1.3: The methodology followed in this thesis.

1.3.2 Classification of Riemannian metrics on SPD matrices

In order to classify Riemannian metrics on SPD matrices, we first characterize the continuous $O(n)$ -invariant Riemannian metrics by means of three continuous multivariate functions $\alpha, \beta, \gamma : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfying three conditions of symmetry, compatibility and positivity (Theorem 4.20). This shows that the class of $O(n)$ -invariant metrics is much wider than the class of affine-invariant metrics, which is indexed by only two real parameters. Secondly, in light of this characterization, we reinterpret the classes of kernel and mean kernel metrics introduced in [Hiai and Petz, 2009, Hiai and Petz, 2012] and we propose intermediate classes of metrics with stability properties. The smaller the class is, the more properties are shared by all the metrics of the class. So we propose a sequence of nested classes defined by conditions on the three functions α, β, γ ranging from the noted metrics (affine-invariant, log-Euclidean, Bures-Wasserstein, BKM, etc.) to the whole class of $O(n)$ -invariant metrics. Notable classes are the families of Mixed-Power-Euclidean and Mixed-Euclidean metrics (Section 5.5) which encompass with few parameters most of the noted metrics. This is represented on Figure 1.4 and on Figure 4.1 in Chapter 4. Thirdly, we also show that affine-invariant metrics are characterized by the pairs of invariance under \mathbb{R}^+ and $SL(n)$, or $\text{Diag}^+(n)$ and $O(n)$, and less evidently under $LT^+(n)$ (lower triangular matrices with positive diagonal) and $\mathfrak{S}(n)$ (permutation matrices). This is shown on Figure 1.5. We characterize $LT^+(n)$ -invariant metrics as pullbacks by the Cholesky map of left-invariant metrics on the Lie group $LT^+(n)$ (Theorem 7.4). This gives a Lie group structure to the manifold of SPD matrices for which the affine-invariant metrics are left-invariant metrics. This is an unexpected result since this Riemannian manifold has only been considered so far as a Riemannian symmetric space.

To prove the continuity of the $O(n)$ -invariant metric defined by the three continuous maps α, β, γ , we were brought to answer the following question: given two symmetric matrices, how can we choose their matrices of eigenvectors so that they are the closest possible? Indeed, the matrix of eigenvectors is not unique. We can reduce the degrees of freedom by imposing the order of the eigenvalues but we are left with rotations and symmetries in the eigenspaces. We answer this question in Chapter 2 as a mathematical preliminary.

These theoretical results came with additional practical results. We provide the complete formula of the sectional curvature of the affine-invariant metrics in Table 4.5 (one term was forgotten in [Skovgaard, 1984]). We give new formulae for the Bures-Wasserstein metric, especially an explicit equation for the parallel transport which allowed to implement it in `geomstats` (Table 4.7 and Proposition 4.12). We derived the curvature of the Mixed-Euclidean metrics (Theorem 5.25) based on the derivation of the BKM curvature in [Michor et al., 2000]. In addition, we prove that all these curvatures take negative values and we show numerically that all the Mixed-Power-Euclidean metrics but the Euclidean, log-Euclidean and affine-invariant metrics also have positive curvature. We show that the necessary and sufficient condition for geodesic completeness on mean kernel metrics [Hiai and Petz, 2009] is still valid for the wider class of extended mean kernel metrics. We also show that the classes of kernel metrics and extended kernel metrics are stable by taking the cometric (if we identify the vector space of symmetric matrices and its dual via the Frobenius inner product). This is an important advantage to compute the geodesics by integration of the Hamiltonian equations of geodesics since they require to compute the cometric. These results can be found in Chapters 4 and 5 in Part III.

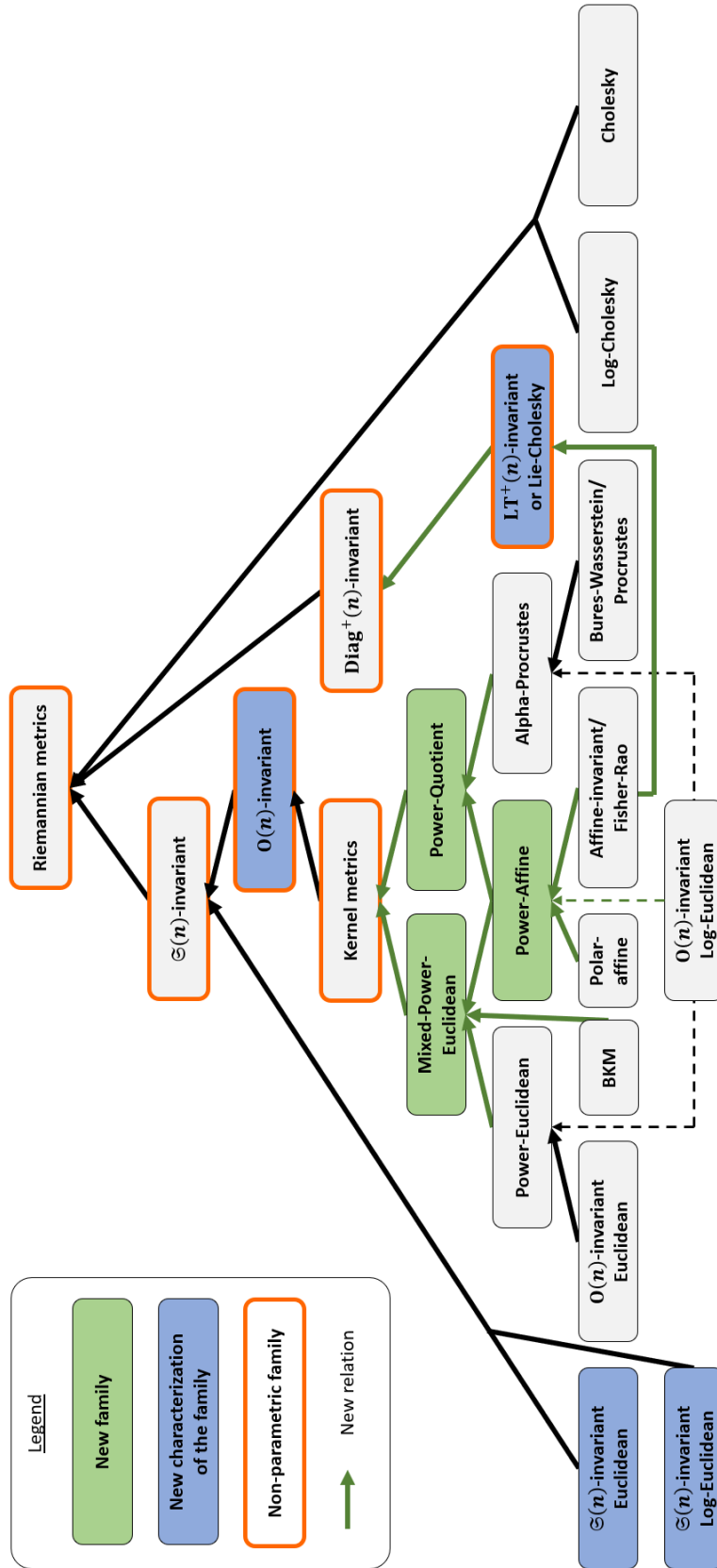


Figure 1.4: Elements of classification of Riemannian metrics on SPD matrices.

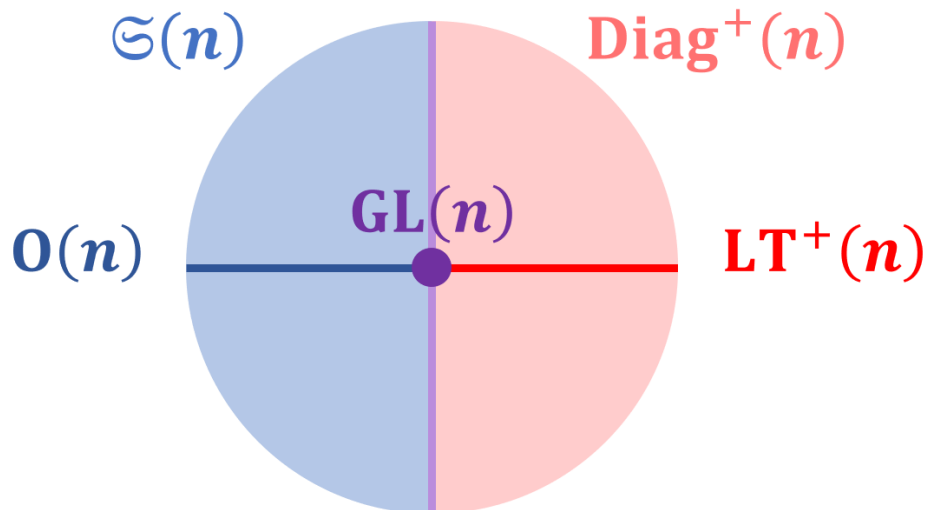


Figure 1.5: Invariance under subgroups of $GL(n)$. Each color represents a class of Riemannian metrics that are invariant under the associated group. Affine-invariant metrics belong to all other classes. Some pairs of other invariances ($O(n)$ and $Diag(n)$; $\mathfrak{S}(n)$ and $LT^+(n)$) imply affine-invariance. This is explained in Chapter 7.

1.3.3 New families of Hadamard metrics on full-rank correlation matrices

We started by studying the recently introduced quotient-affine metric on full-rank correlation matrices [David and Gu, 2019, David, 2019]. It relies on the quotient space structure of the open elliptope. Indeed, the orbits of the action of positive diagonal matrices on SPD matrices are in bijection with correlation matrices so it suffices to find a metric invariant under the positive diagonal group to define a metric on the open elliptope, for example the affine-invariant metric [David and Gu, 2019]. Thanks to O’Neill’s fundamental equations of Riemannian submersions, we compute the quotient geometry, the Riemannian metric, the geodesics, the Levi-Civita connection and the curvature in Chapter 6. We also show that Riemannian logarithms exist but the uniqueness remains an open question as well as closed-form formulae. Furthermore, we show that the curvature is not of constant sign and that it is unbounded from above (Theorem 7.1). This argues for the definition of other metrics with a better behavior, for example Hadamard or even flat since the open elliptope is diffeomorphic to a Euclidean space. We show that the construction of quotient-affine metrics generalize to $LT^+(n)$ -invariant metrics (Section 7.3), which unfortunately suffer the same drawbacks. However, this generalization puts in light the interesting idea of transforming correlation matrices by the Cholesky map. This was done for SPD matrices but not precisely for full-rank correlation matrices. This allows us to define very convenient geometries by pullback via the Cholesky map in Section 7.4: one Riemannian symmetric space called poly-hyperbolic-Cholesky, two Euclidean spaces called Euclidean-Cholesky and log-Euclidean-Cholesky, and one Lie group structure called Lie-Cholesky. All the operations associated to the Riemannian metrics are given in closed form. Of course, each metric can be extended into a family with the same properties.

Although these Riemannian metrics are very practical, they miss an ingredient that could be useful in some applications: the invariance by permutations. Therefore, we propose two permutation-invariant metrics defined completely differently in Chapter 8. The first one relies on a bijective parametrization recently proposed thanks to the matrix exponential map [Archakov and Hansen, 2021]. The authors proved that given a symmetric matrix, there exists a unique diagonal matrix such that the exponential of their sum is a correlation matrix. This defines a bijection between symmetric hollow (with null diagonal) matrices and full-rank correlation matrices. We prove that this bijection is a smooth diffeomorphism (Theorem 8.6), therefore it allows to push permutation-invariant inner products forward to the open ellipsope. We call these Riemannian metrics off-log metrics and, as a log-Euclidean metric, all the geometric operations are known in closed form modulo the computation of the bijection. It is proved in [Archakov and Hansen, 2021] that the bijection can be computed in logarithmic time complexity.

To find all the permutation-invariant inner products on the vector space of symmetric hollow matrices, we were brought to answer the more general question of finding all the G -invariant inner products on a vector space, knowing one of them. Although this seems to be a well-known result in representation theory, we were not able to find this result in reference books. Therefore, we answer this question and we present some examples in Chapter 3 as another mathematical preliminary.

In parallel, we explored an analogous idea which is somehow the symmetric or the dual of the previous one. We showed that given a correlation matrix there exists a unique positive diagonal matrix such that the logarithm of their composition by congruence belongs to a vector space, namely the vector space of symmetric matrices with null row sums (Theorem 8.15). This defines a bijection between full-rank correlation matrices and symmetric matrices with null row sums. Analogously, we show that it is a diffeomorphism (Theorem 8.19) and thus, permutation-invariant inner products can be pulled back to the ellipsope. We call these pullback metrics log-scaled metrics because our result is based on the well-known concept of “scaling” of an SPD matrix [Marshall and Olkin, 1968]. These metrics have an additional characteristic that only the quotient-affine metric has: they are inverse-consistent (Theorem 8.18). Correlation matrices are not stable by inversion but the correlation matrix of the inverse is of course a correlation matrix. This leads to a map that we call the cor-inversion and that is closely related to the notion of partial correlation, well known in signal processing or Gaussian graphical networks. Indeed, the opposite off-diagonal coefficients of the cor-inversion are exactly the partial correlation coefficients (or parcors). This important map defines notions of intrinsic inversion of correlation matrices and inverse-consistency of Riemannian metrics, that are satisfied by the log-scaled metrics. Quotient-affine metrics and log-scaled metrics intrinsically satisfy this property because their construction relies on the action of positive diagonal matrices. The properties of all these metrics are summarized on Table 1.3.

All the geometric operations of log-scaled metrics are known in closed form modulo the computation of the bijection, that is the computation of the scaling of an SPD matrix. Many algorithms exist to compute the scaling. However, most of them are not specific to SPD matrices. Surprisingly, the specific algorithms seem not to rely on the proof of existence of the scaling [Marshall and Olkin, 1968] although it involves the minimization of a strictly convex map. We show that the optimization problem amounts to minimize this even strongly

Reference	Riemannian metric	Comp.	Structure	Curvature	$\mathfrak{S}(n)$ -inv.	Inv.-cons.
[D19] & Ch. 6	Quotient-affine	Yes	Quotient space	$[-\frac{1}{2}, +\infty)$	Yes	Yes
Ch. 7	Quotient-Lie-Chol.		Quotient space	?	Yes	No
Ch. 7	Poly-hyperbolic-Chol.		Symmetric space	≤ 0	No	No
Ch. 7	Euclidean-Chol.		Vector space	Flat	No	No
Ch. 7	Log-Euclidean-Chol.		Vector space	Flat	No	No
[A21] & Ch. 8	Off-log		Vector space	Flat	Yes	No
Ch. 8	Log-scaled		Vector space	Flat	Yes	Yes

Table 1.3: Properties of Riemannian metrics on the open elliptope of full-rank correlation matrices. Abbreviations: [D19] is [David and Gu, 2019], [A21] is [Archakov and Hansen, 2021], Ch. is Chapter, Comp. is Complete, $\mathfrak{S}(n)$ -inv. is $\mathfrak{S}(n)$ -invariant, Inv.-cons. is Inverse-consistent.

convex map on a convex set. Thus we propose in Section 8.5 a method relying on the well-known Projected Gradient Descent (PGD) algorithm. Since the projection is not known in closed form, we contribute a simple algorithm to compute the projection that we plug in the PGD algorithm. We prove the convergence of our approximated PGD algorithm in logarithmic time and we argue that its total complexity and precision is competitive with respect to SPD-specific and unspecific algorithms.

Finally, we show that all these metrics split in two groups in dimension 2, that is with only one correlation coefficient (Theorem 8.23). This provides two different interpolations of the correlation coefficient, including one closely related to the Fisher z -transformation. These results are presented in Chapters 6, 7 and 8 in Part IV.

1.3.4 Bures-Wasserstein geodesics in the orbit space of covariance matrices

Before studying the geometry of stratified spaces, it is important to notice that it seems unrealistic to build a distance on the whole space that coincides with a complete Riemannian distance on the principal stratum. Indeed, the boundary is considered as a singularity rejected infinitely far so these viewpoints seem irreconcilable. Therefore, we are now more interested in non-complete metrics on SPD matrices that should be completed thanks to the stratified boundary of singular matrices. Many Riemannian metrics on SPD matrices such as the Euclidean or the power-Euclidean metrics have geodesics that leave the space of covariance matrices. We know only one metric whose geodesics bounce on the boundary and come back into SPD matrices: this is the Bures-Wasserstein metric. Thus we study the same metric stratified space as in [Takatsu, 2011].

We propose to apply the richer framework of Riemannian orbit spaces [Alekseevsky et al., 2001] to the quotient of the Euclidean space of square matrices equipped with the Frobenius inner product by the right action of the orthogonal group. With this viewpoint, covariance matrices form a stratified space whose strata are the manifolds of covariance matrices of fixed rank, as explained in Section 9.3. It is well known that the quotient distance is exactly the Bures-Wasserstein distance. The geodesics were studied in the principal stratum of SPD

matrices [Malagò et al., 2018, Bhatia et al., 2019] and in other strata [Massart and Absil, 2020]. However, we noticed that several questions were still open on the injectivity domain of the exponential map, the set of preimages or the set of logarithms. We answer precisely all these questions in Sections 9.4 and 9.5. We also compute the horizontal lift, which allows us to give explicit formulae of the Riemannian metric and the exponential map.

Furthermore, we characterize precisely the minimizing geodesics of the total metric space between any two covariance matrices (Theorem 9.31). We prove that they are of constant rank on the interior of the segment. We show that the minimizing geodesic segments between two covariance matrices Σ and Λ are parametrized by the closed unit ball or $\mathbb{R}^{(k-r) \times (l-r)}$ for the spectral norm, where k, l, r are the respective ranks of $\Sigma, \Lambda, \Sigma\Lambda$ (Theorem 9.34). We give their expression in function of this parameter. In particular, we show that there is either exactly one or an infinity of geodesics between two covariance matrices, depending on the rank of their product. We also give the number of geodesics of minimal rank and we give a simple formula of the minimizing geodesic when it is unique. Finally, we show that this formula actually defines a minimizing geodesic between any two matrices so we call it the canonical geodesic. These results are presented in Chapter 9.

1.4 Organization of the thesis

1.4.1 Overview

This thesis is organized in parts divided in chapters. In Part II, we give two mathematical preliminaries. They result from the resolution of two problems that came naturally throughout our research and that revealed to be non-trivial. We put them aside to ensure subject homogeneity within chapters. Chapter 2 tackles the problem of regularity of eigenvalues and eigenvectors of symmetric matrices. It allows us to show the continuity of a Riemannian metric defined via eigenvalue decomposition in Chapter 4. Chapter 3 answers the question of finding all G -invariant inner products on a vector space, based on representation theory. We use this result to build families of permutation-invariant metrics in Chapter 8.

Part III focuses on the geometries of the open cone of SPD matrices, that is full-rank covariance matrices. Since many geometries were previously defined on this manifold, we survey the literature in Chapter 4 and we complete it wherever needed with new formulae. Since these geometries were essentially $O(n)$ -invariant, we characterize this general family and we show that it is much wider than the family of affine-invariant metrics. We promote the idea of defining families of Riemannian metrics encompassing existing ones and having nice stability properties. Thus, we rely on the previous work on kernel metrics of [Hiai and Petz, 2009] to define new subclasses and super-classes of metrics. In particular in Chapter 5, we explore the information geometry and the Riemannian geometry of the families of Mixed-Power-Euclidean and Mixed-Euclidean metrics, which encompass many noted metrics with few parameters.

We identified in the introduction the need to define geometric tools on correlation matrices for intrinsic computation and also to endow covariance matrices with product metrics. Therefore in Part IV, we define new geometries on the open ellipsope of full-rank correlation matrices. After computing the Riemannian operations of the recently introduced quotient-

affine metric [David and Gu, 2019] in Chapter 6 and generalizing it to quotient-Lie-Cholesky metrics, we show that their geometry is quite complex. Since the manifold is diffeomorphic to $\mathbb{R}^{n(n-1)/2}$, we argue that there should exist simpler Hadamard or flat metrics on this manifold. Therefore, we define three new geometries satisfying these requirements in Chapter 7 and two flat geometries that are additionally invariant under permutations in Chapter 8. The geometric operations of the former metrics can be computed in closed form modulo Cholesky decomposition, as well as the geometric operations of the latter metrics modulo the computation in logarithmic time complexity of the diffeomorphisms towards the Euclidean spaces.

Beyond Riemannian geometry, we study the Bures-Wasserstein stratified geometry of covariance matrices in Part V. We complete the work of [Takatsu, 2011] by viewing this space as the Riemannian orbit space [Alekseevsky et al., 2001] of square matrices quotiented by the right action of the orthogonal group. After giving complements on the geometry of each stratum following [Bhatia et al., 2019] and [Massart and Absil, 2020], we characterize all the minimizing geodesic segments between two covariance matrices of any rank.

We conclude the thesis in Part VI and we identify some perspectives for future works on correlation matrices, stratification by eigenvalue multiplicity or spaces of matrices of different sizes.

Part VII is an appendix composed of Chapter 11, which gathers all the proofs of the results of the thesis. We decided to defer to this chapter almost all the proofs for the sake of brevity and readability of the other chapters. Only the very short proofs or the ones that we considered as essential to understand the chapter were kept in their original place. Chapter 11 is self-contained, all the results are recalled before their proofs. For readers interested in reading the proof immediately, it suffices to click on “ Y ” in the sentence “*See the proof of Result X in Section Y .*” below the result. Then it suffices to click on the number “ X ” of the result to come back to the result within the concerned chapter.

1.4.2 Summary of contributions by chapter

Part II. Mathematical preliminaries

Chapter 2. Regularity of eigenvalues and eigenvectors. On the set of symmetric matrices with distinct eigenvalues $\text{Sym}^{\neq}(n)$, we prove that the function of ordered eigenvalues $val : \text{Sym}^{\neq}(n) \rightarrow \text{Diag}(n)$ is smooth and that there exist local smooth maps $vec : \mathcal{U} \subset \text{Sym}^{\neq}(n) \rightarrow \text{O}(n)$ satisfying $vec(\Sigma)val(\Sigma)vec(\Sigma)^{\top} = \Sigma$ for all $\Sigma \in \text{Sym}^{\neq}(n)$ that cannot be extended globally to $\text{Sym}^{\neq}(n)$ in a continuous way. We show an inequality on $\text{Sym}(n)$ that is weaker than a continuity result but that allows us to prove the continuity of our Riemannian metrics in Chapter 4.

Chapter 3. Characterization of invariant inner products. We give a general characterization of invariant inner products on a completely reducible Euclidean space. We apply it to the congruence action on square matrices of the orthogonal group $\text{O}(n)$, the permutation group $\mathfrak{S}(n)$ and the block-orthogonal group $\text{O}(I)$ where $I = \{k_1, \dots, k_p\}$ is a partition of n , i.e. $k_i \in \{1, \dots, n\}$ such that $\sum_{i=1}^p k_i = n$. We use the characterization of $\mathfrak{S}(n)$ -invariant inner

products on the vector space of symmetric hollow matrices $\text{Hol}(n)$ in Chapter 8.

Part III. Covariance matrices of full rank: the open cone of symmetric positive definite matrices

Chapter 4. $O(n)$ -invariant Riemannian metrics. We characterize the continuous $O(n)$ -invariant Riemannian metrics on SPD matrices by means of three continuous multivariate maps $\alpha, \beta, \gamma : \mathbb{R}^n \rightarrow \mathbb{R}$. We build a sequence of nested classes between mean kernel metrics and $O(n)$ -invariant metrics by imposing constraints on these functions and we extend some results to these super-classes. We also summarize and complete the knowledge on five noted Riemannian metrics: Euclidean, log-Euclidean, affine-invariant, Bures-Wasserstein and Bogoliubov-Kubo-Mori.

This chapter was resubmitted in January 2022 to the journal Linear Algebra and its Applications after minor revisions [Thanwerdas and Pennec, 2022b].

Chapter 5. Geometry of Mixed-Euclidean metrics. We formalize two principles of construction of Riemannian metrics: the principle of deformed metrics and the principle of balanced metrics. With the former, we define the new family of deformed-Wasserstein metrics and we give a sufficient condition under which the power-Wasserstein (or alpha-Procrustes) metric is a mean kernel metric. With the latter, we define the new family of Mixed-Euclidean metrics encompassing the Mixed-Power-Euclidean metrics. We unveil their relation with (u, v) -divergences and (α, β) -divergences in information geometry and we compute their Riemann curvature.

This chapter was published in the journal Differential Geometry and its Applications in April 2022 under the title “The geometry of mixed-Euclidean metrics on symmetric positive definite matrices” [Thanwerdas and Pennec, 2022a].

Part IV. Correlation matrices of full rank: the open elliptope

Chapter 6. Geometry of Quotient-affine metrics. We compute several Riemannian operations of the quotient-affine metric on full-rank correlation matrices: vertical and horizontal distributions, vertical and horizontal lifts, Riemannian metric, geodesics, Levi-Civita connection and sectional curvature. We also prove that there exists a Riemannian logarithm. We give explicit formulae of the geodesics, the logarithm and the distance in dimension 2 and we compare product metrics on SPD matrices with classical ones. In particular, we show that the correlation coefficient is interpolated monotonically in contrast with the affine-invariant and the log-Euclidean metrics.

This chapter was presented and published in the proceedings of the conference Geometric Science of Information 2021 under the title “Geodesics and Curvature of the Quotient-Affine Metrics on Full-Rank Correlation Matrices” [Thanwerdas and Pennec, 2021].

Chapter 7. Theoretically and computationally convenient Cholesky-based geometries. We show that the sectional curvature of the quotient-affine metric is of non-constant

sign and is unbounded from above. We show that the manifold of SPD matrices can be endowed via the Cholesky map with a Lie group structure under which the affine-invariant metric is a left-invariant metric. This allows us to generalize the construction of quotient-affine metrics. However, unsatisfied by these complex geometries, we propose three families of much more convenient Hadamard metrics on full-rank correlation matrices by pullback via the Cholesky map.

This chapter was submitted to the SIAM Journal on Matrix Analysis and its Applications in January 2022 under the title “Theoretically and computationally convenient geometries on full-rank correlation matrices” [Thanwerdas and Pennec, 2022c].

Chapter 8. Permutation-invariant Log-Euclidean metrics. We define two new families of permutation-invariant log-Euclidean metrics on full-rank correlation matrices called the off-log metrics and the log-scaled metrics. We define a natural involution called the cor-inversion and we show that the log-scaled metrics are inverse-consistent. We provide all the Riemannian operations in closed form modulo the computation of the bijections between the open ellipsope and the vector spaces in consideration. We propose a new algorithm to compute the scaling of an SPD matrix and thus the log-scaled bijection.

Part V. Stratified spaces of covariance and correlation matrices

Chapter 9. Bures-Wasserstein stratified geometry of covariance matrices. We explain the global stratified geometry of covariance matrices endowed with the Bures-Wasserstein distance relying on the theory of Riemannian orbit spaces. We complete the literature on geodesics in the strata of covariance matrices of fixed rank, specifying the injectivity domain of the exponential map, the set of preimages, the set of logarithms and giving explicit expressions based on the expression of the horizontal lift. We contribute the expression of all minimizing geodesics between two covariance matrices of any rank, we give the exact number of geodesics and we provide the expression of a canonical geodesic between any two covariance matrices.

Part VI. Conclusion

Chapter 10. Conclusion and perspectives. We conclude the thesis and we propose some perspectives for future works.

Part VII. Appendix

Chapter 11. Proofs. For readability, we decided to defer to this chapter almost all the proofs that were not essential to the comprehension of the chapters, that is most of them. After each result, a sentence “*See the proof of Result X in Section Y.*” invites the reader to click on “Y” to see the proof. The result is restated there with the complete proof. It suffices to click on “X” there to go back to the result in the chapter.

Part II

Mathematical preliminaries

Chapter 2

Regularity of eigenvalues and eigenvectors

Abstract

Some functions on symmetric matrices are defined via an eigenvalue decomposition. In this context, it is important to know the regularity of eigenvalues and eigenvectors to infer the regularity of such functions. In this work, we recall and show several results on this topic with an elementary approach. We show that there is no global continuous map of eigenvectors on the subspace of symmetric matrices with distinct eigenvalues. However, we show that there exist a global smooth map of eigenvalues and local smooth maps of eigenvectors on this subspace. We explain how it completes the literature. We also show an inequality on the whole space of symmetric matrices involving their eigenvector matrices, chosen to satisfy a condition of proximity. This inequality allows us to prove the continuity of a map defined via an eigenvalue decomposition, which seems to be new.

Following our policy on when to keep or to defer proofs, we kept the majority of the proofs within this chapter because they help understand the results.

2.1 Introduction

In the context of differential geometry on subspaces of symmetric matrices such as Symmetric Positive Definite (SPD) matrices or correlation matrices, one may want to define objects that depend on the eigenvalues or the eigenvectors of a symmetric matrix. For example, let \mathcal{F} be a generic space, e.g. a topological space or a smooth manifold, and let $f : \text{O}(n) \times \text{Diag}(n) \rightarrow \mathcal{F}$ be a function with some regularity, e.g. continuous or smooth. We assume that f is invariant under the choice of an eigenvalue decomposition: for all $P, Q \in \text{O}(n)$, for all $D, \Delta \in \text{Diag}(n)$, if $PDP^\top = Q\Delta Q^\top$, then $f(P, D) = f(Q, \Delta)$. We say that f is a spectral map. Therefore, one can define the quotient map $\tilde{f} : \Sigma = \text{PDP}^\top \in \text{Sym}(n) \mapsto f(P, D) \in \mathcal{F}$. To study the regularity of \tilde{f} in function of the regularity of f , it seems useful to study the regularity of eigenvalues and eigenvectors.

However, eigenvalues and eigenvectors are a priori multi-valued functions. Indeed, although symmetric matrices are diagonalizable in an orthonormal basis and their eigenvalues are real which are very nice properties, one symmetric matrix has several vectors of eigenvalues (up to permutation) and several bases of eigenvectors (up to symmetry and rotation in the eigenspaces). Thus it is worth formalizing eigenvalues and eigenvectors in terms of functions before studying their regularity.

The eigenvalues of symmetric matrices are real so it suffices to order them to neutralize the degree of freedom, that is the permutation. Therefore, we can define the map $val : \text{Sym}(n) \rightarrow \text{Diag}(n)$ which associates the diagonal matrix of non-decreasingly ordered eigenvalues to any symmetric matrix. It is well known that this map is 1-Lipschitz [Bhatia, 1997, Theorem III.4.4, Equation IV.62], in particular it is continuous.

Now we can call “map of eigenvectors” any map $vec : \mathcal{U} \subseteq \text{Sym}(n) \rightarrow O(n)$ satisfying $vec(\Sigma)val(\Sigma)vec(\Sigma)^\top = \Sigma$ for all $\Sigma \in \mathcal{U}$, where \mathcal{U} is an open subset of symmetric matrices. By analogy with eigenvalues, one could also neutralize the degree of freedom to define a map of eigenvectors $vec : \text{Sym}(n) \rightarrow O(n)$. However, such a map cannot be continuous at any $\Sigma \in \text{Sym}(n)$ having two equal eigenvalues [Katō, 1995]. Denoting $\text{Diag}^\neq(n)$ and $\text{Sym}^\neq(n) = val^{-1}(\text{Diag}^\neq(n))$ the open subsets of diagonal matrices and symmetric matrices with n distinct eigenvalues, we show in this work that there is no continuous map of eigenvectors on $\text{Sym}^\neq(n)$ either. Nevertheless, we show that for all $\Sigma \in \text{Sym}^\neq(n)$, there exists an open neighborhood $\mathcal{U} \subset \text{Sym}^\neq(n)$ of Σ and a continuous map of eigenvectors $vec : \mathcal{U} \rightarrow O(n)$ which is even smooth. In other words, we show that there exist local smooth maps of eigenvectors on $\text{Sym}^\neq(n)$. This also allows us to show that the global map of increasingly ordered eigenvalues is smooth on $\text{Sym}^\neq(n)$.

Going back to our initial problem with the spectral map $f : O(n) \times \text{Diag}(n) \rightarrow \mathcal{F}$ and the quotient map $\tilde{f} : \Sigma = PDP^\top \in \text{Sym}(n) \mapsto f(P, D) \in \mathcal{F}$, the above results show that if f is smooth, then \tilde{f} is smooth on $\text{Sym}^\neq(n)$. Stating properties on the whole vector space $\text{Sym}(n)$ seems much more difficult. We propose the beginning of an answer by showing a result on eigenvectors of any pair of symmetric matrices, including those with repeated eigenvalues. Given two symmetric matrices $\Sigma, \Lambda \in \text{Sym}(n)$, given a matrix of eigenvectors $Q \in O(n)$ of Λ , the result essentially explains how to choose the matrix of eigenvectors $P \in O(n)$ of Σ that is the closest to Q . For example, if $\Sigma = I_n$, the closest-to- Q matrix of eigenvalues $P \in O(n)$ of Σ is clearly $P = Q$. These matrices P and Q satisfy an inequality that is useful to study the regularity of \tilde{f} on $\text{Sym}(n)$. In particular, we show that if f is continuous, then \tilde{f} is continuous.

These results are close to already known results. In perturbation theory [Katō, 1995, Chapter II], these results are often formulated in function of a real parameter $t \in \mathbb{R}$ which captures the perturbation of the matrix $\Sigma \in \text{Sym}(n)$ into $\Sigma(t) = \Sigma + tX \in \text{Sym}(n)$. For example in [Katō, 1995, Theorem II.6.8] and in [Serre, 2010, Section 6.2, Comment 3], it is stated that if a curve on symmetric matrices $t \in I \mapsto \Sigma(t) \in \text{Sym}(n)$ is of class \mathcal{C}^k , then there exists a map of (non necessarily ordered) eigenvalues $(\lambda_1, \dots, \lambda_n) : I \rightarrow \mathbb{R}^n$ of class \mathcal{C}^k . It is also stated that this does not generalize to two parameters or to eigenvectors. This is probably why we did not manage to use this kind of results to prove the continuity of \tilde{f} . Here in contrast, we prove regularity results depending directly on the matrix Σ (not on a real parameter), using differential calculus.

Besides perturbation theory, the book of Denis Serre [Serre, 2010] summarizes some results on eigenvalues and eigenvectors of complex/real matrices in general and more specifically of Hermitian/symmetric matrices. In [Serre, 2010, Theorem 5.3], it is stated that given a matrix M_0 with complex coefficients, given an algebraically simple eigenvalue $\lambda_0 \in \mathbb{C}$ of M_0 , there exist a neighborhood \mathcal{U} of M_0 and analytic functions $\lambda : \mathcal{U} \rightarrow \mathbb{C}$ and $x : \mathcal{U} \rightarrow \mathbb{C}^n$ such that $\lambda(M)$ is an eigenvalue of M , $x(M)$ is an eigenvector of M associated to $\lambda(M)$, and $\lambda(M_0) = \lambda_0$. When $M_0 \in \text{Sym}^\neq(n)$, all eigenvalues are algebraically simple so we prove in

this specific case that the map λ can even be defined globally. Another difference is that we don't use the implicit function theorem to define the map x . On the contrary, we give a semi-constructive approach in the sense that once chosen an $x_0 = x(M_0)$, we explain how to build the neighborhood \mathcal{U} and how to make a consistent choice of sign for the eigenvector $x(M)$ so that $x : \mathcal{U} \rightarrow \mathbb{C}^n$ is smooth.

Thus, our main contribution is probably the inequality we mentioned on the whole space $\text{Sym}(n)$ allowing to show the continuity of \tilde{f} . The other results bring complements to known results with an elementary point of view. In the remainder of this section, we introduce some notations. In Section 2.2, we prove that there is no continuous map of eigenvectors $\text{vec} : \text{Sym}^\neq(n) \rightarrow \text{O}(n)$. In Section 2.3, we prove that there exist local Lipschitz maps of eigenvectors on $\text{Sym}^\neq(n)$. In Section 2.4, we prove that the map of ordered eigenvalues on $\text{Sym}^\neq(n)$ and the local maps of eigenvectors are not only Lipschitz but also smooth. In Section 2.5, we prove that \tilde{f} is continuous if f is continuous. We conclude in Section 2.6.

Notations and preliminary notions We denote \mathbb{R}^+ (resp. \mathbb{R}^*) the set of positive (resp. non null) real numbers. We denote $\text{Mat}(n)$ the vector space of real square matrices of size n , $\text{Sym}(n)$ the vector subspace of real symmetric matrices, $\text{Diag}(n)$ the vector subspace of real diagonal matrices. We denote $\text{Diag} : \text{Mat}(n) \rightarrow \text{Diag}(n)$ the linear operator pinching the diagonal of a matrix. The Frobenius norm is denoted $\|\cdot\|$ on $\text{Mat}(n)$ and on subspaces. We denote $\text{O}(n)$ the orthogonal group.

Definition 2.1 (Spectral map, quotient map) Let \mathcal{F} be a set. We say that $f : \text{O}(n) \times \text{Diag}(n) \rightarrow \mathcal{F}$ is a spectral map if for all $P, Q \in \text{O}(n)$, for all $D, \Delta \in \text{Diag}(n)$, if $PDP^\top = Q\Delta Q^\top$, then $f(P, D) = f(Q, \Delta)$. This automatically defines the quotient map $\tilde{f} : \text{Sym}(n) \rightarrow \mathcal{F}$.

We denote $\text{Diag}^\neq(n)$ the subset of diagonal matrices with n distinct eigenvalues, $\text{Diag}^<(n)$ the subset of diagonal matrices with n distinct eigenvalues ordered increasingly, $\text{Diag}^\leq(n)$ the subset of diagonal matrices with non-decreasingly ordered eigenvalues, and $\text{Diag}^=(n) = (\text{Diag}^\neq(n))^c$ the subset of diagonal matrices with repeated eigenvalues. In other words, if $D \in \text{Diag}^=(n)$, then there exist $i \neq j \in \{1, \dots, n\}$ such that $d_i = d_j$. Note that $\text{Diag}^\neq(n)$ and $\text{Diag}^<(n)$ are open in $\text{Diag}(n)$ since they are the respective preimages of the open sets $(\mathbb{R}^*)^{n-1}$ and $(\mathbb{R}^+)^{n-1}$ by the continuous map $D \in \text{Diag}(n) \mapsto (d_2 - d_1, \dots, d_n - d_{n-1}) \in \mathbb{R}^{n-1}$.

We denote analogously $\text{Sym}^\neq(n)$ the subset of symmetric matrices with n distinct eigenvalues, $\text{Sym}^=(n) = (\text{Sym}^\neq(n))^c$ the subset of symmetric matrices with repeated eigenvalues.

We introduce the map $\text{val} : \text{Sym}(n) \rightarrow \text{Diag}^\leq(n)$ associating to a matrix $\Sigma \in \text{Sym}(n)$ the diagonal matrix $D = \text{val}(\Sigma) = \text{Diag}(d_1, \dots, d_n)$ of its (real) eigenvalues (with multiplicity) ordered non-decreasingly, i.e. $d_1 \leq \dots \leq d_n$. We recall that val is a 1-Lipschitz map.

Theorem 2.2 (val is 1-Lipschitz) [Bhatia, 1997, Theorem III.4.4, Equation IV.62] For all $\Sigma, \Lambda \in \text{Sym}(n)$, $\|\text{val}(\Sigma) - \text{val}(\Lambda)\| \leq \|\Sigma - \Lambda\|$. In other words, the map $\text{val} : \text{Sym}(n) \rightarrow \text{Diag}^\leq(n)$ is 1-Lipschitz.

In particular, val is continuous so the set of symmetric matrices with n distinct eigenvalues $\text{Sym}^\neq(n) = \text{val}^{-1}(\text{Diag}^<(n)) = \text{val}^{-1}(\text{Diag}^\neq(n))$ is open in $\text{Sym}(n)$.

For $\Sigma \in \text{Sym}(n)$, we denote $\text{sp}(\Sigma)$ the set of (real) eigenvalues of Σ .

Definition 2.3 (Partition of n) A partition of n is a vector $I = (k_1, \dots, k_p) \in \{1, \dots, n\}^p$ such that $k_1 + \dots + k_p = n$, for some $p \in \{1, \dots, n\}$. We say that $I = (k_1, \dots, k_p)$ is finer than (or is a subdivision of) $J = (l_1, \dots, l_q)$ and we denote $I \succcurlyeq J$ when there exists a partition of p denoted (m_1, \dots, m_q) such that $l_i = k_{m_1+\dots+m_{i-1}+1} + \dots + k_{m_1+\dots+m_i}$ for all $i \in \{1, \dots, q\}$, denoting $m_0 = 0$. It is a partial order on the set of partitions of n . The finest partition of n is $(1, \dots, 1) \in \mathbb{R}^n$ and the coarsest is (n) .

Note that our notion of partition of n slightly differs from the ones in [Groisser et al., 2017] because our diagonal matrices are ordered while they also want to handle matrices such as $\text{Diag}(0, 1, 0)$.

Definition 2.4 (Signature) The signature of a symmetric matrix $\Sigma \in \text{Sym}(n)$ is the vector of multiplicities of the ordered eigenvalues in $\text{val}(\Sigma) \in \text{Diag}^{\leq}(n)$. It is a partition of n . Signatures inherit the order of partitions of n .

Definition 2.5 (Maximal radius to coarser signature) The maximal radius to coarser signature is the map defined by $\rho : \Sigma \in \text{Sym}(n) \mapsto \frac{1}{\sqrt{2}} \min_{\lambda \neq \mu \in \text{sp}(\Sigma)} |\lambda - \mu| \in (0, +\infty]$. It is the largest number r such that the symmetric matrices in the open ball of $\text{Sym}(n)$ centered on Σ of radius r have a finer signature than Σ . The closed ball of radius $\rho(\Sigma)$ contains a symmetric matrix of strictly coarser signature. For all $\Sigma \in \text{Sym}^{\neq}(n)$, it means that $d(\Sigma, \text{Sym}^-(n)) = \rho(\Sigma)$ [Breiding et al., 2018]. Note that $\rho(\lambda I_n) = +\infty$.

More precisely, the open ball of center $\Sigma \in \text{Sym}^{\neq}(n)$ and radius $r > 0$:

- is included in $\text{Sym}^{\neq}(n)$ if and only if $r \leq \rho(\Sigma)$,
- is adherent to $\text{Sym}^-(n)$ without intersecting it if and only if $r = \rho(\Sigma)$,
- intersects $\text{Sym}^-(n)$ if and only if $r > \rho(\Sigma)$.

2.2 No continuous map of eigenvectors on $\text{Sym}^{\neq}(n)$

We start with a lemma that we use several times in the chapter.

Lemma 2.6 Let $P, Q \in O(n)$ such that $\|P - Q\| < \sqrt{2}$. Then $\text{Diag}(P^{\top}Q) > 0$.

Proof. We denote $P = [p_1 \cdots p_n]$ and $Q = [q_1 \cdots q_n]$ with $p_i, q_i \in \mathbb{R}^n$. Since $\|p_i - q_i\|^2 = \|p_i\|^2 + \|q_i\|^2 - 2\langle p_i | q_i \rangle = 2 - 2\langle p_i | q_i \rangle$ and $\|p_i - q_i\| \leq \|P - Q\| < \sqrt{2}$, we have $1 - \langle p_i | q_i \rangle < 1$ so $[P^{\top}Q]_{ii} = p_i^{\top} q_i = \langle p_i | q_i \rangle > 0$. Hence $\text{Diag}(P^{\top}Q) \in \text{Diag}^+(n)$. \square

Theorem 2.7 (No continuous map of eigenvectors on $\text{Sym}^{\neq}(n)$) Let $n \geq 2$. There exists no continuous map $\text{vec} : \text{Sym}^{\neq}(n) \rightarrow O(n)$ such that for all $\Sigma \in \text{Sym}^{\neq}(n)$, $\text{vec}(\Sigma)\text{val}(\Sigma)\text{vec}(\Sigma)^{\top} = \Sigma$.

See the proof of Theorem 2.7 in Section 11.1.

This forces to define maps of eigenvectors only locally around symmetric matrices with n distinct eigenvalues. We use the distance from $\Sigma \in \text{Sym}^\neq(n)$ to $\text{Sym}^=(n)$ to define appropriate neighborhoods around symmetric matrices with n distinct eigenvalues and local Lipschitz maps of eigenvectors in $\text{Sym}^\neq(n)$.

2.3 Local Lipschitz maps of eigenvectors in $\text{Sym}^\neq(n)$

Given $\Sigma \in \text{Sym}^\neq(n)$, we would like to define a Lipschitz map of eigenvectors $vec : \mathcal{U} \subset \text{Sym}^\neq(n) \rightarrow O(n)$ on a neighborhood \mathcal{U} around Σ . In particular, for all $\Lambda \in \mathcal{U}$, we would like to have an equality of the form $\|vec(\Sigma) - vec(\Lambda)\| \leq c_{\mathcal{U}}\|\Sigma - \Lambda\|$. First, we find a majoration of this type. Second, we show how it helps to define a map of eigenvectors around Σ .

2.3.1 Majoration of the distance between eigenvector matrices in $\text{Sym}^\neq(n)$

Theorem 2.8 (Majoration of the distance between eigenvector matrices in $\text{Sym}^\neq(n)$) For all $\Sigma \in \text{Sym}^\neq(n)$, for all $\Lambda \in \text{Sym}(n)$, for all matrices of eigenvectors $P, Q \in O(n)$ such that $\Sigma = P \text{val}(\Sigma) P^\top$ and $\Lambda = Q \text{val}(\Lambda) Q^\top$, if the matrix $U = P^\top Q$ has non-negative diagonal entries, then $\|P - Q\| \leq \frac{2}{\rho(\Sigma)}\|\Sigma - \Lambda\|$.

Corollary 2.9 (Existence of eigenvector matrices satisfying the majoration in $\text{Sym}^\neq(n)$) Given $\Sigma \in \text{Sym}^+(n)$, $\Lambda \in \text{Sym}(n)$ and $P \in O(n)$ (resp. $Q \in O(n)$) satisfying $\Sigma = P \text{val}(\Sigma) P^\top$ (resp. $\Lambda = Q \text{val}(\Lambda) Q^\top$), there exists $Q = Q_{\Lambda, P} \in O(n)$ (resp. $P = P_{\Sigma, Q} \in O(n)$) such that $\Lambda = Q \text{val}(\Lambda) Q^\top$ (resp. $\Sigma = P \text{val}(\Sigma) P^\top$) and $\|P - Q\| \leq \frac{2}{\rho(\Sigma)}\|\Sigma - \Lambda\|$.

Proof of Corollary 2.9. Given Σ, Λ and $P = [p_1 \cdots p_n]$, take $Q = [q_1 \cdots q_n] \in O(n)$ such that $\Lambda = Q \text{val}(\Lambda) Q^\top$ and change the signs of the q_i 's such that for all $i \in \{1, \dots, n\}$, $p_i^\top q_i \geq 0$. If Q is given instead, choose the signs of p_i 's the same way. \square

Proof of Theorem 2.8. We denote $D = \text{val}(\Sigma)$ and $\Delta = \text{val}(\Lambda)$. On the one hand, using Theorem 2.2 at line (2.1) and triangular inequality at line (2.2):

$$2\|\Sigma - \Lambda\| \geq \|\Sigma - \Lambda\| + \|D - \Delta\| = \|PDP^\top - Q\Delta Q^\top\| + \|Q(D - \Delta)Q^\top\| \quad (2.1)$$

$$\geq \|PDP^\top - QDQ^\top\| = \|P^\top(PDP^\top - QDQ^\top)Q\| = \|DU - UD\|. \quad (2.2)$$

Hence:

$$4\|\Sigma - \Lambda\|^2 \geq \|DU - UD\|^2 = \sum_{i,j} [DU - UD]_{ij}^2 = \sum_{i,j} (d_i - d_j)^2 U_{ij}^2 \quad (2.3)$$

$$= \sum_{i \neq j} (d_i - d_j)^2 U_{ij}^2 \quad (2.4)$$

$$\geq 2\rho(\Sigma)^2 \sum_{i \neq j} U_{ij}^2 = 2\rho(\Sigma)^2 \left(\sum_{i,j} U_{ij}^2 - \sum_i U_{ii}^2 \right) = 2\rho(\Sigma)^2 \text{tr}(I_n - \text{Diag}(U)^2). \quad (2.5)$$

On the other hand:

$$\|P - Q\|^2 = \|P^\top(P - Q)\|^2 = \|I_n - U\|^2 = 2\text{tr}(I_n - U) \quad (2.6)$$

Since $U \in \text{O}(n)$, each column of U is of norm 1 so all entries belong to $[-1, 1]$. The hypothesis ensures that for all $i \in \{1, \dots, n\}$, $U_{ii} \geq 0$ so $U_{ii} \in [0, 1]$. Hence $U_{ii}^2 \leq U_{ii}$ so $\text{tr}(\text{Diag}(U)^2) = \sum_i U_{ii}^2 \leq \sum_i U_{ii} = \text{tr}(U)$. Finally:

$$\frac{4}{\rho(\Sigma)^2} \|\Sigma - \Lambda\|^2 \geq 2\text{tr}(I_n - \text{Diag}(U)^2) \geq 2\text{tr}(I_n - U) = \|P - Q\|^2, \quad (2.7)$$

so $\|P - Q\| \leq \frac{2}{\rho(\Sigma)} \|\Sigma - \Lambda\|$. \square

Remark 2.10 Note that these results are valid for $\Lambda \in \text{Sym}(n)$, not only for $\Lambda \in \text{Sym}^\neq(n)$.

2.3.2 Existence of local Lipschitz maps of eigenvectors in $\text{Sym}^\neq(n)$

Theorem 2.11 (Existence of local Lipschitz maps of eigenvectors in $\text{Sym}^\neq(n)$) For all $\Sigma \in \text{Sym}^\neq(n)$, there exists an open ball $\mathcal{U} = \mathcal{B}(\Sigma, \frac{\rho(\Sigma)}{2\sqrt{2}}) \subset \text{Sym}^\neq(n)$ around Σ on which there exists a Lipschitz map of eigenvectors $vec = vec_{\mathcal{U}} : \mathcal{U} \rightarrow \text{O}(n)$, i.e. for all $\Lambda \in \mathcal{U}$, $vec(\Lambda)val(\Lambda)vec(\Lambda)^\top = \Lambda$. In other words, there exists a constant $c_{\mathcal{U}} \in \mathbb{R}^+$ such that for all $\Lambda, \Lambda' \in \mathcal{U}$, $\|vec(\Lambda) - vec(\Lambda')\| \leq c_{\mathcal{U}} \|\Lambda - \Lambda'\|$.

Proof. Let $\Sigma \in \text{Sym}^\neq(n)$. Let \mathcal{U} be the open ball centered on Σ of radius $\varepsilon = \frac{\rho(\Sigma)}{2\sqrt{2}}$. Since $\varepsilon \leq \rho(\Sigma)$, $\mathcal{U} \subset \text{Sym}^\neq$ by Definition 2.5. We choose $P = [p_1 \cdots p_n] \in \text{O}(n)$ such that $\Sigma = Pval(\Sigma)P^\top$. There are 2^n such matrices because each p_i can be multiplied by -1 : for example, we can choose the one for which on each column, the first non null coefficient is positive.

Let $\Lambda \in \mathcal{U}$. Among the 2^n matrices $Q = [q_1 \cdots q_n] \in \text{O}(n)$ such that $\Lambda = Qval(\Lambda)Q^\top$, there is only one such that $\text{Diag}(P^\top Q) \geq 0$. Indeed by Corollary 2.9, such a matrix Q exists and by Theorem 2.8, it satisfies $\|P - Q\| \leq \frac{2}{\rho(\Sigma)} \|\Sigma - \Lambda\| < \frac{2}{\rho(\Sigma)} \varepsilon = \frac{1}{\sqrt{2}} < \sqrt{2}$ so by Lemma 2.6, $\text{Diag}(P^\top Q) > 0$. In other words, $p_i^\top q_i$ cannot be null (because $\Lambda \in \mathcal{U}$) so the signs of q_i 's are imposed by P . Thus we can define the map of eigenvectors $vec = vec_{\mathcal{U}} : \Lambda \in \mathcal{U} \mapsto Q \in \text{O}(n)$ where Q is the unique orthogonal matrix such that $\Lambda = Qval(\Lambda)Q^\top$ and $\text{Diag}(P^\top Q) > 0$.

Let us show that vec is Lipschitz. Let $\Lambda, \Lambda' \in \mathcal{U}$ and $Q = vec(\Lambda), Q' = vec(\Lambda') \in \text{O}(n)$. By triangular inequality and by Theorem 2.8 applied to (Σ, Λ, P, Q) and $(\Sigma, \Lambda', P, Q')$, we have:

$$\|Q - Q'\| \leq \|P - Q\| + \|P - Q'\| \leq \frac{2}{\rho(\Sigma)} (\|\Sigma - \Lambda\| + \|\Sigma - \Lambda'\|) < \frac{4\varepsilon}{\rho(\Sigma)} = \sqrt{2}. \quad (2.8)$$

Hence, by Lemma 2.6, for all $i \in \{1, \dots, n\}$, $\langle q_i | q'_i \rangle > 0$. Therefore, by Theorem 2.8 applied to $(\Lambda, \Lambda', Q, Q')$, we have $\|Q - Q'\| \leq \frac{2}{\rho(\Lambda)} \|\Lambda - \Lambda'\|$. Finally, $\inf_{\Lambda \in \mathcal{U}} \rho(\Lambda) = \inf_{\Lambda \in \mathcal{U}} d(\Lambda, \text{Sym}^=(n)) = \rho(\Sigma) - \varepsilon = \frac{2\sqrt{2}-1}{2\sqrt{2}} \rho(\Sigma) > 0$ so we can define $c_{\mathcal{U}} = \sup_{\Lambda \in \mathcal{U}} \frac{2}{\rho(\Lambda)} = \frac{4\sqrt{2}}{2\sqrt{2}-1} \frac{1}{\rho(\Sigma)} < +\infty$ and conclude that the map $vec : \mathcal{U} \rightarrow \text{O}(n)$ is $c_{\mathcal{U}}$ -Lipschitz. Note that $c_{\mathcal{U}} \geq \frac{2}{\rho(\Sigma)}$ tends to $+\infty$ when Σ tends to a matrix in $\text{Sym}^=(n)$ since $\rho(\Sigma)$ tends to 0. \square

2.4 Smoothness of eigenvalues and local eigenvectors on $\text{Sym}^\neq(n)$

In this section, we prove the smoothness of the map of eigenvalues on symmetric matrices with distinct eigenvalues, $val|_{\text{Sym}^\neq(n)} : \text{Sym}^\neq(n) \rightarrow \text{Diag}^<(n)$, and the smoothness of the local maps of eigenvectors $vec : \mathcal{U} = \mathcal{B}(\Sigma, \frac{\rho(\Sigma)}{2\sqrt{2}}) \subset \text{Sym}^\neq(n) \rightarrow \text{O}(n)$.

Theorem 2.12 (Smoothness of the map val on $\text{Sym}^\neq(n)$ and smoothness of the maps vec) The maps $val|_{\text{Sym}^\neq(n)} : \text{Sym}^\neq(n) \rightarrow \text{Diag}^<(n)$ and $vec : \mathcal{U} = \mathcal{B}(\Sigma, \frac{\rho(\Sigma)}{2\sqrt{2}}) \rightarrow \text{O}(n)$ are smooth. Their differentials at $\Sigma \in \mathcal{U} \subset \text{Sym}^\neq(n)$ are for all $X \in \text{Sym}(n)$, denoting $D = val(\Sigma) = \text{diag}(d_1, \dots, d_n)$ and $P = vec(\Sigma)$:

$$d_\Sigma val(X) = \text{Diag}(P^\top X P), \quad (2.9)$$

$$d_\Sigma vec(X) = P A, \quad (2.10)$$

where $A \in \text{Skew}(n)$ is defined by $A_{ij} = \frac{[P^\top X P]_{ij}}{d_j - d_i}$ for all $i \neq j$ (and $A_{ii} = 0$).

Proof. Let us show that if $val|_{\text{Sym}^\neq(n)}$ and vec are differentiable, then necessarily their differentials satisfy Equations (2.9) and (2.10). We differentiate the equality $\Sigma = vec(\Sigma) val(\Sigma) vec(\Sigma)^\top$ in the direction $X \in \text{Sym}(n)$, denoting $\dot{P} = d_\Sigma vec(X) \in T_P \text{O}(n) = P \text{Skew}(n)$ and $\dot{D} = d_\Sigma val(X) \in T_D \text{Diag}(n) = \text{Diag}(n)$:

$$X = \dot{P} D P^\top + P \dot{D} P^\top + P \dot{D} P^\top, \quad (2.11)$$

$$P^\top X P = P^\top \dot{P} D + D \dot{P}^\top P + \dot{D}. \quad (2.12)$$

We denote $A = P^\top \dot{P} \in \text{Skew}(n)$. Then $P^\top X P = AD - DA + \dot{D}$ so \dot{D} is the diagonal part of $P^\top X P$ and $AD - DA = [A_{ij}(d_j - d_i)]_{1 \leq i, j \leq n}$ is the off-diagonal part, as expected.

Let us show that $val|_{\text{Sym}^\neq(n)}$ and vec are differentiable. Let $\Sigma \in \mathcal{U}$ and $X \in \text{Sym}(n)$. Now we denote $\dot{D} = \text{Diag}(P^\top X P)$ and $\dot{P} = P A$ with $A \in \text{Skew}(n)$ defined by $A_{ij} = \frac{[P^\top X P]_{ij}}{d_j - d_i}$ for all $i \neq j$. Let us show that:

$$val(\Sigma + tX) = D + t\dot{D} + o(t), \quad (2.13)$$

$$vec(\Sigma + tX) = P + t\dot{P} + o(t). \quad (2.14)$$

Since it is difficult to see what are $val(\Sigma + tX)$ and $vec(\Sigma + tX)$, let us define a smooth curve $\Sigma(t) \in \mathcal{U}$ for t sufficiently small such that $val(\Sigma(t))$ and $vec(\Sigma(t))$ are easier to determine and such that $\Sigma(0) = \Sigma$ and $\dot{\Sigma} = X$ so that $\Sigma(t) = \Sigma + tX + o(t)$. Since val and vec are Lipschitz on \mathcal{U} , we will have $val(\Sigma(t)) = val(\Sigma + tX) + o(t)$ and $vec(\Sigma(t)) = vec(\Sigma + tX) + o(t)$. To do so, it suffices to define $\Sigma(t) = P(t)D(t)P(t)^\top$ with $P(0) = P$, $D(0) = D$, $\dot{P}(0) = \dot{P} = P A$ and $\dot{D}(0) = \dot{D}$. For example, let $P(t) = P \exp(tA)$ and $D(t) = D + t\dot{D}$. For t sufficiently small, $D(t) \in \text{Diag}^<(n)$ since $\text{Diag}^<(n)$ is open so $D(t) = val(\Sigma(t))$. For t sufficiently small, $P^\top P(t) = \exp(tA)$ has positive diagonal since $\exp(tA) = I_n + o(1)$, so $P(t) = vec(\Sigma(t))$. Hence:

$$val(\Sigma + tX) = val(\Sigma(t)) + o(t) = D(t) + o(t) = D + t\dot{D} + o(t), \quad (2.15)$$

$$vec(\Sigma + tX) = vec(\Sigma(t)) + o(t) = P(t) + o(t) = P + t\dot{P} + o(t) \quad (2.16)$$

So val and vec are differentiable on \mathcal{U} .

Let us assume that val and vec are differentiable k times on \mathcal{U} . We denote $L(E, F)$ the vector space of linear maps from the vector space E to the vector space F . Then for all $i \neq j$, the map $A_{ij} : \Sigma \in \mathcal{U} \mapsto \left(X \in \text{Sym}(n) \mapsto \frac{[vec(\Sigma)^\top X vec(\Sigma)]_{ij}}{[val(\Sigma)]_{jj} - [val(\Sigma)]_{ii}} \right) \in L(\text{Sym}(n), \mathbb{R})$ is differentiable k times so $A : \mathcal{U} \rightarrow L(\text{Sym}(n), \text{Skew}(n))$ is differentiable k times. Since $d_\Sigma val(X) = \text{Diag}(vec(\Sigma)^\top X vec(\Sigma))$ and $d_\Sigma vec(X) = vec(\Sigma)A(\Sigma, X)$, the maps $dval : \mathcal{U} \rightarrow L(\text{Sym}(n), \text{Diag}(n))$ and $dvec : \mathcal{U} \rightarrow L(\text{Sym}(n), \text{Mat}(n))$ are differentiable k times, so val and vec are differentiable $k+1$ times. Since val and vec are differentiable on \mathcal{U} , by recurrence they are smooth on \mathcal{U} . Since $val|_{\text{Sym}^\neq(n)}$ is locally smooth, it is globally smooth. \square

We have the following immediate corollary.

Corollary 2.13 (Smoothness of a spectral map) Let \mathcal{M} be a smooth manifold, let $f : O(n) \times \text{Diag}(n) \rightarrow \mathcal{M}$ be a smooth spectral map and let $\tilde{f} : \text{Sym}(n) \rightarrow \mathcal{M}$ be the associated quotient map. Then $\tilde{f}|_{\text{Sym}^\neq(n)}$ is smooth.

Proof. It suffices to write $\tilde{f}|_{\mathcal{U}} = f \circ (vec_{\mathcal{U}}, val|_{\mathcal{U}})$. \square

Note that Corollary 2.13 could already be deduced from [Serre, 2010, Theorem 5.3]. The notable results up to now are the existence of a *global* smooth map of eigenvalues on $\text{Sym}^\neq(n)$ and the fact that the choice of the neighborhood and the definition of the smooth map of eigenvectors rely on the assumption $\text{Diag}(vec(\Sigma)^\top vec(\Lambda)) > 0$. In the next section, we generalize Theorem 2.8 to symmetric matrices with repeated eigenvalues, that is we give an analogous inequality with a more general condition.

2.5 Majoration of the distance between eigenvector matrices in $\text{Sym}(n)$

Back to our initial problem, when one wants to prove regularity properties on objects defined via eigenvalue decomposition, it is important to know how to choose the decomposition. The answer is now rather clear on $\text{Sym}^\neq(n)$. However, the regularity on the whole vector space $\text{Sym}(n)$ is more difficult to handle than on the open set $\text{Sym}^\neq(n)$. Although there is no local continuous map of eigenvectors around $\Sigma \in \text{Sym}^\neq(n)$, given a symmetric matrix Λ close to Σ , there is a way of choosing matrices of eigenvectors P, Q that are the closest possible. We explain this choice in the following result. It allows us to give the beginning of an answer to our initial problem. Indeed, we further prove that if the spectral map is continuous, then the corresponding function defined on symmetric matrices is continuous on $\text{Sym}(n)$.

Theorem 2.14 (Majoration of the distance between eigenvector matrices in $\text{Sym}(n)$) Let $\Sigma, \Lambda \in \text{Sym}(n)$, let $P, Q \in O(n)$ such that $\Sigma = P val(\Sigma) P^\top$ and $\Lambda = Q val(\Lambda) Q^\top$. We denote $p = \text{card}(\text{sp}(\Sigma))$, $\lambda_1 < \dots < \lambda_p$ the eigenvalues of Σ without multiplicity and n_1, \dots, n_p their respective multiplicity, with $n_1 + \dots + n_p = n$. Let $U = P^\top Q \in O(n)$ and let $U_1 \in \text{Mat}(n_1), \dots, U_p \in \text{Mat}(n_p)$ the diagonal blocks of U . If the U_i 's are symmetric positive semi-definite matrices, then $\|P - Q\| \leq \frac{2}{\rho(\Sigma)} \|\Sigma - \Lambda\|$.

This condition generalizes $\text{Diag}(P^\top Q) > 0$. The following corollary ensures that given Q , there exists P satisfying this condition.

Corollary 2.15 (Existence of eigenvector matrices satisfying the majoration in $\text{Sym}(n)$) Given such $\Sigma, \Lambda \in \text{Sym}(n)$ and $Q \in O(n)$ such that $\Lambda = Q \text{vec}(\Lambda) Q^\top$, there exists $P = P_{\Sigma, Q} \in O(n)$ such that $\Sigma = P \text{val}(\Sigma) P^\top$ and $\|P - Q\| \leq \frac{2}{\rho(\Sigma)} \|\Sigma - \Lambda\|$.

The proofs consist in replacing $\text{Diag}(U) = \text{diag}(U_{11}, \dots, U_{nn})$ in the proof of Theorem 2.8 by the block-diagonal matrix $W = \text{Diag}(U_1, \dots, U_p)$ pinched from U .

Proof of Corollary 2.15. Given Σ, Λ, Q , take P such that $\Sigma = P \text{val}(\Sigma) P^\top$. Let $S \in \text{Sym}^+(n)$ and $R \in O(n)$ be block diagonal matrices of same signature as W such that $W^\top = SR^\top$. Hence $R \text{val}(\Sigma) R^\top = \text{val}(\Sigma)$ so P can be replaced by PR . Since $R^\top W = S$, the diagonal blocks of the new $U' := (PR)^\top Q = R^\top U$ are the diagonal blocks of $R^\top W$. Hence PR satisfies the hypothesis of Theorem 2.14. \square

Remark 2.16 Beware that it is not true that given $\Sigma, \Lambda \in \text{Sym}(n)$ and $P \in O(n)$ such that $\Sigma = P \text{val}(\Sigma) P^\top$, there always exists an appropriate Q . This is because the symmetry between Σ and Λ is broken by the (arbitrary) choice of $\rho(\Sigma)$ rather than $\rho(\Lambda)$. Therefore, the condition on $P^\top Q$ relies on the signature (n_1, \dots, n_p) of Σ (see proof of Theorem 2.14). Since the signature of Λ can be completely different, once any eigenvector matrices P and Q of Σ and Λ are respectively chosen, we can only modify P in general so that $P^\top Q$ satisfies the condition of positivity (see proof of Corollary 2.15).

A counter-example of the existence of $Q = Q_{\Lambda, P}$ is $\Sigma = 0$ (hence $\rho(\Sigma) = +\infty$), $\Lambda = \text{diag}(t, -t)$ with $t \neq 0$ and $P = \begin{pmatrix} \cos \theta & \pm \sin \theta \\ \sin \theta & \mp \cos \theta \end{pmatrix}$ with $\theta \in \mathbb{R} \setminus \frac{\pi}{2}\mathbb{Z}$.

However, if the signature of Σ is finer than the signature of Λ , then Q can be adapted so that $P^\top Q$ satisfies the condition. Hence in this case, it is true that given Σ, Λ, P , there exists an appropriate Q . This was the case in Theorem 2.8 because $\Sigma \in \text{Sym}^\neq(n)$, i.e. the signature of Σ was $(1, \dots, 1)$ which is the finest signature. This allows to explain Remark 2.10.

Proof of Theorem 2.14. Lines (2.1) to (2.3) of the previous proof are still valid. Hence:

$$4\|\Sigma - \Lambda\|^2 \geq \sum_{i,j|d_i \neq d_j} (d_i - d_j)^2 U_{ij}^2 \quad (2.17)$$

$$\geq 2\rho(\Sigma)^2 \sum_{i,j|d_i \neq d_j} U_{ij}^2 = 2\rho(\Sigma)^2 \left(\sum_{i,j} U_{ij}^2 - \sum_{\lambda \in \text{sp}(\Sigma)} \sum_{i,j|d_i=d_j=\lambda} U_{ij}^2 \right) \quad (2.18)$$

$$= 2\rho(\Sigma)^2 \text{tr} \left(UU^\top - \sum_{k=1}^p U_k U_k^\top \right) = 2\rho(\Sigma)^2 \text{tr}(I_n - WW^\top). \quad (2.19)$$

Moreover, line (2.6) is still valid so $\|P - Q\|^2 = 2\text{tr}(I_n - U) = 2\text{tr}(I_n - W)$.

The equality $UU^\top = I_n$ ensures that for all $k \in \{1, \dots, p\}$, $0 \leq U_k U_k^\top \leq I_{n_k}$ for the Loewner order. Indeed, for $k = 1$ for example, if we write $U = \begin{pmatrix} U_1 & A \\ B & C \end{pmatrix}$ with A, B, C of appropriate sizes, then $I_n = UU^\top = \begin{pmatrix} U_1 U_1^\top + AA^\top & * \\ * & * \end{pmatrix}$ so $U_1 U_1^\top + AA^\top = I_{n_1}$. Hence $U_1 U_1^\top \geq 0$

and $I_{n_1} - U_1 U_1^\top \geq 0$ as claimed. Otherwise said, $0 \leq WW^\top \leq I_n$, i.e. the eigenvalues of WW^\top are in $[0, 1]$. Hence $WW^\top \leq \sqrt{WW^\top}$. The hypothesis ensures that W is symmetric positive semi-definite so $W = \sqrt{WW^\top}$. Finally, $\text{tr}(W) = \text{tr}(\sqrt{WW^\top}) \geq \text{tr}(WW^\top)$ so $\frac{4}{\rho(\Sigma)^2} \|\Sigma - \Lambda\|^2 \geq 2\text{tr}(I_n - WW^\top) \geq 2\text{tr}(I_n - W) = \|P - Q\|^2$ as expected. \square

Remark 2.17 Theorem 2.14 is not a continuity result. Let $Q(t)$ be a curve on $O(n)$ for $t \in (0, 1]$ and $D(t)$ be a curve on $\text{Diag}(n)$ for $t \in [0, 1]$ with $D = D(0) \in \text{Diag}^-(n)$. Assume that $\Lambda(t) = Q(t)D(t)Q(t)^\top$ converges to $\Sigma \in \text{Sym}^-(n)$ when t tends to 0. The result does not mean that $Q(t)$ converges. It only states that for all $t \in (0, 1]$, one can find $P \in O(n)$ such that $\Sigma = PDP^\top$ and $\|P - Q(t)\| \leq \frac{2}{\rho(\Sigma)} \|\Sigma - \Lambda(t)\|$. So the choice of P depends on t ! Thus, this does not contradict the fact that eigenvectors are not continuous at symmetric matrices with repeated eigenvalues $\Sigma \in \text{Sym}^-(n)$.

Theorem 2.18 (Continuity of a spectral function) Let \mathcal{F} be a topological space, let $f : O(n) \times \text{Diag}(n) \rightarrow \mathcal{F}$ be a continuous spectral map and let $\tilde{f} : \text{Sym}(n) \rightarrow \mathcal{F}$ be the associated quotient map. Then \tilde{f} is continuous.

Proof. Let $\Sigma \in \text{Sym}(n)$ and let $\varepsilon > 0$. By Heine's theorem, f is uniformly continuous on $O(n) \times \mathcal{B}(\text{val}(\Sigma), 1)$ where \mathcal{B} is a closed ball in $\text{Diag}(n)$. Hence, let $\delta > 0$ such that if $\|(P, D) - (Q, \Delta)\| = \sqrt{\|P - Q\|^2 + \|D - \Delta\|^2} < \delta$, then $\|f(P, D) - f(Q, \Delta)\| < \varepsilon$.

Let $\eta = \min(1, \frac{\rho(\Sigma)}{2}) \frac{\delta}{\sqrt{2}}$. Let $\Lambda \in \text{Sym}(n)$ such that $\|\Sigma - \Lambda\| < \eta$. By Corollary 2.15, let $P, Q \in O(n)$ such that $\Sigma = P \text{val}(\Sigma) P^\top$, $\Lambda = Q \text{val}(\Lambda) Q^\top$ and $\|P - Q\| \leq \frac{2}{\rho(\Sigma)} \|\Sigma - \Lambda\|$. Then $\|P - Q\| \leq \frac{\delta}{\sqrt{2}}$ and $\|\text{val}(\Sigma) - \text{val}(\Lambda)\| \leq \|\Sigma - \Lambda\| \leq \frac{\delta}{\sqrt{2}}$. Therefore, $\|(P, D) - (Q, \Delta)\| = \sqrt{\|P - Q\|^2 + \|D - \Delta\|^2} < \delta$ so $\|f(P, D) - f(Q, \Delta)\| < \varepsilon$.

For all $\Lambda \in \text{Sym}(n)$, if $\|\Sigma - \Lambda\| < \eta$, then $\|\tilde{f}(\Sigma) - \tilde{f}(\Lambda)\| < \varepsilon$ so \tilde{f} is continuous at Σ . So \tilde{f} is continuous. \square

We said that Theorem 2.14 only gives the beginning of an answer because it would be interesting to know sufficient conditions on f so that \tilde{f} is of class \mathcal{C}^k on $\text{Sym}(n)$. We think that there must be conditions on the limit on the differential of f at points $(P, D) \in O(n) \times \text{Diag}^-(n)$.

2.6 Conclusion

We proved that increasingly ordered eigenvalues are smooth on the open subset of symmetric matrices with distinct eigenvalues $\text{Sym}^\neq(n)$. We showed that there is no global continuous map of eigenvectors but there exist local smooth maps of eigenvectors. This allowed to show that the quotient map of a smooth spectral map is smooth on $\text{Sym}^\neq(n)$. We also proved an inequality on the whole vector space of symmetric matrices, allowing to show that the quotient map of a continuous spectral map is continuous. We will use these results in Chapter 4. A more difficult problem is to find conditions on the smooth spectral map f and on its successive differentials for the associated spectral map \tilde{f} to be smooth. Such a result would be very appealing.

Chapter 3

Characterization of invariant inner products

Abstract

In several situations in differential geometry, one can be interested in determining all inner products on a vector space that are invariant under a given group action. For example, bi-invariant Riemannian metrics on a Lie group G are characterized by $\text{Ad}(G)$ -invariant inner products on the Lie algebra \mathfrak{g} . Analogously, G -invariant Riemannian metrics on a homogeneous space $\mathcal{M} = G/H$ are characterized by $\text{Ad}(H)$ -invariant inner products on the tangent space $T_H\mathcal{M}$. In addition, given a G -equivariant diffeomorphism between a manifold \mathcal{M} and a Euclidean space V , G -invariant log-Euclidean metrics can be defined on \mathcal{M} by pullback of G -invariant inner products on V . There exists a general procedure based on representation theory to find all invariant inner products on a completely reducible Euclidean space. It consists in changing the viewpoint from invariant inner products to equivariant automorphisms. The goal of this work is to diffuse this method to communities of applied mathematics which use differential geometry. Therefore, in this work, we recall this general method that we did not find elsewhere, along with an elementary presentation of the basics of representation theory. Besides, we illustrate the method on several examples that can be useful in further works, notably on the congruence action of orthogonal matrices, permutation matrices and block orthogonal matrices on squared matrices.

3.1 Introduction

When one looks for appropriate metrics on a given space representing some data, it is natural to require them to be invariant under a certain group action. For example, when data are represented by Symmetric Positive Definite (SPD) matrices, one can use Riemannian metrics that are invariant under the congruence action of the general linear group (affine-invariant metrics [Skovgaard, 1984, Pennec et al., 2006, Lenglet et al., 2006, Fletcher and Joshi, 2007, Moakher, 2005, Batchelor et al., 2005]), the orthogonal group [Thanwerdas and Pennec, 2022b] (e.g. log-Euclidean [Arsigny et al., 2006], Bures-Wasserstein [Bhatia et al., 2019, Dryden et al., 2009, Takatsu, 2011, Malagò et al., 2018], Bogoliubov-Kubo-Mori metrics [Petz and Toth, 1993, Michor et al., 2000]), the group of positive diagonal matrices [David and Gu, 2019, David, 2019, Thanwerdas and Pennec, 2021], the permutation group. In several situations, the question of finding all G -invariant Riemannian metrics on a manifold reduces to finding all H -invariant inner products on a vector space where H is a Lie subgroup of G . For example, bi-invariant metrics (or pseudo-metrics) on a Lie group G are characterized

by $\text{Ad}(G)$ -invariant inner products (or non-degenerate symmetric bilinear forms) on its Lie algebra \mathfrak{g} . On a homogeneous space $\mathcal{M} = G/H$, G -invariant metrics are characterized by $\text{Ad}(H)$ -invariant inner products on the tangent space $\mathfrak{m} = T_H\mathcal{M}$ at the equivalence class $H \in G/H$.

Another frequent case is when the manifold \mathcal{M} is diffeomorphic to a vector space V . For example, the cone of SPD matrices is diffeomorphic to the vector space of symmetric matrices by the symmetric matrix logarithm $\log : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$. Similarly, the ellipsope of full-rank correlation matrices $\text{Cor}^+(n)$ is diffeomorphic to a vector space of dimension $\frac{n(n-1)}{2}$. Hence, given a G -equivariant diffeomorphism $\phi : \mathcal{M} \rightarrow V$ between the manifold \mathcal{M} and a vector space V , the G -invariant inner products on V provide natural flat G -invariant Riemannian metrics on \mathcal{M} by pullback. These three examples of bi-invariant metrics on Lie groups, Riemannian homogeneous manifolds and “Euclideanized” manifolds motivate to find all G -invariant inner products on a vector space V . Indeed, if one wants to impose an invariance on a space, this requirement defines a family of metrics in general, rarely a unique metric. Therefore, there is no reason a priori to distinguish between all the metrics that satisfy this requirement. Then, it is possible to reduce the choice by requiring other invariance or constraints, or to optimize within the family in function of the data for example.

To answer this question, the central notion is the reducibility or irreducibility of a vector space under a group action. We say that V is G -irreducible when there is no other subvector space than $\{0\}$ and V that is stable under the action of G . It can easily be proved that inner products on an irreducible space are positive scalings of one another. Hence, when V is completely reducible, i.e. V can be expressed into a direct sum of irreducible subspaces, then any positive linear combination of inner products on each irreducible subspace is an invariant inner product. For example, if $G = \{-1, 1\} \times \{-1, 1\}$ acts on $V = \mathbb{R}^2 = \mathbb{R} \oplus \mathbb{R}$ component by component, then one can easily check that all G -invariant inner products are given by $\varphi((x, y), (x, y)) = \alpha x^2 + \beta y^2$ with $\alpha, \beta > 0$.

However, there are other invariant inner products in general. Take another example where $G = \{-1, 1\}$ acts on $V = \mathbb{R}^2 = \mathbb{R} \oplus \mathbb{R}$ globally. It is clear that all the G -invariant inner products are given by $\varphi((x, y), (x, y)) = \alpha x^2 + 2\gamma xy + \beta y^2$ with $\alpha, \beta > 0$ and $\alpha\beta > \gamma^2$. In the first example, G acts differently on each copy of \mathbb{R} : for example $(1, -1) \cdot (x, y) = (x, -y)$. In the second example, G acts identically on each copy of \mathbb{R} , they are indistinguishable with respect to the action. This is why another coefficient is allowed between the two components. Therefore, in the irreducible decomposition, one has to group all the irreducible spaces on which G acts the same way to find all G -invariant inner products. That is what we explain in Section 3.2 and what is summarized in Theorem 3.4.

In this work, we assume that we know a G -invariant inner product on V and that V is completely reducible, i.e. there exists an irreducible decomposition of V . This is the case when G is finite [Artin, 2011, Maschke’s Theorem 10.2.10] and also when the group action is unitary and continuous [Artin, 2011, Corollary 10.3.5]. In particular, when G is finite or compact, it is well known that one can find *one* G -invariant inner product by taking any inner product and averaging over G with the counting measure or the Haar measure μ : $\langle x|y \rangle := \frac{1}{|G|} \sum_{a \in G} ax \cdot ay$ or $\int_G (ax \cdot ay) \mu(da)$. So the two hypotheses are automatically satisfied when G is finite or when G is compact and the action is continuous. The examples we give all belong to one of these two situations although the method we give applies more generally as soon as the two mentioned hypotheses are satisfied together.

Our method relies on basic representation theory. Indeed, any G -invariant inner product on V is characterized, via the given inner product $\langle \cdot | \cdot \rangle$, by a G -invariant automorphism of V . Then, representation theory allows to find the general form of this automorphism, hence the general form of G -invariant inner products on V . Note that this characterization of invariant inner products is well known in algebra although we could not find it explicitly in the references we read. Yet, it seems to be much less known in the communities of applied mathematics which use differential geometry, although it seems quite useful. That is why we present it with the fewest details on representation theory and why we study thoroughly some examples.

The general method is exposed in Section 3.2. In the examples, we focus on the actions of the orthogonal group, the block orthogonal group and the permutation group. In Section 3.3, we apply this method to several examples. In Section 3.3.1 we recall the general form of invariant inner products on \mathbb{R}^n under these three group actions. In Section 3.3.2, we recall the general form of invariant inner products on square matrices under the congruence action of the orthogonal group. Then in Section 3.3.3, we characterize invariant inner products on square matrices under the block orthogonal group. In particular, this allows to characterize $O(n)$ -invariant metrics on the homogeneous space of flags $\text{Flag}(I) = O(n)/O(I)$ where I is a partition of n , which is new. Finally in Section 3.3.4, we characterize permutation-invariant inner products on square matrices. We give a natural instantiation of the irreducible decomposition with coordinate-free orthogonal projections. This allows to define Euclidean metrics on SPD matrices and correlation matrices. We summarize and conclude in Section 3.4.

3.2 Characterization of invariant inner products of a Euclidean space using representation theory

A representation of a group G on a vector space V is a group homomorphism $\rho : G \rightarrow \text{GL}(V)$. It is equivalent to a linear group action of G on V , i.e. a map $\rho : G \times V \rightarrow V$ such that for all $a, b \in G$ and all $x \in V$, $\rho(a, \rho(b, x)) = \rho(ab, x)$, $\rho(e, x) = x$ (where e is the neutral element of G) and $\rho_a : x \in V \mapsto \rho(a, x) \in V$ is linear.

Let $(V, \langle \cdot | \cdot \rangle)$ be a Euclidean space (i.e. a real vector space of finite dimension endowed with an inner product), $V \neq \{0\}$. We denote $\|\cdot\|$ the associated norm. The set $O(V) = \{f \in \text{GL}(V) | \forall x \in V, \|f(x)\| = \|x\|\}$ is the subgroup of isometries of $\text{GL}(V)$. Let $\rho : G \rightarrow O(V)$ be a representation of a group G acting isometrically on V .

3.2.1 Preliminaries

3.2.1.1 Vocabulary of representation theory

The following terminology is neither unique in representation theory nor exclusive to this branch of mathematics. We use the term “module” for simplicity, in reference to “ G -module” which is frequent. (In particular, a “module” here does not designate an abelian group $(M, +)$ endowed with a monoid action $(R, \cdot) \times M \rightarrow M$ that is distributive on additions, given a ring $(R, +, \cdot)$.)

Definition 3.1 (Vocabulary of representation theory)

- A (G -)module is a vector space W on which G acts continuously. For instance, the vector space V introduced above is a G -module. We omit the group G when it is clear from the context.
- A *submodule* of W is a G -stable subvector space of W .
- A *module homomorphism* (resp. *endomorphism*, *isomorphism*, *automorphism*) is a ρ -equivariant linear map (resp. endomorphism, isomorphism, automorphism) between modules. Given modules W, W' , we denote $W \simeq W'$ when there exists a module isomorphism from W to W' and we say that W and W' are isomorphic modules.
- A module W is *irreducible* if it has no other submodules than $\{0\}$ and W . Otherwise, it is *reducible*.
- A module W is *completely reducible* if it is the direct sum of irreducible modules.

From now on, we assume that V is completely reducible. For example, this is the case when G is finite [Artin, 2011, Maschke's Theorem 10.2.10] or when ρ is unitary and continuous [Artin, 2011, Corollary 10.3.5].

Note that if W is a submodule of V , then W^\perp (where orthogonality refers to the inner product $\langle \cdot | \cdot \rangle$) is also a submodule of V . Indeed, if $x \in W^\perp$ and $a \in G$, then for all $y \in W$, $\rho(a^{-1})(y) \in W$ because W is a module so $\langle \rho(a)(x) | y \rangle = \langle x | \rho(a^{-1})(y) \rangle = 0$, so $\rho(a)(x) \in W^\perp$. Therefore, we can assume that the decomposition of V into irreducible submodules is orthogonal.

3.2.1.2 The big picture

We split the method into three steps:

1. Transform the problem of finding all ρ -invariant inner products on V into finding all ρ -equivariant automorphisms of V .
2. Find all ρ -equivariant automorphisms of V using representation theory.
3. Go back to the initial problem.

Before the general case (V completely reducible), it is natural to start with the particular case where V is irreducible (Section 3.2.3.1). Then, we explain why the general case does not reduce to the irreducible case but only to an intermediate case (Section 3.2.3.2). Therefore, we treat this intermediate case (Section 3.2.3.3) and we conclude with the general case (Section 3.2.3.4). So we do the 1st step globally (Section 3.2.2), then we do the 2nd and 3rd steps for each case (Section 3.2.3).

3.2.2 Step 1: from invariant inner products to equivariant automorphisms

Let φ be an inner product on V . We introduce the musical isomorphisms of φ :

- “Flat” $\flat_\varphi : x \in V \mapsto \varphi(x, \cdot) \in V^*$ which is a linear isomorphism by the Riesz representation theorem,

- “Sharp” $\#_\varphi = \flat_\varphi^{-1} : V^* \longrightarrow V$ the inverse linear isomorphism.

We denote $\flat = \flat_{\langle \cdot | \cdot \rangle}$ and $\# = \#_{\langle \cdot | \cdot \rangle}$. Then φ is a ρ -invariant inner product on V if and only if $f_\varphi := \# \circ \flat_\varphi : V \longrightarrow V$ is a ρ -equivariant automorphism of V . We retrieve φ from $f := f_\varphi$ by $\varphi(x, y) = \langle f(x) | y \rangle$ for all $x, y \in V$.

The core of the method is to determine the general form of f .

3.2.3 Steps 2 and 3: find all equivariant automorphisms and all invariant inner products

3.2.3.1 The particular case: V irreducible

We start with the simplest case where V is irreducible. The key result to characterize all ρ -equivariant automorphisms of V is Schur’s lemma.

Lemma 3.2 [Artin, 2011, Schur’s Lemma 10.7.6] Let V, W be irreducible modules.

1. A module homomorphism $f : V \longrightarrow W$ is either null or a module isomorphism.
2. A module endomorphism $f : V \longrightarrow V$ is a scaling, i.e. there exists $\alpha \in \mathbb{R}$ such that $f = \alpha \text{Id}_V$.

Indeed, the first statement holds because $f(V)$ is a submodule of W so $f(V) = \{0\}$ or W and $\ker(f)$ is a submodule of V so $\ker(f) = V$ or $\{0\}$. The second statement holds because if $\alpha \in \mathbb{C}$ is an eigenvalue of f , then $f - \alpha \text{Id}_V$ is not a module isomorphism so it is null by the first statement. Since V is a real vector space and $V \neq \{0\}$, α has to be in \mathbb{R} .

As a consequence, if V is an irreducible module, then:

- Module automorphisms on V are the non null scalings.
- Module isomorphisms between V and an irreducible module W are unique up to scaling. Indeed, if $f, f_0 : V \longrightarrow W$ are module isomorphisms, then $f_0^{-1} \circ f : V \longrightarrow V$ is a scaling, i.e. there exists $\alpha \in \mathbb{R}^*$ such that $f = \alpha f_0$.
- The ρ -invariant inner products on V are positive multiples of $\langle \cdot | \cdot \rangle$.

3.2.3.2 The general case does not reduce to the previous case

We continue with the case where V is completely reducible. Therefore, let V_1, \dots, V_m be irreducible submodules of V such that $V = V_1 \oplus \dots \oplus V_m$. From the previous case, it is clear that the map $f : x = x_1 + \dots + x_m \in V \longmapsto \alpha_1 x_1 + \dots + \alpha_m x_m \in V$ with $\alpha_1, \dots, \alpha_m \in \mathbb{R}^*$ is a module automorphism of V . The question is: are they the only ones? The answer is no and the following consequence of Schur’s lemma allows to explain precisely why.

Lemma 3.3 (Consequence of Schur’s lemma on the irreducible decomposition) We group V_1, \dots, V_m by classes $\mathcal{C}^1, \dots, \mathcal{C}^p$ of isomorphic irreducible modules. Each class \mathcal{C}^k is of the form $\mathcal{C}(W) = \{j \in \{1, \dots, k\} | V_j \simeq W\}$ for $W \in \{V_1, \dots, V_m\}$. The decomposition becomes $V = \mathcal{V}^1 \oplus \dots \oplus \mathcal{V}^p$ with $\mathcal{V}^k = \bigoplus_{i \in \mathcal{C}^k} V_i$. Let $f : V \longrightarrow V$ be a module automorphism. Then $f(\mathcal{V}^k) = \mathcal{V}^k$ for all $k \in \{1, \dots, p\}$.

Proof. For $i, j \in \{1, \dots, m\}$, let $f_{ij} = \text{proj}_{V_j} \circ f|_{V_i} : V_i \rightarrow V_j$. Let $k \in \{1, \dots, p\}$ and let $i \in \{1, \dots, n\}$ such that $V_i \in \mathcal{C}^k$. By Schur's lemma, for all $j \in \{1, \dots, n\}$, f_{ij} is null or it is an isomorphism. Since f is an isomorphism, there exists $j \in \{1, \dots, m\}$ such that f_{ij} is non null, thus an isomorphism. Hence $V_j \in \mathcal{C}^k$ so $V_j \subseteq \mathcal{V}^k$. Therefore, $f(V_i) = V_j \subseteq \mathcal{V}^k$. This inclusion is valid for all $V_i \in \mathcal{C}^k$ so $f(\mathcal{V}^k) \subseteq \mathcal{V}^k$ and $f(\mathcal{V}^k) = \mathcal{V}^k$ by equality of dimensions because f is bijective. \square

In other words, the study of ρ -invariant automorphisms of V cannot be reduced to the V_i 's but only to the \mathcal{V}^k 's: there is no reason that $f(V_i) = V_i$ for all $i \in \{1, \dots, m\}$ (unless all classes are singletons). So we need to study the case $V = \mathcal{V} \simeq mW$ with W irreducible, where mW is a notation for the direct sum of m irreducible modules isomorphic to W .

3.2.3.3 The intermediate case: $V \simeq mW$ with W irreducible

We assume that $V = V_1 \overset{\perp}{\oplus} \dots \overset{\perp}{\oplus} V_m$ where $V_1 \simeq \dots \simeq V_m$ are isomorphic irreducible modules. Let W be an irreducible module isomorphic to them, endowed with a G -invariant inner product $\langle \cdot | \cdot \rangle$. Let $\psi_i : V_i \rightarrow W$ be the unique module isomorphism which is an isometry. Indeed, the module isomorphism is unique up to scaling and the pullback of the inner product $\langle \cdot | \cdot \rangle$ onto V_i is necessarily a scaling of the restriction of $\langle \cdot | \cdot \rangle$ to V_i so there is a unique choice of ψ_i such that it is an isometry. For example, W can be taken as one of the V_i 's.

Let f be a module automorphism of V . We define the module endomorphisms $f_{ij} = \psi_j \circ \text{proj}_{V_j} \circ f \circ \psi_i^{-1} : W \rightarrow W$. By Schur's lemma, they are scalings: there exists $S_{ij} \in \mathbb{R}$ such that $f_{ij} = S_{ij} \text{Id}_W$. This defines a matrix $S = (S_{ij})_{1 \leq i, j \leq m} \in \text{Mat}(m)$. Then, f writes:

$$\begin{aligned}
 f(x) &= (\text{Id}_V \circ f)(x) \\
 &= \left(\sum_{j=1}^m \text{proj}_{V_j} \right) \circ f \left(\sum_{i=1}^m x_i \right) \\
 &= \sum_{i=1}^m \sum_{j=1}^m (\text{proj}_{V_j} \circ f)(x_i) \\
 &= \sum_{i=1}^m \sum_{j=1}^m (\psi_j^{-1} \circ f_{ij} \circ \psi_i)(x_i) \\
 &= \sum_{i=1}^m \sum_{j=1}^m S_{ij} \psi_j^{-1} \circ \psi_i(x_i). \tag{3.1}
 \end{aligned}$$

When f comes from an inner product φ as explained in Section 3.2.2, we have for all $x, y \in V$:

$$\begin{aligned}
 \varphi(x, y) &= \langle f(x) | y \rangle \\
 &= \sum_{i=1}^m \sum_{j=1}^m S_{ij} \langle \psi_j^{-1} \circ \psi_i(x_i) | y_j \rangle \\
 &= \sum_{i=1}^m \sum_{j=1}^m S_{ij} (\psi_i(x_i) | \psi_j(y_j)),
 \end{aligned}$$

because the V_i 's are orthogonal and $\psi_j : (V_j, \langle \cdot | \cdot \rangle) \longrightarrow (W, \langle \cdot | \cdot \rangle)$ is an isometry. This implies that S is symmetric and positive definite. Indeed, let $w \in W \setminus \{0\}$, $a \in \mathbb{R}^m$ and $x = \sum_{i=1}^k a_i \psi_i^{-1}(w)$. Then:

- if $a = (1, \dots, 1)$, then $\varphi(x_i, x_j) = S_{ij} \|w\|^2$ and by symmetry of φ , we have $\varphi(x_i, x_j) = \varphi(x_j, x_i) = S_{ji} \|w\|^2$ so $S_{ij} = S_{ji}$,
- we have $\varphi(x, x) = \sum_{i,j} S_{ij} a_i a_j \|w\|^2$ so for all $a \in \mathbb{R}^m \setminus \{0\}$, we have $\sum_{i,j} S_{ij} a_i a_j > 0$.

Conversely, if $S \in \text{Sym}^+(m)$, then $\varphi(x, y) = \langle f(x) | y \rangle$ defines an inner product. It is clearly symmetric and if $x \neq 0$, there exists $w \in W \setminus \{0\}$ and $a \in \mathbb{R}^m \setminus \{0\}$ such that $x = \sum_{i=1}^k a_i \psi_i^{-1}(w)$ so the equality above proves that $\varphi(x, x) > 0$ (for the existence of w and a , take $i \in \{1, \dots, p\}$ such that $x_i \neq 0$ and define $w = \psi_i(x_i)$ and $a_j = \frac{\psi_j(x_j)}{\psi_i(x_i)}$ for $j \in \{1, \dots, m\}$).

So f is a module isomorphism of $V = V_1 \overset{\perp}{\oplus} \dots \overset{\perp}{\oplus} V_m$ if and only if there exists $S \in \text{Sym}^+(m)$ such that f writes as in Equation (3.1). Now we have all the ingredients to state the global result.

3.2.3.4 The general case: V completely reducible

Theorem 3.4 (General form of a ρ -invariant inner product on V) Let $V = \bigoplus_{k=1}^p \mathcal{V}^k$, with $\mathcal{V}^k = \bigoplus_{i=1}^{m_k} V_i^k$, be an orthogonal decomposition where $V_1^k \simeq \dots \simeq V_{m_k}^k$ are irreducible modules. For all $k \in \{1, \dots, p\}$, let $(W^k, (\cdot | \cdot)_k)$ be a Euclidean space and $\psi_i^k : V_i^k \longrightarrow W^k$ be the unique module isomorphism which is an isometry. Then, an inner product φ on V is ρ -invariant if and only if there exist p SPD matrices $S^k \in \text{Sym}^+(m_k)$ for $k \in \{1, \dots, p\}$ such that for all $x = \sum_{k=1}^p \sum_{i=1}^{m_k} x_i^k \in V$ and $y = \sum_{k=1}^p \sum_{i=1}^{m_k} y_i^k \in V$:

$$\varphi(x, y) = \sum_{k=1}^p \sum_{1 \leq i, j \leq m_k} S_{ij}^k (\psi_i^k(x_i^k) | \psi_j^k(y_j^k))_k. \quad (3.2)$$

The number of parameters is $\sum_{k=1}^p \frac{m_k(m_k+1)}{2}$ and the number of positivity constraints is $\sum_{k=1}^p m_k$.

Proof. We assemble the pieces of demonstration of the previous sections together. Let φ be a ρ -invariant inner product on V . Then, the map $f = \# \circ \flat_\varphi : V \longrightarrow V$ is a module automorphism (Section 3.2.2). Hence for all $k \in \{1, \dots, p\}$, $f(\mathcal{V}^k) = \mathcal{V}^k$ (Section 3.2.3.2). Therefore, there exists a matrix $S^k \in \text{Sym}^+(m_k)$ such that for all $x^k, y^k \in \mathcal{V}^k$, $\langle f(x^k) | y^k \rangle = \sum_{i=1}^{m_k} \sum_{j=1}^{m_k} S_{ij}^k (\psi_i^k(x_i^k) | \psi_j^k(y_j^k))_k$ (Section 3.2.3.3). Hence, the inner product φ writes:

$$\begin{aligned} \varphi(x, y) &= \langle f(x) | y \rangle \\ &= \sum_{k=1}^p \langle f(x^k) | y^k \rangle \\ &= \sum_{k=1}^p \sum_{i=1}^{m_k} \sum_{j=1}^{m_k} S_{ij}^k (\psi_i^k(x_i^k) | \psi_j^k(y_j^k))_k. \end{aligned} \quad (3.3)$$

Conversely, this bilinear form is an inner product since it is the sum of inner products on the \mathcal{V}^k 's (Section 3.2.3.3) which are supplementary. It is ρ -invariant because the orthogonal

projections are equivariant (because $\langle \cdot | \cdot \rangle$ is invariant), the ψ_i 's are equivariant and the inner products $(\cdot | \cdot)_k$ are invariant. So this bilinear form is a ρ -invariant inner product on V . \square

Note that the choice of the inner product $(\cdot | \cdot)_k$ on W^k instead of $\lambda(\cdot | \cdot)_k$ with $\lambda > 0$ does not affect the general form of the inner product: it suffices to replace $S^k \in \text{Sym}^+(m_k)$ by $\lambda S^k \in \text{Sym}^+(m_k)$. Neither does the choice of the isometric parameterization $\psi_i^k : (V_i^k, \langle \cdot | \cdot \rangle) \rightarrow (W^k, (\cdot | \cdot)_k)$ instead of $\lambda_i \psi_i^k$ with $\lambda_i > 0$ for all $i \in \{1, \dots, m_k\}$: it suffices to replace $S^k \in \text{Sym}^+(m_k)$ by $\Lambda S^k \Lambda \in \text{Sym}^+(m_k)$ where $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_{m_k})$.

3.3 Examples

We recall that this theorem is based on the complete reducibility of V . To be applied on examples, it means that we have to find explicit irreducible decompositions. This is our main task in this section. From now on, we apply the previous theorem to several situations. The groups we focus on are:

- the orthogonal group $O(n) = \{U \in \text{Mat}(n) | UU^\top = I_n\}$, which is a compact subgroup of the real general linear group $\text{GL}(n)$,
- the block orthogonal group:

$$O(I) = \left\{ \begin{pmatrix} R_1 & & 0 \\ & \ddots & \\ 0 & & R_p \end{pmatrix} \in O(n) | \forall i \in \{1, \dots, p\}, R_i \in O(k_i) \right\} \simeq O(k_1) \times \dots \times O(k_p),$$

which is a subgroup of $O(n)$ where $I = (k_1, \dots, k_p)$ is a ‘‘partition of n ’’, i.e. such that $k_1 + \dots + k_p = n$ with $k_1, \dots, k_p \in \mathbb{N}^*$,

- the permutation group $\mathfrak{S}(n)$, which is a finite subgroup of $O(n)$ denoting either the set of permutations σ of $\{1, \dots, n\}$ or the set of permutation matrices P_σ defined by $[P_\sigma]_{ij} = \delta_{i, \sigma(j)}$. (Beware that $\mathfrak{S}(n) \not\subseteq \text{SO}(n)$, especially transpositions are not in $\text{SO}(n)$.)

The Euclidean modules we focus on are:

- \mathbb{R}^n with the canonical inner product and the isometric action $(R, x) \in O(n) \times \mathbb{R}^n \mapsto Rx \in \mathbb{R}^n$,
- the vector space $\text{Mat}(n)$ of $n \times n$ squared real matrices endowed with the Frobenius inner product and the isometric congruence action $(R, X) \in O(n) \times \text{Mat}(n) \mapsto RMR^\top \in \text{Mat}(n)$, which is also the tangent space of the general linear group $\text{GL}(n)$,
- the Euclidean submodule $\text{Skew}(n) = \{X \in \text{Mat}(n) | X^\top = -X\}$ of skew-symmetric matrices, which is the tangent space at I_n of the orthogonal group $O(n)$,
- the Euclidean submodule $\text{Sym}(n) = \{X \in \text{Mat}(n) | X^\top = X\}$ of symmetric matrices, which is the tangent space of the manifold of Symmetric Positive Definite (SPD) matrices $\text{Sym}^+(n) = \{X \in \text{Sym}(n) | X > 0\}$ (where $>$ denotes the strict Löwner order),
- (only for the action of the permutation group) the Euclidean submodule $\text{Hol}(n) = \{X \in \text{Sym}(n) | \text{Diag}(X) = 0\}$, which is the tangent space of the manifold of full-rank correlation matrices $\text{Cor}^+(n) = \{\Sigma \in \text{Sym}^+(n) | \text{Diag}(\Sigma) = I_n\}$

We denote (e_1, \dots, e_n) the canonical basis of \mathbb{R}^n and $E_{ij} = e_i e_j^\top + e_j e_i^\top \in \text{Sym}(n)$ which form a basis of $\text{Sym}(n)$ for $1 \leq i \leq j \leq n$. These examples being trivial for $n = 1$, we consider $n \geq 2$. We structure each example as follows:

- Find the **Direct sum** and compute the projections,
- Prove the stability and **Irreducibility** of the terms in the decomposition,
- Determine classes of **Isomorphic modules**,
- Write the **General form** of invariant inner product/quadratic form.

For readability, some proofs are moved in appendix of the thesis.

3.3.1 Invariant inner products on \mathbb{R}^n

In this section, we illustrate the previous theorem on simple examples. The irreducible orthogonal decompositions of \mathbb{R}^n with respect to the action of $O(n)$, $\mathfrak{S}(n)$ and $O(I)$ are given in Table 3.1 for $I = (k_1, \dots, k_p)$. The following notations are used:

- $\mathbf{1} \in \mathbb{R}^n$ is the vector with all entries equal to one,
- $\mathbb{R}_0^n = \text{span}(\mathbf{1})^\perp = \{x \in \mathbb{R}^n \mid \text{sum}(x) := \sum_{i=1}^n x_i = x^\top \mathbf{1} = 0\}$,
- \mathbb{R}^{k_i} abusively denotes the subspace of \mathbb{R}^n where the $(k_1 + \dots + k_{i-1})$ first entries and the $(k_{i+1} + \dots + k_p)$ last entries are null,
- $p_{k_i} : \mathbb{R}^n \longrightarrow \mathbb{R}^{k_i}$ denotes the canonical projection,
- $O_n(k_i) \simeq O(k_i)$ denotes the subgroup of $O(I)$ acting on \mathbb{R}^{k_i} .

Group	Decomposition
$O(n)$	\mathbb{R}^n
$O(I)$	$\bigoplus_{i=1}^p \mathbb{R}^{k_i}$
$\mathfrak{S}(n)$	$\mathbb{R}_0^n \oplus \text{span}(\mathbf{1})$

Table 3.1: Irreducible orthogonal decompositions of \mathbb{R}^n with respect to several groups.

Direct sum The decompositions are clearly direct orthogonal sums. The orthogonal projections on \mathbb{R}_0^n and $\text{span}(\mathbf{1})$ are $x \in \mathbb{R}^n \mapsto x - \frac{\langle x, \mathbf{1} \rangle}{n} \mathbf{1} \in \mathbb{R}_0^n$ and $x \in \mathbb{R}^n \mapsto \frac{\langle x, \mathbf{1} \rangle}{n} \mathbf{1} \in \text{span}(\mathbf{1})$.

Irreducibility The action of $O(n)$ on \mathbb{R}^n is transitive so \mathbb{R}^n is stable and irreducible with respect to the action of $O(n)$. Therefore, \mathbb{R}^{k_i} is stable and irreducible with respect to $O_n(k_i)$, thus to $O(I)$.

Regarding the action of $\mathfrak{S}(n)$, the vector spaces \mathbb{R}_0^n and $\text{span}(\mathbf{1})$ are stable and $\text{span}(\mathbf{1})$ is irreducible because of dimension 1. Moreover, let $W \subseteq \mathbb{R}_0^n$ be a submodule such that $W \neq \{0\}$. Let $x \in W$, then there exist $i < j$ such that $x_i \neq x_j$. Hence, applying the permutation $\sigma = (i, j)$, the vector $e_{ij} = e_i - e_j = \frac{1}{x_i - x_j}(x - \sigma \cdot x) \in W$. Hence with a permutation $\tau \in \mathfrak{S}(n)$ such that $\tau(i) = k$ and $\tau(j) = k + 1$, we have $e_{k, k+1} = \tau \cdot e_{ij} \in W$ for all $k \in \{1, \dots, n - 1\}$. As $(e_{k, k+1})_{k \in \{1, \dots, n-1\}}$ is a basis of \mathbb{R}_0^n , we have $\mathbb{R}_0^n \subseteq W$ so $W = \mathbb{R}_0^n$ and \mathbb{R}_0^n is irreducible.

Isomorphic modules Regarding $\mathfrak{S}(n)$, the two modules are clearly not isomorphic for $n \geq 3$ because they have different dimensions. For $n = 2$, $\mathbb{R}_0^2 = \text{span}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and the permutation $(1, 2)$ does not act the same way on $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\mathbf{1}$. Indeed, $(1, 2) \cdot \mathbf{1} = \mathbf{1}$ and $(1, 2) \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. So \mathbb{R}_0^2 and $\text{span}(\mathbf{1})$ are not isomorphic as modules even if they are isomorphic as vector spaces.

Analogously, even if $k_i = k_j$ with $i \neq j$, the group $O(I)$ does not act the same way on \mathbb{R}^{k_i} and \mathbb{R}^{k_j} . Thus, the modules are not isomorphic.

General form Thus, the $O(n)$ -invariant inner products on \mathbb{R}^n are positive scalings of the canonical one. The $O(I)$ -invariant and $\mathfrak{S}(n)$ -invariant inner products on \mathbb{R}^n are respectively given for all $x \in \mathbb{R}^n$ by:

$$\varphi(x, x) = \sum_{i=1}^p \alpha_{k_i} \|p_{k_i}(x)\|^2, \quad (3.4)$$

$$\varphi(x, x) = \alpha_1 \left\| x - \frac{\langle x | \mathbf{1} \rangle}{n} \mathbf{1} \right\|^2 + \alpha_2 \frac{\langle x | \mathbf{1} \rangle^2}{n} = \alpha \|x\|^2 + \beta \langle x | \mathbf{1} \rangle^2, \quad (3.5)$$

with $\alpha_{k_1}, \dots, \alpha_{k_p} > 0$, $\alpha_1 = \alpha > 0$ and $\alpha_2 = \alpha + n\beta > 0$.

3.3.2 $O(n)$ -invariant inner products on $\text{Mat}(n)$

In this section, we illustrate the theorem with the congruence action of the orthogonal group $O(n)$ on squared matrices $\text{Mat}(n)$. This example is also well known although we did not find a reference book containing it. It allows us to prepare the following sections.

3.3.2.1 $O(n)$ -invariant inner products on $\text{Sym}(n)$

The irreducible decomposition of $\text{Sym}(n)$ with respect to the congruence action of $O(n)$ (and $\text{SO}(n)$) is $\text{Sym}(n) = \text{Sym}^0(n) \oplus \text{span}(I_n)$ where $\text{Sym}^0(n) = \text{Sym}(n) \cap \ker \text{tr}$.

Direct sum The decomposition is clearly a direct irreducible sum. The orthogonal projections are $X \in \text{Sym}(n) \mapsto X - \frac{\text{tr}(X)}{n} I_n \in \text{Sym}^0(n)$ and $X \in \text{Sym}(n) \mapsto \frac{\text{tr}(X)}{n} I_n \in \text{span}(I_n)$.

Irreducibility Firstly, $\text{Sym}^0(n)$ and $\text{span}(I_n)$ are stable and $\text{span}(I_n)$ is irreducible because of dimension 1. Secondly, let $W \subseteq \text{Sym}^0(n)$ be a submodule such that $W \neq \{0\}$. Let $X \in W$, $X \neq 0$. Then there exist $P \in \text{SO}(n)$ and $D \in \text{Diag}(n) \cap \ker \text{tr}$ such that $X = PDP^T$ so $D \in W \setminus \{0\}$. If $n = 2$, then $d_1 \neq 0$ and $E_{11} - E_{22} = \frac{2}{d_1} D \in W$. By applying the rotation $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \text{SO}(2)$, we get $E_{12} \in W$ so $\text{Sym}^0(2) = W$ and $\text{Sym}^0(2)$ is irreducible. Let us prove it for $n \geq 3$.

There exist $i < j$ such that $d_i \neq d_j$ because $\text{tr}(D) = 0$ and $D \neq 0$. It is tempting to apply the same method as in the example of $\mathfrak{S}(n)$ acting on \mathbb{R}^n to conclude that $\text{Diag}(n) \cap \ker \text{tr} \subseteq W$. However, permutations are not all in $\text{SO}(n)$, especially transpositions do not belong to $\text{SO}(n)$. So we need to find another way. We define the matrix $R = R^{ij}(\theta) \in \text{SO}(n)$ by $R_{ii} = R_{jj} = \cos \theta$, $R_{ij} = -R_{ji} = \sin \theta$, $R_{kk} = 1$ for $k \notin \{i, j\}$ and zero elsewhere.

With $\theta = \frac{\pi}{4}$, the submatrix $\begin{pmatrix} d_i & 0 \\ 0 & d_j \end{pmatrix}$ is transformed into $\frac{1}{2} \begin{pmatrix} d_i + d_j & d_j - d_i \\ d_j - d_i & d_i + d_j \end{pmatrix}$. Hence $D_1 = \frac{1}{d_j - d_i}(D - RDR^\top) = \frac{1}{2}(e_i e_i^\top + e_i e_j^\top + e_j e_i^\top - e_j e_j^\top) \in W$. Since $n \geq 3$, we can take a third index $k \neq i, j$ and act by the diagonal matrix $\text{diag}(1, \dots, 1, -1, 1, \dots, 1, -1, 1, \dots, 1) \in \text{SO}(n)$ where the -1 are at indexes i and k . We obtain $D_2 = \frac{1}{2}(e_i e_i^\top - e_i e_j^\top - e_j e_i^\top - e_j e_j^\top)$. Hence $e_i e_i^\top - e_j e_j^\top = D_1 + D_2 \in W$. Then for all $k \in \{1, \dots, n-1\}$, it is easy to find even permutations sending i on k and j on $k+1$ so that $e_k e_k^\top - e_{k+1} e_{k+1}^\top \in W$ and $\text{Diag}(n) \cap \ker \text{tr} \subseteq W$. Moreover, $E_{ij} = D_1 - D_2 \in W$ so with appropriate even permutations, we can prove that all E_{kl} with $k \neq l$ belongs to W . Hence, $\text{Sym}^0(n) \subseteq W$ and $\text{Sym}^0(n)$ is irreducible.

Isomorphic modules The two modules are clearly not isomorphic because they have different dimensions: $\dim(\text{Sym}^0(n)) = \frac{n(n+1)}{2} - 1 > 1 = \dim \text{span}(I_n)$ for $n \geq 2$.

General form Thus, the $\text{O}(n)$ -invariant (and $\text{SO}(n)$ -invariant) inner products on $\text{Sym}(n)$ are given for all $X \in \text{Sym}(n)$ by:

$$\varphi(X, X) = \alpha_1 \left\| X - \frac{\text{tr}(X)}{n} I_n \right\|^2 + \alpha_2 \frac{\text{tr}(X)^2}{n} = \alpha \text{tr}(X^2) + \beta \text{tr}(X)^2, \quad (3.6)$$

with $\alpha_1 = \alpha$ and $\alpha_2 = \alpha + n\beta > 0$.

3.3.2.2 $\text{O}(n)$ -invariant inner products on $\text{Mat}(n)$

The irreducible decomposition of $\text{Mat}(n)$ with respect to the congruence action of $\text{O}(n)$ is

$$\text{Mat}(n) = \underbrace{\text{Sym}^0(n) \oplus \text{Skew}(n)}_{\ker \text{tr}} \oplus \text{span}(I_n).$$

Direct sum The decomposition is clearly a direct orthogonal sum. The orthogonal projections are $X \in \text{Mat}(n) \mapsto \frac{X+X^\top}{2} - \frac{\text{tr}(X)}{n} I_n \in \text{Sym}^0(n)$, $X \in \text{Mat}(n) \mapsto \frac{X-X^\top}{2} \in \text{Skew}(n)$ and $X \in \text{Mat}(n) \mapsto \frac{\text{tr}(X)}{n} I_n \in \text{span}(I_n)$.

Irreducibility Let us prove that $\text{Skew}(n)$ is $\text{O}(n)$ -irreducible. Indeed, let $W \neq \{0\}$ be a submodule of $\text{Skew}(n)$ and let $M \in \text{Skew}(n)$, $M \neq 0$. Then, M is orthogonally congruent

$$\text{to } B = \begin{pmatrix} B_{\lambda_1} & & & 0 \\ & \ddots & & \\ & & B_{\lambda_p} & \\ 0 & & & \mathbf{0}_{n-2p} \end{pmatrix} \in \text{Skew}(n) \text{ where } \pm i\lambda_1, \dots, \pm i\lambda_p \in i\mathbb{R}^* \text{ (and } 0 \text{ if } n \neq 2p)$$

are its complex eigenvalues and $B_\lambda = \begin{pmatrix} 0 & \lambda \\ -\lambda & 0 \end{pmatrix} \in \text{Skew}(2)$. Applying the permutation

$\sigma = (1, 2)$, we get $e_1 e_2^\top - e_2 e_1^\top = \begin{pmatrix} B_1 & 0 \\ 0 & 0 \end{pmatrix} = \frac{1}{2\lambda_1}(M - P_\sigma M P_\sigma^\top) \in W$. With appropriate permutations, all $e_i e_j^\top - e_j e_i^\top \in W$ so $W = \text{Skew}(n)$ and $\text{Skew}(n)$ is $\text{O}(n)$ -irreducible.

Isomorphic modules For $n \geq 3$, the modules have different dimensions: $\dim(\text{Sym}^0(n)) = \frac{n(n+1)}{2} - 1 > \dim \text{Skew}(n) = \frac{n(n-1)}{2} > \dim \text{span}(I_n) = 1$. So they are clearly not isomorphic. For $n = 2$, $\dim(\text{Sym}^0(2)) = 2 > 1 = \dim \text{Skew}(2) = \dim \text{span}(I_2)$, but the permutation $(1, 2)$ acts differently on $e_1 e_2^\top - e_2 e_1^\top$ and I_2 so the modules are not isomorphic either.

General form Thus, the $O(n)$ -invariant inner products on $\text{Mat}(n)$ are given for all $X \in \text{Mat}(n)$ by:

$$\begin{aligned} \varphi(X, X) &= \alpha_1 \left\| \frac{X + X^\top}{2} - \frac{\text{tr}(X)}{n} I_n \right\|^2 + \alpha_2 \left\| \frac{X - X^\top}{2} \right\|^2 + \alpha_3 \frac{\text{tr}(X)^2}{n} \\ &= \alpha_1 \text{tr} \left(\frac{X^2 + (X^\top)^2 + 2XX^\top}{4} + \frac{\text{tr}(X)^2}{n^2} I_n - \frac{\text{tr}(X)}{n} (X + X^\top) \right) \\ &\quad + \frac{\alpha_2}{4} \text{tr}(XX^\top + X^\top X - X^2 - (X^\top)^2) + \alpha_3 \frac{\text{tr}(X)^2}{n} \\ &= \alpha \text{tr}(XX^\top) + \beta \text{tr}(X^2) + \gamma \text{tr}(X)^2, \end{aligned}$$

with $\alpha = \frac{\alpha_1 + \alpha_2}{2}$, $\beta = \frac{\alpha_1 - \alpha_2}{2}$ and $\gamma = \frac{\alpha_3 - \alpha_1}{n}$. The inverse relations are $\alpha_1 = \alpha + \beta > 0$, $\alpha_2 = \alpha - \beta > 0$ and $\alpha_3 = \alpha + \beta + n\gamma > 0$, as expected.

3.3.3 $O(I)$ -invariant inner products on $\text{Mat}(n)$

In this section, we consider the congruence action of the block orthogonal group $O(I)$ on skew-symmetric matrices $\text{Skew}(n)$, on symmetric matrices $\text{Sym}(n)$ and finally on squared matrices $\text{Mat}(n)$. In particular, $O(I)$ -invariant inner products on $\text{Skew}(n) = T_{I_n} O(n)$ characterize $O(n)$ -invariant Riemannian metrics on the homogeneous space of flags of signature I denoted $\text{Flag}(I) = O(n)/O(I)$. In this context, the signature is sometimes called the nationality of the flag.

We recall that $I = (k_1, \dots, k_p)$ is a partition of n , i.e. $k_1, \dots, k_p \in \mathbb{N}^*$ with $k_1 + \dots + k_p = n$. We consider the group of block orthogonal matrices of signature I :

$$O(I) = \left\{ \begin{pmatrix} R_1 & & 0 \\ & \ddots & \\ 0 & & R_p \end{pmatrix} \in O(n) \mid \forall i \in \{1, \dots, p\}, R_i \in O(k_i) \right\} \simeq O(k_1) \times \dots \times O(k_p).$$

We denote $O_n(k_1) = \left\{ \begin{pmatrix} R & 0 \\ 0 & I_{n-k_1} \end{pmatrix} \in O(I) \mid R \in O(k_1) \right\} \simeq O(k_1)$ and analogously $O_n(k_i)$ for $i \in \{1, \dots, p\}$. They are subgroups of $O(I)$. In the same block-wise spirit, we also denote for $i \neq j$:

- $\text{Mat}_n(k_1) = \left\{ \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} \in \text{Mat}(n) \mid R \in \text{Mat}(k_1) \right\}$ with appropriate sizes of null matrices, and analogously $\text{Mat}_n(k_i) \simeq \text{Mat}(k_i)$ for $i \in \{1, \dots, p\}$,
- $\text{Mat}_n(k_1, k_2) = \left\{ \begin{pmatrix} 0 & X & 0 \\ 0 & 0 & 0 \end{pmatrix} \in \text{Mat}(n) \mid X \in \text{Mat}(k_1, k_2) \right\}$ again with appropriate sizes of null matrices, and analogously $\text{Mat}_n(k_i, k_j) \simeq \text{Mat}(k_i, k_j)$ for $i, j \in \{1, \dots, p\}$,

- $\text{Skew}_n(k_i) = \text{Skew}(n) \cap \text{Mat}_n(k_i) \simeq \text{Skew}(k_i)$,
- $\text{Skew}_n(k_i, k_j) = \text{Skew}(n) \cap (\text{Mat}_n(k_i, k_j) \oplus \text{Mat}_n(k_j, k_i)) \simeq \text{Mat}_n(k_i, k_j)$,
- $\text{Sym}_n(k_i) = \text{Sym}(n) \cap \text{Mat}_n(k_i)$,
- $\text{Sym}_n(k_i, k_j) = \text{Sym}(n) \cap (\text{Mat}_n(k_i, k_j) \oplus \text{Mat}_n(k_j, k_i)) \simeq \text{Mat}_n(k_i, k_j)$,
- $\text{Sym}_n^0(k_i) = \text{Sym}_n(k_i) \cap \ker \text{tr}$,
- $\text{Sym}_n^0(k_i, k_j) = \text{Sym}_n(k_i, k_j) \cap \ker \text{tr}$.

Whenever $i = j$, we also denote $\text{Mat}_n(k_i, k_i) = \text{Mat}_n(k_i)$, $\text{Sym}_n(k_i, k_i) = \text{Sym}_n(k_i)$ and $\text{Skew}_n(k_i, k_i) = \text{Skew}_n(k_i)$.

3.3.3.1 $O(I)$ -invariant inner products on $\text{Skew}(n)$

We denote $X \in \text{Skew}(n) \mapsto X^{ij} \in \text{Skew}_n(k_i, k_j)$ the canonical projection for $i, j \in \{1, \dots, p\}$.

Example 3.5 ($O(I)$ -invariant inner products on $\text{Skew}(n)$) The irreducible decomposition of $\text{Skew}(n)$ with respect to the congruence action of $O(I)$ is $\text{Skew}(n) = \bigoplus_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \text{Skew}_n(k_i) \oplus \bigoplus_{1 \leq i < j \leq p} \text{Skew}_n(k_i, k_j)$. The decomposition is orthogonal for the Frobenius inner product. The $O(I)$ -invariant inner products on $\text{Skew}(n)$ are given for all $X \in \text{Skew}(n)$ by:

$$\varphi(X, X) = \sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \alpha_{ii} \|X^{ii}\|^2 + \sum_{1 \leq i < j \leq p} \alpha_{ij} \|X^{ij}\|^2, \quad (3.7)$$

where $\alpha_{ij} > 0$ for all $1 \leq i \leq j \leq p$.

See the proof of Example 3.5 in Section 11.2.

3.3.3.2 $O(n)$ -invariant Riemannian metrics on $\text{Flag}(I)$

The homogeneous manifold of flags of signature I is $\text{Flag}(I) = O(n)/O(I) = \{UO(I), U \in O(n)\}$. Since $T_{I_n} O(n) = \text{Skew}(n)$ and $T_{I_n} O(I) = \text{Skew}(I) := \bigoplus_{1 \leq i \leq p} \text{Skew}_n(k_i) = \bigoplus_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \text{Skew}_n(k_i)$, the tangent space $T_{O(I)} \text{Flag}(I)$ can be identified to $\mathfrak{m} := \bigoplus_{1 \leq i < j \leq p} \text{Skew}_n(k_i, k_j)$. Therefore, a Riemannian metric g on $\text{Flag}(I)$ is $O(n)$ -invariant if and only if $g_{O(I)} : T_{O(I)} \text{Flag}(I) \times T_{O(I)} \text{Flag}(I) \rightarrow \mathbb{R}$ is $O(I)$ -invariant. Such inner products are characterized by the $p(p-1)/2$ positive coefficients $\alpha_{ij} > 0$ for $1 \leq i < j \leq p$. From Equation (3.7), the general $O(n)$ -invariant Riemannian metric on $\text{Flag}(I)$ is given for all $X \in T_{UO(I)} \text{Flag}(I) \simeq U\mathfrak{m}$ by:

$$g_{UO(I)}(X, X) = \sum_{1 \leq i < j \leq p} \alpha_{ij} \|(U^{-1}X)^{ij}\|^2. \quad (3.8)$$

The manifold of flags $\text{Flag}(I)$ endowed with one of these Riemannian metrics is a Riemannian symmetric space. They all share the same Levi-Civita connection, thus the same geodesics.

We retrieve a result on Grassmannians $\text{Gr}(n, k) = \text{Flag}(k, n-k) = O(n)/(O(k) \times O(n-k))$ (for which $p = 2$): there exists a unique $O(n)$ -invariant Riemannian metric on $\text{Gr}(n, k)$ up to a scaling factor $\alpha_{12} > 0$.

3.3.3.3 $O(I)$ -invariant inner products on $\text{Sym}(n)$

We denote $X \in \text{Sym}(n) \mapsto X^{ij} \in \text{Sym}_n(k_i, k_j)$ the canonical projection for $i, j \in \{1, \dots, p\}$.

Example 3.6 ($O(I)$ -invariant inner products on $\text{Sym}(n)$) The irreducible decomposition of $\text{Sym}(n)$ with respect to the congruence action of $O(I)$ is $\text{Sym}(n) = \bigoplus_{i=1}^p \text{span}(I_n^{ii}) \oplus \bigoplus_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \text{Sym}_n^0(k_i) \oplus \bigoplus_{1 \leq i < j \leq p} \text{Sym}_n(k_i, k_j)$. The decomposition is orthogonal for the Frobenius inner product and \mathcal{V} . The $O(I)$ -invariant inner products on $\text{Sym}(n)$ are given for all $X \in \text{Sym}(n)$ by:

$$\varphi(X, X) = \sum_{1 \leq i, j \leq p} S_{ij} \frac{\text{tr}(X^{ii})\text{tr}(X^{jj})}{\sqrt{k_i k_j}} + \sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \alpha_{ii} \left\| X^{ii} - \frac{\text{tr}(X^{ii})}{k_i} I_n^{ii} \right\|^2 + \sum_{1 \leq i < j \leq p} \alpha_{ij} \|X^{ij}\|^2, \quad (3.9)$$

where $S \in \text{Sym}^+(p)$ and $\alpha_{ij} > 0$ for all $1 \leq i \leq j \leq p$.

See the proof of Example 3.6 in Section 11.2.

3.3.3.4 $O(I)$ -invariant inner products on $\text{Mat}(n)$

To distinguish between the projections $X \mapsto X^{ij}$ onto symmetric matrices and skew-symmetric matrices for $1 \leq i \leq j \leq p$, we denote:

- $X \in \text{Mat}(n) \mapsto X_{\text{sym}}^{ij} = \frac{1}{2}(X + X^\top)^{ij} \in \text{Sym}_n(k_i, k_j)$,
- $X \in \text{Mat}(n) \mapsto X_{\text{skew}}^{ij} = \frac{1}{2}(X - X^\top)^{ij} \in \text{Skew}_n(k_i, k_j)$.

Example 3.7 ($O(I)$ -invariant inner products on $\text{Mat}(n)$) The irreducible decomposition of $\text{Mat}(n)$ with respect to the congruence action of $O(I)$ is:

$$\begin{aligned} \text{Mat}(n) = & \bigoplus_{i=1}^p \text{span}(I_n^{ii}) \oplus \bigoplus_{1 \leq i < j \leq p} (\text{Skew}_n(k_i, k_j) \oplus \text{Sym}_n(k_i, k_j)) \\ & \oplus \bigoplus_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \text{Skew}_n(k_i) \oplus \bigoplus_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \text{Sym}_n^0(k_i) \end{aligned} \quad (3.10)$$

Direct sum It is the direct sum of the two previous decompositions.

Irreducibility We already know that all the terms are stable and irreducible.

Isomorphic modules We grouped $\text{Skew}_n(k_i, k_j)$ and $\text{Sym}_n(k_i, k_j)$ because they are isomorphic modules. According to the previous study of the action of $O(n)$ and $O(I)$ on $\text{Skew}(n)$ and $\text{Sym}(n)$, it is clear that there is no module isomorphism between other terms in the direct sum.

General form Hence, the $O(I)$ -invariant inner products on $\text{Mat}(n)$ are given for all $X \in \text{Mat}(n)$ by:

$$\begin{aligned} \varphi(X, X) = & \sum_{1 \leq i, j \leq p} S_{ij} \frac{\text{tr}(X^{ii})\text{tr}(X^{jj})}{\sqrt{k_i k_j}} + \sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \left(\alpha_{ii} \left\| X_{\text{sym}}^{ii} - \frac{\text{tr}(X^{ii})}{k_i} I_n \right\|^2 + \beta_{ii} \|X_{\text{skew}}^{ii}\|^2 \right) \\ & + \sum_{1 \leq i < j \leq p} \left(\alpha_{ij} \|X_{\text{sym}}^{ij}\|^2 + \beta_{ij} \|X_{\text{skew}}^{ij}\|^2 + \gamma_{ij} \langle X_{\text{sym}}^{ij} | X_{\text{skew}}^{ij} \rangle \right), \end{aligned} \quad (3.11)$$

where $S \in \text{Sym}^+(p)$, $\alpha_{ii}, \beta_{ii} > 0$ for all $i \in \{1, \dots, p\}$ such that $k_i \geq 2$, $\alpha_{ij} > 0$ and $\alpha_{ij}\beta_{ij} - \gamma_{ij}^2 > 0$ for all $1 \leq i < j \leq p$.

3.3.4 $\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$

In this section, we consider the congruence action of the group of permutation matrices $\mathfrak{S}(n) \equiv \{P_\sigma = (\delta_{i, \sigma(j)})_{1 \leq i, j \leq n} \in O(n), \sigma \in \mathfrak{S}(n)\}$ on squared matrices $\text{Mat}(n)$. We especially focus on the action on symmetric hollow matrices $\text{Hol}(n) = \text{Sym}(n) \cap \ker \text{Diag}$ (without isomorphic irreducible submodules) and symmetric matrices $\text{Sym}(n)$ (with isomorphic irreducible submodules). Since $\text{Hol}(n)$ and $\text{Sym}(n)$ are respectively the tangent spaces of full-rank correlation matrices $\text{Cor}^+(n) = \{\Sigma \in \text{Sym}^+(n) | \text{Diag}(\Sigma) = I_n\}$ and SPD matrices $\text{Sym}^+(n)$, this provides permutation-invariant Euclidean metrics on full-rank correlation and covariance matrices.

The case $n = 2$ is degenerate so we treat it here. We have:

$$\text{Mat}(2) = \text{span}(I_2) \oplus \text{span} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \text{span} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \oplus \text{span} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (3.12)$$

The two first terms are isomorphic modules with isotropy group $\mathfrak{S}(2)$, the two last terms are isomorphic modules with isotropy group $\{\text{Id}\}$. So the $\mathfrak{S}(2)$ -invariant inner products on $\text{Mat}(2)$ are given for all $X = \begin{pmatrix} x & y \\ z & t \end{pmatrix} \in \text{Mat}(2)$ by:

$$\begin{aligned} \varphi(X, X) = & \alpha (x + t)^2 + \beta (y + z)^2 + 2\gamma (x + t)(y + z) \\ & + \alpha' (x - t)^2 + \beta' (y - z)^2 + 2\gamma' (x - t)(y - z), \end{aligned} \quad (3.13)$$

where $\begin{pmatrix} \alpha & \gamma \\ \gamma & \beta \end{pmatrix} \in \text{Sym}^+(2)$ and $\begin{pmatrix} \alpha' & \gamma' \\ \gamma' & \beta' \end{pmatrix} \in \text{Sym}^+(2)$.

In the following, we assume that $n \geq 3$.

3.3.4.1 $\mathfrak{S}(n)$ -invariant inner products on $\text{Hol}(n)$, $n \geq 3$

We denote $\text{sum}(X) = \mathbf{1}^\top X \mathbf{1} = \sum_{i,j} X_{ij}$ the sum of entries of $X \in \text{Mat}(n)$ and analogously $\text{sum}(\mu) = \mathbf{1}^\top \mu$ for $\mu \in \mathbb{R}^n$.

Example 3.8 ($\mathfrak{S}(n)$ -invariant inner products on $\text{Hol}(n)$) The irreducible decomposition of $\text{Hol}(n)$ with respect to the congruence action of $\mathfrak{S}(n)$, i.e. of permutation matrices, is $\text{Hol}(n) = \ker \phi_1 \oplus \text{im } \psi_2 \oplus \text{im } \psi_3$ where:

1. $\phi_1 : X \in \text{Hol}(n) \mapsto X\mathbf{1} \in \mathbb{R}^n$,
2. $\psi_2 : \mu \in \mathbb{R}_0^n \mapsto \frac{1}{n-2}(\mu\mathbf{1}^\top + \mathbf{1}\mu^\top - 2\text{diag}(\mu)) \in \text{Hol}(n)$,
3. $\psi_3 : x \in \mathbb{R} \mapsto \frac{x}{n(n-1)}(\mathbf{1}\mathbf{1}^\top - I_n) \in \text{Hol}(n)$,

with ϕ_1 surjective, ψ_2 and ψ_3 injective. The corresponding orthogonal projections are:

- $p_3 : X \in \text{Hol}(n) \mapsto X_3 = \psi_3(\text{sum}(X)) \in \text{im } \psi_3$,
- $p_2 : X \in \text{Hol}(n) \mapsto X_2 = \psi_2((X - X_3)\mathbf{1}) \in \text{im } \psi_2$,
- $p_1 : X \in \text{Hol}(n) \mapsto X_1 = X - X_2 - X_3 \in \text{ker } \phi_1$.

The $\mathfrak{S}(n)$ -invariant inner products on $\text{Hol}(n)$ are given for all $X \in \text{Hol}(n)$ by:

$$\varphi(X, X) = \alpha_1 \text{tr}(X_1^2) + \alpha_2 \text{tr}(X_2^2) + \alpha_3 \text{tr}(X_3^2) \quad (3.14)$$

$$= \alpha \text{tr}(X^2) + \beta \text{sum}(X^2) + \gamma \text{sum}(X)^2, \quad (3.15)$$

where $\alpha_1 = \alpha > 0$, $\alpha_2 = \alpha + \frac{n-2}{2}\beta$ and $\alpha_3 = \alpha + (n-1)(\beta + n\gamma) > 0$.
See the proof of Example 3.8 in Section 11.2.

Remark 3.9 The case $n = 3$ is also degenerate because $\text{ker } \phi_1 = \{0\}$. The inner product writes $\varphi(X, X) = \beta \text{sum}(X^2) + \gamma \text{sum}(X)^2$ with $\beta > 0$ and $\beta + 3\gamma > 0$.

3.3.4.2 $\mathfrak{S}(n)$ -invariant inner products on $\text{Sym}(n)$, $n \geq 3$

Example 3.10 ($\mathfrak{S}(n)$ -invariant inner products on $\text{Sym}(n)$) The irreducible decomposition of $\text{Sym}(n)$ with respect to the congruence action of $\mathfrak{S}(n)$, i.e. of permutation matrices, is $\text{Sym}(n) = \underbrace{\text{ker } \phi_1 \oplus \text{im } \psi_2 \oplus \text{im } \psi_3}_{\text{Hol}(n)} \oplus \underbrace{\text{im } \psi_4 \oplus \text{im } \psi_5}_{\text{Diag}(n)} \simeq \text{ker } \phi_1 \oplus 2\mathbb{R}_0^n \oplus 2\mathbb{R}$ where ϕ_1, ψ_2, ψ_3 were defined in the previous section and:

4. $\psi_4 : \lambda \in \mathbb{R}_0^n \mapsto \text{diag}(\lambda) \in \text{Diag}(n)$,
5. $\psi_5 : y \in \mathbb{R} \mapsto \frac{y}{n}I_n \in \text{Diag}(n)$,

with ψ_4, ψ_5 injective. The corresponding orthogonal projections are:

- $\pi_5 : X \in \text{Sym}(n) \mapsto X_5 = \psi_5(\text{tr}(X)) \in \text{im } \psi_5$,
- $\pi_4 : X \in \text{Sym}(n) \mapsto X_4 = \text{Diag}(X - X_5) \in \text{im } \psi_4$,
- $\pi_3 : X \in \text{Sym}(n) \mapsto X_3 = p_3(X - \text{Diag}(X)) \in \text{im } \psi_3$,
- $\pi_2 : X \in \text{Sym}(n) \mapsto X_2 = p_2(X - \text{Diag}(X)) \in \text{im } \psi_2$,
- $\pi_1 : X \in \text{Sym}(n) \mapsto X_1 = p_1(X - \text{Diag}(X)) \in \text{ker } \phi_1$.

The $\mathfrak{S}(n)$ -invariant inner products on $\text{Sym}(n)$ are given for all $X = X_1 + \psi_2(\mu) + \psi_3(x) + \psi_4(\lambda) + \psi_5(y) \in \text{Sym}(n)$ by:

$$\varphi(X, X) = \alpha_1 \text{tr}(X_1^2) + S_{11} \|\mu\|^2 + \Sigma_{11} x^2 + S_{22} \|\lambda\|^2 + \Sigma_{22} y^2 + 2S_{12} \lambda^\top \mu + 2\Sigma_{12} xy \quad (3.16)$$

$$= \alpha \text{tr}(X^2) + \beta \text{sum}(X^2) + \gamma \text{sum}(X)^2 + \delta \text{tr}(\text{Diag}(X)^2) + \varepsilon \text{sum}(\text{Diag}(X)X) + \zeta \text{tr}(X)^2 + \eta \text{tr}(X) \text{sum}(X), \quad (3.17)$$

where $\alpha_1 = \alpha > 0$ and the SPD matrices are $S = \begin{pmatrix} \frac{2}{n-2}\alpha + \beta & \beta + \frac{\varepsilon}{2} \\ \beta + \frac{\varepsilon}{2} & \alpha + \beta + \delta + \varepsilon \end{pmatrix} \in \text{Sym}^+(2)$ and $\Sigma = \frac{1}{n} \begin{pmatrix} \frac{1}{n-1}\alpha + \beta + n\gamma & \beta + \frac{\varepsilon}{2} + n(\gamma + \frac{\eta}{2}) \\ \beta + \frac{\varepsilon}{2} + n(\gamma + \frac{\eta}{2}) & \alpha + \beta + \delta + \varepsilon + n(\gamma + \zeta + \eta) \end{pmatrix} \in \text{Sym}^+(2)$. See the proof of Example 3.10 in Section 11.2.

Remark 3.11 The case $n = 3$ is degenerate because $\ker \phi_1 = \{0\}$. In this case, one has to replace α_1 and α by 0 (and n by 3) in the previous formulae.

3.3.4.3 $\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$, $n \geq 3$

Example 3.12 ($\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$) The irreducible decomposition of $\text{Mat}(n)$ with respect to the congruence action of $\mathfrak{S}(n)$, i.e. of permutation matrices, is:

$$\text{Mat}(n) = \underbrace{\ker \phi_1 \oplus \text{im } \psi_2 \oplus \text{im } \psi_3}_{\text{Hol}(n)} \oplus \underbrace{\text{im } \psi_4 \oplus \text{im } \psi_5}_{\text{Diag}(n)} \oplus \underbrace{\ker \phi_6 \oplus \text{im } \psi_7}_{\text{Skew}(n)} \quad (3.18)$$

$$\underbrace{\hspace{15em}}_{\text{Sym}(n)}$$

$$\simeq \ker \phi_1 \oplus \ker \phi_6 \oplus 3\mathbb{R}_0^n \oplus 2\mathbb{R}, \quad (3.19)$$

where $\phi_1, \psi_2, \psi_3, \psi_4, \psi_5$ were defined in the previous sections and:

6. $\phi_6 : X \in \text{Skew}(n) \mapsto X\mathbf{1} \in \mathbb{R}_0^n$,
7. $\psi_7 : \nu \in \mathbb{R}_0^n \mapsto \frac{1}{n}(\nu\mathbf{1}^\top - \mathbf{1}\nu^\top) \in \text{Skew}(n)$,

with ϕ_6 surjective and ψ_7 injective. The corresponding orthogonal projections are:

- $\Pi_7 : X \in \text{Mat}(n) \mapsto X_7 = \psi_7 \left(\frac{X-X^\top}{2} \mathbf{1} \right) \in \text{im } \psi_7$,
- $\Pi_6 : X \in \text{Mat}(n) \mapsto X_6 = \frac{X-X^\top}{2} - X_7 \in \ker \phi_6$,
- $\Pi_5 : X \in \text{Mat}(n) \mapsto X_5 = \pi_5 \left(\frac{X+X^\top}{2} \right) = \psi_5(\text{tr}(X)) \in \text{im } \psi_5$,
- $\Pi_4 : X \in \text{Mat}(n) \mapsto X_4 = \pi_4 \left(\frac{X+X^\top}{2} \right) = \text{Diag}(X - X_5) \in \text{im } \psi_4$,
- $\Pi_3 : X \in \text{Mat}(n) \mapsto X_3 = \pi_3 \left(\frac{X+X^\top}{2} \right) \in \text{im } \psi_3$,
- $\Pi_2 : X \in \text{Mat}(n) \mapsto X_2 = \pi_2 \left(\frac{X+X^\top}{2} \right) \in \text{im } \psi_2$,
- $\Pi_1 : X \in \text{Mat}(n) \mapsto X_1 = \pi_1 \left(\frac{X+X^\top}{2} \right) \in \ker \phi_1$.

The $\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$ are given for all $X = X_1 + \psi_2(\mu) + \psi_3(x) + \psi_4(\lambda) + \psi_5(y) + X_6 + \psi_7(\nu) \in \text{Mat}(n)$ by:

$$\begin{aligned} \varphi(X, X) &= \alpha_1 \text{tr}(X_1^2) + S_{11} \|\mu\|^2 + \Sigma_{11} x^2 + S_{22} \|\lambda\|^2 + \Sigma_{22} y^2 + \alpha_2 \text{tr}(X_6^2) + S_{33} \|\nu\|^2 \\ &\quad + 2S_{12} \lambda^\top \mu + 2S_{23} \nu^\top \lambda + 2S_{13} \mu^\top \nu + 2\Sigma_{12} xy \end{aligned} \quad (3.20)$$

$$\begin{aligned} &= \alpha \text{tr}(XX^\top) + \alpha' \text{tr}(X^2) + \beta \text{sum}(X^2) + \beta' \text{sum}(XX^\top) + \beta'' \text{sum}(X^\top X) \\ &\quad + \gamma \text{sum}(X)^2 + \delta \text{tr}(\text{Diag}(X)^2) + \zeta \text{tr}(X)^2 + \eta \text{tr}(X) \text{sum}(X) \\ &\quad + \varepsilon \text{sum}(\text{Diag}(X)X) + \varepsilon' \text{sum}(\text{Diag}(X)X^\top), \end{aligned} \quad (3.21)$$

where $\alpha_1 = \frac{\alpha+\alpha'}{2} > 0$, $\alpha_2 = \frac{\alpha-\alpha'}{2} > 0$ and the SPD matrices are:

$$S = \begin{pmatrix} \frac{2}{n-2} \frac{\alpha+\alpha'}{2} + \beta + \beta' + \beta'' & \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') & \frac{\beta''-\beta'}{4} \\ \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') & \frac{\alpha+\alpha'}{2} + \beta + \beta' + \beta'' + \delta + \varepsilon + \varepsilon' & \frac{\beta''-\beta'}{4} + \frac{\varepsilon'-\varepsilon}{2} \\ \frac{\beta''-\beta'}{4} & \frac{\beta''-\beta'}{4} + \frac{\varepsilon'-\varepsilon}{2} & 2\alpha + \beta - \beta' - \beta'' \end{pmatrix} \in \text{Sym}^+(3),$$

$$\Sigma = \frac{1}{n} \begin{pmatrix} \frac{1}{n-1} \frac{\alpha+\alpha'}{2} + \beta + \beta' + \beta'' + n\gamma & \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') + n(\gamma + \frac{\eta}{2}) \\ \beta + \frac{1}{2}(\varepsilon + \varepsilon') + n(\gamma + \frac{\eta}{2}) & \frac{\alpha+\alpha'}{2} + \beta + \beta' + \beta'' + \delta + \varepsilon + \varepsilon' + n(\gamma + \zeta + \eta) \end{pmatrix} \in \text{Sym}^+(2).$$

See the proof of Example 3.12 in Section 11.2.

Remark 3.13 The case $n = 3$ is degenerate because $\ker \phi_1 = \{0\}$. In this case, one has to replace α_1 and α by 0 (and n by 3) in the previous formulae.

Remark 3.14 Equation (3.19) can be found in [Ramgoolam, 2019]. It gives the irreducible decomposition of $\text{Mat}(n)$ up to module isomorphism. Our Equation (3.18) instantiates an irreducible decomposition by choosing specific spaces. The advantage is that it allows to manipulate coordinate-free equations and explicit spaces. Moreover, it puts the light on interesting spaces that are stable by permutations.

3.4 Conclusion

We formalized a general method to determine all G -invariant inner products on a completely reducible Euclidean space V . We gave the general form of inner products on \mathbb{R}^n and $\text{Mat}(n)$ that are invariant under the orthogonal group $O(n)$, the block orthogonal group $O(I)$ and the permutation group $\mathfrak{S}(n)$. They are summarized in Table 3.2.

Beyond linear algebra, this characterization is interesting when one wants to characterize invariant Riemannian metrics on Lie groups and homogeneous spaces. It is well known that the $O(n)$ -invariant inner product on Grassmannians $\text{Gr}(n, k) = O(n)/(O(k) \times O(n-k))$ is unique up to scaling. More generally, our method allowed us to characterize all $O(n)$ -invariant metrics on the homogeneous manifold $\text{Flag}(I) = O(n)/O(I)$ of flags of signature I (see Equation (3.8)). The characterization is also interesting when there exists a global diffeomorphism from a manifold to a vector space, such as for symmetric positive definite matrices or full-rank correlation matrices.

This method is based on representation theory. It would be interesting to investigate non-linear methods to characterize invariant Riemannian metrics on manifolds when this problem does not reduce to a linear problem.

Action	Inner product $\varphi(x, x)$ or $\varphi(X, X)$	Positivity condition
$O(n) \curvearrowright \mathbb{R}^n$	$\alpha \ x\ ^2$	$\alpha > 0$
$O(I) \curvearrowright \mathbb{R}^n$	$\sum_{i=1}^p \alpha_{k_i} \ p_{k_i}(x)\ ^2$	$\alpha_{k_i} > 0$
$\mathfrak{S}(n) \curvearrowright \mathbb{R}^n$	$\alpha \ x\ ^2 + \beta \sum(x)^2$	$\alpha > 0, \alpha + n\beta > 0$
$O(n) \curvearrowright \text{Mat}(n)$	$\alpha \text{tr}(XX^\top) + \beta \text{tr}(X^2) + \gamma \text{tr}(X)^2$	$\alpha > \beta , \alpha + \beta + n\gamma > 0$
$O(I) \curvearrowright \text{Skew}(n)$	$\sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \alpha_{ii} \ X^{ii}\ ^2 + \sum_{1 \leq i < j \leq p} \alpha_{ij} \ X^{ij}\ ^2$	$\alpha_{ij} > 0$
$O(I) \curvearrowright \text{Sym}(n)$	$\sum_{1 \leq i, j \leq p} S_{ij} \frac{\text{tr}(X^{ii})\text{tr}(X^{jj})}{\sqrt{k_i k_j}} + \sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \alpha_{ii} \ X^{ii} - \frac{\text{tr}(X^{ii})}{k_i} I_n\ ^2 + \sum_{1 \leq i < j \leq p} \alpha_{ij} \ X^{ij}\ ^2$	$S \in \text{Sym}^+(p), \alpha_{ij} > 0$
$O(I) \curvearrowright \text{Mat}(n)$	$\sum_{1 \leq i, j \leq p} S_{ij} \frac{\text{tr}(X^{ii})\text{tr}(X^{jj})}{\sqrt{k_i k_j}} + \sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \left(\alpha_{ii} \ X_{\text{sym}}^{ii} - \frac{\text{tr}(X^{ii})}{k_i} I_n\ ^2 + \beta_{ii} \ X_{\text{skew}}^{ii}\ ^2 \right) + \sum_{1 \leq i < j \leq p} \left(\alpha_{ij} \ X_{\text{sym}}^{ij}\ ^2 + \beta_{ij} \ X_{\text{skew}}^{ij}\ ^2 + \gamma_{ij} \langle X_{\text{sym}}^{ij}, X_{\text{skew}}^{ij} \rangle \right)$	$S \in \text{Sym}^+(p), \alpha_{ii} > 0, \beta_{ii} > 0$ $\alpha_{ij} > 0, \alpha_{ij}\beta_{ij} - \gamma_{ij}^2 > 0$
$\mathfrak{S}(n) \curvearrowright \text{Hol}(n)$	$\alpha \text{tr}(X^2) + \beta \text{sum}(X^2) + \gamma \text{sum}(X)^2$	$\alpha > 0, 2\alpha + (n-2)\beta > 0,$ $\alpha + (n-1)(\beta + n\gamma) > 0$
$\mathfrak{S}(n) \curvearrowright \text{Sym}(n)$	$\alpha \text{tr}(X^2) + \beta \text{sum}(X^2) + \gamma \text{sum}(X)^2 + \delta \text{tr}(\text{Diag}(X)^2) + \varepsilon \text{sum}(\text{Diag}(X)X) + \zeta \text{tr}(X)^2 + \eta \text{tr}(X) \text{sum}(X)$	$\alpha > 0, S_1, \Sigma_1 \in \text{Sym}^+(2)$
$\mathfrak{S}(n) \curvearrowright \text{Mat}(n)$	$\alpha \text{tr}(XX^\top) + \alpha' \text{tr}(X^2) + \beta \text{sum}(X^2) + \beta' \text{sum}(XX^\top) + \beta'' \text{sum}(X^\top X) + \gamma \text{sum}(X)^2 + \delta \text{tr}(\text{Diag}(X)^2) + \zeta \text{tr}(X)^2 + \eta \text{tr}(X) \text{sum}(X) + \varepsilon \text{sum}(\text{Diag}(X)X) + \varepsilon' \text{sum}(\text{Diag}(X)X^\top)$	$\alpha > \alpha' $ $S_2 \in \text{Sym}^+(3)$ $\Sigma_2 \in \text{Sym}^+(2)$

$$S_1 = \begin{pmatrix} \frac{2}{n-2}\alpha + \beta & \beta + \frac{\varepsilon}{2} \\ \beta + \frac{\varepsilon}{2} & \alpha + \beta + \delta + \varepsilon \end{pmatrix} \text{ and } \Sigma_1 = \frac{1}{n} \begin{pmatrix} \frac{1}{n-1}\alpha + \beta + n\gamma & \beta + \frac{\varepsilon}{2} + n(\gamma + \frac{\eta}{2}) \\ \beta + \frac{\varepsilon}{2} + n(\gamma + \frac{\eta}{2}) & \alpha + \beta + \delta + \varepsilon + n(\gamma + \zeta + \eta) \end{pmatrix},$$

$$S_2 = \begin{pmatrix} \frac{2}{n-2} \frac{\alpha + \alpha'}{2} + \beta + \beta' + \beta'' & \beta + \beta' + \beta'' & \beta + \beta' + \beta'' + \frac{\beta'' - \beta'}{4} \\ \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') & \frac{\alpha + \alpha'}{2} + \beta + \beta' + \beta'' + \delta + \varepsilon + \varepsilon' & \frac{\beta'' - \beta'}{4} + \frac{\varepsilon' - \varepsilon}{2} \\ \frac{\beta'' - \beta'}{4} + \frac{\varepsilon' - \varepsilon}{2} & \frac{\beta'' - \beta'}{4} + \frac{\varepsilon' - \varepsilon}{2} & 2\alpha + \beta - \beta' - \beta'' \end{pmatrix},$$

$$\Sigma_2 = \frac{1}{n} \begin{pmatrix} \frac{1}{n-1} \frac{\alpha + \alpha'}{2} + \beta + \beta' + \beta'' + n\gamma & \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') + n(\gamma + \frac{\eta}{2}) \\ \beta + \frac{1}{2}(\varepsilon + \varepsilon') + n(\gamma + \frac{\eta}{2}) & \alpha + \alpha' + \beta + \beta' + \beta'' + \delta + \varepsilon + \varepsilon' + n(\gamma + \zeta + \eta) \end{pmatrix}.$$

 Table 3.2: General form of invariant inner products, where $x \in \mathbb{R}^n$ and $X \in \text{Mat}(n)$.

Part III

**Covariance matrices of full rank:
the open cone of symmetric
positive definite matrices**

Chapter 4

$O(n)$ -invariant Riemannian metrics

Abstract

Symmetric Positive Definite (SPD) matrices are ubiquitous in data analysis under the form of covariance matrices or correlation matrices. Several $O(n)$ -invariant Riemannian metrics were defined on the SPD cone, in particular the kernel metrics introduced by Hiai and Petz. The class of kernel metrics interpolates between many classical $O(n)$ -invariant metrics and it satisfies key results of stability and completeness. However, it does not contain all the classical $O(n)$ -invariant metrics. Therefore in this work, we investigate super-classes of kernel metrics and we study which key results remain true. We also introduce an additional key result called cometric-stability, a crucial property to implement geodesics with a Hamiltonian formulation. Our method to build intermediate embedded classes between $O(n)$ -invariant metrics and kernel metrics is to give a characterization of the whole class of $O(n)$ -invariant metrics on SPD matrices and to specify requirements on metrics one by one until we reach kernel metrics. As a secondary contribution, we synthesize the literature on the main $O(n)$ -invariant metrics, we provide the complete formula of the sectional curvature of the affine-invariant metric and the formula of the geodesic parallel transport between commuting matrices for the Bures-Wasserstein metric.

*This chapter was resubmitted in January 2022 to the journal *Linear Algebra and its Applications* after minor revisions [Thanwerdas and Pennec, 2022b]. Here is the revised version with an additional modification specific to this manuscript. Indeed, our proof of continuity of the $O(n)$ -invariant metrics in Section 4.5.2 is trivialized by our Theorem 2.18 in Chapter 2 so we replace the sequence of majorations that proved the continuity by a simple reference to this Theorem 2.18.*

4.1 Introduction

Symmetric Positive Definite (SPD) matrices are ubiquitous in data analysis because in many situations, the data (signals, images, diffusion coefficients...) can be represented by their covariance matrices. This is the case in the domains of Brain-Computer Interfaces, diffusion and functional MRI, Computer Vision, Diffusion Tensor Imaging (DTI), etc. SPD matrices form a cone in the vector space of symmetric matrices so a first idea to compute with SPD matrices could be to perform Euclidean computations on symmetric matrices. However, this method has several drawbacks. As geodesics are straight lines, they leave the SPD cone

at finite time so extrapolation methods could lead to non admissible matrices, namely with negative eigenvalues. Moreover, the trace is linearly interpolated but other invariants such as the determinant are not monotonically interpolated along geodesics. For example in DTI, where SPD matrices are represented by 3D ellipsoids, the ellipsoids along the geodesic can have a larger volume than the two ellipsoids at extremities, which leads to non realistic predictions in fiber tracking (swelling effect).

Hence, other Riemannian metrics were used in applications to solve these problems. The affine-invariant/Fisher-Rao metric [Skovgaard, 1984, Pennec et al., 2006, Lenglet et al., 2006, Fletcher and Joshi, 2007, Moakher, 2005, Batchelor et al., 2005, Varoquaux et al., 2010, Barachant et al., 2013] provides a Riemannian symmetric structure to the SPD manifold: it is negatively curved, geodesically complete (matrices with null eigenvalues are rejected to infinity), it is invariant under the congruence action (which, in the context of covariance matrices, corresponds to the invariance of the feature vector under affine transformations) and it is inverse-consistent. The log-Euclidean metric [Arsigny et al., 2006] is diffeomorphic to a Euclidean inner product: it also provides a Riemannian symmetric space, it is geodesically complete and inverse-consistent. It is not curved and it is not affine-invariant although it is still invariant under orthogonal transformations and dilations. The Bures-Wasserstein/Procrustes metric [Bhatia et al., 2019, Dryden et al., 2009, Takatsu, 2011, Malagò et al., 2018] is a positively curved quotient metric which is also invariant under orthogonal transformations. It is not geodesically complete but geodesics remain in the cone with boundaries: this means that this metric is suited for computing with Positive Semi-Definite (PSD) matrices. Many other interesting metrics exist with different properties: Bogoliubov-Kubo-Mori [Petz and Toth, 1993, Michor et al., 2000], polar-affine [Su et al., 2012], Euclidean-Cholesky [Wang et al., 2004], log-Euclidean-Cholesky [Li et al., 2017], log-Cholesky [Pinheiro and Bates, 1996, Lin, 2019], power-Euclidean [Dryden et al., 2010], and more recently power-affine [Thanwerdas and Pennec, 2019b], alpha-Procrustes [Hà Quang, 2019], mixed-power-Euclidean [Thanwerdas and Pennec, 2019a].

Except those named after Cholesky, all the other Riemannian metrics cited above are invariant under orthogonal transformations. If we consider SPD matrices as covariance matrices, this transformation corresponds to a rigid-body transformation of the feature vector $X \in \mathbb{R}^n \mapsto RX + X_0$ where R is an orthogonal matrix. In 2009, Hiai and Petz introduced the subclass of *kernel metrics* [Hiai and Petz, 2009], which are $O(n)$ -invariant metrics indexed by smooth symmetric maps $\phi : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$. This class satisfies key results: it contains most of the cited $O(n)$ -invariant metrics, it is stable under a certain class of diffeomorphisms and it provides a sufficient condition for geodesic completeness. This sufficient condition becomes necessary if we restrict the class to the subclass of *mean kernel metrics* which is indexed by kernel maps of the form $\phi = m^\theta$ where $m : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$ is a symmetric homogeneous mean and $\theta \in \mathbb{R}$ is a power. However, the class of kernel metrics does not contain *all* the aforementioned $O(n)$ -invariant metrics. The main goal of this chapter is to study the super-classes of kernel metrics, especially the whole class of $O(n)$ -invariant metrics for which we give a characterization. More precisely, our objective is to determine which key results on kernel metrics can be generalized and thus to understand better the specificity of kernel metrics within these super-classes.

4.1.1 Results and organization of the chapter

In the remainder of the Introduction, we give the notations and conventions used in the chapter. In Section 4.2, we introduce two preliminary concepts and one result. The first concept is the notion of $O(n)$ -equivariant map on symmetric matrices. We especially explain how to build them from a map defined on diagonal matrices via the spectral theorem because this is a procedure we need several times in the chapter. Then the second concept is a particular case of the previous one, called *univariate map*. These are maps characterized by a map on positive real numbers. They are particularly interesting because their differential is known in closed form modulo eigenvalue decomposition and because the class of kernel metrics is stable under univariate diffeomorphisms. Finally the result is the characterization of $O(n)$ -invariant inner products on symmetric matrices. These inner products are composed of two terms, the Frobenius term and the trace term, which have different weights so they form a two-parameter family. In the proof, we give elementary tools that we reuse when we characterize $O(n)$ -invariant metrics on SPD matrices.

To explain why kernel metrics do not encompass all the $O(n)$ -invariant metrics cited above, we need to present them or at least the most important ones. One can notice that many metrics and families of metrics are actually based on five of them, namely the Euclidean, the log-Euclidean, the affine-invariant, the Bures-Wasserstein and the Bogoliubov-Kubo-Mori metrics. That is why in Section 4.3, we synthesize the literature on these five noted metrics. For each of them, we give the fundamental Riemannian operations (squared distance, Levi-Civita connection, curvature, geodesics, logarithm map, parallel transport map) when they are known. As a secondary contribution, we give the complete formula of the sectional curvature of the affine-invariant metric and we also give, for the Bures-Wasserstein metric, the new formula of the parallel transport between commuting matrices and simpler formulae of the Levi-Civita connection, the curvature and the parallel transport equation.

In Section 4.4, after reviewing kernel metrics and their key properties, we give two new observations on them. Firstly, the cometric of a metric on SPD matrices can be considered itself as a metric on SPD matrices by identifying the vector space of symmetric matrices and its dual via the Frobenius inner product. Therefore we observe that the cometric of a kernel metric defined by the kernel map ϕ is a kernel metric characterized by $1/\phi$. This remarkable result has an important consequence for the numerical computation of geodesics. Indeed, the geodesic equation $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$, which is a second order equation, has a Hamiltonian version which is a first order equation that only involves the cometric, not the Christoffel symbols. The Hamiltonian equation is much simpler to integrate and numerically more stable, that is why it is often preferred in numerical implementations, for instance in the Python package `geomstats` [Miolane et al., 2020a]. Hence knowing a simple explicit formula for the cometric helps to compute numerically the geodesics. Secondly, there is a natural extension of kernel metrics that encompasses all the aforementioned $O(n)$ -invariant metrics, which still satisfies the key properties of kernel metrics including the cometric stability. Roughly speaking, kernel metrics look like the Frobenius inner product on symmetric matrices where the elementary quadratic forms (the X_{ij}^2) are weighted by a coefficient involving the kernel map ϕ and depending on the point. Since the Frobenius inner product is not the only $O(n)$ -invariant inner product on symmetric matrices as explained above, the trace term can be added to the framework of kernel metrics to form extended kernel metrics.

In Section 4.5, we characterize the class of $O(n)$ -invariant metrics on SPD matrices by means of three multivariate maps $\alpha, \beta, \gamma : (\mathbb{R}^+)^n \rightarrow \mathbb{R}$ operating on the eigenvalues (d_1, \dots, d_n) of the SPD matrix and which satisfy three conditions of symmetry, compatibility and positivity (Theorem 4.5). Then, we observe that kernel metrics are characterized by two properties within this family. They are *ortho-diagonal*: it means that the metric matrix is diagonal, i.e. $\beta = 0$. They are *bivariate*: it means that the remaining functions α and γ do not depend on their $n - 2$ last terms, and the compatibility condition imposes that they are equal so we can write $\gamma = \alpha = 1/\phi : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$. Since the term “kernel” is quite overloaded in many different contexts (such as in Reproducing Kernel Hilbert Spaces in machine learning or in kernel density estimation/regression in statistics), we propose to designate them as Bivariate Ortho-Diagonal (BOD) metrics. Afterwards, we give key properties of $O(n)$ -invariant metrics in analogy with the key properties of BOD (kernel) metrics. Since we do not have a closed-form expression for the cometric anymore, we introduce the intermediate class of bivariate separable metrics which is cometric-stable and we give the expression of the cometric. A summary of the classes of metrics defined in this chapter is shown on Figure 4.1.

Section 4.6 is dedicated to the conclusion.

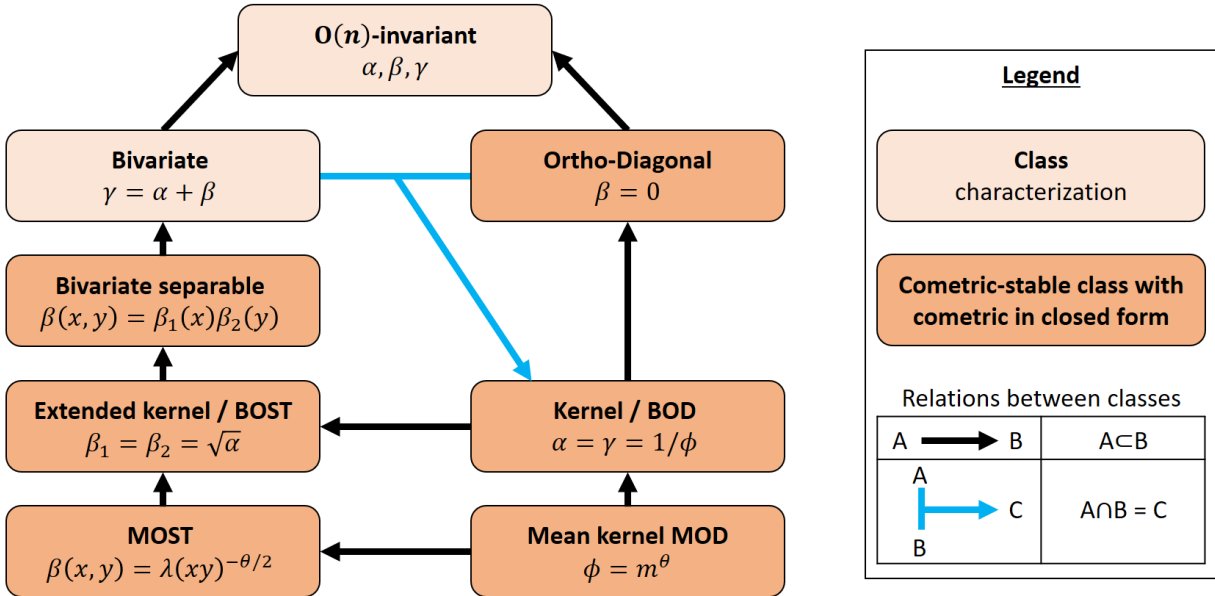


Figure 4.1: Super-classes of kernel metrics.

4.1.2 Notations and conventions

Manifolds Our manifold-related notations are summarized in Table 4.1. A chart $\varphi : \mathcal{U} \subset \mathcal{M} \rightarrow \mathbb{R}^N$ provides a local basis of vectors $(\partial_1, \dots, \partial_N)$ where $\partial_k = \frac{\partial}{\partial \varphi^k}$ is a short notation defined for all differentiable maps $f : \mathcal{M} \rightarrow \mathbb{R}$ and at each point $x \in \mathcal{U}$ by $(\partial_k f)|_x = \frac{\partial(f \circ \varphi^{-1})}{\partial x^k} \Big|_{\varphi(x)}$. A vector field X can be locally decomposed on this basis, $X = X^k \partial_k$, where $X^k : \mathcal{U} \rightarrow \mathbb{R}$ are the coordinate functions of X and where we used Einstein’s summation

convention. As we deal with matrices, the coordinates often have two indices: $X = X^{ij}\partial_{ij}$.

$T_x\mathcal{M}, T\mathcal{M}$	Tangent space at x , tangent bundle
$d_x f, df$	Differential of map f at x , differential of map f
f^*, f_*	Pullback via f , pushforward via f
$\dot{\gamma}$	Derivative of curve γ
g, G	Metric on $\text{Sym}^+(n)$, metric on another space
d	Riemannian distance on $\text{Sym}^+(n)$
∇	Levi-Civita connection
R	Curvature $R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z$
$\gamma_{(\Sigma, X)}(t)$	Geodesic at time t with $\gamma(0) = \Sigma$ and $\dot{\gamma}(0) = X$
Exp, Log	Riemannian exponential and logarithm maps
$\Pi_{\gamma; \Sigma \rightarrow \Lambda} X$	Parallel transport of X along curve γ from Σ to Λ

Table 4.1: Notations in a manifold.

Manifolds of matrices We denote the matrix spaces as shown in Table 4.2. The (i, j) -coefficient of a matrix M is denoted M_{ij} , $[M]_{ij}$ or $M(i, j)$ depending on the context. To build a matrix from its coefficients, we denote $M = [M_{ij}]_{1 \leq i, j \leq n}$ or simply $M = [M_{ij}]_{i, j}$. We denote (C_{ij}) the canonical basis of matrices, $E_{ii} = C_{ii}$, $E_{ij} = \frac{1}{\sqrt{2}}(C_{ij} + C_{ji})$ and $F_{kl} = \frac{1}{2}(C_{kl} + C_{lk})$ for $i \neq j$ and $k, l \in \{1, \dots, n\}$. The norms are denoted $\|M\|_1 = \sum_{i, j} |M_{ij}|$ and $\|M\|_2 = \sqrt{\text{tr}(MM^T)}$.

Vector space of matrices		Manifold of matrices	
$\text{Mat}(n)$	$n \times n$ real matrices	$\text{GL}(n)$ $\text{GL}^+(n)$	General Linear group Positive determinant
$\text{Sym}(n)$	Real symmetric	$\text{Sym}^+(n)$	Symmetric positive definite
$\text{Skew}(n)$	Real skew-symmetric	$O(n)$ $\text{SO}(n)$	Orthogonal group Rotation group
$\text{Diag}(n)$	Diagonal	$\text{Diag}^+(n)$	Positive diagonal

Table 4.2: Notations for matrix spaces.

The congruence action is the following action of the general linear group on matrices $\star : (A, M) \in \text{GL}(n) \times \text{Mat}(n) \mapsto AMA^T \in \text{Mat}(n)$ which leaves stable the spaces of symmetric matrices and SPD matrices. Then, $A \in \text{GL}(n)$ naturally acts on $M = M^{i_1 j_1, \dots, i_q j_q} C_{i_1 j_1} \otimes \dots \otimes C_{i_q j_q} \in \text{Mat}(n)^{\otimes q}$ by:

$$A \star M = M^{i_1 j_1, \dots, i_q j_q} (A \star C_{i_1 j_1}) \otimes \dots \otimes (A \star C_{i_q j_q}) \in \text{Mat}(n)^{\otimes q}.$$

$\text{GL}(n)$ also acts by \star on any Cartesian product $\prod_{i=1}^r \text{Sym}(n)^{\otimes q_i}$ component-wise, especially on $\text{Sym}(n)^p$.

Let \mathcal{E} and \mathcal{F} be two spaces on which $\text{GL}(n)$ acts by \star . Let $\mathcal{G} \subseteq \text{GL}(n)$ be a subgroup of $\text{GL}(n)$. A map $f : \mathcal{E} \rightarrow \mathcal{F}$ is:

- \mathcal{G} -equivariant if $f(A \star M) = A \star f(M)$ for all $A \in \mathcal{G}$, for all $M \in \mathcal{E}$,
- \mathcal{G} -invariant if $f(A \star M) = f(M)$ for all $A \in \mathcal{G}$, for all $M \in \mathcal{E}$.

In particular, a Riemannian metric $g : \text{Sym}^+(n) \times \text{Sym}(n) \times \text{Sym}(n) \rightarrow \mathbb{R}$ (or an inner product) is \mathcal{G} -invariant if $g_{A\Sigma A^\top}(AXA^\top, AXA^\top) = g_\Sigma(X, X)$ for all $A \in \mathcal{G}$, $\Sigma \in \text{Sym}^+(n)$ and $X \in \text{Sym}(n)$.

The symmetric group of order n is denoted by \mathfrak{S}_n and the permutations by small greek letters σ, τ, \dots . The permutation matrix associated to the permutation σ , which sends any basis (e_1, \dots, e_n) of \mathbb{R}^n to the permuted basis $(e_{\sigma(1)}, \dots, e_{\sigma(n)})$, is denoted P_σ . We have $P_\sigma(i, j) = \delta_{\sigma(i), j}$ where δ is the Kronecker symbol. Given a matrix $M \in \text{Mat}(n)$, we have $(P_\sigma^\top M P_\sigma)(i, j) = M(\sigma(i), \sigma(j))$.

The manifold of SPD matrices The manifold $\text{Sym}^+(n)$ is an open set of the vector space of symmetric matrices $\text{Sym}(n)$. Hence, the canonical immersion $\text{id} : \text{Sym}^+(n) \hookrightarrow \text{Sym}(n)$ provides:

- An identification between the tangent space $T_\Sigma \text{Sym}^+(n)$ and the vector space $\text{Sym}(n)$ at any point $\Sigma \in \text{Sym}^+(n)$ by $d_\Sigma \text{id} : T_\Sigma \text{Sym}^+(n) \xrightarrow{\sim} \text{Sym}(n)$. Thus, any tangent vector $X \in T_\Sigma \text{Sym}^+(n)$ is considered as a symmetric matrix: $X \equiv d_\Sigma \text{id}(X) \in \text{Sym}(n)$.
- A global chart $(\text{id}, \text{Sym}^+(n))$ of the manifold $\text{Sym}^+(n)$, thus a global derivation $\partial_X Y = X^{ij}(\partial_{ij} Y^{kl})\partial_{kl}$ defined by derivation of coordinates in this global chart. More generally, if $f : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$ is a diffeomorphism on its image, it provides a global derivation denoted ∂^f .

Another important tool is the matrix exponential $\exp(X) = \sum_{k=0}^{+\infty} \frac{X^k}{k!}$ which is a diffeomorphism between $\text{Sym}(n)$ and $\text{Sym}^+(n)$, and therefore its inverse, the symmetric matrix logarithm $\log : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$.

The spectral theorem ensures that symmetric matrices are orthogonally congruent to a diagonal matrix. If the symmetric matrix is SPD, then the diagonal matrix has positive elements on the diagonal. Most of the time in this chapter, for an SPD matrix $\Sigma \in \text{Sym}^+(n)$, we denote $\Sigma = PDP^\top$ one spectral decomposition with $P \in O(n)$ and $D = \text{diag}(d_1, \dots, d_n) \in \text{Diag}^+(n)$. When we consider tangent vectors $X, Y, \dots \in T_\Sigma \text{Sym}^+(n)$, we denote $X' = P^\top X P$ so that every matrix expressed in the orthogonal basis given by P is denoted with a prime: $X = P X' P^\top$, $Y = P Y' P^\top$, etc.

Products of symmetric matrices share two nice properties with symmetric matrices. First, if $X, Y \in \text{Sym}(n)$, then $\text{sp}(XY) \subset \mathbb{R}$ where sp denotes the spectrum, i.e. the set of complex eigenvalues. Second, if $\Sigma, \Lambda \in \text{Sym}^+(n)$, then $\Sigma\Lambda$ has a unique square-root matrix that represents a positive definite self-adjoint endomorphism, it is denoted $(\Sigma\Lambda)^{1/2} = \sqrt{\Sigma\Lambda} = \Sigma^{1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}\Sigma^{-1/2} = \Lambda^{-1/2}(\Lambda^{1/2}\Sigma\Lambda^{1/2})^{1/2}\Lambda^{1/2}$. This definition is used in [Bhatia et al., 2019, Equation (10)] for example.

4.2 Preliminary concepts and results

4.2.1 Extending maps defined on diagonal matrices

Thanks to the spectral theorem, $O(n)$ -equivariant maps $f : \text{Sym}^+(n) \rightarrow \mathcal{F}$ are characterized by their values on positive diagonal matrices. A question that arises several times in this chapter is: are we allowed to extend a map $f : \text{Diag}^+(n) \rightarrow \mathcal{F}$ into an $O(n)$ -equivariant map $f : \text{Sym}^+(n) \rightarrow \mathcal{F}$ by the formula $f(PDP^\top) = P \star f(D)$? To do so, we need to show that given two eigenvalue decompositions $\Sigma = PDP^\top = Q\Delta Q^\top$, we have $P \star f(D) = Q \star f(\Delta)$. Note that (Q, Δ) is highly constrained by (P, D) . The following lemma gives explicitly the possible cases, hence it tells exactly what is to be checked in such an extension process.

Lemma 4.1 (Relation between two eigenvalue decompositions of an SPD matrix) Let $D, \Delta \in \text{Diag}^+(n)$ and $P, Q \in O(n)$ such that $PDP^\top = Q\Delta Q^\top$. Let $\tau \in \mathfrak{S}(n)$ be a permutation that orders the values of D decreasingly, i.e. such that $D = P_\tau \text{Diag}(\lambda_1 I_{m_1}, \dots, \lambda_p I_{m_p}) P_\tau^\top$ with $\lambda_1 > \dots > \lambda_p > 0$. Then, there exists a permutation $\sigma \in \mathfrak{S}(n)$ and a block-diagonal orthogonal matrix $R = \text{Diag}(R_1, \dots, R_p) \in O(n)$ with j -th block $R_j \in O(m_j)$ such that $\Delta = P_\sigma^\top P_\tau^\top D P_\tau P_\sigma$ and $Q = P P_\tau R P_\sigma$.

Proof. Δ is clearly a permutation of D so there exists $\sigma \in \mathfrak{S}(n)$ such that $\Delta = P_\sigma^\top P_\tau^\top D P_\tau P_\sigma$. Let $R = P_\tau^\top P^\top Q P_\sigma^\top$. Then $PDP^\top = Q\Delta Q^\top$ is equivalent to $\text{Diag}(\lambda_1 I_{m_1}, \dots, \lambda_p I_{m_p}) R = R \text{Diag}(\lambda_1 I_{m_1}, \dots, \lambda_p I_{m_p})$. Decomposing R by blocks, the off-diagonal blocks have to be null since the λ_i 's are distinct. Since $RR^\top = I_n$, the diagonal blocks are orthogonal. \square

This result tells what is to be checked to extend $f : \text{Diag}^+(n) \rightarrow \mathcal{F}$. In this chapter, we only need to extend tensorial maps $T : \text{Diag}^+(n) \rightarrow \text{Sym}(n)^{\otimes q} \otimes (\text{Sym}(n)^*)^{\otimes p}$ or equivalently $T : \text{Diag}^+(n) \times \text{Sym}(n)^p \rightarrow \text{Sym}(n)^{\otimes q}$ for $p, q \in \mathbb{N}$. Hence, we state the result in this particular case though it is valid for \mathcal{F} .

Lemma 4.2 (Spectral extension) Let $T : \text{Diag}^+(n) \times \text{Sym}(n)^p \rightarrow \text{Sym}(n)^{\otimes q}$ be a map such that for all $D_0 = \text{Diag}(\lambda_1 I_{m_1}, \dots, \lambda_p I_{m_p})$ with $\lambda_1 > \dots > \lambda_p > 0$ and for all $X \in \text{Sym}(n)^p$:

- (a) $T(D_0, X) = P_\sigma \star T(P_\sigma^\top D_0 P_\sigma, P_\sigma^\top \star X)$ for all permutations $\sigma \in \mathfrak{S}(n)$,
- (b) $T(D_0, X) = R \star T(D_0, R^\top \star X)$ for all block-diagonal orthogonal matrices $R \in O(n)$, $R = \text{Diag}(R_1, \dots, R_p)$ with $R_j \in O(m_j)$.

Then, $T : \text{Sym}^+(n) \times \text{Sym}(n)^p \rightarrow \text{Sym}(n)^{\otimes q}$ defined by $T(PDP^\top, X) := P \star T(D, P^\top \star X)$ extends T , with $D \in \text{Diag}^+(n)$, $P \in O(n)$ and $X \in \text{Sym}(n)^p$.

Proof. Assume that $PDP^\top = Q\Delta Q^\top$. Then by Lemma 4.1, let $\sigma, \tau \in \mathfrak{S}(n)$ and R as in (b) such that $D_0 = P_\tau^\top D P_\tau = \text{Diag}(\lambda_1 I_{m_1}, \dots, \lambda_p I_{m_p})$, $\Delta = P_\sigma^\top D_0 P_\sigma$ and $Q = P P_\tau R P_\sigma$. Then, by applying (a) with σ , (b) with R and (a) with τ , we easily see that $Q \star T(\Delta, Q^\top \star X) = P \star T(D, P^\top \star X)$. Thus $T : \text{Sym}^+(n) \times \text{Sym}(n)^p \rightarrow \text{Sym}(n)^{\otimes q}$ is well defined. \square

In practice in this chapter, we use Lemma 4.2 for:

- $p = 0, q = 1$ for $f : \text{Diag}^+(n) \rightarrow \text{Sym}(n)$ in Section 4.2.2,

- $p = 1, q = 1$ for $\Phi : \text{Diag}^+(n) \times \text{Sym}(n) \longrightarrow \text{Sym}(n)$ in Section 4.4.2.3,
- $p = 2, q = 0$ for $g : \text{Diag}^+(n) \times \text{Sym}(n) \times \text{Sym}(n) \longrightarrow \mathbb{R}$ in Section 4.5.2.

4.2.2 Univariate maps

We apply Lemma 4.2 to a map defined on positive real numbers $f : \mathbb{R}^+ \longrightarrow \mathbb{R}$ and extended to positive diagonal matrices $f : \text{Diag}^+(n) \longrightarrow \text{Diag}(n)$ by $f(\text{Diag}(d_1, \dots, d_n)) := \text{Diag}(f(d_1), \dots, f(d_n))$.

- (a) Since f is defined component-wise, we have $f(D) = P_\sigma f(P_\sigma^\top D P_\sigma) P_\sigma^\top$.
- (b) As $f(\lambda I_{m_j}) = f(\lambda) I_{m_j}$, the matrix $Rf(D)R^\top$ is a block diagonal matrix with j -th block $f(\lambda_j) R_j R_j^\top = f(\lambda_j) I_{m_j}$, which corresponds to $f(D)$'s j -th block so $Rf(D)R^\top = f(D)$.

Therefore f can be extended into an $O(n)$ -equivariant map $f : \text{Sym}^+(n) \longrightarrow \text{Sym}(n)$ by $f(PDP^\top) = Pf(D)P^\top$. This extension is called the functional calculus of f in Functional Analysis. We call it a univariate map. The symmetric matrix logarithm $\log : \text{Sym}^+(n) \longrightarrow \text{Sym}(n)$, the power diffeomorphisms $\text{pow}_p : \text{Sym}^+(n) \longrightarrow \text{Sym}^+(n)$ with $p \neq 0$ or the constant map $\text{pow}_0 : \Sigma \in \text{Sym}^+(n) \longmapsto I_n \in \text{Sym}(n)$ are examples of univariate maps.

Definition 4.3 (Univariate maps) A univariate map is the extension of a map on positive real numbers $f : \mathbb{R}^+ \longrightarrow \mathbb{R}$ into an $O(n)$ -equivariant map $f : \text{Sym}^+(n) \longrightarrow \text{Sym}(n)$ by the equality $f(PDP^\top) = P \text{Diag}(f(d_1), \dots, f(d_n)) P^\top$. Moreover [Bhatia, 1997, Theorem V.3.3], if $f \in \mathcal{C}^1(\mathbb{R}^+)$, then its extension f is differentiable and the differential $df : \text{Sym}^+(n) \times \text{Sym}(n) \longrightarrow \text{Sym}(n)$ is $O(n)$ -equivariant, thus it is characterized by its values at diagonal matrices $D \in \text{Diag}^+(n)$, given by:

$$\forall X \in \text{Sym}(n), [d_D f(X)]_{ij} = f^{[1]}(d_i, d_j) X_{ij}, \quad (4.1)$$

where $f^{[1]}$ is the first divided difference defined below. Thus, a \mathcal{C}^1 -diffeomorphism $f : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is extended into a diffeomorphism $f : \text{Sym}^+(n) \longrightarrow \text{Sym}^+(n)$.

Definition 4.4 (First divided difference) [Bhatia, 1997] Let $f \in \mathcal{C}^1(\mathbb{R}^+)$. The first divided difference of f is the continuous symmetric map $f^{[1]} : (\mathbb{R}^+)^2 \longrightarrow \mathbb{R}$ defined for all $x, y \in \mathbb{R}$ by:

$$f^{[1]}(x, y) = \left\{ \begin{array}{ll} \frac{f(x)-f(y)}{x-y} & \text{if } x \neq y \\ f'(x) & \text{if } x = y \end{array} \right\}. \quad (4.2)$$

4.2.3 $O(n)$ -invariant inner products on symmetric matrices

To characterize the $O(n)$ -invariant metrics on SPD matrices, an appropriate starting point is the characterization of $O(n)$ -invariant inner products on the tangent space, i.e. on symmetric matrices. The following theorem states that such inner products form a two-parameter family indexed by a Scaling factor $\alpha > 0$ and a Trace factor $\beta > -\alpha/n$.

Theorem 4.5 (Characterization of $O(n)$ -invariant inner products on symmetric matrices) Let $\langle \cdot | \cdot \rangle : \text{Sym}(n) \times \text{Sym}(n) \rightarrow \mathbb{R}$ be an inner product on symmetric matrices. It is $O(n)$ -invariant if and only if there exists $(\alpha, \beta) \in \mathbf{ST} := \{(\alpha, \beta) \in \mathbb{R}^2 \mid \min(\alpha, \alpha + n\beta) > 0\}$ such that:

$$\forall X \in \text{Sym}(n), \langle X | X \rangle = \alpha \text{tr}(X^2) + \beta \text{tr}(X)^2. \quad (4.3)$$

Moreover, the linear isometry that pulls the Frobenius inner product back onto this one is $F_{p,q}(X) = qX + \frac{p-q}{n}\text{tr}(X)I_n$ with $p = \sqrt{\alpha + n\beta}$ and $q = \sqrt{\alpha}$.

There are several proofs of this elementary result. We already gave one in Section 3.3.2.1. Here we give one based on the following lemma because we reuse it to characterize $O(n)$ -invariant metrics on SPD matrices. This lemma gives the characterization of inner products on symmetric matrices which are respectively invariant under two subgroups of $O(n)$:

- (a) the group $\mathcal{D}^\pm(n) := \{\varepsilon = \text{Diag}(\pm 1, \dots, \pm 1)\} \cong \{-1, +1\}^n$ of diagonal matrices taking their diagonal values in $\{-1, +1\}$,
- (b) the group $\mathfrak{S}^\pm(n) := \{\varepsilon P_\sigma \in \text{Mat}(n) \mid (\varepsilon, \sigma) \in \mathcal{D}^\pm(n) \times \mathfrak{S}(n)\} \cong \mathcal{D}^\pm(n) \times \mathfrak{S}(n)$ of signed permutation matrices.

Lemma 4.6 (Characterization of inner products on symmetric matrices invariant under $\mathcal{D}^\pm(n)$ or $\mathfrak{S}^\pm(n)$) Let $\langle \cdot | \cdot \rangle : \text{Sym}(n) \times \text{Sym}(n) \rightarrow \mathbb{R}$ be an inner product on symmetric matrices.

- (a) It is $\mathcal{D}^\pm(n)$ -invariant if and only if there exist $\frac{n(n-1)}{2}$ positive real numbers $\alpha_{ij} = \alpha_{ji} > 0$ for $i \neq j$ and a matrix $S \in \text{Sym}^+(n)$ such that:

$$\forall X \in \text{Sym}(n), \langle X | X \rangle = \sum_{i \neq j} \alpha_{ij} X_{ij}^2 + \sum_{i,j} S_{ij} X_{ii} X_{jj}. \quad (4.4)$$

- (b) It is $\mathfrak{S}^\pm(n)$ -invariant if and only if there exist $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ with $\alpha > 0$, $\gamma > \beta$ and $\gamma + (n-1)\beta > 0$ such that:

$$\forall X \in \text{Sym}(n), \langle X | X \rangle = \gamma \sum_{i=1}^n X_{ii}^2 + \alpha \sum_{i \neq j} X_{ij}^2 + \beta \sum_{i \neq j} X_{ii} X_{jj}. \quad (4.5)$$

See the proof of Lemma 4.6 in Section 11.3.

Proof of Theorem 4.5. An $O(n)$ -invariant inner product on symmetric matrices is $\mathfrak{S}^\pm(n)$ -invariant so it is of the form of Equation (4.5). We define the rotation matrix $R = \begin{pmatrix} R_{\pi/4} & 0 \\ 0 & I_{n-2} \end{pmatrix} \in O(n)$ with $R_{\pi/4} = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \in O(2)$ and we apply it to the matrix $X = \begin{pmatrix} M & Y \\ Y^\top & Z \end{pmatrix} \in \text{Sym}(n)$ with $M = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \text{Sym}(2)$. Since $R_{\pi/4} M R_{\pi/4}^\top = \frac{1}{2} \begin{pmatrix} a+c+2b & c-a \\ c-a & a+c-2b \end{pmatrix}$, the coefficient of b^2 in $\langle X | X \rangle$ in Equation (4.5) is 2α and the coefficient of b^2 in $\langle R X R^\top | R X R^\top \rangle$ is $2\gamma - 2\beta$. Hence by invariance, $\gamma = \alpha + \beta$ and the positivity condition becomes $\alpha > 0$ and $\alpha + n\beta > 0$. Conversely, Equation (4.3) clearly defines $O(n)$ -invariant inner products. \square

4.3 Main $O(n)$ -invariant metrics on SPD matrices with new formulae

The goal of this section is to describe the main $O(n)$ -invariant metrics on SPD matrices that can be found in the literature, namely the Euclidean (abbreviated ‘E’, Section 4.3.1), the Log-Euclidean (‘LE’, Section 4.3.2), the Affine-invariant (‘A’, Section 4.3.3), the Bures-Wasserstein (‘BW’, Section 4.3.4) and the Bogoliubov-Kubo-Mori (‘BKM’, Section 4.3.5) metrics. For each metric, we give a short explanation on the way it was introduced, some useful references and a synthetic table that summarizes its fundamental Riemannian operations: squared distance, Levi-Civita connection, curvature, geodesics, logarithm map, parallel transport map.

Our contributions are (1) the synthesis of many results scattered in the literature especially for the Bures-Wasserstein metric, (2) the complete formula of the sectional curvature of the affine-invariant metric, (3) the new formula of the parallel transport between commuting matrices and new expressions of the Levi-Civita connection, the curvature and the parallel transport equation of the Bures-Wasserstein metric.

4.3.1 $O(n)$ -invariant Euclidean metrics

A Euclidean metric on SPD matrices is the pullback of an inner product $\langle \cdot | \cdot \rangle$ on symmetric matrices by the canonical immersion $\text{id} : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$. As we know $O(n)$ -invariant inner products on symmetric matrices from Theorem 4.5, we know all the $O(n)$ -invariant Euclidean metrics on SPD matrices.

Definition 4.7 ($O(n)$ -invariant Euclidean metrics on SPD matrices) An $O(n)$ -invariant Euclidean metric on SPD matrices is a Riemannian metric of the following form for all $\Sigma \in \text{Sym}^+(n)$ and $X \in \text{Sym}(n)$:

$$g_{\Sigma}^{\text{E}(\alpha, \beta)}(X, X) = \alpha \text{tr}(X^2) + \beta \text{tr}(X)^2, \quad (4.6)$$

with $(\alpha, \beta) \in \mathbf{ST}$, i.e. $\alpha > 0$ and $\beta > -\alpha/n$. Its Riemannian operations are detailed in Table 4.3.

4.3.2 $O(n)$ -invariant log-Euclidean metrics

A log-Euclidean metric on SPD matrices [Arsigny et al., 2006] is the pullback of an inner product $\langle \cdot | \cdot \rangle$ on symmetric matrices by the symmetric matrix logarithm $\log : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$. Hence the SPD manifold endowed with the log-Euclidean metric is isometric to a Euclidean space, thus geodesically complete. From Theorem 4.5 and the fact that $d\log : \text{Sym}^+(n) \times \text{Sym}(n) \rightarrow \text{Sym}(n)$ is $O(n)$ -equivariant, we know all the $O(n)$ -invariant log-Euclidean metrics.

Definition 4.8 ($O(n)$ -invariant log-Euclidean metrics on SPD matrices) An $O(n)$ -invariant log-Euclidean metric on SPD matrices is a Riemannian metric of the following form for all $\Sigma \in \text{Sym}^+(n)$ and $X \in \text{Sym}(n)$:

$$g_{\Sigma}^{\text{LE}(\alpha, \beta)}(X, X) = \alpha \text{tr}(d_{\Sigma} \log(X)^2) + \beta \text{tr}(\Sigma^{-1} X)^2, \quad (4.7)$$

Metric	$g_\Sigma(X, X) = \alpha \ X\ _2^2 + \beta \operatorname{tr}(X)^2$
Squared distance	$d(\Sigma, \Lambda)^2 = \alpha \ \Lambda - \Sigma\ _2^2 + \beta (\operatorname{tr}(\Lambda) - \operatorname{tr}(\Sigma))^2$
Levi-Civita	$\nabla_X Y = \partial_X Y$
Curvature	$R = 0$
Geodesics	$\gamma_{(\Sigma, X)}(t) = \Sigma + tX$ for $t \in I$ where I depends on $\lambda_{\min} = \min \operatorname{sp}(\Sigma^{-1}X)$ and $\lambda_{\max} = \max \operatorname{sp}(\Sigma^{-1}X)$ as follows: <ul style="list-style-type: none"> • If $\lambda_{\min} < 0 < \lambda_{\max}$, then $I = (-1/\lambda_{\max}, -1/\lambda_{\min})$. • If $0 \leq \lambda_{\min}$, then $I = (-1/\lambda_{\max}, +\infty)$. • If $\lambda_{\max} \leq 0$, then $I = (-\infty, -1/\lambda_{\min})$.
Logarithm	$\operatorname{Log}_\Sigma(\Lambda) = \Lambda - \Sigma$
Parallel transport	Does not depend on the curve: $\Pi_{\Sigma \rightarrow \Lambda} : \begin{cases} T_\Sigma \operatorname{Sym}^+(n) & \longrightarrow & T_\Lambda \operatorname{Sym}^+(n) \\ X & \longmapsto & (d_\Lambda \operatorname{id})^{-1}(d_\Sigma \operatorname{id}(X)) \equiv X \end{cases}$

Table 4.3: Riemannian operations of $O(n)$ -invariant Euclidean metrics on SPD matrices.

with $(\alpha, \beta) \in \mathbf{ST}$, i.e. $\alpha > 0$ and $\beta > -\alpha/n$. Moreover, this metric is the pullback of the Frobenius log-Euclidean metric ($\alpha = 1$ and $\beta = 0$) by the isometry $f_{p,q} : \Sigma \in \operatorname{Sym}^+(n) \mapsto \exp(F_{p,q}(\log \Sigma)) = \det(\Sigma)^{\frac{p-q}{n}} \Sigma^q \in \operatorname{Sym}^+(n)$ with $p = \sqrt{\alpha + n\beta}$ and $q = \sqrt{\alpha}$, where $F_{p,q}$ was defined in Theorem 4.5. Its Riemannian operations are detailed in Table 4.4.

Metric	$g_\Sigma(X, X) = \alpha \ d_\Sigma \log(X)\ _2^2 + \beta \operatorname{tr}(\Sigma^{-1}X)^2$
Squared distance	$d(\Sigma, \Lambda)^2 = \alpha \ \log \Lambda - \log \Sigma\ _2^2 + \beta \log(\det(\Lambda)/\det(\Sigma))^2$
Levi-Civita	$\nabla_X Y = \partial_X^{\log} Y$
Curvature	$R = 0$
Geodesics	$\forall t \in \mathbb{R}, \gamma_{(\Sigma, X)}(t) = \exp(\log(\Sigma) + t d_\Sigma \log(X))$
Logarithm	$\operatorname{Log}_\Sigma(\Lambda) = (d_\Sigma \log)^{-1}(\log \Lambda - \log \Sigma)$
Parallel transport	Does not depend on the curve: $\Pi_{\Sigma \rightarrow \Lambda} : \begin{cases} T_\Sigma \operatorname{Sym}^+(n) & \longrightarrow & T_\Lambda \operatorname{Sym}^+(n) \\ X & \longmapsto & (d_\Lambda \log)^{-1}(d_\Sigma \log(X)) \end{cases}$

Table 4.4: Riemannian operations of $O(n)$ -invariant log-Euclidean metrics on SPD matrices.

4.3.3 Affine-invariant metrics

Affine-invariant metrics were introduced in many different ways. We adopt here the most recent viewpoint [Pennec, 2009], which underlies the term ‘‘affine-invariant’’. Consider SPD matrices $\Sigma \in \operatorname{Sym}^+(n)$ as empirical covariance matrices of a random vector $X \in \mathbb{R}^n$, namely $\Sigma = \frac{1}{n}(X - \bar{X})(X - \bar{X})^\top$. Define the affine action on vectors $((A, B), X) \in (\operatorname{GL}(n) \times \mathbb{R}^n) \times \mathbb{R}^n \mapsto AX + B \in \mathbb{R}^n$. Then, the induced action on SPD matrices is $((A, B), \Sigma) \in (\operatorname{GL}(n) \times \mathbb{R}^n) \times \operatorname{Sym}^+(n) \mapsto A\Sigma A^\top \in \operatorname{Sym}^+(n)$. It is simply the congruence action of $\operatorname{GL}(n)$ on matrices. Hence an affine-invariant metric on SPD matrices simply designates a $\operatorname{GL}(n)$ -invariant metric.

Historically, Siegel introduced a metric on the half space $\mathcal{S} = \{X + i\Sigma \mid X \in \text{Sym}(n), \Sigma \in \text{Sym}^+(n)\}$ which is invariant under the action of the symplectic group [Siegel, 1943]. As a consequence, the restriction of this metric to SPD matrices by the immersion $\Sigma \in \text{Sym}^+(n) \mapsto i\Sigma \in \mathcal{S}$ was proved to be invariant under $\text{GL}(n)$ and under inversion and to provide a Riemannian homogeneous structure to $\text{Sym}^+(n)$. The expression of this metric is $g_\Sigma(X, Y) = \text{tr}(\Sigma^{-1}X\Sigma^{-1}Y)$.

Rao considered the Fisher information of a family of densities as a Riemannian metric on the space of parameters [Rao, 1945] and Skovgaard detailed all the properties of the Fisher-Rao metric of the family of multivariate Gaussian densities [Skovgaard, 1984]. By restriction to the family of *centered* multivariate Gaussian densities, we get the same metric as Siegel's scaled by a factor 1/2, namely $g_\Sigma(X, Y) = \frac{1}{2}\text{tr}(\Sigma^{-1}X\Sigma^{-1}Y)$. In addition, Amari and Nagaoka stated that the canonical immersion $\text{id} : \Sigma \in \text{Sym}^+(n) \mapsto \Sigma \in \text{Sym}(n)$ and the inversion $\text{inv} : \Sigma \in \text{Sym}^+(n) \mapsto \Sigma^{-1} \in \text{Sym}(n)$ give two dual coordinate systems with respect to this metric [Amari and Nagaoka, 2000].

Between 2005 and 2007, this metric was used in many computational methods for Diffusion Tensor Imaging [Pennec et al., 2006, Lenglet et al., 2006, Fletcher and Joshi, 2007, Moakher, 2005, Batchelor et al., 2005], in functional MRI [Varoquaux et al., 2010] and in Brain-Computer Interfaces [Barachant et al., 2013]. It was claimed to be the unique affine-invariant metric. However, Pennec showed that $\text{GL}(n)$ -invariant metrics are characterized by $\text{O}(n)$ -invariant inner products on the tangent space at I_n , that is on symmetric matrices. Hence from Theorem 4.5, there is actually a two-parameter family of affine-invariant metrics [Pennec, 2009].

Definition 4.9 (Affine-invariant metrics on SPD matrices) An *affine-invariant metric* on SPD matrices is a $\text{GL}(n)$ -invariant Riemannian metric. It is of the following form for all $\Sigma \in \text{Sym}^+(n)$ and $X \in \text{Sym}(n)$:

$$g_\Sigma^{A(\alpha, \beta)}(X, X) = \alpha \text{tr}((\Sigma^{-1}X)^2) + \beta \text{tr}(\Sigma^{-1}X)^2, \quad (4.8)$$

with $(\alpha, \beta) \in \mathbf{ST}$, i.e. $\alpha > 0$ and $\beta > -\alpha/n$. The Fisher-Rao metric often refers to the affine-invariant metric with $(\alpha, \beta) = (1/2, 0)$. Moreover, given $\alpha > 0$, this metric is the pullback of the affine-invariant metric with $\beta = 0$ by the isometry $f_{p,1} : \Sigma \in \text{Sym}^+(n) \mapsto \det(\Sigma)^{\frac{p-1}{n}} \Sigma \in \text{Sym}^+(n)$ with $p = \sqrt{\frac{\alpha+n\beta}{\alpha}}$.

The following proposition details the characteristics of homogeneity and symmetry of these Riemannian metrics. The Riemannian operations, essentially due to Skovgaard [Skovgaard, 1984], are detailed in Table 4.5. The second term of the sectional curvature is part of our contributions as it seems to be forgotten in [Skovgaard, 1984]. We prove the formula in Section 11.3.

Proposition 4.10 (Riemannian symmetric structure of the affine-invariant metric) The Riemannian manifold $(\text{Sym}^+(n), g^{A(\alpha, \beta)})$ is a Riemannian symmetric space, hence it is geodesically complete. The underlying homogeneous space is $\text{GL}^+(n)/\text{SO}(n)$ and $g^{A(\alpha, \beta)}$ is a quotient metric obtained by the submersion $\pi : A \in \text{GL}^+(n) \mapsto AA^\top \in \text{Sym}^+(n)$ from the left-invariant metric $G_A(M, M) = 4\alpha \text{tr}(A^{-1}M(A^{-1}M)^\top) + 4\beta \text{tr}(A^{-1}M)^2$ for $A \in \text{GL}^+(n)$ and $M \in T_A\text{GL}^+(n)$. The symmetries are $s_\Sigma : \Lambda \in \text{Sym}^+(n) \mapsto \Sigma\Lambda^{-1}\Sigma \in \text{Sym}^+(n)$.

Metric	$g_\Sigma(X, X) = \alpha \ \Sigma^{-1}X\ _2^2 + \beta \operatorname{tr}(\Sigma^{-1}X)^2$
Squared distance	$d(\Sigma, \Lambda)^2 = \alpha \ \log(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})\ _2^2 + \beta \log(\det(\Sigma^{-1}\Lambda))^2$
Levi-Civita	$(\nabla_X Y) _\Sigma = (\partial_X Y) _\Sigma - \frac{1}{2}(X\Sigma^{-1}Y + Y\Sigma^{-1}X)$
Curvature	The sectional curvature $\kappa \in [-1/2\alpha; 0]$. More precisely, the Riemann and sectional curvatures are: $R_\Sigma(X, Y, Z, T) = \frac{\alpha}{2}(X\Sigma^{-1}Y\Sigma^{-1}(Z\Sigma^{-1}T - T\Sigma^{-1}Z)\Sigma^{-1})$ $\kappa_\Sigma(\Sigma^{1/2}E_{ii}^\beta\Sigma^{1/2}, \Sigma^{1/2}E_{jj}^\beta\Sigma^{1/2}) = -1/4\alpha$ for $i \neq j$ $\kappa_\Sigma(\Sigma^{1/2}E_{ij}^\beta\Sigma^{1/2}, \Sigma^{1/2}E_{ik}^\beta\Sigma^{1/2}) = -1/8\alpha$ for $i \neq j \neq k \neq i$ where $E_{ij}^\beta = E_{ij} - \frac{1-p}{np}\delta_{ij}I_n$. Other terms are null.
Geodesics	$\forall t \in \mathbb{R}, \gamma_{(\Sigma, X)}(t) = \Sigma^{1/2} \exp(t\Sigma^{-1/2}X\Sigma^{-1/2})\Sigma^{1/2}$
Logarithm	$\operatorname{Log}_\Sigma(\Lambda) = \Sigma^{1/2} \log(\Sigma^{-1/2}\Lambda\Sigma^{-1/2})\Sigma^{1/2}$
Parallel transport	Depends on the curve. Along a geodesic: $\Pi_{\Sigma \rightarrow \Lambda} : \begin{cases} T_\Sigma \operatorname{Sym}^+(n) & \longrightarrow T_\Lambda \operatorname{Sym}^+(n) \\ X & \longmapsto (\Lambda\Sigma^{-1})^{1/2}X(\Sigma^{-1}\Lambda)^{1/2} \end{cases}$

Table 4.5: Riemannian operations of affine-invariant metrics on SPD matrices.

Another metric that also provides a Riemannian symmetric structure on $\operatorname{Sym}^+(n)$ was used in [Su et al., 2012, Zhang et al., 2018]. It was introduced directly by the quotient structure detailed in Proposition 4.10 but with the submersion $\sqrt{\pi} : A \in \operatorname{GL}^+(n) \mapsto \sqrt{AA^\top} \in \operatorname{Sym}^+(n)$ based on the polar decomposition of A (and without the coefficient 4). We called it the polar-affine metric in [Thanwerdas and Pennec, 2019b]. It is $\operatorname{GL}(n)$ -invariant with respect to the action $(A, \Sigma) \in \operatorname{GL}(n) \times \operatorname{Sym}^+(n) \mapsto \sqrt{A\Sigma^2A^\top} \in \operatorname{Sym}^+(n)$. Hence it is $O(n)$ -invariant in the usual sense. It is the pullback metric of the affine-invariant metric via the square diffeomorphism $\operatorname{pow}_2 : \Sigma \mapsto \Sigma^2$ [Thanwerdas and Pennec, 2019b].

4.3.4 Bures-Wasserstein metric

The L^2 -Wasserstein distance between multivariate centered Gaussian distributions is given by $d(\Sigma, \Lambda)^2 = \operatorname{tr}(\Sigma) + \operatorname{tr}(\Lambda) - 2\operatorname{tr}((\Sigma\Lambda)^{1/2})$. It corresponds to the Procrustes distance between square-root matrices, namely $d(\Sigma, \Lambda)^2 = \inf_{U \in O(n)} \|\Sigma^{1/2} - \Lambda^{1/2}U\|_{\operatorname{Frob}}^2$. The second order approximation of this squared distance defines a Riemannian metric called the Bures metric (or the Helstrom metric) in quantum physics. All these viewpoints are explained in details with modern notations in [Bhatia et al., 2019]. In particular, the expression of the Riemannian metric is derived in [Bhatia et al., 2019] and we take it as a definition.

Definition 4.11 (Bures-Wasserstein metric) The *Bures-Wasserstein metric* is the Riemannian metric associated to the Bures-Wasserstein distance. It is $O(n)$ -invariant and given an eigenvalue decomposition $\Sigma = PDP^\top \in \operatorname{Sym}^+(n)$ with $P \in O(n)$ and $D = \operatorname{diag}(d_1, \dots, d_n)$ and $X = PX'P^\top$, its expression is:

$$g_\Sigma^{\operatorname{BW}}(X, X) = g_D^{\operatorname{BW}}(X', X') = \frac{1}{2} \sum_{i,j} \frac{1}{d_i + d_j} X'_{ij}{}^2. \quad (4.9)$$

The Bures-Wasserstein metric can also be expressed by means of the linear map $\mathcal{S}_\Sigma : \operatorname{Sym}(n) \rightarrow \operatorname{Sym}(n)$ implicitly defined by the Sylvester equation $X = \Sigma\mathcal{S}_\Sigma(X) + \mathcal{S}_\Sigma(X)\Sigma$ for

$X \in \text{Sym}(n)$. More explicitly with the previous notations, we have $\mathcal{S}_\Sigma(X) = P \left[\frac{X'_{ij}}{d_i+d_j} \right]_{i,j} P^\top$. Then we have $g_\Sigma^{\text{BW}}(X, Y) = \frac{1}{2} \text{tr}(X \mathcal{S}_\Sigma(Y)) = \text{tr}(\mathcal{S}_\Sigma(X) \Sigma \mathcal{S}_\Sigma(Y))$, where $X, Y \in T_\Sigma \text{Sym}^+(n)$ are canonically identified with $d_\Sigma \text{id}(X), d_\Sigma \text{id}(Y) \in \text{Sym}(n)$, as explained in the introduction. This is a common expression in recent papers [Malagò et al., 2018, van Oostrum, 2020]. However, in [Takatsu, 2011] which is a reference paper on the Bures-Wasserstein metric, Takatsu gives the expression $g_\Sigma(X, Y) = \text{tr}(\mathbf{X} \Sigma \mathbf{Y})$. The trick comes from the identification $\mathcal{S}_\Sigma(X) \equiv \mathbf{X} \in \text{Sym}(n)$ that differs from the canonical one $d_\Sigma \text{id}(X) \equiv \mathbf{X} \in \text{Sym}(n)$. As this could be confusing when the formula is written without this precision (and without bold letters), we adopt the same formalism as in [Malagò et al., 2018, Bhatia et al., 2019, van Oostrum, 2020].

Bundle	$\text{GL}(n)$
Group action	$\rho : (A, U) \in \text{GL}(n) \times \text{O}(n) \mapsto AU \in \text{GL}(n)$
Submersion	$\pi : A \in \text{GL}(n) \mapsto \Sigma := AA^\top \in \text{Sym}^+(n)$
Vertical space	$\mathcal{V}_A = \ker d_A \pi = \text{Skew}(n) A^{-\top}$
Bundle metric	$G_A(M, M) = \text{tr}(MM^\top)$
Hor. space	$\mathcal{H}_A = \mathcal{V}_A^\perp = \text{Sym}(n)A$
Hor. isometry	$(d_A \pi) _{\mathcal{H}_A} : \begin{cases} \mathcal{H}_A = \text{Sym}(n)A & \longrightarrow & T_\Sigma \text{Sym}^+(n) \\ X^h = X^0 A & \longmapsto & X = \Sigma X^0 + X^0 \Sigma \end{cases}$
Sym. lift X^0	$\mathcal{S}_\Sigma : \begin{cases} T_\Sigma \text{Sym}^+(n) & \longrightarrow & \mathcal{H}_{I_n} = \text{Sym}(n) \\ X & \longmapsto & X^0 = P X^{0'} P^\top \text{ with } X^{0'}_{ij} = \frac{X'_{ij}}{d_i+d_j} \end{cases}$
Hor. lift X^h	$X \in T_\Sigma \text{Sym}^+(n) \mapsto X^h = X^0 A \in \mathcal{H}_A$

Table 4.6: Quotient structure of the Bures-Wasserstein metric.

We recall the quotient structure of the Bures-Wasserstein metric [Bhatia et al., 2019] in Table 4.6. The Riemannian operations are detailed in Table 4.7. Let us precise what was known and what is new in Table 4.7.

The proofs of the formulae of the distance and the logarithm can be found in [Bhatia et al., 2019]. The Levi-Civita connection and the exponential map were computed in [Malagò et al., 2018]. We computed the Levi-Civita connection independently using a more geometric proof that we provide in Section 11.3. We get a simpler formula.

Takatsu computed the curvature in [Takatsu, 2010] in a basis of vectors and gave a general formula in [Takatsu, 2011]. However, we argued above that the notations of [Takatsu, 2011] could be confusing because of the chosen identification. Moreover, the expression of the curvature given there is a bit implicit since it is $R_\Sigma(X, Y, X, Y) = \frac{3}{4} \text{tr}((|Y, X] - S) \Sigma (|Y, X] - S)^\top)$ where $S = \mathcal{S}_\Sigma([X, Y] \Sigma + \Sigma [Y, X]) \in \text{Sym}(n)$. Therefore, we prove in Section 11.3 the compact and explicit formula provided in Table 4.7 using the same method: O’Neill’s equations of submersions [O’Neill, 1966].

Finally, the geodesic parallel transport between commuting SPD matrices is new. The proof is given in Section 11.3. We provide a new formulation of the equation of the parallel transport between any two SPD matrices in the following proposition. It is used in the Python package `geomstats` [Miolane et al., 2020a] to compute the parallel transport.

Metric	$g_{\Sigma}(X, X) = g_{\Sigma^{1/2}}(X^h, X^h) = \frac{1}{2} \sum_{i,j} \frac{1}{d_i+d_j} X_{ij}^2$
Squared distance	$d(\Sigma, \Lambda)^2 = \text{tr}\Sigma + \text{tr}\Lambda - 2\text{tr}((\Sigma\Lambda)^{1/2})$
Levi-Civita	$(\nabla_X Y) _{\Sigma} = (\partial_X Y) _{\Sigma} - (X^0 \Sigma Y^0 + Y^0 \Sigma X^0)$
Curvature	The sectional curvature is non-negative. More precisely $R_{\Sigma}(X, Y, X, Y) = \frac{3}{2} \sum_{i,j} \frac{d_i d_j}{d_i+d_j} [X^{0'}, Y^{0'}]_{ij}^2$ where $[V, W] = VW - WV$ is the Lie bracket of matrices.
Geodesics	$\gamma_{(\Sigma, X)}(t) = \Sigma + tX + t^2 X^0 \Sigma X^0$ for $t \in I$ where I depends on $\lambda_{\max} = \max \text{sp}(X^0)$ and $\lambda_{\min} = \min \text{sp}(X^0)$ as follows: <ul style="list-style-type: none"> • If $\lambda_{\min} < 0 < \lambda_{\max}$, then $I = (-1/\lambda_{\max}, -1/\lambda_{\min})$. • If $0 \leq \lambda_{\min}$, then $I = (-1/\lambda_{\max}, +\infty)$. • If $\lambda_{\max} \leq 0$, then $I = (-\infty, -1/\lambda_{\min})$.
Logarithm	$\text{Log}_{\Sigma}(\Lambda) = (\Sigma\Lambda)^{1/2} + (\Lambda\Sigma)^{1/2} - 2\Sigma$
Parallel transport	Depends on the curve. Along a geodesic between commuting matrices $\Sigma = PDP^{\top}$ and $\Lambda = P\Delta P^{\top}$: $\Pi_{\Sigma \rightarrow \Lambda} : \begin{cases} T_{\Sigma} \text{Sym}^+(n) & \longrightarrow T_{\Lambda} \text{Sym}^+(n) \\ X & \longmapsto P \left[\sqrt{\frac{\delta_i + \delta_j}{d_i + d_j}} [P^{\top} X P]_{ij} \right] P^{\top} \end{cases}$

Table 4.7: Riemannian operations of the Bures-Wasserstein metric on SPD matrices.

Proposition 4.12 (Parallel transport equation of Bures-Wasserstein metric) Let $\gamma(t)$ the geodesic between $\gamma(0) = \Sigma$ and $\gamma(1) = \Lambda$, and a vector $X \in T_{\Sigma} \text{Sym}^+(n)$. We denote $\gamma^h(t) = (1-t)\Sigma^{1/2} + t\Sigma^{-1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}$ the horizontal lift of the geodesic γ . The two following statements are equivalent.

- (i) The vector field $X(t)$ defined along $\gamma(t)$ is the parallel transport of X .
- (ii) $X(t) = \gamma(t)X^0(t) + X^0(t)\gamma(t)$ where $X^0(t)$ is a curve in $\text{Sym}(n)$ satisfying the following ODE:

$$\gamma(t)\dot{X}^0(t) + \dot{X}^0(t)\gamma(t) + \gamma^h(t)\dot{\gamma}^{h\top} X^0(t) + X^0(t)\dot{\gamma}^h \gamma^h(t)^{\top} = 0. \quad (4.10)$$

See the proof of Proposition 4.12 in Section 11.3.

Interestingly, this equation resembles the parallel transport equation on the Kendall shape space, which is also a quotient space, derived in [Kim et al., 2021, Proposition 3.1].

4.3.5 Bogoliubov-Kubo-Mori metric

The Bogoliubov-Kubo-Mori metric is a Riemannian metric used in quantum physics [Petz and Toth, 1993], given by $g_{\Sigma}^{\text{BKM}}(X, X) = \text{tr}(\int_0^{\infty} (\Sigma + tI_n)^{-1} X (\Sigma + tI_n)^{-1} X dt)$. It can be seen as the integration of the affine-invariant metric on a half-line included in the SPD cone. It can be rewritten thanks to the differential of the logarithm and we take this other expression as a definition.

Definition 4.13 (Bogoliubov-Kubo-Mori (BKM) metric) The *Bogoliubov-Kubo-Mori metric*

is the $O(n)$ -invariant Riemannian metric defined for $\Sigma \in \text{Sym}^+(n)$ and $X \in T_\Sigma \text{Sym}^+(n)$ by:

$$g_\Sigma^{\text{BKM}}(X, X) = \text{tr}(X d_\Sigma \log(X)). \quad (4.11)$$

Important functions related to this metric are defined by [Michor et al., 2000] to get simple expressions of the Levi-Civita connection and the curvature. Given $\Sigma = PDP^\top \in \text{Sym}^+(n)$, they define $m_{ij} = \int_0^\infty (d_i + t)^{-1}(d_j + t)^{-1} dt$ which is symmetric in (i, j) and $m_{ijk} = \int_0^\infty (d_i + t)^{-1}(d_j + t)^{-1}(d_k + t)^{-1} dt$ which is symmetric in (i, j, k) . They also denote $g_\Sigma(X) = d_\Sigma \log(X)$ whose expression is $g_\Sigma(X) = P g_D(X') P^\top$ and $[g_D(X')]_{ij} = m_{ij} X'_{ij}$ where $X' = P^\top X P$. This g_Σ is defined so that $g_\Sigma(X, Y) = \text{tr}(X g_\Sigma(Y))$. By differentiating this equality and using the definition of the BKM metric, they get the differential of $\Sigma \mapsto g_\Sigma$:

$$\begin{aligned} d_\Sigma g(PF_{ij}P^\top)(PF_{kl}P^\top) &= d_D g(F_{ij})(F_{kl}) \\ &= -\frac{1}{2}(\delta_{jk} m_{ilj} F_{il} + \delta_{jl} m_{ikj} F_{ik} + \delta_{il} m_{jki} F_{jk} + \delta_{ik} m_{jli} F_{jl}), \end{aligned}$$

or more compactly $[d_\Sigma g(PXP^\top)(PXP^\top)]_{ij} = -2 \sum_{k=1}^n m_{ijk} X_{ik} X_{jk}$. The Levi-Civita connection and the curvature can be expressed in closed forms by means of g and dg , as shown in Table 4.8. Note that the sign of the sectional curvature is not known. The distance, exponential, logarithm and parallel transport maps are not known either.

Metric	$g_\Sigma(X, X) = \text{tr}(X d_\Sigma \log(X))$
Levi-Civita	$(\nabla_X Y) _\Sigma = (\partial_X Y) _\Sigma + \frac{1}{2} g_\Sigma^{-1}(d_\Sigma g(X)(Y))$
Curvature	$R_\Sigma(X, Y)Z = -\frac{1}{4} g_\Sigma^{-1}(d_\Sigma g(X)(g_\Sigma^{-1}(d_\Sigma g(Y)(Z))))$ $+ \frac{1}{4} g_\Sigma^{-1}(d_\Sigma g(Y)(g_\Sigma^{-1}(d_\Sigma g(X)(Z))))$

Table 4.8: Riemannian operations of the BKM metric on SPD matrices.

In this section, we reviewed five of the mainly used $O(n)$ -invariant Riemannian metrics and we contributed new formulae. We also highlighted that the $O(n)$ -invariant Euclidean, the $O(n)$ -invariant log-Euclidean and the affine-invariant metrics are actually two-parameter families of Riemannian metrics indexed by $(\alpha, \beta) \in \mathbf{ST}$ while this extra term weighted by the trace factor β is never defined in the literature for the Bures-Wasserstein and the Bogoliubov-Kubo-Mori metrics. Actually, there does not seem to exist a natural way of extending them with a trace term. Indeed, under the Bures-Wasserstein metric, there is a choice of an $O(n)$ -right-invariant inner product on $\text{GL}(n)$ but they differ from $O(n)$ -invariant inner products on symmetric matrices given in Theorem 4.5. Indeed, any inner product on $\text{GL}(n)$ of the form $\langle X|X \rangle = \text{tr}(X^\top S X)$ with $S \in \text{Sym}^+(n)$ is $O(n)$ -right-invariant. As for the BKM metric, we could change the inner product in the integral but after computation, we would obtain this metric: $\alpha g_\Sigma^{\text{BKM}}(X, X) + \beta \sum_{i,j} \log^{[1]}(d_i, d_j) X'_{ii} X'_{jj}$. The fact that we cannot separate the indices i and j in the trace term differs from the previous situations.

In the next section, we recall the definition of the class of kernel metrics [Hiai and Petz, 2009, Hiai and Petz, 2012] and a selection of its key properties. Since this class of Riemannian metrics contains all the previously introduced metrics without trace term, we show that this is the right framework to define the trace term extension. We show that this new class of extended kernel metrics still satisfies the key results on kernel metrics we selected. We also prove another property of these two classes: the stability under the cometric.

4.4 The interpolating class of kernel metrics: new observations

Kernel metrics were introduced by Hiai and Petz in 2009 [Hiai and Petz, 2009]. It is a family of $O(n)$ -invariant metrics indexed by smooth bivariate functions $\phi : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$ called kernels. It has several key properties and it encompasses all the $O(n)$ -invariant metrics introduced in Section 4.3 without trace term ($\beta = 0$). After recalling these key results (Section 4.4.1), we provide new observations on kernel metrics (Section 4.4.2), especially the trace term extension and the stability under the cometric.

4.4.1 The general class of kernel metrics

Definition 4.14 (Kernel metrics, mean kernel metrics) [Hiai and Petz, 2009] A *kernel metric* is an $O(n)$ -invariant metric for which there is a smooth bivariate map $\phi : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$ such that $g_\Sigma(X, X) = g_D(X', X') = \sum_{i,j} \frac{1}{\phi(d_i, d_j)} X'_{ij}{}^2$, where $\Sigma = PDP^\top$ with $P \in O(n)$ and $D = \text{Diag}(d_1, \dots, d_n)$, and $X = PX'P^\top$.

A *mean kernel metric* is a kernel metric characterized by a bivariate map ϕ of the form $\phi(x, y) = a m(x, y)^\theta$ where $a > 0$ is a positive coefficient, $\theta \in \mathbb{R}$ is a homogeneity power and $m : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$ is a symmetric homogeneous mean, that is:

1. symmetric, i.e. $m(x, y) = m(y, x)$ for all $x, y > 0$,
2. homogeneous, i.e. $m(\lambda x, \lambda y) = \lambda m(x, y)$ for all $\lambda, x, y > 0$,
3. non-decreasing in both variables,
4. $\min(x, y) \leq m(x, y) \leq \max(x, y)$ for all $x, y > 0$. It implies $m(x, x) = x$.

As the goal of this chapter is to extend the class of kernel metrics, we selected from [Hiai and Petz, 2009, Hiai and Petz, 2012] the results that we found simple and powerful to be able to generalize them later on. It would be interesting to study other properties such as monotonicity and comparison properties but it is beyond our scope. Our selection of results is in Proposition 4.15.

Proposition 4.15 (Key results on kernel metrics) [Hiai and Petz, 2009]

1. (Generality) The Euclidean, log-Euclidean and affine-invariant metrics without trace term ($\beta = 0$), the polar-affine, the Bures-Wasserstein and the Bogoliubov-Kubo-Mori metrics are mean kernel metrics. The kernels and the names of the corresponding means are given in Table 4.9.
2. (Stability) The class of kernel metrics is stable under univariate diffeomorphisms. More precisely, if g is a kernel metric with kernel function ϕ and if f is a univariate diffeomorphism (defined in Section 4.2.2), then the pullback metric f^*g is a kernel metric with bivariate function $(x, y) \mapsto \frac{\phi(f(x), f(y))}{f^{[1]}(x, y)^2}$. Note that the class of *mean* kernel metrics is not stable under univariate diffeomorphisms because of the non-decreasing property required for mean kernel metrics.

Metric	$\phi(x, y)$	Mean m	θ
Euclidean	1	Any mean	0
Log-Euclidean	$(\frac{x-y}{\log(x)-\log(y)})^2$	Logarithmic mean	2
Affine-invariant	xy	Geometric mean	2
Polar-affine	$(\frac{2xy}{x+y})^2$	Harmonic mean	2
Bures-Wasserstein	$4 \frac{x+y}{2}$	Arithmetic mean	1
BKM	$\frac{x-y}{\log(x)-\log(y)}$	Logarithmic mean	1

Table 4.9: Bivariate functions of all the $O(n)$ -invariant metrics of Section 4.3.

3. (Completeness) A mean kernel metric with homogeneity power θ is geodesically complete if and only if $\theta = 2$. Therefore this result provides a sufficient condition for kernel metrics to be geodesically complete.

Another property that we left for a different reason is the attractivity of the Log-Euclidean metric, i.e. the fact that the log-Euclidean metric is the limit when p tends to 0 of the pullback of a kernel metric by a power diffeomorphism $\text{pow}_p : \Sigma \in \text{Sym}^+(n) \mapsto \Sigma^p \in \text{Sym}^+(n)$, scaled by $\frac{1}{p^2}$. However, it is not specific to kernel metrics since this is the case for any metric g .

4.4.2 New observations on kernel metrics

4.4.2.1 Kernel metrics form a cone

The class of kernel metrics is a sub-cone of the cone of Riemannian metrics on the SPD manifold. Indeed, it is stable by positive scaling and it is convex because if g, g' are kernel metrics associated to ϕ, ϕ' , then $(1-t)g + tg'$ is a kernel metric associated to $\phi\phi' / ((1-t)\phi' + t\phi) > 0$ for $t \in [0, 1]$.

4.4.2.2 Cometric stability of the class of kernel metrics

A Riemannian metric $g : T\mathcal{M} \times T\mathcal{M} \rightarrow \mathbb{R}$ on a manifold \mathcal{M} defines a cometric $g^* : T^*\mathcal{M} \times T^*\mathcal{M} \rightarrow \mathbb{R}$ defined for all covectors $\omega, \omega' \in T^*\mathcal{M}$ by $g^*(\omega, \omega') = \omega(x')$ where $x' \in T\mathcal{M}$ is the unique vector such that for all vectors $x \in T\mathcal{M}$, $g(x, x') = \omega'(x)$ (Riesz's theorem).

On the manifold of SPD matrices $\mathcal{M} = \text{Sym}^+(n)$, we have a canonical identification of $T_\Sigma\mathcal{M}$ with $\text{Sym}(n)$ given by $d_\Sigma\text{id}$. Hence by duality, we also have a canonical identification between $T_\Sigma^*\mathcal{M}$ and $\text{Sym}(n)^*$. So to identify $T_\Sigma\mathcal{M}$ with $T_\Sigma^*\mathcal{M}$, we only need an identification between $\text{Sym}(n)$ and $\text{Sym}(n)^*$. This is provided by the Frobenius inner product. To summarize, there is a natural identification between the tangent space and the cotangent space given by:

$$\left\{ \begin{array}{l} T_\Sigma\text{Sym}^+(n) \longrightarrow T_\Sigma^*\text{Sym}^+(n) \\ X \longmapsto (Y \in T_\Sigma\text{Sym}^+(n) \longmapsto \text{tr}(d_\Sigma\text{id}(X)d_\Sigma\text{id}(Y))) \end{array} \right. \quad (4.12)$$

Hence, a cometric on SPD matrices can be seen as a metric.

Back to kernel metrics, it is interesting to note that this class is stable under taking the cometric and that the cometric has a simple expression.

Proposition 4.16 (Cometric stability of kernel metrics) Let g be a kernel metric with kernel function ϕ . Then the cometric g^* seen as a metric through the identification explained above is a kernel metric with kernel function $\phi^* = 1/\phi$.

This elementary fact is interesting from a numerical point of view. Indeed, to compute numerically the geodesics, one can either integrate the geodesic equation involving the Christoffel symbols (which is of second order) or integrate its Hamiltonian version involving the cometric (which is of first order). Hence, the fact that the cometric of a kernel metric is available is a quite important result that appeared to be previously unnoticed. More precisely, the geodesic equation writes $\ddot{x}^k + \Gamma_{ij}^k \dot{x}^i \dot{x}^j = 0$ where $x(t)$ is a curve on the manifold \mathcal{M} and Γ_{ij}^k are the Christoffel symbols related to the metric by $\Gamma_{ij}^k = \frac{1}{2}g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})$. By considering a curve $p(t)$ on the cotangent bundle $T^*\mathcal{M}$ instead, and $x(t)$ the curve on the manifold \mathcal{M} such that $p(t) \in T_{x(t)}^*\mathcal{M}$, the geodesic equation admits the following Hamiltonian formulation:

$$\begin{cases} \dot{x}^k = g^{kl} p_l \\ \dot{p}_l = -\frac{1}{2} \frac{\partial g^{ij}}{\partial x^l} p_i p_j \end{cases} \quad (4.13)$$

The Hamiltonian equation is often preferred to compute the geodesics numerically since the integration is simpler and more stable. It only involves the cometric $g^* = (g^{ij})_{i,j}$, which is very easy to compute for a kernel metric. This is how geodesics are generically computed in the python package `geomstats` for example.

4.4.2.3 Canonical Frobenius-like expression of a kernel metric

An expression of kernel metrics was given in [Hiai and Petz, 2009] by means of the operators $\mathbb{L}_\Sigma : X \mapsto \Sigma X$, $\mathbb{R}_\Sigma : X \mapsto X \Sigma$ and $\phi(\mathbb{L}_\Sigma, \mathbb{R}_\Sigma) : \text{Sym}(n) \rightarrow \text{Sym}(n)$ defined for $\Sigma = PDP^\top \in \text{Sym}^+(n)$ by $\phi(\mathbb{L}_\Sigma, \mathbb{R}_\Sigma)X = P([\phi(d_i, d_j)]_{i,j} \circ (P^\top X P)) P^\top$, where \circ denotes the Schur (entry-wise) product. This expression is $g_\Sigma^\phi(X, X) = \text{tr}(X \phi(\mathbb{L}_\Sigma, \mathbb{R}_\Sigma)^{-1}(X))$. The existence of the map $\Phi : \text{Sym}^+(n) \times \text{Sym}(n) \rightarrow \text{Sym}(n)$ hidden in $\phi(\mathbb{L}_\Sigma, \mathbb{R}_\Sigma) := \Phi_\Sigma$ is ensured by extending the $O(n)$ -equivariant map $\Phi : \text{Diag}^+(n) \times \text{Sym}(n) \rightarrow \text{Sym}(n)$ defined by $[\Phi_D(X)]_{ij} = \phi(d_i, d_j)X_{ij}$. Indeed, one can easily check that Φ satisfies the two hypotheses of Lemma 4.2. In this work, we even prefer to define the bivariate map $\psi = \phi^{-1/2}$ and define in an analogous way the map $\Psi : \text{Sym}^+(n) \times \text{Sym}(n) \rightarrow \text{Sym}(n)$ so that we can write the kernel metric with a suitable Frobenius-like expression:

$$g_\Sigma^\phi(X, X) = \text{tr}(\Psi_\Sigma(X)^2). \quad (4.14)$$

We can give explicitly Ψ in some particular cases:

1. Euclidean metric: $\Psi_\Sigma^E(X) = X$;
2. log-Euclidean metric: $\Psi_\Sigma^{\text{LE}}(X) = d_\Sigma \log(X)$;
3. affine-invariant metric: $\Psi_\Sigma^A(X) = \Sigma^{-1/2} X \Sigma^{-1/2}$.

This is an important step towards the trace term extension.

4.4.2.4 Kernel metrics with a trace term

The class of kernel metrics does not encompass the $O(n)$ -invariant Euclidean, $O(n)$ -invariant log-Euclidean and affine-invariant metrics with a trace factor $\beta \neq 0$. However, thanks to the previous canonical expression, we can define a natural extension of a kernel metric with a trace term.

Definition 4.17 (Extended kernel metrics) Let g^ϕ be a kernel metric associated to the kernel function $\phi : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$. We define the map $\psi = \phi^{-1/2}$ and the map $\Psi : \text{Sym}^+(n) \times \text{Sym}(n) \rightarrow \text{Sym}(n)$ as described above so that $g_\Sigma(X, X) = \text{tr}(\Psi_\Sigma(X)^2)$. We define a two-parameter family which extends the kernel metric g^ϕ for all $\Sigma \in \text{Sym}^+(n)$ and $X \in \text{Sym}(n)$ by:

$$g_\Sigma^{\phi, \alpha, \beta}(X, X) = \alpha \text{tr}(\Psi_\Sigma(X)^2) + \beta \text{tr}(\Psi_\Sigma(X))^2, \quad (4.15)$$

where $(\alpha, \beta) \in \mathbf{ST}$, i.e. $\alpha > 0$ and $\alpha + n\beta > 0$.

We can apply this definition to the Bures-Wasserstein and the BKM metrics. One can show that the trace term such defined is $\beta \text{tr}(\Sigma^{-1/2}X)^2$. Contrarily to the log-Euclidean and the affine-invariant cases, there is no isometry a priori between two metrics of the family. It is interesting to note that Propositions 4.15 and 4.16 are still valid for these extended kernel metrics. We omit the proofs since they are analogous to the ones given for kernel metrics in [Hiai and Petz, 2009].

Proposition 4.18 (Key results on extended kernel metrics)

1. (Generality) All the metrics in Section 4.3 are extended kernel metrics.
2. (Stability) The class of extended kernel metrics is stable under univariate diffeomorphisms and the transformation is the same as in Proposition 4.15.
3. (Completeness) An extended mean kernel metric with homogeneity power θ is geodesically complete if and only if $\theta = 2$.
4. (Cometric) The class of extended kernel metrics is cometric-stable and the corresponding transformation is $(\phi, \alpha, \beta) \mapsto (\frac{1}{\phi}, \frac{1}{\alpha}, -\frac{\beta}{\alpha(\alpha+n\beta)})$.

In this section, we recalled the definition of kernel metrics and three key properties. We added the property of stability under the cometric with an explicit expression and we argued that it is an interesting property from a numerical point of view to compute geodesics. We found a wider class of metrics which satisfies the same key properties and which encompasses all the $O(n)$ -invariant metrics defined in Section 4.3. It is now tempting to look for wider classes of $O(n)$ -invariant metrics and to determine if these properties are still valid.

In the next section, we characterize $O(n)$ -invariant metrics by means of three multivariate functions satisfying conditions of symmetry, compatibility and positivity. This result allows to understand better the specificity of kernel metrics and extended kernel metrics within the whole class of $O(n)$ -invariant metrics. Then we give a counterpart of Proposition 4.18 and we propose a new intermediate class of $O(n)$ -invariant metrics which is cometric stable.

4.5 Characterization of $O(n)$ -invariant metrics

In this section, we give a characterization of $O(n)$ -invariant metrics on SPD matrices. We present it as an extension of Theorem 4.5 characterizing $O(n)$ -invariant inner products on symmetric matrices. Instead of two parameters α, β which satisfy a positivity condition, an $O(n)$ -invariant metric is characterized by three multivariate functions $\alpha, \beta, \gamma : (\mathbb{R}^+)^n \rightarrow \mathbb{R}$ which satisfy a positivity condition plus a symmetry condition and a compatibility condition. This is explained in Section 4.5.1. We also give two corollary results which characterize two subclasses of $O(n)$ -invariant metrics with additional invariances: scaling invariance and inverse-consistency. Section 4.5.2 is dedicated to the proof of the theorem. In Section 4.5.3, we reinterpret kernel metrics in light of the theorem. In Section 4.5.4, we give key results on $O(n)$ -invariant metrics and we compare them to those on kernel metrics given in Proposition 4.15. In particular, we state that the cometric can be difficult to compute. Hence in Section 4.5.5, we introduce the class of bivariate separable metrics which is an intermediate class between $O(n)$ -invariant and extended kernel metrics, which is cometric-stable and for which the cometric is known in closed-form.

4.5.1 Theorem and corollaries

Let us rephrase the characterization of $O(n)$ -invariant inner products on $\text{Sym}(n)$ (Theorem 4.5). An inner product $\langle \cdot | \cdot \rangle$ on $\text{Sym}(n)$ is $O(n)$ -invariant if and only if there exist real numbers $\gamma, \alpha > 0$ and $\beta \in \mathbb{R}$ such that:

$$\langle X | X \rangle = \gamma \sum_i X_{ii}^2 + \alpha \sum_{i \neq j} X_{ij}^2 + \beta \sum_{i \neq j} X_{ii} X_{jj}, \quad (4.16)$$

1. (Compatibility) $\gamma = \alpha + \beta$,
2. (Positivity) the symmetric matrix S defined by $S_{ii} = \gamma$ and $S_{ij} = \beta$ is positive definite.

The characterization of $O(n)$ -invariant metrics on $\text{Sym}^+(n)$ has an analogous form where real numbers are replaced by n -multivariate functions and where there is an additional property of symmetry of these functions. We introduce this notion of symmetry before stating the theorem. The proof is in Section 4.5.2.

Definition 4.19 ($(k, n - k)$ -symmetric functions) We say that a function $f : (\mathbb{R}^+)^n \rightarrow \mathbb{R}$ is $(k, n - k)$ -symmetric if it is symmetric in its k first variables and symmetric in its $n - k$ last variables. In other words, f is invariant under permutations $\sigma = \sigma_1 \sigma_2$ where σ_1 has support in $\{1, \dots, k\}$ and σ_2 has support in $\{k + 1, \dots, n\}$. Hence, given a set $I \subseteq \{1, \dots, n\}$ of cardinal k and $d \in (\mathbb{R}^+)^n$, we denote $f(d_{i \in I}, d_{i \notin I}) := f(\sigma \cdot d)$ where $\sigma(\{1, \dots, k\}) = I$ and $(\sigma \cdot d)_i = d_{\sigma(i)}$.

Theorem 4.20 (Characterization of $O(n)$ -invariant metrics) Let g be a Riemannian metric on $\text{Sym}^+(n)$. If g is $O(n)$ -invariant, then there exist three maps $\gamma, \alpha : (\mathbb{R}^+)^n \rightarrow \mathbb{R}^+$ and $\beta : (\mathbb{R}^+)^n \rightarrow \mathbb{R}$ such that for all $\Sigma = PDP^\top \in \text{Sym}^+(n)$ and $X = PX'P^\top \in T_\Sigma \text{Sym}^+(n)$:

$$\begin{aligned} g_\Sigma(X, X) &= g_D(X', X') \\ &= \sum_i \gamma(d_i, d_{k \neq i}) X_{ii}'^2 + \sum_{i \neq j} \alpha(d_i, d_j, d_{k \neq i, j}) X_{ij}'^2 + \sum_{i \neq j} \beta(d_i, d_j, d_{k \neq i, j}) X_{ii}' X_{jj}', \end{aligned} \quad (4.17)$$

0. (Symmetry) γ is $(1, n - 1)$ -symmetric and α, β are $(2, n - 2)$ -symmetric,
1. (Compatibility) γ equals $\alpha + \beta$ on the set $\mathcal{D} = \{d \in (\mathbb{R}^+)^n \mid d_1 = d_2\}$,
2. (Positivity) for all $d \in (\mathbb{R}^+)^n$, the symmetric matrix $S(d)$ defined by $S_{ii}(d) = \gamma(d_i, d_{k \neq i})$ and $S_{ij}(d) = \beta(d_i, d_j, d_{k \neq i, j})$ is positive definite.

Conversely, if there exist such maps α, β, γ , then Equation (4.17) correctly defines an $O(n)$ -invariant Riemannian metric that we denote $g^{\alpha, \beta, \gamma}$ or equivalently $g^{\alpha, S}$. Moreover, g is continuous if and only if α, β, γ are continuous.

Before giving the proof, we observe that this theorem allows to characterize subclasses of $O(n)$ -invariant metrics as well. Here we give the general form of $O(n)$ -invariant metrics that are invariant under scaling and under inversion respectively. We omit the proof.

Proposition 4.21 (Characterizations of subclasses of $O(n)$ -invariant metrics) Let g be an $O(n)$ -invariant metric characterized by the maps α, β, γ .

1. g is invariant under scaling if and only if $f(\lambda d) = \frac{1}{\lambda^2} f(d)$ for $f \in \{\alpha, \beta, \gamma\}$, for all $d \in (\mathbb{R}^+)^n$ and for all $\lambda > 0$.
2. g is invariant under inversion if and only if $\gamma(d_1^{-1}, \dots, d_n^{-1}) = d_1^4 \gamma(d_1, \dots, d_n)$ and $f(d_1^{-1}, \dots, d_n^{-1}) = d_1^2 d_2^2 f(d_1, \dots, d_n)$ for $f \in \{\alpha, \beta\}$, for all $d \in (\mathbb{R}^+)^n$.

4.5.2 Proof of the theorem

Proof of Theorem 4.20 (Characterization of $O(n)$ -invariant metrics). Let g be an $O(n)$ -invariant metric on $\text{Sym}^+(n)$. Since any diagonal matrix D is invariant under the subgroup $\mathcal{D}^\pm(n)$, the inner product g_D is $\mathcal{D}^\pm(n)$ -invariant. Hence, Lemma 4.6 (a) ensures that there are positive coefficients $\alpha_{ij}(D) = \alpha_{ji}(D)$ and a matrix $S(D) \in \text{Sym}^+(n)$ s.t. $g_D(X, X) = \sum_{i \neq j} \alpha_{ij}(D) X_{ij}^2 + \sum_{i, j} S_{ij}(D) X_{ii} X_{jj}$. Then, we define the three maps:

- $\alpha : d \in (\mathbb{R}^+)^n \mapsto \alpha_{12}(\text{Diag}(d)) > 0$,
- $\beta : d \in (\mathbb{R}^+)^n \mapsto S_{12}(\text{Diag}(d))$,
- $\gamma : d \in (\mathbb{R}^+)^n \mapsto S_{11}(\text{Diag}(d)) > 0$.

Following the same idea as in the proof of Lemma 4.6 (b), we use the invariance under permutations since $\text{Diag}^+(n)$ is stable under this action. Then, one easily checks that α, β are $(2, n - 2)$ -symmetric and γ is $(1, n - 1)$ -symmetric and that we can express the other coefficients in function of α, β, γ by permuting the d_i 's. We get for $i \neq j$:

- $\alpha_{ij}(\text{Diag}(d)) = \alpha(d_i, d_j, d_{k \neq i, j})$,
- $S_{ij}(\text{Diag}(d)) = \beta(d_i, d_j, d_{k \neq i, j})$,
- $S_{ii}(\text{Diag}(d)) = \gamma(d_i, d_{k \neq i})$.

So we get the expression (4.17), the symmetry and the positivity conditions. We only miss the compatibility condition so let $d = (d_1, \dots, d_n) \in (\mathbb{R}^+)^n$ such that $d_1 = d_2$. Since $D = \text{Diag}(d)$ is stable under any block-diagonal orthogonal matrix $R = \text{Diag}(R_\theta, I_{n-2}) \in O(n)$ with $R_\theta \in O(2)$, with the same computations as in the proof of Theorem 4.5, we get $\gamma(d) = \alpha(d) + \beta(d)$.

Conversely, if α, β, γ are three maps satisfying the conditions of symmetry, compatibility and positivity, then we define $g_D(X, X) = \sum_i \gamma(d_i, d_{k \neq i}) X_{ii}^2 + \sum_{i \neq j} \alpha(d_i, d_j, d_{k \neq i, j}) X_{ij}^2 + \sum_{i \neq j} \beta(d_i, d_j, d_{k \neq i, j}) X_{ii} X_{jj}$. In other words, we define a map $g : \text{Diag}^+(n) \times \text{Sym}(n) \times \text{Sym}(n) \rightarrow \mathbb{R}$ and we would like to extend it by defining $g_{PDP^\top}(X, X) = g_D(P^\top X P, P^\top X P)$. According to Lemma 4.2, we have two cases to study. One can easily show that the first condition with permutations is satisfied. The non-trivial condition is the second one, involving a diagonal matrix $D = \text{Diag}(\lambda_1 I_{m_1}, \dots, \lambda_p I_{m_p})$ with sorted diagonal values $\lambda_1 > \dots > \lambda_p > 0$ and a block-diagonal orthogonal matrix $R = \text{Diag}(R_1, \dots, R_p) \in O(n)$ with $R_k \in O(m_k)$. So we have to show that $g_D(R^\top X R, R^\top X R) = g_D(X, X)$ for all matrix $X \in \text{Sym}(n)$, since $R^\top D R = D$. We denote $\bar{X}^{kl} \in \text{Mat}(m_k, m_l)$ the (k, l) block matrix defined by $\bar{X}_{ij}^{kl} = X_{n_{k-1}+i, n_{l-1}+j}$ where $n_k = \sum_{j=1}^k m_j$. Note that $\bar{X}^{kk} \in \text{Sym}(m_k)$ is the k -th diagonal block of X and $\bar{X}^{lk} = (\bar{X}^{kl})^\top$. Therefore $\overline{R^\top X R}^{kl} = R_k^\top \bar{X}^{kl} R_l$. In the following, we split the sums between the blocks with multiplicity 1 and the blocks with higher multiplicity and we use the compatibility condition. The notation $\alpha(\lambda_k, \lambda_l, \dots)$ stands for $\alpha(d_i, d_j, d_{m \neq i, j})$ where $\lambda_k = d_i$ and $\lambda_l = d_j$, i.e. $n_{k-1} + 1 \leq i \leq n_k$ and $n_{l-1} + 1 \leq j \leq n_l$. We compute the difference:

$$\begin{aligned}
& g_D(R^\top X R, R^\top X R) - g_D(X, X) \\
&= \sum_{k:m_k=1} \gamma(d_{n_k}, d_{m \neq n_k}) \underbrace{((R^\top X R)_{n_k n_k}^2 - X_{n_k n_k}^2)}_0 \\
&+ \sum_{\substack{k \neq l \\ m_k = m_l = 1}} \alpha(\lambda_k, \lambda_l, \dots) \underbrace{((R^\top X R)_{n_k n_l}^2 - X_{n_k n_l}^2)}_0 \\
&+ \sum_{\substack{k \neq l \\ m_k = m_l = 1}} \beta(\lambda_k, \lambda_l, \dots) \underbrace{((R^\top X R)_{n_k n_k} (R^\top X R)_{n_l n_l} - X_{n_k n_k} X_{n_l n_l})}_0 \\
&+ \sum_{k:m_k > 1} \underbrace{\gamma(\lambda_k, \lambda_k, \dots)}_{\alpha(\lambda_k, \lambda_k, \dots) + \beta(\lambda_k, \lambda_k, \dots)} \sum_{i=n_{k-1}+1}^{n_k} ((R^\top X R)_{ii}^2 - X_{ii}^2) \\
&+ \sum_{\substack{k, l \\ m_k \text{ or } m_l > 1}} \alpha(\lambda_k, \lambda_l, \dots) \sum_{\substack{n_{k-1}+1 \leq i \leq n_k \\ n_{l-1}+1 \leq j \leq n_l \\ i \neq j}} ((R^\top X R)_{ij}^2 - X_{ij}^2) \\
&+ \sum_{\substack{k, l \\ m_k \text{ or } m_l > 1}} \beta(\lambda_k, \lambda_l, \dots) \sum_{\substack{n_{k-1}+1 \leq i \leq n_k \\ n_{l-1}+1 \leq j \leq n_l \\ i \neq j}} ((R^\top X R)_{ii} (R^\top X R)_{jj} - X_{ii} X_{jj}).
\end{aligned}$$

Hence the missing term $i = j$ in the two last sums is provided by the sum weighted by γ .

After a change of indexes based on the equality $\overline{R^\top X R}^{kl} = R_k^\top \bar{X}^{kl} R_l$, we get:

$$\begin{aligned}
& g_D(R^\top X R, R^\top X R) - g_D(X, X) \\
&= \sum_{\substack{k,l \\ m_k \text{ or } m_l > 1}} \alpha(\lambda_k, \lambda_l, \dots) \underbrace{\sum_{i=1}^{m_k} \sum_{j=1}^{m_l} ((R_k^\top \bar{X}^{kl} R_l)_{ij}^2 - (\bar{X}^{kl})_{ij}^2)}_{\text{tr}(R_k^\top \bar{X}^{kl} R_l (R_k^\top \bar{X}^{kl} R_l)^\top) - \text{tr}(\bar{X}^{kl} (\bar{X}^{kl})^\top) = 0} \\
&+ \sum_{\substack{k,l \\ m_k \text{ or } m_l > 1}} \beta(\lambda_k, \lambda_l, \dots) \underbrace{\sum_{i=1}^{m_k} \sum_{j=1}^{m_l} ((R_k^\top \bar{X}^{kk} R_k)_{ii} (R_l^\top \bar{X}^{ll} R_l)_{jj} - \bar{X}_{ii}^{kk} \bar{X}_{jj}^{ll})}_{\text{tr}(R_k^\top \bar{X}^{kk} R_k) \text{tr}(R_l^\top \bar{X}^{ll} R_l) - \text{tr}(\bar{X}^{kk}) \text{tr}(\bar{X}^{ll}) = 0} \\
&= 0.
\end{aligned}$$

This proves that g_Σ is well defined for all $\Sigma \in \text{Sym}^+(n)$ and $O(n)$ -invariant by construction. The positivity condition ensures that g is a metric.

Finally, it is clear that α, β, γ have at least the same regularity as the metric g since they are coordinates of the map $D \in \text{Diag}^+(n) \mapsto g_D$. By Theorem 2.18, if α, β, γ are continuous, then the metric g is continuous. \square

The smoothness seems to be more complicated to study. We suspect additional conditions of compatibility on the derivatives of the smooth maps α, β, γ at the singular set of SPD matrices with repeated eigenvalues in order to make the metric g is smooth.

4.5.3 Reinterpretation of kernel metrics

Theorem 4.20 allows to reinterpret kernel metrics. The curiosity of this theorem is the function γ because we have no information on it as soon as the d_i 's are distinct. If α, β, γ do not depend on their $n - 2$ last arguments, i.e. if they are *bivariate*, then γ does not depend on its second argument and $\gamma(d_1)$ must be equal to $\alpha(d_1, d_1) + \beta(d_1, d_1)$. Hence $g_\Sigma(X, X) = \sum_{i,j} \alpha(d_i, d_j) X_{ij}'^2 + \sum_{i,j} \beta(d_i, d_j) X_{ii}' X_{jj}'$, with $\alpha > 0$ and $\alpha + n\beta > 0$, which is much more tractable. Moreover, if $\beta = 0$, then the quadratic form has a *diagonal* expression (sum of squares $X_{ij}'^2$, no mixed terms $X_{ii}' X_{jj}'$) in the basis of matrices induced by the *orthogonal* matrix $P \in O(n)$ in the eigenvalue decomposition of Σ . In this case, we say that the metric is *ortho-diagonal*.

To sum up, the subclass of kernel metrics has two fundamental properties: it is bivariate ($\alpha = \gamma - \beta = 1/\phi$) and ortho-diagonal ($\beta = 0$). This is the reason why we propose to designate kernel (resp. mean kernel) metrics as Bivariate Ortho-Diagonal or BOD metrics (resp. Mean Ortho-Diagonal or MOD metrics), as summarized in Table 4.10. We say that the metric is Bivariate Ortho-ST when it is the extension by Definition 4.17 of a Bivariate Ortho-Diagonal metric with the Scaling and Trace factors $\alpha > 0$ and $\beta > -\alpha/n$. Hence, the extended (mean) kernel metrics can also be designated as BOST (and MOST) metrics.

4.5.4 Key results on $O(n)$ -invariant metrics

In Section 4.4, we gave four key results on BOD/MOD metrics in Propositions 4.15 and 4.16, and four key results on BOST/MOST metrics in Proposition 4.18. Here we give the

Previous description	New designation
Kernel metric	BOD metric
Mean kernel metric	MOD metric
Extended kernel metric	BOST metric
Extended mean kernel metric	MOST metric

Table 4.10: Name correspondences for kernel metrics and sub/super-classes.

counterpart of these propositions for $O(n)$ -invariant metrics.

Proposition 4.22 (Key results on $O(n)$ -invariant metrics)

- (Generality) The class of $O(n)$ -invariant metrics obviously contains the classes of BOD, MOD, BOST, MOST metrics, hence it contains all the metrics in Section 4.3.
- (Stability) The class of $O(n)$ -invariant metrics is obviously stable by $O(n)$ -equivariant diffeomorphisms of $\text{Sym}^+(n)$. Hence it is stable by univariate diffeomorphisms $f : \text{Sym}^+(n) \rightarrow \text{Sym}^+(n)$ and in this case, the pullback metric $f^*g^{\alpha,\beta,\gamma}$ is characterized by the three maps:

$$(a) \quad \alpha_f : d \in (\mathbb{R}^+)^n \mapsto \frac{\alpha(f(d))}{f^{[1]}(d_1, d_2)^2},$$

$$(b) \quad \beta_f : d \in (\mathbb{R}^+)^n \mapsto \frac{\beta(f(d))}{f^{[1]}(d_1, d_2)^2},$$

$$(c) \quad \gamma_f : d \in (\mathbb{R}^+)^n \mapsto \frac{\gamma(f(d))}{f'(d_1)^2}.$$

- (Completeness) Let $g = g^{\alpha,\beta,\gamma}$ be an $O(n)$ -invariant metric. We assume that α, β, γ satisfy a homogeneity property which is similar to the one assumed for mean kernel metrics: there exists $\theta \in \mathbb{R}$ such that for $f \in \{\alpha, \beta, \gamma\}$, $x \in (\mathbb{R}^+)^n$ and $\lambda > 0$, we have $f(\lambda x) = \lambda^{-\theta} f(x)$. If the metric g is geodesically complete, then $\theta = 2$.
- (Cometric) The class of $O(n)$ -invariant metrics is obviously cometric-stable. The cometric is characterized by $\alpha^* = 1/\alpha$ and $S^* = S^{-1}$ where $S(d) \in \text{Sym}^+(n)$ is defined by $S_{ij}(d) = \beta(d_i, d_j, d_{k \neq i, j})$ and $S_{ii}(d) = \gamma(d_i, d_{k \neq i})$ for all $d \in \mathbb{R}^+$ and $i \neq j$.

We omit the proof since it consists in elementary verifications for all but the third statement, whose proof is analogous to the one given in [Hiai and Petz, 2009].

About completeness, the result is much weaker for general $O(n)$ -invariant metrics. Indeed, we lost the converse implication: “if $\theta = 2$, then the metric is geodesically complete”. According to the proof of [Hiai and Petz, 2009], the key element to prove this converse implication is exactly the bivariate, plus the fact that a symmetric homogeneous mean satisfies $m(x, x) = x$. It is worth noticing that $\theta = 2$ is still necessary though.

About the cometric, we lost the closed-form expression we had for BOD and BOST metrics. Computing the cometric is numerically quite heavy in general because it is equivalent to invert the matrix $S(d)$ for all $d \in (\mathbb{R}^+)^n$. However, note that when $\beta = 0$, the cometric is obviously given by the triple $(1/\alpha, 0, 1/\gamma)$. These ortho-diagonal metrics can be seen as the multivariate generalization of BOD metrics. In the next section, we give a cometric-stable extension of the class of BOST metrics for which the cometric can be computed in closed form: the class of bivariate separable metrics.

4.5.5 Bivariate separable metrics

We argued in Section 4.5.3 that *bivariate* metrics are of the form $g_\Sigma(X, X) = \sum_{i,j} \alpha(d_i, d_j) X_{ij}^{\prime 2} + \sum_{i,j} \beta(d_i, d_j) X'_{ii} X'_{jj}$ with $\alpha > 0$ and $\alpha + n\beta > 0$. Then, the first term corresponds to a BOD metric and it can be rewritten $\text{tr}(\Psi_\Sigma(X)^2)$, but it is still difficult to write the second term in a more compact way. If the function β is *separable*, i.e. if β can be written $\beta(x, y) = \psi^{(1)}(x)\psi^{(2)}(y)$, then the second term is simply $\text{tr}(\Psi_\Sigma^{(1)}(X))\text{tr}(\Psi_\Sigma^{(2)}(X))$. Indeed, we can define $\Psi_D^{(k)}(X) = \text{Diag}(\psi^{(k)}(d_i)X_{ii})$ and extend it into $\Psi_\Sigma^{(k)}$ as explained in Section 4.4.2.3. In particular, BOST metrics correspond to the case when $\beta(x, y) = \lambda\sqrt{\alpha(x, x)\alpha(y, y)}$ with $1 + n\lambda > 0$. The wider class of *bivariate separable* metrics is actually cometric-stable and the cometric can be computed quite easily. This is stated in Proposition 4.23.

Proposition 4.23 (Cometric of bivariate separable metrics) Let $\psi : (\mathbb{R}^+)^2 \rightarrow \mathbb{R}^+$ be a symmetric map and let $\psi^{(1)}, \psi^{(2)} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be two maps on positive real numbers. As explained above, we define their extensions $\Psi, \Psi^{(1)}, \Psi^{(2)} : \text{Sym}^+(n) \times \text{Sym}(n) \rightarrow \text{Sym}(n)$. The quadratic form defined by $g_\Sigma(X, X) = \text{tr}(\Psi_\Sigma(X)^2) + \text{tr}(\Psi_\Sigma^{(1)}(X))\text{tr}(\Psi_\Sigma^{(2)}(X))$ automatically satisfies the symmetry and compatibility conditions of Theorem 4.20. Then g is positive definite if and only if the vectors $x = x(d) = \left(\frac{\psi^{(1)}(d_i)}{\psi(d_i, d_i)}\right)_{1 \leq i \leq n}$ and $y = y(d) = \left(\frac{\psi^{(2)}(d_i)}{\psi(d_i, d_i)}\right)_{1 \leq i \leq n}$ satisfy the inequality $\|x\|\|y\| - \langle x|y \rangle < 2$ for all $d \in (\mathbb{R}^+)^n$.

In this case, we say that g is a Bivariate Separable metric. As an $O(n)$ -invariant metric, it is characterized by $\alpha(d) = \psi(d_1, d_2)^2$ and the matrix $S = S(d) = \Delta(I_n + \frac{1}{2}(xy^\top + yx^\top))\Delta$ with $\Delta = \text{Diag}(\psi(d_i, d_i))$. This class of metrics is cometric-stable. If $x = 0$ or $y = 0$, the cometric at Σ is simply characterized by $S^{-1} = \Delta^{-2}$. Otherwise, the cometric is given by:

$$S^{-1} = \Delta^{-1} \left[I_n - \frac{1}{4c}(2 + \langle x|y \rangle)(xy^\top + yx^\top) + \frac{1}{4c}(\|y\|^2 xx^\top + \|x\|^2 yy^\top) \right] \Delta^{-1}, \quad (4.18)$$

with $c = 1 + \langle x|y \rangle - \frac{1}{4}(\|x\|^2\|y\|^2 - \langle x|y \rangle^2) > 0$.

Proof of Proposition 4.23. To determine when g is a metric, we express the functions α, β, γ, S of Theorem 4.20 in function of $\psi, \psi^{(1)}, \psi^{(2)}$:

1. $\alpha(d_1, \dots, d_n) = \psi(d_1, d_2)^2 > 0$,
2. $\beta(d_1, \dots, d_n) = \frac{1}{2}(\psi^{(1)}(d_1)\psi^{(2)}(d_2) + \psi^{(1)}(d_2)\psi^{(2)}(d_1))$,
3. $\gamma(d_1, \dots, d_n) = \psi(d_1, d_1)^2 + \psi^{(1)}(d_1)\psi^{(2)}(d_1)$,
4. hence $S_{ij}(d) = \Delta_{ij}^2 + \frac{1}{2}(\psi^{(1)}(d_i)\psi^{(2)}(d_j) + \psi^{(2)}(d_i)\psi^{(1)}(d_j))$, so we have $S = \Delta(I_n + \frac{1}{2}(xy^\top + yx^\top))\Delta$ with the notations of the proposition.

The symmetry and compatibility conditions of Theorem 4.20 are trivially satisfied. The positivity condition reduces to $S \in \text{Sym}^+(n)$, i.e. $I_n + \frac{1}{2}(xy^\top + yx^\top) \in \text{Sym}^+(n)$. As the eigenvalues of $M = xy^\top + yx^\top$ are 0 (with multiplicity $n-2$) and $\langle x|y \rangle \pm \|x\|\|y\|$, S is positive definite if and only if $2 + \langle x|y \rangle \pm \|x\|\|y\| > 0$. But $\langle x|y \rangle + \|x\|\|y\| \geq 0$ so there is only one condition: $2 > \|x\|\|y\| - \langle x|y \rangle (\geq 0)$, as announced.

Now, we want to compute S^{-1} . If $x = 0$ or $y = 0$, the result is obvious so we assume that $x, y \neq 0$. As M is of rank 2 at most, there exists a polynomial P of degree 3 at most such that

$P(I_n + \frac{1}{2}M) = 0$. Let us find such a polynomial to compute S^{-1} . Since $M^2 = \langle x|y \rangle M + N$ with $N = \|y\|^2 x x^\top + \|x\|^2 y y^\top$ and $NM = \|x\|^2 \|y\|^2 M + \langle x|y \rangle N$, we have:

$$\begin{aligned} \left(I_n + \frac{1}{2}M\right)^2 &= I_n + \left(1 + \frac{\langle x|y \rangle}{4}\right) M + \frac{1}{4}N, \\ \left(I_n + \frac{1}{2}M\right)^3 &= \left(I_n + \frac{1}{2}M\right)^2 + \frac{1}{2} \left(I_n + \frac{1}{2}M\right)^2 M \\ &= \left(I_n + \frac{1}{2}M\right)^2 + \frac{1}{2}M + \frac{1}{2} \left(1 + \frac{\langle x|y \rangle}{4}\right) M^2 + \frac{1}{8}NM \\ &= \left(I_n + \frac{1}{2}M\right)^2 + \frac{4 + 4\langle x|y \rangle + \langle x|y \rangle^2 + \|x\|^2 \|y\|^2}{8} M + \frac{1}{4}(2 + \langle x|y \rangle)N \\ &= a \left(I_n + \frac{1}{2}M\right)^2 + \frac{b}{2}M - (2 + \langle x|y \rangle)I_n \\ &= a \left(I_n + \frac{1}{2}M\right)^2 + b \left(I_n + \frac{1}{2}M\right) + c I_n, \end{aligned}$$

with $\begin{cases} a = 3 + \langle x|y \rangle \\ b = \frac{-12 - 8\langle x|y \rangle - \langle x|y \rangle^2 + \|x\|^2 \|y\|^2}{4} \\ c = 1 + \langle x|y \rangle + \frac{\langle x|y \rangle^2 - \|x\|^2 \|y\|^2}{4} = 1 - a - b > 0 \end{cases}$. Indeed, $c > 1 + \langle x|y \rangle - \frac{1}{2}(\|x\| \|y\| + \langle x|y \rangle) = 1 - \frac{1}{2}(\|x\| \|y\| - \langle x|y \rangle) > 0$. Hence, denoting $S_0 := I_n + \frac{1}{2}M$, we have $S_0^{-1} = \frac{1}{c}(S_0^2 - a S_0 - b I_n) = I_n + \frac{1}{4c}(N - (2 + \langle x|y \rangle)M)$ and $S^{-1} = \Delta^{-1} \left(I_n + \frac{1}{4c}(N - (2 + \langle x|y \rangle)M)\right) \Delta^{-1}$ which is exactly Equation (4.18).

Finally, we want to prove that the cometric is bivariate separable. Regarding Equation (4.18), we look for $x' = \frac{Ax+By}{4c}$ and $y' = Cx + Dy$ for $A, B, C, D \in \mathbb{R}$ such that:

$$x' y'^\top + y' x'^\top = -\frac{1}{2c}(2 + \langle x|y \rangle)(x y^\top + y x^\top) + \frac{1}{2c}(\|y\|^2 x x^\top + \|x\|^2 y y^\top) \quad (4.19)$$

It is satisfied if $AC = \|y\|^2$, $BD = \|x\|^2$ and $AD + BC = -2(2 + \langle x|y \rangle)$, or equivalently $(AX + B)(CX + D) = \|y\|^2 X^2 - 2(2 + \langle x|y \rangle)X + \|x\|^2$. This is a second-order polynomial with roots $\lambda = \frac{2 + \langle x|y \rangle + \sqrt{\delta}}{\|y\|^2}$ and $\mu = \frac{2 + \langle x|y \rangle - \sqrt{\delta}}{\|y\|^2}$ where $\delta = (2 + \langle x|y \rangle + \|x\| \|y\|)(2 + \langle x|y \rangle - \|x\| \|y\|) > 0$ is the discriminant. Hence, it suffices to define $A = \|y\|$, $B = -\lambda \|y\|$, $C = \|y\|$ and $D = -\mu \|y\|$, so that $S^{-1} = \Delta^{-1} \left(I_n + \frac{1}{2}(x' y'^\top + y' x'^\top)\right) \Delta^{-1}$. Hence, the cometric is bivariate separable and this class of metrics is cometric-stable. \square

4.6 Conclusion

To encompass all the $O(n)$ -invariant metrics summarized in Section 4.3, including the ones with a trace term ($\beta \neq 0$), we defined the class of extended kernel metrics. This class satisfies the key results of stability and completeness we selected from [Hiai and Petz, 2009] plus the cometric-stability with cometric in closed form, which is important to compute geodesics numerically via the Hamiltonian formulation. Then, from the characterization of $O(n)$ -invariant metrics in terms of three continuous maps $\alpha, \beta, \gamma : (\mathbb{R}^+)^n \rightarrow \mathbb{R}^+$ satisfying properties of symmetry, compatibility and positivity, we were able to characterize kernel

metrics as Bivariate Ortho-Diagonal (BOD) metrics. Among the key results on mean kernel metrics, the sufficient condition of completeness and the closed-form expression of the cometric disappear for general $O(n)$ -invariant metrics. We finally defined the intermediate class of bivariate separable metrics which is cometric-stable and for which the cometric has a simple expression.

Since kernel metrics encompass very different metrics regarding curvature and completeness, it would be nice to introduce some more requirements on metrics to perform the opposite work of defining principled sub-classes of (mean) kernel metrics. In the next chapter, we propose some principled subfamilies of kernel metrics. It would also be interesting to rely on the cometric-stability of kernel metrics or super-classes to effectively compute the geodesics numerically and to investigate their properties regarding statistical analyses.

Another interesting direction would be to consider other properties of kernel metrics that were described in [Hiai and Petz, 2009], namely monotonicity and comparison properties. It would be challenging to understand how they could be generalized to BOST metrics or even to $O(n)$ -invariant metrics. Furthermore, to our knowledge there is no trace of families of non $O(n)$ -invariant metrics in the literature. However, there exist some situations where the $O(n)$ -invariance is not relevant, for example on correlation matrices because the space is not stable under this group action. This is investigated in Part IV.

Chapter 5

Geometry of Mixed-Euclidean metrics

Abstract

Several Riemannian metrics and families of Riemannian metrics were defined on the manifold of Symmetric Positive Definite (SPD) matrices. Firstly, we formalize a common general process to define families of metrics: the principle of deformed metrics. We relate the recently introduced family of alpha-Procrustes metrics to the general class of mean kernel metrics by providing a sufficient condition under which elements of the former belongs to the latter. Secondly, we focus on the principle of balanced bilinear forms that we recently introduced. We give a new sufficient condition under which the balanced bilinear form is a metric. It allows us to introduce the Mixed-Euclidean (ME) metrics which generalize the Mixed-Power-Euclidean (MPE) metrics. We unveil their link with the (u, v) -divergences and the (α, β) -divergences of information geometry and we provide an explicit formula of the Riemann curvature tensor. We show that the sectional curvature of all ME metrics can take negative values and we show experimentally that the sectional curvature of all MPE metrics but the log-Euclidean, power-Euclidean and power-affine metrics can take positive values.

This chapter was published in the journal Differential Geometry and its Applications in April 2022 under the title “The geometry of mixed-Euclidean metrics on symmetric positive definite matrices” [Thanwerdas and Pennec, 2022a]. All the proofs, originally in appendix of the paper, are deferred to Section 11.4.

5.1 Introduction

The convex cone of Symmetric Positive Definite (SPD) matrices is a manifold on which several Riemannian metrics were defined: Euclidean, Fisher-Rao/affine-invariant [Skovgaard, 1984, Amari and Nagaoka, 2000, Moakher, 2005, Pennec et al., 2006, Lenglet et al., 2006, Fletcher and Joshi, 2007], log-Euclidean [Arsigny et al., 2006], Bures-Wasserstein [Dowson and Landau, 1982, Olkin and Pukelsheim, 1982, Dryden et al., 2009, Takatsu, 2010, Takatsu, 2011, Bhatia et al., 2019], Bogoliubov-Kubo-Mori [Petz and Toth, 1993, Michor et al., 2000], log-Cholesky [Lin, 2019], etc. Several families of metrics encompassing them were defined to understand their common properties, their differences and the level of generality of each property: kernel metrics and mean kernel metrics [Hiai and Petz, 2009, Hiai and Petz, 2012], power-Euclidean [Dryden et al., 2010], alpha-Procrustes [Hà Quang, 2019], deformed-affine

[Thanwerdas and Pennec, 2019b], mixed-power-Euclidean [Thanwerdas and Pennec, 2019a], extended kernel metrics, bivariate separable metrics [Thanwerdas and Pennec, 2022b], etc. In particular, kernel metrics form a very general family of $O(n)$ -invariant metrics indexed by kernel maps $\phi : (0, \infty)^2 \rightarrow (0, \infty)$ acting on the eigenvalues of SPD matrices. This family contains many $O(n)$ -invariant metrics and it has good stability properties. The subclass of mean kernel metrics, for which the kernel maps have monotonicity properties, is interesting because it provides a necessary and sufficient condition for geodesic completeness. Hence, kernel metrics and mean kernel metrics appear as sufficiently general families with interesting properties so it is a natural framework to work in. However, this class contains metrics with very different geometries so it motivates us to define subfamilies of metrics which share more geometric properties with one another.

In previous works, we introduced two principles for building families of Riemannian metrics that share interesting properties: the principle of deformed metrics [Thanwerdas and Pennec, 2019b] and the principle of balanced bilinear forms [Thanwerdas and Pennec, 2019a]. Deforming metrics (or datasets of SPD matrices) via a diffeomorphism is a very common procedure to define families of metrics. In particular, kernel metrics are stable by univariate diffeomorphisms, those which are characterized by their action on eigenvalues. However, mean kernel metrics are not stable by all univariate diffeomorphisms because of the monotonicity requirement. In this work, we gather many constructions of deformed metrics and we contribute a sufficient condition under which alpha-Procrustes metrics are mean kernel metrics.

The balanced bilinear form of two flat metrics is defined by composing the Frobenius inner product with the parallel transport of each flat metric [Thanwerdas and Pennec, 2019a]. When the bilinear form is a metric, it forms a dually-flat manifold along with the two flat Levi-Civita connections of the flat metrics. In the case where the two flat metrics are power-Euclidean metrics, the balanced bilinear form is a metric called the mixed-power-Euclidean metric. In this work, we give a new sufficient condition for a balanced bilinear form to be a metric, namely that the flat metrics are univariately-deformed-Euclidean metrics, which allows to define the new family of Mixed-Euclidean metrics. Then, we provide the geometric operations of Mixed-Euclidean metrics regarding information geometry and Riemannian geometry. In particular, our main contributions are on the one hand the link we establish between Mixed-Euclidean/Mixed-Power-Euclidean metrics and the $(u, v)/(\alpha, \beta)$ -divergences of information geometry, and on the other hand the expression of the Riemann curvature tensor of Mixed-Euclidean metrics.

In Section 5.2, we present our notations and the preliminary concepts of univariate maps and kernel metrics. In Section 5.3, we study deformed metrics and we relate the family of alpha-Procrustes metrics to the class of mean kernel metrics. In Section 5.4, we recall the main concepts of information geometry, we state the principle of balanced bilinear forms and we explain the relation between the two. In Section 5.5, we introduce the new family of Mixed-Euclidean metrics and we study its geometry. We conclude and discuss some perspectives in Section 5.6. The proofs of the results are presented in appendix.

5.2 Notations and preliminary concepts

Section 5.2 partly summarizes Chapter 4 to allow Chapter 5 to be read independently. The skilled reader may, after a glance at the definition of the second divided difference (Definition 5.2), skip this part to resume at Section 5.3.

In this section, we introduce some notations and we recall two concepts that are used throughout the chapter. The first one is the concept of univariate map on SPD matrices: it is a map acting on the eigenvalues, such as the symmetric matrix logarithm or the power maps. The successive differentials of smooth univariate maps can be expressed in closed form modulo eigenvalue decomposition thanks to the functions called divided differences [Bhatia, 1997]. This main advantage explains why they are ubiquitous as indexing collections of families of metrics. Secondly, we recall the main facts about classes of kernel and mean kernel metrics introduced in [Hiai and Petz, 2009, Hiai and Petz, 2012].

5.2.1 Notations

We denote $\text{Sym}(n)$ the vector space of real symmetric matrices of size n , $\text{Sym}^+(n)$ the manifold of SPD matrices, $\text{O}(n)$ the orthogonal group, $\text{Diag}^+(n)$ the group of positive diagonal matrices.

On the manifold $\text{Sym}^+(n)$, we denote $T_\Sigma \text{Sym}^+(n)$ the tangent space at $\Sigma \in \text{Sym}^+(n)$. Given a metric g^I on the manifold $\text{Sym}^+(n)$ where I is any index characterizing the metric, we denote ∇^I its Levi-Civita connection, R^I the Riemann curvature tensor, T^I the torsion tensor, Π^I the parallel transport. We omit the index when the context is clear.

Given a matrix M , we denote M_{ij} or $[M]_{ij}$ the (i, j) -th coefficient of M . Given coefficients $(M_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n^2}$, we denote $[M_{ij}]_{i, j}$ the matrix with (i, j) -th entry M_{ij} . Given $(d_1, \dots, d_n) \in \mathbb{R}^n$, we denote $\text{diag}(d_1, \dots, d_n)$ the corresponding diagonal matrix.

We recall that $\exp : \Sigma \in \text{Sym}(n) \mapsto \sum_{k=0}^{+\infty} \frac{1}{k!} \Sigma^k \in \text{Sym}^+(n)$ is a diffeomorphism whose inverse is the symmetric matrix logarithm denoted $\log : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$.

5.2.2 Univariate maps

We call $\text{O}(n)$ -equivariant map a map $f : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$ such that $f(R\Sigma R^\top) = Rf(\Sigma)R^\top$ for all $\Sigma \in \text{Sym}^+(n)$ and $R \in \text{O}(n)$. Among $\text{O}(n)$ -equivariant maps, we focus on the class of univariate maps.

Definition 5.1 (Univariate maps) A univariate map is an $\text{O}(n)$ -equivariant map $f : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$ such that there exists a map on positive real numbers also denoted $f : (0, \infty) \rightarrow \mathbb{R}$ such that $f(PDP^\top) = P \text{Diag}(f(d_1), \dots, f(d_n)) P^\top$ for all $P \in \text{O}(n)$ and $D \in \text{Diag}^+(n)$ with $D = \text{Diag}(d_1, \dots, d_n)$.

Any $f : (0, \infty) \rightarrow \mathbb{R}$ can be extended into a univariate map and if the former is of class \mathcal{C}^1 (resp. \mathcal{C}^2 , resp. a \mathcal{C}^1 -diffeomorphism), then the latter is differentiable (resp. two times differentiable, resp. a diffeomorphism) [Bhatia, 1997, Thanwerdas and Pennec, 2022b]. We denote \mathcal{Univ} the set of smooth univariate diffeomorphisms. In addition, the differential and the Hessian of a smooth univariate map can be expressed thanks to the first and second divided differences as follows.

Definition 5.2 (Divided differences) [Bhatia, 1997]

1. Let $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$. The first divided difference of f is the continuous symmetric map $f^{[1]} : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined for $x, y \in \mathbb{R}$ by:

$$f^{[1]}(x, y) = \begin{cases} \frac{f(x)-f(y)}{x-y} & \text{if } x \neq y \\ f'(x) & \text{if } x = y \end{cases}. \quad (5.1)$$

2. Let $f \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$. The second divided difference of f is the continuous symmetric map $f^{[2]} : \mathbb{R}^3 \rightarrow \mathbb{R}$ defined for $x, y, z \in \mathbb{R}$ by:

$$f^{[2]}(x, y, z) = \begin{cases} (f^{[1]}(x, \cdot))^{[1]}(y, z) = \frac{f^{[1]}(x, z) - f^{[1]}(x, y)}{z - y} & \text{if } y \neq z \\ (f^{[1]}(y, \cdot))^{[1]}(z, x) = \frac{f^{[1]}(y, x) - f^{[1]}(y, z)}{x - z} & \text{if } z \neq x \\ (f^{[1]}(z, \cdot))^{[1]}(x, y) = \frac{f^{[1]}(z, y) - f^{[1]}(z, x)}{y - x} & \text{if } x \neq y \\ \frac{1}{2}f''(x) & \text{if } x = y = z \end{cases}. \quad (5.2)$$

If $f \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$, then one can check that the differential of $f^{[1]}$ at $(x, y) \in \mathbb{R}^2$ is:

$$d_{(x,y)}f^{[1]}(h, k) = \begin{cases} \frac{f'(x)h - f'(y)k}{x-y} + \frac{f(x)-f(y)}{(x-y)^2}(h-k) & \text{if } x \neq y \\ \frac{f''(x)}{2}(h+k) & \text{if } x = y \end{cases}, \quad (5.3)$$

so one can prove that $df^{[1]}$ is continuous and $f^{[1]} \in \mathcal{C}^1(\mathbb{R}^2, \mathbb{R})$. This also proves that $\frac{\partial f^{[1]}}{\partial x}(x, x) = \frac{f''(x)}{2}$ and that $f^{[2]}$ is continuous.

From now on, all univariate maps are assumed to be smooth.

Lemma 5.3 (Differential and Hessian of a univariate map) [Bhatia, 1997] The differential and the Hessian of a univariate map f are $O(n)$ -equivariant: $d_{PDP^\top}f(PXP^\top) = P d_D f(X) P^\top$ and $H_{PDP^\top}f(PXP^\top, PYP^\top) = P H_D f(X, Y) P^\top$. Hence, they are determined by their values at diagonal matrices $D \in \text{Diag}^+(n)$, which are given by the following formulae:

$$[d_D f(X)]_{ij} = f^{[1]}(d_i, d_j) X_{ij}, \quad (5.4)$$

$$[H_D f(X, X)]_{ij} = 2 \sum_{k=1}^n f^{[2]}(d_i, d_j, d_k) X_{ik} X_{jk}. \quad (5.5)$$

5.2.3 Classes of $O(n)$ -invariant metrics

The class of kernel metrics is a subclass of $O(n)$ -invariant metrics on SPD matrices indexed by smooth bivariate symmetric maps $\phi : (0, \infty)^2 \rightarrow (0, \infty)$ [Hiai and Petz, 2009]. The advantages of this class are the simple formulation of its elements, some important results on the metrics (completeness, cometric) and some important stability properties of the class. Hence, it is a good ambient class to define subfamilies of metrics. Therefore in this section, we recall the definition of kernel metrics, the refinement of mean kernel metrics and the results on completeness, cometric and stability.

5.2.3.1 Kernel metrics

Definition 5.4 (Kernel metric) A *kernel metric* [Hiai and Petz, 2009] is an $O(n)$ -invariant metric for which there is a smooth bivariate map $\phi : (0, \infty)^2 \rightarrow (0, \infty)$ such that $g_\Sigma(X, X) = g_D(X', X') = \sum_{i,j} \frac{1}{\phi(d_i, d_j)} X'_{ij}{}^2$, where $X = PX'P^\top$, $\Sigma = PDP^\top$ with $P \in O(n)$ and $D = \text{Diag}(d_1, \dots, d_n)$.

Important examples of kernel metrics are the Euclidean, the log-Euclidean, the affine-invariant, the Bures-Wasserstein and the Bogoliubov-Kubo-Mori metrics:

$$\text{(Euclidean)} \quad g_\Sigma^E(X, X) = \text{tr}(X^2), \quad (5.6)$$

$$\text{(Log-Euclidean)} \quad g_\Sigma^{\text{LE}}(X, X) = \text{tr}(d_\Sigma \log(X)^2), \quad (5.7)$$

$$\text{(Affine-invariant)} \quad g_\Sigma^A(X, X) = \text{tr}((\Sigma^{-1}X)^2), \quad (5.8)$$

$$\text{(Bures-Wasserstein)} \quad g_\Sigma^{\text{BW}}(X, X) = \text{tr}(\Sigma \mathcal{S}_\Sigma(X)^2), \quad (5.9)$$

$$\text{(Bogoliubov-Kubo-Mori)} \quad g_\Sigma^{\text{BKM}}(X, X) = \text{tr}(d_\Sigma \log(X)X), \quad (5.10)$$

where $\mathcal{S}_\Sigma(X)$ denotes the solution of the Sylvester equation $X = \Sigma \mathcal{S}_\Sigma(X) + \mathcal{S}_\Sigma(X) \Sigma$. A review of the definitions, geometric properties and main references on these five metrics can be found in [Thanwerdas and Pennec, 2022b].

5.2.3.2 The refinement of mean kernel metrics

There is a refinement of kernel metrics where the bivariate function ϕ relies on a function called a symmetric homogeneous mean [Hiai and Petz, 2009]. These subclasses provide a nice necessary and sufficient condition for geodesic completeness.

Definition 5.5 (Mean kernel metrics) [Hiai and Petz, 2009] A *mean kernel metric* is a kernel metric characterized by a bivariate map ϕ of the form $\phi(x, y) = a m(x, y)^\theta$ where $a > 0$ is a positive coefficient, $\theta \in \mathbb{R}$ is a homogeneity power and $m : (0, \infty)^2 \rightarrow (0, \infty)$ is a symmetric homogeneous mean, that is:

1. symmetric, i.e. $m(x, y) = m(y, x)$ for all $x, y > 0$,
2. homogeneous, i.e. $m(cx, cy) = c m(x, y)$ for all $c, x, y > 0$,
3. non-decreasing in both variables,
4. $\min(x, y) \leq m(x, y) \leq \max(x, y)$ for all $x, y > 0$. It implies $m(x, x) = x$.

Theorem 5.6 (Completeness of mean kernel metrics) [Hiai and Petz, 2009] Mean kernel metrics are geodesically complete if and only if $\theta = 2$.

5.2.3.3 Main results

The five kernel metrics cited above are mean kernel metrics. The mean functions are summarized in Table 5.1.

Moreover, the class of kernel metrics is stable under pullback by univariate diffeomorphisms [Hiai and Petz, 2009]. However, the class of *mean* kernel metrics is not stable under

Metric	Kernel $\phi(x, y)$	Mean m	Power θ
Euclidean	1	Any mean	0
Log-Euclidean	$(\frac{x-y}{\log(x)-\log(y)})^2$	Logarithmic mean	2
Affine-invariant	xy	Geometric mean	2
Bures-Wasserstein	$4 \frac{x+y}{2}$	Arithmetic mean	1
Bogoliubov-Kubo-Mori	$\frac{x-y}{\log(x)-\log(y)}$	Logarithmic mean	1

Table 5.1: Bivariate functions of the main $O(n)$ -invariant metrics on SPD matrices.

univariate diffeomorphisms, essentially because of the third condition of a symmetric homogeneous mean. Indeed, the mean has to be non-decreasing in both variables which is neither a differential nor a Riemannian property.

In addition, the class of kernel metrics is cometric stable [Thanwerdas and Pennec, 2022b]. Indeed, the cometric is a metric on the cotangent bundle $T^*\text{Sym}^+(n) \simeq \text{Sym}^+(n) \times \text{Sym}(n)^*$. Thanks to the Riesz theorem, the Frobenius inner product provides the identification $\text{Sym}(n)^* \simeq \text{Sym}(n)$ so the cometric can be considered as a metric. The cometric of the kernel metric characterized by ϕ is the kernel metric characterized by $\frac{1}{\phi}$.

5.3 Deformed metrics

Log-Euclidean metrics on SPD matrices are pullback metrics of Euclidean metrics on the vector space of symmetric matrices via the symmetric matrix logarithm $\log : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$. This geometric construction of a metric on SPD matrices based on a diffeomorphism f is commonly used to define families of metrics on SPD matrices indexed by automorphisms of SPD matrices. Indeed, even if these metrics are isometric, they do not give the same results in data analyses. It is actually equivalent to compute with the metric g on the transformed dataset $[f(\Sigma_1), \dots, f(\Sigma_N)]$ or to compute with the pullback metric f^*g on the initial dataset $[\Sigma_1, \dots, \Sigma_N]$.

In this section, we give examples of situations in the literature where such transformations are applied to the data (Section 5.3.1), then we unify them into our principle of deformed metrics and we give the fundamental Riemannian operations (distance, geodesics, curvature, parallel transport) of the deformed metrics (Section 5.3.2). In Section 5.3.3, we contribute the new family of deformed-Wasserstein metrics based on this principle which comprises the family of alpha-Procrustes metrics [Hà Quang, 2019]. We also give a sufficient condition under which alpha-Procrustes metrics are mean kernel metrics.

5.3.1 Use of a deformation in the literature

As mentioned before, the class of kernel metrics is stable by pullback under univariate diffeomorphisms [Hiai and Petz, 2009, Thanwerdas and Pennec, 2022b]. In particular, pullbacks of the Euclidean metric and the affine-invariant metric under power diffeomorphisms are detailed in the original paper on kernel metrics [Hiai and Petz, 2009]. They were later called power-Euclidean [Dryden et al., 2010] and power-affine metrics, or more generally deformed-Euclidean and deformed-affine metrics for an arbitrary diffeomorphism [Thanwerdas and

Penec, 2019b]. Moreover, power-Euclidean metrics are *mean* kernel metrics for any power and power-affine metrics are *mean* kernel metrics if and only if the power belongs to $[-2, 2]$ [Hiai and Petz, 2009]. Since power-Euclidean metrics interpolate between the log-Euclidean, the Wigner-Yanase/square-root and the Euclidean metrics, an optimization procedure was proposed on the parameter to choose the most appropriate metric on a dataset of covariance matrices for Diffusion Tensor Imaging (DTI) [Dryden et al., 2010]. It is common in DTI to compute with precision matrices which are the inverses of covariance matrices, $\text{inv}(\Sigma) = \Sigma^{-1}$ [Lenglet et al., 2004], or with other transformations of the covariance matrices such as the adjugate function $\text{adj}(\Sigma) = \det(\Sigma)\Sigma^{-1}$ [Fuster et al., 2016]. More recently, the family of alpha-Procrustes metrics was introduced by pullback under power diffeomorphisms of the Bures-Wasserstein metric, as for power-Euclidean and power-affine metrics, and it was extended to the infinite dimension in the context of Reproducing Kernel Hilbert Spaces (RKHS) [Hà Quang, 2019].

In papers where the power diffeomorphisms are used to define power-Euclidean, power-affine and alpha-Procrustes metrics [Hiai and Petz, 2009, Dryden et al., 2010, Hiai and Petz, 2012, Thanwerdas and Penec, 2019b, Hà Quang, 2019], it is often noticed that the limit when the power tends to 0 is the log-Euclidean metric. This is actually a general fact. Indeed, from a Riemannian metric g , it is possible to construct a one-parameter family of metrics $(g^{(p)})_{p \in \mathbb{R}^*}$ by taking the pullback by the power diffeomorphism $\text{pow}_p : \Sigma \in \text{Sym}^+(n) \mapsto \Sigma^p \in \text{Sym}^+(n)$ for $p \neq 0$ and to scale it by $\frac{1}{p^2}$, that is $g_\Sigma^{(p)}(X, X) = \frac{1}{p^2} g_{\Sigma^p}(d_\Sigma \text{pow}_p(X), d_\Sigma \text{pow}_p(X))$ for all $\Sigma \in \text{Sym}^+(n)$ and all $X \in T_\Sigma \text{Sym}^+(n)$. Then when p tends to 0, $g^{(p)}$ tends to the log-Euclidean metric associated to the inner product g_{I_n} , that is $g_\Sigma^{(p)}(X, X) \xrightarrow{p \rightarrow 0} g_{I_n}(d_\Sigma \log(X), d_\Sigma \log(X))$ for all $\Sigma \in \text{Sym}^+(n)$ and all $X \in T_\Sigma \text{Sym}^+(n)$.

5.3.2 Principle of deformed metrics

Principle 5.7 (Principle of deformed metrics) Let g be a Riemannian metric on $\text{Sym}^+(n)$ and $f : \text{Sym}^+(n) \rightarrow \text{Sym}^+(n)$ be a diffeomorphism. Then the f -deformed metric is defined as the pullback metric f^*g . It is a Riemannian metric on $\text{Sym}^+(n)$ which is isometric to g and whose expression is:

$$(f^*g)_\Sigma(X, X) = g_{f(\Sigma)}(d_\Sigma f(X), d_\Sigma f(X)). \quad (5.11)$$

All the Riemannian operations of a deformed metric are obtained by pulling back the formulae that are known for the initial metric, as shown in Table 5.2.

A fundamental stability property is that if g is $O(n)$ -invariant and if f is $O(n)$ -equivariant, then the deformed metric f^*g is also $O(n)$ -invariant. Moreover, as mentioned before, if g is a kernel metric and if f is univariate, then the deformed metric f^*g is a kernel metric and the set $\{f^*g, f \in \mathcal{Univ}\}$ forms a family of kernel metrics that is closed under pullback by univariate diffeomorphisms [Hiai and Petz, 2009]. The Riemannian operations that are known in closed form for g are also known in closed form for f^*g .

Metric	$g_{\Sigma}^f(X, X) = g_{f(\Sigma)}(d_{\Sigma}f(X), d_{\Sigma}f(X))$
Distance	$d^f(\Sigma, \Lambda) = d(f(\Sigma), f(\Lambda))$
Levi-Civita	$d_{\Sigma}f(\nabla_{X_{\Sigma}}^f Y) = \nabla_{d_{\Sigma}f(X)}(df(Y))$
Curvature	$d_{\Sigma}f(R_{\Sigma}^f(X, Y)Z) = R_{f(\Sigma)}(d_{\Sigma}f(X), d_{\Sigma}f(Y))d_{\Sigma}f(Z)$
Geodesics	$f(\gamma_{(\Sigma, X)}^f(t)) = \gamma_{(f(\Sigma), d_{\Sigma}f(X))}(t)$
Logarithm	$d_{\Sigma}f(\text{Log}_{\Sigma}^f(\Lambda)) = \text{Log}_{f(\Sigma)}(f(\Lambda))$
Parallel transport	$d_{\Lambda}f(\Pi_{\gamma; \Sigma \rightarrow \Lambda}^f X) = \Pi_{f \circ \gamma; f(\Sigma) \rightarrow f(\Lambda)}(d_{\Sigma}f(X))$

Table 5.2: Riemannian operations of deformed metrics on SPD matrices.

5.3.3 The new family of deformed-Wasserstein metrics

Definition 5.8 (Deformed-Wasserstein metrics) A deformed-Wasserstein metric is the pullback metric by a univariate diffeomorphism of the Bures-Wasserstein metric (Formula (5.9)).

The family of deformed-Wasserstein metrics contains the family of alpha-Procrustes metrics since they are pullbacks of the Bures-Wasserstein metric by the power diffeomorphism $\text{pow}_{2\alpha}$ scaled by $\frac{1}{4\alpha^2}$ [Hà Quang, 2019]. In this work, we designate alpha-Procrustes metrics as power-Wasserstein metrics to be consistent with power-Euclidean and power-affine metrics and to parameterize the family by $p \in \mathbb{R}^*$, the correspondence being $p = 2\alpha$. As argued earlier, we can say that the log-Euclidean metric belongs to deformed-Wasserstein metrics so we can designate it as power-Wasserstein with power $p = 0$.

Since the Bures-Wasserstein metric is a mean kernel metric, it is tempting to determine when a power-Wasserstein metric is a mean kernel metric, in analogy to the work done for the power-Euclidean and the power-affine metrics [Hiai and Petz, 2009]. Here we give a sufficient condition under which a power-Wasserstein metric is a mean kernel metric. The proof is in Section 11.4.

Theorem 5.9 (Sufficient condition for power-Wasserstein to be mean kernel) The power-Wasserstein metric of parameter $p \leq 1$ is a mean kernel metric.

See the proof of Theorem 5.9 in Section 11.4.

This condition does not seem to be sufficient. Indeed, after numerical simulations, we conjecture that there exists $p_0 \in (2.61, 2.611)$ such that the power-Wasserstein metric of parameter p is a mean kernel metric if and only if $p \in (-\infty, 1] \cup [p_0, +\infty)$. Moreover, the proof actually tells that if $p \in (1, 2]$, then it is not a mean kernel metric.

In this section, we gathered the deformations of Riemannian metrics or of SPD datasets under our principle of deformed metrics. Therefore from a metric we can define the family of power deformations of this metric, which tends to a log-Euclidean metric when the power tends to 0. Moreover, the family of univariate deformations of a kernel metric is a stable subfamily of kernel metrics and it is interesting to determine when these metrics are mean kernel metrics. It seems to be a quite difficult problem for general univariate deformations. On the example of power deformations of the Bures-Wasserstein metric, we gave the sufficient condition $p \leq 1$. To the best of our knowledge, determining necessary and sufficient conditions for deformed-Euclidean, deformed-affine and deformed-Wasserstein metrics to be

mean kernel metrics remains an open problem.

5.4 Balanced metrics

The affine-invariant and the Bogoliubov-Kubo-Mori metrics were shown to provide a dually-flat structure, that is a couple of flat affine connections which are dual with respect to the metric. This is a rich geometric structure which provides a so called canonical divergence, potentials and specific algorithms [Amari and Nagaoka, 2000, Banerjee et al., 2005, Nielsen and Nock, 2009]. A dually-flat manifold is a Hessian manifold where the potential is defined globally [Shima and Yagi, 1997, Amari and Armstrong, 2014]. Inspired by the characterization of the duality based on parallel transport, we introduced a preliminary version of the principle of balanced bilinear forms in [Thanwerdas and Pennec, 2019a] which allows to define a bilinear form g^0 on SPD matrices from two flat Riemannian metrics g, g^* by $g_\Sigma^0(X, Y) = \text{tr}((\Pi_{\Sigma \rightarrow I_n} X)(\Pi_{\Sigma \rightarrow I_n}^* Y))$ where Π, Π^* denote the respective parallel transports. The term “balanced” was chosen because the bilinear form relies half on each of the two flat metrics. We showed that if the two flat metrics are power-Euclidean metrics, then the balanced bilinear form is symmetric and positive definite, i.e. a metric. In this section, we give a weaker condition under which the bilinear form is a metric.

To ease the comprehension of this section, we recall the main concepts of information geometry, especially dually-flat manifolds and related notions, in Section 5.4.1. In Section 5.4.2, we provide a new condition on the two flat metrics so that the balanced bilinear form is a metric: it is sufficient to assume that the flat metrics are univariately-deformed-Euclidean metrics.

5.4.1 Information geometry and dually-flat manifolds

Before introducing the specific concepts of information geometry, we recall the definition of an affine map between manifolds equipped with affine connections and the definition of a flat affine connection. We denote ∂ the canonical affine connection on a vector space.

Definition 5.10 (Affine map) Let $\mathcal{M}, \mathcal{M}'$ be two manifolds with respective affine connections ∇, ∇' . We say that $f : \mathcal{M} \rightarrow \mathcal{M}'$ is an affine map if for all vector fields X, Y on \mathcal{M} , we have $\nabla'_{f_*(X)} f_*(Y) = f_*(\nabla_X Y)$.

Definition 5.11 (Flat affine connection) Depending on domains of research and authors, a flat affine connection is a connection such that:

1. (Affine geometry) $R = 0$ and $T = 0$,
2. (Information geometry) [Amari and Nagaoka, 2000, Section 1.7] there exists a global chart $f : (\mathcal{M}, \nabla) \rightarrow (\mathbb{R}^{\dim \mathcal{M}}, \partial)$ which is an affine map, i.e. \mathcal{M} can be seen as an open set of $\mathbb{R}^{\dim \mathcal{M}}$ via f .

In the following, we call 1-flat (resp. 2-flat) a connection that is flat according to the sense 1 (resp. 2) of the previous definition. A 2-flat connection is clearly 1-flat. Conversely, a 1-flat connection is locally 2-flat, i.e. each point has a neighborhood \mathcal{U} such that ∇ is 2-flat

on \mathcal{U} . A 1-flat connection is a priori not globally 2-flat because the manifold need not be an open set of \mathbb{R}^n (e.g. the circle \mathbb{S}^1). The obstruction is topological.

In this work, the flat metrics we introduce on $\text{Sym}^+(n)$ (which is an open set of the vector space of symmetric matrices) are actually 2-flat.

5.4.1.1 Dual connections with respect to a metric

Definition 5.12 (Dual connections) [Amari and Nagaoka, 2000] Let (\mathcal{M}, g) be a Riemannian manifold, ∇^g the Levi-Civita connection of g and ∇, ∇^* be affine connections on \mathcal{M} . We say that ∇^* is the *dual connection* of ∇ with respect to g if one of the following equivalent requirements is satisfied:

1. $\partial_k g_{ij} = g_{lj} \Gamma_{ki}^l + g_{il} (\Gamma^*)_{kj}^l$ for all $i, j, k \in \{1, \dots, \dim \mathcal{M}\}$ in any chart, where Γ_{ij}^k and $(\Gamma^*)_{ij}^k$ are the Christoffel symbols of ∇ and ∇^* ,
2. $Z g(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla_Z^* Y)$ for all vector fields X, Y, Z on \mathcal{M} ,
3. $g(X, Y) = g(\Pi X, \Pi^* Y)$ for all vector fields X, Y on \mathcal{M} .

Hence given a metric g , ∇ uniquely determines ∇^* and $(\nabla^*)^* = \nabla$ so that $\frac{\nabla + \nabla^*}{2}$ is a metric connection. We call (g, ∇, ∇^*) a *dualistic structure*.

Note that if ∇ and ∇^* are torsion-free, then $\frac{\nabla + \nabla^*}{2} = \nabla^g$.

Definition 5.13 (Dually-flat manifold) [Amari and Nagaoka, 2000] We say that $(\mathcal{M}, g, \nabla, \nabla^*)$ is a dually-flat manifold (or a Hessian manifold) when ∇ and ∇^* are dual with respect to the metric g and when ∇ and ∇^* are flat (in the sense 2 of Definition 5.11).

5.4.1.2 Divergence

Definition 5.14 (Divergence) [Amari and Nagaoka, 2000] A *divergence* is a distance-like smooth map $D : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+$ such that:

1. (separation) $D(x, y) = 0$ if and only if $x = y$,
2. (non-degenerate) the symmetric positive semi-definite bilinear form $g^D : z \in \mathcal{M} \mapsto -\partial_x|_{x=z} \partial_y|_{y=z} D$ is positive definite. It is called the *induced Riemannian metric*. We denote $\flat : T\mathcal{M} \rightarrow T^*\mathcal{M}$ and $\sharp = \flat^{-1} : T^*\mathcal{M} \rightarrow T\mathcal{M}$ the musical isomorphisms associated to the metric g^D , defined by $\flat(X)(Y) = g(X, Y)$.

We can also define the *dual divergence* $D^* : (x, y) \mapsto D(y, x)$ and the *induced connection* by $\nabla_X^D Y : z \in \mathcal{M} \mapsto \sharp(Z \mapsto \partial_x^2|_{x=z} \partial_y|_{y=z} D(X, Y, Z))$.

Lemma 5.15 (Dual connections induced by a divergence) [Amari and Nagaoka, 2000] Let D be a divergence on \mathcal{M} . Then the connections $\nabla := \nabla^D$ and $\nabla^* := \nabla^{D^*}$ are dual with respect to the induced metric g^D : a divergence induces a dualistic structure.

In general, there is not a canonical way to define a divergence from a dualistic structure, except if it is dually-flat.

5.4.1.3 Canonical divergence of a dually-flat manifold

Definition 5.16 (Canonical divergence) [Amari and Nagaoka, 2000] Let $(\mathcal{M}, g, \nabla, \nabla^*)$ be a dually-flat manifold where \mathcal{M} is simply connected. Let $u, v : \mathcal{M} \rightarrow \mathbb{R}^n$ be two smooth coordinate systems such that u is ∇ -affine, v is ∇^* -affine and $g(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial v^j}) = \delta_{ij}$. The canonical divergence D is defined by $D(x, y) = \psi(x) + \varphi(y) - \langle u(x)|v(y) \rangle$ for all $x, y \in \mathcal{M}$ where $\langle \cdot | \cdot \rangle$ is the canonical inner product on \mathbb{R}^n and $\psi, \varphi : \mathcal{M} \rightarrow \mathbb{R}$ are smooth maps called potentials defined as follows:

1. $d\psi = \sum_i v^i du^i$ for all $i \in \{1, \dots, n\}$ or equivalently without coordinates $d_x \psi(X) = \langle v(x)|d_x u(X) \rangle$ for all $x \in \mathcal{M}$ and $X \in T_x \mathcal{M}$,
2. $d\varphi = \sum_i u^i dv^i$ for all $i \in \{1, \dots, n\}$ or $d_x \varphi(X) = \langle u(x)|d_x v(X) \rangle$,
3. $\psi(x) + \varphi(x) = \langle u(x)|v(x) \rangle$ for all $x \in \mathcal{M}$.

The equation $d\psi = \sum_i v^i du^i$ has a solution by Poincaré's lemma because \mathcal{M} is simply connected and the differential form $\omega = \sum_i v^i du^i$ is closed. Indeed, $g(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}) = \frac{\partial v^k}{\partial u^j} g(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial v^k}) = \frac{\partial v^i}{\partial u^j}$ and by symmetry of g , $g(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}) = \frac{\partial v^j}{\partial u^i}$ so $\frac{\partial v^i}{\partial u^j} = \frac{\partial v^j}{\partial u^i}$. So ψ is well defined up to an additive constant and φ as well. Finally, $d_x(\psi + \varphi)(X) = \langle v(x)|d_x u(X) \rangle + \langle u(x)|d_x v(X) \rangle = d_x(\langle u|v \rangle)(X)$ so there exists a constant $c \in \mathbb{R}$ such that $\psi(x) + \varphi(x) = \langle u(x)|v(x) \rangle + c$ for all $x \in \mathcal{M}$. We can impose $c = 0$ by choosing the constant in φ appropriately.

5.4.2 Principle of balanced bilinear forms

The principle of balanced bilinear forms [Thanwerdas and Pennec, 2019a] provides a bilinear form by combining the parallel transports of two flat metrics via the Frobenius inner product. We can give a more general definition of a balanced bilinear form by choosing any inner product on symmetric matrices, although we focus on the Frobenius inner product afterwards.

Principle 5.17 (Principle of balanced bilinear forms) We fix $\langle \cdot | \cdot \rangle$ an inner product on $\text{Sym}(n)$. Let g^+, g^- be two flat Riemannian metrics on $\text{Sym}^+(n)$. We denote ∇^+, ∇^- their Levi-Civita connections and Π^+, Π^- their associated parallel transport maps that do not depend on the curve since the metrics are flat. Then the balanced bilinear form associated to g^+ and g^- is defined by:

$$g_\Sigma^0(X, Y) = \langle \Pi_{\Sigma \rightarrow I_n}^+ X | \Pi_{\Sigma \rightarrow I_n}^- Y \rangle. \quad (5.12)$$

Theorem 5.18 (Relation between balanced metric and dually-flat manifold) [Thanwerdas and Pennec, 2019a] Let g^+, g^- be two flat Riemannian metrics on $\text{Sym}^+(n)$. We denote ∇^+, ∇^- their Levi-Civita connections. If the balanced bilinear form g^0 is a metric, then $(\text{Sym}^+(n), g^0, \nabla^+, \nabla^-)$ is a dually-flat manifold, which automatically comes with a canonical divergence D according to the previous section.

It would be nice to have a sufficient condition under which a balanced bilinear form is a metric. In [Thanwerdas and Pennec, 2019a], we proved that, with the Frobenius inner product, if g^+ and g^- are power-Euclidean metrics with powers α and β , then g^0 is a metric. In the following theorem, we give a weaker sufficient condition which allows to define the new family of Mixed-Euclidean metrics. The proof is in Section 11.4.

Theorem 5.19 (Sufficient condition for a balanced bilinear form to be a metric) Let $\langle \cdot | \cdot \rangle = \text{Frob}$ be the Frobenius inner product. Let g^+, g^- be deformed-Euclidean metrics respectively associated to univariate diffeomorphisms u and v . Then the balanced bilinear form g^0 is a metric.

See the proof of Theorem 5.19 in Section 11.4.

5.5 The new family of mixed-Euclidean metrics

5.5.1 Definition

Definition 5.20 (Mixed-Euclidean metric $\text{ME}(u, v)$) The (u, v) -Mixed-Euclidean metric is the balanced metric g^0 defined in Theorem 5.19. It is given by:

$$g_{\Sigma}^{\text{ME}(u,v)}(X, X) = \frac{1}{u'(1)v'(1)} \sum_{i,j} u^{[1]}(d_i, d_j) v^{[1]}(d_i, d_j) X_{ij}^2, \quad (5.13)$$

where $X = PX'P^{\top}$, $\Sigma = PDP^{\top}$ with $P \in O(n)$, $D = \text{diag}(d_1, \dots, d_n)$.

Remark 5.21 We notice that if we denote $\phi_u = \frac{u'(1)}{u^{[1]}}$ and $\phi_v = \frac{v'(1)}{v^{[1]}}$ the kernel maps associated to the u, v -deformed Euclidean metrics, the balanced metric is a kernel metric characterized by $\phi_{u,v} = \sqrt{\phi_u \phi_v}$. Hence, the principle of balanced bilinear forms seems to appear as a principle of mean of metrics.

The family of Mixed-Euclidean metrics contains the family of Mixed-Power-Euclidean metrics [Thanwerdas and Pennec, 2019a] for $u = F_{\alpha}$ and $v = F_{\beta}$ where $F_{\alpha} = \text{pow}_{\alpha}$ if $\alpha \neq 0$ and $F_0 = \log$.

$$\text{(Log-Euclidean)} \quad g_{\Sigma}^{\text{MPE}(0,0)}(X, X) = \text{tr}(d_{\Sigma} \log(X)^2), \quad (5.14)$$

$$\text{(Power-Euclidean)} \quad g_{\Sigma}^{\text{MPE}(\alpha,\alpha)}(X, X) = \frac{1}{\alpha^2} \text{tr}(d_{\Sigma} \text{pow}_{\alpha}(X)^2), \quad (5.15)$$

$$\text{(Power-affine)} \quad g_{\Sigma}^{\text{MPE}(\alpha,-\alpha)}(X, X) = \frac{1}{\alpha^2} \text{tr}((\Sigma^{-\alpha} d_{\Sigma} \text{pow}_{\alpha}(X))^2), \quad (5.16)$$

$$\text{("Power-BKM")} \quad g_{\Sigma}^{\text{MPE}(\alpha,0)}(X, X) = \frac{1}{\alpha} \text{tr}(d_{\Sigma} \text{pow}_{\alpha}(X) d_{\Sigma} \log(X)), \quad (5.17)$$

$$\text{(General MPE)} \quad g_{\Sigma}^{\text{MPE}(\alpha,\beta)}(X, X) = \frac{1}{\alpha\beta} \text{tr}(d_{\Sigma} \text{pow}_{\alpha}(X) d_{\Sigma} \text{pow}_{\beta}(X)). \quad (5.18)$$

As mentioned in [Thanwerdas and Pennec, 2019a], this family interpolates between the log-Euclidean metric $(0, 0)$, the power-Euclidean metrics (α, α) , the power-affine metrics $(\alpha, -\alpha)$ (including the affine-invariant metric $(1, -1)$) and the Bogoliubov-Kubo-Mori metric $(1, 0)$.

5.5.2 Information geometry of Mixed-Euclidean metrics

As said in Theorem 5.18, balanced metrics come with a canonical divergence. As Mixed-Euclidean metrics are the balanced metrics of two deformed-Euclidean metrics $u^* g^E$ and $v^* g^E$, it is straightforward that $u : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$ and $v : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$ provide flat

coordinate systems for these respective metrics. The canonical divergence of this structure is known as the (u, v) -divergence in Information Geometry [Amari, 2016, Section 4.5.2]. The novelty here is the relation we establish between Mixed-Euclidean metrics and (u, v) -divergences. In particular, the Mixed-Power-Euclidean metrics come with the so-called (α, β) -divergences on SPD matrices [Amari, 2014]. This family contains the well known families of α -divergences and β -divergences [Amari, 2014, Formulae 69,70]. We state the correspondence between Mixed-Euclidean metrics and (u, v) -divergences in the following corollary of Theorem 5.18. We recall the formulae of (α, β) -divergences with the corresponding potentials and we illustrate the correspondence with two charts.

Corollary 5.22 (Mixed-Euclidean metrics and (u, v) -divergences) Let u, v be two univariate diffeomorphisms $u, v : \text{Sym}^+(n) \rightarrow \text{Sym}^+(n)$. Then the manifold $(\text{Sym}^+(n), g^{\text{ME}(u,v)}, u^* \nabla^E, v^* \nabla^E)$ is dually-flat and its canonical divergence is the (u, v) -divergence of Information Geometry [Amari, 2016]. In particular, the manifold $(\text{Sym}^+(n), g^{\text{MPE}(p,q)}, \nabla^{\text{PE}(p)}, \nabla^{\text{PE}(q)})$ is dually-flat and its canonical divergence is the (α, β) -divergence [Amari, 2014, Formulae 51,54,56,66]. The (α, β) -divergences and the corresponding potentials (up to an additive constant, see Section 5.4.1.3) are:

$$(\alpha = \beta = 0) \quad D^{0,0}(\Sigma|\Sigma') = \frac{1}{2} \|\log(\Sigma) - \log(\Sigma')\|_{\text{Frob}}^2, \quad (5.19)$$

$$(\alpha = \beta \neq 0) \quad D^{\alpha,\alpha}(\Sigma|\Sigma') = \frac{1}{2\alpha^2} \|\Sigma^\alpha - \Sigma'^\alpha\|_{\text{Frob}}^2, \quad (5.20)$$

$$(\alpha = -\beta \neq 0) \quad D^{\alpha,-\alpha}(\Sigma|\Sigma') = -\frac{1}{\alpha^2} \text{tr} \left[(I_n + \alpha \log \Sigma) - \alpha \log \Sigma' - \Sigma^\alpha \Sigma'^{-\alpha} \right], \quad (5.21)$$

$$(\alpha \neq \beta = 0) \quad D^{\alpha,0}(\Sigma|\Sigma') = \frac{1}{\alpha} \text{tr} \left[\left(\Sigma^\alpha \log \Sigma - \frac{1}{\alpha} \Sigma^\alpha \right) + \frac{1}{\alpha} \Sigma'^\alpha - \Sigma^\alpha \log \Sigma' \right], \quad (5.22)$$

$$(\alpha, \beta, \alpha \pm \beta \neq 0) \quad D^{\alpha,\beta}(\Sigma|\Sigma') = \frac{1}{\alpha\beta} \text{tr} \left[\frac{\alpha}{\alpha + \beta} \Sigma^{\alpha+\beta} + \frac{\beta}{\alpha + \beta} \Sigma'^{\alpha+\beta} - \Sigma^\alpha \Sigma'^\beta \right], \quad (5.23)$$

$$(\alpha = \beta = 0) \quad \psi^{0,0}(\Sigma) = \frac{1}{2} \text{tr}(\log(\Sigma)^2), \quad (5.24)$$

$$(\alpha = \beta \neq 0) \quad \psi^{\alpha,\alpha}(\Sigma) = \frac{1}{2\alpha^2} \text{tr}(\Sigma^{2\alpha}), \quad (5.25)$$

$$(\alpha = -\beta \neq 0) \quad \psi^{\alpha,-\alpha}(\Sigma) = -\frac{1}{\alpha} \text{tr}(\log \Sigma) = -\frac{1}{\alpha} \log(\det \Sigma) \quad (5.26)$$

$$(\alpha \neq \beta = 0) \quad \psi^{\alpha,0}(\Sigma) = \frac{1}{\alpha} \text{tr}(\Sigma^\alpha \log \Sigma - \frac{1}{\alpha} \Sigma^\alpha), \quad (5.27)$$

$$(\alpha, \beta, \alpha \pm \beta \neq 0) \quad \psi^{\alpha,\beta}(\Sigma) = \frac{1}{\beta(\alpha + \beta)} \text{tr}(\Sigma^{\alpha+\beta}). \quad (5.28)$$

The (α, β) -divergences on SPD matrices can also be obtained by extending the (α, β) -divergences **on positive discrete measures** [Cichocki et al., 2011]. Indeed, a positive discrete measure is a vector of positive numbers so the diagonal matrix of eigenvalues of an SPD matrix can be considered as a positive discrete measure. Then the (α, β) -potential on positive diagonal matrices is extended by $O(n)$ -invariance, which defines the (α, β) -potential and the (α, β) -divergence on SPD matrices [Amari, 2014]. Conversely, the (α, β) -divergences

on SPD matrices define divergences on positive discrete measures when restricted to positive diagonal matrices. So there is a one-to-one correspondence between (α, β) -divergences on SPD matrices [Amari, 2014] (or Mixed-Power-Euclidean metrics) and (α, β) -divergences on positive discrete measures [Cichocki et al., 2011]. This correspondence is given on Figure 5.1. The graph on the right is essentially borrowed from [Cichocki et al., 2011] with complements from [Cichocki and Amari, 2010].

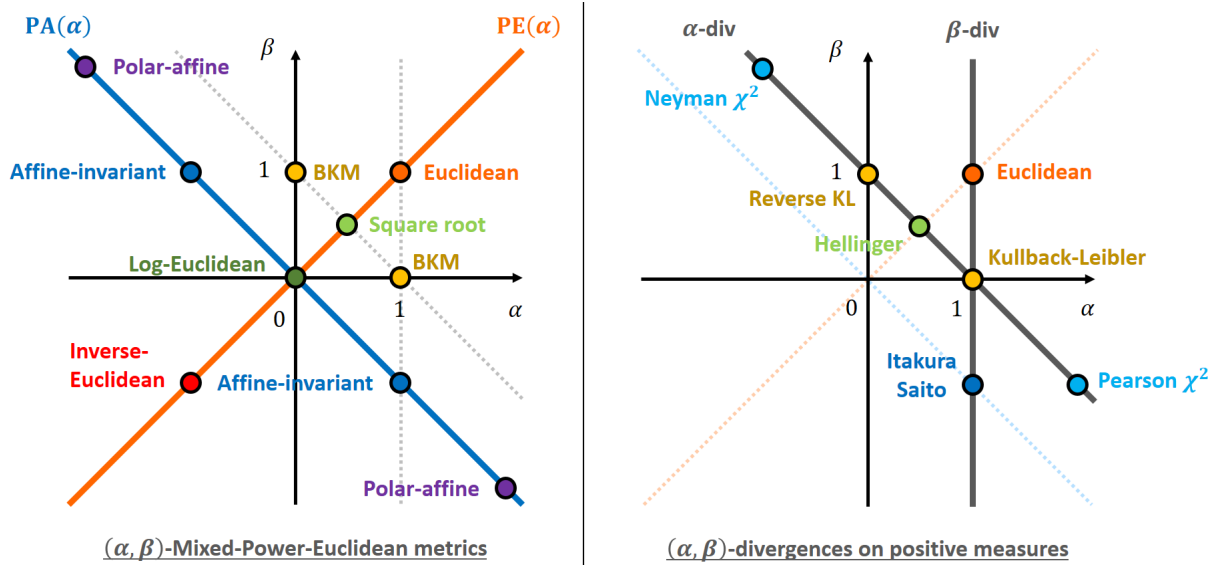


Figure 5.1: Correspondence between MPE metrics on SPD matrices and (α, β) -divergences on positive discrete measures.

Remark 5.23 The affine-invariant/Fisher-Rao metric is associated to the Kullback-Leibler divergence of **centered multivariate Gaussian densities**, which differs from the $O(n)$ -invariant extension of the Kullback-Leibler divergence of **positive discrete measures** represented on Figure 5.1.

The (u, v) -divergences can be expressed via an integral formula [Amari, 2016, Formula (4.170)] following Definition 5.16. The formulae of the previous corollary can thus be computed either from that formula or directly.

5.5.3 Riemannian geometry of Mixed-Euclidean metrics

Another immediate consequence of the relation between balanced metrics and dually-flat manifolds is that the Levi-Civita connection of the Mixed-Euclidean metric $\text{MPE}(u, v)$ is simply the arithmetic mean of the Levi-Civita connections of the deformed-Euclidean metrics u^*g^E and v^*g^E .

Corollary 5.24 (Levi-Civita connection of Mixed-Euclidean metrics)

$$\nabla_{X_\Sigma}^{\text{ME}(u,v)} Y = \partial_{X_\Sigma} Y + \frac{1}{2}((d_\Sigma u)^{-1}(H_\Sigma u(X, Y)) + (d_\Sigma v)^{-1}(H_\Sigma v(X, Y))) \quad (5.29)$$

It is even possible to compute the curvature following the same ideas as for the BKM metric in [Michor et al., 2000]. The proof is given in Section 11.4.

Theorem 5.25 (Curvature of Mixed-Euclidean metrics) Let $u, v : \text{Sym}^+(n) \rightarrow \text{Sym}^+(n)$ be two univariate diffeomorphisms. We define the univariate diffeomorphism $w = v \circ u^{-1}$ so that $u : (\text{Sym}^+(n), g^{\text{ME}(u,v)}) \rightarrow (\text{Sym}^+(n), \frac{w'(1)}{u'(1)v'(1)} g^{\text{ME}(\text{Id},w)})$ is an isometry. For $\Sigma = PDP^\top \in \text{Sym}^+(n)$, we denote $X = PX'P^\top \in T_\Sigma \text{Sym}^+(n)$ and analogously for $Y, Z, T \in T_\Sigma \text{Sym}^+(n)$, we denote $u_{ij} = u^{[1]}(d_i, d_j)$, $u_{ijk} = u^{[2]}(d_i, d_j, d_k)$ and analogously for v, w . We denote $m_{ij} = w^{[1]}(u(d_i), u(d_j)) = \frac{v_{ij}}{u_{ij}}$ and $m_{ijk} = w^{[2]}(u(d_i), u(d_j), u(d_k))$. Then the curvature of the mixed-Euclidean metric $g^{\text{ME}(u,v)}$ is:

$$R_\Sigma^{\text{ME}(u,v)}(X, Y, Z, T) = \frac{1}{u'(1)v'(1)} \sum_{i,j,k,l} \rho_{ijkl} (X'_{ij} Y'_{jk} Z'_{kl} T'_{li} - Y'_{ij} X'_{jk} Z'_{kl} T'_{li} + X'_{ij} Z'_{jk} Y'_{kl} T'_{li} - Y'_{ij} Z'_{jk} X'_{kl} T'_{li}), \quad (5.30)$$

where $\rho_{ijkl} = \frac{m_{ijl}m_{jlk}}{2m_{jl}} u_{ij} u_{jk} u_{kl} u_{li} = \frac{1}{2u_{jl}v_{jl}} (u_{ij}v_{ijl} - v_{ij}u_{ijl})(u_{jk}v_{jkl} - v_{jk}u_{jkl})$ is symmetric in $i \leftrightarrow k$, in $j \leftrightarrow l$ and in $u \leftrightarrow v$. In particular, at $\Sigma = I_n$, the curvature is:

$$R_{I_n}^{\text{ME}(u,v)}(X, Y, Z, T) = \frac{1}{4} \left[\left(\ln \left| \frac{v'}{u'} \right| \right)' (1) \right]^2 R_{I_n}^A(X, Y, Z, T), \quad (5.31)$$

where A stands for the affine-invariant metric (Formula 5.8). Therefore, the sectional curvature of the mixed-Euclidean metric at I_n takes non-positive values. In particular, for mixed-power-Euclidean metrics $\text{MPE}(\alpha, \beta)$ with $\alpha^2 \neq \beta^2$ (thus excluding log-Euclidean, power-Euclidean and power-affine metrics), since $\kappa_{\lambda\Sigma}^{\text{MPE}(\alpha,\beta)}(X, Y) = \lambda^{-(\alpha+\beta)} \times \kappa_\Sigma^{\text{MPE}(\alpha,\beta)}(X, Y)$ for all $\lambda > 0$, the lower bound of the sectional curvature is $-\infty$.

See the proof of Theorem 5.25 in Section 11.4.

It seems difficult to determine theoretically whether the sectional curvature of mixed-Euclidean metrics (again, excluding $\text{MPE}(\alpha, \beta)$ with $\alpha^2 = \beta^2$) can take positive values. On Figure 5.2, we show numerical results which make us think that this is the case. Indeed, we observe numerically that for all $\alpha, \beta \in \{0.05k \mid k \in \{-40, \dots, 40\}\}$ such that $\alpha^2 \neq \beta^2$, we have $\kappa_{\min}^{\text{MPE}(\alpha,\beta)} < 0$ and $\kappa_{\max}^{\text{MPE}(\alpha,\beta)} > 0$. These simulations also tend to show that, at a given point Σ , the negative values taken by the sectional curvature are much larger in absolute value than the positive values taken by the sectional curvature.

From Figure 5.2, it appears that power-Euclidean metrics (flat), power-affine metrics (Hadamard) and the log-Euclidean metric at the intersection play a special role among the family of Mixed-Power-Euclidean since all others apparently admit positive and negative sectional curvature.

In addition, for Mixed-Power-Euclidean metrics, we can also compute the geodesics, the logarithm map and the distance between commuting matrices. These formulae are proved in Section 11.4.

Theorem 5.26 (Riemannian operations of MPE metrics) Let $\alpha, \beta \in \mathbb{R}$ such that $\alpha + \beta \neq 0$, thus excluding log-Euclidean and power-affine metrics. Table 5.3 summarizes the formulae

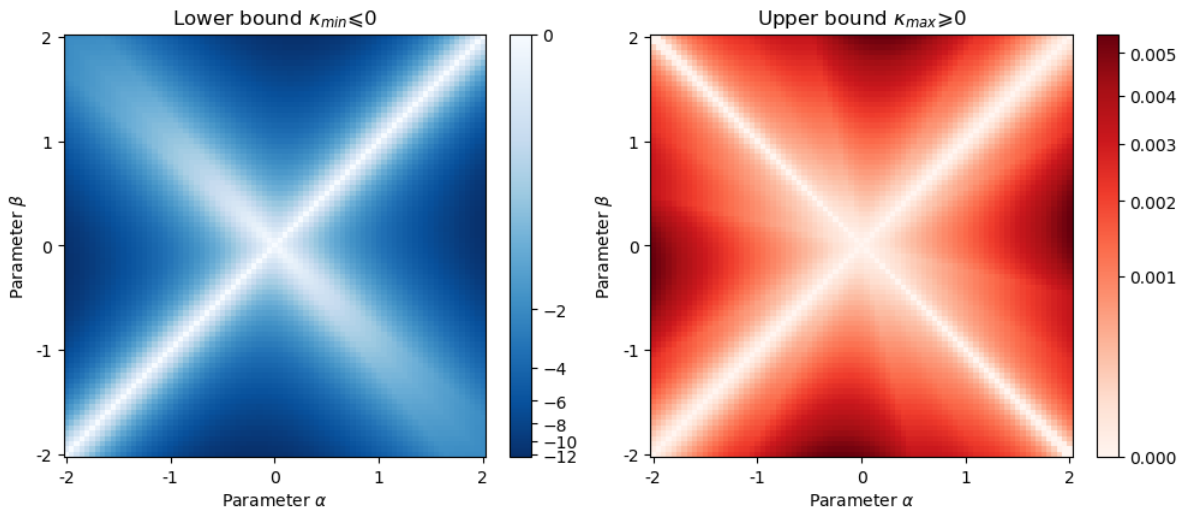


Figure 5.2: **Lower and upper bounds of the sectional curvature of the mixed-power-Euclidean metrics. Left: lower bound. Right: upper bound.** The lower bound of the power-affine metrics ($\beta = -\alpha$) is known as $-\frac{\alpha^2}{2}$ [Thanwerdas and Pennek, 2022b]. The bounds were obtained in dimension 3 by taking 1000 random positive diagonal matrices D of determinant 1 (to avoid scaling effects), 1000 random pairs of symmetric matrices (X, Y) and computing the sectional curvature $\kappa_D^{\text{MPE}(\alpha, \beta)}(X, Y) = \frac{R_D(X, Y, X, Y)}{g_D(X, X)g_D(Y, Y) - g_D(X, Y)^2}$ for $(\alpha, \beta) \in [-2, 2]^2$ with a step $\Delta\alpha = \Delta\beta = 0.05$. Diagonal matrices are taken instead of SPD matrices because the MPE metrics are $O(n)$ -invariant.

of the geodesics, the logarithm map and the distance in the particular case where $\Sigma, \Lambda \in \text{Sym}^+(n)$ and $V \in T_\Sigma \text{Sym}^+(n)$ **commute**. They essentially reduce to the formulae of the α_0 -power-Euclidean metric with $\alpha_0 = \frac{\alpha + \beta}{2}$. These formulae are generally not valid for non-commuting matrices.

See the proof of Theorem 5.26 in Section 11.4.

Geodesics	$\gamma_{(\Sigma, V)}(t) = (\Sigma^{\alpha_0} + t d_{\Sigma} \text{pow}_{\alpha_0}(V))^{1/\alpha_0}$
Logarithm	$\text{Log}_\Sigma(\Lambda) = (d_{\Sigma} \text{pow}_{\alpha_0})^{-1}(\Lambda^{\alpha_0} - \Sigma^{\alpha_0})$
Distance	$d(\Sigma, \Lambda) = \frac{1}{\alpha_0} \ \Lambda^{\alpha_0} - \Sigma^{\alpha_0}\ _{\text{Frob}}$

Table 5.3: Riemannian operations of Mixed-Power-Euclidean metrics for commuting matrices.

It would be tempting to generalize the formulae of geodesics, logarithm and distance between commuting matrices to Mixed-Euclidean metrics. However, if we consider two diffeomorphisms $u, v : (0, \infty) \rightarrow (0, \infty)$, the map $f = \sqrt{uv}$ (which generalizes $\text{pow}_{\alpha_0} = \sqrt{\text{pow}_\alpha \text{pow}_\beta}$) is not a diffeomorphism of $(0, \infty)$ in general. For example, take $u(x) = x(x+1)$ and $v(x) = \frac{1}{x^2}$. So the generalization is not straightforward.

5.6 Conclusion

Deforming a Riemannian metric is a general way of defining new metrics and new families of metrics on SPD matrices. In particular, using power diffeomorphisms defines one-parameter families which tend to the log-Euclidean metric when the power tends to 0. The class of kernel metrics is stable by univariate diffeomorphism whereas the class of mean kernel metrics is not. We showed that the alpha-Procrustes (or power-Wasserstein) metrics are mean kernel metrics when the power $p = 2\alpha \leq 1$.

We extended the principle of balanced bilinear forms and we gave a new sufficient condition under which the bilinear form is a metric. This allowed to define the new family of Mixed-Euclidean metrics which extends the two-parameter family of Mixed-Power-Euclidean metrics. Since balanced metrics define dually-flat manifolds which are characterized by a canonical divergence, Mixed-Euclidean metrics are in one-to-one correspondence with the (u, v) -divergences of information geometry. In particular, Mixed-Power-Euclidean metrics are in bijection with (α, β) -divergences. Finally, we computed the curvature of all Mixed-Euclidean metrics.

Some questions remain open. What are the conditions on the univariate diffeomorphisms u, v for the u -deformed Euclidean, affine, Wasserstein or the (u, v) -mixed-Euclidean metric to be a mean kernel metric? Are there more general conditions on two flat metrics for their balanced bilinear form to be a metric? What if we replace the Frobenius inner product by another one? Does the operation $(\phi, \phi') \mapsto \sqrt{\phi\phi'}$ on flat kernel metrics generalize the principle of balanced metrics?

More generally, since the two-parameter family of (α, β) -Mixed-Power-Euclidean metrics interpolate between the Euclidean, the log-Euclidean, the affine-invariant and the Bogoliubov-Kubo-Mori metrics, does there exist a principled family with three or four parameters which additionally includes the Bures-Wasserstein metric? It is not difficult to build parametric families of metrics which interpolate between all of them whereas it is more difficult to find interpolations with an interesting geometry such as the dually-flat geometry.

Part IV

Correlation matrices of full rank: the open ellipsope

Chapter 6

Geometry of Quotient-affine metrics

Abstract

Correlation matrices are used in many domains of neurosciences such as fMRI, EEG, MEG. However, statistical analyses often rely on embeddings into a Euclidean space or into Symmetric Positive Definite matrices which do not provide intrinsic tools. The quotient-affine metric was recently introduced as the quotient of the affine-invariant metric on SPD matrices by the action of diagonal matrices. In this work, we provide most of the fundamental Riemannian operations of the quotient-affine metric: the expression of the metric itself, the geodesics with initial tangent vector, the Levi-Civita connection and the curvature.

This chapter was presented and published in the proceedings of the conference Geometric Science of Information 2021 under the title “Geodesics and Curvature of the Quotient-Affine Metrics on Full-Rank Correlation Matrices” [Thanwerdas and Pennec, 2021]. All the proofs, originally in appendix of the paper, are deferred to Section 11.5. Besides, one result proved with Jonas Lueg at the conference GSI 2021 is added. Minor modifications were additionally brought for harmonization purposes.

6.1 Introduction

Correlation matrices are used in many domains with time series data such as functional brain connectivity in functional MRI, electroencephalography (EEG) or magnetoencephalography (MEG) signals. Full-rank correlation matrices form a strict sub-manifold of the cone of Symmetric Positive Definite (SPD) matrices sometimes called the (open) elliptope [Tropp, 2018]. However, very few geometric tools were defined for intrinsic computations with correlation matrices. For example, [Rebonato and Jaeckel, 2001] rely on a surjection from a product of n spheres of dimension $n - 1$ onto the space of correlation matrices in order to sample valid correlation matrices for financial applications: a point in the former space can be represented by an $n \times n$ matrix A with normed rows and therefore encodes a correlation matrix AA^\top . Since low-rank matrices have null measure, one gets a full-rank correlation matrix almost surely. More recently, the open elliptope was endowed with the Hilbert projective geometry [Nielsen and Sun, 2019] which relies on its convexity.

Since there exist efficient tools on SPD matrices (affine-invariant/Fisher-Rao metric, log-Euclidean metric, etc.), correlation matrices are often treated as SPD matrices [Varoquaux et al., 2010]. Nevertheless, these extrinsic tools do not respect the geometry of correlation matrices. Moreover, most of these tools on SPD matrices are invariant under orthogonal

transformations which is not compatible with correlation matrices. The ellipsope is not even stable by the orthogonal action. Hence using the tools designed for SPD matrices may not be relevant for correlation matrices. One could restrict a Riemannian metric from the cone of SPD matrices to the open ellipsope to define intrinsic tools by embedding. To the best of our knowledge, this hasn't been studied in depth. However, a key property of correlation matrices in applications with respect to covariance matrices is that the scale is renormalized independently for each axis. This physically corresponds to an invariance under a group action, namely the action of diagonal matrices on SPD matrices. The previously cited structures do not rely on this physical reality. This is why we are more interested in the recently introduced quotient-affine metric which corresponds to the quotient of the affine-invariant metric on SPD matrices by the action of diagonal matrices [David and Gu, 2019, David, 2019].

In this work, we investigate the geometry of this quotient-affine metric and we contribute additional tools to compute on this manifold. Based on the formalization of the quotient manifold with vertical and horizontal distributions, we compute in closed form some fundamental Riemannian operations of the quotient-affine metrics, notably the exponential map and the curvature. In Section 6.2, we recall how quotient-affine metrics are introduced, as well as the basics on quotient manifolds. In Section 6.3, we provide the following fundamental quotient and Riemannian operations of quotient-affine metrics: the vertical and horizontal projections, the metric, the exponential map, the Levi-Civita connection and the sectional curvature. This opens the way to many practical algorithms on the open ellipsope for different applications. Considering SPD matrices as the Cartesian product of positive diagonal matrices and full-rank correlation matrices, it also allows to define new Riemannian metrics which preserve the quotient-affine geometry on correlation matrices. Thus in Section 6.4, we illustrate the quotient-affine metric in dimension 2 by coupling it with the diagonal power-Euclidean metrics $g_D^{E(p)}(\Delta, \Delta) = \text{tr}(D^{2(p-1)}\Delta^2)$ for $p \in \{-1, 0, 1, 2\}$, where D is positive diagonal and Δ diagonal, and then by comparing it with the affine-invariant and the log-Euclidean metrics on SPD matrices.

6.2 Quotient-affine metrics

6.2.1 The quotient manifold of full-rank correlation matrices

The group of positive diagonal matrices $\text{Diag}^+(n)$ acts on the manifold of SPD matrices $\text{Sym}^+(n)$ via the congruence action $(\Sigma, D) \in \text{Sym}^+(n) \times \text{Diag}^+(n) \mapsto D\Sigma D \in \text{Sym}^+(n)$. The manifold of full-rank correlation matrices $\text{Cor}^+(n)$ can be seen as the quotient manifold $\text{Sym}^+(n)/\text{Diag}^+(n)$ via the invariant submersion Cor which computes the correlation matrix from a covariance matrix, $\text{Cor} : \Sigma \in \text{Sym}^+(n) \mapsto \text{Diag}(\Sigma)^{-1/2} \Sigma \text{Diag}(\Sigma)^{-1/2} \in \text{Cor}^+(n)$, where $\text{Diag}(\Sigma) = \text{diag}(\Sigma_{11}, \dots, \Sigma_{nn})$. Hence, any Riemannian metric G on $\text{Sym}^+(n)$ which is invariant under $\text{Diag}^+(n)$ induces a quotient metric g on $\text{Cor}^+(n)$. The steps to define it are the following.

1. Vertical distribution. $\mathcal{V}_\Sigma = \ker d_\Sigma \text{Cor}$ for all $\Sigma \in \text{Sym}^+(n)$.
2. Horizontal distribution. $\mathcal{H}_\Sigma := \mathcal{V}_\Sigma^\perp$ for all $\Sigma \in \text{Sym}^+(n)$, where the orthogonality \perp refers to the inner product G_Σ .

3. Horizontal lift. The linear map $d_\Sigma \text{Cor}$ restricted to the horizontal space \mathcal{H}_Σ is a linear isomorphism onto the tangent space of full-rank correlation matrices $(d_\Sigma \text{Cor})|_{\mathcal{H}_\Sigma} : \mathcal{H}_\Sigma \xrightarrow{\sim} T_{\text{Cor}(\Sigma)} \text{Cor}^+(n)$. The horizontal lift $\#$ is its inverse:

$$\# : X \in T_{\text{Cor}(\Sigma)} \text{Cor}^+(n) \xrightarrow{\sim} X^\# \in \mathcal{H}_\Sigma. \quad (6.1)$$

4. Quotient metric. It is defined by pullback through the horizontal lift:

$$\forall C \in \text{Cor}^+(n), \forall X \in T_C \text{Cor}^+(n), g_C(X, X) = G_\Sigma(X^\#, X^\#), \quad (6.2)$$

where $\Sigma \in \text{Cor}^{-1}(C)$ and the definition does not depend on the chosen Σ .

So the only missing ingredient is a Riemannian metric on SPD matrices which is invariant under the congruence action of positive diagonal matrices. In [David and Gu, 2019, David, 2019], the authors chose to use the famous affine-invariant/Fisher-Rao metric.

6.2.2 The affine-invariant metrics and the quotient-affine metrics

The affine-invariant metric is the Riemannian metric defined on SPD matrices by $G_\Sigma(V, V) = \text{tr}(\Sigma^{-1}V\Sigma^{-1}V)$ for all $\Sigma \in \text{Sym}^+(n)$ and $V \in T_\Sigma \text{Sym}^+(n)$ [Siegel, 1943, Skovgaard, 1984, Amari and Nagaoka, 2000]. It is invariant under the congruence action of the whole real general linear group $\text{GL}(n)$ which contains $\text{Diag}^+(n)$ as a subgroup. It provides a Riemannian symmetric structure to the manifold of SPD matrices, hence it is geodesically complete and the geodesics are given by the group action of one-parameter subgroups. We recall the exponential map, the Levi-Civita connection and the sectional curvature below for all $\Sigma \in \text{Sym}^+(n)$ and vector fields $V, W \in \Gamma(T\text{Sym}^+(n))$ where we also denote $V \equiv V_\Sigma$ and $W \equiv W_\Sigma \in T_\Sigma \text{Sym}^+(n)$:

$$\text{Exp}_\Sigma^G(V) = \Sigma^{1/2} \exp(\Sigma^{-1/2}V\Sigma^{-1/2})\Sigma^{1/2}, \quad (6.3)$$

$$(\nabla_V^G W)|_\Sigma = d_\Sigma W(V) - \frac{1}{2}(V\Sigma^{-1}W + W\Sigma^{-1}V), \quad (6.4)$$

$$\kappa_\Sigma^G(V, W) = \frac{1}{4} \frac{\text{tr}((\Sigma^{-1}V\Sigma^{-1}W - \Sigma^{-1}W\Sigma^{-1}V)^2)}{G(V, V)G(W, W) - G(V, W)^2} \leq 0. \quad (6.5)$$

The metrics that are invariant under the congruence action of the general linear group $\text{GL}(n)$ actually form a two-parameter family of metrics indexed by $\alpha > 0$ and $\beta > -\alpha/n$ [Pennec, 2009]: $G_\Sigma^{\alpha, \beta}(V, V) = \alpha \text{tr}(\Sigma^{-1}V\Sigma^{-1}V) + \beta \text{tr}(\Sigma^{-1}V)^2$. We call them all affine-invariant metrics. In particular, these metrics are invariant under the congruence action of diagonal matrices so they are good candidates to define Riemannian metrics on full-rank correlation matrices by quotient. In [David and Gu, 2019, David, 2019], the authors rely on the ‘‘classical’’ affine-invariant metric ($\alpha = 1, \beta = 0$). We generalize their definition below.

Definition 6.1 (Quotient-affine metrics on full-rank correlation matrices) The quotient-affine metric of parameters $\alpha > 0$ and $\beta > -\alpha/n$ is the quotient metric $g^{\alpha, \beta}$ on $\text{Cor}^+(n)$ induced by the affine-invariant metric $G^{\alpha, \beta}$ via the submersion $\text{Cor} : \Sigma \in \text{Sym}^+(n) \mapsto \text{Diag}(\Sigma)^{-1/2} \Sigma \text{Diag}(\Sigma)^{-1/2} \in \text{Cor}^+(n)$.

6.3 Fundamental Riemannian operations

In this section, we detail the quotient geometry of quotient-affine metrics $g^{\alpha,\beta}$. We give the vertical and horizontal distributions and projections in Section 6.3.1. We contribute the formulae of the metric itself in Section 6.3.2, the exponential map in Section 6.3.3, and finally the Levi-Civita connection and the sectional curvature in Section 6.3.4. To the best of our knowledge, all these formulae are new. They are proved in Section 11.5.

6.3.1 Vertical and horizontal distributions and projections

- Let \bullet be the Hadamard/Schur product on matrices defined by $[X \bullet Y]_{ij} = X_{ij}Y_{ij}$.
- Let $A : \Sigma \in \text{Sym}^+(n) \mapsto A(\Sigma) = \Sigma \bullet \Sigma^{-1} \in \text{Sym}^+(n)$. This smooth map is invariant under the action of positive diagonal matrices. The Schur product theorem ensures that $A(\Sigma) \in \text{Sym}^+(n)$. A fortiori, $I_n + A(\Sigma) \in \text{Sym}^+(n)$.
- Let $\psi : \mu \in \mathbb{R}^n \mapsto (\mu \mathbf{1}^\top + \mathbf{1} \mu^\top) \in \text{Sym}(n)$. This is an injective linear map. *This function was met in Chapter 3 in the $\mathfrak{S}(n)$ -irreducible decomposition of $\text{Sym}(n)$.*
- Let $\mathcal{S}_\Sigma(V)$ the unique solution of the Sylvester equation $\Sigma \mathcal{S}_\Sigma(V) + \mathcal{S}_\Sigma(V) \Sigma = V$ for $\Sigma \in \text{Sym}^+(n)$ and $V \in \text{Sym}(n)$.
- Let $\text{Hol}(n)$ be the vector space of symmetric matrices with vanishing diagonal (symmetric hollow matrices). Each tangent space of the manifold of full-rank correlation matrices can be seen as a copy of this vector space.

Theorem 6.2 (Vertical and horizontal distributions and projections) The vertical distribution is given by $\mathcal{V}_\Sigma = \Sigma \bullet \psi(\mathbb{R}^n)$ and the horizontal distribution is given by $\mathcal{H}_\Sigma = \mathcal{S}_{\Sigma^{-1}}(\text{Hol}(n))$. The vertical projection is:

$$\text{ver} : V \in T_\Sigma \text{Sym}^+(n) \mapsto \Sigma \bullet \psi((I_n + A(\Sigma))^{-1} \text{Diag}(\Sigma^{-1} V) \mathbf{1}) \in \mathcal{V}_\Sigma. \quad (6.6)$$

Then, the horizontal projection is simply $\text{hor}(V) = V - \text{ver}(V)$.

See the proof of Theorem 6.2 in Section 11.5.

6.3.2 Horizontal lift and metric

Theorem 6.3 (Horizontal lift) Let $\Sigma \in \text{Sym}^+(n)$ and $C = \text{Cor}(\Sigma) \in \text{Cor}^+(n)$. The horizontal lift at Σ of $X \in T_C \text{Cor}^+(n)$ is $X^\# = \text{hor}(\Delta_\Sigma X \Delta_\Sigma)$ with $\Delta_\Sigma = \text{Diag}(\Sigma)^{1/2}$. In particular, the horizontal lift at $C \in \text{Sym}^+(n)$ is $X^\# = \text{hor}(X)$.

See the proof of Theorem 6.3 in Section 11.5.

Theorem 6.4 (Expression of quotient-affine metrics) For all $C \in \text{Cor}^+(n)$ and $X \in T_C \text{Cor}^+(n)$, $g_C^{\alpha,\beta}(X, X) = \alpha g_C^{\text{QA}}(X, X)$ (independent from β) where:

$$g_C^{\text{QA}}(X, X) = \text{tr}((C^{-1} X)^2) - 2 \mathbf{1}^\top \text{Diag}(C^{-1} X) (I_n + A(C))^{-1} \text{Diag}(C^{-1} X) \mathbf{1}. \quad (6.7)$$

See the proof of Theorem 6.4 in Section 11.5.

6.3.3 Geodesics

The geodesics of a quotient metric are the projections of the horizontal geodesics of the original metric. This allows us to obtain the exponential map of the quotient-affine metrics.

Theorem 6.5 (Geodesics of quotient-affine metrics) The geodesic from $C \in \text{Cor}^+(n)$ with initial tangent vector $X \in T_C \text{Cor}^+(n)$ is:

$$\forall t \in \mathbb{R}, \gamma_{(C,X)}^{\text{QA}}(t) = \text{Exp}_C^{\text{QA}}(tX) = \text{Cor}(C^{1/2} \exp(tC^{-1/2} \text{hor}(X)C^{-1/2})C^{1/2}). \quad (6.8)$$

In particular, the quotient-affine metric is geodesically complete.

See the proof of Theorem 6.5 in Section 11.5.

The Riemannian logarithm between C_1 and $C_2 \in \text{Cor}^+(n)$ is much more complicated to compute. It amounts to find $\Sigma \in \text{Sym}^+(n)$ in the fiber above C_2 such that $\text{Log}_{C_1}^G(\Sigma)$ is horizontal. Then we have $\text{Log}_{C_1}^{\text{QA}}(C_2) = d_{C_1} \text{Cor}(\text{Log}_{C_1}^G(\Sigma))$. This means finding Σ that minimizes the affine-invariant distance in the fiber:

$$D = \arg \min_{D \in \text{Diag}^+(n)} d(C_1, DC_2D),$$

from which we get $\Sigma = DC_2D$. This is the method used in [David and Gu, 2019, David, 2019].

During the conference GSI 2021, Jonas Lueg had the idea to prove the existence of a minimizer by showing the coercivity of the smooth map $f : D \in \text{Diag}^+(n) \mapsto \text{tr}(\log(\Sigma D \Lambda D \Sigma)^2)$ for all $\Sigma, \Lambda \in \text{Sym}^+(n)$, i.e. show that $f(D) \rightarrow +\infty$ when one of the d_i 's tends to 0 or to $+\infty$. Thus we proved the following theorem. Note that the uniqueness remains an open problem.

Theorem 6.6 (Existence of a logarithm) For all $C_1, C_2 \in \text{Cor}^+(n)$, there exists $X \in T_C \text{Cor}^+(n)$ such that $\text{Exp}_{C_1}^{\text{QA}}(X) = C_2$.

See the proof of Theorem 6.6 in Section 11.5.

6.3.4 Levi-Civita connection and sectional curvature

In this section, we give the Levi-Civita connection and the curvature. The computations are based on the fundamental equations of submersions [O'Neill, 1966]. We denote $\text{sym}(M) = \frac{1}{2}(M + M^\top)$ the symmetric part of a matrix.

Theorem 6.7 (Levi-Civita connection and sectional curvature of quotient-affine metrics) The Levi-Civita connection of quotient-affine metrics is:

$$\begin{aligned} (\nabla_X^{\text{QA}} Y)|_C &= d_C Y(X) + \text{sym}[\text{Diag}(X^\#)Y^\# + \text{Diag}(Y^\#)X^\# + \text{Diag}(X^\#C^{-1}Y^\#)C \\ &\quad - X^\#C^{-1}Y^\# - \frac{1}{2}\text{Diag}(X^\#)C\text{Diag}(Y^\#) - \frac{3}{2}\text{Diag}(X^\#)\text{Diag}(Y^\#)C]. \end{aligned} \quad (6.9)$$

The curvature of quotient-affine metrics is:

$$\kappa_C^{\text{QA}}(X, Y) = \kappa_C^G(X^\#, Y^\#) + \frac{3}{4} \frac{G_C(\text{ver}[X^\#, Y^\#], \text{ver}[X^\#, Y^\#])}{g_C(X, X)g_C(Y, Y) - g_C(X, Y)^2} \quad (6.10)$$

$$= \frac{2 \text{tr}((C^{-1}X^\#C^{-1}Y^\# - C^{-1}Y^\#C^{-1}X^\#)^2) + 3 \mathbf{1}^\top D(I_n + A(C))^{-1} D \mathbf{1}}{8(g_C(X, X)g_C(Y, Y) - g_C(X, Y)^2)}, \quad (6.11)$$

where $[V, W] = dW(V) - dV(W)$ is the Lie bracket on $\text{Sym}^+(n)$ and $D = D(X, Y) - D(Y, X)$ with $D(X, Y) = \text{Diag}(C^{-1}\text{Diag}(X^\#)Y^\# - C^{-1}Y^\#C^{-1}\text{Diag}(X^\#)C)$. There is a slight abuse of notation because $\text{ver}[X^\#, Y^\#]$ induces that $X^\#$ and $Y^\#$ are vector fields. Indeed here, they are horizontal vector fields extending the horizontal lifts at C .

See the proof of Theorem 6.7 in Section 11.5.

The first term of the sectional curvature is negative, the second one is positive so we don't know in general the sign of the curvature of quotient-affine metrics.

The quotient-affine metrics not only provide a Riemannian framework on correlation matrices but also provide correlation-compatible statistical tools on SPD matrices if we consider that the space of SPD matrices is the Cartesian product of positive diagonal matrices and full-rank correlation matrices. We give a taste of such a construction in the next section.

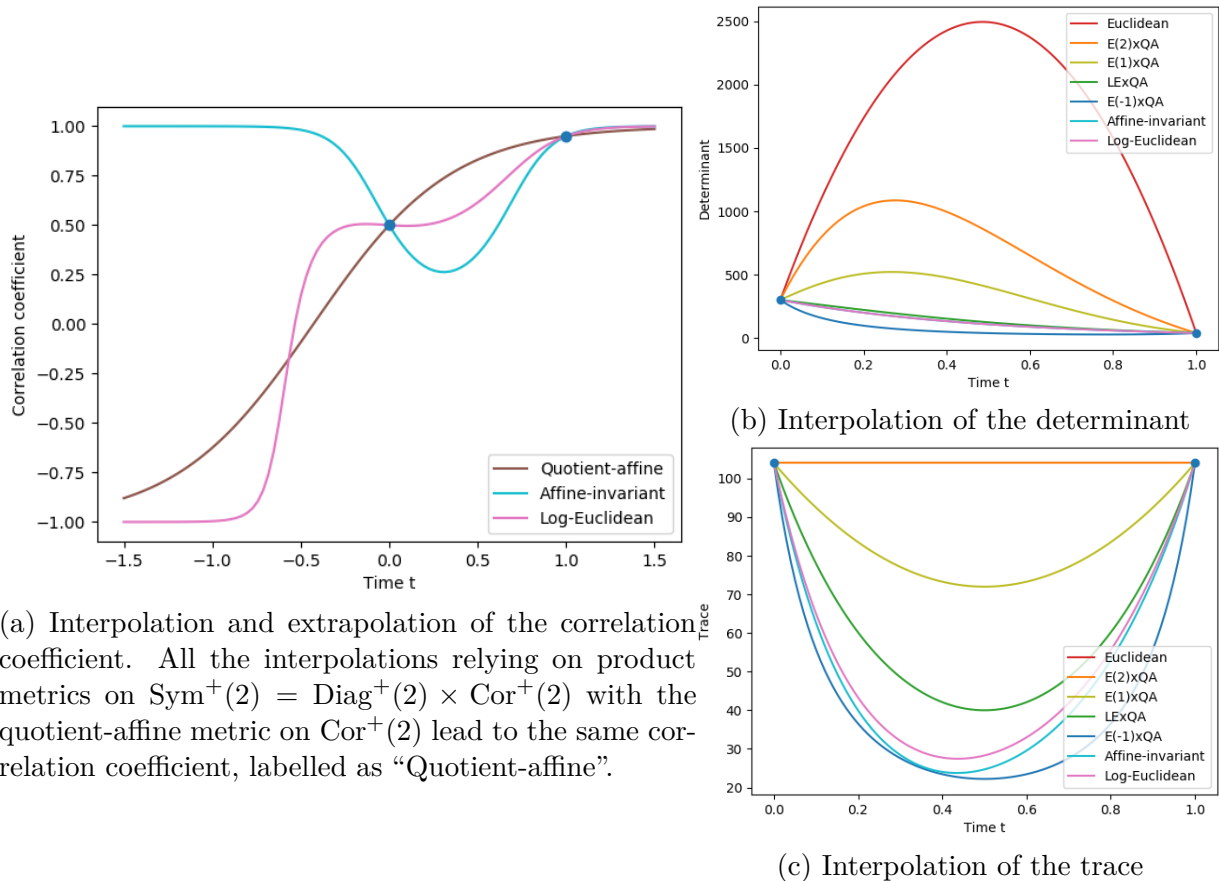
6.4 Illustration in dimension 2

In dimension 2, any correlation matrix $C \in \text{Cor}^+(2)$ writes $C = C(\rho) := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ where $\rho \in (-1, 1)$ is the correlation coefficient between the two variables. In the following theorem, we give explicitly the logarithm, the distance and the interpolating geodesics of the quotient-affine metric.

Theorem 6.8 (Quotient-affine metrics in dimension 2) Let $C_1 = C(\rho_1), C_2 = C(\rho_2) \in \text{Cor}^+(n)$ with $\rho_1, \rho_2 \in (-1, 1)$. We denote $f : \rho \in (-1, 1) \mapsto \frac{1+\rho}{1-\rho} \in (0, \infty)$ which is a smooth increasing map. We denote $\lambda = \lambda(\rho_1, \rho_2) = \frac{1}{2} \log \frac{f(\rho_2)}{f(\rho_1)}$ which has the same sign as $\rho_2 - \rho_1$. Then the quotient-affine operations are:

1. (Logarithm) $\text{Log}_{C_1}^{\text{QA}}(C_2) = \lambda \begin{pmatrix} 0 & 1 - \rho_1^2 \\ 1 - \rho_1^2 & 0 \end{pmatrix}$,
2. (Distance) $d^{\text{QA}}(C_1, C_2) = \sqrt{2}|\lambda|$,
3. (Geodesics) $\gamma_{(C_1, C_2)}^{\text{QA}}(t) = C(\rho^{\text{QA}}(t))$ where $\rho^{\text{QA}}(t) = \frac{\rho_1 \cosh(\lambda t) + \sinh(\lambda t)}{\rho_1 \sinh(\lambda t) + \cosh(\lambda t)} \in (-1, 1)$ is monotonic (increasing if and only if $\rho_2 - \rho_1 > 0$).

Let $\Sigma_1, \Sigma_2 \in \text{Sym}^+(2)$ and C_1, C_2 their respective correlation matrices. We denote $\gamma^{\text{AI}}, \gamma^{\text{LE}}$ the geodesics between Σ_1 and Σ_2 for the affine-invariant and the log-Euclidean metrics respectively. We define $\rho^{\text{AI}}, \rho^{\text{LE}}$ such that the correlation matrices of $\gamma^{\text{AI}}(t), \gamma^{\text{LE}}(t)$ are $C(\rho^{\text{AI}}(t)), C(\rho^{\text{LE}}(t))$. Figure 6.1.(a) shows $\rho^{\text{AI}}, \rho^{\text{LE}}, \rho^{\text{QA}}$ with $\Sigma_1 = \begin{pmatrix} 4 & 1 \\ 1 & 100 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 100 & 19 \\ 19 & 4 \end{pmatrix}$. When varying numerically Σ_1 and Σ_2 , it seems that ρ^{LE} and ρ^{AI} always have three inflection points. In contrast, ρ^{QA} always has one inflection point since $(\rho^{\text{QA}})'' = -2\lambda^2 \rho^{\text{QA}}(1 - (\rho^{\text{QA}})^2)$. Analogously, we compare the interpolations of the determinant (Fig. 6.1.(b)) and the trace (Fig. 6.1.(c)) using several Riemannian metrics: Euclidean (trace-monotonic); log-Euclidean and affine-invariant (determinant-monotonic); power-Euclidean \times quotient-affine (correlation-monotonic).



(a) Interpolation and extrapolation of the correlation coefficient. All the interpolations relying on product metrics on $\text{Sym}^+(2) = \text{Diag}^+(2) \times \text{Cor}^+(2)$ with the quotient-affine metric on $\text{Cor}^+(2)$ lead to the same correlation coefficient, labelled as “Quotient-affine”.

Figure 6.1: Extrapolation and interpolation between the SPD matrices Σ_1 and Σ_2 using various Riemannian metrics. The metrics $E(p) \times \text{QA}$ refer to the p -power-Euclidean metric on the diagonal part and the quotient-affine metric on the correlation part. When p tends to 0, $E(p)$ tends to the log-Euclidean metric $\text{LE} \equiv E(0)$.

On Figure 6.2, we compare the geodesics of these metrics. The determinant is the area of the ellipsis and the trace is the sum of the lengths of the axes. Thus the product metrics of the form power-Euclidean on the diagonal part and quotient-affine on the correlation part can be seen as performing a correlation-monotonic trade-off between the trace-monotonicity and the determinant-monotonicity.

6.5 Conclusion

We investigated in this chapter the very nice idea of quotienting the affine-invariant metrics on SPD matrices by the action of the positive diagonal group to obtain the principled quotient-affine metrics on full-rank correlation matrices. The quotient-affine metric with $\alpha = 1$ and $\beta = 0$ was first proposed in Paul David’s thesis [David, 2019] and in the subsequent journal paper [David and Gu, 2019]. We contribute here exact formulae for the main Riemannian operations, including the exponential map, the connection and the sectional curvature. The exponential map is particularly interesting to rigorously project tangent computations to

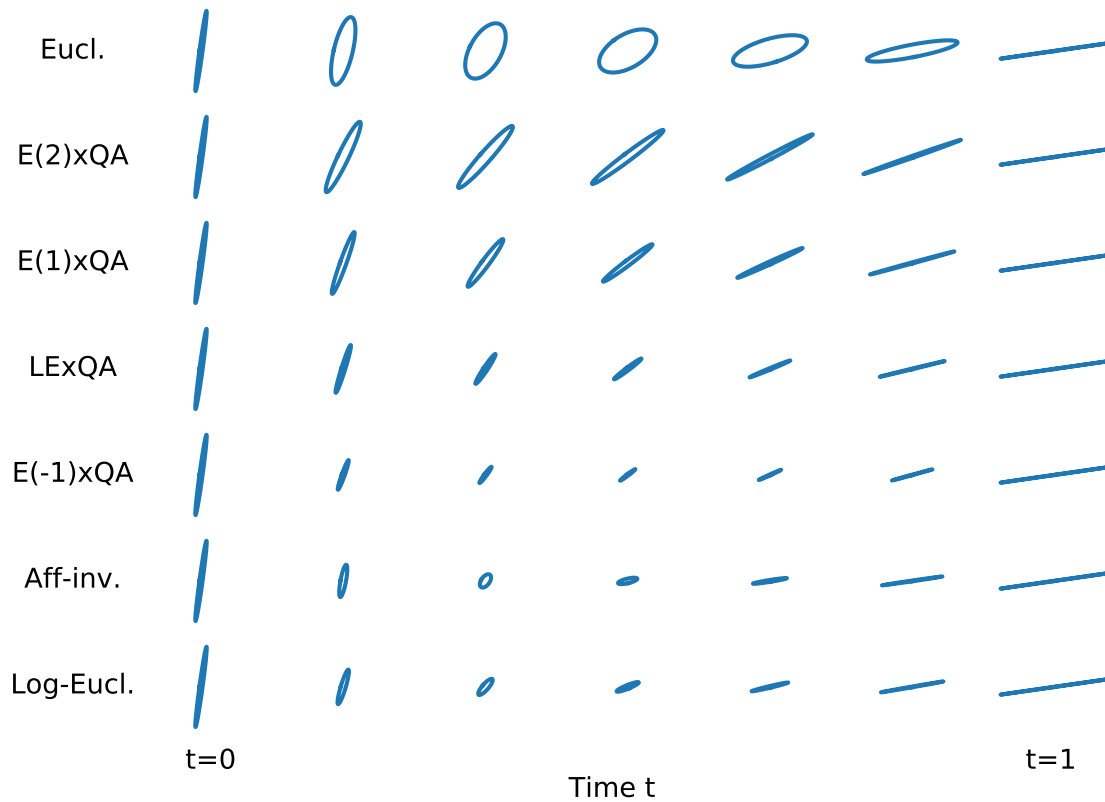


Figure 6.2: Interpolations between SPD matrices Σ_1 and Σ_2 .

the space of full-rank correlation matrices. This opens the way to the implementation of a number of generic algorithms on Riemannian manifolds. However, we could not find a closed form expression for the logarithm which remains to be computed through an optimization procedure. Thus, computing distances with these metrics remains computationally expensive. In order to obtain more efficient methods, this leads us to look for other principled Riemannian metrics on correlation matrices for which the logarithm could be expressed in closed form.

Chapter 7

Theoretically and computationally convenient Cholesky-based geometries

Abstract

In contrast to SPD matrices, few tools exist to perform Riemannian statistics on the open ellipsope of full-rank correlation matrices. The quotient-affine metric was recently built as the quotient of the affine-invariant metric by the congruence action of positive diagonal matrices. The space of SPD matrices had always been thought of as a Riemannian homogeneous space. In contrast, we view in this work SPD matrices as a Lie group and the affine-invariant metric as a left-invariant metric. This unexpected new viewpoint allows us to generalize the construction of the quotient-affine metric and to show that the main Riemannian operations can be computed numerically. However, the uniqueness of the Riemannian logarithm or the Fréchet mean are not ensured, which is bad for computing on the ellipsope. Hence, we define three new families of Riemannian metrics on full-rank correlation matrices which provide Hadamard structures, including two flat. Thus the Riemannian logarithm and the Fréchet mean are unique. We also define a nilpotent group structure for which the affine logarithm and the group mean are unique. We provide the main Riemannian/group operations of these four structures in closed form.

This chapter was submitted to the SIAM Journal on Matrix Analysis and its Applications in January 2022 under the title “Theoretically and computationally convenient geometries on full-rank correlation matrices” [Thanwerdas and Pennec, 2022c]. All the proofs, whatever their location in the submitted version, are deferred to Section 11.6.

7.1 Introduction

The (open) ellipsope is the set of full-rank correlation matrices, it is open in the affine space of symmetric matrices with unit diagonal. Its geometry has been much less studied than the one of the cone of Symmetric Positive Definite (SPD) matrices. Nevertheless, several applications could benefit from well suited geometries on this manifold: graphical networks, brain connectomes [Varoquaux et al., 2010, Dodero et al., 2015], finance [Rebonato and Jaeckel, 2001, Marti et al., 2021] or phylogenetic trees [Garba et al., 2021]. A few operations have been proposed such as sampling by projecting samples on spheres [Rebonato and Jaeckel, 2001, Kercheval, 2008], computing distances via the Hilbert geometry of convex sets [Nielsen

and Sun, 2019] or via a recent parametrization by $\mathbb{R}^{n(n-1)/2}$ [Archakov and Hansen, 2021].

The open ellipsope was also recently characterized as the quotient manifold of SPD matrices by the smooth, proper and free congruence action of positive diagonal matrices [David and Gu, 2019, David, 2019]. Indeed, a correlation matrix is obtained from the covariance matrix by dividing by the standard deviation of each variable. This is equivalent to multiply the covariance matrix on left and right by the diagonal matrix of the inverse square roots of the variances. Hence all covariance matrices that are congruent up to a diagonal matrix represent the same correlation matrix. This structure allowed to quotient the well known affine-invariant metrics on SPD matrices to the so called quotient-affine metrics on full-rank correlation matrices [David and Gu, 2019, David, 2019, Thanwerdas and Pennec, 2021]. They only differ by a scaling factor so they are often referred as *the* quotient-affine metric. It offers promising perspectives since it is geodesically complete with closed form expressions for the exponential map, the Levi-Civita connection and the sectional curvature. However, it was difficult to determine the sign and potential bounds of the curvature of the quotient-affine metric. In this chapter, we show that the sectional curvature can take both positive and negative values, that it is bounded from below and unbounded from above. Therefore, the ellipsope of full-rank correlation matrices endowed with the quotient-affine metric is neither a Hadamard space nor a CAT(k) space for any $k \in \mathbb{R}$.

This calls for new Riemannian metrics on the ellipsope. Indeed, we could expect to find suitable metrics with non-positive or even null curvature since the ellipsope is an open set of an affine space. The structure of the space is an important element of modeling because simple structures often bring good theoretical properties and better computability of the geometric operations. For example, in a Hadamard space, the Fréchet mean is unique. Recall that in a metric space (\mathcal{M}, d) , a Fréchet mean of points $x_1, \dots, x_k \in \mathcal{M}$ is a point $x \in \mathcal{M}$ which minimizes the function $x \in \mathcal{M} \mapsto \sum_{i=1}^k d(x, x_i)^2 \geq 0$. This is the case of the affine-invariant metric on SPD matrices [Skovgaard, 1984] or the Fisher metric of beta and Dirichlet distributions [Le Brigant et al., 2021]. Moreover, if one has a Euclideanization of the space, i.e. a smooth diffeomorphism to a Euclidean space, then all operations become trivial. This is the case of the log-Euclidean [Arsigny et al., 2006] or the log-Cholesky [Lin, 2019] metrics on SPD matrices.

Another strong element of modeling is the invariance of the geometry under a given group action. Let us give some examples. First, on SPD matrices, one could require the invariance of the geometry under all affine transformations of the feature vector. Indeed in EEG, if we assume that there is an affine transformation from one brain to another at first order, then electromagnetic fields are transformed similarly since they satisfy linear equations. The affine-invariant metric significantly improved the results of classification in BCI [Barachant et al., 2013]. Second, one could want the analysis to be invariant from the scale of each variable. As explained previously, this corresponds to the invariance under the congruence action of positive diagonal matrices and it suggests to focus more on correlation matrices than on covariance matrices. Third, all the operations mentioned above on correlation matrices are invariant under permutations. It means that the statistical analyses are invariant under any joint permutation of the rows and columns of the correlation matrices in the dataset. It is an advantage if the way the variables (or channels) are ordered is arbitrary. It is a drawback if the order is chosen for a certain reason. For instance, in the auto-correlation matrix of a signal, the variables represent different times, which are not exchangeable. One

could want a structure on correlation matrices that respects the time structure instead of being invariant under permutations. In this chapter, we focus on non-permutation-invariant geometric structures.

Moreover, an invariance may also lead to simple structures with good properties. For example, a Riemannian homogeneous space is geodesically complete, which is well suited for interpolation and extrapolation. The geodesics of naturally reductive homogeneous spaces are the orbits of the one-parameter subgroups so they are known in closed form. In a Riemannian symmetric space, the parallel transport is obtained in closed form by composition of two symmetries. So it is another good reason to study Lie group actions on our space.

7.1.1 Overview of the results

In this chapter, we define new non-permutation-invariant geometric structures on full-rank correlation matrices following two directions. Firstly, we define a non-permutation-invariant generalization of quotient-affine metrics. Our approach consists in studying the congruence action of several matrix Lie groups on SPD matrices. Our main result is a characterization of affine-invariant metrics, i.e. metrics that are invariant under the congruence action of the general linear group $GL(n)$, by the joint invariance of a pair of subgroups. These are the group of permutation matrices $\mathfrak{S}(n)$ and the group of lower triangular matrices with positive diagonal $LT^+(n)$. In other words, the affine-invariant metrics are the unique $(\mathfrak{S}(n) \times LT^+(n))$ -invariant metrics on SPD matrices. Therefore, the family of $LT^+(n)$ -invariant metrics appears as a natural non-permutation-invariant generalization of affine-invariant metrics. Moreover, we show that such metrics are exactly the pullback metrics of left-invariant metrics on the Lie group $LT^+(n)$ by the Cholesky map, so we call them Lie-Cholesky metrics. In particular, there exists a Lie group structure on SPD matrices (namely, the product of the Cholesky factors) such that affine-invariant metrics are left-invariant metrics on that Lie group. Finally, since the Lie group $LT^+(n)$ contains positive diagonal matrices, Lie-Cholesky metrics descend to quotient metrics on full-rank correlation matrices by the same procedure as affine-invariant metrics. We show that several Riemannian operations are numerically computable such as the Riemannian metric, the exponential map, the logarithm map or the Riemannian distance. However, despite the nice theoretical results and the computability of the main geometric operations, quotient-Lie-Cholesky metrics don't have obvious nice geometric properties. For example, it is not clear whether the Riemannian logarithm and the Fréchet mean are unique or not.

Hence, secondly, to solve the drawbacks of quotient-affine and quotient-Lie-Cholesky metrics, we propose a series of new geometries on full-rank correlation matrices for which the mean of finite samples is unique. They are built in a different way than the Lie-Cholesky metrics, the only common point being the use of the Cholesky map to convey the structures from triangular matrices to correlation matrices. Hence they are not permutation-invariant either for $n \geq 3$. We define the poly-hyperbolic-Cholesky metrics as the pullbacks of weighted products of hyperbolic spaces $\mathbb{H}^1 \times \cdots \times \mathbb{H}^{n-1}$. The metrics of this family provide a structure of Riemannian symmetric space with non-positive bounded curvature, thus Hadamard. All operations are known in closed form. We define the Euclidean-Cholesky and the log-Euclidean-Cholesky metrics as the pullbacks by two maps derived from the Cholesky map of Euclidean structures on the vector space $LT^0(n)$ of strictly lower triangular matrices. The

metrics in these two families are flat and geodesically complete so all Riemannian operations are known in closed form, the former being less costly than the latter. Finally, we define a Lie group structure as the pullback of the Lie group (for matrix multiplication) $LT^1(n)$ of lower triangular matrices with unit diagonal. Here, we consider the geometric structure induced by the canonical Cartan-Schouten connection rather than by a Riemannian metric. Indeed, the (group) exponential map allows to define a group mean, which is unique here because the Lie algebra of the Lie group is nilpotent [Buser and Karcher, 1981]. Finally, in dimension 2, we show that on the one hand, the quotient-affine and the poly-hyperbolic-Cholesky metrics coincide, and on the other hand, the Euclidean-Cholesky and the log-Euclidean-Cholesky metrics coincide and their geodesics coincide with the group geodesics. In particular, these geodesics provide a new interpolation of the correlation coefficient, different than the one proposed in [Thanwerdas and Penneec, 2021].

In the remainder of this section, we introduce the matrix notations, the notions of covariance and correlation matrices and the Cholesky map. In Section 7.2, we recall the definition of the quotient-affine metrics and we prove that its sectional curvature takes both negative and positive values and is unbounded from above. In Section 7.3, we give a characterization of affine-invariant metrics in function of the congruence action of other groups. This allows us to introduce Lie-Cholesky metrics on SPD matrices and quotient-Lie-Cholesky metrics on full-rank correlation matrices. In Section 7.4, we introduce four new non-permutation-invariant structures on full-rank correlation matrices for which the mean is unique. We conclude in Section 7.5.

7.1.2 Concepts and notations

7.1.2.1 Matrix notations

Our main matrix space notations are given in Table 7.1. In this chapter, we also use the following linear maps:

- $(A, B) \in \text{Mat}(n) \times \text{Mat}(n) \mapsto A \bullet B = [A_{ij}B_{ij}]_{1 \leq i, j \leq n} \in \text{Mat}(n)$ is the Hadamard/Schur product of matrices,
- $\text{Diag} : M \in \text{Mat}(n) \mapsto [\delta_{ij}M_{ij}]_{1 \leq i, j \leq n} \in \text{Diag}(n)$ selects the diagonal terms,
- $\text{Off} = \text{Id}_{\text{Mat}(n)} - \text{Diag} : \text{Mat}(n) \longrightarrow \ker \text{Diag}$ selects the off-diagonal terms,
- $\text{Low} : M \in \text{Mat}(n) \mapsto [\delta_{i \geq j}M_{ij}]_{1 \leq i, j \leq n} \in \text{LT}(n)$ selects the lower triangular terms, *including* the diagonal terms,
- $\text{Low}_0 = \text{Low} - \text{Diag} : \text{Mat}(n) \longrightarrow \text{LT}^0(n)$ selects the strictly lower triangular terms, *excluding* the diagonal terms,
- $\text{Low}_S = \text{Low}_0 + \frac{1}{2}\text{Diag} : \text{Sym}(n) \longrightarrow \text{LT}(n)$ selects the strictly lower triangular terms and half of the diagonal terms, so that when a symmetric matrix writes $M = L + L^\top \in \text{Sym}(n)$ with $L \in \text{LT}(n)$, then $L = \text{Low}_S(M)$,
- $\text{Sum} : M \in \text{Mat}(n) \mapsto \sum_{i,j} M_{ij} = \mathbf{1}^\top M \mathbf{1}$ sums the terms of the matrix,
- $\text{sum} : x \in \mathbb{R}^n \mapsto \sum_i x_i = \mathbf{1}^\top x$ sums the terms of the vector,

Matrix vector spaces		Matrix manifolds	
$\text{Mat}(n)$	Squared	$\text{GL}(n)$	General linear group
$\text{Skew}(n)$	Skew-symmetric	$\text{O}(n)$	Orthogonal group
		$\text{SO}(n)$	Special orthogonal group
$\text{Sym}(n)$	Symmetric	$\text{Sym}^+(n)$	Sym Positive Definite cone
$\text{Hol}(n)$	Symmetric hollow	$\text{Cor}^+(n)$	Full-rank correlation ellipptope
$\text{LT}(n)$	Lower Triangular	$\text{LT}^+(n)$	LT with positive diagonal
$\text{LT}^0(n)$	LT with null diagonal	$\text{LT}^1(n)$	LT with unit diagonal
$\text{Diag}(n)$	Diagonal	$\text{Diag}^+(n)$	Positive diagonal group

Table 7.1: Matrix space notations.

where $\mathbb{1} = (1, \dots, 1)^\top \in \mathbb{R}^n$.

The possibly uncommon notations in Table 7.1 are:

- $\text{Sym}^+(n) = \{\Sigma \in \text{Sym}(n) \mid \Sigma > 0\}$ where $>$ is the Loewner order,
- $\text{Cor}^+(n) = \{\Sigma \in \text{Sym}^+(n) \mid \text{Diag}(\Sigma) = I_n\}$,
- $\text{Hol}(n) = \{M \in \text{Sym}(n) \mid \text{Diag}(M) = 0\}$ where “hollow” means vanishing diagonal,
- $\text{LT}(n) = \{\text{Low}(M) \mid M \in \text{Mat}(n)\}$,
- $\text{LT}^+(n) = \{L \in \text{LT}(n) \mid \text{Diag}(L) \in \text{Diag}^+(n)\}$,
- $\text{LT}^0(n) = \{L \in \text{LT}(n) \mid \text{Diag}(L) = 0\}$,
- $\text{LT}^1(n) = \{L \in \text{LT}(n) \mid \text{Diag}(L) = I_n\} = I_n + \text{LT}^0(n)$.

We denote $\mathfrak{S}(n)$ the permutation group of order n .

7.1.2.2 Covariance and correlation matrices

Given an invertible covariance matrix $\Sigma = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq n} \in \text{Sym}^+(n)$ of a random vector X , the corresponding correlation matrix is defined by $C = \text{Cor}(\Sigma) = (\text{Cor}(X_i, X_j))_{1 \leq i, j \leq n}$ where:

$$\text{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i) \mathbb{E}(X_j), \quad (7.1)$$

$$\text{Cor}(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Cov}(X_i, X_i)} \sqrt{\text{Cov}(X_j, X_j)}} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}} \sqrt{\Sigma_{jj}}} \quad (7.2)$$

$$= [\text{Diag}(\Sigma)^{-1/2} \Sigma \text{Diag}(\Sigma)^{-1/2}]_{ij}, \quad (7.3)$$

with $\text{Diag}(\Sigma) = \text{diag}(\Sigma_{11}, \dots, \Sigma_{nn}) \in \text{Diag}^+(n)$. Moreover, if $D \in \text{Diag}^+(n)$, then the correlation matrix associated to $D\Sigma D$ is again $\text{Cor}(\Sigma)$. It is known as the invariance of the correlation matrix under the scaling of each component of the random vector. In other words, the surjective map $\text{Cor} : \Sigma \in \text{Sym}^+(n) \mapsto \text{Cor}(\Sigma) \in \text{Cor}^+(n)$ is invariant under the group action $(D, \Sigma) \in \text{Diag}^+(n) \times \text{Sym}^+(n) \mapsto D\Sigma D \in \text{Sym}^+(n)$ and the orbit space $\text{Sym}^+(n)/\text{Diag}^+(n)$ can be identified with $\text{Cor}^+(n)$. Note that, via this identification, the induced topology $\text{Cor}^+(n) \leftrightarrow \text{Sym}(n)$ coincides with the quotient topology of $\text{Sym}^+(n)/\text{Diag}^+(n)$. The action being proper [David and Gu, 2019], the quotient manifold theorem [Lee, 2012] states

that the orbit space $\text{Sym}^+(n)/\text{Diag}^+(n)$ has a unique smooth manifold structure such that the canonical surjection $\text{Sym}^+(n) \rightarrow \text{Sym}^+(n)/\text{Diag}^+(n)$ is a smooth submersion. Since $\text{Cor} : \text{Sym}^+(n) \rightarrow \text{Cor}^+(n)$ is a smooth submersion and since $\text{Cor}^+(n)$ is $\text{Sym}^+(n)/\text{Diag}^+(n)$ as a topological space, this smooth structure coincides with the induced smooth structure of $\text{Cor}^+(n)$.

7.1.2.3 The Cholesky map

A Cholesky decomposition of a symmetric positive semi-definite matrix Σ is a factorization of the form $\Sigma = LL^\top$ where $L \in \text{LT}(n)$ is a lower triangular matrix. If Σ is positive definite, then there exists a unique triangular matrix with positive diagonal $L \in \text{LT}^+(n)$ such that $\Sigma = LL^\top$. This allows to define the Cholesky bijective map:

$$\text{Chol} : \Sigma \in \text{Sym}^+(n) \mapsto L \in \text{LT}^+(n), \quad (7.4)$$

whose inverse is the smooth map $\phi : L \in \text{LT}^+(n) \mapsto LL^\top \in \text{Sym}^+(n)$. Moreover, the Cholesky map is smooth. Indeed, it is the product of two smooth maps $\text{Chol}(\Sigma) = L(\Sigma)\sqrt{D(\Sigma)}$ where $L \equiv L(\Sigma) \in \text{LT}^1(n)$ and $D \equiv D(\Sigma) \in \text{Diag}^+(n)$ are recursively defined for all $(i, j) \in \{1, \dots, n\}^2$ with $i > j$ by:

$$D_{ii} = \Sigma_{ii} - \sum_{j=1}^{i-1} L_{ij}^2 D_{jj} > 0, \quad (7.5)$$

$$L_{ij} = \frac{1}{D_{jj}} \left(\Sigma_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk} D_{kk} \right), \quad (7.6)$$

the order of the computations being $\{D_{11}\} \rightarrow \dots \rightarrow \{L_{i1} \rightarrow L_{i2} \rightarrow \dots \rightarrow L_{i,i-1} \rightarrow D_{ii}\} \rightarrow \dots \rightarrow \{L_{n1} \rightarrow \dots \rightarrow L_{n,n-1} \rightarrow D_{nn}\}$.

We denote:

$$\mathcal{L} = \text{Chol}(\text{Cor}^+(n)) = \left\{ L = \begin{pmatrix} L_1 \\ \vdots \\ L_n \end{pmatrix} \in \text{LT}^+(n) \mid \forall i \in \{1, \dots, n\}, \|L_i\|^2 = L_i L_i^\top = 1 \right\},$$

which is diffeomorphic to $\text{Cor}^+(n)$ via the Cholesky map. We also define a Cholesky-based diffeomorphism between $\text{Cor}^+(n)$ and $\text{LT}^1(n)$:

$$\Theta : C \in \text{Cor}^+(n) \mapsto \Gamma = \text{Diag}(\text{Chol}(C))^{-1} \text{Chol}(C) \in \text{LT}^1(n), \quad (7.7)$$

$$\Phi : \Gamma \in \text{LT}^1(n) \mapsto C = \text{Diag}(\Gamma \Gamma^\top)^{-1/2} \Gamma \Gamma^\top \text{Diag}(\Gamma \Gamma^\top)^{-1/2} \in \text{Cor}^+(n). \quad (7.8)$$

We clearly have the relations $\Phi = \Theta^{-1} = \text{Cor} \circ \phi$.

We compute the differentials of $\phi : \text{LT}^+(n) \rightarrow \text{Sym}^+(n)$ and $\text{Chol} : \text{Sym}^+(n) \rightarrow \text{LT}^+(n)$. For all $Z \in T_L \text{LT}^+(n) \simeq \text{LT}(n)$:

$$V := d_L \phi(Z) = ZL^\top + LZ^\top, \quad (7.9)$$

$$L^{-1}VL^{-\top} = L^{-1}Z + (L^{-1}Z)^\top,$$

$$\text{Low}_S(L^{-1}VL^{-\top}) = L^{-1}Z,$$

$$Z = d_\Sigma \text{Chol}(V) = L \text{Low}_S(L^{-1}VL^{-\top}). \quad (7.10)$$

In particular, $d_{I_n} \text{Chol} = \text{Low}_S$.

7.2 Quotient-affine metrics

In this section, we briefly recall how to build quotient metrics to fix the notations (Section 7.2.1), then we recall the definition of quotient-affine metrics (Section 7.2.2) and finally we show that the sectional curvature takes both negative and positive curvature and is unbounded from above (Section 7.2.3).

Sections 7.2.1 and 7.2.2 partly summarize Chapter 6 to allow Chapter 7 to be read independently. The skilled reader may skip these sections to resume at Section 7.2.3.

7.2.1 Quotient metrics

Given smooth manifolds $\mathcal{M}, \mathcal{M}'$ and a smooth submersion $\pi : \mathcal{M} \rightarrow \mathcal{M}'$, we can define the vertical space $\mathcal{V}_x = \ker d_x\pi \subset T_x\mathcal{M}$, where $d_x\pi : T_x\mathcal{M} \rightarrow T_{\pi(x)}\mathcal{M}'$ is the differential of the map π at point $x \in \mathcal{M}$. The horizontal space \mathcal{H}_x can be any supplementary vector space, i.e. such that $\mathcal{V}_x \oplus \mathcal{H}_x = T_x\mathcal{M}$. Given a horizontal distribution $x \mapsto \mathcal{H}_x$, there exist vertical and horizontal projections $\text{ver}_x : T_x\mathcal{M} \rightarrow \mathcal{V}_x$ and $\text{hor}_x : T_x\mathcal{M} \rightarrow \mathcal{H}_x$. Moreover, the linear map $(d_x\pi)|_{\mathcal{H}_x} : \mathcal{H}_x \rightarrow T_{\pi(x)}\mathcal{M}'$ is an isomorphism. Its inverse isomorphism is called the horizontal lift and denoted $\#_x : X \in T_{\pi(x)}\mathcal{M}' \rightarrow X_x^\# \in \mathcal{H}_x$. When \mathcal{M} is endowed with a Riemannian metric, there is a canonical choice of horizontal space which is the orthogonal $\mathcal{H}_x = \mathcal{V}_x^\perp$. In this case, the projections are orthogonal.

Given a smooth manifold \mathcal{M} on which a Lie group G acts smoothly, properly and freely, the quotient space \mathcal{M}/G admits a unique smooth manifold structure that turns the canonical projection $\pi : \mathcal{M} \rightarrow \mathcal{M}/G$ into a smooth submersion [Lee, 2012, Theorem 21.10]. The vertical distribution is G -equivariant, i.e. $\mathcal{V}_{a \cdot x} = a \cdot \mathcal{V}_x$ for all $a \in G$ (where \cdot is the group action on \mathcal{M} and $T\mathcal{M}$). Given a G -invariant Riemannian metric g on \mathcal{M} , the horizontal distribution is G -equivariant and the metric descends to a metric g' on \mathcal{M}/G defined by $g'_{\pi(x)}(X, X) = g_x(X_x^\#, X_x^\#)$.

7.2.2 Definition of quotient-affine metrics

Applying this to $\mathcal{M} = \text{Sym}^+(n)$, $G = \text{Diag}^+(n)$ and $\mathcal{M}/G \simeq \text{Cor}^+(n)$, the submersion $\text{Cor} : \Sigma \in \text{Sym}^+(n) \mapsto \text{Diag}(\Sigma)^{-1/2} \Sigma \text{Diag}(\Sigma)^{-1/2} \in \text{Cor}^+(n)$ [David and Gu, 2019, David, 2019] allows to descend any $\text{Diag}^+(n)$ -invariant Riemannian metric on $\text{Sym}^+(n)$ to a Riemannian metric on $\text{Cor}^+(n)$ [O'Neill, 1966].

A natural example of $\text{Diag}^+(n)$ -invariant Riemannian metric on SPD matrices is provided by the affine-invariant metric defined for all $\Sigma \in \text{Sym}^+(n)$ and $V \in T_\Sigma \text{Sym}^+(n) \simeq \text{Sym}(n)$ by:

$$g_\Sigma^{\text{AI}(\alpha, \beta)}(V, V) = \alpha \text{tr}(\Sigma^{-1}V\Sigma^{-1}V) + \beta \text{tr}(\Sigma^{-1}V)^2 \quad (7.11)$$

where $\alpha > 0$ and $\beta > -\frac{\alpha}{n}$. Its sectional curvature is for $V, W \in \text{Sym}(n)$ [Skovgaard, 1984, Thanwerdas and Penec, 2022b]:

$$\kappa_\Sigma^{\text{AI}(\alpha, \beta)}(V, W) = \frac{1}{4\alpha} \text{tr}((\Sigma^{-1}V\Sigma^{-1}W - \Sigma^{-1}W\Sigma^{-1}V)^2) \in \left[-\frac{1}{2\alpha}; 0 \right] \quad (7.12)$$

The quotient-affine metric is the quotient of the affine-invariant metric via the submersion $\text{Cor} : \text{Sym}^+(n) \rightarrow \text{Cor}^+(n)$. It does not depend on β and it writes αg^{QA} with, for all

$C \in \text{Cor}^+(n)$ and $X \in T_C \text{Cor}^+(n) \simeq \text{Hol}(n)$:

$$g_C^{\text{QA}}(X, X) = \text{tr}((C^{-1}X)^2) - 2\mathbf{1}^\top \text{Diag}(C^{-1}X)(I_n + C \bullet C^{-1})^{-1} \text{Diag}(C^{-1}X)\mathbf{1}. \quad (7.13)$$

Note that it is invariant under permutations.

The vertical/horizontal distributions/projections and the horizontal lift are for all $\Sigma \in \text{Sym}^+(n)$, $V \in T_\Sigma \text{Sym}^+(n) \simeq \text{Sym}(n)$ and $X \in T_{\text{Cor}(\Sigma)} \text{Cor}^+(n) \simeq \text{Hol}(n)$ [Thanwerdas and Pennec, 2021]:

- $\mathcal{V}_\Sigma = \{D\Sigma + \Sigma D \mid D \in \text{Diag}(n)\}$,
- $\mathcal{H}_\Sigma = \{V \in \text{Sym}(n) \mid \Sigma^{-1}V + V\Sigma^{-1} \in \text{Hol}(n)\} = \mathcal{S}_{\Sigma^{-1}}(\text{Hol}(n))$,
- $\text{ver}_\Sigma(V) = D\Sigma + \Sigma D \in \mathcal{V}_\Sigma$ with $D = \text{diag}((I_n + \Sigma \bullet \Sigma^{-1})^{-1} \text{Diag}(\Sigma^{-1}V)\mathbf{1})$,
- $\text{hor}_\Sigma(V) = V - \text{ver}_\Sigma(V) \in \mathcal{H}_\Sigma$,
- $X_\Sigma^\# = \text{hor}_\Sigma(\text{Diag}(\Sigma)^{1/2} X \text{Diag}(\Sigma)^{1/2}) \in \mathcal{H}_\Sigma$,

where $\mathcal{S}_A(X)$ is the unique solution of the Sylvester equation $A\mathcal{S}_A(X) + \mathcal{S}_A(X)A = X$.

These operations allow to write the sectional curvature for all $C \in \text{Cor}^+(n)$ and $X, Y \in T_C \text{Cor}^+(n) \simeq \text{Hol}(n)$ [Thanwerdas and Pennec, 2021]:

$$\kappa_C^{\text{QA}}(X, Y) = \underbrace{\kappa_C^{\text{AI}}(X_C^\#, Y_C^\#)}_{\in [-\frac{1}{2}; 0]} + \underbrace{\frac{3}{8} \frac{\mu^\top (I_n + C \bullet C^{-1})^{-1} \mu}{g_C^{\text{QA}}(X, X)g_C^{\text{QA}}(Y, Y) - g_C^{\text{QA}}(X, Y)^2}}_{\geq 0}, \quad (7.14)$$

with $\mu \in \mathbb{R}^n$ defined by $\mu = [D(X, Y) - D(Y, X)]\mathbf{1} \in \mathbb{R}^n$ and $D(X, Y) \in \text{Diag}(n)$ defined by $D(X, Y) = \text{Diag}([C^{-1} \text{Diag}(X_C^\#)C, C^{-1}Y_C^\#])$, where $[A, B] = AB - BA$ is the commutator of squared matrices.

7.2.3 Complement: bounds of curvature

Theorem 7.1 (Bounds of curvature) The sectional curvature of the quotient-affine metric takes positive and negative values. It is bounded from below and unbounded from above.

See complete proof of Theorem 7.1 in Section 11.6.

Sketch of the proof. First of all, $\kappa_C(X, Y) \geq \kappa_C^{\text{AI}}(X, Y) \geq -\frac{1}{2}$ so the curvature is bounded from below. Second, at $C = I_n$, $X^\# = X$ and $Y^\# = Y$ so $\text{Diag}(X^\#) = \text{Diag}(Y^\#) = 0$ and $\mu = 0$. Hence, $\kappa_{I_n}(X, Y) = \kappa_{I_n}^{\text{AI}}(X, Y) \leq 0$ and for example $\kappa_{I_n}(E_{ij}, E_{ik}) = -\frac{1}{8} < 0$ with $i \neq j \neq k \neq i \in \{1, \dots, n\}$ [Thanwerdas and Pennec, 2022b]. So the curvature takes negative values. Third, let $X = \mathbf{1}\mathbf{1}^\top - I_n$ and $Y = \mu\mathbf{1}^\top + \mathbf{1}\mu^\top - 2\text{diag}(\mu)$ with $\text{sum}(\mu) = \mathbf{1}^\top \mu = 0$ where $\mu \in \mathbb{R}^n$. Let $C = (1 - \rho)I_n + \rho\mathbf{1}\mathbf{1}^\top \in \text{Cor}^+(n)$ for $\rho \in (-\frac{1}{n-1}, 1)$. We show in Section 11.6 that $\kappa_C(X, Y)$ tends to $+\infty$ when $\rho \rightarrow -\frac{1}{n-1}$, which proves that the curvature takes positive values and it is not bounded from above. \square

Hence, the Riemannian manifold of full-rank correlation matrices endowed with the quotient-affine metric is geodesically complete but it is not a $\text{CAT}(k)$ space for any $k \in \mathbb{R}$. In particular, it is not a Hadamard space as one could have hoped. It may be a problem for

many algorithms because the Riemannian logarithm and the Fréchet mean are not ensured to be unique.

In Section 7.3, we relax the invariance under permutations that might be unsuitable in some contexts. In Section 7.4, we define non-permutation-invariant metrics which in addition bring uniqueness of the mean.

7.3 Generalization of quotient-affine metrics

In this section, we define a family of quotient metrics which generalize quotient-affine metrics and which are not invariant by permutations. In Section 7.3.1, we start by giving an overview of the congruence action of several subgroups of the general linear group $GL(n)$ on SPD matrices. In particular, we show that affine-invariant metrics are exactly $(\mathfrak{S}(n) \times LT^+(n))$ -invariant metrics so the family of $LT^+(n)$ -invariant metrics appears as a natural generalization of affine-invariant metrics. In Section 7.3.2, we show that $LT^+(n)$ -invariant metrics are exactly pullbacks by the Cholesky map of left-invariant metrics on the Lie group $LT^+(n)$. We call them Lie-Cholesky metrics. In particular, we define a Lie group structure on SPD matrices such that the affine-invariant metric is a left-invariant metric on that Lie group. This is an unexpected result as the SPD cone endowed with an affine-invariant metric is always seen as a Riemannian homogeneous (symmetric) space. In Section 7.3.3, we define quotient-Lie-Cholesky metrics and we show that we can compute numerically the exponential map, the logarithm map and the Riemannian distance.

7.3.1 Congruence actions of matrix Lie groups on SPD matrices

The action of congruence of the real general linear group $GL(n)$ on SPD matrices is:

$$\begin{cases} GL(n) \times \text{Sym}^+(n) & \longrightarrow & \text{Sym}^+(n) \\ (A, \Sigma) & \longmapsto & A\Sigma A^\top \end{cases} \quad (7.15)$$

In the following theorem, we focus on the subactions given by the following subgroups.

1. The group of matrices with positive determinant $GL^+(n)$.
2. The special linear group $SL(n) = \{A \in GL(n) | \det(A) = 1\} \subset GL^+(n)$, which is interesting for the invariance of covariance matrices under volume-preserving linear transformations of the feature vector.
3. The orthogonal group $O(n) = \{R \in GL(n) | RR^\top = I_n\}$ and the special orthogonal group $SO(n) = O(n) \cap SL(n)$, which are interesting for the invariance of covariance matrices under rotations and symmetries.
4. The group of lower triangular matrices with positive diagonal $LT^+(n)$, which is interesting because an $LT^+(n)$ -invariant metric on SPD matrices is a left-invariant metric on a Lie group (cf. Section 7.3.2).
5. The positive diagonal group $\text{Diag}^+(n) \subset LT^+(n)$, which is interesting for the invariance of covariance matrices under scalings on each variable.

6. The group of positive real numbers \mathbb{R}^+ which injects itself into the general linear group via the map $\lambda \in \mathbb{R}^+ \mapsto \lambda I_n \in \text{GL}^+(n)$. It is interesting for the invariance of covariance matrices under global scaling.
7. The permutation group $\mathfrak{S}(n)$ which injects itself into the orthogonal group via the map $\sigma \in \mathfrak{S}(n) \mapsto P_\sigma \in \text{O}(n)$ defined by $[P_\sigma]_{ij} = \delta_{i,\sigma(j)}$ for all $i, j \in \{1, \dots, n\}$. It is interesting for the invariance of covariance matrices under permutation of the axes.

In the proof, we also use the following notations: \mathbb{R}^* is the group of invertible scalar matrices, $\text{Diag}^*(n)$ is the group of invertible diagonal matrices, $\text{LT}^*(n)$ is the group of invertible lower triangular matrices, $\text{UT}^+(n)$ is the group of upper triangular matrices with positive diagonal.

Theorem 7.2 (Characterization of affine-invariant metrics) Let g be a Riemannian metric on SPD matrices. The following statements are equivalent:

1. g is $\text{GL}(n)$ -invariant,
2. g is $\text{GL}^+(n)$ -invariant,
3. g is $\text{SL}(n)$ -invariant and \mathbb{R}^+ -invariant,
4. g is $\text{SO}(n)$ -invariant and $\text{Diag}^+(n)$ -invariant,
5. g is $\mathfrak{S}(n)$ -invariant and $\text{LT}^+(n)$ -invariant.

See the proof of Theorem 7.2 in Section 11.6.

One interpretation of this result is that these pairs of invariance ($\text{SL}(n)$ and \mathbb{R}^+ ; $\text{SO}(n)$ and $\text{Diag}^+(n)$; $\mathfrak{S}(n)$ and $\text{LT}^+(n)$) are incompatible except in the affine-invariant metrics. In other words, the family of $\text{LT}^+(n)$ -invariant metrics is a natural non-permutation-invariant extension of the family of affine-invariant metrics. In the next section, we characterize them as pullback metrics of left-invariant metrics on the Lie group $\text{LT}^+(n)$ by the Cholesky map.

7.3.2 Lie-Cholesky metrics

7.3.2.1 $\text{LT}^+(n)$ -invariant metrics are Lie-Cholesky metrics

The manifold $\text{LT}^+(n)$ is an open set of the vector space $\text{LT}(n)$. It is a Lie group where the internal law is the matrix multiplication.

Definition 7.3 (Lie-Cholesky metrics on $\text{Sym}^+(n)$) A Lie-Cholesky metric on $\text{Sym}^+(n)$ is the pullback of a left-invariant metric on the Lie group $(\text{LT}^+(n), \times)$ via the Cholesky diffeomorphism. It is geodesically complete.

Theorem 7.4 (Lie-Cholesky are $\text{LT}^+(n)$ -invariant metrics) A Riemannian metric on $\text{Sym}^+(n)$ is a Lie-Cholesky metric if and only if it is $\text{LT}^+(n)$ -invariant.

See the proof of Theorem 7.4 in Section 11.6.

7.3.2.2 Consequences

Corollary 7.5 The Riemannian metrics on $\text{Sym}^+(n)$ which are $\text{LT}^+(n)$ -invariant are geodesically complete.

Corollary 7.6 Permutation-invariant Lie-Cholesky metrics are affine-invariant metrics.

Corollary 7.7 (Affine-invariant metrics on SPD matrices are left-invariant metrics on a Lie group) The manifold $\text{Sym}^+(n)$ has a structure of Lie group given by the matrix multiplication of the Cholesky factors, i.e. for all $\Sigma, \Sigma' \in \text{Sym}^+(n)$, denoting $L = \text{Chol}(\Sigma)$ and $L' = \text{Chol}(\Sigma')$, the internal law is:

$$\Sigma \star \Sigma' = (LL')(LL')^\top = L\Sigma'L^\top. \quad (7.16)$$

The affine-invariant metrics $g_\Sigma(X, X) = \alpha \text{tr}(\Sigma^{-1}X\Sigma^{-1}X) + \beta \text{tr}(\Sigma^{-1}X)^2$ are left-invariant metrics for this Lie group structure. They are pullbacks of the left-invariant metrics on the Lie group $\text{LT}^+(n)$ characterized by the following inner products at $I_n \in \text{LT}^+(n)$, where $Z \in T_{I_n}\text{LT}^+(n) \simeq \text{LT}(n)$:

$$\begin{aligned} \langle Z|Z \rangle_{\alpha, \beta} &= g_{I_n}(Z + Z^\top, Z + Z^\top) \\ &= \alpha \text{tr}((Z + Z^\top)(Z + Z^\top)) + \beta \text{tr}(Z + Z^\top)^2 \\ &= 2\alpha \text{tr}(ZZ^\top) + 2\alpha \text{tr}(\text{Diag}(Z)^2) + 4\beta \text{tr}(Z)^2. \end{aligned} \quad (7.17)$$

We denote $\langle \cdot | \cdot \rangle = \langle \cdot | \cdot \rangle_{1/2, 0}$, i.e. $\langle Z|Z' \rangle = \text{tr}(ZZ^\top) + \text{tr}(\text{Diag}(Z)^2)$ for $Z \in \text{LT}(n)$.

The left-invariant metrics on $\text{LT}^+(n)$ are parametrized by the inner products at I_n . Hence, they can be parametrized by self-adjoint positive definite linear maps $f : \text{LT}(n) \rightarrow \text{LT}(n)$ as follows, for all $Z, Z' \in \text{LT}(n)$:

$$Z \cdot Z' = \langle f(Z)|Z' \rangle = \langle Z|f(Z') \rangle. \quad (7.18)$$

In other words, if $\text{vec} : \text{LT}(n) \rightarrow \mathbb{R}^{n(n+1)/2}$ denotes the linear map defined by $\text{vec}(Z) = (Z_{11}, Z_{21}, Z_{22}, \dots, Z_{n, n-1}, Z_{nn})^\top$ for all $Z \in \text{LT}(n)$, then there exists an SPD matrix $A \in \text{Sym}^+(\frac{n(n+1)}{2})$ such that $Z \cdot Z' = \text{vec}(Z)^\top A \text{vec}(Z')$. Thus the Lie-Cholesky metrics are parametrized by self-adjoint positive definite linear maps $f : \text{LT}(n) \rightarrow \text{LT}(n)$ or SPD matrices $A \in \text{Sym}^+(\frac{n(n+1)}{2})$. They are denoted $\text{LC}(f)$ or $\text{LC}(A)$.

Therefore, we can see the family of Lie-Cholesky metrics as a natural non-permutation-invariant extension of affine-invariant metrics. Since they are left-invariant metrics on a Lie group, the numerical computation of the geodesics (exponential map, logarithm map) and the parallel transport is simpler and much more stable and precise than for other metrics [Guigui and Pennec, 2021b]. The curvature is known in closed form modulo the computation of the operator $\text{ad}^* : \text{LT}(n) \times \text{LT}(n) \rightarrow \text{LT}(n)$ defined for all $X, Y, Z \in \text{LT}(n)$ by $\langle \text{ad}^*(X)(Y)|Z \rangle = \langle Y|[X, Z] \rangle$ [Besse, 1987, Theorem 7.30]. The operator ad^* for Lie-Cholesky metrics is given in the following lemma.

Lemma 7.8 (Operator ad^* for Lie-Cholesky metrics) Let $\text{LC}(A)$ be a Lie-Cholesky metric characterized by $A \in \text{Sym}^+(\frac{n(n+1)}{2})$. Then for $X, Y \in \text{LT}(n)$, $\text{vec}(\text{ad}^*(X)(Y)) = A(I_n \otimes X^\top - X \otimes I_n)A^{-1}\text{vec}(Y)$.

See the proof of Lemma 7.8 in Section 11.6.

7.3.3 Quotient-Lie-Cholesky metrics

Lie-Cholesky metrics being $LT^+(n)$ -invariant, they are in particular $\text{Diag}^+(n)$ -invariant as the affine-invariant metric. As recalled earlier, the smooth map $\text{Cor} : \text{Sym}^+(n) \rightarrow \text{Cor}^+(n)$ is a $\text{Diag}^+(n)$ -invariant submersion so Lie-Cholesky metrics descend to quotient-Lie-Cholesky metrics on full-rank correlation matrices. Quotient-Lie-Cholesky metrics naturally extend quotient-affine metrics. To use them, one needs to compute the vertical and horizontal projections. After the definition of quotient-Lie-Cholesky metrics, the following lemma gives the expression of the vertical and horizontal distributions.

Definition 7.9 (Quotient-Lie-Cholesky metric) A quotient-Lie-Cholesky metric is the quotient metric of a Lie-Cholesky metric by the submersion $\text{Cor} : \text{Sym}^+(n) \rightarrow \text{Cor}^+(n)$. Equivalently, it is the pushforward metric by the diffeomorphism $\Phi : L \in LT^1(n) \mapsto \text{Diag}(LL^\top)^{-1/2}LL^\top\text{Diag}(LL^\top)^{-1/2} \in \text{Cor}^+(n)$ of a quotient metric on $LT^1(n) \simeq \text{Diag}(n) \backslash LT^+(n)$ defined by quotient via the submersion $\pi : L \in LT^+(n) \mapsto \text{Diag}(L)^{-1}L \in LT^1(n)$.

In this section, we express all the Riemannian operations of quotient-Lie-Cholesky metrics $LT^1(n)$. It suffices to push them forward by Φ to get them on $\text{Cor}^+(n)$.

Lemma 7.10 (Vertical distribution, horizontal distribution) The vertical and horizontal distributions associated to the quotient-Lie-Cholesky associated to f are $\mathcal{V}_L = \text{Diag}(n)L$ and $\mathcal{H}_L^{\text{LC}(f)} = Lf^{-1}(L^{-1}d_{LL^\top}\text{Chol}(\mathcal{S}_{(LL^\top)^{-1}}(\text{Hol}(n))))$.
See the proof of Lemma 7.10 in Section 11.6.

Then, to compute the vertical and horizontal projections, it suffices to take bases of the vertical space \mathcal{V}_L and the horizontal space \mathcal{H}_L , to orthonormalize them by Gram-Schmidt process and to project onto these orthonormal bases. Then the horizontal lift can be computed as follows, which allows to compute the metric and the exponential map.

Lemma 7.11 (Horizontal lift, Riemannian metric, exponential map) Given the horizontal projection $\text{hor}_L : T_L LT^+(n) \rightarrow \mathcal{H}_L$ and the exponential map of the Lie-Cholesky metric $\text{Exp}^{\text{LC}} : LT(n) \rightarrow LT^+(n)$, we have for all $L \in LT^+(n)$, for all $\Gamma = \text{Diag}(L)^{-1}L \in LT^1(n)$, for all $\xi \in T_\Gamma LT^1(n) \simeq LT^0(n)$:

- (Horizontal lift) $\xi^\# = \text{hor}_L(\text{Diag}(L)\xi)$,
- (Riemannian metric) $g_\Gamma^{\text{QLC}}(\xi, \xi) = \langle f(\Gamma^{-1}\text{hor}_\Gamma(\xi)) | \Gamma^{-1}\text{hor}_\Gamma(\xi) \rangle$,
- (Exponential map) $\text{Exp}_\Gamma^{\text{QLC}}(t\xi) = \text{Diag}(\text{Exp}_\Gamma^{\text{LC}}(t\text{hor}(\xi)))^{-1}\text{Exp}_\Gamma^{\text{LC}}(t\text{hor}(\xi))$.

In particular, quotient-Lie-Cholesky metrics are geodesically complete.
See the proof of Lemma 7.11 in Section 11.6.

The Riemannian logarithm and the Riemannian distance can then be computed by minimizing the Lie-Cholesky distance along a fiber. The Lie-Cholesky distance is itself computed numerically with efficient tools on Lie groups [Guigui and Pennec, 2021b].

In this section, we gave an overview of the congruence action of several matrix Lie groups on SPD matrices. It allowed to understand that the natural extension of affine-invariant

metrics to non-permutation-invariant metrics is the family of $\text{LT}^+(n)$ -invariant metrics. We showed that such metrics are pullbacks of left-invariant metrics on the Lie group $\text{LT}^+(n)$ by the Cholesky diffeomorphism. Hence, the space of lower triangular matrices and the Cholesky map naturally appear when one wants to get rid of the invariance under permutations on SPD matrices. Following the same construction as for the quotient-affine metric, we built quotient-Lie-Cholesky metrics and we showed that most of the interesting Riemannian operations can be computed numerically. However, this geometry is not completely satisfying since we have no formula in closed form and no obvious nice geometric properties except geodesic completeness. Moreover, the numerical computation of the Riemannian logarithm and the Riemannian distance are done with the generic methods in quotient manifolds which can be heavy and unstable. Furthermore, the quotient-affine metric has unbounded curvature as shown in Section 7.2 and it is difficult to say something about the curvature of general quotient-Lie-Cholesky metrics.

Hence in the following section, we continue to rely on lower triangular matrices and the Cholesky map to define Riemannian metrics with simpler geometries than the general quotient geometry of quotient-Lie-Cholesky metrics.

7.4 New geometric structures with unique mean

In this section, we define new families of metrics on the open ellipsope of full-rank correlation matrices. They are based on simple geometries of subspaces of lower triangular matrices and transported to the ellipsope via the Cholesky map or the derived map Θ . In Section 7.4.1, we introduce the poly-hyperbolic-Cholesky metrics which provide Riemannian symmetric structures. Then we introduce the Euclidean-Cholesky metrics (Section 7.4.2) and the log-Euclidean-Cholesky metrics (Section 7.4.3) that provide Euclidean structures. Equipped with a metric of one of these three families, the open ellipsope is Hadamard so the Riemannian logarithm and the Fréchet mean are unique. In Section 7.4.4, we introduce a Lie group structure that allows to define Cartan-Schouten affine connections and left-invariant metrics. The group mean of the canonical Cartan-Schouten connection is unique. In Section 7.4.5, we give the geodesics in dimension 2, which correspond to interpolations of one correlation coefficient.

7.4.1 Symmetric space: poly-hyperbolic-Cholesky metrics

In this section, we use the diffeomorphism $\text{Chol}_{|\text{Cor}^+(n)} : \text{Cor}^+(n) \longrightarrow \mathcal{L}$ where \mathcal{L} is the set of lower triangular matrices with positive diagonal such that each row is unit normed for the canonical Euclidean norm. Thus, the k -th row of $L = \text{Chol}(C) \in \mathcal{L}$ writes $(L_{k1}, \dots, L_{k,k-1}, L_{kk}, 0, \dots, 0)$ with $L_{kk} > 0$. It belongs to the open hemisphere $\text{HS}^{k-1} = \{x \in \mathbb{R}^k \mid \|x\| = 1 \text{ and } x_k > 0\}$. So each $L \in \mathcal{L}$ is a point in $\text{HS}^0 \times \dots \times \text{HS}^{n-1}$. This construction is clearly bijective and diffeomorphic. Note that since $L_{11} = 1$ and $\text{HS}^0 = \{1\}$, we can remove it from the Cartesian product. Hence, we can define the diffeomorphism $\Psi : \mathcal{L} \longrightarrow \text{HS}^1 \times \dots \times \text{HS}^{n-1}$. Moreover, an open hemisphere is one of the avatars of the hyperbolic space, which is the Riemannian manifold of negative constant curvature.

Before introducing the Riemannian metric induced by the diffeomorphism Ψ , we recall the definition and Riemannian operations of the hyperbolic space in the model of the hyperboloid.

Theorem 7.12 (Hyperbolic geometry of the hyperboloid) The vector space \mathbb{R}^{k+1} is endowed with the non-degenerate quadratic form $Q(x) = \sum_{i=1}^k x_i^2 - x_{k+1}^2$. We denote $(x, y) \mapsto Q(x, y)$ the associated symmetric bilinear form. The hyperboloid \mathbb{H}^k is the Riemannian manifold defined by $\mathbb{H}^k = \{x \in \mathbb{R}^{k+1} | Q(x) = -1\}$ endowed with the induced pseudo-metric, which is a Riemannian metric on \mathbb{H}^k . The tangent space writes $T_x \mathbb{H}^k = \{v \in \mathbb{R}^{k+1} | Q(x, v) = 0\}$. For all $x, y \in \mathbb{H}^k$, $v \in T_x \mathbb{H}^k$ and $w \in T_x \mathbb{H}^k$ non colinear to v , the Riemannian operations are:

- (Riemannian metric) $g_x^{\mathbb{H}}(v, v) = Q(v)$,
- (Riemannian norm) $\|v\|_{\mathbb{H}} = \sqrt{Q(v)}$,
- (Riemannian distance) $d_{\mathbb{H}}(x, y) = \operatorname{arccosh}(-Q(x, y))$,
- (Exponential map) $\operatorname{Exp}_x^{\mathbb{H}}(v) = \cosh(\|v\|_{\mathbb{H}})x + \sinh(\|v\|_{\mathbb{H}}) \frac{v}{\|v\|_{\mathbb{H}}}$,
- (Logarithm map) $\operatorname{Log}_x^{\mathbb{H}}(y) = d^{\mathbb{H}}(x, y) \frac{y + Q(x, y)x}{\|y + Q(x, y)x\|_{\mathbb{H}}}$,
- (Sectional curvature) $\kappa_x^{\mathbb{H}}(v, w) = -1$.

The formulae on the other models of the hyperbolic space can be obtained by pullback via the appropriate diffeomorphism. In particular, the diffeomorphism between the open hemisphere $\operatorname{HS}^k = \{(x_1, \dots, x_{k+1}) \in \mathbb{R}^k \times \mathbb{R}^+ | \sum_{i=1}^{k+1} x_i^2 = 1\}$ and the hyperbolic space \mathbb{H}^k is:

$$\varphi^{\operatorname{SH}} : (x_1, \dots, x_{k+1}) \in \operatorname{HS}^k \mapsto \frac{1}{x_{k+1}}(x_1, \dots, x_k, 1) \in \mathbb{H}^k. \quad (7.19)$$

Thus the natural metric on the open hemisphere is the pullback metric $g^{\operatorname{HS}} = (\varphi^{\operatorname{SH}})^* g^{\mathbb{H}}$.

Definition 7.13 (Poly-hyperbolic-Cholesky metrics) Let $\alpha_1, \dots, \alpha_{n-1} > 0$ be positive coefficients. A poly-hyperbolic-Cholesky metric on $\operatorname{Cor}^+(n)$ is the pullback metric $g^{\operatorname{PHC}} = (\Psi \circ \operatorname{Chol})^*(\alpha_1 g^{\operatorname{HS}^1} \oplus \dots \oplus \alpha_{n-1} g^{\operatorname{HS}^{n-1}})$ by the map $\operatorname{Chol} \circ \Psi$ of a weighted product metric on the product of hyperbolic spaces $\operatorname{HS}^1 \times \dots \times \operatorname{HS}^{n-1}$. The PHC metric with all weights equal to 1 is called the canonical PHC metric.

Theorem 7.14 (Symmetric space structure) The manifold of full-rank correlation matrices $\operatorname{Cor}^+(n)$ equipped with a poly-hyperbolic-Cholesky metric is a Riemannian symmetric space of non-positive sectional curvature bounded by $[a, 0]$ with $a = -\frac{1}{\min_{i \geq 2} \alpha_i}$. For $n \geq 3$, it is not of constant curvature. The canonical PHC metric writes for all $C \in \operatorname{Cor}^+(n)$ and $X \in T_C \operatorname{Cor}^+(n) \simeq \operatorname{Hol}(n)$:

$$g_C^{\operatorname{CPHC}}(X, X) = \|\operatorname{Diag}(L)^{-1} L \operatorname{Low}_S(L^{-1} X L^{-\top})\|^2, \quad (7.20)$$

where $L = \operatorname{Chol}(C) \in \mathcal{L}$. The square distance between C and $C' = \phi(L')$ writes:

$$d_{\operatorname{CPHC}}(C, C')^2 = \sum_{i=2}^n \operatorname{arccosh}(-Q(L_i^\top, L'_i{}^\top))^2, \quad (7.21)$$

where L_i, L'_i are the i -th rows of L, L' respectively.

See the proof of Theorem 7.14 in Section 11.6.

7.4.2 Vector space: Euclidean-Cholesky metrics

The Cholesky map was already used on SPD matrices [Wang et al., 2004]. The Euclidean-Cholesky metric is defined as the Euclidean metric on the Cholesky factor belonging to $\text{LT}^+(n)$. However, it is not complete because $\text{LT}^+(n)$ is open in $\text{LT}(n)$. A solution to this problem is to take the logarithm of the diagonal before taking the Euclidean metric [Pinheiro and Bates, 1996, Lin, 2019]. It amounts to define the product metric of a Euclidean metric on the strictly lower part and a log-Euclidean metric on the diagonal part.

For correlation matrices, we can use the diffeomorphism $\Theta : \text{Cor}^+(n) \rightarrow \text{LT}^1(n)$. Since the diagonal is I_n , the two mentioned metrics reduce to the same metric on the open ellipsope. We call it the Euclidean-Cholesky metric.

Definition 7.15 (Euclidean-Cholesky metrics) The Euclidean-Cholesky metrics on full-rank correlation matrices are pullback metrics by $\Theta : \text{Cor}^+(n) \rightarrow \text{LT}^1(n)$ of inner products on $\text{LT}^1(n) = I_n + \text{LT}^0(n)$.

Theorem 7.16 (Riemannian operations) Let $\|\cdot\|$ be a Euclidean norm on $\text{LT}^0(n)$. The Riemannian operations of the Euclidean-Cholesky metric associated to this norm are, for all $C, C', C_i \in \text{Cor}^+(n)$, $X \in T_C \text{Cor}^+(n) \simeq \text{Hol}(n)$, $t \in \mathbb{R}$:

- (Exponential map) $\text{Exp}_C(tX) = \Theta^{-1}(\Theta(C) + t d_C \Theta(X))$,
- (Logarithm map) $\text{Log}_C(C') = (d_C \Theta)^{-1}(\Theta(C') - \Theta(C))$,
- (Geodesic) $\gamma_{C \rightarrow C'}(t) = \Theta^{-1}((1-t)\Theta(C) + t\Theta(C'))$,
- (Distance) $d(C, C') = \|\Theta(C') - \Theta(C)\|$,
- (Parallel transport) $\Pi_{C \rightarrow C'} X = (d_{C'} \Theta)^{-1}(d_C \Theta(X))$,
- (Curvature) Null,
- (Fréchet mean) $\bar{C} = \Theta^{-1}(\frac{1}{n} \sum_{i=1}^n \Theta(C_i))$,

where $d_C \Theta(X) = \Theta(C) \text{Low}_S(L^{-1} X L^{-\top}) - \frac{1}{2} \text{Diag}(L^{-1} X L^{-\top}) \Theta(C)$ and $L = \text{Chol}(C)$. The Euclidean-Cholesky metrics are geodesically complete.

These metrics are flat, geodesically complete and the Riemannian operations are trivial. Since they reduce to (the pullback of) an inner product on a *vector space*, we prefer not to use the term *Lie group* for them, contrarily to the terminology of [Li et al., 2017, Lin, 2019]. We prefer to reserve it for Lie groups that are not vector spaces, such as the natural Lie group structure of $\text{LT}^+(n)$ (with matrix multiplication) underlying Lie-Cholesky metrics.

7.4.3 Vector space: log-Euclidean-Cholesky metrics

Another map was used to Euclideanize the manifold $\text{LT}^+(n)$: the matrix logarithm [Li et al., 2017]. Indeed, the matrix exponential is a smooth diffeomorphism from $\text{LT}(n)$ to $\text{LT}^+(n)$ [Gallier, 2008]. We can use the same idea for correlation matrices since the matrix exponential is a smooth diffeomorphism from $\text{LT}^0(n)$ to $\text{LT}^1(n)$. Moreover it has a particularly simple expression:

$$\exp(\xi) = \sum_{k=0}^{n-1} \frac{1}{k!} \xi^k, \quad (7.22)$$

because $LT^0(n)$ is a nilpotent algebra. Then the logarithm $\log : LT^1(n) \rightarrow LT^0(n)$ is simply:

$$\log(Z) = \sum_{k=1}^{n-1} \frac{(-1)^{k-1}}{k} (Z - I_n)^k. \quad (7.23)$$

Therefore, the differential of the logarithm writes:

$$d_Z \log(\xi) = \sum_{k=1}^{n-1} \frac{(-1)^k}{k} [(Z - I_n)^{k-1} \xi + (Z - I_n)^{k-2} \xi (Z - I_n) + \dots + \xi (Z - I_n)^{k-1}]. \quad (7.24)$$

Definition 7.17 (Log-Euclidean-Cholesky metrics) The log-Euclidean-Cholesky metrics on full-rank correlation matrices are pullback metrics by $\log \circ \Theta : \text{Cor}^+(n) \rightarrow LT^0(n)$ of inner products on $LT^0(n)$.

Theorem 7.18 (Riemannian operations) Let $\|\cdot\|$ be a Euclidean norm on $LT^0(n)$. The Riemannian operations of log-Euclidean-Cholesky metrics associated to this norm are, for all $C, C', C_i \in \text{Cor}^+(n)$, $X \in T_C \text{Cor}^+(n) \simeq \text{Hol}(n)$, $t \in \mathbb{R}$:

- (Exponential map) $\text{Exp}_C(tX) = \Theta^{-1} \circ \exp(\log(\Theta(C)) + t d_C(\log \circ \Theta)(X))$,
- (Logarithm map) $\text{Log}_C(C') = (d_C(\log \circ \Theta))^{-1}(\log(\Theta(C')) - \log(\Theta(C)))$,
- (Geodesic) $\gamma_{C \rightarrow C'}(t) = \Theta^{-1} \circ \exp((1-t)\log(\Theta(C)) + t\log(\Theta(C')))$,
- (Distance) $d(C, C') = \|\log(\Theta(C')) - \log(\Theta(C))\|$,
- (Parallel transport) $\Pi_{C \rightarrow C'} X = (d_{C'}(\log \circ \Theta))^{-1}(d_C \log \circ \Theta(X))$,
- (Curvature) Null,
- (Fréchet mean) $\bar{C} = \Theta^{-1} \circ \exp(\frac{1}{n} \sum_{i=1}^n \log(\Theta(C_i)))$,

where $d_C(\log \circ \Theta)(X) = d_{\Theta(C)} \log(d_C \Theta(X))$. The log-Euclidean-Cholesky metrics are geodesically complete.

The computation of the Riemannian operations of the Euclidean-Cholesky metrics is more straightforward than those of the log-Euclidean-Cholesky because one has to compute the differential of the triangular matrix logarithm for the latter.

7.4.4 Nilpotent Lie group structure

Another interesting structure is given by the natural Lie group structure of $LT^1(n)$ for the matrix multiplication. This equips full-rank correlation matrices with a Lie group structure via the diffeomorphism $\Theta : \text{Cor}^+(n) \rightarrow LT^1(n)$. Hence, left-invariant metrics can be defined. In analogy with $\text{Sym}^+(n) \simeq LT^+(n)$, they can also be called Lie-Cholesky metrics. Then all Riemannian operations can be computed numerically [Guigui and Pennec, 2021b] and the space is ensured to be geodesically complete. However, this doesn't give information on the sign of the curvature.

More interestingly, one can rely on the canonical Cartan-Schouten connection to define the group exponential and the notion of group mean. We can also name them after Lie-Cholesky.

Theorem 7.19 (Group operations) The group operations associated to the Lie-Cholesky group structure on full-rank correlation matrices are, for all $C, C', C_i \in \text{Cor}^+(n)$, $X \in T_C \text{Cor}^+(n) \simeq \text{Hol}(n)$, $t \in \mathbb{R}$:

- (Exponential map) $\text{Exp}_C(tX) = \Theta^{-1}(\Theta(C) \exp(t \Theta(C)^{-1} d_C \Theta(X)))$,
- (Logarithm map) $\text{Log}_C(C') = (d_C \Theta)^{-1}(\Theta(C) \log(\Theta(C)^{-1} \Theta(C')))$,
- (Geodesic) $\gamma_{C \rightarrow C'}(t) = \Theta^{-1}(\Theta(C)(\Theta(C)^{-1} \Theta(C'))^t)$,
- (Group mean) Unique, characterized by $\sum_{i=1}^k \log(\Theta(\bar{C})^{-1} \Theta(C_i)) = 0$.

See the proof of Theorem 7.19 in Section 11.6.

7.4.5 Explicit geodesics in dimension 2

In dimension 2, the ellipsope is reduced to one parameter. All full-rank correlation matrices write $C = C(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $\rho \in (-1, 1)$. Therefore, the quotient-affine metric and the metrics defined in Section 7.4 only depend on one scaling parameter. They actually split in two groups and the geodesics can be computed in closed forms. The two formulae in the following result provide two different interpolations of the correlation coefficient. The proof is in the supplementary material.

Theorem 7.20 (Geodesics in dimension 2) Let $C_1 = C(\rho_1), C_2 = C(\rho_2) \in \text{Cor}^+(2)$ with $\rho_1, \rho_2 \in (-1, 1)$.

1. Quotient-affine metrics and poly-hyperbolic-Cholesky metrics coincide (up to a scaling factor). The geodesic between C_1 and C_2 is $C(\rho(t))$ for $t \in \mathbb{R}$ where:

$$\rho(t) = \frac{\rho_1 \cosh(\lambda t) + \sinh(\lambda t)}{\rho_1 \sinh(\lambda t) + \cosh(\lambda t)}, \quad (7.25)$$

where $\lambda = \log \sqrt{\frac{1+\rho_2}{1-\rho_2}} - \log \sqrt{\frac{1+\rho_1}{1-\rho_1}}$ is known as the difference of the Fisher transformation of the correlation coefficients ρ_1 and ρ_2 .

2. Euclidean-Cholesky and log-Euclidean-Cholesky metrics coincide. The geodesic between C_1 and C_2 is $C(\rho(t))$ for $t \in \mathbb{R}$ where:

$$\rho(t) = \frac{F(t)}{\sqrt{1 + F(t)^2}}, \quad (7.26)$$

where $F(t) = (1-t) \frac{\rho_1}{\sqrt{1-\rho_1^2}} + t \frac{\rho_2}{\sqrt{1-\rho_2^2}}$. This geodesic also coincides with the Lie-Cholesky group geodesic of Section 7.4.4.

See the proof of Theorem 7.20 in Section 11.6.

7.5 Conclusion

In this work, we proposed new Riemannian metrics on the open ellipsope of full-rank correlation matrices that are not invariant under permutations. To the best of our knowledge, all the existing geometric structures were invariant under permutations. Thus the geometries we propose significantly departs from the classical structures. This can be a good assumption in some applications and an irrelevant characteristic in some others. We generalized the recently introduced quotient-affine metrics by studying the congruence action of several matrix Lie groups on SPD matrices. We showed that the family of $LT^+(n)$ -invariant metrics is a natural non-permutation-invariant generalization of affine-invariant metrics. Moreover, they are pullbacks of left-invariant metrics on the Lie group $LT^+(n)$ by the Cholesky map. They are invariant under the congruence action of positive diagonal matrices so they descend to the ellipsope. We explained that the main Riemannian operations can be computed numerically for these quotient-Lie-Cholesky metrics. However, we also showed that the curvature of quotient-affine metrics is unbounded and we can conjecture that the situation is not better for quotient-Lie-Cholesky metrics. In addition, the Riemannian operations are not computable in closed form a priori.

That is why we introduced new Riemannian metrics on the ellipsope in a different way. We kept the Cholesky map which seems to be a good alternative to the invariance under permutations since the space of lower triangular matrices is not stable by permutations. Thus we defined the poly-hyperbolic-Cholesky (PHC) metrics which provide non-positively curved Riemannian symmetric space structures. We also defined two kinds of vector space structures that are flat, geodesically complete and for which all operations are known in closed form. Thus, these three families of metrics provide a Hadamard structure, in particular the Riemannian logarithm and the Fréchet mean are unique. We also put forward a nilpotent Lie group structure for which the group mean is unique. Finally, we proved that in dimension 2, the PHC geodesics are the quotient-affine geodesics and the geodesics of the three last structures coincide. This provides a new interpolation of the correlation coefficient.

It would be nice to test these new metrics on different kinds of data in future works. Moreover, all metrics on correlation matrices provide new product metrics on covariance matrices by decoupling the scales of the variables and the correlations between them. This approach seems promising since in many problems, the correlation gives more information than the covariance on the strength of the relations between the variables, although the scales can remain interesting. Thus, one question could be to adjust the weights between the two components and also between the scales of the variables. The possibilities are multiplied now we have many metrics on correlation matrices. Another direction of research is to investigate permutation-invariant Riemannian metrics on correlation matrices with a simpler geometry than the one of the quotient-affine metrics, for example Hadamard or even flat.

Chapter 8

Permutation-invariant Log-Euclidean metrics

Abstract

There is a growing interest in defining specific tools on correlation matrices which depart from those suited to SPD matrices. Several geometries have been defined on the open ellipsope of full-rank correlation matrices: some are permutation-invariant, some others are log-Euclidean, i.e. diffeomorphic to a Euclidean space. In this work, we merge these two properties by defining the families of off-log metrics and log-scaled metrics. Firstly, we prove that the recently introduced off-log bijection is a diffeomorphism, allowing to pullback (permutation-invariant) inner products. We introduce the “cor-inverse” involution on the open ellipsope which can be seen as analogous to the inversion of SPD matrices. We show that off-log metrics are not inverse-consistent. That is why secondly, we define the log-scaling diffeomorphism between the open ellipsope and the vector space of symmetric matrices with null row sums. This map is based on the congruence action of positive diagonal matrices on SPD matrices, more precisely on the existence and uniqueness of a “scaling”, i.e. an SPD matrix with unit row sums within an orbit. Thanks to this multiplicative approach, log-scaled metrics are inverse-consistent. We provide the main Riemannian operations in closed form for the two families modulo the computation of the respective bijections. In particular, we define a new algorithm that computes the scaling of an SPD matrix in logarithmic time complexity.

8.1 Introduction

In many domains such as Diffusion Tensor Imaging, Brain-Computer Interfaces, brain connectomes or radar signals, the data are time series which are often represented by their covariance matrices. They encode the dependence between the variables and the scale of intensity of these variables. Many Riemannian geometries were proposed to compute with covariance matrices with more natural tools than the Euclidean ones. The use of the affine-invariant metric was shown to outperform many results based on Euclidean metrics such as fiber reconstruction in DTI [Pennec et al., 2006], movement classification in BCI [Barachant et al., 2013] or detection of brain functional connectivity [Varoquaux et al., 2010]. Shortly after, the log-Euclidean metric [Arsigny et al., 2006, Fillard et al., 2007] was shown to be a more efficient alternative to the affine-invariant metric with similar results. The Bures-Wasserstein metric was also proposed to deal with low-rank matrices since the two previous ones are only defined on the space $\text{Sym}^+(n)$ of Symmetric Positive Definite (SPD) matrices. All these metrics belong to the wide families of kernel metrics [Hiai and Petz, 2009]

and $O(n)$ -invariant metrics [Thanwerdas and Pennec, 2022b]. Non $O(n)$ -invariant metrics were also proposed such as the Cholesky [Wang et al., 2004, Grubišić and Pietersz, 2007], log-Euclidean-Cholesky [Li et al., 2017] and log-Cholesky [Lin, 2019] metrics.

In the previously cited domains and in other ones such as phylogenetic trees [Garba et al., 2021] or Gaussian graphical networks [Lauritzen, 1996, Epskamp and Fried, 2018], another possible and natural choice to represent the data is the correlation matrix instead of the covariance matrix. However, the geometries of correlation matrices have been much less studied. Hence they are often considered as covariance matrices on which one can use the classical tools. Nevertheless, these tools are not adapted to correlation matrices, at least the $O(n)$ -invariant ones. Indeed, firstly, the manifold of full-rank correlation matrices is not stable by the congruence action of the orthogonal group so this action has no sense for them. Secondly, it is not a totally geodesic submanifold for neither of the noted $O(n)$ -invariant metrics on SPD matrices, except the Euclidean metric. This motivates the study of intrinsic geometries of correlation matrices. Such geometries could also have a great impact on applications with covariance matrices since they would provide product metrics with one part on diagonal matrices and the other part on correlation matrices. It would thus allow to decouple the scale of the variables from the dependence between the variables.

Among the geometries proposed on correlation matrices, one involves a surjection from a product of spheres [Rebonato and Jaekel, 2001, Kercheval, 2008]. It is an orbit space [Alekseevsky et al., 2001], this construction is quite analogous to the Bures-Wasserstein geometry of covariance matrices. However, to our knowledge, it has not been precisely described yet. A metric space structure called the Hilbert geometry relies on the convexity of the set [Nielsen and Sun, 2019]. Among the Riemannian structures, the recently introduced quotient-affine metric is obtained by taking the quotient of the affine-invariant metric under the congruence action of positive diagonal matrices $\star : (\Delta, \Sigma) \in \text{Diag}^+(n) \times \text{Sym}^+(n) \mapsto \Delta \Sigma \Delta \in \text{Sym}^+(n)$ [David and Gu, 2019, Thanwerdas and Pennec, 2021]. Indeed, full-rank correlation matrices can be seen as the orbits of this action so the space $\text{Cor}^+(n)$ of such matrices is the quotient manifold $\text{Sym}^+(n)/\text{Diag}^+(n)$ and any invariant metric on $\text{Sym}^+(n) \simeq \text{Cor}^+(n) \times \text{Diag}^+(n)$ descends to a Riemannian metric on $\text{Cor}^+(n)$. These constructions have the common property to be invariant under permutations. It means that the statistical analysis is invariant under reordering the variables, which can be a relevant hypothesis when the order is arbitrary. When the order of the variables is meaningfully chosen depending on the application (e.g. for auto-correlation matrices), other Riemannian metrics that are not permutation-invariant can be considered. The metrics proposed in [Thanwerdas and Pennec, 2022c] provide a Hadamard structure or even a vector space structure, which are very convenient for computing with correlation matrices.

Given this short survey on geometries of correlation matrices, there is an obvious gap to fill in: no permutation-invariant log-Euclidean metrics have been derived yet on the space $\text{Cor}^+(n)$ of full-rank correlation matrices. By *log-Euclidean*, we mean the pullback of an inner product on a vector space \mathcal{V} by a diffeomorphism referred to as a logarithm and denoted $\text{Log} : \text{Cor}^+(n) \rightarrow \mathcal{V}$. In this work, we propose two approaches to define such a diffeomorphism. The first one is based on a recent bijective parametrization of full-rank correlation matrices by the space $\text{LT}^0(n)$ of lower triangular matrices with null diagonal introduced by Archakov and Hansen [Archakov and Hansen, 2021]. The second one is entirely new. We rephrase their framework to present the two approaches in a similar way to facilitate the comparison

between them and the comprehension of the second one by analogy with the first one. These two methods are summarized in Table 8.1 and explained below.

	First approach (additive) Off-log diffeomorphism	Second approach (multiplicative) Cor-exp diffeomorphism
Action	$+$: $\begin{cases} \text{Diag}(n) \times \text{Sym}(n) & \longrightarrow \text{Sym}(n) \\ (D, S) & \longmapsto D + S \end{cases}$	\star : $\begin{cases} \text{Diag}^+(n) \times \text{Sym}^+(n) & \longrightarrow \text{Sym}^+(n) \\ (\Delta, \Sigma) & \longmapsto \Delta \Sigma \Delta \end{cases}$
Claim	$\forall S \in \text{Sym}(n), \exists! D := \mathcal{D}(S) \in \text{Diag}(n) :$ $\exp(D + S) \in \text{Cor}^+(n)$	$\forall \Sigma \in \text{Sym}^+(n), \exists! \Delta := \mathcal{D}^*(\Sigma) \in \text{Diag}^+(n) :$ $\log(\Delta \star \Sigma) \in \mathcal{V}^*$
Status	Claim proved in [Archakov and Hansen, 2021]	Claim proved for $\mathcal{V}^* = \text{Row}_0(n)$ in Section 3
Diagram		
Outline	Sec. 2.1. Definition of \mathcal{D} Sec. 2.2. Definition of pullback metrics Sec. 2.3. Numerical computation of \mathcal{D}	Sec. 3. Def. of \mathcal{D}^* (choice of appropriate \mathcal{V}^*) Sec. 4. Definition of pullback metrics Sec. 5. Numerical computation of \mathcal{D}^*

Table 8.1: Two approaches to define permutation-invariant log-Euclidean metrics on $\text{Cor}^+(n)$. The left part is based on [Archakov and Hansen, 2021]. The right part is new; \mathcal{V}^* is a vector space stable by permutations satisfying $\text{Sym}(n) = \mathcal{V}^* \oplus \text{Diag}(n)$.

8.1.1 Results and organization of the chapter

In the remainder of this section, we introduce the necessary notations. We also define a natural involution on the open elliptope called the cor-inversion, which allows to define a notion of inverse-consistency for Riemannian metrics.

In [Archakov and Hansen, 2021], the authors show the claim on the left of Table 8.1, that is for all symmetric matrix S , there exists a unique diagonal matrix $D = \mathcal{D}(S)$ such that $\exp(D + S)$ is a full-rank correlation matrix. Thus it defines a surjective map $\pi : S \in \text{Sym}(n) \longmapsto \exp \circ (\mathcal{D}(S) + S) \in \text{Cor}^+(n)$ which is equivariant under permutations, and a bijective map $L \in \text{LT}^0(n) \longmapsto \pi(L + L^\top) \in \text{Cor}^+(n)$. We astutely replace the space $\text{LT}^0(n)$ by the space $\mathcal{V} = \text{Hol}(n)$ of symmetric matrices with null diagonal (which is of same dimension) so that the restriction $\text{Exp} = \pi|_{\mathcal{V}} : \mathcal{V} \longrightarrow \text{Cor}^+(n)$ is also equivariant under permutations. Note that $\text{Hol}(n)$ is actually the tangent space of $\text{Cor}^+(n)$. Moreover, we trivially observe that π is invariant by the additive action of a diagonal matrix. Our contribution is to show that the bijection Exp is a smooth (that is \mathcal{C}^∞) diffeomorphism, to define by pullback the family of *off-log metrics* and to provide all the Riemannian operations in closed form modulo the computation of \mathcal{D} . An algorithm with logarithmic time complexity is already defined in [Archakov and Hansen, 2021] to compute \mathcal{D} . This additive approach is summarized on the left part of Table 8.1 and exposed in Section 8.2.

Our second approach consists in inverting the roles played by the vector spaces $\text{Sym}(n) = \mathcal{V} \oplus \text{Diag}(n)$ and the manifolds $\text{Sym}^+(n) = \text{Cor}^+(n) \times \text{Diag}^+(n)$, as well as the matrix exponential and the matrix logarithm, and especially to replace the additive action of $\text{Diag}(n)$ on $\text{Sym}(n)$ by the congruence action $\star : (\Delta, \Sigma) \in \text{Diag}^+(n) \times \text{Sym}^+(n) \longmapsto \Delta \Sigma \Delta \in \text{Sym}^+(n)$. In this work, we find a vector space \mathcal{V}^* such that for all $\Sigma \in \text{Sym}^+(n)$, there exists a unique

$\Delta = \mathcal{D}^*(\Sigma) \in \text{Diag}^+(n)$ such that $\log(\Delta\Sigma\Delta) \in \mathcal{V}^*$. This allows to define the surjective map $\pi^* : \Sigma \in \text{Sym}^+(n) \mapsto \log(\mathcal{D}^*(\Sigma) \star \Sigma) \in \mathcal{V}^*$ and the bijective map $\text{Log}^* = \pi|_{\text{Cor}^+(n)} : \text{Cor}^+(n) \rightarrow \mathcal{V}^*$ which are equivariant under permutations. This multiplicative approach is summarized on the right part of Table 8.1 and exposed in Section 8.3. One major advantage of this multiplicative approach is that it intrinsically respects the structure of correlation matrices since $\text{Cor}^+(n) = \text{Sym}^+(n)/\text{Diag}^+(n)$, contrarily to the additive approach. The main consequence is the compatibility with the inversion, i.e. $\pi^*(C^{-1}) = -\pi^*(C)$ for all $C \in \text{Cor}^+(n)$.

More precisely, we try to prove the claim with $\mathcal{V}^* = \text{Hol}(n)$ and $\mathcal{V}^* = \text{Row}_0(n)$, where $\text{Row}_0(n)$ is the vector space of symmetric matrices with null row sums. With the first choice, we only manage to prove the existence. We actually prove that the uniqueness would imply the uniqueness of the Riemannian logarithm at identity of the quotient-affine metric mentioned above, which is an open problem. This is a secondary contribution that relates two problems on full-rank correlation matrices. In contrast, we prove the claim with $\mathcal{V}^* = \text{Row}_0(n)$. Indeed, we show that $\exp(\text{Row}_0(n)) = \text{Row}_1^+(n)$, where $\text{Row}_1^+(n)$ is the submanifold of SPD matrices with unit row sums. This reduces our question to the famous problem of scaling an SPD matrix to prescribed row sums by congruence of a positive diagonal matrix: for all SPD matrix Σ , does there exist a positive diagonal matrix $\Delta = \mathcal{D}^*(\Sigma)$ such that $\Delta\Sigma\Delta \in \text{Row}_1^+(n)$. The answer is yes [Marshall and Olkin, 1968, Johnson and Reams, 2009] so the claim is true.

In Section 8.4, similarly to the additive approach, we prove that the bijection Log^* is a smooth diffeomorphism and we define by pushforward the family of *log-scaled metrics*. Contrarily to off-log metrics, they are inverse-consistent. We provide all the Riemannian operations in closed form modulo the computation of \mathcal{D}^* , that is the computation of the scaling of an SPD matrix. In Section 8.5, we design a new algorithm to compute the scaling of an SPD matrix which is more efficient than generic algorithms to compute the scaling such as [Khachiyan and Kalantari, 1992, O’Leary, 2003] (see surveys [Johnson and Reams, 2009, Idel, 2016, Allen-Zhu et al., 2017]). We build it by following the proof of existence given by [Marshall and Olkin, 1968] and adapting the efficient Projected Gradient Descent (PGD) algorithm [Bubeck, 2015]. We conclude in Section 8.6.

8.1.2 Notations

8.1.2.1 Matrices

Tables 8.2 and 8.3 summarize our notations for matrix spaces. We also use the following constant and linear operators on vectors and matrices:

- $\mathbf{1} \in \mathbb{R}^n$ is the vector with all entries equal to 1;
- $\text{diag} : \mathbb{R}^n \rightarrow \text{Diag}(n)$ makes a diagonal matrix from a vector;
- $\text{sum} : \mathbb{R}^n \rightarrow \mathbb{R}$ computes the sum of the entries of a vector;
- $\text{Diag} : \text{Mat}(n) \rightarrow \text{Diag}(n)$ extracts the diagonal matrix from a matrix;
- $\text{Off} : \text{Mat}(n) \rightarrow \ker \text{Diag}$ subtracts the diagonal matrix from a matrix;
- $\text{Sum} : \text{Mat}(n) \rightarrow \mathbb{R}$ computes the sum of entries of a matrix;

- $\cdot : \text{Mat}(n) \times \text{Mat}(n) \longrightarrow \text{Mat}(n)$ computes the Hadamard/Schur (entry-wise) product of matrices.

Squared of size n	$\text{Mat}(n) = \{M = [M_{ij}]_{1 \leq i, j \leq n} M_{ij} \in \mathbb{R}\}$
Skew-symmetric	$\text{Skew}(n) = \{Y \in \text{Mat}(n) Y^\top = -Y\}$
Symmetric	$\text{Sym}(n) = \{X \in \text{Mat}(n) X^\top = X\}$
Diagonal	$\text{Diag}(n) = \{\text{Diag}(X) X \in \text{Mat}(n)\}$
Symmetric hollow (null diagonal)	$\text{Hol}(n) = \{X \in \text{Sym}(n) \text{Diag}(X) = 0\}$
Symmetric null-row-sum	$\text{Row}_0(n) = \{X \in \text{Sym}(n) X\mathbf{1} = 0\}$

Table 8.2: Matrix vector spaces.

Invertible	$\text{GL}(n) = \{A \in \text{Mat}(n) \det(A) \neq 0\}$
Orthogonal	$\text{O}(n) = \{A \in \text{GL}(n) AA^\top = I_n\}$
Symmetric Positive Definite	$\text{Sym}^+(n) = \{AA^\top A \in \text{GL}(n)\}$
Positive diagonal	$\text{Diag}^+(n) = \text{Sym}^+(n) \cap \text{Diag}(n)$
Full-rank correlation	$\text{Cor}^+(n) = \{C \in \text{Sym}^+(n) \text{Diag}(C) = I_n\}$
SPD unit-row-sum	$\text{Row}_\mathbf{1}^+(n) = \{\Sigma \in \text{Sym}^+(n) \Sigma\mathbf{1} = \mathbf{1}\}$

Table 8.3: Matrix manifolds.

We recall the definition of the congruence action $\star : (A, M) \in \text{GL}(n) \times \text{Mat}(n) \mapsto AMA^\top \in \text{Mat}(n)$. Since the permutation group $\mathfrak{S}(n)$ is a subgroup of the orthogonal group via the canonical injection $\sigma \in \mathfrak{S}(n) \mapsto P_\sigma = [\delta_{i, \sigma(j)}]_{1 \leq i, j \leq n} \in \text{O}(n)$, the permutation action $(\sigma, M) \in \mathfrak{S}(n) \times \text{Mat}(n) \mapsto \sigma \cdot M = P_\sigma M P_\sigma^\top \in \text{Mat}(n)$ is the congruence action of the subgroup $\mathfrak{S}(n)$.

We recall the definition of the matrix exponential $\exp : M \in \text{Mat}(n) \longrightarrow \sum_{k=0}^{+\infty} \frac{1}{k!} M^k \in \text{GL}(n)$ which is a smooth map. Its restriction to symmetric matrices is a smooth diffeomorphism onto SPD matrices, $\exp : \text{Sym}(n) \longrightarrow \text{Sym}^+(n)$. The symmetric matrix logarithm is its smooth inverse, $\log : \text{Sym}^+(n) \longrightarrow \text{Sym}(n)$. The computation of \exp , \log and their differentials are particularly simple modulo eigenvalue decomposition. Given $\Sigma = PDP^\top \in \text{Sym}^+(n)$, $X = Q\Delta Q^\top$, $Y \in \text{Sym}(n)$ where $P, Q \in \text{O}(n)$, $D \in \text{Diag}^+(n)$ and $\Delta \in \text{Diag}(n)$:

$$\exp(X) = Q \exp(\Delta) Q^\top, \quad (8.1)$$

$$\log(\Sigma) = P \log(D) P^\top, \quad (8.2)$$

$$d_X \exp(Y) = Q \left([\exp^{[1]}(\delta_i, \delta_j)]_{1 \leq i, j \leq n} \bullet (Q^\top Y Q) \right) Q^\top, \quad (8.3)$$

$$d_\Sigma \log(Y) = P \left([\log^{[1]}(d_i, d_j)]_{1 \leq i, j \leq n} \bullet (P^\top Y P) \right) P^\top, \quad (8.4)$$

where $f^{[1]}(x, y) = \begin{cases} \frac{f(x)-f(y)}{x-y} & \text{if } x \neq y \\ f'(x) & \text{if } x = y \end{cases}$ is the first divided difference of $f \in \{\exp, \log\}$ [Bhatia, 1997]. In other words, the maps $\exp, \log, d\exp, d\log$ are $\text{O}(n)$ -equivariant, and therefore $\mathfrak{S}(n)$ -equivariant.

8.1.2.2 Correlation matrices

The manifold of full-rank correlation matrices is called the open elliptope. It is relatively open in $\text{Sym}(n)$, i.e. open in $I_n + \text{Hol}(n)$.

We introduce the smooth submersion $\text{Cor} : \Sigma \in \text{Sym}^+(n) \mapsto \text{Diag}(\Sigma)^{-1/2} \Sigma \text{Diag}(\Sigma)^{-1/2} \in \text{Cor}^+(n)$. Given $\Sigma \in \text{Sym}^+(n)$, $X \in \text{Sym}(n)$, denoting $\Delta = \text{Diag}(\Sigma)^{-1/2}$, its differential is:

$$d_{\Sigma} \text{Cor}(X) = \Delta \left[X - \frac{1}{2} (\Delta^2 \text{Diag}(X) \Sigma + \Sigma \text{Diag}(X) \Delta^2) \right] \Delta. \quad (8.5)$$

We introduce a notation for equicorrelation matrices $C(\rho) = (1 - \rho)I_n + \rho \mathbf{1}\mathbf{1}^T \in \text{Cor}^+(n)$ where $\rho \in (-\frac{1}{n-1}, 1)$. Given a correlation matrix $C \in \text{Cor}^+(n)$, there exist partitions of n , i.e. sets $I = \{i_1, \dots, i_p\}$ satisfying $i_1, \dots, i_p \geq 1$ and $i_1 + \dots + i_p = n$, partitioning the matrix C into equicorrelation diagonal blocks and constant off-diagonal blocks. The signature of C is the maximum I_C of such sets I with respect to the natural order on partitions of n . We say that C is a block equicorrelation matrix of signature I_C (see Table 8.4). For example, an equicorrelation matrix is a block equicorrelation matrix with signature $\{n\}$. The maps introduced in this chapter preserve the signature.

$$C(\rho) = \left(\begin{array}{cccc} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{array} \right) \left| \begin{array}{cccc} C(\rho_1) & \rho_{12} \mathbf{1}\mathbf{1}^T & \cdots & \rho_{1p} \mathbf{1}\mathbf{1}^T \\ \rho_{12} \mathbf{1}\mathbf{1}^T & C(\rho_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{p-1,p} \mathbf{1}\mathbf{1}^T \\ \rho_{1p} \mathbf{1}\mathbf{1}^T & \cdots & \rho_{p-1,p} \mathbf{1}\mathbf{1}^T & C(\rho_p) \end{array} \right)$$

Table 8.4: Equicorrelation and block equicorrelation matrices.

The elliptope is stable by the permutation action, it is not stable by the congruence action of the orthogonal group $O(n)$. It is not stable by inversion either. However, the correlation matrix of its inverse, namely $\text{Cor}(C^{-1})$, contains the same information as the partial correlation matrix defined below.

Definition 8.1 (Cor-inversion, partial correlation matrix)

- The cor-inversion is the smooth involution $\mathcal{I} : C \in \text{Cor}^+(n) \mapsto \text{Cor}(C^{-1}) \in \text{Cor}^+(n)$.
- The partial correlation matrix of an SPD matrix $\Sigma \in \text{Sym}^+(n)$ is the matrix $\Gamma \in \text{Sym}(n)$ defined by $\Gamma = I_n - \text{Off}(\mathcal{I}(C)) = 2I_n - \mathcal{I}(C)$, where $C = \text{Cor}(\Sigma)$.

The bijective parametrization $C \mapsto \Gamma$ is used in the theory of stationary stochastic processes where the (potentially complex and infinite-dimensional) matrices are Toeplitz. The set of partial correlation coefficients (along with the common variance) is considered as an alternative “representation of the second-order statistics” [Burg, 1975, Section II.B.5] of the process with respect to the traditional “auto-correlation” (or auto-covariance) function. This characterization is used in signal processing, especially in radar signal processing where the manifold of SPD Toeplitz matrices is traditionally endowed with the Poincaré polydisk geometry [Barbaresco, 2013]. In Gaussian graphical networks, the partial correlation between two variables indicates the correlation between them conditionally to the other variables.

Thus the partial correlations are the weights of the arrows in the network [Lauritzen, 1996, Koller and Friedman, 2009, Epskamp and Fried, 2018]. This approach is applied in many domains such as genomics [de la Fuente et al., 2004, Peng et al., 2009] or brain connectomics [Marrelec et al., 2006]. Hence the importance of partial correlations confirms that the cor-inversion is a relevant concept.

We consider the cor-inversion as analogous to the matrix inversion for SPD matrices $\text{inv} : \text{Sym}^+(n) \rightarrow \text{Sym}^+(n)$. The cor-inversion commutes with permutations on full-rank correlation matrices as well as the inversion commutes with the congruence by $O(n)$ on SPD matrices. Moreover, for all signature I , the space of block equicorrelation matrices of signature I is stable by the cor-inversion. In addition, in relation to the problem of this chapter, we have the following result.

Theorem 8.2 (Compatibility between the second approach and the cor-inversion) Let \mathcal{V}^* be a vector space stable by permutations and such that $\text{Sym}(n) = \mathcal{V}^* \oplus \text{Diag}(n)$. We assume that the claim on the right part of Table 8.1 is true, i.e. for all $C \in \text{Cor}^+(n)$, there exists a unique $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta C \Delta) \in \mathcal{V}^*$. This defines the inverse bijections $\text{Log}^* : C \in \text{Cor}^+(n) \mapsto \log(\Delta C \Delta) \in \mathcal{V}^*$ and $\text{Exp}^* = \text{Cor} \circ \text{exp} : \mathcal{V}^* \rightarrow \text{Cor}^+(n)$. Then we automatically have $\text{Log}^*(\mathcal{I}(C)) = -\text{Log}^*(C)$, i.e. the following diagram commutes.

$$\begin{array}{ccc} \text{Cor}^+(n) & \xrightarrow{\mathcal{I}} & \text{Cor}^+(n) \\ \downarrow \text{Log}^* & & \downarrow \text{Log}^* \\ \mathcal{V}^* & \xrightarrow{-\text{Id}} & \mathcal{V}^* \end{array} \quad (8.6)$$

Proof. Note that for all $\Sigma \in \text{Sym}^+(n)$, we have $\text{Cor}(\Sigma^{-1}) = \text{Cor}(\text{Cor}(\Sigma)^{-1})$. Indeed, if $\Sigma = DCD$ with $C = \text{Cor}(\Sigma) \in \text{Cor}^+(n)$, then $\text{Cor}(\Sigma^{-1}) = \text{Cor}(D^{-1}C^{-1}D^{-1}) = \text{Cor}(C^{-1})$. Therefore, for all $X \in \mathcal{V}^*$ we have $\text{Exp}^*(-X) = \text{Cor}(\text{exp}(-X)) = \text{Cor}(\text{exp}(X)^{-1}) = \text{Cor}(\text{Exp}^*(X)^{-1}) = \mathcal{I}(\text{Exp}^*(X))$. Thus with $C = \text{Exp}^*(X)$, we have $\text{Log}^*(\mathcal{I}(X)) = -\text{Log}^*(C)$. \square

Otherwise said, the multiplicative approach is automatically compatible with the cor-inversion. This is due to the use of the congruence action of positive diagonal matrices on SPD matrices instead of the additive action of diagonal matrices on symmetric matrices. Indeed, the former is intrinsically related to the definition of a correlation matrix. On the contrary, we can expect that the bijections built via the additive approach are not compatible with the cor-inversion in general.

Thus, if one finds a vector space \mathcal{V}^* satisfying the claim and if the bijections are smooth, the log-Euclidean metrics defined by pullback will automatically be inverse-consistent. This is quite satisfying for log-Euclidean metrics on full-rank correlation matrices in analogy with SPD matrices.

8.2 Permutation-invariant log-Euclidean metrics via the off-log diffeomorphism

In this section, we rephrase the framework of [Archakov and Hansen, 2021] (Section 8.2.1) to ease the comprehension of the next sections by analogy. In Section 8.2.2, we prove that the

bijection they define, that we call the off-log bijection and that we denote $\text{Log} : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$, is actually a diffeomorphism. It allows to pullback inner products on full-rank correlation matrices. Since the off-log diffeomorphism is equivariant under permutations, we give a characterization of permutation-invariant inner products on $\text{Hol}(n)$ so that their pullbacks provide permutation-invariant log-Euclidean metrics on $\text{Cor}^+(n)$. Then, we detail the Riemannian operations of these metrics. We prove that, as expected, the log-Euclidean metrics such defined are not inverse-consistent with respect to the cor-inversion. In Section 8.2.3, we simply recall the algorithm of [Archakov and Hansen, 2021] to compute the inverse diffeomorphism $\text{Exp} = \text{Log}^{-1}$, the speed of convergence and the complexity.

8.2.1 The off-log bijection

Theorem 8.3 states that the claim in the left part of Table 8.1 is true. It allows to define the off-log bijection $\text{Log} : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$. Theorem 8.4 states some interesting properties of the off-log bijection. These results are due to Archakov and Hansen [Archakov and Hansen, 2021].

Theorem 8.3 (Definition of \mathcal{D}) [Archakov and Hansen, 2021] For all $S \in \text{Sym}(n)$, there exists a unique $D \in \text{Diag}(n)$ such that $\exp(D + S) \in \text{Cor}^+(n)$. This allows to define:

- the surjective map $\mathcal{D} : S \in \text{Sym}(n) \mapsto D \in \text{Diag}(n)$,
- the surjective map $\pi : S \in \text{Sym}(n) \mapsto \exp(\mathcal{D}(S) + S) \in \text{Cor}^+(n)$ which is invariant under the additive group action $+$: $\text{Diag}(n) \times \text{Sym}(n) \rightarrow \text{Sym}(n)$,
- the bijective map $\text{Exp} = \pi|_{\text{Hol}(n)} : \text{Hol}(n) \rightarrow \text{Cor}^+(n)$ (note that $\pi = \text{Exp} \circ \text{Off}$),
- the smooth bijective inverse map $\text{Log} = \text{Exp}^{-1} = \text{Off} \circ \log : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$ that we call the *off-log bijection*.

Theorem 8.4 (Properties of the off-log bijection) [Archakov and Hansen, 2021]

1. (Equivariance) Log and Exp are equivariant under permutations.
2. (Equicorrelation matrix) For all $\rho \in (-\frac{1}{n-1}, 1)$, $\text{Log}(C(\rho)) = \frac{1}{n} \ln \left(\frac{1+(n-1)\rho}{1-\rho} \right) (\mathbb{1}\mathbb{1}^\top - I_n)$.
In dimension $n = 2$, $\text{Log}(C(\rho)) = \begin{pmatrix} 0 & F(\rho) \\ F(\rho) & 0 \end{pmatrix}$ where $F(\rho) = \frac{1}{2} \log\left(\frac{1+\rho}{1-\rho}\right) \in \mathbb{R}$ is the Fisher transformation of the correlation coefficient $\rho \in (-1, 1)$.
3. (Block equicorrelation matrix) If C is a block equicorrelation matrix of signature $I = \{i_1, \dots, i_p\}$, then $\text{Log}(C)$ is a block symmetric hollow matrix of signature I with multiples of $\mathbb{1}_{i_j} \mathbb{1}_{i_j}^\top - I_{i_j}$ on diagonal blocks and multiples of $\mathbb{1}_{i_j} \mathbb{1}_{i_k}^\top$ on off-diagonal blocks.
4. (Generalization) For all $\Delta \in \text{Diag}^+(n)$, for all $S \in \text{Sym}(n)$, there exists a unique $D \in \text{Diag}(n)$ such that $\text{Diag}(\exp(D + S)) = \Delta$.

Note that Theorem 8.3 is a particular case of Theorem 8.4 item 4 with $\Delta = I_n$. The result in dimension 2 was stated as a motivation in [Archakov and Hansen, 2021] to use the map $\text{Log} = \text{Off} \circ \log$ in higher dimensions since it gives in dimension 2 a well known transformation

of the correlation coefficient. Interestingly, the same coefficient appears in dimension 2 for the quotient-affine metric [Thanwerdas and Pennec, 2021].

By analogy with the symmetric matrix logarithm $\log : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$ satisfying $\log(\Sigma^{-1}) = -\log(\Sigma)$, one could expect that the off-log bijection $\text{Log} : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$ “commutes with inversion”, i.e. satisfies $\text{Log}(\mathcal{I}(C)) = -\text{Log}(C)$. We show that it is not the case as we argued in the introduction.

Theorem 8.5 (Incompatibility between cor-inversion and off-log bijection) Let $n \geq 3$. There exists $C \in \text{Cor}^+(n)$, such that $\text{Log}(\mathcal{I}(C)) \neq -\text{Log}(C)$. Otherwise said, the following diagram does not commute.

$$\begin{array}{ccc} \text{Cor}^+(n) & \xrightarrow{\mathcal{I}} & \text{Cor}^+(n) \\ \downarrow \text{Log} & & \downarrow \text{Log} \\ \text{Hol}(n) & \xrightarrow{-\text{Id}} & \text{Hol}(n) \end{array} \quad (8.7)$$

See the proof of Theorem 8.5 in Section 11.7.

This incompatibility is one of the justifications of the multiplicative approach that we present in Section 8.3. Still, this bijection $\text{Log} : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$ remains a very nice tool that allows to define permutation-invariant log-Euclidean metrics on full-rank correlation matrices. Let us show this.

8.2.2 Permutation-invariant pullback metrics via the off-log diffeomorphism

This section is part of our contributions. We prove that the off-log bijection $\text{Log} : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$ is actually a diffeomorphism (Section 8.2.2.1). Then we characterize all permutation-invariant inner products on $\text{Hol}(n)$ (Section 8.2.2.2) and we pull them back to permutation-invariant log-Euclidean metrics on $\text{Cor}^+(n)$ (Section 8.2.2.3).

8.2.2.1 The off-log bijection is a diffeomorphism

Theorem 8.6 ($\text{Log} = \text{Off} \circ \log$ is a diffeomorphism) The off-log bijection $\text{Log} : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$ is a smooth diffeomorphism. We give the differentials of Log and Exp in function of the differentials of the symmetric matrix logarithm and exponential maps \log and \exp . For all $C \in \text{Cor}^+(n)$ and $S, X, Y \in \text{Hol}(n)$:

$$d_C \text{Log}(X) = \text{Off}(d_C \log(X)), \quad (8.8)$$

$$d_S \text{Exp}(Y) = d_{\log(\text{Exp}(S))} \exp(Y - \text{Diag}(\text{Exp}(S)Y)). \quad (8.9)$$

See the proof of Theorem 8.6 in Section 11.7.

8.2.2.2 Permutation-invariant inner products on $\text{Hol}(n)$

The characterization of permutation-invariant inner products on $\text{Hol}(n)$ can be found in Section 3.3.4.1 of Chapter 3.

Theorem 8.7 (Permutation-invariant inner products on $\text{Hol}(n)$) [Section 3.3.4.1] For $n \geq 4$, permutation-invariant inner products on $\text{Hol}(n)$ are the symmetric bilinear forms associated to the following positive definite quadratic forms defined for $X \in \text{Hol}(n)$:

$$q(X) = \alpha \text{tr}(X^2) + \beta \text{Sum}(X^2) + \gamma \text{Sum}(X)^2 \quad (8.10)$$

with $\alpha > 0$, $2\alpha + (n-2)\beta > 0$ and $\alpha + (n-1)(\beta + n\gamma) > 0$. For $n = 3$, the permutation-invariant inner products have the same form with $\alpha = 0$, i.e. $q(X) = \beta \text{Sum}(X^2) + \gamma \text{Sum}(X)^2$ with $\beta > 0$ and $\beta + 3\gamma > 0$. For $n = 2$, they have the same form with $\alpha = \beta = 0$, i.e. $q(X) = \gamma \text{Sum}(X)^2$ with $\gamma > 0$.

8.2.2.3 Pullback metrics via the off-log diffeomorphism

Definition 8.8 (Off-log metrics) An off-log metric on $\text{Cor}^+(n)$ is the pullback metric of an inner product characterized by a quadratic form q as in Definition 8.7. For all $C \in \text{Cor}^+(n)$ and $X \in T_C \text{Cor}^+(n) = \text{Hol}(n)$, it writes $g_C(X, X) = q(d_C \text{Log}(X))$ where $d_C \text{Log}(X) = \text{Off}(d_C \log(X))$.

Theorem 8.9 (Riemannian operations of off-log metrics) We consider an off-log metric characterized by the quadratic form q . Let $C, C', C_1, \dots, C_k \in \text{Cor}^+(n)$, $X \in \text{Hol}(n)$. The Riemannian operations of this metric are summarized in Table 8.5.

Exponential map	$\text{Exp}_C(X) = \text{Exp}(\text{Log}(C) + d_C \text{Log}(X))$
Logarithm map	$\text{Log}_C(C') = d_{\text{Log}(C)} \text{Exp}(\text{Log}(C') - \text{Log}(C))$
Geodesic	$\gamma(t) = \text{Exp}((1-t)\text{Log}(C) + t\text{Log}(C'))$
Squared distance	$d(C, C')^2 = q(\text{Log}(C') - \text{Log}(C))$
Fréchet mean	$C = \text{Exp}(\frac{1}{k} \sum_{i=1}^k \text{Log}(C_i))$
Curvature	$R = 0$
Parallel transport	$\Pi_{C \rightarrow C'} X = (d_{C'} \text{Log})^{-1}(d_C \text{Log}(X))$

Table 8.5: Riemannian operations of off-log metrics.

Beware that the Riemannian exponential and logarithm maps only coincide with the diffeomorphisms $\text{Exp} : \text{Hol}(n) \rightarrow \text{Cor}^+(n)$ and $\text{Log} : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$ at $C = I_n$ introduced in Theorem 8.3. They differ from the symmetric matrix diffeomorphisms $\text{exp} : \text{Sym}(n) \rightarrow \text{Sym}^+(n)$ and $\text{log} : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$.

Therefore, the off-log diffeomorphism provides a closed-form distance between two full-rank correlation matrices (modulo the computation of a symmetric matrix logarithm, i.e. modulo an eigenvalue decomposition). Moreover, all the other Riemannian operations can be computed in closed form modulo the computation of Exp , i.e. the computation of \mathcal{D} . In the next section, we recall how it is computed in [Archakov and Hansen, 2021].

8.2.3 Numerical computation of the off-log inverse map

The implicit functions \mathcal{D} and $\text{Exp} = \text{exp} \circ (\mathcal{D} + \text{Id}_{\text{Sym}(n)})$ can be computed very efficiently, as proved in [Archakov and Hansen, 2021].

Theorem 8.10 (Algorithm to compute \mathcal{D} : convergence, speed, complexity) [Archakov and Hansen, 2021] Let $S \in \text{Sym}(n)$ and $D_0 \in \text{Diag}(n)$, e.g. $D_0 = 0$. The sequence $(D_k)_{k \in \mathbb{N}} \subset \text{Diag}(n)$ is recursively defined by $D_{k+1} = \varphi_S(D_k)$ where $\varphi_S : D \in \text{Diag}(n) \mapsto D - \log(\text{Diag}(\exp(D + S)))$ is a smooth L -contractant map with $L \in [0, 1)$. Then:

1. (Convergence) (D_k) converges to the fixed point $D^* = \mathcal{D}(S) \in \text{Diag}(n)$ of φ_S .
2. (Speed of convergence) The convergence is exponential: $\|D_k - D^*\| \leq \frac{L^k}{1-L} \|D_1 - D_0\|$.
3. (Time complexity) For all $\varepsilon > 0$, the sufficient number of iterations so that $\|D_k - D^*\| < \varepsilon$ is $k_\varepsilon = O(\ln(\frac{n}{\varepsilon}))$.
4. (Complexity) At each iteration, a symmetric matrix exponential is computed which is a $O(n^3)$. Hence the complexity with precision $\varepsilon > 0$ is $O(n^3 k_\varepsilon) = O(n^3 \ln(\frac{n}{\varepsilon}))$.

In this section, we recalled the main facts on the off-log parametrization introduced in [Archakov and Hansen, 2021] and we transformed it as a geometric tool to introduce log-Euclidean metrics on full-rank correlation matrices. We also formalized this tool in terms of invariance under a group action and we showed that off-log metrics are not inverse-consistent. In the next sections, we rely on this formalization to introduce the family of log-scaled metrics which are permutation-invariant, log-Euclidean and inverse-consistent.

8.3 The log-scaling bijection

In this section, we examine two versions of the following conjecture: for all $\Sigma \in \text{Sym}^+(n)$, there exists a unique $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta \Sigma \Delta) \in \mathcal{V}^*$. This conjecture depends on \mathcal{V}^* , which is a vector space stable by permutations satisfying $\text{Sym}(n) = \mathcal{V}^* \oplus \text{Diag}(n)$. In Section 8.3.1, we relate the conjecture with $\mathcal{V}^* = \text{Hol}(n)$ to the problem of the quotient-affine logarithm [David and Gu, 2019, Thanwerdas and Pennec, 2021]. We prove the existence and we explain why the uniqueness remains difficult to prove. In Section 8.3.2, we explain why $\mathcal{V}^* = \text{Row}_0(n)$ is a good candidate for the conjecture to be true and in Section 8.3.3, we prove the conjecture thanks to a result known as the existence and uniqueness of the *scaling* of SPD matrices [Marshall and Olkin, 1968, Johnson and Reams, 2009]. We explain the proof of existence because it can be difficult to understand in [Marshall and Olkin, 1968] and because we need it in Section 8.5 to design the algorithm that computes the scaling. In Section 8.3.4, we give the properties of our new Euclideanization called the log-scaling bijection.

8.3.1 Is the conjecture true with $\mathcal{V}^* = \text{Hol}(n)$?

Before relating the conjecture with $\mathcal{V}^* = \text{Hol}(n)$ to the problem of existence and uniqueness of the Riemannian logarithm at I_n of the quotient-affine metric, we first recall the conjecture and the definition of the quotient-affine metric.

Conjecture 8.11 (The result is true with $\mathcal{V}^* = \text{Hol}(n)$) For all $\Sigma \in \text{Sym}^+(n)$, there exists a unique matrix $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta \Sigma \Delta) \in \text{Hol}(n)$.

Definition 8.12 (Quotient-affine metric) The quotient-affine metric on $\text{Cor}^+(n)$ is the quotient metric of the affine-invariant metric on $\text{Sym}^+(n)$ by the congruence action of $\text{Diag}^+(n)$

[David and Gu, 2019]. At I_n , the horizontal space is $\mathcal{H}_{I_n}^{\text{QA}} = \text{Hol}(n)$, the quotient-affine metric writes $g_{I_n}^{\text{QA}}(X, X) = \text{tr}(X^2)$ and the exponential map writes $\text{Exp}_{I_n}^{\text{QA}}(X) = \text{Cor}(\exp(X))$ for all $X \in \text{Hol}(n)$ [Thanwerdas and Penne, 2021].

Theorem 8.13 (Existence and equivalence of conjectures) We define the smooth map $f : \Delta \in \text{Diag}^+(n) \mapsto d^{\text{AI}}(I_n, \Delta C \Delta)^2 = \text{tr}(\log(\Delta C \Delta)^2)$. It gives the affine-invariant squared distance between I_n and all points of the fiber $\text{Cor}^{-1}(C) = \{\Delta C \Delta \in \text{Sym}^+(n) | \Delta \in \text{Diag}^+(n)\}$.

1. The smooth map f has a global minimizer.
2. For all $\Sigma \in \text{Sym}^+(n)$, there exists $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta \Sigma \Delta) \in \text{Hol}(n)$.
3. The following conjectures are equivalent for all $C \in \text{Cor}^+(n)$.
 - (i) There exists a unique $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta C \Delta) \in \text{Hol}(n)$ (Conjecture 8.11).
 - (ii) There exists a unique $X \in \text{Hol}(n)$ such that $\text{Exp}_{I_n}^{\text{QA}}(X) = C$.
 - (iii) There exists a unique local minimizer of the smooth map f , which is actually the global minimizer ensured by statement 1.
4. The previous conjectures imply the uniqueness of the quotient-affine logarithm at I_n .

See the proof of Theorem 8.13 in Section 11.7.

Otherwise said, Conjecture 8.11 is stronger than the conjecture stating the uniqueness of the quotient-affine logarithm at I_n . On the one hand, this could provide a new path to prove the latter. However, Conjecture 8.11 seems difficult to prove because the manifold $\exp(\text{Hol}(n))$ is hard to describe in terms of properties on the coefficients of the matrices. Thus it is difficult to determine whether its intersection with the fiber $\text{Diag}^+(n) \star \Sigma = \text{Cor}^{-1}(\text{Cor}(\Sigma))$ is reduced to one point or not. On the other hand, this could also help to show that Conjecture 8.11 is false. Indeed, the quotient-affine metric has both positive and negative curvature so the quotient-affine logarithm might not be unique. Hence, this seems to be a difficult problem.

8.3.2 Why $\mathcal{V}^* = \text{Row}_0(n)$ seems to be a better choice

Nevertheless, another interesting decomposition of symmetric matrices where each subspace is stable by permutations is given by $\text{Sym}(n) = \text{Row}_0(n) \oplus \text{Diag}(n)$, where $\text{Row}_0(n) = \{S \in \text{Sym}(n) | S\mathbf{1} = 0\}$ is the vector space of symmetric matrices with null row sums. That is why we propose to examine $\mathcal{V}^* = \text{Row}_0(n)$. In the following theorem, we show that $\exp(\text{Row}_0(n))$ has a nice form.

Theorem 8.14 ($\exp : \text{Row}_0(n) \longrightarrow \text{Row}_{\mathbf{1}}^+(n)$ is a smooth diffeomorphism) The symmetric matrix logarithm is a smooth diffeomorphism from $\text{Row}_{\mathbf{1}}^+(n) = \{\Sigma \in \text{Sym}^+(n) | \Sigma\mathbf{1} = \mathbf{1}\}$ onto $\text{Row}_0(n) = \{S \in \text{Sym}(n) | S\mathbf{1} = 0\}$.

Proof. It is clear that $\exp(\text{Row}_0(n)) \subset \text{Row}_{\mathbf{1}}^+(n)$ since if $S \in \text{Row}_0(n)$, then $\exp(S)\mathbf{1} = \sum_{k=0}^{\infty} \frac{1}{k!} S^k \mathbf{1} = \mathbf{1}$. Conversely, let $\Sigma \in \text{Row}_{\mathbf{1}}^+(n)$. Then the Lagrange polynomial $P(X) =$

$\sum_{\lambda \in \text{eig}(\Sigma)} \log(\lambda) \prod_{\mu \in \text{eig}(\Sigma), \mu \neq \lambda} \frac{\lambda+1-\mu}{\lambda-\mu}$ satisfies $P(\Sigma - I_n) = \log \Sigma$. Since $1 \in \text{eig}(\Sigma)$, $P(0) = \log(1) + \sum_{\lambda \neq 1} \log(\lambda) \frac{1-\lambda}{\lambda-1} \prod_{\mu \neq \lambda, 1} \frac{1-\mu}{\lambda-\mu} = 0$. Hence $\log(\Sigma)\mathbf{1} = P(\Sigma - I_n)\mathbf{1} = P(0)\mathbf{1} = 0$. So $\log(\text{Row}_{\mathbf{1}}^+(n)) \subset \text{Row}_0(n)$. Finally, $\text{Row}_{\mathbf{1}}^+(n) = \exp(\text{Row}_0(n))$ so $\log : \text{Row}_{\mathbf{1}}^+(n) \rightarrow \text{Row}_0(n)$ is a smooth diffeomorphism. \square

Hence, the question becomes: for all $\Sigma \in \text{Sym}^+(n)$, does there exist a unique $\Delta \in \text{Diag}^+(n)$ such that $\Delta\Sigma\Delta \in \text{Row}_{\mathbf{1}}^+(n)$? The answer is yes [Marshall and Olkin, 1968, Johnson and Reams, 2009], let us explain why.

8.3.3 The conjecture is true with $\mathcal{V}^* = \text{Row}_0(n)$

We recall that we denote $\star : \text{Diag}^+(n) \times \text{Sym}^+(n) \rightarrow \text{Sym}^+(n)$ the congruence action of positive diagonal matrices on SPD matrices.

Theorem 8.15 (Definition of \mathcal{D}^*) For all $\Sigma \in \text{Sym}^+(n)$, there exists a unique $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta\Sigma\Delta) \in \text{Row}_0(n)$ [Marshall and Olkin, 1968, Johnson and Reams, 2009]. This allows to define:

- the surjective map $\mathcal{D}^* : \Sigma \in \text{Sym}^+(n) \mapsto \Delta \in \text{Diag}^+(n)$,
- the surjective map $\pi^* : \Sigma \in \text{Sym}^+(n) \mapsto \log(\mathcal{D}^*(\Sigma) \star \Sigma) \in \text{Row}_0(n)$ which is invariant under the congruence group action of $\text{Diag}^+(n)$ on $\text{Sym}^+(n)$,
- the bijective map $\text{Log}^* = \pi^*_{|\text{Cor}^+(n)} : \text{Cor}^+(n) \rightarrow \text{Row}_0(n)$ that we call the *log-scaling* (note that $\pi^* = \text{Log}^* \circ \text{Cor}$),
- the smooth bijective inverse map $\text{Exp}^* = (\text{Log}^*)^{-1} = \text{Cor} \circ \exp : \text{Row}_0(n) \rightarrow \text{Cor}^+(n)$.

Proof of Theorem 8.15 (Definition of \mathcal{D}^).* The existence and uniqueness are due to [Marshall and Olkin, 1968], the uniqueness has been proved differently later in [Johnson and Reams, 2009]. Before explaining the proof of existence which is important to compute the solution numerically, let us explain the properties of the functions \mathcal{D}^* , π^* , Log^* and Exp^* .

- The map \mathcal{D}^* is surjective because $\mathcal{D}^*(\Delta^{-2}) = \Delta$ for all $\Delta \in \text{Diag}^+(n)$.
- The map π^* is surjective because if $S \in \text{Row}_0(n)$, then $\pi^*(\exp(S)) = S$.
- The map Log^* is surjective because $\text{Log}^*(\text{Cor}(\exp(S))) = S$ and injective because if $C, C' \in \text{Cor}^+(n)$ are such that $\text{Log}^*(C) = \text{Log}^*(C')$, then $\mathcal{D}^*(C) \star C = \mathcal{D}^*(C') \star C'$ so $C = \text{Cor}(\mathcal{D}^*(C) \star C) = \text{Cor}(\mathcal{D}^*(C') \star C') = C'$.
- We just showed that $(\text{Log}^*)^{-1} = \text{Cor} \circ \exp$ so $\text{Exp}^* = (\text{Log}^*)^{-1}$ is bijective and smooth.

The existence of $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta\Sigma\Delta) \in \text{Row}_0(n)$ comes from the following more general result from Marshall and Olkin [Marshall and Olkin, 1968] that we explain below.

Theorem 8.16 (Existence of a scaling for prescribed row sum) [Marshall and Olkin, 1968] Let $S \in \text{Sym}(n)$ and $\mu \in (\mathbb{R}^+)^n$. We define $f : x \in (\mathbb{R}^+)^n \mapsto x^\top S x \in \mathbb{R}$. We denote $\mathcal{M}^\mu = \{x \in (\mathbb{R}^+)^n \mid \prod_{i=1}^n x_i^{\mu_i} = 1\}$ which is a smooth hypersurface of \mathbb{R}^n . We assume that $f|_{\mathcal{M}^\mu}$ is:

1. (bounded away from 0) $\exists c > 0, \forall x \in \mathcal{M}^\mu, |f(x)| \geq c$,
2. (unbounded above when $\|x\| \rightarrow \infty$) $\forall C \geq 0, \exists M \geq 0, \forall x \in \mathcal{M}^\mu, \|x\| > M \Rightarrow f(x) > C$.

Then there exists $\Delta = \Delta_\mu \in \text{Diag}^+(n)$ such that $\Delta S \Delta \mathbf{1} = \mu$.

Theorem 8.16 implies the existence in Theorem 8.15 with $S = \Sigma \in \text{Sym}^+(n)$. Indeed, for all $x \in \mathcal{M}^\mu$, $f(x) \geq \lambda_{\min} \|x\|^2$ where $\lambda_{\min} = \min \text{eig}(\Sigma) > 0$ so Assumption 2 is satisfied. Moreover, there exists $\varepsilon > 0$ such that the ball $B(0, \varepsilon)$ satisfies $B(0, \varepsilon) \cap \mathcal{M}^\mu = \emptyset$ so if $x \in \mathcal{M}^\mu$, then $\|x\| \geq \varepsilon$. Thus for all $x \in \mathcal{M}^\mu$, $f(x) \geq \lambda_{\min} \varepsilon^2 > 0$ so Assumption 1 is satisfied. These assumptions are actually satisfied for any strictly copositive matrix [Marshall and Olkin, 1968, Theorem 1]. \square

Detailed proof of Theorem 8.16. We follow the path sketched in [Marshall and Olkin, 1968] giving more details. The proof consists in rewriting the equation $\Delta S \Delta \mathbf{1} = \mu$ into a minimization problem:

$$S\delta - \left(\frac{\mu_1}{\delta_1}, \dots, \frac{\mu_n}{\delta_n} \right)^\top = 0 \quad (8.11)$$

where $\delta = \Delta \mathbf{1} = (\delta_1, \dots, \delta_n)^\top$. Now we can easily find a function whose gradient is (twice) the left side of this equation, namely:

$$F(\Delta) = \delta^\top S \delta - 2 \left(\sum_{i=1}^n \mu_i \log(\delta_i) - \log(c) \right), \quad (8.12)$$

where $c > 0$ is a constant. Equivalently, if we denote $m = \sum_{i=1}^n \mu_i$, $x_i = c^{-1/m} \delta_i$ and $\lambda = c^{-2/m}$, then the following function has the same gradient up to a scaling factor:

$$G(\Delta) = \lambda F(\Delta) = x^\top S x - 2\lambda \sum_{i=1}^n \mu_i \log(x_i). \quad (8.13)$$

This function can be seen as the Lagrangian $\mathcal{L}(x, \lambda)$ of the minimization problem $\min_{x \in \mathcal{M}^\mu} f(x) = \min f|_{\mathcal{M}^\mu}$. Therefore, we prove that $f|_{\mathcal{M}^\mu}$ reaches its infimum at a certain $x^* \in \mathcal{M}^\mu$ (thanks to Assumption 2) and that this provides a solution $\Delta^* \in \text{Diag}^+(n)$ to our problem (thanks to Assumption 1).

To prove that $f|_{\mathcal{M}^\mu}$ reaches its infimum, it suffices to prove that the level sets $\mathcal{S}_{f|_{\mathcal{M}^\mu}}(a) = \{x \in \mathcal{M}^\mu \mid f(x) \leq a\}$ are compact. Indeed, the level sets $\mathcal{S}_{f|_{\mathcal{M}^\mu}}(a) = f|_{\mathcal{M}^\mu}^{-1}(] - \infty, a])$ are closed in \mathcal{M}^μ since $f|_{\mathcal{M}^\mu}$ is continuous on \mathcal{M}^μ and $] - \infty, a]$ is closed in \mathbb{R} . As \mathcal{M}^μ is closed in \mathbb{R}^n , $\mathcal{S}_{f|_{\mathcal{M}^\mu}}(a)$ is closed in \mathbb{R}^n . The level sets are bounded because according to Assumption 2, there exists $M \geq 0$ such that for all $x \in \mathcal{S}_{f|_{\mathcal{M}^\mu}}(a)$, $\|x\| \leq M$. Hence the level sets are compact. Since at least one of them is non-empty and f is continuous on this compact set, it reaches its minimum at a certain $x^* \in \mathcal{M}^\mu$.

According to the Lagrange multiplier theorem, there exists a unique λ^* such that (x^*, λ^*) is a critical point of the Lagrangian. Since $\nabla \mathcal{L}(x, \lambda) = (2Sx - 2\lambda(\frac{\mu_1}{x_1}, \dots, \frac{\mu_n}{x_n}), -2 \sum_{i=1}^n \mu_i \log(x_i))$, the critical point (x^*, λ^*) has to satisfy $\lambda^* = \frac{f(x^*)}{m}$. Thus according to Assumption 1, we have $\lambda^* \neq 0$. Hence it provides a solution $\Delta^* = \frac{1}{\sqrt{\lambda^*}} \text{diag}(x^*)$ to Equation (8.11). \square

This demonstration not only allows to prove the result but also shows that $\mathcal{D}^*(\Sigma) = \frac{1}{\sqrt{\lambda^*}} \text{diag}(x^*)$ with $x^* = \min_{\mathcal{M}} f$ and $\lambda^* = \frac{f(x^*)}{n}$ where $f : x \in (\mathbb{R}^+)^n \mapsto x^\top \Sigma x \in \mathbb{R}^+$ and $\mathcal{M} = \mathcal{M}^{\mathbf{1}} = \{x \in (\mathbb{R}^+)^n \mid \prod_{i=1}^n x_i = 1\}$. We use this fact in Section 8.5 to compute $\mathcal{D}^*(\Sigma)$.

8.3.4 Properties of the log-scaling bijection

Let us give properties of Log^* that are analogous to the properties of Log .

Theorem 8.17 (Properties of the log-scaling bijection) The log-scaling bijection satisfies the following properties.

1. (Equivariance) Log^* and Exp^* are equivariant under permutations.
2. (Equicorrelation) For all $\rho \in (-\frac{1}{n-1}, 1)$, $\text{Log}^*(C(\rho)) = \frac{1}{n} \ln \left(\frac{1+(n-1)\rho}{1-\rho} \right) (\mathbf{1}\mathbf{1}^\top - nI_n)$. In dimension $n = 2$, $\text{Log}^*(C(\rho)) = \begin{pmatrix} -F(\rho) & F(\rho) \\ F(\rho) & -F(\rho) \end{pmatrix}$ where $F(\rho) = \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right) \in \mathbb{R}$ is the Fisher transformation of the correlation coefficient $\rho \in (-1, 1)$.
3. (Block equicorrelation matrix) If C is a block equicorrelation matrix of signature $I = \{i_1, \dots, i_p\}$, then $\text{Log}^*(C)$ is a block symmetric matrix with null row sum of signature I with diagonal blocks of the form $(\alpha_j - \beta_j)I_{i_j} + \beta_j \mathbf{1}_{i_j} \mathbf{1}_{i_j}^\top$ and off-diagonal blocks of the form $\beta_{jk} \mathbf{1}_{i_j} \mathbf{1}_{i_k}^\top$.
4. (Generalization) For all $x \in (\mathbb{R}^+)^n$, for all $\Sigma \in \text{Sym}^+(n)$, there exists a unique $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta \Sigma \Delta)x = 0$.

See the proof of Theorem 8.17 in Section 11.7.

The log-scaling bijection has an additional property: it is compatible with the cor-inversion. It is a corollary of Theorem 8.2, the key reason being the use of the congruence action of positive diagonal matrices instead of the additive action of diagonal matrices.

Theorem 8.18 (Compatibility between inversion and log-scaling bijection) For all $C \in \text{Cor}^+(n)$, $\text{Log}^*(\mathcal{I}(C)) = -\text{Log}^*(C)$. Otherwise said, the following diagram commutes.

$$\begin{array}{ccc} \text{Cor}^+(n) & \xrightarrow{\mathcal{I}} & \text{Cor}^+(n) \\ \downarrow \text{Log}^* & & \downarrow \text{Log}^* \\ \text{Row}_0(n) & \xrightarrow{-\text{Id}} & \text{Row}_0(n) \end{array} \quad (8.14)$$

In this section, we proved the existence in the conjecture with $\mathcal{V}^* = \text{Hol}(n)$ and we proved the conjecture for $\mathcal{V}^* = \text{Row}_0(n)$. This provides a bijective map $\text{Log}^* : \text{Cor}^+(n) \rightarrow \text{Row}_0(n)$ called the log-scaling bijection.

8.4 Permutation-invariant log-Euclidean metrics via the log-scaling bijection

In this section, we use the log-scaling bijection to define log-Euclidean metrics on $\text{Cor}^+(n)$. More precisely, in Section 8.4.1, we prove that the log-scaling bijection is a diffeomorphism. In Section 8.4.2, we characterize all permutation-invariant inner products on $\text{Row}_0(n)$. In Section 8.4.3, we define permutation-invariant log-Euclidean metrics by pullback and we give their geometric properties.

8.4.1 The cor-exp bijection is a diffeomorphism

Theorem 8.19 ($\text{Exp}^* = \text{Cor} \circ \text{exp}$ is a diffeomorphism) The log-scaling bijection $\text{Log}^* : \text{Cor}^+(n) \rightarrow \text{Row}_0(n)$ is a smooth diffeomorphism. We give the differentials of Log^* and Exp^* in function of the differentials of the symmetric matrix logarithm and exponential maps \log and exp . For all $C \in \text{Cor}^+(n)$, $S, Y \in \text{Row}_0(n)$ and $X \in \text{Hol}(n)$ such that $\Sigma = \mathcal{D}^*(C) \star C = \text{exp}(S)$:

$$d_S \text{Exp}^*(Y) = \Delta^{-1} \left[d_S \text{exp}(Y) - \frac{1}{2} (\Delta^{-2} \text{Diag}(d_S \text{exp}(Y)) \Sigma + \Sigma \text{Diag}(d_S \text{exp}(Y)) \Delta^{-2}) \right] \Delta^{-1}, \quad (8.15)$$

$$d_C \text{Log}^*(X) = d_\Sigma \log \left(\Delta X \Delta + \frac{1}{2} (X^0 \Sigma + \Sigma X^0) \right), \quad (8.16)$$

where $\Delta = \text{Diag}(\Sigma)^{1/2}$ and $X^0 = -2 \text{diag}((I_n + \Sigma)^{-1} \Delta X \Delta \mathbf{1})$.

See the proof of Theorem 8.19 in Section 11.7.

8.4.2 Permutation-invariant inner products on $\text{Row}_0(n)$

Theorem 8.20 (Permutation-invariant inner products on $\text{Row}_0(n)$) For $n \geq 4$, permutation-invariant inner products on $\text{Row}_0(n)$ are the symmetric bilinear forms associated to the following positive definite quadratic forms q^* defined for $Y \in \text{Row}_0(n)$:

$$q^*(Y) = \alpha \text{tr}(Y^2) + \delta \text{tr}(\text{Diag}(Y)^2) + \zeta \text{tr}(Y)^2, \quad (8.17)$$

with $\alpha > 0$, $n\alpha + (n-2)\delta > 0$ and $n\alpha + (n-1)(\delta + n\zeta) > 0$. For $n = 3$, the permutation-invariant inner products have the same form with $\alpha = 0$. For $n = 2$, they have the same form with $\alpha = \delta = 0$.

See the proof of Theorem 8.20 in Section 11.7.

8.4.3 Pullback metrics via the log-scaling diffeomorphism

Definition 8.21 (Log-scaled metrics) An log-scaled metric on $\text{Cor}^+(n)$ is the pullback metric of an inner product characterized by a quadratic form q^* as in Definition 8.20. For all $C \in \text{Cor}^+(n)$ and $X \in T_C \text{Cor}^+(n) = \text{Hol}(n)$, it writes $g_C^*(X, X) = q^*(d_C \text{Log}^*(X))$ where $d_C \text{Log}^*(X) = d_C \log(X + \frac{1}{2}(X^0 C + C X^0))$ with $X^0 = -2 \text{diag}((I_n + C)^{-1} X \mathbf{1})$.

Theorem 8.22 (Riemannian operations of log-scaled metrics) We consider a log-scaled metric characterized by the quadratic form q^* . Let $C, C', C_1, \dots, C_n \in \text{Cor}^+(n)$, $X \in \text{Row}_0(n)$. The Riemannian operations of this metric are summarized in Table 8.6. Moreover, the metric is permutation-invariant and inverse-consistent, i.e. it is invariant under the pullback by the cor-inversion $\mathcal{I} : \text{Cor}^+(n) \rightarrow \text{Cor}^+(n)$.

Beware not to confuse the Riemannian maps $\text{Exp}_C^* : T_C \text{Cor}^+(n) = \text{Hol}(n) \rightarrow \text{Cor}^+(n)$ and $\text{Log}_C^* : \text{Cor}^+(n) \rightarrow T_C \text{Cor}^+(n) = \text{Hol}(n)$ with the diffeomorphisms $\text{Exp}^* : \text{Row}_0(n) \rightarrow \text{Cor}^+(n)$ and $\text{Log}^* : \text{Cor}^+(n) \rightarrow \text{Row}_0(n)$.

Since one motivation behind the off-log bijection in [Archakov and Hansen, 2021] was the closed form expression in dimension 2 (cf. Theorem 8.4.2), it seems important to note the following result.

Exponential map	$\text{Exp}_C^*(X) = \text{Exp}^*(\text{Log}^*(C) + d_C \text{Log}^*(X))$
Logarithm map	$\text{Log}_C^*(C') = d_{\text{Log}^*(C)} \text{Exp}^*(\text{Log}^*(C') - \text{Log}^*(C))$
Geodesic	$\gamma^*(t) = \text{Exp}^*((1-t)\text{Log}^*(C) + t\text{Log}^*(C'))$
Squared distance	$d^*(C, C')^2 = q^*(\text{Log}^*(C') - \text{Log}^*(C))$
Fréchet mean	$C^* = \text{Exp}^*(\frac{1}{k} \sum_{i=1}^k \text{Log}^*(C_i))$
Curvature	$R^* = 0$
Parallel transport	$\Pi_{C \rightarrow C'}^* X = (d_{C'} \text{Log}^*)^{-1}(d_C \text{Log}^*(X))$

Table 8.6: Riemannian operations of log-scaled metrics.

Theorem 8.23 (Coincidence of the metrics in dimension 2) In dimension 2, up to a positive scaling factor, the quotient-affine metric, the off-log metric and the log-scaled metric coincide. We recall that the Fisher transformation is the increasing map $F : \rho \in (-1, 1) \mapsto \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right) \in \mathbb{R}^+$. Let $C = C(\rho)$ and $X = \begin{pmatrix} 0 & x \\ x & 0 \end{pmatrix}$ with $\rho \in (-1, 1)$ and $x \in \mathbb{R}$. Then:

1. (Metric) $g_C(X, X) = \frac{x^2}{(1-\rho^2)^2}$ (up to a scaling factor $\alpha > 0$),
2. (Geodesic) $\gamma(t) = C(\rho(t))$ where $\rho(t) = \frac{\rho_1 \cosh(\lambda t) + \sinh(\lambda t)}{\rho_1 \sinh(\lambda t) + \cosh(\lambda t)}$ with $\lambda = F(\rho_2) - F(\rho_1)$ is monotonic (increasing if $\rho_1 < \rho_2$, decreasing if $\rho_1 > \rho_2$, constant if $\rho_1 = \rho_2$),
3. (Distance) $d(C_1, C_2) = |\lambda| = |F(\rho_2) - F(\rho_1)|$ (up to a scaling factor $\sqrt{\alpha}$).

See the proof of Theorem 8.23 in Section 11.7.

We showed in this section that the log-scaling bijection is a diffeomorphism. Therefore it provides a family of permutation-invariant inverse-consistent log-Euclidean metrics by pulling back permutation-invariant inner products on $\text{Row}_0(n)$. Thus, the Riemannian operations are trivial. As shown in Table 8.6, they are known in closed form modulo the computation of $\text{Log}^* : \text{Cor}^+(n) \rightarrow \text{Row}_0(n)$ or equivalently the computation of $\mathcal{D}^* : \text{Sym}^+(n) \rightarrow \text{Diag}^+(n)$. That is the goal of the next section.

8.5 Numerical computation of the log-scaling

The topic of scaling a matrix to prescribed row and column sums is much broader than the use we make of it in this chapter. The research on this topic in linear algebra started in the 1960's with the Sinkhorn theorem stating that squared positive matrices $A \in \text{Mat}(n)$, i.e. with positive entries $A_{ij} > 0$, admit a bistochastic scaling $D_1 A D_2$ where D_1, D_2 are positive diagonal matrices. After several extensions of this result, the novelty of [Marshall and Olkin, 1968] was to gather three ingredients: to tackle symmetric matrices $A \in \text{Sym}(n)$, to impose $D_1 = D_2$ and to abandon the positivity of the coefficients. As recalled in Theorem 8.16, they proved the sufficiency of two hypotheses which are in particular satisfied for positive definite matrices and even for strictly copositive matrices. See the introduction of [Johnson and Reams, 2009] for a brief survey or [Idel, 2016] for a thorough one.

According to [Allen-Zhu et al., 2017], this problem actually appeared in different scientific communities from telephone traffic computation to probability theory. Moreover, the results

have been applied to many fields from image reconstruction to computer science. Thus, the question of the computation of the scaling has always been a burning topic. Examining the literature, there seems to exist very few algorithms specifically designed for SPD matrices with respect to positive matrices or other classes of matrices. That is surprising since the proof of Theorem 8.16 [Marshall and Olkin, 1968] suggests that the positive definiteness could improve significantly the performance of an optimization algorithm since the map to minimize $f : x \in \mathbb{R}^n \mapsto x^\top \Sigma x \in \mathbb{R}^+$ is strictly convex. We can cite [O’Leary, 2003] with polynomial complexity (without more information) and [Khachiyan and Kalantari, 1992] with time complexity $O(\sqrt{n} \ln(n) + \ln(\ln(\frac{1}{\varepsilon})))$ via a Newton method (as proved in [Kalantari, 2005, Theorem 8.4 SP]), where ε is the desired precision. For the latter, each step requires to solve a linear system, which is $O(n^2)$, so the total complexity is $O(n^{5/2} \ln(n) + n^2 \ln(\ln(\frac{1}{\varepsilon})))$. As for generic methods, the best one seems to be from [Allen-Zhu et al., 2017] where the total complexity is $O(m + n^{1/4})$ where m is the number of non-zero entries in the matrix.

In this section, we provide an algorithm to compute the scaling of an SPD matrix with precision ε , therefore to compute $\mathcal{D}^* : \text{Sym}^+(n) \rightarrow \text{Diag}^+(n)$ and $\text{Log}^* : \text{Cor}^+(n) \rightarrow \text{Row}_0(n)$, with complexity $O(m \ln(\frac{n}{\varepsilon}) + n \ln(\frac{n}{\varepsilon}) \ln(\ln(\frac{n}{\varepsilon})))$ which seems quite competitive. In Section 8.5.1, we explain the convex optimization problem we need to solve to compute \mathcal{D}^* , based on the proof of Theorem 8.16 [Marshall and Olkin, 1968]. In Section 8.5.2, we explain that this kind of problem can be solved with logarithmic time complexity by Projected Gradient Descent (PGD) if one knows the projection [Bubeck, 2015]. Since we don’t have the projection in closed form, we propose in Section 8.5.3 an efficient algorithm to compute it. Finally in Section 8.5.4, we modify the PGD algorithm with the projection approximation and we show that it still converges with logarithmic time complexity.

8.5.1 The optimization problem

From the proof of Theorem 8.16, we know that $\mathcal{D}^*(\Sigma) = \frac{1}{\sqrt{\lambda^*}} \text{diag}(x^*)$ with $x^* = \min_{\mathcal{M}} f$ and $\lambda^* = \frac{f(x^*)}{n}$ where $f : x \in (\mathbb{R}^+)^n \mapsto x^\top \Sigma x \in \mathbb{R}^+$ and $\mathcal{M} = \{x \in (\mathbb{R}^+)^n \mid \prod_{i=1}^n x_i = 1\}$. Since the map f is a positive definite quadratic form, it is natural to think about convex optimization to compute x^* . However, this map has to be minimized on the set \mathcal{M} which is not convex. In the following result, we show that x^* is also the minimum of f on the convex hull of \mathcal{M} , which is $\Omega = \{x \in (\mathbb{R}^+)^n \mid \prod_{i=1}^n x_i \geq 1\}$.

Theorem 8.24 (Optimization problem to compute the scaling) $\arg \min_{\Omega} f \in \mathcal{M}$ and $\Omega = \text{conv}(\mathcal{M})$. Thus x^* is the global minimizer of the convex map f on the closed convex set Ω . See the proof of Theorem 8.24 in Section 11.7.

8.5.2 Optimizing an α -strongly convex β -smooth map on a closed convex set

There are many algorithms to optimize convex maps on closed convex sets, as described in [Bubeck, 2015]. The optimal algorithm depends on the properties of the map. The simplest one is the Projected Gradient Descent (PGD): it consists in alternating an unconstrained gradient descent and a projection onto the convex set of constraints. More precisely, if $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a map to minimize on the closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$, the algorithm takes

$x_0 \in \mathcal{C}$, computes $y_{k+1} = x_k - \eta \nabla f|_{x_k}$ and $x_{k+1} = \Pi_{\mathcal{C}} y_{k+1}$ where $\eta > 0$ is a fixed parameter and $\Pi_{\mathcal{C}}$ is the projection onto the closed convex set \mathcal{C} . The PGD algorithm is shown to converge with logarithmic time complexity when the map is α -strongly convex and β -smooth. [Bubeck, 2015]

In this section, we recall the definitions of an α -strongly convex map and a β -smooth map and we show that $f : x \mapsto x^\top \Sigma x$ satisfies these two properties. Then we recall the definition of the projection onto a closed convex set \mathcal{C} and the theorem of convergence of the PGD algorithm for an α -strongly convex and β -smooth map.

8.5.2.1 α -strongly convex β -smooth maps

Definition 8.25 (α -strongly convex map) [Bubeck, 2015] Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex set. We say that the map $f : \mathcal{C} \rightarrow \mathbb{R}$ is α -strongly convex if one of the following equivalent assertions is satisfied:

- (i) $f((1-t)x_1 + tx_2) \leq (1-t)f(x_1) + tf(x_2) - \frac{\alpha}{2}t(1-t)\|x_1 - x_2\|^2$ for all $t \in [0, 1]$, for all $x_1, x_2 \in \mathcal{C}$,
- (ii) the function $f - \frac{\alpha}{2}\|\cdot\|^2$ is convex.

If f is differentiable, then these assertions are equivalent to a third one:

- (iii) $f(x_1) - f(x_2) \leq \nabla f|_{x_1}^\top (x_1 - x_2) - \frac{\alpha}{2}\|x_1 - x_2\|^2$ for all $x_1, x_2 \in \mathcal{C}$.

Definition 8.26 (β -smooth map) [Bubeck, 2015] We say that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is β -smooth if f is of class \mathcal{C}^1 and its gradient is β -Lipschitz, i.e. for all $x_1, x_2 \in \mathbb{R}^n$, we have $\|\nabla f|_{x_1} - \nabla f|_{x_2}\| \leq \beta\|x_1 - x_2\|$.

We state an important inequality satisfied by β -smooth maps from [Bubeck, 2015], that we use later.

Lemma 8.27 (Important result on β -smooth maps) [Bubeck, 2015] Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a β -Lipschitz map. Then for all $x_1, x_2 \in \mathbb{R}^n$, $f(x_1) - f(x_2) - \nabla f|_{x_2}^\top (x_1 - x_2) \leq \frac{\beta}{2}\|x_1 - x_2\|^2$.

Now, we state that $f : x \mapsto x^\top \Sigma x$ satisfies these two properties.

Lemma 8.28 (The map f is $2\lambda_{\min}$ -strongly convex and $2\lambda_{\max}$ -smooth) Let $\Sigma \in \text{Sym}^+(n)$. We denote $\lambda_{\min}, \lambda_{\max}$ the minimum and maximum eigenvalues of Σ . The map $f : x \in \mathbb{R}^n \mapsto x^\top \Sigma x \in \mathbb{R}$ is $2\lambda_{\min}$ -strongly convex and $2\lambda_{\max}$ -smooth.

Proof. The map $f - \lambda_{\min}\|\cdot\|^2 : x \in \mathbb{R}^n \mapsto x^\top (\Sigma - \lambda_{\min}I_n)x$ is convex because $\Sigma - \lambda_{\min}I_n$ is symmetric positive semi-definite. So f is $2\lambda_{\min}$ -strongly convex. Moreover, $\nabla f|_x = 2\Sigma x$ so $\|\nabla f|_{x_1} - \nabla f|_{x_2}\| = 2\|\Sigma(x_1 - x_2)\| \leq 2\lambda_{\max}\|x_1 - x_2\|$. Hence, f is $2\lambda_{\max}$ -smooth. \square

8.5.2.2 Projection and Projected Gradient Descent

We recall the definition of the projection onto a closed convex set and the theorem of convergence of the Projected Gradient Descent for α -strongly convex β -smooth maps.

Definition 8.29 (Projection onto a closed convex set) Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a closed convex set. For all $z \in \mathbb{R}^n$, there exists a unique $x \in \mathcal{C}$ satisfying the following equivalent conditions:

- (i) for all $y \in \mathcal{C}$, $\|z - x\| \leq \|z - y\|$,
- (ii) for all $y \in \mathcal{C}$, $\langle z - x | y - x \rangle \leq 0$.

This point $x \in \mathcal{C}$ is called the projection of z onto \mathcal{C} and denoted $\Pi_{\mathcal{C}}(z)$. It defines a 1-Lipschitz map $\Pi_{\mathcal{C}} : \mathbb{R}^n \rightarrow \mathcal{C}$ which satisfies $(\Pi_{\mathcal{C}})|_{\mathcal{C}} = \text{Id}_{\mathcal{C}}$ and $\Pi_{\mathcal{C}}(\mathbb{R}^n \setminus \mathcal{C}) \subseteq \partial\mathcal{C}$.

Theorem 8.30 (Projected gradient descent for α -strongly convex β -smooth maps) [Bubeck, 2015] Let $f : \mathcal{C} \rightarrow \mathbb{R}$ be a convex map on a convex set \mathcal{C} . We assume that f is α -strongly convex and β -smooth and we denote $\kappa = \beta/\alpha$. Let $x_0 \in \mathcal{C}$. We define $y_{k+1} = x_k - \frac{1}{\beta} \nabla f|_{x_k}$ and $x_{k+1} = \Pi_{\mathcal{C}}(y_{k+1})$. Then $(x_k)_{k \in \mathbb{N}}$ converges to $x^* = \arg \min_{\mathcal{C}} f$. Moreover, $\|x_k - x^*\| \leq \exp(-\frac{k}{2\kappa}) \|x_0 - x^*\|$ for all $k \in \mathbb{N}$.

Thus, this theorem provides a very efficient algorithm to compute $\arg \min_{\Omega} f$, and thus \mathcal{D}^* , if we know the projection onto $\Omega = \{x \in (\mathbb{R}^+)^n \mid \prod_{i=1}^n x_i \geq 1\}$. However, we don't know it in closed form. Hence, we have to approximate it and to show that this approximation preserves the previous theorem, i.e. that the new algorithm with approximated projection still converges, hopefully with logarithmic time complexity. That is the goal of the two next sections.

8.5.3 Approximate the projection

In this section, we define an approximation of the projection $\Pi_{\Omega} : \mathbb{R}^n \rightarrow \Omega$ which is not generic but very specific to the space $\Omega = \{x \in (\mathbb{R}^+)^n \mid \prod_{i=1}^n x_i \geq 1\}$.

We consider the hypersurfaces $\mathcal{M}_{\alpha} = \partial\Omega_{\alpha} = \{x \in (\mathbb{R}^+)^n \mid \prod_{i=1}^n x_i = \alpha\}$ for $\alpha \in \mathbb{R}^+$. Our method to project $z \in \mathbb{R}^n \setminus \Omega$ onto Ω consists in defining a map $F_z : a \in \mathbb{R}^+ \mapsto x(a) \in \mathcal{M}_{\alpha(a)}$ such that $\alpha(a) = \prod_{i=1}^n x_i(a)$ increases with a . Thus, it suffices to do a line search on $a > 0$ to find the one such that $\alpha(a) = 1$. This is illustrated on Figure 8.1.

The tangent hyperplane at $x \in \mathcal{M}_{\alpha}$ is $T_x \mathcal{M}_{\alpha} = (\frac{1}{x})^{\perp}$ where $\frac{1}{x} = (\frac{1}{x_1}, \dots, \frac{1}{x_n})$. Thus, the orthogonal projection $x \in \mathcal{M}_{\alpha}$ of $z \notin \Omega$ onto $\Omega_{\alpha} = \text{conv}(\mathcal{M}_{\alpha})$ satisfies $x - z \parallel \frac{1}{x}$. Otherwise said, there exists $a \in \mathbb{R}^+$ such that $x - z = a \frac{1}{x}$. Given $z \notin \Omega$ and $\alpha \in \mathbb{R}^+$, it is difficult to compute $x \in \mathcal{M}_{\alpha}$ and $a \in \mathbb{R}^+$ such that $x - z = a \frac{1}{x}$. On the contrary, given $z \notin \Omega$ and $a \in \mathbb{R}^+$, it is very easy to find $\alpha \in \mathbb{R}^+$ and $x \in \mathcal{M}_{\alpha}$ such that $x - z = a \frac{1}{x}$. Indeed, x_i satisfies the equation $x_i^2 - z_i x_i - a = 0$ so $x_i = \frac{1}{2}(z_i + \sqrt{z_i^2 + 4a})$ and $\alpha = \prod_{i=1}^n x_i$.

Therefore, for $z \in \mathbb{R}^n \setminus \Omega$, we define the smooth map $F_z : a \in \mathbb{R}^+ \mapsto x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and the smooth increasing map $PF_z : a \in \mathbb{R}^+ \mapsto \prod_{i=1}^n x_i \in \mathbb{R}^+$ where $x_i = \frac{1}{2}(z_i + \sqrt{z_i^2 + 4a})$ for all $i \in \{1, \dots, n\}$. Thus, $x := F_z(a)$ and $\alpha := PF_z(a)$ satisfy $x - z = a \frac{1}{x}$ and $x \in \mathcal{M}_{\alpha}$. In other words, $x = \Pi_{\Omega_{\alpha}}(z)$ or $F_z(a) = \Pi_{\Omega_{PF_z(a)}}(z)$.

Theorem 8.31 (Approximate the projection) We use the above notations.

1. There exists a unique $a^* \in \mathbb{R}^+$ such that $PF_z(a^*) = 1$. It satisfies $F_z(a^*) = \Pi_{\Omega}(z)$ and for all $a \geq a^*$, $PF_z(a) \geq 1$ and $F_z(a) \in \Omega$.

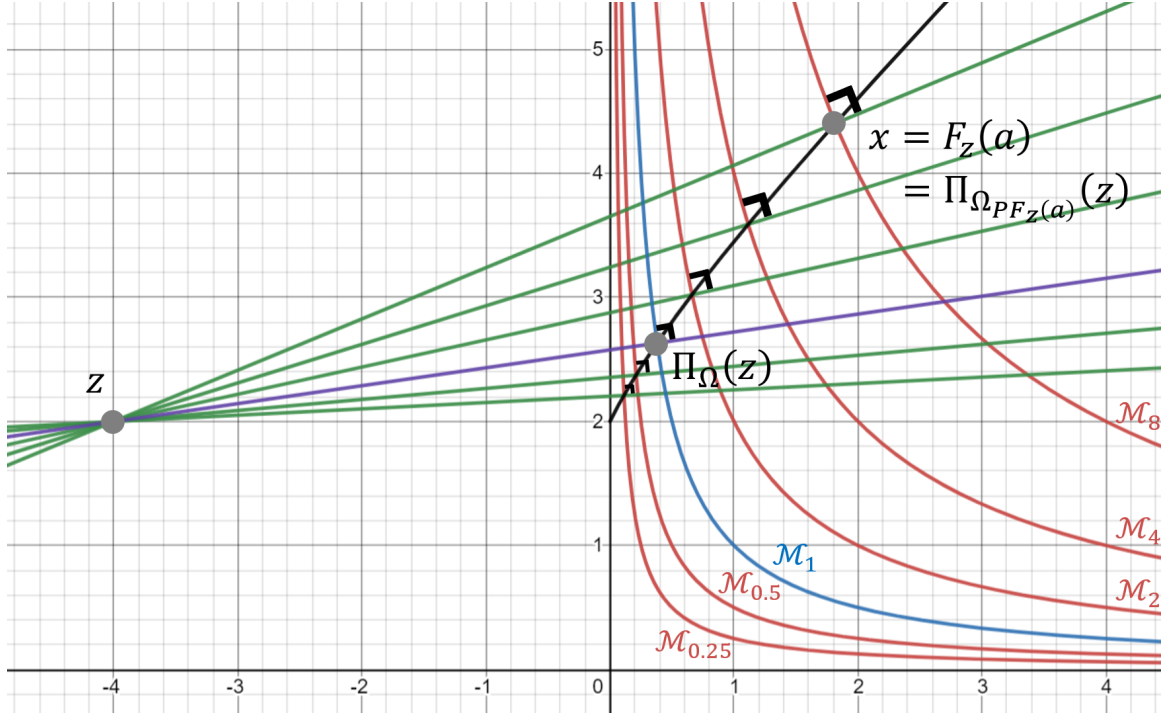


Figure 8.1: Approximation of the projection.

2. We define $I_0 = [b_0; c_0]$ where $b_0 = 0$ and $c_0 = 1 - \min_i z_i > 0$ so that $PF_z(b_0) < 1$ and $PF_z(c_0) \geq 1$. We define recursively for all $k \in \mathbb{N}$:

$$a_k = \frac{b_k + c_k}{2}, \quad (8.18)$$

$$I_{k+1} = [b_{k+1}; c_{k+1}] = \begin{cases} [b_k; a_k] & \text{if } PF_z(a_k) \geq 1 \\ [a_k; c_k] & \text{if } PF_z(a_k) < 1 \end{cases}. \quad (8.19)$$

Then $\lim_{k \rightarrow \infty} a_k = a^*$, $\lim_{k \rightarrow \infty} PF_z(a_k) = 1$ and $\lim_{k \rightarrow \infty} F_z(a_k) = \Pi_\Omega(z)$. Moreover, $|a_k - a^*| \leq c_0 2^{-k}$. The time complexity is $k_\varepsilon = \frac{1}{\ln(2)} \ln(\frac{c_0}{\varepsilon})$ and the complexity is $O(n \ln(\frac{c_0}{\varepsilon}))$.

Proof of Theorem 8.31 (Approximate the projection).

1. Note that $\lim_{a \rightarrow 0} PF_z(a) = \prod_{i=1}^n \frac{z_i + |z_i|}{2}$. It is null (thus inferior to 1) if one of the z_i 's is non-positive. Otherwise if all $z_i > 0$, it is equal to $\prod_{i=1}^n z_i < 1$ since $z \notin \Omega$. Thus, since PF_z is continuous and increasing from $\lim_{a \rightarrow 0} PF_z(a) < 1$ to $+\infty$, there exists a unique $a^* \in \mathbb{R}^+$ such that $PF_z(a^*) = 1$. Therefore, $F_z(a^*) = \Pi_{\Omega_1}(z)$ and for all $a \geq a^*$, $PF_z(a) \geq PF_z(a^*) = 1$ so $F_z(a) = \Pi_{\Omega_{PF_z(a)}}(z) \in \Omega_{PF_z(a)} \subseteq \Omega$.
2. We just proved that $PF_z(b_0) < 1$. Without loss of generality, let us assume that $\min_i z_i = z_1$. We have $z_1 < 1$ because $\prod_{i=1}^n z_i < 1$. Hence $c_0 = 1 - z_1 > 0$ and $z_1 + \sqrt{z_1^2 + 4c_0} = 1$. Since for all $i \in \{1, \dots, n\}$, $z_i + \sqrt{z_i^2 + 4c_0} \geq z_1 + \sqrt{z_1^2 + 4c_0}$, we have $PF_z(c_0) \geq 1$. Moreover, the decreasing sequence of segments $(I_k)_{k \in \mathbb{N}}$ has its diameter tending to 0 so their intersection is a singleton. Since $PF_z(a^*) = 1$, $a^* \in I_k$

for all $k \in \mathbb{N}$ so $\bigcap_{k \in \mathbb{N}} I_k = \{a^*\}$. For all $k \in \mathbb{N}$, $|a_k - a^*| \leq c_k - b_k = \frac{c_0}{2^k}$ so the sequence (a_k) converges to a^* . By continuity, $\lim_{k \rightarrow \infty} PF_z(a_k) = 1$ and $\lim_{k \rightarrow \infty} F_z(a_k) = \Pi_\Omega(z)$.

The time complexity k_ε for precision ε is given by $c_0 2^{-k_\varepsilon} = \varepsilon$, i.e. $k_\varepsilon = \frac{1}{\ln(2)} \ln(\frac{c_0}{\varepsilon})$. (We keep c_0 because $c_0 = c_0(z)$ will change at each step of the minimization algorithm since z will change.) At each step, we only compute $PF_z(a_k)$ which requires $4n + n - 1 = 5n - 1 = O(n)$ operations. Hence, the complexity is the product $O(n \ln(\frac{c_0}{\varepsilon}))$. \square

Theorem 8.32 (Stopping criterion) $\forall z \in \mathbb{R}^n \setminus \Omega, \forall \alpha > 1, \|\Pi_{\Omega_\alpha} z - \Pi_\Omega z\| \leq (\alpha^{1/n} - 1) \|\Pi_\Omega z\|$.

Proof. Let $z \in \mathbb{R}^n \setminus \Omega$. We denote $x = \Pi_\Omega z$ and $a \in \mathbb{R}^+$ such that $x - z = a \frac{x}{x}$. Thus it is equivalent to prove that for all $x \in \mathcal{M}$ and $a \in \mathbb{R}^+$, $\|\Pi_{\Omega_\alpha}(x - \frac{a}{x}) - x\| \leq (\alpha^{1/n} - 1) \|x\|$. Hence, we define the smooth maps:

$$\begin{aligned} z : a \in \mathbb{R}^+ &\mapsto x - \frac{a}{x} \in \mathbb{R}^n \setminus \Omega, \\ y : a \in \mathbb{R}^+ &\mapsto \Pi_{\Omega_\alpha}(z(a)) \in \mathcal{M}_\alpha, \\ D : a \in \mathbb{R}^+ &\mapsto \|y(a) - x\|^2 = \sum_{i=1}^n (y_i(a) - x_i)^2 \in \mathbb{R}^+. \end{aligned}$$

We want to prove that for all $a \in \mathbb{R}^+$, $D(a) \leq (\alpha^{1/n} - 1)^2 \|x\|^2$. We proceed in two steps. Firstly, we prove that $\lim_{a \rightarrow +\infty} D(a) = (\alpha^{1/n} - 1)^2 \|x\|^2$. Secondly, we prove that D is increasing. The proofs of these two facts being quite technical, we put them in Section 11.7. \square

This theorem states that given $\varepsilon > 0$, one can compute a_k 's until $1 \leq PF_z(a_k) \leq 1 + \varepsilon^n$ to ensure that $F_z(a_k) \in \Omega$ and $\|F_z(a_k) - \Pi_\Omega z\| \leq \varepsilon \|\Pi_\Omega(z)\|$. Thus we define the projection approximation as follows.

Definition 8.33 (Projection approximation) Let $\varepsilon > 0$. We define $\Pi_\Omega^\varepsilon : z \in \mathbb{R}^n \setminus \Omega \mapsto F_z(a_k) \in \Omega$ where $k = \min\{i \in \mathbb{N} | 1 \leq PF_z(a_i) \leq 1 + \varepsilon^n\}$. Note that $\Pi_\Omega^\varepsilon(z) = \Pi_{\Omega_{PF_z(a_k)}}(z)$.

Now we can use this projection approximation to modify the algorithm in Theorem 8.30 and prove the convergence of this new algorithm.

8.5.4 Approximated projection gradient descent

In this section, we define the new algorithm and we prove that it still converges at exponential rate. We recall that $\Omega = \{x \in (\mathbb{R}^+)^n | \prod_{i=1}^n x_i = 1\}$ and $f : x \in \mathbb{R}^n \mapsto x^\top \Sigma x \in \mathbb{R}$ where $\Sigma \in \text{Sym}^+(n)$. We denote $\lambda_{\min}, \lambda_{\max}$ the minimum and maximum eigenvalues of Σ .

Theorem 8.34 (Algorithm to compute \mathcal{D}^* : convergence, speed, complexity) Let $x_0 = \mathbf{1} \in \Omega$. We denote $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$ and $R = \sqrt{n\kappa}$. Let $\varepsilon \in (0, 1)$. Let $\delta = \frac{\varepsilon^2}{72n\kappa^2}$. Let $(u_k)_{k \in \mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$ is a decreasing sequence with $\sum_{k=0}^{\infty} u_k \leq 1$, e.g. $u_k = \frac{6}{\pi^2(k+1)^2}$, and $\delta_k = \delta u_k$. We define $y_{k+1} = (I_n - \frac{1}{\lambda_{\max}} \Sigma)x_k$ and $x_{k+1} = \Pi_\Omega^{\delta_{k+1}}(y_{k+1})$. Then:

1. (Convergence) $(x_k)_{k \in \mathbb{N}}$ converges to $x^* = \arg \min_{\Omega} f$.
2. (Speed of convergence) The convergence is exponential: $\|x_k - x^*\| \leq \varepsilon + \exp(-\frac{k}{2\kappa})\|x_0 - x^*\|$.
3. (Time complexity) For all $\varepsilon > 0$, the sufficient number of iterations so that $\|x_k - x^*\| < 2\varepsilon$ is $k_\varepsilon = O(\ln(\frac{n}{\varepsilon}))$.
4. (Complexity) The total complexity with precision 2ε is $O(n^2 \ln(\frac{n}{\varepsilon}) + n \ln(\frac{n}{\varepsilon}) \ln \ln(\frac{n}{\varepsilon}))$.

We recall that given x^* , we then have $\mathcal{D}^*(\Sigma) = \frac{1}{\sqrt{\lambda^*}} \text{diag}(x^*)$ where $\lambda^* = \frac{1}{n}(x^*)^\top \Sigma x^*$.

Proof. The proof of Theorem 8.30 strongly relies on the characterization of the projection onto a closed convex set (item 2 of Definition 8.29) [Bubeck, 2015]. Therefore, it is not obvious that replacing the projection by an approximation leads to the same convergence. There is actually an extra term that appears and that we need to bound carefully. We proceed in three steps:

1. We bound all what should be bounded.
2. We define the extra term and we bound it.
3. We prove the convergence.

We need to introduce the theoretical sequences of Theorem 8.30 by $x_0^0 = x_0 = \mathbb{1}$, $y_{k+1}^0 = (I_n - \frac{1}{\lambda_{\max}} \Sigma)x_k^0$ and $x_{k+1}^0 = \Pi_{\Omega} y_{k+1}^0$. We also introduce the sequences $(a_k)_{k \in \mathbb{N}}$, $(\alpha_k)_{k \in \mathbb{N}}$ and $(\bar{x}_k)_{k \in \mathbb{N}}$ defined by $x_k = F_{y_k}(a_k)$, $\alpha_k = PF_{y_k}(a_k)$ and $\bar{x}_k = \Pi_{\Omega_{\alpha_k}} y_k^0$. Note that $x_k = \Pi_{\Omega_{\alpha_k}} y_k$.

First step: bounds

Note that $\|x_0\| \leq R$ since $\|x_0\|^2 = n \leq n\kappa$. Let us show that $\|x^*\| \leq R$. We denote $\mathcal{E}_{\Sigma} = \{x \in \mathbb{R}^n | x^\top \Sigma x \leq 1\}$ the ellipsoid characterizing the SPD matrix Σ . We clearly have $B(0, \frac{1}{\sqrt{\lambda_{\max}}}) \subseteq \mathcal{E}_{\Sigma} \subseteq B(0, \frac{1}{\sqrt{\lambda_{\min}}})$, where $B(0, d) = \{x \in \mathbb{R}^n | \|x\| \leq d\}$ is the centered closed ball of radius d . Thus $\frac{1}{2}\sqrt{n\lambda_{\min}}\mathcal{E}_{\Sigma} \cap \Omega \subseteq B(0, \frac{\sqrt{n}}{2}) \cap \Omega = \emptyset$. In addition, $r \leq r'$ if and only if $r\mathcal{E}_{\Sigma} \subseteq r'\mathcal{E}_{\Sigma}$. Therefore, there exists (a unique) $r^* \in \mathbb{R}^+$ such that for all $r \in \mathbb{R}^+$, $r < r^*$ if and only if $r\mathcal{E}_{\Sigma} \cap \Omega = \emptyset$. Note that $r\mathcal{E}_{\Sigma} = f^{-1}((0, r^2])$. Since x^* is the unique minimizer of f on Ω , we have $r^* = \sqrt{\min_{\Omega} f}$ and $r^*\mathcal{E}_{\Sigma} \cap \Omega = \{x^*\}$. Since $\mathbb{1} \in B(0, \sqrt{n}) \cap \Omega = \sqrt{n\lambda_{\max}}B(0, \frac{1}{\sqrt{\lambda_{\max}}}) \cap \Omega \subseteq \sqrt{n\lambda_{\max}}\mathcal{E}_{\Sigma} \cap \Omega$, this proves that $\sqrt{n\lambda_{\max}}\mathcal{E}_{\Sigma} \cap \Omega$ is non-empty so $\sqrt{n\lambda_{\max}} \geq r^*$. Hence $x^* \in r^*\mathcal{E}_{\Sigma} \subseteq \sqrt{n\lambda_{\max}}\mathcal{E}_{\Sigma} \subseteq B(0, R)$ so $\|x^*\| \leq R$.

From Theorem 8.30, we have $\|x_k^0 - x^*\| \leq \exp(-\frac{k}{2\kappa})\|x_0^0 - x^*\| \leq 2R$ so $\|x_k^0\| \leq \|x_k^0 - x^*\| + \|x^*\| \leq 3R$ and $\|y_{k+1}^0\| \leq \|x_k^0\| \leq 3R$ because the eigenvalues of $I_n - \frac{1}{\lambda_{\max}}\Sigma$ are in $[0, 1)$.

Moreover, using Theorem 8.32 and the fact that $\Pi_{\Omega_{\alpha_{k+1}}}$ is 1-Lipschitz, we have:

$$\begin{aligned}
\|x_{k+1}^0 - x_{k+1}\| &\leq \|x_{k+1}^0 - \bar{x}_{k+1}\| + \|\bar{x}_{k+1} - x_{k+1}\| \\
&= \|\Pi_{\Omega} y_{k+1}^0 - \Pi_{\Omega_{\alpha_{k+1}}} y_{k+1}^0\| + \|\Pi_{\Omega_{\alpha_{k+1}}} y_{k+1}^0 - \Pi_{\Omega_{\alpha_{k+1}}} y_{k+1}\| \\
&\leq (\alpha_{k+1}^{1/n} - 1) \|x_{k+1}^0\| + \|y_{k+1}^0 - y_{k+1}\| \\
&\leq \delta_{k+1} 3R + \|x_k^0 - x_k\| = 3R\delta u_{k+1} + \|x_k^0 - x_k\|, \\
\|x_k^0 - x_k\| &\leq \sum_{i=0}^{k-1} 3R\delta u_{i+1} \leq 3R\delta \leq R, \\
\|x_k\| &\leq \|x_k - x_k^0\| + \|x_k^0\| \leq 4R, \\
\|y_{k+1}\| &\leq \|x_k\| \leq 4R.
\end{aligned}$$

Second step: extra term.

By Definition 8.29, we have $(x_{k+1}^0 - y_{k+1}^0)^\top (x_{k+1}^0 - x^*) \leq 0$. We would like to have an equality of the form $(x_{k+1} - y_{k+1})^\top (x_{k+1} - x^*) \leq c_\varepsilon$ where $c_\varepsilon \underset{\varepsilon \rightarrow 0}{=} o(1)$. We introduce x_{k+1}^0 and y_{k+1}^0 in this expression to use the inequality:

$$\begin{aligned}
&(x_{k+1} - y_{k+1})^\top (x_{k+1} - x^*) \\
&= (x_{k+1} - x_{k+1}^0 + x_{k+1}^0 - y_{k+1}^0 + y_{k+1}^0 - y_{k+1})^\top (x_{k+1} - x_{k+1}^0 + x_{k+1}^0 - x^*) \\
&\leq (x_{k+1} - x_{k+1}^0)^\top (x_{k+1} - y_{k+1} + x_{k+1}^0 - x^*) + (y_{k+1}^0 - y_{k+1})^\top (x_{k+1}^0 - x^*) \\
&\leq \|x_{k+1} - x_{k+1}^0\| (\|x_{k+1}\| + \|y_{k+1}\| + \|x_{k+1}^0 - x^*\|) + \|y_{k+1}^0 - y_{k+1}\| \|x_{k+1}^0 - x^*\| \\
&\leq \|x_{k+1} - x_{k+1}^0\| (\|x_{k+1}\| + \|y_{k+1}\| + 2\|x_{k+1}^0 - x^*\|) \\
&\leq 3R\delta(4R + 4R + 4R) = 36R^2\delta = \frac{\varepsilon^2}{2\kappa} =: c_\varepsilon.
\end{aligned}$$

This implies an inequality that we use to prove the convergence and which is a modification of a similar inequality (namely with $c_\varepsilon = 0$) from [Bubeck, 2015]. Note that $y_{k+1} = (I_n - \frac{1}{\lambda_{\max}} \Sigma)x_k = x_k - \frac{1}{\beta} \nabla f|_{x_k}$ since $\nabla f|_x = 2\Sigma x$ and $\beta = 2\lambda_{\max}$. Hence the previous inequality rewrites:

$$\begin{aligned}
&(x_{k+1} - (x_k - \frac{1}{\beta} \nabla f|_{x_k}))^\top (x_{k+1} - x^*) \leq c_\varepsilon \\
&\iff \nabla f|_{x_k}^\top (x_{k+1} - x^*) \leq \beta c_\varepsilon + \beta(x_k - x_{k+1})^\top (x_{k+1} - x^*).
\end{aligned}$$

Third step: convergence.

We use the fact that f is α -strongly convex and β -smooth with $\alpha = 2\lambda_{\min}$ and $\beta = 2\lambda_{\max}$ as in [Bubeck, 2015]:

$$\begin{aligned}
0 &\leq f(x_{k+1}) - f(x^*) = f(x_{k+1}) - f(x_k) + f(x_k) - f(x^*) \\
&\leq \nabla f|_{x_k}^\top (x_{k+1} - x_k) + \frac{\beta}{2} \|x_{k+1} - x_k\|^2 + \nabla f|_{x_k}^\top (x_k - x^*) - \frac{\alpha}{2} \|x_k - x^*\|^2 \\
&= \nabla f|_{x_k}^\top (x_{k+1} - x^*) + \frac{\beta}{2} \|x_{k+1} - x_k\|^2 - \frac{\alpha}{2} \|x_k - x^*\|^2 \\
&\leq \beta c_\varepsilon + \beta(x_k - x_{k+1})^\top (x_{k+1} - x^*) + \frac{\beta}{2} \|x_{k+1} - x_k\|^2 - \frac{\alpha}{2} \|x_k - x^*\|^2 \\
&\leq \beta \frac{\varepsilon^2}{2\kappa} + \beta(x_k - x_{k+1})^\top (x_k - x^*) - \frac{\beta}{2} \|x_{k+1} - x_k\|^2 - \frac{\alpha}{2} \|x_k - x^*\|^2,
\end{aligned}$$

so:

$$\|x_{k+1} - x_k\|^2 - 2(x_k - x^*)^\top (x_k - x_{k+1}) \leq \frac{\varepsilon^2}{\kappa} - \frac{\alpha}{\beta} \|x_k - x^*\|^2.$$

Hence:

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \|x_{k+1} - x_k + x_k - x^*\|^2 \\ &= \|x_k - x^*\|^2 + \|x_{k+1} - x_k\|^2 - 2(x_k - x^*)^\top (x_k - x_{k+1}) \\ &\leq \frac{\varepsilon^2}{\kappa} + \left(1 - \frac{1}{\kappa}\right) \|x_k - x^*\|^2, \\ \|x_{k+1} - x^*\|^2 - \varepsilon^2 &\leq \left(1 - \frac{1}{\kappa}\right) (\|x_k - x^*\|^2 - \varepsilon^2). \end{aligned}$$

Let $k_0 = \min\{k \in \mathbb{N} \mid \|x_k - x^*\| \leq \varepsilon\}$. Then for all $k \leq k_0$:

$$\begin{aligned} \|x_k - x^*\|^2 - \varepsilon^2 &\leq \left(1 - \frac{1}{\kappa}\right)^k (\|x_0 - x^*\|^2 - \varepsilon^2) \\ &\leq \|x_0 - x^*\|^2 \exp\left(-\frac{k}{\kappa}\right), \end{aligned}$$

and this inequality is trivially true for $k \geq k_0$ since $\|x_k - x^*\| - \varepsilon \leq 0 \leq \|x_0 - x^*\|^2 \exp(-\frac{k}{\kappa})$. To conclude, $\|x_k - x^*\| \leq \varepsilon + \|x_0 - x^*\| \exp(-\frac{k}{2\kappa})$.

Finally, the time complexity k_ε is obtained by $\|x_0 - x^*\| \exp(-\frac{k_\varepsilon}{2\kappa}) = \varepsilon$, i.e. $k_\varepsilon = 2\kappa \ln(\frac{\|x_0 - x^*\|}{\varepsilon}) \leq 2\kappa \ln(\frac{2R}{\varepsilon}) = O(\ln(\frac{n}{\varepsilon}))$. At each step, there is a multiplication matrix-vector (n^2 operations) and a projection approximation with complexity $\frac{5n-1}{\ln(2)} \ln(\frac{n}{\delta_j})$ since $(c_0)_j = 1 - \min_i [y_k]_i \leq 1 + 4R = 1 + \sqrt{n\kappa}$ (with the notation c_0 from Theorem 8.31). Therefore, the total complexity is bounded by the following quantity:

$$\sum_{j=1}^{k_\varepsilon} \left(n^2 + n \ln\left(\frac{n}{\delta_j}\right) \right) \leq n^2 k_\varepsilon + n k_\varepsilon \ln\left(\frac{n}{\delta u_{k_\varepsilon}}\right) = O\left(n^2 \ln\left(\frac{n}{\varepsilon}\right) + n \ln\left(\frac{n}{\varepsilon}\right) \ln \ln\left(\frac{n}{\varepsilon}\right) \right).$$

□

From the proof, we can notice that the choice of $\delta \propto \varepsilon^2$ seems necessary to get an approximation of the result up to ε . Numerically, it means that if real numbers are implemented with precision 10^{-16} , then the precision on \mathcal{D}^* cannot exceed 10^{-8} . Moreover, the other parameters n , κ and u_k 's strengthen this observation. For example, if $n = 100$, $\kappa = 10$ and $u_{10} = \frac{1}{100}$, then $\delta_{10} < \varepsilon^2 \times 10^{-6}$ so $\varepsilon > 10^{-5}$. However, the inequalities in the proof are far from optimal. It means that at the beginning of the algorithm, x_k might be far from x_{k+1} and x^* so the inequalities used in the third step are large. In addition, when x_k becomes close to x^* , the inequalities used in the first step are overestimated of a factor about 18. These observations can help gain in precision in practice.

The combination of this approach with the Newton method of [Khachiyan and Kalantari, 1992] can also help gain in precision. Indeed, as soon as $\|\Delta_k \Sigma \Delta_k \mathbf{1} - \mathbf{1}\| \leq c < 1$ (where $\Delta_k = \frac{1}{\sqrt{x_k^\top \Sigma x_k}} \text{diag}(x_k)$), the Newton method ensures that for all $s \in \mathbb{N}$, $\|\Delta_{k+s} \Sigma \Delta_{k+s} \mathbf{1} - \mathbf{1}\| \leq c^{2^s}$ [Khachiyan and Kalantari, 1992] so one could reach any precision ε' in an additional

$O(\ln(\ln(\frac{1}{\varepsilon'})))$ of steps. The total complexity becomes $O(n^2 \ln(n \ln(\frac{1}{\varepsilon'})))$ since each step of Newton requires to solve a linear system which is a $O(n^2)$. Although the complexity is worse, the precision could be better. However, this is conditioned to the possibility to reach $\|\Delta_k \Sigma \Delta_k \mathbf{1} - \mathbf{1}\| < 1$ with our method. One can show that $\|\Delta_k \Sigma \Delta_k \mathbf{1} - \mathbf{1}\| \leq 2\kappa R \|x_k - x^*\|$ so one has to choose $\varepsilon < \frac{1}{8\kappa R}$. With the values of the previous paragraph, $\frac{1}{8\kappa R} = 5 \times 10^{-4}$ which is greater than 10^{-5} so one can actually reach any precision ε' by plugging Khachiyan's method [Khachiyan and Kalantari, 1992] after ours, with a better complexity than Khachiyan's. Note that the precision ε' is limited by the resolution of the linear system $(\Delta_{k+s}^{-2} + \Sigma)h = -\Sigma \Delta_{k+s} \mathbf{1}$ where $h \in \mathbb{R}^n$ is unknown, the next step being $\Delta_{k+s+1} = \Delta_{k+s} + \text{diag}(h)$.

To compare with algorithms for which the complexity depends on the number of non-zero entries $m \leq n^2$ of Σ , we can notice that the multiplication matrix-vector at each step of the algorithm only requires m operations. Therefore, the complexity of our algorithm is actually $O(m \ln(\frac{n}{\varepsilon}) + n \ln(\frac{n}{\varepsilon}) \ln(\ln(\frac{n}{\varepsilon})))$. In comparison with the generic algorithm with complexity $O(m + n^{4/3})$ of [Allen-Zhu et al., 2017], it is competitive for sparse SPD matrices with $m < n^{4/3}$.

8.6 Conclusion

In this work, we introduced two families of permutation-invariant log-Euclidean metrics on full-rank correlation matrices. The first family was built via the off-log diffeomorphism whose algebraic properties were introduced in [Archakov and Hansen, 2021]. The second family was built via the log-scaling diffeomorphism thanks to the result on the scaling of symmetric matrices [Marshall and Olkin, 1968]. In addition, the log-scaled metrics are inverse-consistent contrarily to the off-log metrics. These metrics allow to compute with correlation matrices very efficiently since they are flat and the main Riemannian operations are known in closed form modulo the computation of the maps \mathcal{D} and \mathcal{D}^* . That is why we also provide efficient algorithms to compute them. These permutation-invariant log-Euclidean metrics fill a gap in the study of Riemannian metrics on the open ellipsope since they gather the invariance under permutations of some structures [Kercheval, 2008, Nielsen and Sun, 2019, David and Gu, 2019, Thanwerdas and Pennec, 2021] and the log-Euclidean geometry of some others [Thanwerdas and Pennec, 2022c]. More generally, we tried to introduce a comprehensive formalism on full-rank correlation matrices in terms of stability under the action of permutations and under the cor-inverse involution. This systematic approach allowed to satisfy intrinsically the requirement of inverse-consistency in the construction of log-Euclidean metrics. We hope that this presentation will help to manipulate correlation matrices as easily as we work with SPD matrices. This approach could also help to define appropriate geometries of block equi-correlation matrices of a given signature.

Part V

Stratified spaces of covariance and correlation matrices

Chapter 9

Bures-Wasserstein stratified geometry of covariance matrices

Abstract

The set of covariance matrices equipped with the Bures-Wasserstein distance is the orbit space of the smooth, proper and isometric action of the orthogonal group on the Euclidean space of square matrices by right multiplication. This construction induces a natural orbit stratification on covariance matrices, which is exactly the stratification by the rank. Thus, the strata are the manifolds of symmetric positive semi-definite (PSD) matrices of fixed rank endowed with the Bures-Wasserstein Riemannian metric. In this work, we study the geodesics of the Bures-Wasserstein distance. Firstly, we complete the literature on geodesics in each stratum by clarifying the set of preimages of the exponential map and by specifying the injectivity domain. We also give explicit formulae of the horizontal lift, the exponential map and the Riemannian logarithms that were kept implicit in previous works. Secondly, we give the expression of all the minimizing geodesic segments joining two covariance matrices of any rank. More precisely, we show that the set of all minimizing geodesics between two covariance matrices Σ and Λ is parametrized by the closed unit ball of $\mathbb{R}^{(k-r) \times (l-r)}$ for the spectral norm, where k, l, r are the respective ranks of $\Sigma, \Lambda, \Sigma\Lambda$. In particular, the minimizing geodesic is unique if and only if $r = \min(k, l)$. Otherwise, there are infinitely many.

9.1 Introduction

Many data can be represented as covariance matrices. They are often assumed to be Symmetric Positive Definite (SPD) because it is much more convenient from the geometric point of view. Indeed, the set of SPD matrices is an open convex cone in the vector space of symmetric matrices so it has a canonical differential structure. The induced Euclidean metric is not satisfying to compute with SPD matrices because geodesics leave the space in finite time and interpolations are often non-realistic. To solve this problem, a lot of Riemannian metrics were proposed on SPD matrices, mainly $O(n)$ -invariant metrics [Thanwerdas and Pennec, 2022b] (affine-invariant [Siegel, 1943, Skovgaard, 1984, Pennec et al., 2006, Lenglet et al., 2006, Fletcher and Joshi, 2007, Moakher, 2005], log-Euclidean [Arsigny et al., 2006, Fillard et al., 2007, Hà Quang et al., 2014], Bures-Wasserstein [Dowson and Landau, 1982, Olkin and Pukelsheim, 1982, Takatsu, 2010, Takatsu, 2011, Malagò et al., 2018, Bhatia et al., 2019], Bogoliubov-Kubo-Mori [Petz and Toth, 1993, Michor et al., 2000], etc.), Cholesky-like metrics

[Li et al., 2017, Lin, 2019] or product metrics with one metric on positive diagonal matrices and one metric on full-rank correlation matrices [Thanwerdas and Pennec, 2021, Thanwerdas and Pennec, 2022c].

However, this viewpoint often forgets about singular covariance matrices, that is covariance matrices with non-full rank. Altogether, they form a closed convex cone which is not anymore a manifold. First, it can be equipped with distances to provide a metric space structure. The Euclidean distance is not satisfying either here because geodesics leave the closed cone in finite time again. A very interesting alternative is the Helstrom/Bures distance defined in quantum information geometry [Helstrom, 1967, Bures, 1969]. It was also introduced in optimal transport as the L^2 Wasserstein/Kantorovitch distance between multivariate centered Gaussian distributions, possibly degenerate [Dowson and Landau, 1982, Olkin and Pukelsheim, 1982]. This is why it is now called the Bures-Wasserstein distance. From our viewpoint, it allows to define a distance on the whole set of covariance matrices, contrarily to the aforementioned affine-invariant and log-Euclidean metrics. The Bures-Wasserstein distance is also the quotient distance of the Euclidean metric on square matrices by the right action of the orthogonal group $(U, M) \in O(n) \times \mathbb{R}^{n \times n} \longrightarrow MU \in \mathbb{R}^{n \times n}$. This is why it was also called the Procrustes distance [Dryden et al., 2009, Hà Quang, 2022]. This viewpoint allows to split the closed cone into strata that are Riemannian manifolds whose induced geodesic distance is precisely the Bures-Wasserstein distance. In particular, the space of covariance matrices equipped with this distance is a complete geodesic metric space.

The geometry of stratified spaces is a topic of interest in the community of statistics in non-linear spaces. Examples of popular stratified spaces are the Kendall shape spaces [Kendall, 1984], the BHV space of trees [Billera et al., 2001], the QED space of trees [Ferguson et al., 2010], the Graph space [Calissano et al., 2020], the Wald space of forests [Garba et al., 2021], the correlation matrices or the symmetric/diagonal matrices stratified by eigenvalue multiplicity. Moreover, the space of covariance matrices with the Bures-Wasserstein distance is a metric space of non-negative curvature [Takatsu, 2011]. Spaces of this type have been much less described than metric spaces of non-positive curvature [Bridson and Haefliger, 1999]. These are two motivations to study the Bures-Wasserstein geometry of covariance matrices. In this work, we focus on geodesics.

There are three new elements with respect to the previous chapters that we would like to highlight. Firstly, most of the previous Riemannian manifolds were geodesically complete. In this chapter, we need the notion of *definition domain* of the exponential map because it is not \mathbb{R} in general. Secondly, the geodesics were minimizing on \mathbb{R} most of the time. Here, we need the notions of *cut time* and injectivity domain to specify when the geodesic stops to be minimizing. Thirdly, we talk about *geodesics in a metric space*, not only in a Riemannian manifold as previously.

In the principal (or regular) stratum of SPD matrices, the Bures-Wasserstein Riemannian metric was extensively studied. Since $X \in GL(n) \longmapsto XX^\top \in \text{Sym}^+(n)$ is a Riemannian submersion, many geometric operations can be computed thanks to O’Neill’s equations [O’Neill, 1966]. Therefore, the curvature was derived in [Takatsu, 2010, Takatsu, 2011, Thanwerdas and Pennec, 2022b], the quotient geometry was described in [Malagò et al., 2018, Bhatia et al., 2019, van Oostrum, 2020], the exponential map was computed in [Malagò et al., 2018], a Riemannian logarithm was given in [Bhatia et al., 2019], the injectivity radius was computed in [Massart and Absil, 2020] and a simplified equation of the geodesic parallel transport

was proposed in [Thanwerdas and Pennec, 2022b]. In contrast to geodesically complete Riemannian metrics, it is important to specify the definition domain of the exponential map. It was characterized in [Malagò et al., 2018] as the connected component of 0 in a subset of \mathbb{R} , which could be specified more explicitly. Moreover, the uniqueness of the Riemannian logarithm is not established and the injectivity domain seems to be unknown. Therefore in Section 9.4, we clarify the definition domain of the exponential map, we prove the uniqueness of the logarithm thanks to a result from [Massart and Absil, 2020] and we prove that the geodesics are minimizing on their domain of definition, which also provides the injectivity domain.

In each other stratum of PSD matrices of fixed rank $k < n$, the Bures-Wasserstein Riemannian metric was studied via the analogous Riemannian submersion defined by $X \in \mathbb{R}_*^{n \times k} \mapsto XX^\top \in \text{Sym}^+(n, k)$, where $\mathbb{R}_*^{n \times k}$ is the open set of matrices of full rank k in $\mathbb{R}^{n \times k}$ and $\text{Sym}^+(n, k)$ is the set of PSD matrices of size n and rank k . The curvature was computed in [Massart et al., 2019], the exponential map, its domain of definition, the logarithm map and the injectivity radius were derived in [Massart and Absil, 2020]. The horizontal lift was kept implicit so these results are formulated in the total space $\mathbb{R}_*^{n \times k}$. We think that it is easier to understand the geometry with formulae depending directly on the tangent vector and not on its horizontal lift. Therefore, in Section 9.5, we compute the horizontal lift and we give the expressions of the Riemannian metric, the exponential map and its definition domain in function of vectors tangent to the manifold $\text{Sym}^+(n, k)$. Moreover, we solve the problem of characterization of the preimages of the exponential map, which is knowingly left open in [Massart and Absil, 2020] where they focus on the characterization of logarithms, that is preimages with minimal norm. In addition, we give an explicit bijective parametrization of the Riemannian logarithms, which allows us to count them. When it is unique, we give an explicit formula of the corresponding minimizing geodesic in function of the end points. This finally allows us to compute the injectivity domain which is kept implicit in [Massart and Absil, 2020].

Beyond the clarification and completion of the literature on geodesics in each stratum, our main objective is to characterize the minimizing geodesics *between* strata. Our main results are the following. (1) Any minimizing geodesic segment between two covariance matrices Σ and Λ is of constant rank on the interior of the segment. It is called the rank of the minimizing geodesic and it is greater than the ranks of Σ and Λ . (2) We give the explicit formula of all the minimizing geodesic segments in Theorem 9.31. (3) They are parametrized by the vectors of the closed unit ball of $\mathbb{R}^{(k-r) \times (l-r)}$ for the spectral norm, where k, l, r are the respective ranks of $\Sigma, \Lambda, \Sigma\Lambda$. In other words, they are parametrized by matrices $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$ with singular values in $[0, 1]$. (4) The minimizing geodesic segment is unique if and only if $r = \min(k, l)$ (this includes $\max(k, l) = n$). Otherwise, there are infinitely many. (5) The number of minimizing geodesics of minimal rank (i.e. of rank equal to $\max(k, l)$) is 1 if $r = \min(k, l)$, otherwise it is 2 if $k = l$, otherwise it is infinite. (6) Assuming $k \geq l$, if R_0 belongs to the Stiefel manifold $\text{St}(k-r, l-r)$, that is $R_0^\top R_0 = I_{l-r}$, then the corresponding geodesic is of minimal rank. (7) The choice of parameter $R_0 = 0$ leads to the geodesic $\gamma_{\Sigma \rightarrow \Lambda}^0(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda)$ whose expression does not depend on the ranks of Σ and Λ . It is called the Bures-Wasserstein canonical geodesic.

In the remainder of this section, we introduce some matrix notations. In Section 9.2, we introduce the important concepts of geodesics and quotient space in metric spaces and

manifolds. We give a particular attention to the notions that characterize if a geodesic (self-parallel curve) is minimizing: cut time, injectivity domain, difference between preimages of the exponential map and Riemannian logarithms. In Section 9.3, we recall the algebraic structure, metric topology and differential geometry of the convex cone of covariance matrices seen as the quotient of square matrices by the orthogonal group. In Sections 9.4 and 9.5, we complete the literature on the geodesics of the Bures-Wasserstein metric on SPD matrices and on singular matrices of fixed rank respectively. In Section 9.6, we give our main results on minimizing geodesics in the whole Bures-Wasserstein metric space of covariance matrices. We conclude in Section 9.7 and we give a glance at an analogous structure on correlation matrices. The main proofs are deferred to Section 11.8.

Matrix notations Let $n, k \in \mathbb{N}$. In this work, we use the following manifolds of matrices:

- the vector space of $n \times k$ matrices $\mathbb{R}^{n \times k}$,
- the open subset $\mathbb{R}_*^{n \times k} \subset \mathbb{R}^{n \times k}$ of full-rank matrices,
- in particular, the vector space of square matrices $\text{Mat}(n) = \mathbb{R}^{n \times n}$ and the general linear group $\text{GL}(n) = \mathbb{R}_*^{n \times n}$,
- the orthogonal group $\text{O}(n)$,
- the Stiefel manifold $\text{St}(n, k) = \text{O}(n)/\text{O}(n - k)$,
- the manifold of symmetric positive definite matrices $\text{Sym}^+(n)$,
- the manifold of symmetric positive semi-definite matrices of fixed rank k , $\text{Sym}^+(n, k)$,
- the vector space of diagonal matrices $\text{Diag}(n)$,
- the groups of invertible diagonal matrices $\text{Diag}^*(n) = \text{Diag}(n) \cap \text{GL}(n)$ and positive diagonal matrices $\text{Diag}^+(n) = \text{Diag}(n) \cap \text{Sym}^+(n)$.

We use the following notations.

- I_n denotes the identity matrix of size n .
- $\mathbf{0}_n$ denotes the null matrix of size $n \times n$. $\mathbf{0}_{n,k}$ denotes the null matrix of size $n \times k$. We may simply denote them 0 when sizes are obvious in the context.
- (Sylvester equation) $\mathcal{S}_A(B)$ is the unique solution Z of the Sylvester equation $AZ + ZA = B$ for $A \in \text{Sym}^+(k)$ and $B \in \text{Sym}(k)$.
- (Löwner order) For all $\Sigma \in \text{Sym}(n)$, we say that Σ is positive definite (resp. positive semi-definite) and we denote $\Sigma > 0$ (resp. $\Sigma \geq 0$) when Σ has positive (resp. non-negative) eigenvalues. Given $\Lambda \in \text{Sym}(n)$, we denote $\Sigma > \Lambda$ (resp. $\Sigma \geq \Lambda$) when $\Sigma - \Lambda > 0$ (resp. $\Sigma - \Lambda \geq 0$).

We recall basic facts on symmetric matrices.

1. (Eigenvalue decomposition) For all $\Sigma \in \text{Sym}(n)$, there exist $U \in \text{O}(n)$ and $D \in \text{Diag}(n)$ such that $\Sigma = UDU^\top$. By removing null eigenvalues, given $r = \text{rk}(\Sigma)$, there also exist $U \in \text{St}(n, r)$ and $D \in \text{Diag}^*(r)$ such that $\Sigma = UDU^\top$.

2. For all $X \in \mathbb{R}^{n \times k}$, XX^\top is symmetric positive semi-definite and $\text{rk}(X) = \text{rk}(XX^\top) = \text{rk}(X^\top X)$. In particular, if $X \in \mathbb{R}_*^{n \times k}$, then $X^\top X \in \text{GL}(k)$.
3. (Singular value decomposition) For all $M \in \mathbb{R}^{n \times k}$, denoting $r = \text{rk}(M) \leq \min(n, k)$, there exist $U \in O(n)$, $V \in O(k)$ and $D = \begin{pmatrix} D_r & \mathbf{0}_{r, k-r} \\ \mathbf{0}_{n-r, r} & \mathbf{0}_{n-r, k-r} \end{pmatrix}$ with $D_r \in \text{Diag}^+(r)$ such that $M = UDV^\top$. The diagonal entries of D , that is the D'_{ii} s for $i \in \{1, \dots, \min(n, k)\}$ are called the singular values of M .
4. (Moore-Penrose inverse) For all $M \in \mathbb{R}^{n \times k}$, the unique matrix $M^\dagger \in \mathbb{R}^{k \times n}$ satisfying $MM^\dagger M = M$, $M^\dagger MM^\dagger = M^\dagger$, $MM^\dagger \in \text{Sym}(n)$ and $M^\dagger M \in \text{Sym}(k)$ is called the pseudoinverse or Moore-Penrose inverse. In this work, we only use it for symmetric matrices $\Sigma \in \text{Sym}(n)$. Given $r = \text{rk}(\Sigma)$ and an eigenvalue decomposition $\Sigma = UDU^\top = P\text{Diag}(D, 0)P^\top$ with $U \in \text{St}(n, r)$, $P = [U \ U_\perp] \in O(n)$, $D \in \text{Diag}^*(r)$, it is easy to check that $\Sigma^\dagger = UD^{-1}U^\top = P\text{Diag}(D^{-1}, 0)P^\top$. Moreover, for all $V \in \text{St}(n, k)$, $(V\Sigma V^\top)^- = V\Sigma^-V^\top$.
5. (Symmetric square root) For all $\Sigma \in \text{Sym}(n)$, if $\Sigma \geq 0$, then there exists a unique matrix $A \in \text{Sym}(n)$, $A \geq 0$ such that $A^2 = \Sigma$. It is called the (symmetric) square root of Σ and it is denoted $\sqrt{\Sigma}$ or $\Sigma^{1/2}$.
6. (Non-symmetric square roots) For all $X, Y \in \mathbb{R}^{n \times k}$, $XX^\top = YY^\top$ if and only if there exists $U \in O(k)$ such that $XU = Y$ [Groetzner and Dür, 2020, Lemma 2.6].

We use the following notations for norms of vectors $x \in \mathbb{R}^k$ and matrices $M \in \mathbb{R}^{n \times k}$:

1. (Euclidean norm) $\|M\|_2 = \text{tr}(M^\top M)^{1/2}$ and $\|x\|_2 = (x^\top x)^{1/2}$.
2. (Spectral norm or Schatten's infinite norm) $\|M\|_S = \sup_{\substack{x \in \mathbb{R}^k \\ \|x\|_2 \leq 1}} \|Mx\|_2 = \max_{1 \leq i \leq n} \sigma_i(M)$, where $\sigma_1(M), \dots, \sigma_n(M)$ are the singular values of M .

Without index, $\|\cdot\|$ generically denotes the norm on the tangent spaces associated to the Riemannian metric at hand.

9.2 Preliminary concepts

In this section, we recall the basic definitions of geodesics and quotient spaces. We depart from the previous chapters where most of the Riemannian metrics were geodesically complete and where the injectivity domain of the exponential map was the entire manifold. We also need to introduce the notions of length and geodesic in a metric space [Bridson and Haefliger, 1999, Paulin, 2014] and the notion of Riemannian orbit space [Alekseevsky et al., 2001].

In Section 9.2.1, we introduce the concepts needed to study geodesics and minimizing geodesics. We recall that geodesics are defined in a metric space, in a manifold endowed with an affine connection and in a Riemannian manifold, where the two previous notions coincide. Then we introduce our notations for preimages of the exponential map, logarithms and related notions. We define the cut time and the injectivity domain, extending the definition usually given in complete metric spaces, and we explain where to be cautious. In Section 9.2.2, we recall the notions of quotient metric space, quotient Riemannian manifold and Riemannian

orbit space. We introduce the vocabulary, notations and results that we use in the next sections.

9.2.1 Geodesics

Definition 9.1 (Curve) Let \mathcal{M} be a topological space. A curve on \mathcal{M} is a continuous map $c : I \rightarrow \mathcal{M}$, where I is an interval of \mathbb{R} . When I is a segment of \mathbb{R} , we may call c a segment.

9.2.1.1 Geodesics in a metric space

Definition 9.2 (Length, length distance, length space) [Paulin, 2014] Let (\mathcal{M}, d) be a metric space.

1. (Length) Let $c : [a, b] \subseteq I \rightarrow \mathcal{M}$ be a curve. The length of c is defined by $L(c) = \sup \sum_{k=0}^p d(c(t_k), c(t_{k+1})) \in [0, +\infty]$ over all subdivisions $a = t_0 \leq t_1 \leq \dots \leq t_p \leq t_{p+1} = b$. We say that $c : I \rightarrow \mathcal{M}$ is rectifiable when for all $a \leq b$ in I , $L(c|_{[a,b]})$ is finite.
2. (Length distance) The length distance between $x \in \mathcal{M}$ and $y \in \mathcal{M}$ is defined by $d_L(x, y) = \inf L(c) \geq d(x, y)$ over all rectifiable curves $c : [0, 1] \rightarrow \mathcal{M}$ from $x = c(0)$ to $y = c(1)$. If \mathcal{M} is connected by rectifiable curves, then d_L is a distance.
3. (Length space) We say that (\mathcal{M}, d) is a length space when $d = d_L$.

Lemma 9.3 (Length is additive and continuous) [Bridson and Haefliger, 1999, Proposition 1.20] Let $c : [a, b] \rightarrow \mathcal{M}$ be a rectifiable curve.

1. For all $t \in [a, b]$, $L(c) = L(c|_{[a,t]}) + L(c|_{[t,b]})$.
2. The map $f : t \in [a, b] \mapsto L(c|_{[a,t]})$ is non-decreasing and continuous.

A classical example is the sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ endowed with the Euclidean distance d . The distances between the north pole $N = (0, 0, 1)$ and the south pole $S = (0, 0, -1)$ are $d(N, S) = 2$ and $d_L(N, S) = \pi$. The metric space (\mathbb{S}^2, d) is not a length space while the metric space (\mathbb{S}^2, d_L) is a length space.

Definition 9.4 (Geodesics in a metric space) [Paulin, 2014] Let (\mathcal{M}, d) be a metric space.

1. (Constant speed) We say that $c : I \rightarrow \mathcal{M}$ is a curve parameterized at constant speed when there exists $v \geq 0$ such that for all $t \leq t'$ in I , $L(c|_{[t,t']}) = v(t' - t)$.
2. (Unit speed) We say that $c : I \rightarrow \mathcal{M}$ is a curve parameterized at unit speed (or by arc length) when for all $t \leq t'$ in I , $L(c|_{[t,t']}) = t' - t$.
3. (Globally minimizing) We say that $c : I \rightarrow \mathcal{M}$ is (globally) minimizing when for all $a \leq b$ in I , $L(c|_{[a,b]}) = d(c(a), c(b))$.
4. (Locally minimizing) We say that $c : I \rightarrow \mathcal{M}$ is locally minimizing when for all $t \in I$, there exists a neighborhood $I_0 \subseteq I$ of t such that $c|_{I_0}$ is minimizing.
5. (Geodesic) A geodesic is a locally minimizing curve of constant speed. A minimizing geodesic is a globally minimizing curve of constant speed.

6. (Geodesic space) A geodesic (metric) space is a length space such that there exists a minimizing geodesic between any two points.

9.2.1.2 Geodesics of an affine connection

In the following notions, the definition interval and the definition domain are of interest when the manifold is not geodesically complete and the set of preimages is of interest when the exponential map is not injective.

Definition 9.5 (Geodesics of an affine connection) [Paulin, 2014] Let (\mathcal{M}, ∇) be a smooth manifold equipped with an affine connection $\nabla : \Gamma(T\mathcal{M}) \times \Gamma(T\mathcal{M}) \rightarrow \Gamma(T\mathcal{M})$.

1. (Geodesic) A geodesic (or self-parallel curve) is a solution $\gamma : I \rightarrow \mathcal{M}$ of the second-order equation $\forall t \in I, \nabla_{\dot{\gamma}(t)} \dot{\gamma} = 0$. The maximal solution satisfying the initial condition $\dot{\gamma}(0) = v$ for $v \in T_x\mathcal{M}$ is denoted $\gamma_{(x,v)} : I_{x,v} \rightarrow \mathcal{M}$.
2. (Definition interval) We call $I_{x,v} \subseteq \mathbb{R}$ the definition interval of $\gamma_{(x,v)}$. It is the maximal interval of \mathbb{R} on which the geodesic $\gamma_{(x,v)}$ is defined.

In a geodesically complete manifold, $I_{x,v} = \mathbb{R}$.

3. (Exponential map) The exponential map is defined by $\text{Exp} : v \mapsto \gamma_{(x,v)}(1)$ on the open set $\bigcup_{x \in \mathcal{M}} \{v \in T_x\mathcal{M} \mid 1 \in I_{x,v}\} \subseteq T\mathcal{M}$. The exponential map at x , defined by $\text{Exp}_x(v) = \text{Exp}(v)$ for $v \in T_x\mathcal{M}$ such that $1 \in I_{x,v}$, is a diffeomorphism from a neighborhood of 0 in $T_x\mathcal{M}$ to a neighborhood of x in \mathcal{M} .
4. (Definition domain) The definition domain of the exponential map is $\mathcal{D}_x = \{v \in T_x\mathcal{M} \mid 1 \in I_{x,v}\}$.
5. (Preimage) A preimage of $y \in \mathcal{M}$ by Exp_x is a vector $v \in T_x\mathcal{M}$ such that $1 \in I_{x,v}$ and $\text{Exp}_x(v) = y$. The set of preimages of y from x is denoted $\mathcal{P}re_x(y) = \text{Exp}_x^{-1}(\{y\}) = \{v \in T_x\mathcal{M} \mid 1 \in I_{x,v} \text{ and } \text{Exp}_x(v) = y\}$.

Note that there might be none, one, several or infinitely many elements in $\mathcal{P}re_x(y)$.

6. (Geodesic from x to y) A geodesic from x to y is a geodesic $\gamma_{(x,v)}$ such that $v \in \mathcal{P}re_x(y)$. They are bijectively indexed by $\mathcal{P}re_x(y)$ so we denote them $\gamma_{x \rightarrow y}^v = \gamma_{(x,v)}$ for $v \in \mathcal{P}re_x(y)$.

When there exists a unique preimage of y from x , we simply denote the geodesic $\gamma_{x \rightarrow y}$.

9.2.1.3 Geodesics of a Riemannian metric

Any Riemannian manifold is equipped with a natural affine connection called the Levi-Civita connection. Moreover, we recall in the following definition that a connected Riemannian manifold is a metric space and even a length space.

Definition 9.6 (Length, Riemannian distance) [Paulin, 2014] Let (\mathcal{M}, g) be a Riemannian manifold. For $v \in T\mathcal{M}$, we denote its norm $\|v\| = \sqrt{g(v, v)}$.

1. (Length) The length of a \mathcal{C}^1 curve $c : [a, b] \rightarrow \mathcal{M}$ is defined by $L(c) = \int_0^1 \|\dot{c}(t)\| dt$. This definition extends to piecewise \mathcal{C}^1 curves. The length is independent from the parametrization of the curve.
2. (Distance) The Riemannian distance between $x, y \in \mathcal{M}$ is defined by $d(x, y) = \inf L(c)$ over all piecewise \mathcal{C}^1 curves $c : [0, 1] \rightarrow \mathcal{M}$ from $x = c(0)$ to $y = c(1)$. If \mathcal{M} is connected, then the Riemannian distance is a distance (defining the topology of \mathcal{M}), the Riemannian length and the metric length coincide on piecewise \mathcal{C}^1 curves, and (\mathcal{M}, d) is a length space.

Fortunately, the two notions of geodesics coincide: a piecewise \mathcal{C}^1 curve is a geodesic for the Levi-Civita connection if and only if it is a geodesic in the metric space (\mathcal{M}, d) . Then, it is natural to ask what are the *globally* minimizing geodesics. A few concepts can be introduced to formalize this question: injectivity radius, cut time, cut locus, injectivity domain. Moreover, among the preimages $v \in \text{Pre}_x(y)$ of the exponential map of the Levi-Civita connection, those which satisfy $\|v\| = d(x, y)$ are called Riemannian logarithms. When there exists a unique Riemannian logarithm, the logarithm map can be defined.

Definition 9.7 (Geodesics of a Riemannian metric) Let (\mathcal{M}, g) be a connected Riemannian manifold. Let d be the Riemannian distance and ∇ be the Levi-Civita connection. We denote $\mathcal{B}_x(\varepsilon) = \{v \in T_x\mathcal{M} \mid \|v\| < \varepsilon\} \subset T_x\mathcal{M}$ the centered open ball of $T_x\mathcal{M}$ of radius $\varepsilon > 0$.

1. (Geodesics) A curve $\gamma : I \rightarrow \mathcal{M}$ is called a geodesic of (\mathcal{M}, g) if one of the two following equivalent statements is satisfied:
 - (a) γ is a geodesic of (\mathcal{M}, d) (locally length-minimizing curve of constant speed),
 - (b) γ is a geodesic of (\mathcal{M}, ∇) (self-parallel curve).

The equivalence is shown in [Paulin, 2014, Proposition 3.14] for example.

2. (Injectivity radius) [Paulin, 2014] The injectivity radius at $x \in \mathcal{M}$ is defined by $\text{inj}(x) = \sup \varepsilon$ over all $\varepsilon > 0$ such that \exp_x is a diffeomorphism from $\mathcal{B}_x(\varepsilon) \subset T_x\mathcal{M}$ to its image. The injectivity ball is $\mathcal{B}_x(\text{inj}(x)) \subset T_x\mathcal{M}$.

For $v \in T_x\mathcal{M}$ of norm 1 and $t_0 \in [0, \text{inj}(x))$, the map $\gamma_{(x,v)} : t \in [0, t_0] \mapsto \exp_x(tv) \in \mathcal{M}$ is the unique geodesic between x and $\gamma_{(x,v)}(t_0)$. It is globally minimizing.

The injectivity radius of \mathcal{M} is defined by $\text{inj}(\mathcal{M}) = \inf_{x \in \mathcal{M}} \text{inj}(x)$.

3. (Cut time) The cut time at $x \in \mathcal{M}$ in the direction $v \in T_x\mathcal{M}$, $\|v\| = 1$, is defined by $t_{\text{cut}}(x, v) = \sup\{t \in I_{x,v} \mid d(x, \text{Exp}_x(tv)) = t\} \in (0, +\infty]$. Note that the geodesic $\gamma_{(x,v)}$ need not be minimizing on $(-t_{\text{cut}}(x, -v), t_{\text{cut}}(x, v))$.
4. (Tangential cut locus) The tangential cut locus at x is the set $TCL(x) = \{t_{\text{cut}}(x, v)v \mid v \in T_x\mathcal{M}, \|v\| = 1, t_{\text{cut}}(x, v) < +\infty\}$.
5. (Injectivity domain) The injectivity domain of the exponential map at x is the set $\text{Inj}(x) = \{tv \mid t \in [0, t_{\text{cut}}(x, v)), v \in T_x\mathcal{M}, \|v\| = 1\} \subseteq \mathcal{D}_x$. The injectivity ball is included in the injectivity domain.

6. (Logarithms) We call (Riemannian) logarithm of $y \in \mathcal{M}$ from $x \in \mathcal{M}$ a preimage $v \in \mathcal{P}re_x(y) \subseteq T_x\mathcal{M}$ of y from x by the exponential map such that $\|v\| = d(x, y)$. We denote $\mathcal{L}og_x(y) \subseteq \mathcal{P}re_x(y)$ the set of logarithms of y from x . In particular, the geodesic $\gamma_{x \rightarrow y}^v$ joining x to y with initial speed $v \in \mathcal{L}og_x$ is minimizing on $[0, 1]$.
7. (Logarithm map) Denoting $\mathcal{U}_x \subseteq \mathcal{M}$ the subset of points y such that there exists a unique logarithm of y from x , this defines a map $\text{Log}_x : \mathcal{U}_x \rightarrow T_x\mathcal{M}$. In particular, $\mathcal{B}_x(\text{inj}(x)) \subseteq \mathcal{U}_x$ and $\text{Log}_x : \text{Exp}_x(\mathcal{B}_x(\varepsilon)) \rightarrow \mathcal{B}_x(\varepsilon)$ is a diffeomorphism for all $\varepsilon < \text{inj}(x)$.

Remark 9.8 The cut time is usually defined in complete manifolds only [do Carmo, 1992]. Although the definition still holds, some basic results may fail in non-complete manifolds. For example, if $I_{x,v} \neq \mathbb{R}$ and $t_{cut}(x, v) < +\infty$, then $t_{cut}(x, v)$ need not belong to $I_{x,v}$ and $\text{Exp}_x(t_{cut}(x, v)v)$ need not be defined. Then the cut locus, which is the image of the tangential cut locus by the exponential map in complete manifolds, should be defined differently. One possible definition could simply forget the vectors $v \in T_x\mathcal{M}$ such that $t_{cut}(x, v) \notin I_{x,v}$ as well as the definition in complete manifolds forgets about the vectors $v \in T_x\mathcal{M}$ such that $t_{cut}(x, v) = +\infty$. A maybe more satisfying definition could rely on the metric completion of the space: the cut point in direction v could be the limit of $\text{Exp}_x(tv)$ when t tends to $t_{cut}(x, v)$ if this limit exists.

Studying the geodesics of a space means (at least) determining precisely an expression of the geodesic $\gamma_{(x,v)}$, the maximal domain $I_{x,v}$, the injectivity radiuses $\text{inj}(x)$ and $\text{inj}(\mathcal{M})$, the cut time $t_{cut}(x, v)$, the preimages of the exponential map $\mathcal{P}re_x(y)$, an expression of the geodesics $\gamma_{x \rightarrow y}^v$, the Riemannian logarithms $\mathcal{L}og_x(y)$, the logarithm map Log_x and its definition domain \mathcal{U}_x .

The goal of Sections 9.4 and 9.5 is to clarify and complete the knowledge on geodesics of the Bures-Wasserstein metric on the manifolds $\text{Sym}^+(n)$ and $\text{Sym}^+(n, k)$.

9.2.2 Quotient spaces

9.2.2.1 Quotient distance

Definition 9.9 (Quotient distance) Let (\mathcal{M}, d) be a metric space. Let G be a group acting isometrically on (\mathcal{M}, d) . We denote $\mathcal{M}^0 = \mathcal{M}/G$ and we define the map $d^0 : (Gx, Gy) \in \mathcal{M}^0 \times \mathcal{M}^0 \mapsto d(Gx, Gy) = \inf_{g \in G} d(gx, y) \in [0, +\infty)$. If the orbits are closed, then d_0 is a distance on \mathcal{M}^0 called the quotient distance.

Remark 9.10 There exists a more general notion of quotient pseudo-distance when the action is not isometric. In this chapter, the actions are isometric so we don't detail it.

Remark 9.11 The hypotheses of the definition are satisfied if the following conditions hold together:

- G is a locally compact topological group (e.g. a Lie group),
- (\mathcal{M}, d) is a metric space (e.g. a Riemannian manifold),

- the action of G on (\mathcal{M}, d) is continuous and isometric,
- the action is proper, i.e. (with the previous assumptions [Bourbaki, 1971, III.4.4 Proposition 7]) for all $x, y \in \mathcal{M}$, there exist respective neighborhoods $\mathcal{V}_x, \mathcal{V}_y$ such that the set $K = \{g \in G | g\mathcal{V}_x \cap \mathcal{V}_y \neq \emptyset\}$ is relatively compact.

Indeed, let $(g_n x)_{n \in \mathbb{N}}$ be a sequence in Gx tending to $y \in \mathcal{M}$. Then there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $g_n x \in \mathcal{V}_y$ so $g_n \mathcal{V}_x \cap \mathcal{V}_y \neq \emptyset$, i.e. $g_n \in K$. Since \mathcal{M} is a metric space and K is relatively compact, there exists a subsequence $(g_{\varphi(n)})$ converging to $g \in G$. Since the action is continuous, $g_{\varphi(n)} x$ converges to gx . Therefore, $y = gx \in Gx$ and Gx is sequentially closed, hence closed.

Definition 9.12 (Registered points) We say that $x, y \in \mathcal{M}$ are registered points when $d(x, y) = d^0(Gx, Gy)$.

Lemma 9.13 (Length in a quotient metric space) We denote L the length on both \mathcal{M} and $\mathcal{M}^0 = \mathcal{M}/G$, and $\pi : \mathcal{M} \rightarrow \mathcal{M}^0$ the canonical projection. For all curve $c : [0, 1] \rightarrow \mathcal{M}$, we have $L(\pi \circ c) \leq L(c)$. In particular, if $x = c(0)$ and $y = c(1)$ are registered and if c is minimizing, then $\pi \circ c$ is minimizing and $L(\pi \circ c) = L(c) = d(x, y) = d^0(\pi(x), \pi(y))$.

Proof. By definition $L(\pi \circ c) \leq \sup \sum_{k=0}^p d^0(\pi(c(t_k)), \pi(c(t_{k+1}))) \leq \sup \sum_{k=0}^p d(c(t_k), c(t_{k+1})) = L(c)$. If $x = c(0)$ and $y = c(1)$ are registered, then $d(x, y) = d^0(\pi(x), \pi(y)) \leq L(\pi \circ c)$. If c is minimizing, then $d(x, y) = L(c) \geq L(\pi \circ c)$. Thus if both hold, then $L(\pi \circ c) = L(c) = d(x, y) = d^0(\pi(x), \pi(y))$ and $\pi \circ c$ is minimizing. \square

9.2.2.2 Quotient Riemannian metric

Definition 9.14 (Quotient metric) Let (\mathcal{M}, g) be a Riemannian manifold. Let G be a Lie group acting smoothly, properly, freely and isometrically on (\mathcal{M}, g) . Then there exists a unique smooth structure on $\mathcal{M}^0 = \mathcal{M}/G$ such that the quotient map $\pi : \mathcal{M} \rightarrow \mathcal{M}^0$ is a submersion [Lee, 2012]. Thus, one can define for all $x \in \mathcal{M}$:

1. (Vertical space) $\mathcal{V}_x = T_x \mathcal{M}_x = \ker d_x \pi$ where $\mathcal{M}_x = \pi^{-1}(x)$ is a submanifold of \mathcal{M} ,
2. (Horizontal space) $\mathcal{H}_x = \mathcal{V}_x^\perp$ so that $T_x \mathcal{M} = \mathcal{V}_x \oplus \mathcal{H}_x$,
3. (Horizontal lift) $\#_x : T_{\pi(x)} \mathcal{M}^0 \rightarrow \mathcal{H}_x$ the inverse isomorphism of $(d_x \pi)|_{\mathcal{H}_x} : \mathcal{H}_x \rightarrow T_{\pi(x)} \mathcal{M}^0$,
4. (Quotient metric) $g_{\pi(x)}^0 : (v, w) \in T_{\pi(x)} \mathcal{M}^0 \times T_{\pi(x)} \mathcal{M}^0 \mapsto g_x(v_x^\#, w_x^\#) \in \mathbb{R}$.

This is a particular case of a Riemannian submersion [O'Neill, 1966]. The Riemannian distance of the quotient metric g^0 is the quotient distance of d , defined by $d^0(\pi(x), \pi(y)) = d(Gx, Gy) = \inf_{g \in G} d(gx, y)$.

The fundamental theorem on geodesics of a quotient metric is the following.

Theorem 9.15 (Geodesics of a quotient metric) [O'Neill, 1966] The projection of a horizontal geodesic is a horizontal geodesic and their lengths coincide on any segment. More precisely, let $x \in \mathcal{M}$ and $v \in T_{\pi(x)} \mathcal{M}^0$. Then $I_{x, v_x^\#} \subseteq I_{\pi(x), v}$ and for all $t \in I_{x, v_x^\#}$, $\text{Exp}_{\pi(x)}(tv) = \pi(\text{Exp}_x(tv_x^\#))$.

9.2.2.3 Riemannian orbit spaces

When the action of G on \mathcal{M} is not free, the orbit space \mathcal{M}/G is not a Riemannian manifold in general. This was studied in [Alekseevsky et al., 2001]. We briefly recall the main facts that we use in this chapter.

Theorem 9.16 (Riemannian geometry of orbit spaces) [Alekseevsky et al., 2001] Let (\mathcal{M}, g) be a connected complete Riemannian manifold. Let G be a Lie group acting smoothly, properly and isometrically on (\mathcal{M}, g) . We denote $\pi : \mathcal{M} \rightarrow \mathcal{M}/G$ the canonical surjection. For a Lie subgroup H of G , we denote $(H) = \{gHg^{-1} | g \in G\}$ the conjugacy class of H and $\mathcal{M}_{(H)}$ the set of points $x \in \mathcal{M}$ such that the stabilizer of x , $\text{Stab}(x) = \{g \in G | gx = x\}$, belongs to (H) .

1. $\mathcal{M}_{(H)}$ is a smooth submanifold of \mathcal{M} .
2. We denote $(\mathcal{M}/G)_{(H)} = \pi(\mathcal{M}_{(H)}) = \mathcal{M}_{(H)}/G$ the (isotropy) stratum of type (H) . Then $\pi_{(H)} := \pi|_{\mathcal{M}_{(H)}} : \mathcal{M}_{(H)} \rightarrow (\mathcal{M}/G)_{(H)}$ is a smooth fiber bundle with fiber type G/H .
3. The isotropy strata form a partition of \mathcal{M}/G .
4. $(\mathcal{M}/G, d^0)$ is a complete metric space and a length space.

The definition of the vertical space still holds while the horizontal space is replaced by the normal space [Michor, 2008, VI.29.2].

Definition 9.17 (Vertical space, normal space) We take the notations of Theorem 9.16. Let $x \in \mathcal{M}_{(H)}$. Thus $\mathcal{M}_x = \pi^{-1}(\pi(x))$ is a submanifold of $\mathcal{M}_{(H)}$.

1. (Vertical space) $\mathcal{V}_x = T_x\mathcal{M}_x \subseteq T_x\mathcal{M}_{(H)} \subseteq T_x\mathcal{M}$.
2. (Normal space) $\mathcal{N}_x = \mathcal{V}_x^\perp \subseteq T_x\mathcal{M}$ so that $T_x\mathcal{M} = \mathcal{V}_x \oplus \mathcal{N}_x$.

Note that \mathcal{V}_x and \mathcal{N}_x need not have a constant dimension. We recall a result that we use later.

Lemma 9.18 (Geodesics in a Riemannian orbit space) [Alekseevsky et al., 2001, Lemma 3.5] We take the notation of Theorem 9.16. Given Lie subgroups H_1, H_2 of G , we denote $(H_1) \leq (H_2)$ if H_1 is conjugated so a subgroup of (H_2) . Let $\gamma : [0, 1] \rightarrow \mathcal{M}/G$ be a minimizing curve. For $t \in [0, 1]$, let $(\mathcal{M}/G)_{(H_t)}$ denote the stratum of $\gamma(t)$. Then, for all $t \in (0, 1)$, $(H_t) \leq (H_0)$ and $(H_t) \leq (H_1)$.

We are now well prepared to study the construction of the orbit space of covariance matrices (Section 9.3), the geodesics within each stratum (Sections 9.4 and 9.5) and the minimizing geodesics in the whole space (Section 9.6).

9.3 Bures-Wasserstein geometry of covariance matrices

We denote the set of symmetric positive semi-definite matrices or covariance matrices by $\text{Cov}(n) = \{\Sigma \in \text{Sym}(n) | \Sigma \geq 0\} = \{XX^\top | X \in \text{Mat}(n)\} \subset \text{Mat}(n)$. It is a complete metric

subspace of the vector space of $n \times n$ square matrices equipped with the Euclidean distance $d^E(\Sigma, \Lambda) = \|\Sigma - \Lambda\|_2 = \text{tr}((\Sigma - \Lambda)^2)^{1/2}$.

In this section, we recall that this set can also be described as the orbit space of the manifold $\text{Mat}(n)$ of $n \times n$ matrices quotiented on the right by the orthogonal group $O(n)$. We recall that the quotient topology coincides with the Euclidean topology and we recall the expression of the quotient distance known as the Bures-Wasserstein distance. We insist on the definition of the strata, which are investigated in the following sections.

9.3.1 The quotient geometry of covariance matrices

The **group action** of the orthogonal group $O(n)$ on the vector space of square matrices $\text{Mat}(n)$ is $(X, U) \in \text{Mat}(n) \times O(n) \mapsto XU \in \text{Mat}(n)$. It is smooth, proper and isometric for the Euclidean distance.

The **stabilizer** (or isotropy group) of a matrix $X \in \text{Mat}(n)$ is $\text{Stab}(X) = \{U \in O(n) | XU = X\}$. If we denote $k = \text{rk}(X)$, it is well known that X is equivalent to the matrix $J_k = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$, i.e. there exist $P \in \text{GL}(n)$ and $Q \in O(n)$ such that $X = PJ_kQ$. Then

it is clear that $H_k := \text{Stab}(J_k) = \left\{ \begin{pmatrix} I_k & 0 \\ 0 & U \end{pmatrix} \mid U \in O(n-k) \right\}$ and $\text{Stab}(X) = Q^\top \text{Stab}(J_k) Q$

with $\dim \text{Stab}(X) = \dim O(n-k) = \frac{(n-k)(n-k-1)}{2}$. Hence two matrices have conjugate stabilizers if and only if they have the same rank. Note that $(H_k) \leq (H_l)$ if and only if $k \geq l$.

The **orbit** of X is $\text{Orb}(X) = \{XU \mid U \in O(n)\} \simeq O(n)/\text{Stab}(X)$. Its dimension is $\dim \text{Orb}(X) = \dim O(n) - \dim \text{Stab}(X) = nk - \frac{k(k+1)}{2}$. Given that there exists $U \in O(n)$ such that $Y = XU$ if and only if $XX^\top = YY^\top$ [Groetzner and Dür, 2020, Lemma 2.6], we have $\text{Orb}(X) = \{Y \in \text{Mat}(n) \mid YY^\top = XX^\top\}$.

The **orbit space** $\text{Mat}(n)/O(n) = \{\text{Orb}(X) \mid X \in \text{Mat}(n)\}$ is thus in bijection with the set of covariance matrices $\text{Cov}(n) = \{XX^\top \mid X \in \text{Mat}(n)\}$ by the map $\text{Orb}(X) \mapsto XX^\top$.

The **orbit strata** of $\text{Mat}(n)/O(n)$ are the sets of points that have conjugate stabilizers [Alekseevsky et al., 2001], i.e. that have the same rank here. This gives a manifold structure to $\text{Mat}(n)_{(H_k)} = \mathbb{R}_k^{n \times n}$. The strata of $\text{Mat}(n)/O(n)$ are $\mathbb{R}_k^{n \times n}/O(n)$, or equivalently the strata of covariance matrices are the sets of symmetric positive semi-definite matrices of fixed rank $\text{Sym}^+(n, k) = \text{Cov}(n) \cap \mathbb{R}_k^{n \times n}$. The principal/regular stratum is the set of Symmetric Positive Definite (SPD) matrices $\text{Sym}^+(n) = \text{Cov}(n) \cap \text{GL}(n)$. Finally, $\pi_k := \pi_{(H_k)} : \mathbb{R}_k^{n \times n} \rightarrow \text{Sym}^+(n, k)$ is a smooth fiber bundle with fiber type $\text{St}(n, k) = O(n)/O(n-k)$, in particular it is a submersion.

9.3.2 The Bures-Wasserstein distance

Since the group action is continuous, proper and isometric, the Euclidean distance descends to a distance on $\text{Mat}(n)/O(n)$. Via the bijection $\text{Orb}(X) \in \text{Mat}(n)/O(n) \mapsto XX^\top \in \text{Cov}(n)$, it is usually expressed as a distance on covariance matrices. It is known as the Bures-Wasserstein distance [Dowson and Landau, 1982, Olkin and Pukelsheim, 1982, Bhatia et al., 2019].

Definition 9.19 (Bures-Wasserstein distance) [Bhatia et al., 2019] The Bures-Wasserstein distance between Σ and Λ is defined by:

$$d^{\text{BW}}(\Sigma, \Lambda) = \inf_{\substack{X, Y \in \text{Mat}(n) \\ XX^\top = \Sigma, YY^\top = \Lambda}} d^{\text{E}}(X, Y) = \inf_{R \in \text{O}(n)} d^{\text{E}}(\Sigma^{1/2}, \Lambda^{1/2}R) \quad (9.1)$$

$$= \text{tr}(\Sigma + \Lambda - 2(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2})^{1/2}. \quad (9.2)$$

If $XX^\top = \Sigma$ and $YY^\top = \Lambda$, let $R \in \text{O}(n)$ such that $X^\top Y = (X^\top \Lambda X)^{1/2}R$. Then $d^{\text{E}}(X, YR^\top)^2 = \|YR^\top - X\|_2^2 = \text{tr}(XX^\top + YY^\top - 2X^\top YR^\top) = d^{\text{BW}}(\Sigma, \Lambda)^2$.

As the quotient of a length space, the space of covariance matrices endowed with the Bures-Wasserstein metric is a length space. It is even a complete geodesic metric space [Alekseevsky et al., 2001, Proposition 3.1.(1)].

The following result seems elementary although we did not find a clear reference in the literature.

Lemma 9.20 (Euclidean and Bures-Wasserstein topologies coincide) The Euclidean distance d^{E} and the Bures-Wasserstein distance d^{BW} define the same topology on $\text{Cov}(n)$.

See the proof of Lemma 9.20 in Section 11.8.

9.3.3 Topology, metric and smooth structure of the strata

The set of symmetric positive definite (SPD) matrices $\text{Sym}^+(n) = \{\Sigma \in \text{Sym}(n) \mid \text{sp}(\Sigma) \subset (0, +\infty)\}$ is an open set of the vector space of symmetric matrices, hence it has a natural structure of smooth manifold. This topology clearly coincides with the topology induced by $(\text{Cov}(n), d^{\text{E}})$, thus it also coincides with the topology induced by $(\text{Cov}(n), d^{\text{BW}})$.

The set $\text{Sym}^+(n, k)$ is in bijection with $\mathbb{R}_k^{n \times n} / \text{O}(n)$. The Euclidean distance on $\mathbb{R}_k^{n \times n}$ descends to the Bures-Wasserstein distance on $\text{Sym}^+(n, k)$ thanks to Equation (9.1). Therefore, the quotient topology coincides with the Bures-Wasserstein topology induced by $(\text{Cov}(n), d^{\text{BW}})$, thus also with the Euclidean topology induced by $(\text{Cov}(n), d^{\text{E}})$ and $(\text{Sym}(n), d^{\text{E}})$.

However, the Riemannian geometry is difficult to study via the submersion $\pi_k : \mathbb{R}_k^{n \times n} \rightarrow \text{Sym}^+(n, k)$. Indeed, it is the projection of a bundle of fiber $\text{St}(n, k) \simeq \text{O}(n) / \text{O}(n - k)$ which is not a Lie group. Fortunately, the set $\text{Sym}^+(n, k)$ is also in bijection with the quotient manifold $\mathbb{R}_*^{n \times k} / \text{O}(k)$ [Massart and Absil, 2020, Proposition 2.1], where $\mathbb{R}_*^{n \times k}$ is the open set of matrices of full rank in $\mathbb{R}^{n \times k}$. We recall this quotient geometry in Table 9.1. The quotient distance induced on $\text{Sym}^+(n, k)$ is the Bures-Wasserstein distance [Massart and Absil, 2020, Proposition 5.1]. In particular, the quotient topology coincides with the previous ones. This bijection naturally provides a smooth structure on $\text{Sym}^+(n, k)$. Above all, the submersion $\tilde{\pi}_k : \mathbb{R}_*^{n \times k} \rightarrow \text{Sym}^+(n, k)$ is the projection of a principal fiber bundle. Hence, it is much more convenient to study the Bures-Wasserstein Riemannian geometry of $\text{Sym}^+(n, k)$ via the Riemannian submersion $\tilde{\pi}_k$. This is exactly what is done in [Massart and Absil, 2020].

To summarize, the strata $\text{Sym}^+(n, k)$ are smooth connected manifolds and the regular stratum $\text{Sym}^+(n)$ is a dense open set in $\text{Cov}(n)$.

The Riemannian geometry of the principal stratum was extensively studied [Takatsu, 2010, Takatsu, 2011, Malagò et al., 2018, Bhatia et al., 2019, van Oostrum, 2020, Thanwerdas and Pennec, 2022b] and the Riemannian geometry of the other strata was recently well

Set	$\text{Sym}^+(n, k) = \text{Cov}(n) \cap \mathbb{R}_k^{n \times n}$
Smooth manifold	$\mathbb{R}_*^{n \times k} / \text{O}(k)$
Group action	$\begin{cases} \mathbb{R}_*^{n \times k} \times \text{O}(k) & \longrightarrow \mathbb{R}_*^{n \times k} \\ (X, U) & \longmapsto XU \end{cases}$
Orbit	$\text{Orb}(X) = \{Y \in \mathbb{R}_*^{n \times k} \mid YY^\top = XX^\top\}$
Identification	$\begin{cases} \mathbb{R}_*^{n \times k} / \text{O}(k) & \longrightarrow \text{Sym}^+(n, k) \\ \text{Orb}(X) & \longmapsto XX^\top \end{cases}$
Submersion	$\pi_{\text{Sym}^+(n, k)} : \begin{cases} \mathbb{R}_*^{n \times k} & \longrightarrow \text{Sym}^+(n, k) \\ X & \longmapsto XX^\top \end{cases}$

Table 9.1: Smooth manifold structure of $\text{Sym}^+(n, k)$.

detailed [Massart et al., 2019, Massart and Absil, 2020]. However, there remain missing formulae and open questions about the geodesics in each stratum: injectivity domain, preimages, and explicit formulae of the horizontal lift, the exponential map and logarithms in the base space $\text{Sym}^+(n, k)$ of the principal fiber bundle $\mathbb{R}_*^{n \times k} \longrightarrow \text{Sym}^+(n, k)$. We precisely answer these questions in Section 9.4 (full-rank matrices) and Section 9.5 (low-rank matrices). Furthermore, we contribute in Section 9.6 the minimizing geodesics for the Bures-Wasserstein *distance* between different strata and the condition of uniqueness of the geodesic between two points.

9.4 Geodesics of the Bures-Wasserstein metric on $\text{Sym}^+(n)$

In this section, we give complements and new results on the Bures-Wasserstein geodesics on SPD matrices $\text{Sym}^+(n)$. The quotient structure is well known, as well as the exponential map [Malagò et al., 2018] and the injectivity radius [Massart and Absil, 2020]. The definition interval of the geodesic was implicitly described in [Malagò et al., 2018] as the connected component of 0 in a subset of \mathbb{R} so we give it explicitly here. It was proved in [Bhatia et al., 2019] that there exists a preimage which is a logarithm. We prove the uniqueness of the preimage and the logarithm based on a result of [Massart and Absil, 2020] on $\text{Sym}^+(n, k)$ applied for $k = n$. Moreover, we contribute the cut time, thus the injectivity domain. The proof is deferred to Section 11.8. After Theorem 9.22 on Bures-Wasserstein geodesics, we show on an example that we already know some geodesics between degenerate matrices that cross the main stratum of SPD matrices.

Definition 9.21 (Bures-Wasserstein metric on $\text{Sym}^+(n)$) [Malagò et al., 2018, Bhatia et al., 2019, van Oostrum, 2020] The Bures-Wasserstein metric on $\text{Sym}^+(n)$ is the quotient Riemannian metric induced by the submersion $\pi : X \in \text{GL}(n) \longmapsto XX^\top \in \text{Sym}^+(n)$ and the Frobenius metric on $\text{GL}(n)$. Let $X \in \text{GL}(n)$ such that $XX^\top = \Sigma$ and let $V \in T_\Sigma \text{Sym}^+(n) \equiv \text{Sym}(n)$. The quotient operations are:

1. (Vertical space) $\mathcal{V}_X = \ker d_X \pi = X \text{Skew}(n)$,
2. (Horizontal space) $\mathcal{H}_X = \text{Sym}(n) X$,
3. (Horizontal lift) $V_X^\# = \mathcal{S}_\Sigma(V) X \in \mathcal{H}_X$,

4. (Bures-Wasserstein metric) $g_{\Sigma}^{BW(n)}(V, V) = \text{tr}(\mathcal{S}_{\Sigma}(V)\Sigma\mathcal{S}_{\Sigma}(V))$,

where $\mathcal{S}_{\Sigma}(V) \in \text{Sym}(n)$ is the unique solution of the Sylvester equation $\Sigma\mathcal{S}_{\Sigma}(V) + \mathcal{S}_{\Sigma}(V)\Sigma = V$.

Theorem 9.22 (Bures-Wasserstein geodesics on $\text{Sym}^+(n)$) Let $\Sigma \in \text{Sym}^+(n)$.

1. (Exponential map) [Malagò et al., 2018] For all $V \in T_{\Sigma}\text{Sym}^+(n) \equiv \text{Sym}(n)$, the geodesic from Σ with initial speed V writes $\gamma_{(\Sigma, V)}(t) = \Sigma + tV + t^2\mathcal{S}_{\Sigma}(V)\Sigma\mathcal{S}_{\Sigma}(V) \in \text{Sym}^+(n)$.
2. (Definition interval) Let $\lambda_{\max} = \max \text{sp}(\mathcal{S}_{\Sigma}(V))$ and $\lambda_{\min} = \min \text{sp}(\mathcal{S}_{\Sigma}(V))$. The definition interval of the geodesic $\gamma_{(\Sigma, V)}$ is the interval $I_{\Sigma, V}$ defined by:

- $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, -\frac{1}{\lambda_{\min}})$ if $\lambda_{\min} < 0 < \lambda_{\max}$,
- $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_{\min}})$ if $\lambda_{\min} < 0$ and $\lambda_{\max} \leq 0$,
- $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, +\infty)$ if $\lambda_{\min} \geq 0$ and $\lambda_{\max} > 0$,
- $I_{\Sigma, V} = \mathbb{R}$ if $\lambda_{\min} = \lambda_{\max} = 0$ (which only happens for $V = 0$).

3. (Cut time) The cut time is $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$ if $\lambda_{\min} < 0$ or $+\infty$ otherwise. The geodesic $\gamma_{(\Sigma, V)} : I_{\Sigma, V} \rightarrow \mathcal{M}$ is even minimizing on $I_{\Sigma, V}$.
4. (Logarithm map) For all $\Lambda \in \text{Sym}^+(n)$, there exists a unique preimage $V \in \mathcal{P}re_{\Sigma}(\Lambda)$. It writes $V = 2\text{sym}(\Sigma^{1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}\Sigma^{-1/2}) - 2\Sigma$, where we denote $\text{sym}(M) = \frac{1}{2}(M + M^{\top})$. The geodesic joining Σ to Λ writes:

$$\gamma_{\Sigma \rightarrow \Lambda}(t) = (1-t)^2\Sigma + t^2\Lambda + 2t(1-t)\text{sym}(\Sigma^{1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}\Sigma^{-1/2}). \quad (9.3)$$

Moreover, it is a logarithm: $V \in \mathcal{L}og_x(y)$. Thus the logarithm map is defined on $\mathcal{U}_{\Sigma} = \text{Sym}^+(n)$ and it writes:

$$\text{Log}_{\Sigma} : \begin{cases} \text{Sym}^+(n) & \longrightarrow & T_{\Sigma}\text{Sym}^+(n) \\ \Lambda & \longmapsto & 2\text{sym}(\Sigma^{1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}\Sigma^{-1/2}) - 2\Sigma \end{cases} \quad (9.4)$$

See the proof of Theorem 9.22 in Section 11.8.

Remark 9.23 The minimizing geodesic $\gamma_{(\Sigma, V)} : I_{\Sigma, V} \rightarrow \text{Sym}^+(n)$ clearly has a limit at the finite boundaries of $I_{\Sigma, V}$. When $I_{\Sigma, V}$ is bounded, we can define $\Sigma_0 = \lim_{t \rightarrow -1/\lambda_{\max}} \gamma_{(\Sigma, V)}(t)$, $\Sigma_1 = \lim_{t \rightarrow -1/\lambda_{\min}} \gamma_{(\Sigma, V)}(t)$ and the extended curve $\gamma : \bar{I}_{\Sigma, V} = [-1/\lambda_{\max}, -1/\lambda_{\min}]$ by $\gamma(t) = \gamma_{(\Sigma, V)}(t)$ for $t \in I_{\Sigma, V}$, $\gamma(-1/\lambda_{\max}) = \Sigma_0$ and $\gamma(-1/\lambda_{\min}) = \Sigma_1$. The curve γ is a minimizing geodesic on $I_{\Sigma, V}$. Thus, by Lemma 9.3 (continuity of the length), the curve γ is a minimizing geodesic on $[-1/\lambda_{\max}, -1/\lambda_{\min}]$. So we already have examples of minimizing geodesics between two degenerate matrices which pass through the principal stratum $\text{Sym}^+(n)$. For instance, we have $\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}$ which are linked by the

geodesic $\text{Exp}_{I_2}(tV) = \begin{pmatrix} (1+t)^2 & 0 \\ 0 & (1-t)^2 \end{pmatrix}$ for $t \in (-1, 1)$ with $V = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$.

Moreover, from the viewpoint of the Levi-Civita connection, the curve $t \mapsto \text{Exp}_\Sigma(tV)$ is a geodesic (self-parallel curve) on each subinterval of the set $J_{\Sigma,V} = \{t \in \mathbb{R} \mid \text{Exp}_\Sigma(tV) \in \text{Sym}^+(n)\} = \mathbb{R} \setminus \{t \in \mathbb{R} \mid -\frac{1}{t} \in \text{sp}(\mathcal{S}_\Sigma(V))\}$, which is \mathbb{R} without a maximum of n points. Theorem 9.22 actually states that every geodesic is minimizing on its domain. Hence, the minimizing geodesic $\gamma_{(\Sigma,V)} : I_{\Sigma,V} \rightarrow \text{Sym}^+(n)$ naturally extends to a curve $\gamma_{(\Sigma,V)} : \mathbb{R} \rightarrow \text{Cov}(n)$ which is a minimizing geodesic on the segments delimited by two consecutive values in $\mathbb{R} \setminus J_{\Sigma,V}$.

9.5 Geodesics of the Bures-Wasserstein metric on $\text{Sym}^+(n, k)$

In this section, we give complements and new results on the Bures-Wasserstein geodesics on the manifold of PSD matrices of fixed rank k , $\text{Sym}^+(n, k)$. They were mainly studied in [Massart and Absil, 2020]. The formulae of the exponential map and its definition domain were kept implicit because they were formulated in function of horizontal vectors in the total space $\mathbb{R}_*^{n \times k}$ of matrices of full-rank k . We compute the horizontal lift, which allows us to express the Bures-Wasserstein metric, the exponential map and the definition interval directly in function of the tangent vector. Moreover, we characterize the preimages of the exponential map. In [Massart and Absil, 2020], they characterize the logarithms by solving the underlying matrix equation and then imposing the condition of minimizing norm. Based on their resolution of the matrix equation, we impose the weaker condition that the preimage $v \in \text{Pre}_x(y)$ must satisfy $1 \in I_{x,v}$; otherwise, the geodesic $\gamma_{x,v} : I_{x,v} \rightarrow \mathcal{M}$ could leave the space before reaching y . This is how we characterize all the preimages. We also give an explicit formula of the minimizing geodesic joining two points when it is unique and we specify the number of minimizing geodesics between two points otherwise. Finally, we compute the injectivity domain that was kept implicit in [Massart and Absil, 2020]. After Theorem 9.26, we precisely specify the novelty of our result with respect to the reference work [Massart and Absil, 2020]. Then, we give examples to illustrate the possible cases for the number of preimages and logarithms.

Definition 9.24 (Bures-Wasserstein metric on $\text{Sym}^+(n, k)$) [Massart and Absil, 2020] The Bures-Wasserstein metric on $\text{Sym}^+(n, k)$ is the quotient Riemannian metric induced by the submersion $\pi : X \in \mathbb{R}_*^{n \times k} \mapsto XX^\top \in \text{Sym}^+(n, k)$ and the Frobenius metric on $\mathbb{R}^{n \times k}$. Let $X \in \mathbb{R}_*^{n \times k}$ such that $XX^\top = \Sigma$ and let $V \in T_\Sigma \text{Sym}^+(n, k)$. The vertical and horizontal spaces are:

1. (Vertical space) $\mathcal{V}_X = \ker d_X \pi = X \text{Skew}(k)$,
2. (Horizontal space) $\mathcal{H}_X = \{X(X^\top X)^{-1}F + X_\perp K, F \in \text{Sym}(k), K \in \text{Mat}(n-k, k)\}$ where $X_\perp \in \text{Mat}(n, n-k)$ has orthonormal columns ($X_\perp^\top X_\perp = I_{n-k}$) that are orthogonal to the columns of X ($X^\top X_\perp = 0$).

Theorem 9.25 (Horizontal lift, tangent space, metric) Let $\Sigma \in \text{Sym}^+(n, k)$, let $X \in \mathbb{R}_*^{n \times k}$ such that $\Sigma = XX^\top \in \text{Sym}^+(n, k)$ and let $V \in T_\Sigma \text{Sym}^+(n, k)$. Let $\Sigma = UDU^\top$ be a singular value decomposition with $D \in \text{Diag}^+(k)$ and $U \in \text{St}(n, k)$. We denote $S = S_{\Sigma,V} = U\mathcal{S}_D(U^\top VU)U^\top$, where $\mathcal{S}_A(B)$ denotes the unique solution Z of the Sylvester equation $AZ + ZA = B$. Note that $S_{\Sigma,V}$ and $(I_n - UU^\top)$ are independent from the chosen decomposition.

1. (Tangent space) $T_\Sigma \text{Sym}^+(n, k) = \{V \in \text{Sym}(n) \mid X_\perp^\top V X_\perp = 0\}$,

2. (Horizontal lift) $V_X^\# = X(X^\top X)^{-1} \mathcal{S}_{X^\top X}(X^\top V X) + \underbrace{(I_n - X(X^\top X)^{-1} X^\top)}_{X_\perp X_\perp^\top} V X(X^\top X)^{-1}$,
3. (Bures-Wasserstein metric) $g_\Sigma^{BW(n,k)}(V, V) = \text{tr}(S_{\Sigma, V} \Sigma S_{\Sigma, V} + V \Sigma^\dagger V (I_n - U U^\top))$.

See the proof of Theorem 9.25 in Section 11.8.

Theorem 9.26 (Bures-Wasserstein geodesics on $\text{Sym}^+(n, k)$) Let $\Sigma, \Lambda \in \text{Sym}^+(n, k)$ and $X, Y \in \mathbb{R}_*^{n \times k}$ such that $XX^\top = \Sigma$ and $YY^\top = \Lambda$. Let $U \in \text{St}(n, k)$ and $D \in \text{Diag}^+(k)$ such that $\Sigma = UDU^\top$.

1. (Exponential map) For all $V \in T_\Sigma \text{Sym}^+(n, k)$, the geodesic from Σ with initial speed V is $\gamma_{(\Sigma, V)} : t \in I_{\Sigma, V} \mapsto \Sigma + tV + t^2 W_{\Sigma, V}$, where $W_{\Sigma, V} = S_{\Sigma, V} \Sigma S_{\Sigma, V} + S_{\Sigma, V} V (I_n - U U^\top) + (I_n - U U^\top) V S_{\Sigma, V} + (I_n - U U^\top) V \Sigma^\dagger V (I_n - U U^\top)$ and $S_{\Sigma, V} = U \mathcal{S}_D(U^\top V U) U^\top$.
2. (Definition interval) Let $F_{X, V}^0 = \mathcal{S}_{X^\top X}((X^\top X)^{-1/2} X^\top V X (X^\top X)^{-1/2})$ and $M_{X, V}^0 = (X^\top X)^{-3/2} X^\top V (I_n - X(X^\top X)^{-1} X^\top) V X (X^\top X)^{-3/2} \in \text{Sym}(n)$. Let $\mathcal{E}_{\Sigma, V} = \{\lambda \in \text{sp}(F_{X, V}^0) \mid \ker(\lambda I_k - F_{X, V}^0) \cap \ker(M_{X, V}^0) \neq \{0\}\} \subseteq \text{sp}(F_{X, V}^0) = \text{sp}(S_{\Sigma, V})$. If $\mathcal{E}_{\Sigma, V}$ is non-empty, then let $\lambda_+ = \max \mathcal{E}_{\Sigma, V}$ and $\lambda_- = \min \mathcal{E}_{\Sigma, V}$. The definition interval of the geodesic $\gamma_{(\Sigma, V)}$ is the interval $I_{\Sigma, V}$ defined by:

- $I_{\Sigma, V} = (-\frac{1}{\lambda_+}, -\frac{1}{\lambda_-})$ if $\lambda_- < 0 < \lambda_+$,
- $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_-})$ if $\lambda_- < 0$ and $\lambda_+ \leq 0$,
- $I_{\Sigma, V} = (-\frac{1}{\lambda_+}, +\infty)$ if $\lambda_- \geq 0$ and $\lambda_+ > 0$,
- $I_{\Sigma, V} = \mathbb{R}$ if $\mathcal{E}_{\Sigma, V}$ is empty.

Applying this to $X = UD^{1/2}$ without loss of generality, $F_{X, V}^0 = \mathcal{S}_D(U^\top V U)$ and $M_{X, V}^0 = D^{-1} U^\top V (I_n - U U^\top) V U D^{-1}$ which is a bit more tractable to compute $\mathcal{E}_{\Sigma, V}$.

3. (Cut time) Let $\lambda_{\max} = \max \text{sp}(F_{X, V}^0)$ and $\lambda_{\min} = \min \text{sp}(F_{X, V}^0)$. Note that if $\mathcal{E}_{\Sigma, V} \neq \emptyset$, then we have $(\lambda_-, \lambda_+) \subseteq (\lambda_{\min}, \lambda_{\max})$. The cut time is $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$ if $\lambda_{\min} < 0$ or $+\infty$ otherwise. Symmetrically, we have $t_{\text{cut}}(\Sigma, -V) = \frac{1}{\lambda_{\max}}$ if $\lambda_{\max} > 0$ or $+\infty$ otherwise.
4. (Preimages) We define the indexing set $\mathcal{I}_{X, Y}^{\text{Pre}}$ by:

$$\begin{aligned} \mathcal{I}_{X, Y}^{\text{Pre}} = \{R \in \text{O}(n) \mid H := X^\top Y R^\top \in \text{Sym}(n) \text{ and} \\ \forall \mu < 0, \ker(\mu I_k - (X^\top X)^{-1/2} H (X^\top X)^{-1/2}) \\ \cap \ker(\mu^2 I_k - (X^\top X)^{-1/2} R Y^\top Y R^\top (X^\top X)^{-1/2}) = \{0\}\}. \end{aligned}$$

For $R \in \mathcal{I}_{X, Y}^{\text{Pre}}$, we denote $H = H_{X, Y, R} = X^\top Y R^\top$ so that $X^\top Y = H R$.

Then, the map $R \in \mathcal{I}_{X, Y}^{\text{Pre}} \mapsto V = 2 \text{sym}(X R Y^\top) - 2 \Sigma \in \mathcal{P}re_\Sigma(\Lambda)$ is a bijection whose inverse is $V \in \mathcal{P}re_\Sigma(\Lambda) \mapsto R = (Y^\top Y)^{-1} Y^\top (X + V_X^\#) \in \mathcal{I}_{X, Y}^{\text{Pre}}$.

The geodesic joining Σ to Λ parametrized by $R \in \mathcal{I}_{X, Y}^{\text{Pre}}$ writes:

$$\forall t \in [0, 1], \gamma_{\Sigma \rightarrow \Lambda}^R(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(X R Y^\top). \quad (9.5)$$

5. (Logarithms) Let $\mathcal{I}_{X,Y}^{\mathcal{L}og} = \{R \in O(n) | H_{X,Y,R} = X^\top Y R^\top \in \text{Cov}(n)\} = \{R \in O(n) | H_{X,Y,R} = (X^\top \Lambda X)^{1/2}\} = \{R \in O(n) | X^\top Y = (X^\top \Lambda X)^{1/2} R\} \subseteq \mathcal{I}_{X,Y}^{\mathcal{P}re}$.

Then, the map $R \in \mathcal{I}_{X,Y}^{\mathcal{L}og} \mapsto V = 2 \text{sym}(X R Y^\top) - 2\Sigma \in \mathcal{L}og_\Sigma(\Lambda)$ is a bijection whose inverse is $V \in \mathcal{L}og_\Sigma(\Lambda) \mapsto R = (Y^\top Y)^{-1} Y^\top (X + V_X^\#) \in \mathcal{I}_{X,Y}^{\mathcal{L}og}$.

6. (Logarithm map) Let $r = \text{rk}(\Sigma \Lambda) = \text{rk}(X^\top Y) = \text{rk}(H)$.

- (a) If $r = k$, then there exists a unique logarithm of Λ from Σ . In this case, the minimizing geodesic joining Σ to Λ writes:

$$\gamma_{\Sigma \rightarrow \Lambda}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2} ((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda). \quad (9.6)$$

- (b) If $r = k - 1$, then there exist exactly two logarithms of Λ from Σ .

- (c) If $r < k - 1$, then there is an infinity of logarithms of Λ from Σ .

Therefore, the logarithm map is defined on $\mathcal{U}_\Sigma = \{\Lambda \in \text{Sym}^+(n, k) | \text{rk}(\Sigma \Lambda) = k\}$ and it writes $\text{Log}_\Sigma : \Lambda \in \mathcal{U}_\Sigma \mapsto 2 \text{sym}(\Sigma^{1/2} ((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda) - 2\Sigma \in T_\Sigma \text{Sym}^+(n, k)$.

See the proof of Theorem 9.26 in Section 11.8.

Remark 9.27 Let us clarify our contributions in Theorem 9.26 with respect to the reference paper [Massart and Absil, 2020].

1. (Exponential map) The formula is new. Only the exponential map of a horizontal vector in the total space $\mathbb{R}_*^{n \times k}$ was given in [Massart and Absil, 2020].
2. (Definition interval) The definition interval was formulated in the total space $\mathbb{R}_*^{n \times k}$ in [Massart and Absil, 2020]. The novelty here is to formulate it in function of V thanks to the horizontal lift.
3. (Cut time) This is new.
4. (Preimages) Note that $\mathcal{I}_{X,Y}^{\mathcal{P}re} \subseteq \mathcal{I}_{X,Y}^{\mathcal{S}ol} := \{R \in O(n) | X^\top Y R^\top \in \text{Sym}(n)\}$. As shown in [Massart and Absil, 2020], the set $\mathcal{I}_{X,Y}^{\mathcal{S}ol}$ indexes the solutions of the logarithm equation $\text{Exp}_\Sigma(V) = \Lambda$. The characterization of $\mathcal{I}_{X,Y}^{\mathcal{P}re} = \{R \in O(n) | 1 \in I_{\Sigma, V}\}$ (with $V = X R Y^\top + Y R^\top X^\top - 2\Sigma$) is explicitly dismissed in [Massart and Absil, 2020]. The curve “ γ^R ” with $R \in \mathcal{I}_{X,Y}^{\mathcal{S}ol}$ may hit the boundary before reaching Λ . Therefore, among these candidate R 's such that $X^\top Y R^\top \in \text{Sym}(n)$, we specify the set of R 's that really define a geodesic from Σ to Λ in the manifold $\text{Sym}^+(n, k)$, based on the condition underlying the definition interval.
5. (Logarithms) It was stated in [Massart and Absil, 2020] that if R leads to a logarithm then $H \geq 0$, and that if $H \geq 0$, then R leads to a preimage which is additionally a logarithm. However, it is not stated clearly that in this case, H has to be equal to $(X^\top \Lambda X)^{1/2}$. It is important for the next point though. The expression of the logarithms is new, although very straightforward. It was not given in [Massart and Absil, 2020] because they prefer to work in the total space $\mathbb{R}_*^{n \times k}$.

6. (Logarithm map) It was stated that the logarithm is unique if and only if $r = k$ in [Massart and Absil, 2020]. However, it was not stated that there are exactly two logarithms when $r = k - 1$ and that there is an infinity of *logarithms* when $r < k - 1$. The expression of the minimizing geodesic when it is unique is also new.

In other words, we have a minor contribution on the reformulation in $\text{Sym}^+(n, k)$ of results stated in the total space, and more important contributions on the injectivity domain (cut time), the expression of the minimizing geodesic when it is unique and the clarification between the three sets $\mathcal{I}_{X,Y}^{\text{Log}} \subseteq \mathcal{I}_{X,Y}^{\text{Pre}} \subseteq \mathcal{I}_{X,Y}^{\text{Sol}} := \{R \in \text{O}(n) \mid X^\top Y R^\top \in \text{Sym}(n)\}$. We give several examples below to illustrate the differences between these three sets.

Remark 9.28 The definitions of the sets $\mathcal{E}_{\Sigma,V}$ and $\mathcal{I}_{\Sigma,V}^{\text{Pre}}$ might seem intricate. It is difficult to simplify them though. Nevertheless, the set $\mathcal{E}_{\Sigma,V}$ is easy to determine numerically. Analogously, it is easy to determine numerically if a candidate $R \in \mathcal{I}_{X,Y}^{\text{Sol}}$ given in [Massart and Absil, 2020, Lemma 4.1] belongs to $\mathcal{I}_{X,Y}^{\text{Pre}}$.

Examples 9.29 Let us denote $R_\theta^- = \begin{pmatrix} -\cos \theta & \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \in \text{O}(2) \cap \text{Sym}(2)$ with $\det(R_\theta^-) = -1$.

1. Let $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 4\Sigma$ in $\text{Sym}^+(3, 2)$ with $r = 2$. Then, let

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } Y = \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}. \text{ Then } XX^\top = \Sigma, YY^\top = \Lambda \text{ and } X^\top Y = 2I_2. \text{ Thus}$$

the candidate pairs for (H, R) are $(2I_2, I_2)$, $(-2I_2, -I_2)$ and $(2R_\theta^-, R_\theta^-)$. One can show that:

- (a) $R = I_2$ leads to the unique minimizing geodesic $\gamma_{\Sigma \rightarrow \Lambda}(t) = (1+t)^2 \Sigma \in \text{Sym}^+(3, 2)$ for $t \in (-1, +\infty) \supset [0, 1]$,
- (b) there is no non-minimizing geodesic,
- (c) $R = -I_2$ or $R = R_\theta^-$ lead to curves that hit $\text{Sym}^+(3, 1)$ at $t = \frac{1}{3} < 1$, e.g. $R = -I_2$ leads to the curve $\gamma(t) = (1-3t)^2 \Sigma \in \text{Sym}^+(3, 2)$ only for $t \in (-\infty, \frac{1}{3})$.

2. Let $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ in $\text{Sym}^+(3, 2)$ with $r = 2$ again. Then, let

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ and } Y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \text{ Then } X^\top Y = I_2 \text{ so the candidate pairs for } (H, R)$$

are (I_2, I_2) , $(-I_2, -I_2)$ and (R_θ^-, R_θ^-) . One can show that:

- (a) $R = I_2$ leads to the geodesic $\gamma_{\Sigma \rightarrow \Lambda}^{I_2}(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & t & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$ for $t \in \mathbb{R}$

which is minimizing on $\left[-\frac{2}{1+\sqrt{2}}, \frac{2}{\sqrt{2}-1}\right] \supset [0, 1]$,

(b) $R = R_\theta^-$ for $\theta \neq 0$ lead to non-minimizing geodesics, e.g. $R = R_\pi^-$ leads to the curve $\gamma_{\Sigma \rightarrow \Lambda}^{R_\pi^-}(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (1-2t)^2 & -t(1-2t) \\ 0 & -t(1-2t) & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$ for $t \in [0, 1]$,

(c) $R = R_0^-$ and $R = -I_2$ lead to curves that hit $\text{Sym}^+(3, 1)$ at $t = \frac{1}{2} < 1$, e.g. $R = R_0^-$ leads to the curve $\gamma(t) = \begin{pmatrix} (1-2t)^2 & 0 & 0 \\ 0 & 1 & t \\ 0 & t & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$ only for $t \in (\infty, \frac{1}{2})$.

3. Let $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ in $\text{Sym}^+(3, 2)$ with $r = 1$. Then, let $X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $Y = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$. Then $X^\top Y = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ so the candidate values of R are $\text{Diag}(\pm 1, \pm 1)$. One can show that:

(a) $R_\pm^0 = \text{Diag}(1, \pm 1)$ lead to two minimizing geodesics whose expressions are $\gamma_{\Sigma \rightarrow \Lambda}^{R_\pm^0}(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (1-t)^2 & \pm t(1-t) \\ 0 & \pm t(1-t) & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$ for $t \in [0, 1]$,

(b) there is no non-minimizing geodesic,

(c) $R_\pm^1 = \text{Diag}(-1, \pm 1)$ lead to curves that hit $\text{Sym}^+(3, 1)$ at $t = \frac{1}{2} < 1$, namely $\gamma_{\Sigma \rightarrow \Lambda}^{R_\pm^1}(t) = \begin{pmatrix} (1-2t)^2 & 0 & 0 \\ 0 & (1-t)^2 & \pm t(1-t) \\ 0 & \pm t(1-t) & t^2 \end{pmatrix} \in \text{Sym}^+(3, 2)$ only for $t \in (-\infty, \frac{1}{2})$.

4. Let $\Sigma = \begin{pmatrix} I_2 & 0 \\ 0 & 0 \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 0 & 0 \\ 0 & I_2 \end{pmatrix}$ in $\text{Sym}^+(4, 2)$ with $r = 0$. Then, let $X = \begin{pmatrix} I_2 \\ 0 \end{pmatrix}$ and $Y = \begin{pmatrix} 0 \\ I_2 \end{pmatrix}$. Then $X^\top Y = 0$ so every $R \in O(2)$ is a candidate. One can show that any $R \in O(2)$ leads to a minimizing geodesic $\gamma_{\Sigma \rightarrow \Lambda}^R(t) = \begin{pmatrix} (1-t)^2 I_2 & t(1-t)R \\ t(1-t)R^\top & t^2 I_2 \end{pmatrix} \in \text{Sym}^+(4, 2)$ for $t \in \mathbb{R}$.

In the two last sections, we studied the geodesics and the minimizing geodesics within each stratum. In the next section, we turn to the study of the minimizing geodesic segments in the Bures-Wasserstein metric space $(\text{Cov}(n), d^{\text{BW}})$, that is between any two covariance matrices of any rank.

9.6 Minimizing geodesics of the Bures-Wasserstein distance on $\text{Cov}(n)$

In this section, we completely characterize the Bures-Wasserstein minimizing geodesic segments between any two covariance matrices. We show that they have constant rank on the interior of the segment and we give an explicit expression. Moreover, we show that the number of geodesics depends on the ranks of the extremities and we give this number in all cases. More precisely, we show that minimizing geodesics between Σ and $\Lambda \in \text{Cov}(n)$ are parametrized by the closed unit ball of $\mathbb{R}^{(k-r) \times (l-r)}$ for the spectral norm, where k, l, r are the respective ranks of $\Sigma, \Lambda, \Sigma\Lambda$. We also give the number of geodesics of minimal rank. Finally, we show that there exists a canonical geodesic with an expression that does not depend on the ranks of the extremities. This expression coincides with the formula in low rank when the minimizing geodesic is unique and with the formula in full rank.

9.6.1 Characterization of minimizing geodesics

The following lemma states that the rank of a minimizing geodesic segment is constant on the interior of the segment. Then, Theorem 9.31 characterizes the Bures-Wasserstein minimizing geodesic segments.

Lemma 9.30 (Rank of minimizing curve) Let $\gamma : [0, 1] \rightarrow \text{Cov}(n)$ be a minimizing curve from Σ to Λ . Then γ has constant rank $p \geq \max(\text{rk}(\Sigma), \text{rk}(\Lambda))$ on $(0, 1)$.

Proof. Let $p = \max_{t \in [0, 1]} \text{rk}(\gamma(t))$ and let $t_0 \in [0, 1]$ such that $\text{rk}(\gamma(t_0)) = p$. By Lemma 9.18, for all $t \in (0, t_0) \cup (t_0, 1)$, $(H_{\text{rk}(\gamma(t))}) \leq H_{\text{rk}(\gamma(t_0))}$ so $\text{rk}(\gamma(t)) \geq \text{rk}(\gamma(t_0)) = p$ so $\text{rk}(\gamma(t)) = p$. \square

Theorem 9.31 (Bures-Wasserstein minimizing geodesic segments in $\text{Cov}(n)$) Let $\Sigma, \Lambda \in \text{Cov}(n)$ with $\text{rk}(\Sigma) = k$ and $\text{rk}(\Lambda) = l$. Let $X, Y \in \mathbb{R}^{n \times n}$ such that $XX^\top = \Sigma$ and $YY^\top = \Lambda$. The two following statements are equivalent:

- (i) the curve $\gamma : [0, 1] \rightarrow \text{Cov}(n)$ is a minimizing geodesic segment from Σ to Λ ,
- (ii) there exists $R \in O(n)$ such that $H_{X,Y,R} := X^\top Y R^\top \in \text{Cov}(n)$ and for all $t \in [0, 1]$, $\gamma(t) = \gamma_{\Sigma \rightarrow \Lambda}^R(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(XRY^\top)$.

Moreover, $H_{X,Y,R} = (X^\top \Lambda X)^{1/2}$ and the minimizing geodesic $\gamma_{\Sigma \rightarrow \Lambda}^R$ is of constant rank $p \geq \max(k, l)$ on $(0, 1)$.

See the proof of Theorem 9.31 in Section 11.8.

Remark 9.32 In the previous theorem, X and Y may be respectively taken as $X = \Sigma^{1/2}$ and $Y = \Lambda^{1/2}$.

9.6.2 Number of minimizing geodesics

In this section, we count the number of minimizing geodesic segments between two covariance matrices. We start with an elementary lemma.

Lemma 9.33 (Elementary algebra) Let $\Sigma, \Lambda \in \text{Cov}(n)$ with $\text{rk}(\Sigma) = k$ and $\text{rk}(\Lambda) = l$. Let $r = \text{rk}(\Sigma\Lambda)$.

1. For all $X, Y \in \text{Mat}(n)$ such that $XX^\top = \Sigma$ and $YY^\top = \Lambda$, $r = \text{rk}(X^\top Y)$.
2. We have $l - r \leq n - k$.

See the proof of Lemma 9.33 in Section 11.8.

In the following theorem and especially in its proof, we need to distinguish cases where matrices may have one or two null dimensions, i.e. belonging to $\mathbb{R}^{n \times 0}$, $\mathbb{R}^{0 \times k}$ or $\mathbb{R}^{0 \times 0}$. Thus we recall that these spaces are equal to the vector space $\{0\}$. Indeed, there is a unique linear map from \mathbb{R}^n to \mathbb{R}^0 or from \mathbb{R}^0 to \mathbb{R}^k , which is the identically null map. The canonical basis of \mathbb{R}^0 is empty and the corresponding matrix in the canonical bases is called the empty matrix. It is practical to treat these spaces as non-trivial spaces to avoid writing particular cases. In particular, $\text{St}(n, 0) = \mathbb{R}^{n \times 0}$ and $\text{O}(0) = \text{GL}(0) = \text{Mat}(0) = \text{Diag}(0)$ are sets of cardinal 1.

Theorem 9.34 (Number of Bures-Wasserstein minimizing geodesic segments in $\text{Cov}(n)$) Let $\Sigma, \Lambda \in \text{Cov}(n)$ with $\text{rk}(\Sigma) = k$ and $\text{rk}(\Lambda) = l$. We assume that $k \geq l$ without loss of generality. We denote $r = \text{rk}(\Sigma\Lambda)$. We have $l - r \leq n - k$.

1. There exists a bijection between the set of minimizing geodesics from Σ to Λ and the closed unit ball of $\mathbb{R}^{(k-r) \times (l-r)}$ for the spectral norm $\bar{\mathcal{B}}_S(0, 1) = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid \|R_0\|_S \leq 1\} = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid 0 \leq R_0^\top R_0 \leq I_{l-r}\}$. More precisely, this bijection is given by:

$$\gamma_{\Sigma \rightarrow \Lambda}^{R_0}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(X_r Y_r^\top + X_{k-r} R_0 Y_{l-r}^\top), \quad (9.7)$$

where $X = (X_r \ X_{k-r} \ 0) \in \mathbb{R}^{n \times n}$ and $Y = (Y_r \ Y_{l-r} \ 0) \in \mathbb{R}^{n \times n}$ are such that $XX^\top = \Sigma$, $YY^\top = \Lambda$ and $X^\top Y = \text{Diag}(D_r, 0)$, and $R = \begin{pmatrix} I_r & 0 & 0 \\ 0 & R_0 & * \\ 0 & * & * \end{pmatrix} \in \text{O}(n)$, with $X_r, Y_r \in \mathbb{R}^{n \times r}$, $X_{k-r} \in \mathbb{R}^{n \times (k-r)}$, $Y_{l-r} \in \mathbb{R}^{n \times (l-r)}$, $D_r \in \text{Diag}^+(r)$, $R_0 \in \bar{\mathcal{B}}_S(0, 1)$.

2. The minimizing geodesic is unique if and only if $r = l$. This includes the cases $k = n$.
3. There are infinitely many minimizing geodesics if and only if $r < l$.
4. The minimizing geodesics corresponding to the choices $R_0 \in \text{St}(k-r, l-r)$ (including the empty matrix if $r = l$) have rank exactly k on $[0, 1)$ (on $[0, 1]$ if $l = k$). Note that $\text{St}(k-r, l-r)$ is included in the unit sphere $\mathcal{S}_S(0, 1) = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid \|R_0\|_S = 1\}$.
5. The minimizing geodesic corresponding to the choice $R_0 = 0$ (or the empty matrix if $r = l$) writes for all $t \in [0, 1]$:

$$\gamma_{\Sigma \rightarrow \Lambda}^0(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda). \quad (9.8)$$

If $r = l$, it has rank exactly k on $[0, 1)$.

The number of minimizing geodesic segments in $\text{Sym}^+(n, k)$ and in $\text{Cov}(n)$ is summarized in Table 9.2 with $n \geq k \geq l \geq r$.

See the proof of Theorem 9.34 in Section 11.8.

$\Sigma \in$	$\Lambda \in$	$r = \text{rk}(\Sigma\Lambda)$	Number of minimizing geodesics	
			in $\text{Sym}^+(n, k)$	in $\text{Cov}(n)$
$\text{Sym}^+(n)$	$\text{Sym}^+(n)$	n	1	1
$\text{Sym}^+(n)$	$\text{Sym}^+(n, k)$	k	1	1
$\text{Sym}^+(n, k)$	$\text{Sym}^+(n, k)$	k	1	1
		$k - 1$	2	∞
		$< k - 1$	∞	∞
$\text{Sym}^+(n, k)$	$\text{Sym}^+(n, l)$	l	1	1
		$< l$	∞	∞

Table 9.2: Number of Bures-Wasserstein minimizing geodesic segments ($n \geq k \geq l \geq r$).

Remark 9.35 The canonical minimizing geodesic (Equation 9.8) is defined between any two covariance matrices. This formula can always be used to interpolate between two covariance matrices. In particular, we easily check that it reduces to the formula of the unique minimizing geodesic between two SPD matrices (Equation 9.3). Moreover, Equation 9.8 is logically the same as the formula of the minimizing geodesic in $\text{Sym}^+(n, k)$ when it is unique (Equation 9.6).

9.7 Conclusion

In this work, we have answered several open questions on geodesics of the Bures-Wasserstein distance on covariance matrices. Beyond geodesics, a very important element of Riemannian geometry is the curvature. We know that the space of covariance matrices with the Bures-Wasserstein distance is an Aleksandrov space of non-negative curvature [Takatsu, 2011] and we know the curvature tensor in each stratum [Takatsu, 2010, Takatsu, 2011, Massart et al., 2019]. However, we lack a comprehensive and global approach of the curvature of the whole metric space. In particular, what is the appropriate notion of curvature to use to go from a stratum to another?

In the community of geometric statistics, most of the stratified spaces that were studied from the viewpoint of geodesics or curvature are very singular (spiders, trees) or a bit complex to start with (BHV space, Wald space) [Feragen and Nye, 2020]. Thus, the familiar example of the Bures-Wasserstein Riemannian orbit space appears to be a good basis to generalize concepts defined in Riemannian statistics. Indeed, after studying the geometry of these non-Riemannian spaces, what statistical tools should we define on them to generalize the Euclidean and Riemannian ones? This is probably the main question to investigate for the future.

As a perspective for future works, we quickly introduce an ongoing work that is closely related to the Bures-Wasserstein geometry of covariance matrices. The set of correlation matrices, known as the elliptope, was mainly studied on the open stratum of full-rank correlation matrices $\text{Cor}^+(n) = \{C \in \text{Sym}^+(n) | \text{Diag}(C) = I_n\}$ [David and Gu, 2019, David, 2019, Nielsen and Sun, 2019, Thanwerdas and Penneç, 2021, Thanwerdas and Penneç, 2022c, Archakov and Hansen, 2021, Thanwerdas, 2022]. The geometry of the whole elliptope $\text{Cor}(n) = \{C \in \text{Cov}(n) | \text{Diag}(C) = I_n\}$ was described topologically in [Kercheval, 2008]

as the quotient of square matrices with unit row length $B(n) = \{M = (X_1 \cdots X_n)^\top \in \text{Mat}(n) \mid \|X_1\| = \cdots = \|X_n\| = 1\} \simeq (\mathbb{S}^{n-1})^n$ by the orthogonal group $O(n)$. Therefore, in light of the work on the Bures-Wasserstein orbit space of covariance matrices, it is natural to go beyond this topological description and define a Riemannian orbit space structure on the elliptope of correlation matrices of any rank $\text{Cor}(n)$. We naturally call it the quotient-polysphere structure.

Indeed, the sets of correlation matrices of fixed rank $\text{Cor}^+(n, k)$ for $k > 0$ are clearly the isotropy strata of this orbit space, the principal stratum being the open elliptope $\text{Cor}^+(n)$. (More precisely, the stratum $\text{Cor}^+(n, 1)$ is discrete and contains the 2^{n-1} vertices of the elliptope so each vertex is itself an orbit stratum [Tropp, 2018].) The natural geometry on $B(n)$ is the product of spheres so the elliptope $\text{Cor}(n)$ inherits the quotient-polysphere distance. It reduces to the quotient-polysphere Riemannian metric on each stratum $\text{Cor}^+(n, k) = (\mathbb{S}^{k-1})^n / O(k)$ so the study can be conducted first on each stratum then globally, as for the Bures-Wasserstein geometry. The computations remain to be done.

Part VI
Conclusion

Chapter 10

Conclusion and perspectives

10.1 Summary of contributions and beyond

In this thesis, we studied geometries of covariance and correlation matrices from different angles. On SPD matrices, we formalized a dictionary of Riemannian metrics enriched with new formulae and with a novel approach of classes based on the characterization of $O(n)$ -invariant metrics. In particular, we interpolated the noted Euclidean, log-Euclidean, power-Euclidean, affine-invariant, power-affine and Bogoliubov-Kubo-Mori metrics by the two-parameter family of Mixed-Power-Euclidean metrics. We enlightened the bridge between the wider family of Mixed-Euclidean metrics and the (u, v) -divergences of information geometry and we computed their curvature by generalizing the computation known for the BKM metric, that is with $u = \text{Id}$ and $v = \log$. Besides $O(n)$ -invariant metrics, we also characterized $\text{LT}^+(n)$ -invariant Riemannian metrics as pullbacks of left-invariant metrics on the Lie group $\text{LT}^+(n)$ via the Cholesky map. These Lie group metrics form a family of Riemannian metrics parametrized by inner products on $\text{LT}(n)$, that is by $\text{Sym}^+(\frac{n(n+1)}{2})$. They belong to the wider family of $\text{Diag}^+(n)$ -invariant metrics which descend to the open ellipsope of full-rank correlation matrices. This exploration of the space of Riemannian metrics on SPD matrices with parametric and non-parametric subfamilies offers a new promising setting for choosing a metric satisfying given requirements or for optimizing the metric.

We also followed this systematic approach on full-rank correlation matrices. Our work on Lie group actions on SPD matrices allowed us to generalize the quotient-affine metric with the family of quotient-Lie-Cholesky metrics. However, we observed that these are probably not the best suited metrics on the open ellipsope. Thus we grasped the relevant concepts behind them such as the permutation action, the positive diagonal action on SPD matrices, the reduced Cholesky map $\Theta : \text{Cor}^+(n) \rightarrow \text{LT}^1(n)$, the cor-inverse involution or the block-equicorrelation matrices. In view of proposing better metrics, the characterization of permutation-invariant inner products on the tangent space $\text{Hol}(n) = T_{I_n} \text{Cor}^+(n)$ turned out to be crucial for understanding the natural vector spaces to work with. This made us realize the module isomorphism between $\text{Hol}(n)$ and $\text{Row}_0(n)$, which supported the idea of an analogous correspondence at the manifold level between $\text{Cor}^+(n)$ and $\text{Row}_{\mathbb{1}}^+(n)$. Thanks to these suited tools, we built five families of Hadamard and flat metrics on full-rank correlation matrices. In particular, the log-scaled metrics are flat, geodesically complete, permutation-invariant and inverse-consistent, which are quite satisfying properties. All the geometric

operations can be computed via the diffeomorphism onto the Euclidean space at hand, in finite time for the ones based on the Cholesky map, in logarithmic time complexity for the others. For the log-scaled metrics, we proposed a new algorithm to compute the scaling of an SPD matrix which seems competitive with respect to existing ones.

These Riemannian metrics also lead to brand new approaches on SPD matrices since they allow to decouple the scales of variables and the correlations between variables. These two components of the data are completely mixed so far because of the use of $O(n)$ -invariant metrics such as the affine-invariant or the log-Euclidean metrics. These new product metrics define new classes of means and midpoints with different monotonicity properties that could be compared with existing ones [Bhatia and Holbrook, 2006, Mostajeran and Sepulchre, 2018, Mostajeran et al., 2020]. In practice, they offer the possibility for researchers to make other types of assumptions than before on their data and experiments. For example, if the scales of variables do not matter in first approximation, the data can be normalized to correlation matrices. Then one could add a metric λg on the scales, that is on positive diagonal matrices, with a parameter $\lambda \geq 0$ to tune in function of the relative importance of scales to correlations.

Let us give the example of single-cell analysis in genetics. Cells are sampled from a tissue, the goal is to identify types of different cells. For each cell i , the number of mRNA segments belonging to the gene j is counted after a polymerase chain reaction (PCR). Thus the raw data are count matrices of size $n_{\text{cells}} \times n_{\text{genes}}$. An independent normalization of each row is usually applied to neglect effects related to the size of the cell: a big cell tends to have more mRNA segments than a smaller cell. Thus the correlation matrix of the count matrix could be a relevant representation of the data. However, the size of the cell may also contain information on the type of cell. Therefore, one could assume that choosing a product metric with a low coefficient on the diagonal part could better fit the application.

It is even possible to tune the metric g to give more or less importance to the scales of some variables. For example, in brain-computer interfaces, we may assume that the scales of the signals at the electrodes located in the brain region of the response to the stimulus contain more information than elsewhere. Therefore, one would rather choose a metric on diagonal matrices that is *not* invariant by permutations to take into account this asymmetry. By analogy with [Guigui et al., 2021, Guigui, 2021], the product metrics could also initiate new methods in the registration of covariance matrices with parallel transport by allowing a step of normalization of the scales, especially when the scales significantly differ within a dataset of covariance matrices.

On covariance matrices of any rank, we promoted the viewpoint of Riemannian orbit spaces to compute all the minimizing geodesics of the Bures-Wasserstein geometry. It was well known that covariance matrices form a quotient space, the strata had been described in [Takatsu, 2011], the geometry had been studied on each stratum [Malagò et al., 2018, Bhatia et al., 2019, Massart and Absil, 2020], and we finally brought a bigger picture with the characterization of minimizing geodesics. This work departs from the current research on stratified spaces for applications. Indeed, it is an example of a non-negatively curved metric stratified space that is not a tree space or a graph space and for which the minimizing geodesics are known in closed form. Besides these three major differences, the simple formulae of geodesics supports the idea of using Riemannian orbit spaces when it is possible to get more structure in the space. This class contains Kendall shape spaces, the space of correlation matrices

stratified by the rank, the space of diagonal matrices stratified by eigenvalue multiplicity and probably other examples. Moreover, Geodesic PCA [Huckemann et al., 2010] is defined in this kind of spaces via generalized geodesics, which are simply projections of horizontal geodesics. We note that the Bures-Wasserstein metric is more and more used in applications [Severn et al., 2019, Chewi et al., 2020, Kroshnin et al., 2021, Han et al., 2021] since its geometric operations were derived. In this aspect, our formulae open new uses of the Bures-Wasserstein distance in applications where covariance matrices may be either regular or singular.

Riemannian orbit spaces could be a nice first class of stratified spaces to implement in `geomstats` with generic geometric and statistical operations. Complete Riemannian orbit spaces should inherit all the methods known in complete metric spaces and should be the right class to generalize many statistical methods used in complete Riemannian (quotient) manifolds. Although mainly ad hoc methods are currently used in stratified spaces such as the BHV space or the QED space, it might be possible to adapt some of them to the richer structure of Riemannian orbit spaces. Therefore, the Bures-Wasserstein geometry of covariance matrices offers a new playground to test generic tools on orbit spaces and stratified spaces.

During my PhD thesis, I also took part in the development of the `geomstats` package with about 100 commits. I also had a less visible role in the reflection of the class architecture, trying to mathematically categorize the types of spaces of interest with my colleague Nicolas Guigui. Thanks to my bibliographic work on Riemannian metrics, especially on the affine-invariant, the Bures-Wasserstein and the quotient-affine metrics, we had long discussions on all sorts of quotient spaces and homogeneous spaces before he implemented the generic class of (principal) fiber bundle. I personally left the code aside momentarily to allocate time for the research of suitable structures on covariance and correlation matrices and to push the computations as far as they could be. I can now implement all these tools in `geomstats` and make them available for everyone.

10.2 Future works

10.2.1 Other geometries for the rank stratification

The first burning perspective to this work is probably the study of the Riemannian orbit space of correlation matrices endowed with the quotient-polysphere structure that we briefly described at the end of Chapter 9. Contrarily to the Bures-Wasserstein geometry, it has not been described on each stratum so it is a very challenging problem.

It would also be nice to introduce *families* of stratified geometries on covariance and correlation matrices to allow some flexibility. In this direction, generalized Bures-Wasserstein metrics were recently introduced on SPD matrices [Han et al., 2021]. They are defined as the quotient metrics of the $O(n)$ -invariant metrics $g_A^{\text{GBW}(S)}(X, X) = \text{tr}(X^\top S X)$ for $A \in \text{GL}(n)$ and $X \in \text{Mat}(n)$, where $S \in \text{Sym}^+(n)$ is a parameter. The Bures-Wasserstein metric corresponds to $S = I_n$. This metric turn out to be simply the pullback metric of the Bures-Wasserstein metric by the diffeomorphism $\Sigma \in \text{Sym}^+(n) \mapsto S^{1/2} \Sigma S^{1/2} \in \text{Sym}^+(n)$. In other words, it is a change of reference point. This construction could be extended to the whole

space of covariance matrices to define the family of generalized Bures-Wasserstein stratified geometries indexed by $S \in \text{Sym}^+(n)$. With $S \in \text{Diag}^+(n) \setminus \mathbb{R}^+ I_n$, this introduces weights on the features, thus the geometry is not anymore invariant under permutations. It could be interesting if one wants to give more or less importance to some features for example.

Analogously, a generalized quotient-polysphere distance could be defined by introducing weights in the product metric of spheres $(\mathbb{S}^{n-1})^n$. Another question that we left aside is the research of geometries for the manifolds $\text{Cor}(n, k)$ of fixed-rank correlation matrices, independently from other strata. It could be useful in applications where the rank is almost surely equal to k . Note that contrarily to the full-rank case, the manifold $\text{Cor}(n, k)$ is not the quotient of $\text{Sym}^+(n, k)$ by positive diagonal matrices. Indeed, singular covariance matrices include covariance matrices with zero variance, which are excluded from $\text{Cor}(n, k)$. We suspect that several quotient geometries on $\text{Sym}^+(n, k) = \mathcal{M}/G$ may pass to $\text{Cor}(n, k)$ not by simple quotient but by an operation of type $(\text{Diag}^+(n) \setminus \mathcal{M})/G$, that is removing the scales *before* quotienting by G . That is exactly the case for the quotient-polysphere structure: the Bures-Wasserstein geometry on covariance matrices is not invariant under positive diagonal matrices so instead of defining something like $\text{Cor}(n) = \text{Diag}^+(n) \setminus \text{Cov}(n) = \text{Diag}^+(n) \setminus (\text{Mat}(n)/\text{O}(n))$, the quotient-polysphere is defined by $\text{Cor}(n) = (\mathbb{S}^{n-1})^n / \text{O}(n) = (\text{Diag}^+(n) \setminus \text{Mat}(n)) / \text{O}(n)$.

Covariance and correlation matrices are also embedding spaces for spaces of trees or graphs such as the Laplacian matrices [Ginestet et al., 2017] or the Wald space [Garba et al., 2021], which are themselves stratified spaces. It could be interesting to study the geometry induced by the Bures-Wasserstein geometry of these spaces. For example, a condition satisfied by Laplacian matrices is that the row sums are null. We notice that this condition is preserved all along a minimizing geodesic if it is satisfied by the end points. We could extend the notion of totally geodesic submanifold to metric spaces and study if these spaces are totally geodesic subspaces of covariance matrices or correlation matrices.

10.2.2 Geometries of matrices of any size and any rank

Distances between Grassmannians and between SPD matrices of different dimensions were proposed in [Ye and Lim, 2016, Lim et al., 2019]. For two SPD matrices $\Sigma \in \text{Sym}^+(n)$ and $\Lambda \in \text{Sym}^+(k)$ with $k \leq n$ represented as ellipsoids \mathcal{E}_Σ and \mathcal{E}_Λ , the general methodology consists in finding the ellipsoid $\mathcal{E}_{\Sigma_0} \subseteq \mathcal{E}_\Sigma$ of dimension k which is the closest to \mathcal{E}_Λ for a given distance d in $\text{Sym}^+(k)$. The distance between Σ and Λ is defined as the distance $d(\Lambda, \Sigma_0)$ in $\text{Sym}^+(k)$. For Grassmannians, it consists in aligning vector spaces in a similar way. This method seems quite general, it could probably be applied to PSD matrices represented as ellipsoids for SPD matrices and ellipsoidal cylinders for singular matrices, as well as for correlation matrices. This would provide distances between covariance matrices of different dimensions and different ranks. If the dimension is fixed, it could be interesting to study the geometry of such defined distances on the spaces of covariance matrices and correlation matrices.

This raises the problem of the geometry of spaces of matrices of different dimensions. Beyond the distance, is it possible to glue the manifolds together in a stratified way? Could they be seen as orbit spaces? These challenging theoretical questions may meet practical needs in graph theory for example, were it is difficult to compare graphs with a different

number of nodes. These graphs are often represented by adjacency or Laplacian matrices so adding an isolated node corresponds to add zeros in the matrix, which artificially introduces singularities. This also obliges to fix the number of nodes beforehand, which is a strong constraint in applications where nodes appear and disappear.

10.2.3 Eigenvalue multiplicity and block-wise transformations

The stratification of SPD matrices by eigenvalue multiplicity was investigated in [Groisser et al., 2017]. To understand eigenvalue multiplicity, it may be easier to consider the simpler stratified space of positive diagonal matrices. It is the orbit space of SPD matrices by the congruence action of the orthogonal group. This framework can probably be extended to covariance matrices of any rank by allowing null eigenvalues. This would create a framework where both multiple eigenvalues and singular matrices are allowed.

Another direction to investigate could be the invariance under the action of block-wise orthogonal transformations or block-wise permutations. This could help identify groups of variables that must be equally treated within groups but potentially not across different groups. This is related to eigenvalue multiplicity since block-orthogonal matrices are precisely the stabilizers of diagonal matrices with repeated (non-decreasingly ordered) eigenvalues. This is also related to the study of flag manifolds, which are quotients of the orthogonal group by a block-orthogonal group [Monk, 1959, Alekseevsky, 1997] and which generalize the Grassmannians. Moreover, thanks to Chapter 3, we have all the necessary tools to determine all block-permutation-invariant inner products on symmetric matrices so this could help define suitable Riemannian metrics to handle this kind of invariance.

10.2.4 Applications

One of the most exciting challenges is to apply these new tools on covariance and correlation matrices in the domains we identified: brain-computer interfaces, brain connectivity, signal processing, genomics, shapes of proteins, finance, etc. Although the affine-invariant and the log-Euclidean metrics were shown to outperform the results with respect to the Euclidean metric in many contexts, there are still many unsolved issues with this kind of data. While discussing with researchers from these domains, the three problems we heard the most were the fact that the rank might be deficient, the fact that the rank might be non-constant within a dataset and the lack of tools for correlation matrices. Therefore, the tools we propose will at least allow to do new operations on the data. Note that many of the new geometries defined in this thesis could probably be extended to complex matrices, which opens the way to other applications in complex signal processing for instance.

Furthermore, applying our metrics to these problems will allow us to test our assumptions and to compare the results obtained with different metrics. There are many simple tests to do on synthetic data of full-rank correlation matrices such as computing means, principal components and distances to do clustering. For example in BCI, it is frequent to classify mental tasks by comparing distances to reference points. Then, the interactions with researchers of these domains will help us get other assumptions on the data and understand at which step of the workflow the geometry may intervene. These are appealing perspectives to

close the loop initiated in the introduction and reproduced on Figure 10.1.

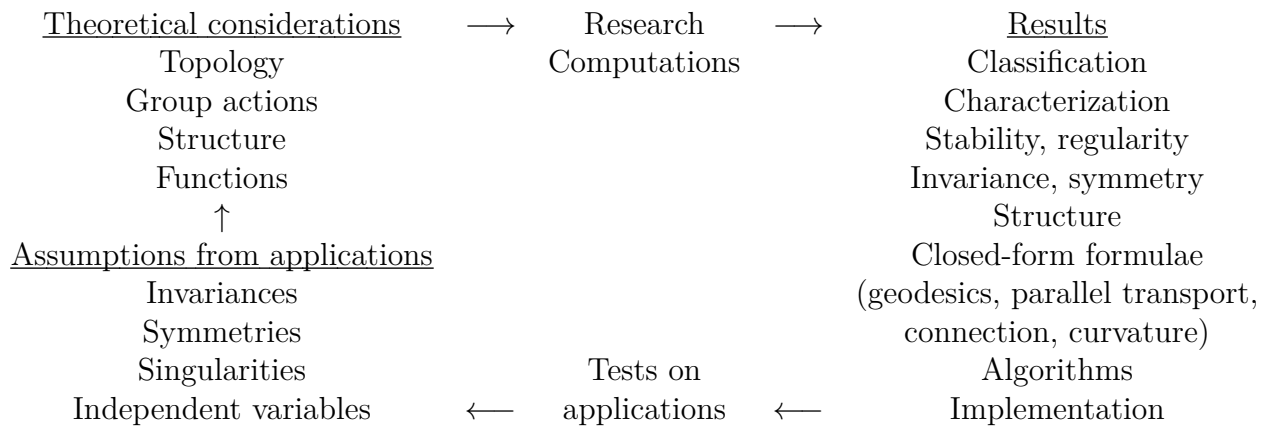


Figure 10.1: Future works: tests on applications.

To conclude, the concepts and geometric tools introduced in this thesis could be the basis of many future developments to tackle challenging theoretical and practical problems.

10.3 Publications

Journal papers

- Thanwerdas, Yann and Penneç, Xavier (2022). The geometry of mixed-Euclidean metrics on symmetric positive definite matrices. *Differential Geometry and its Applications*, 81.
- Miolane, N., Guigui, N., Le Brigant, A., Mathe, J., Hou, B., Thanwerdas, Y., Heyder, S., Peltre, O., Koep, N., Zaatiti, H., Hajri, H., Cabanes, Y., Gerald, T., Chauchat, P., Shewmake, C., Brooks, D., Kainz, B., Donnat, C., Holmes, S., and Penneç, X. (2020). Geomstats: A Python Package for Riemannian Geometry in Machine Learning. *Journal of Machine Learning Research*, 21(223):1–9.

Journal submissions

- Thanwerdas, Yann and Penneç, Xavier (2022). Theoretically and computationally convenient geometries on full-rank correlation matrices. Submitted to the SIAM Journal on Matrix Analysis and Applications.
- Thanwerdas, Yann and Penneç, Xavier (2022). $O(n)$ -invariant Riemannian metrics on SPD matrices. Submitted to Linear Algebra and its Applications (2nd round after minor revisions).

Preprints

- Thanwerdas, Yann and Pennec, Xavier (2022). Bures-Wasserstein minimizing geodesics between covariance matrices of different ranks. Preprint. Submission planned in April 2022 to SIMAX.
- Thanwerdas, Yann (2022). Permutation-invariant log-Euclidean geometries on full-rank correlation matrices. Preprint. Submission planned in May 2022 to SIMAX.

Conference articles

- Thanwerdas, Yann and Pennec, Xavier (2021). Geodesics and Curvature of the Quotient-Affine Metrics on Full-Rank Correlation Matrices. In *Proceedings of GSI 2021 - 5th conference on Geometric Science of Information*, volume 12829 of *Lecture Notes in Computer Science*, pages 93-102, Toulouse, France. Springer International Publishing.
- Miolane, N., Guigui, N., Zaatiti, H., Shewmake, C., Hajri, H., Brooks, D., Le Brigant, A., Mathe, J., Hou, B., Thanwerdas, Y., Heyder, S., Peltre, O., Koep, N., Cabanes, Y., Gerald, T., Chauchat, P., Kainz, B., Donnat, C., Holmes, S., and Pennec, X. (2020). Introduction to Geometric Learning in Python with Geomstats. In *SciPy 2020 - 19th Python in Science Conference*, pages 48–57, Austin, Texas, United States.
- Thanwerdas, Yann and Pennec, Xavier (2019). Exploration of Balanced Metrics on Symmetric Positive Definite Matrices. In *Proceedings of GSI 2019 - 4th conference on Geometric Science of Information*, volume 11712 of *Lecture Notes in Computer Science*, pages 484–493, Toulouse, France. Springer International Publishing.
Among the five running papers for the best paper award.
- Thanwerdas, Yann and Pennec, Xavier (2019). Is affine-invariance well defined on SPD matrices? A principled continuum of metrics. In *Proceedings of GSI 2019 - 4th conference on Geometric Science of Information*, volume 11712 of *Lecture Notes in Computer Science*, pages 502–510, Toulouse, France. Springer International Publishing.

Part VII
Appendix

Chapter 11

Proofs

11.1 Proofs of Chapter 2

11.1.1 Proof of Theorem 2.7

Theorem 2.7 (No continuous map of eigenvectors on $\text{Sym}^{\neq}(n)$) Let $n \geq 2$. There exists no continuous map $vec : \text{Sym}^{\neq}(n) \rightarrow O(n)$ such that for all $\Sigma \in \text{Sym}^{\neq}(n)$, $vec(\Sigma)val(\Sigma)vec(\Sigma)^{\top} = \Sigma$.

Proof of Theorem 2.7 (No continuous map of eigenvectors on $\text{Sym}^{\neq}(n)$). It is well known that there is no continuous map of eigenvectors on the whole vector space of symmetric matrices $\text{Sym}(n)$. Indeed, if $\Sigma(t) = t \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = tI_2 \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} I_2$ and $\Lambda(t) = t \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{t}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, then $vec(\Sigma(t)) = \varepsilon$ where $\varepsilon = \text{Diag}(\pm 1, \pm 1)$ and $vec(\Lambda(t)) = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \varepsilon'$ where $\varepsilon' = \text{Diag}(\pm 1, \pm 1)$. If the map vec were continuous, then $vec(0) = \lim_{t \rightarrow 0} vec(\Sigma(t)) = \varepsilon$ and $vec(0) = \lim_{t \rightarrow 0} vec(\Lambda(t)) = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \varepsilon'$, which is a contradiction. In other words, one can reach a symmetric matrix with repeated eigenvalues (here 0) from different bases in the eigenspace corresponding to the repeated eigenvalue.

Removing the matrices with repeated eigenvalues is not sufficient because one can still turn around the axis $\mathbb{R}I_2$. That is what we prove now. Assume there exists a continuous map of eigenvectors $vec : \text{Sym}^{\neq}(n) \rightarrow O(n)$. Let $D = \text{Diag}(-1, 1) \in \text{Diag}(2)$, $P(t) = \begin{pmatrix} \cos(\pi t) & \sin(\pi t) \\ -\sin(\pi t) & \cos(\pi t) \end{pmatrix} \in O(2)$ and $\Sigma(t) = P(t)DP(t)^{\top} \in \text{Sym}^{\neq}(n)$. Then $P(0) = I_2$, $P(1) = -I_2$ and $\Sigma(0) = \Sigma(1) = D$. We define the continuous maps $Q : t \in \mathbb{R} \mapsto vec(\Sigma(t)) \in O(2)$ and $R : t \in \mathbb{R} \mapsto P(t)Q(0) \in O(2)$. Note that $Q(0) = Q(1) = R(0) = -R(1) \in \text{Diag}(\pm 1, \pm 1)$ and $\Sigma(t) = Q(t)DQ(t)^{\top} = R(t)DR(t)^{\top}$ for all $t \in [0, 1]$. Thanks to Lemma 2.6, we will build a sequence $0 = t_0 < \dots < t_N \leq t_{N+1} = 1$ and show by recurrence that $Q(t_k) = R(t_k)$ for all $k \in \{0, \dots, N+1\}$. Thus we will show that $Q(1) = R(1)$ which will be a contradiction.

By the Heine theorem applied to Q on $[0, 1]$, let $\delta \in (0, \frac{1}{4})$ such that if $t, t' \in [0, 1]$ satisfy $|t - t'| < 2\delta$, then $\|Q(t) - Q(t')\| < \sqrt{2}$. We denote $t_k = k\delta$ for $k \in \{0, \dots, N\}$

where $N = \lfloor \frac{1}{\delta} \rfloor$ and $t_{N+1} = 1$. Since $N \leq \frac{1}{\delta} < N + 1$, we have $t_N \leq 1 < t_N + \delta$ so $0 \leq t_{N+1} - t_N < \delta$. Assume that $k \in \{0, \dots, N\}$ is such that Q and R coincide on t_0, \dots, t_k . Since $Q(t_{k+1})DQ(t_{k+1})^\top = R(t_{k+1})DR(t_{k+1})^\top$ and D has distinct eigenvalues, we have $Q(t_{k+1}) = R(t_{k+1})\varepsilon$ with $\varepsilon \in \text{Diag}(\pm 1, \pm 1)$. Let us show that $\varepsilon = I_2$. Since $|t_{k+1} - t_k| \leq \delta < 2\delta$, we have $\|Q(t_{k+1}) - Q(t_k)\| < \sqrt{2}$ so by Lemma 2.6, $\text{Diag}(Q(t_k)^\top Q(t_{k+1})) > 0$. We also have $\text{Diag}(R(t_k)^\top R(t_{k+1})) = Q(0)\text{Diag}(P(t_k)^\top P(t_{k+1}))Q(0) = \cos(\pi(t_{k+1} - t_k))I_2 > 0$ since $|t_{k+1} - t_k| < \frac{1}{2}$. Since $\text{Diag}(Q(t_k)^\top Q(t_{k+1})) = \text{Diag}(R(t_k)^\top R(t_{k+1}))\varepsilon$, it is only possible if $\varepsilon = I_2$ so $Q(t_{k+1}) = R(t_{k+1})$. By recurrence, since $R(0) = Q(0)$, we have $R(t_{N+1}) = Q(t_{N+1})$, i.e. $R(1) = Q(1)$. This is a contradiction. \square

11.2 Proofs of Chapter 3

In this section, we prove the general form of invariant inner products given in Chapter 3.

11.2.1 Proof of Example 3.5

Example 3.5 ($O(I)$ -invariant inner products on $\text{Skew}(n)$) The irreducible decomposition of $\text{Skew}(n)$ with respect to the congruence action of $O(I)$ is $\text{Skew}(n) = \bigoplus_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \text{Skew}_n(k_i) \oplus \bigoplus_{1 \leq i < j \leq p} \text{Skew}_n(k_i, k_j)$. The decomposition is orthogonal for the Frobenius inner product. The $O(I)$ -invariant inner products on $\text{Skew}(n)$ are given for all $X \in \text{Skew}(n)$ by:

$$\varphi(X, X) = \sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \alpha_{ii} \|X^{ii}\|^2 + \sum_{1 \leq i < j \leq p} \alpha_{ij} \|X^{ij}\|^2,$$

where $\alpha_{ij} > 0$ for all $1 \leq i \leq j \leq p$.

Proof of Example 3.5: $O(I)$ -invariant inner products on $\text{Skew}(n)$.

Direct sum The orthogonal direct sum is the sum of blocks and the orthogonal projections are the canonical projections. When $k_i = 1$, $\text{Skew}_n(k_i) = \{0\}$ so these spaces are removed from the direct sum.

Irreducibility By block matrix multiplication, the terms are clearly stable by $O(I)$. Moreover, we already showed that $\text{Skew}(n)$ is $O(n)$ -irreducible in the previous section for $n \geq 2$. Hence, for $k_i \geq 2$, $\text{Skew}_n(k_i)$ is clearly irreducible with respect to the group $O_n(k_i)$. Since it is a subgroup of $O(I)$ and since $\text{Skew}_n(k_i)$ is stable under $O(I)$, $\text{Skew}_n(k_i)$ is $O(I)$ -irreducible.

Moreover, we can show that $\text{Mat}(k, l)$ is $O(k) \times O(l)$ -irreducible for the action $((U, V), X) \in (O(k) \times O(l)) \times \text{Mat}(k, l) \mapsto UXV^\top$ and this will analogously prove that $\text{Skew}_n(k_i, k_j) \simeq \text{Mat}(k_i, k_j)$ is $O_n(k_i) \times O_n(k_j)$ -irreducible, hence $O(I)$ -irreducible. Indeed, if $k \leq l$ (without loss of generality), then by singular value decomposition, for all $X \in \text{Mat}(k, l)$, there exists $U, V \in O(k) \times O(l)$ such that $UXV^\top = (D \ 0)$ with $D \in \text{Diag}(k)$. If $X \neq 0$, then there exists $i \in \{1, \dots, k\}$ such that $d_i \neq 0$, for example $i = 1$. Hence, applying the action of $(I_k, \text{diag}(-1, 1, \dots, 1)) \in O(k) \times O(l)$, we get $D' = \text{diag}(-d_1, d_2, \dots, d_n)$. Then $E_{11} = \frac{1}{d_i}(D - D')$ and all the matrices $e_i e_j^\top$ of the canonical basis can be obtained by permuting the lines and

columns by $(P_{(1,i)}, P_{(1,j)})$. So from any $X \in \text{Mat}(k, l) \setminus \{0\}$, we retrieve $\text{Mat}(k, l)$ so $\text{Mat}(k, l)$ is $O(k) \times O(l)$ -irreducible.

Isomorphic modules The group clearly acts differently on each block so the irreducible modules of the decomposition are not isomorphic.

General form Therefore, we are again in the simple case where an invariant inner product is the positive weighted sum of the restrictions of the Frobenius inner product to each irreducible submodule in the decomposition. \square

11.2.2 Proof of Example 3.6

Example 3.6 ($O(I)$ -invariant inner products on $\text{Sym}(n)$) The irreducible decomposition of $\text{Sym}(n)$ with respect to the congruence action of $O(I)$ is $\text{Sym}(n) = \bigoplus_{i=1}^p \text{span}(I_n^{ii}) \oplus \bigoplus_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \text{Sym}_n^0(k_i) \oplus \bigoplus_{1 \leq i < j \leq p} \text{Sym}_n(k_i, k_j)$. The decomposition is orthogonal for the Frobenius inner product and \mathcal{V} . The $O(I)$ -invariant inner products on $\text{Sym}(n)$ are given for all $X \in \text{Sym}(n)$ by:

$$\varphi(X, X) = \sum_{1 \leq i, j \leq p} S_{ij} \frac{\text{tr}(X^{ii})\text{tr}(X^{jj})}{\sqrt{k_i k_j}} + \sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \alpha_{ii} \left\| X^{ii} - \frac{\text{tr}(X^{ii})}{k_i} I_n^{ii} \right\|^2 + \sum_{1 \leq i < j \leq p} \alpha_{ij} \|X^{ij}\|^2,$$

where $S \in \text{Sym}^+(p)$ and $\alpha_{ij} > 0$ for all $1 \leq i \leq j \leq p$.

Proof of Example 3.6: $O(I)$ -invariant inner products on $\text{Sym}(n)$.

Direct sum The orthogonal direct sum is the sum of blocks, refined on the diagonal blocks with the $O(k_i)$ -irreducible decomposition on $\text{Sym}_n(k_i) \simeq \text{Sym}(k_i)$, i.e. the trace part $\text{span}(I_n^{ii})$ and the traceless part $\text{Sym}_n^0(k_i)$. When $k_i = 1$, $\text{Sym}_n^0(k_i) = \{0\}$ so these spaces are removed from the direct sum.

Irreducibility The terms of the direct sum are stable. The terms $\text{span}(I_n^{ii})$ and $\text{Sym}_n^0(k_i)$ are $O(I)$ -irreducible because they are $O_n(k_i)$ -irreducible. The modules $\text{Sym}_n(k_i, k_j)$ and $\text{Skew}_n(k_i, k_j)$ are isomorphic as $O(k_i) \times O(k_j)$ -modules to $\text{Mat}(k_i, k_j)$: it suffices to keep the Upper Triangular (UT) part and take the opposite of the Lower Triangular (LT) part of $X \in \text{Sym}_n(k_i, k_j)$ to get $\text{UT}(X) - \text{LT}(X) \in \text{Skew}_n(k_i, k_j)$. We just proved that $\text{Skew}_n(k_i, k_j)$ is $O(I)$ -irreducible, hence so does $\text{Sym}_n(k_i, k_j)$.

Isomorphic modules The main difference with the previous examples is that the spaces $\text{span}(I_n^{ii})$ are all isomorphic as modules. Indeed, $\psi_i : \frac{\lambda}{\sqrt{k_i}} I_n^{ii} \in \text{span}(I_n^{ii}) \mapsto \lambda \in \mathbb{R}$ for $i \in \{1, \dots, p\}$ are clearly module isomorphisms with respect to $O(I)$, where $O(I)$ trivially acts on \mathbb{R} by $(R, \lambda) \in O(I) \times \mathbb{R} \mapsto \lambda \in \mathbb{R}$. They are isometries since $\text{tr} \left(\left(\frac{\lambda}{\sqrt{k_i}} I_n^{ii} \right)^2 \right) = \lambda^2$. The other terms of the irreducible decomposition are clearly not isomorphic as modules because the action of $O(I)$ differs on each block. If there were an off-diagonal block of dimension 1, it couldn't be isomorphic as a module to $\text{span}(I_n^{ii})$ since it can be changed into its opposite

by an appropriate element of $O(I)$.

General form Therefore, according to Theorem 3.4, the $O(I)$ -invariant inner products on $\text{Sym}(n)$ are given for all $X \in \text{Sym}(n)$ by:

$$\begin{aligned} \varphi(X, X) &= \sum_{\substack{1 \leq i \leq p \\ k_i \geq 2}} \alpha_{ii} \left\| X^{ii} - \frac{\text{tr}(X^{ii})}{k_i} I_n^{ii} \right\|^2 + \sum_{1 \leq i < j \leq p} \alpha_{ij} \|X^{ij}\|^2 \\ &+ \sum_{1 \leq i, j \leq p} S_{ij} \underbrace{\psi_i \left(\frac{\text{tr}(X^{ii})}{k_i} I_n^{ii} \right)}_{\frac{\text{tr}(X^{ii})}{\sqrt{k_i}}} \underbrace{\psi_j \left(\frac{\text{tr}(X^{jj})}{k_j} I_n^{jj} \right)}_{\frac{\text{tr}(X^{jj})}{\sqrt{k_j}}}, \end{aligned} \quad (11.1)$$

where $\alpha_{ij} > 0$ for $1 \leq i \leq j \leq p$ and $S = (S_{ij})_{1 \leq i \leq j \leq p} \in \text{Sym}^+(p)$, as expected. \square

11.2.3 Proof of Example 3.8

Example 3.8 ($\mathfrak{S}(n)$ -invariant inner products on $\text{Hol}(n)$) The irreducible decomposition of $\text{Hol}(n)$ with respect to the congruence action of $\mathfrak{S}(n)$, i.e. of permutation matrices, is $\text{Hol}(n) = \ker \phi_1 \oplus \text{im } \psi_2 \oplus \text{im } \psi_3$ where:

1. $\phi_1 : X \in \text{Hol}(n) \mapsto X\mathbf{1} \in \mathbb{R}^n$,
2. $\psi_2 : \mu \in \mathbb{R}_0^n \mapsto \frac{1}{n-2}(\mu\mathbf{1}^\top + \mathbf{1}\mu^\top - 2\text{diag}(\mu)) \in \text{Hol}(n)$,
3. $\psi_3 : x \in \mathbb{R} \mapsto \frac{x}{n(n-1)}(\mathbf{1}\mathbf{1}^\top - I_n) \in \text{Hol}(n)$,

with ϕ_1 surjective, ψ_2 and ψ_3 injective. The corresponding orthogonal projections are:

- $p_3 : X \in \text{Hol}(n) \mapsto X_3 = \psi_3(\text{sum}(X)) \in \text{im } \psi_3$,
- $p_2 : X \in \text{Hol}(n) \mapsto X_2 = \psi_2((X - X_3)\mathbf{1}) \in \text{im } \psi_2$,
- $p_1 : X \in \text{Hol}(n) \mapsto X_1 = X - X_2 - X_3 \in \ker \phi_1$.

The $\mathfrak{S}(n)$ -invariant inner products on $\text{Hol}(n)$ are given for all $X \in \text{Hol}(n)$ by:

$$\begin{aligned} \varphi(X, X) &= \alpha_1 \text{tr}(X_1^2) + \alpha_2 \text{tr}(X_2^2) + \alpha_3 \text{tr}(X_3^2) \\ &= \alpha \text{tr}(X^2) + \beta \text{sum}(X^2) + \gamma \text{sum}(X)^2, \end{aligned}$$

where $\alpha_1 = \alpha > 0$, $\alpha_2 = \alpha + \frac{n-2}{2}\beta$ and $\alpha_3 = \alpha + (n-1)(\beta + n\gamma) > 0$.

Proof of Example 3.8: $\mathfrak{S}(n)$ -invariant inner products on $\text{Hol}(n)$, $n \geq 3$.

Direct sum Here, the decomposition and the projections are not trivial so let us check all the statements made above. Indeed, ψ_3 is clearly injective and for all $\mu \in \mathbb{R}^n$, with $\mu_0 = \mu - \frac{\text{sum}(\mu)}{n}\mathbf{1} \in \mathbb{R}_0^n$, we have $\psi_2(\mu_0)\mathbf{1} = \mu_0$ and $[\psi_2(\mu_0) + \psi_3(\text{sum}(\mu))]\mathbf{1} = \mu$, which proves that ψ_2 is injective and ψ_1 is surjective. By definition of X_3 , $\mathbf{1}^\top(X - X_3)\mathbf{1} = \text{sum}(X) - \text{sum}(X_3) = 0$ so $(X - X_3)\mathbf{1} \in \mathbb{R}_0^n$ and p_2 is well defined. Then $X\mathbf{1} = (X - X_3)\mathbf{1} - \psi_2((X - X_3)\mathbf{1}) = 0$ so p_1

is well defined and $\text{Hol}(n) = \ker \phi_1 + \text{im } \psi_2 + \text{im } \psi_3$. Since $\dim \ker \phi_1 + \dim \text{im } \psi_2 + \dim \text{im } \psi_3 = (\dim \text{Hol}(n) - n) + (n - 1) + 1 = \dim \text{Hol}(n)$, the sum is a direct sum. Now if $X \in \ker \phi_1$, $Y = \psi_2(\mu)$ and $Z = \psi_3(x)$, we have:

$$\begin{aligned} \text{tr}(XY^\top) &= \frac{1}{n-2} \text{tr}(X\mu\mathbf{1}^\top + X\mathbf{1}\mu^\top - 2X\text{diag}(\mu)) \\ &= -\frac{2}{n-2} \text{tr}(X\text{diag}(\mu)) = -\frac{2}{n-2} \sum_{i=1}^n \mu_i \sum_{j=1}^n X_{ij} = 0, \\ \text{tr}(XZ^\top) &= \frac{x}{n(n-1)} \text{tr}(X\mathbf{1}\mathbf{1}^\top - X) = 0, \\ \text{tr}(YZ^\top) &= \frac{x}{n(n-1)(n-2)} \text{tr}(\psi_2(\mu)\mathbf{1}\mathbf{1}^\top - \psi_2(\mu)) \\ &= \frac{x}{n(n-1)(n-2)} \text{tr}(\mu\mathbf{1}^\top) = 0, \end{aligned}$$

so the direct sum is orthogonal for the Frobenius inner product.

Irreducibility The vector spaces $\ker \phi_1$, $\text{im } \psi_2$, $\text{im } \psi_3$ are stable because ϕ_1, ψ_2, ψ_3 are equivariant. Then, $\text{im } \psi_2, \text{im } \psi_3$ are irreducible as images of irreducible modules by injective maps. Let us show that $\ker \phi_1 = \{X \in \text{Hol}(n) | X\mathbf{1} = 0\}$ is irreducible. Note that $\ker \phi_1$ is generated by matrices E_{ijkl} with distinct $i, j, k, l \in \{1, \dots, n\}$, defined by:

$$\begin{aligned} E_{ijkl} &= E_{ij} + E_{kl} - E_{il} - E_{jk} \\ &= (e_i - e_k)(e_j - e_l)^\top + (e_j - e_l)(e_i - e_k)^\top \\ &= E_{jilk} = -E_{kjil} = -E_{ilkj}. \end{aligned}$$

Indeed, the subfamily $\{E_{12kl} | 3 \leq k < l \leq n\} \cup \{E_{132l} | 4 \leq l \leq n\}$ is free because if $(\lambda_{kl})_{3 \leq k < l \leq n} \in \mathbb{R}^{\binom{n-2}{2}}$ and $(\lambda_{2l})_{4 \leq l \leq n} \in \mathbb{R}^{n-3}$ are coefficients such that $\sum_{3 \leq k < l \leq n} \lambda_{kl} E_{12kl} + \sum_{4 \leq l \leq n} \lambda_{2l} E_{132l} = 0$, then for $3 \leq k < l \leq n$, the (k, l) -th coefficient is $\lambda_{kl} = 0$ and the $(2, l)$ -th coefficient is $\lambda_{2l} = 0$. Since it is of cardinal $\binom{n-2}{2} + n - 3 = \frac{n(n-3)}{2} = \dim \ker \phi_1$, this subfamily is a basis so the whole family generates $\ker \phi_1$. Moreover, all the matrices E_{ijkl} are in the same orbit under the action of the permutation group $\mathfrak{S}(n)$. Indeed, if E_{rstu} is another one, it suffices to apply a permutation $\sigma \in \mathfrak{S}(n)$ such that $\sigma(i) = r$, $\sigma(j) = s$, $\sigma(k) = t$ and $\sigma(l) = u$ to get $E_{rstu} = P_\sigma E_{ijkl} P_\sigma^\top$.

We are ready to prove that $\ker \phi_1$ is irreducible. Let $W \subseteq \ker \phi_1$, $W \neq \{0\}$, and let $X \in W$, $X \neq 0$. It suffices to show that one $E_{ijkl} \in W$ and we will have $\ker \phi_1 \subseteq W$. First, there exist distinct indexes $i, j, k \in \{1, \dots, n\}$ such that $X_{ik} \neq X_{jk}$, otherwise X would belong to $\text{span}(\mathbf{1}\mathbf{1}^\top - I_n)$ so it would be null. Up to permutation, we can assume that $i = 1$, $j = 2$ and $k = 3$. We define $Y = X - (1, 2) \cdot X = \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix} \in W$

where $A = \begin{pmatrix} a_3 & \cdots & a_n \\ -a_3 & \cdots & -a_n \end{pmatrix} \in \text{Mat}(2, n-2)$ with $\sum_{i=3}^n a_i = 0$ and $(a_3, \dots, a_n) \neq 0$ since $a_3 = X_{13} - X_{12} \neq 0$. Hence, there exist $3 \leq k < l \leq n$ such that $a_k \neq a_l$ so we can define $Z = \frac{1}{a_k - a_l} (Y - (k, l) \cdot Y) = E_{1k2l} \in W$. Therefore, $W = \ker \phi_1$ and $\ker \phi_1$ is irreducible.

Isomorphic modules A simple verification on the dimensions shows that the three terms cannot be isomorphic. Indeed, $\dim \ker \phi_1 = \frac{n(n-3)}{2}$, $\dim \operatorname{im} \psi_2 = n - 1 > \dim \operatorname{im} \psi_3 = 1$ so for $n \geq 5$, $\dim \ker \phi_1 > \dim \operatorname{im} \psi_2$ and for $n \in \{3, 4\}$, $\dim \operatorname{im} \psi_2 > \dim \ker \phi_1 \neq 1$.

General form So the invariant inner products are positive weighted sums of the restriction of the Frobenius inner product to each term. Hence, the $\mathfrak{S}(n)$ -invariant inner products on $\operatorname{Hol}(n)$ are given for all $X \in \operatorname{Hol}(n)$ by:

$$\varphi(X, X) = \alpha_1 \operatorname{tr}(X_1^2) + \alpha_2 \operatorname{tr}(X_2^2) + \alpha_3 \operatorname{tr}(X_3^2) \quad (11.2)$$

$$= \alpha_1 \operatorname{tr}(X^2 - X_2^2 - X_3^2) + \alpha_2 \operatorname{tr}(X_2^2) + \alpha_3 \operatorname{tr}(X_3^2) \quad (11.3)$$

$$= \alpha_1 \operatorname{tr}(X^2) + (\alpha_2 - \alpha_1) \operatorname{tr}(X_2^2) + (\alpha_3 - \alpha_1) \operatorname{tr}(X_3^2), \quad (11.4)$$

with $\alpha_1, \alpha_2, \alpha_3 > 0$. We would like to get a more comprehensive formula without X_i 's, with only X . To do so, we compute $\operatorname{tr}(X_2^2)$ and $\operatorname{tr}(X_3^2)$ in function of X in two steps. First, we compute these quantities in function of $\mu \in \mathbb{R}_0^n$ and $x \in \mathbb{R}$, taking $X_2 = \psi_2(\mu)$ and $X_3 = \psi_3(x)$:

$$\begin{aligned} \operatorname{tr}(X_2^2) &= \frac{1}{(n-2)^2} \operatorname{tr}((\mu \mathbf{1}^\top + \mathbf{1} \mu^\top - 2 \operatorname{diag}(\mu))^2) \\ &= \frac{1}{(n-2)^2} \operatorname{tr}[\underbrace{\operatorname{sum}(\mu)}_0 (\mu \mathbf{1}^\top + \mathbf{1} \mu^\top) + 4 \operatorname{diag}(\mu)^2 + n \mu \mu^\top + \|\mu\|^2 \mathbf{1} \mathbf{1}^\top \\ &\quad - 4 \mu \mu^\top - 2(\mu^2 \mathbf{1}^\top + \mathbf{1}(\mu^2)^\top)] \\ &= \frac{1}{(n-2)^2} (2n-4) \|\mu\|^2 = \frac{2}{n-2} \|\mu\|^2, \\ \operatorname{tr}(X_3^2) &= \frac{x^2}{n^2(n-1)^2} \operatorname{tr}((\mathbf{1} \mathbf{1}^\top - I_n)^2) = \frac{x^2}{n^2(n-1)^2} n(n-1) = \frac{x^2}{n(n-1)}. \end{aligned}$$

Second, we need to express μ and x in function of $X = X_1 + \psi_2(\mu) + \psi_3(x)$. On the one hand, $\operatorname{sum}(X) = 0 + 0 + x$ so $x = \operatorname{sum}(X)$. On the other hand, $X \mathbf{1} = 0 + \mu + \frac{x}{n} \mathbf{1}$ so $\mu = X \mathbf{1} - \frac{\operatorname{sum}(X)}{n} \mathbf{1}$. We finish by computing $\|\mu\|^2$ in function of X :

$$\begin{aligned} \|\mu\|^2 &= \mu^\top \mu = \left(\mathbf{1}^\top X - \frac{\operatorname{sum}(X)}{n} \mathbf{1}^\top \right) \left(X \mathbf{1} - \frac{\operatorname{sum}(X)}{n} \mathbf{1} \right) \\ &= \operatorname{sum}(X^2) - \frac{1}{n} \operatorname{sum}(X)^2. \end{aligned}$$

Altogether, the inner product writes:

$$\varphi(X, X) = \alpha \operatorname{tr}(X^2) + \beta \operatorname{sum}(X^2) + \gamma \operatorname{sum}(X)^2, \quad (11.5)$$

with $\alpha = \alpha_1$, $\beta = \frac{2}{n-2}(\alpha_2 - \alpha_1)$ and $\gamma = \frac{1}{n(n-1)}(\alpha_3 - \alpha_1) - \frac{2}{n(n-2)}(\alpha_2 - \alpha_1)$. The inverse relations write $\alpha_1 = \alpha$, $\alpha_2 = \alpha + \frac{n-2}{2}\beta$ and $\alpha_3 = \alpha + (n-1)(\beta + n\gamma)$ as expected. \square

11.2.4 Proof of Example 3.10

Example 3.10 ($\mathfrak{S}(n)$ -invariant inner products on $\operatorname{Sym}(n)$) The irreducible decomposition of $\operatorname{Sym}(n)$ with respect to the congruence action of $\mathfrak{S}(n)$, i.e. of permutation matrices, is

$\text{Sym}(n) = \underbrace{\ker \phi_1 \oplus \text{im } \psi_2 \oplus \text{im } \psi_3}_{\text{Hol}(n)} \oplus \underbrace{\text{im } \psi_4 \oplus \text{im } \psi_5}_{\text{Diag}(n)} \simeq \ker \phi_1 \oplus 2\mathbb{R}_0^n \oplus 2\mathbb{R}$ where ϕ_1, ψ_2, ψ_3 were

defined in the previous section and:

4. $\psi_4 : \lambda \in \mathbb{R}_0^n \mapsto \text{diag}(\lambda) \in \text{Diag}(n)$,
5. $\psi_5 : y \in \mathbb{R} \mapsto \frac{y}{n}I_n \in \text{Diag}(n)$,

with ψ_4, ψ_5 injective. The corresponding orthogonal projections are:

- $\pi_5 : X \in \text{Sym}(n) \mapsto X_5 = \psi_5(\text{tr}(X)) \in \text{im } \psi_5$,
- $\pi_4 : X \in \text{Sym}(n) \mapsto X_4 = \text{Diag}(X - X_5) \in \text{im } \psi_4$,
- $\pi_3 : X \in \text{Sym}(n) \mapsto X_3 = p_3(X - \text{Diag}(X)) \in \text{im } \psi_3$,
- $\pi_2 : X \in \text{Sym}(n) \mapsto X_2 = p_2(X - \text{Diag}(X)) \in \text{im } \psi_2$,
- $\pi_1 : X \in \text{Sym}(n) \mapsto X_1 = p_1(X - \text{Diag}(X)) \in \ker \phi_1$.

The $\mathfrak{S}(n)$ -invariant inner products on $\text{Sym}(n)$ are given for all $X = X_1 + \psi_2(\mu) + \psi_3(x) + \psi_4(\lambda) + \psi_5(y) \in \text{Sym}(n)$ by:

$$\begin{aligned} \varphi(X, X) &= \alpha_1 \text{tr}(X_1^2) + S_{11} \|\mu\|^2 + \Sigma_{11} x^2 + S_{22} \|\lambda\|^2 + \Sigma_{22} y^2 + 2S_{12} \lambda^\top \mu + 2\Sigma_{12} xy \\ &= \alpha \text{tr}(X^2) + \beta \text{sum}(X^2) + \gamma \text{sum}(X)^2 + \delta \text{tr}(\text{Diag}(X)^2) \\ &\quad + \varepsilon \text{sum}(\text{Diag}(X)X) + \zeta \text{tr}(X)^2 + \eta \text{tr}(X) \text{sum}(X), \end{aligned}$$

where $\alpha_1 = \alpha > 0$ and the SPD matrices are $S = \begin{pmatrix} \frac{2}{n-2}\alpha + \beta & \beta + \frac{\varepsilon}{2} \\ \beta + \frac{\varepsilon}{2} & \alpha + \beta + \delta + \varepsilon \end{pmatrix} \in \text{Sym}^+(2)$

and $\Sigma = \frac{1}{n} \begin{pmatrix} \frac{1}{n-1}\alpha + \beta + n\gamma & \beta + \frac{\varepsilon}{2} + n(\gamma + \frac{\eta}{2}) \\ \beta + \frac{\varepsilon}{2} + n(\gamma + \frac{\eta}{2}) & \alpha + \beta + \delta + \varepsilon + n(\gamma + \zeta + \eta) \end{pmatrix} \in \text{Sym}^+(2)$.

Proof of Example 3.10: $\mathfrak{S}(n)$ -invariant inner products on $\text{Sym}(n)$, $n \geq 3$.

Direct sum Based on the study of the action of $\mathfrak{S}(n)$ on $\text{Hol}(n)$ and $\mathbb{R}^n \simeq \text{Diag}(n)$, the direct sum and the orthogonal projections are clear.

Irreducibility The terms in $\text{Hol}(n)$ were treated in the previous example. The spaces $\text{im } \psi_4$ and $\text{im } \psi_5$ are irreducible modules as images by equivariant maps of irreducible modules.

Isomorphic modules On the one hand, $\text{im } \psi_2$ and $\text{im } \psi_4$ are isomorphic to \mathbb{R}_0^n as modules. On the other hand, $\text{im } \psi_3$ and $\text{im } \psi_5$ are isomorphic to \mathbb{R} as modules. None of them is isomorphic to $\ker \phi_1$ because they have different dimensions.

General form Therefore, Theorem 3.4 ensures that the $\mathfrak{S}(n)$ -invariant inner products on $\text{Sym}(n)$ are given for all $X = X_1 + \psi_2(\mu) + \psi_3(x) + \psi_4(\lambda) + \psi_5(y) \in \text{Sym}(n)$ by:

$$\varphi(X, X) = \alpha_1 \text{tr}(X_1^2) + S_{11} \|\mu\|^2 + \Sigma_{11} x^2 + S_{22} \|\lambda\|^2 + \Sigma_{22} y^2 + 2S_{12} \lambda^\top \mu + 2\Sigma_{12} xy \quad (11.6)$$

$$\begin{aligned} &= \alpha_1 \text{tr}(X^2) + \left(S_{11} - \frac{2}{n-2} \alpha_1 \right) \|\mu\|^2 + \left(\Sigma_{11} - \frac{1}{n(n-1)} \alpha_1 \right) x^2 \\ &\quad + (S_{22} - \alpha_1) \|\lambda\|^2 + \left(\Sigma_{22} - \frac{1}{n} \alpha_1 \right) y^2 + 2S_{12} \lambda^\top \mu + 2\Sigma_{12} xy, \end{aligned} \quad (11.7)$$

where $\alpha_1 > 0$, $S \in \text{Sym}^+(2)$ and $\Sigma \in \text{Sym}^+(2)$, because $\text{tr}(X_1^2) = \text{tr}(X^2) - \sum_{i=2}^5 \text{tr}(X_i^2)$, because the computations of $\text{tr}(\psi_2(\mu)^2) = \frac{2}{n-2}\|\mu\|^2$ and $\text{tr}(\psi_3(x)^2) = \frac{x^2}{n(n-1)}$ made previously are still valid and because $\text{tr}(X_4^2) = \|\lambda\|^2$ and $\text{tr}(X_5^2) = \frac{y^2}{n}$. To get an explicit formula in function of X , we need to compute $\mu, \lambda \in \mathbb{R}_0^n$ and $x, y \in \mathbb{R}$:

$$\begin{aligned}\text{tr}(X) &= 0 + 0 + 0 + 0 + y, \\ \text{Diag}(X) &= 0 + 0 + 0 + \text{diag}(\lambda) + \frac{y}{n}I_n, \\ \text{sum}(X) &= 0 + 0 + x + 0 + y, \\ X\mathbf{1} &= 0 + \mu + \frac{x}{n}\mathbf{1} + \lambda + \frac{y}{n}\mathbf{1},\end{aligned}$$

so $y = \text{tr}(X)$, $\text{diag}(\lambda) = \text{Diag}(X) - \frac{y}{n}I_n$, $x = \text{sum}(X) - \text{tr}(X)$ and $\mu = X\mathbf{1} - \lambda - \frac{x+y}{n}\mathbf{1}$. Then:

$$\|\mu\|^2 = \left(\mathbf{1}^\top X - \lambda^\top - \frac{x+y}{n}\mathbf{1}^\top\right) \left(X\mathbf{1} - \lambda - \frac{x+y}{n}\mathbf{1}\right) \quad (11.8)$$

$$= \left(\text{sum}(X^2) + \|\lambda\|^2 - \frac{(x+y)^2}{n} - 2\lambda^\top X\mathbf{1}\right), \quad (11.9)$$

where:

$$\|\lambda\|^2 = (\mathbf{1}^\top \text{Diag}(X) - \frac{y}{n}\mathbf{1}^\top)(\text{Diag}(X)\mathbf{1} - \frac{y}{n}\mathbf{1}) \quad (11.10)$$

$$= \text{tr}(\text{Diag}(X)^2) - \frac{1}{n}\text{tr}(X)^2, \quad (11.11)$$

$$\lambda^\top X\mathbf{1} = (\mathbf{1}^\top \text{Diag}(X) - \frac{y}{n}\mathbf{1}^\top)X\mathbf{1} = \text{sum}(X\text{Diag}(X)) - \frac{1}{n}\text{sum}(X)\text{tr}(X). \quad (11.12)$$

So we obtain:

$$\begin{aligned}\|\mu\|^2 &= \text{sum}(X^2) + \text{tr}(\text{Diag}(X)^2) - \frac{1}{n}\text{tr}(X)^2 - \frac{1}{n}\text{sum}(X)^2 \\ &\quad - 2\text{sum}(X\text{Diag}(X)) + \frac{2}{n}\text{sum}(X)\text{tr}(X),\end{aligned} \quad (11.13)$$

$$x^2 = \text{sum}(X)^2 + \text{tr}(X)^2 - 2\text{sum}(X)\text{tr}(X), \quad (11.14)$$

$$\|\lambda\|^2 = \text{tr}(\text{Diag}(X)^2) - \frac{1}{n}\text{tr}(X)^2, \quad (11.15)$$

$$y^2 = \text{tr}(X)^2, \quad (11.16)$$

$$\lambda^\top \mu = \lambda^\top X\mathbf{1} - \|\lambda\|^2 \quad (11.17)$$

$$= \text{sum}(X\text{Diag}(X)) - \frac{1}{n}\text{sum}(X)\text{tr}(X) - \text{tr}(\text{Diag}(X)^2) + \frac{1}{n}\text{tr}(X)^2, \quad (11.18)$$

$$xy = \text{sum}(X)\text{tr}(X) - \text{tr}(X)^2. \quad (11.19)$$

Altogether, the inner product writes:

$$\begin{aligned}\varphi(X, X) &= \alpha \text{tr}(X^2) + \beta \text{sum}(X^2) + \gamma \text{sum}(X)^2 + \delta \text{tr}(\text{Diag}(X)^2) \\ &\quad + \varepsilon \text{sum}(\text{Diag}(X)X) + \zeta \text{tr}(X)^2 + \eta \text{tr}(X) \text{sum}(X),\end{aligned}$$

with:

$$\begin{aligned}
\alpha &= \alpha_1, \\
\beta &= -\frac{2}{n-2}\alpha_1 + S_{11}, \\
\gamma &= \underbrace{\left(\frac{2}{n(n-2)} - \frac{1}{n(n-1)}\right)}_{\frac{1}{(n-2)(n-1)}}\alpha_1 - \frac{1}{n}S_{11} + \Sigma_{11}, \\
\delta &= -\underbrace{\left(1 + \frac{2}{n-2}\right)}_{\frac{n}{n-2}}\alpha_1 + S_{11} + S_{22} - 2S_{12}, \\
\varepsilon &= \frac{4}{n-2}\alpha_1 - 2S_{11} + 2S_{12}, \\
\zeta &= \underbrace{\left(\frac{2}{n(n-2)} - \frac{1}{n(n-1)} + \frac{1}{n} - \frac{1}{n}\right)}_{\frac{1}{(n-2)(n-1)}}\alpha_1 - \frac{1}{n}S_{11} + \Sigma_{11} - \frac{1}{n}S_{22} + \Sigma_{22} + \frac{2}{n}S_{12} - 2\Sigma_{12}, \\
\eta &= \underbrace{\left(-\frac{4}{n(n-2)} + \frac{2}{n(n-1)}\right)}_{\frac{-2}{(n-2)(n-1)}}\alpha_1 + \frac{2}{n}S_{11} - 2\Sigma_{11} - \frac{2}{n}S_{12} + 2\Sigma_{12}.
\end{aligned}$$

The inverse relations are:

$$\begin{aligned}
\alpha_1 &= \alpha, \\
S_{11} &= \frac{2}{n-2}\alpha + \beta, \\
\Sigma_{11} &= \gamma + \frac{1}{n}\beta + \frac{1}{n(n-1)}\alpha, \\
S_{12} &= \beta + \frac{\varepsilon}{2}, \\
S_{22} &= \alpha + \beta + \delta + \varepsilon, \\
\Sigma_{12} &= \frac{1}{n}\left(\beta + \frac{\varepsilon}{2}\right) + \gamma + \frac{\eta}{2}, \\
\Sigma_{22} &= \frac{1}{n}(\alpha + \beta + \delta + \varepsilon) + \gamma + \zeta + \eta,
\end{aligned}$$

as expected. □

11.2.5 Proof of Example 3.12

Example 3.12 ($\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$) The irreducible decomposition

of $\text{Mat}(n)$ with respect to the congruence action of $\mathfrak{S}(n)$, i.e. of permutation matrices, is:

$$\begin{aligned} \text{Mat}(n) &= \underbrace{\ker \phi_1 \oplus \text{im } \psi_2 \oplus \text{im } \psi_3}_{\text{Hol}(n)} \oplus \underbrace{\text{im } \psi_4 \oplus \text{im } \psi_5}_{\text{Diag}(n)} \oplus \underbrace{\ker \phi_6 \oplus \text{im } \psi_7}_{\text{Skew}(n)} \\ &\quad \underbrace{\hspace{10em}}_{\text{Sym}(n)} \\ &\simeq \ker \phi_1 \oplus \ker \phi_6 \oplus 3\mathbb{R}_0^n \oplus 2\mathbb{R}, \end{aligned}$$

where $\phi_1, \psi_2, \psi_3, \psi_4, \psi_5$ were defined in the previous sections and:

6. $\phi_6 : X \in \text{Skew}(n) \mapsto X\mathbf{1} \in \mathbb{R}_0^n$,
7. $\psi_7 : \nu \in \mathbb{R}_0^n \mapsto \frac{1}{n}(\nu\mathbf{1}^\top - \mathbf{1}\nu^\top) \in \text{Skew}(n)$,

with ϕ_6 surjective and ψ_7 injective. The corresponding orthogonal projections are:

- $\Pi_7 : X \in \text{Mat}(n) \mapsto X_7 = \psi_7 \left(\frac{X-X^\top}{2} \mathbf{1} \right) \in \text{im } \psi_7$,
- $\Pi_6 : X \in \text{Mat}(n) \mapsto X_6 = \frac{X-X^\top}{2} - X_7 \in \ker \phi_6$,
- $\Pi_5 : X \in \text{Mat}(n) \mapsto X_5 = \pi_5 \left(\frac{X+X^\top}{2} \right) = \psi_5(\text{tr}(X)) \in \text{im } \psi_5$,
- $\Pi_4 : X \in \text{Mat}(n) \mapsto X_4 = \pi_4 \left(\frac{X+X^\top}{2} \right) = \text{Diag}(X - X_5) \in \text{im } \psi_4$,
- $\Pi_3 : X \in \text{Mat}(n) \mapsto X_3 = \pi_3 \left(\frac{X+X^\top}{2} \right) \in \text{im } \psi_3$,
- $\Pi_2 : X \in \text{Mat}(n) \mapsto X_2 = \pi_2 \left(\frac{X+X^\top}{2} \right) \in \text{im } \psi_2$,
- $\Pi_1 : X \in \text{Mat}(n) \mapsto X_1 = \pi_1 \left(\frac{X+X^\top}{2} \right) \in \ker \phi_1$.

The $\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$ are given for all $X = X_1 + \psi_2(\mu) + \psi_3(x) + \psi_4(\lambda) + \psi_5(y) + X_6 + \psi_7(\nu) \in \text{Mat}(n)$ by:

$$\begin{aligned} \varphi(X, X) &= \alpha_1 \text{tr}(X_1^2) + S_{11} \|\mu\|^2 + \Sigma_{11} x^2 + S_{22} \|\lambda\|^2 + \Sigma_{22} y^2 + \alpha_2 \text{tr}(X_6^2) + S_{33} \|\nu\|^2 \\ &\quad + 2S_{12} \lambda^\top \mu + 2S_{23} \nu^\top \lambda + 2S_{13} \mu^\top \nu + 2\Sigma_{12} xy \\ &= \alpha \text{tr}(XX^\top) + \alpha' \text{tr}(X^2) + \beta \text{sum}(X^2) + \beta' \text{sum}(XX^\top) + \beta'' \text{sum}(X^\top X) \\ &\quad + \gamma \text{sum}(X)^2 + \delta \text{tr}(\text{Diag}(X)^2) + \zeta \text{tr}(X)^2 + \eta \text{tr}(X) \text{sum}(X) \\ &\quad + \varepsilon \text{sum}(\text{Diag}(X)X) + \varepsilon' \text{sum}(\text{Diag}(X)X^\top), \end{aligned}$$

where $\alpha_1 = \frac{\alpha+\alpha'}{2} > 0$, $\alpha_2 = \frac{\alpha-\alpha'}{2} > 0$ and the SPD matrices are:

$$\begin{aligned} S &= \begin{pmatrix} \frac{2}{n-2} \frac{\alpha+\alpha'}{2} + \beta + \beta' + \beta'' & \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') & \frac{\beta''-\beta'}{4} \\ \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') & \frac{\alpha+\alpha'}{2} + \beta + \beta' + \beta'' + \delta + \varepsilon + \varepsilon' & \frac{\beta''-\beta'}{4} + \frac{\varepsilon'-\varepsilon}{2} \\ \frac{\beta''-\beta'}{4} & \frac{\beta''-\beta'}{4} + \frac{\varepsilon'-\varepsilon}{2} & 2\alpha + \beta - \beta' - \beta'' \end{pmatrix} \in \text{Sym}^+(3), \\ \Sigma &= \frac{1}{n} \begin{pmatrix} \frac{1}{n-1} \frac{\alpha+\alpha'}{2} + \beta + \beta' + \beta'' + n\gamma & \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') + n(\gamma + \frac{\eta}{2}) \\ \beta + \frac{1}{2}(\varepsilon + \varepsilon') + n(\gamma + \frac{\eta}{2}) & \frac{\alpha+\alpha'}{2} + \beta + \beta' + \beta'' + \delta + \varepsilon + \varepsilon' + n(\gamma + \zeta + \eta) \end{pmatrix} \in \text{Sym}^+(2). \end{aligned}$$

Proof of Example 3.12: $\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$, $n \geq 3$.

Direct sum Given the previous sections, we only need to check the direct sum $\text{Skew}(n) = \ker \phi_6 \oplus \text{im } \psi_7$. Since $\psi_7(\nu)\mathbf{1} = \nu$ for all $\nu \in \mathbb{R}_0^n$, ϕ_6 is surjective and ψ_7 is injective. Since $\mathbf{1}^\top \frac{X-X^\top}{2} \mathbf{1} = \frac{1}{2}(\text{sum}(X) - \text{sum}(X^\top)) = 0$, $\frac{X-X^\top}{2} \mathbf{1} \in \mathbb{R}_0^n$ so Π_7 is well defined. Hence, $\left(\frac{X-X^\top}{2} - X_7\right) \mathbf{1} = \frac{X-X^\top}{2} \mathbf{1} - \psi_7\left(\frac{X-X^\top}{2} \mathbf{1}\right) \mathbf{1} = 0$ so Π_6 is well defined and $\text{Skew}(n) = \ker \phi_6 + \text{im } \psi_7$. Since $\dim \ker \phi_6 + \dim \text{im } \psi_7 = (\dim \text{Skew}(n) - (n-1)) + (n-1) = \dim \text{Skew}(n)$, the sum is a direct sum. Moreover, if $X \in \ker \phi_6$ and $Y = \psi_7(\nu) \in \text{im } \psi_7$ with $\nu \in \mathbb{R}_0^n$, then:

$$\text{tr}(YX^\top) = \frac{1}{n} \text{tr}(\nu \mathbf{1}^\top X^\top - \mathbf{1} \nu^\top X^\top) = \frac{1}{n} \text{tr}(\mathbf{1} \nu^\top X) = 0,$$

so the sum is orthogonal for the Frobenius inner product. The other projections Π_1, \dots, Π_5 are the compositions of the previous projections π_1, \dots, π_5 with the projection onto symmetric matrices $X \in \text{Mat}(n) \mapsto \frac{X+X^\top}{2}$.

Irreducibility The vector spaces $\ker \phi_6$ and $\text{im } \psi_7$ are stable because ϕ_6 and ψ_7 are equivariant. Then, $\text{im } \psi_7$ is irreducible as image of an irreducible module by an injective map. Let us show that $\ker \phi_6 = \{X \in \text{Skew}(n) | X\mathbf{1} = 0\}$ is irreducible. Note that $\ker \phi_6$ is generated by matrices F_{ijkl} with $i, j, k, l \in \{1, \dots, n\}$, $i \neq k$ and $j \neq l$ and $(i \neq j \text{ or } k \neq l)$, defined by:

$$\begin{aligned} F_{ijkl} &= (e_i - e_k)(e_j - e_l)^\top - (e_j - e_l)(e_i - e_k)^\top \\ &= (e_i e_j^\top - e_j e_i^\top) + (e_k e_l^\top - e_l e_k^\top) - (e_i e_l^\top - e_l e_i^\top) - (e_k e_j^\top - e_j e_k^\top) \\ &= F_{jilk} = -F_{kjil} = -F_{ilkj}. \end{aligned}$$

Indeed, the subfamily $(F_{11kl})_{2 \leq k < l \leq n}$ is free because if $(\lambda_{kl})_{2 \leq k < l \leq n} \in \mathbb{R}^{\binom{n-1}{2}}$ are coefficients such that $\sum_{2 \leq k < l \leq n} \lambda_{kl} F_{11kl} = 0$, then the (k, l) -th coefficient is $\lambda_{kl} = 0$. Since it is of cardinal $\binom{n-1}{2} = \frac{(n-1)(n-2)}{2} = \dim \ker \phi_6$, this subfamily is a basis so the whole family generates $\ker \phi_6$.

For $n = 3$, $\dim \ker \phi_6 = 1$ so $\ker \phi_6$ is irreducible. We assume that $n \geq 4$. Let $W \subseteq \ker \phi_6$ be a submodule, $W \neq \{0\}$. Let $X \in W$, $X \neq 0$. By absurdum, if for all distinct $i, j, k, l \in$

$\{1, \dots, n\}$, $X_{ik} - X_{jk} = X_{il} - X_{jl}$, then:

$$\begin{aligned}
\forall i \neq j \neq l \neq i, 0 &= \sum_{k=1}^n (X_{ik} - X_{jk}) \\
&= -X_{ji} + X_{ij} + \sum_{k \neq i, j} (X_{ik} - X_{jk}) \\
&= 2X_{ij} + \sum_{k \neq i, j} (X_{il} - X_{jl}) \\
&= 2X_{ij} + (n-2)(X_{il} - X_{jl}), \\
\forall j \neq l, 0 &= \sum_{i=1}^n 2X_{ij} \\
&= 2X_{lj} + \sum_{i \neq j, l} 2X_{ij} \\
&= 2X_{lj} - \sum_{i \neq j, l} (n-2)(X_{il} - X_{jl}) \\
&= 2X_{lj} - (n-2)[(0 - X_{jl}) - (n-2)X_{jl}] \\
&= n(n-3)X_{jl},
\end{aligned}$$

so X would be 0. Therefore, there exist distinct $i, j, k, l \in \{1, \dots, n\}$ such that $X_{ik} - X_{jk} = X_{il} - X_{jl}$. Up to a permutation, we can assume that $(i, j, k, l) = (1, 2, 3, 4)$. Let $Y = X - (1, 2) \cdot X = \begin{pmatrix} B & A \\ -A^\top & 0 \end{pmatrix} \in W$ where $B = \begin{pmatrix} 0 & a_2 \\ -a_2 & 0 \end{pmatrix} \in \text{Skew}(2)$ and $A = \begin{pmatrix} a_3 & a_4 & \cdots & a_n \\ -a_3 & -a_4 & \cdots & -a_n \end{pmatrix} \in \text{Mat}(2, n-2)$. Hence $a_3 = X_{13} - X_{23} \neq X_{14} - X_{24} = a_4$ so $F_{1324} = \frac{1}{a_3 - a_4}(Y - (3, 4) \cdot Y) \in W$. By permutations, all $F_{ijkl} \in W$ with distinct $i, j, k, l \in \{1, \dots, n\}$.

Moreover, there exist $i \neq j$ such that $X_{ij} \neq 0$. Up to a permutation, we can assume that $(i, j) = (1, 2)$. Let $Y' = X - (1, 2) \cdot X = \begin{pmatrix} B' & A' \\ -A'^\top & 0 \end{pmatrix} \in W$ where $B' = \begin{pmatrix} 0 & a'_2 \\ -a'_2 & 0 \end{pmatrix} \in \text{Skew}(2)$ and $A' = \begin{pmatrix} a'_3 & a'_4 & \cdots & a'_n \\ -a'_3 & -a'_4 & \cdots & -a'_n \end{pmatrix} \in \text{Mat}(2, n-2)$ with $a'_2 = 2X_{12} \neq 0$. Let $\sigma = (3, 4, \dots, n) \in \mathfrak{S}(n)$, let $Z_k = \sigma^k \cdot Y'$ for $k \in \{0, \dots, n-3\}$ and let $Z = \frac{1}{a'_2} \sum_{k=0}^{n-3} Z_k =$

$$\begin{pmatrix} 0 & n-2 & -1 & \cdots & -1 \\ -(n-2) & 0 & 1 & \cdots & 1 \\ 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & 0 & \cdots & 0 \end{pmatrix}. \text{ Finally, } F_{1123} = \frac{1}{n-2}(Z - \sum_{k=4}^n F_{132k}) \in W \text{ and by}$$

permutations, all $F_{iikl} \in W$ for distinct $i, k, l \in \{1, \dots, n\}$. So $W = \ker \phi_6$ and $\ker \phi_6$ is irreducible.

Isomorphic modules The module $\text{im } \psi_7$ is isomorphic to \mathbb{R}_0^n , thus to $\text{im } \psi_2$ and $\text{im } \psi_4$. We still have $\text{im } \psi_3 \simeq \psi_5$. For $n \geq 4$, the other dimensions don't match. For $n = 3$, $\dim \ker \phi_6 = 1$ but the permutation $(1, 2)$ changes $X \in \ker \phi_6$ into $-X$ while it does not act on $\text{im } \psi_3$ or $\text{im } \psi_5$. So there is no other isomorphism.

General form Therefore, Theorem 3.4 ensures that the $\mathfrak{S}(n)$ -invariant inner products on $\text{Mat}(n)$ are given for all $X = X_1 + \psi_2(\mu) + \psi_3(x) + \psi_4(\lambda) + \psi_5(y) + X_6 + \psi_7(\nu) \in \text{Mat}(n)$ by:

$$\begin{aligned} \varphi(X, X) &= \alpha_1 \text{tr}(X_1^2) + S_{11} \|\mu\|^2 + \Sigma_{11} x^2 + S_{22} \|\lambda\|^2 + \Sigma_{22} y^2 + \alpha_2 \text{tr}(X_6^2) + S_{33} \|\nu\|^2 \\ &\quad + 2S_{12} \lambda^\top \mu + 2S_{23} \nu^\top \lambda + 2S_{13} \mu^\top \nu + 2\Sigma_{12} xy \\ &= (\alpha_1 + \alpha_2) \text{tr}(XX^\top) + (\alpha_1 - \alpha_2) \text{tr}(X^2) + \left(S_{11} - \frac{2}{n-2} \alpha_1 \right) \|\mu\|^2 \\ &\quad + \left(\Sigma_{11} - \frac{1}{n(n-1)} \alpha_1 \right) x^2 + (S_{22} - \alpha_1) \|\lambda\|^2 + \left(\Sigma_{22} - \frac{1}{n} \alpha_1 \right) y^2 \\ &\quad + (S_{33} - 2\alpha_1 - 2\alpha_2) \|\nu\|^2 + 2S_{12} \lambda^\top \mu + 2S_{13} \nu^\top \lambda + 2S_{23} \mu^\top \nu + 2\Sigma_{12} xy, \end{aligned}$$

where $\alpha_1, \alpha_2 > 0$, $S \in \text{Sym}^+(3)$ and $\Sigma \in \text{Sym}^+(2)$, because:

- $\text{tr}(X_1^2) = \text{tr}(XX^\top) - \sum_{i=2}^5 \text{tr}(X_i^2) - \left\| \frac{X-X^\top}{2} \right\|^2 = \frac{\text{tr}(XX^\top) + \text{tr}(X^2)}{2} - \sum_{i=2}^5 \text{tr}(X_i^2)$,
- $\text{tr}(\psi_2(\mu)^2) = \frac{2}{n-2} \|\mu\|^2$, $\text{tr}(\psi_3(x)^2) = \frac{x^2}{n(n-1)}$, $\text{tr}(X_4^2) = \|\lambda\|^2$, $\text{tr}(X_5^2) = \frac{y^2}{n}$ as seen earlier,
- $\text{tr}(X_6 X_6^\top) = \left\| \frac{X-X^\top}{2} \right\|^2 - \text{tr}(X_7 X_7^\top) = \frac{\text{tr}(XX^\top) - \text{tr}(X^2)}{2} - \text{tr}(X_7 X_7^\top)$ and $\text{tr}(X_7 X_7^\top) = 2\|\nu\|^2$.

The previous expressions of $\mu, \lambda \in \mathbb{R}_0^n$ and $x, y \in \mathbb{R}$ in function of X are still valid if we replace X by $\frac{X+X^\top}{2} \in \text{Sym}(n)$. So $y = \text{tr}\left(\frac{X+X^\top}{2}\right) = \text{tr}(X)$, $\text{diag}(\lambda) = \text{Diag}\left(\frac{X+X^\top}{2}\right) - \frac{y}{n} I_n = \text{Diag}(X) - \frac{y}{n} I_n$, $x = \text{sum}\left(\frac{X+X^\top}{2}\right) - \text{tr}\left(\frac{X+X^\top}{2}\right) = \text{sum}(X) - \text{tr}(X)$ and $\mu = \frac{X+X^\top}{2} \mathbf{1} - \lambda - \frac{x+y}{n} \mathbf{1}$. Updating Equations (11.8) to (11.19), we get:

$$\begin{aligned} \|\mu\|^2 &= \frac{1}{2} \text{sum}(X^2) + \frac{1}{4} \text{sum}(XX^\top) + \frac{1}{4} \text{sum}(X^\top X) + \text{tr}(\text{Diag}(X)^2) - \frac{1}{n} \text{tr}(X)^2 - \frac{1}{n} \text{sum}(X)^2 \\ &\quad - \text{sum}(\text{Diag}(X)X) - \text{sum}(\text{Diag}(X)X^\top) + \frac{2}{n} \text{sum}(X) \text{tr}(X), \\ x^2 &= \text{sum}(X)^2 + \text{tr}(X)^2 - 2 \text{sum}(X) \text{tr}(X), \\ \|\lambda\|^2 &= \text{tr}(\text{Diag}(X)^2) - \frac{1}{n} \text{tr}(X)^2, \\ y^2 &= \text{tr}(X)^2, \\ \lambda^\top \mu &= \frac{1}{2} \text{sum}(\text{Diag}(X)X) + \frac{1}{2} \text{sum}(\text{Diag}(X)X^\top) - \frac{1}{n} \text{sum}(X) \text{tr}(X) - \text{tr}(\text{Diag}(X)^2) + \frac{1}{n} \text{tr}(X)^2, \\ xy &= \text{sum}(X) \text{tr}(X) - \text{tr}(X)^2. \end{aligned}$$

In addition, $\nu = \frac{X-X^\top}{2} \mathbf{1}$ so:

$$\begin{aligned}
\|\nu\|^2 &= \frac{1}{4} \text{sum}((X - X^\top)^\top (X - X^\top)) \\
&= \frac{1}{2} \text{sum}(X^2) - \frac{1}{4} \text{sum}(XX^\top) - \frac{1}{4} \text{sum}(X^\top X), \\
\nu^\top \lambda &= \frac{1}{2} \text{sum} \left((X - X^\top)^\top \left(\text{Diag}(X) - \frac{\text{tr}(X)}{n} I_n \right) \right) \\
&= \frac{1}{2} \text{sum}(\text{Diag}(X)X^\top) - \frac{1}{2} \text{sum}(\text{Diag}(X)X), \\
\nu^\top \mu &= -\nu^\top \lambda + \frac{1}{2} \text{sum} \left((X - X^\top)^\top \left(\frac{X + X^\top}{2} - \frac{x+y}{n} I_n \right) \right) \\
&= \frac{1}{2} \text{sum}(\text{Diag}(X)X) - \frac{1}{2} \text{sum}(\text{Diag}(X)X^\top) + \frac{1}{4} \text{sum}(X^\top X) - \frac{1}{4} \text{sum}(XX^\top).
\end{aligned}$$

Altogether, the inner product writes:

$$\begin{aligned}
\varphi(X, X) &= \alpha \text{tr}(XX^\top) + \alpha' \text{tr}(X^2) + \beta \text{sum}(X^2) + \beta' \text{sum}(XX^\top) + \beta'' \text{sum}(X^\top X) \\
&\quad + \gamma \text{sum}(X)^2 + \delta \text{tr}(\text{Diag}(X)^2) + \zeta \text{tr}(X)^2 + \eta \text{tr}(X) \text{sum}(X) \\
&\quad + \varepsilon \text{sum}(\text{Diag}(X)X) + \varepsilon' \text{sum}(\text{Diag}(X)X^\top), \tag{11.20}
\end{aligned}$$

where:

$$\begin{aligned}
\alpha &= \alpha_1 + \alpha_2, \\
\alpha' &= \alpha_1 - \alpha_2, \\
\beta &= \left(-\frac{1}{n-2} - 1 \right) \alpha_1 - \alpha_2 + \frac{1}{2} (S_{11} + S_{33}), \\
\beta' &= \frac{1}{2} \left(-\frac{1}{n-2} + 1 \right) \alpha_1 + \frac{1}{2} \alpha_2 + \frac{1}{4} (S_{11} - S_{33} - 2S_{13}), \\
\beta'' &= \frac{1}{2} \left(-\frac{1}{n-2} + 1 \right) \alpha_1 + \frac{1}{2} \alpha_2 + \frac{1}{4} (S_{11} - S_{33} + 2S_{13}), \\
\gamma &= \left(\frac{2}{n(n-2)} - \frac{1}{n(n-1)} \right) \alpha_1 - \frac{1}{n} S_{11} + \Sigma_{11}, \\
\delta &= \left(-\frac{2}{n-2} - 1 \right) \alpha_1 + S_{11} + S_{22} - 2S_{12}, \\
\varepsilon &= \frac{2}{n-2} \alpha_1 - S_{11} + S_{12} - S_{23} + S_{13}, \\
\varepsilon' &= \frac{2}{n-2} \alpha_1 - S_{11} + S_{12} + S_{23} - S_{13}, \\
\zeta &= \left(\frac{2}{n(n-2)} - \frac{1}{n(n-1)} + \frac{1}{n} - \frac{1}{n} \right) \alpha_1 - \frac{1}{n} S_{11} + \Sigma_{11} - \frac{1}{n} S_{22} + \Sigma_{22} + \frac{2}{n} S_{12} - 2\Sigma_{12}, \\
\eta &= \left(-\frac{4}{n(n-2)} + \frac{2}{n(n-1)} \right) \alpha_1 + \frac{2}{n} S_{11} - 2\Sigma_{11} - \frac{2}{n} S_{12} + 2\Sigma_{12}.
\end{aligned}$$

The inverse relations are:

$$\begin{aligned}\alpha_1 &= \frac{\alpha + \alpha'}{2}, \\ \alpha_2 &= \frac{\alpha - \alpha'}{2}, \\ S_{13} &= \frac{\beta'' - \beta'}{4}, \\ S_{11} &= \frac{2}{n-2} \frac{\alpha + \alpha'}{2} + \beta + \beta' + \beta'', \\ S_{33} &= 2\alpha + \beta - \beta' - \beta'', \\ \Sigma_{11} &= \frac{1}{n(n-1)} \frac{\alpha + \alpha'}{2} + \frac{1}{n}(\beta + \beta' + \beta'') + \gamma, \\ S_{12} &= \beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon'), \\ S_{22} &= \frac{\alpha + \alpha'}{2} + \beta + \beta' + \beta'' + \delta + \varepsilon + \varepsilon', \\ S_{23} &= \frac{\beta'' - \beta'}{4} + \frac{\varepsilon' - \varepsilon}{2}, \\ \Sigma_{12} &= \frac{1}{n} \left(\beta + \beta' + \beta'' + \frac{1}{2}(\varepsilon + \varepsilon') \right) + \gamma + \frac{\eta}{2}, \\ \Sigma_{22} &= \frac{1}{n} \left(\frac{\alpha + \alpha'}{2} + \beta + \beta' + \beta'' + \delta + \varepsilon + \varepsilon' \right) + \gamma + \zeta + \eta.\end{aligned}$$

□

11.3 Proofs of Chapter 4

In this section, we prove the results of Chapter 4: $O(n)$ -invariant metrics.

11.3.1 Proof of Lemma 4.6

Lemma 4.6 (Characterization of inner products on symmetric matrices invariant under $\mathcal{D}^\pm(n)$ or $\mathfrak{S}^\pm(n)$) Let $\langle \cdot | \cdot \rangle : \text{Sym}(n) \times \text{Sym}(n) \rightarrow \mathbb{R}$ be an inner product on symmetric matrices.

- (a) It is $\mathcal{D}^\pm(n)$ -invariant if and only if there exist $\frac{n(n-1)}{2}$ positive real numbers $\alpha_{ij} = \alpha_{ji} > 0$ for $i \neq j$ and a matrix $S \in \text{Sym}^+(n)$ such that:

$$\forall X \in \text{Sym}(n), \langle X | X \rangle = \sum_{i \neq j} \alpha_{ij} X_{ij}^2 + \sum_{i,j} S_{ij} X_{ii} X_{jj}.$$

- (b) It is $\mathfrak{S}^\pm(n)$ -invariant if and only if there exist $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ with $\alpha > 0$, $\gamma > \beta$ and $\gamma + (n-1)\beta > 0$ such that:

$$\forall X \in \text{Sym}(n), \langle X | X \rangle = \gamma \sum_{i=1}^n X_{ii}^2 + \alpha \sum_{i \neq j} X_{ij}^2 + \beta \sum_{i \neq j} X_{ii} X_{jj}.$$

Proof of Lemma 4.6.

(a) We write $\langle X|X \rangle = \sum_{i,j,k,l} a_{ij,kl} X_{ij} X_{kl}$ a general inner product. Note that $a_{ij,kl} = a_{ji,kl} = a_{ji,lk} = a_{ij,lk}$ by symmetry of X and $a_{ij,kl} = a_{kl,ij}$ by symmetry of the inner product. We use the invariance under the matrix $\varepsilon_m \in \mathcal{D}^\pm(n)$ with -1 on the m -th component and 1 elsewhere, for $m \in \{1, \dots, n\}$. We denote $\mathcal{P} \text{ XOR } \mathcal{Q} = 1$ if the “exclusive or” between propositions \mathcal{P} and \mathcal{Q} holds, and otherwise $\mathcal{P} \text{ XOR } \mathcal{Q} = 0$. Thus, we have $[\varepsilon_m X \varepsilon_m]_{ij} = (-1)^{(i=m) \text{ XOR } (j=m)} X_{ij}$ and $[\varepsilon_m X \varepsilon_m]_{ij} [\varepsilon_m X \varepsilon_m]_{kl} = \theta_{ijklm} X_{ij} X_{kl}$ with $\theta_{ijklm} = (-1)^{[(i=m) \text{ XOR } (j=m)] \text{ XOR } [(k=m) \text{ XOR } (l=m)]} \in \{-1, 1\}$. Then the equality $\langle X|X \rangle = \langle \varepsilon_m X \varepsilon_m | \varepsilon_m X \varepsilon_m \rangle$ leads to $a_{ij,kl} = \theta_{ijklm} a_{ij,kl}$. Therefore, if there exists $m \in \{1, \dots, n\}$ such that $\theta_{ijklm} = -1$, then $a_{ij,kl} = 0$. One can easily show that $\theta_{ijklm} = -1$ if and only if m equals exactly one or exactly three index(es) among i, j, k, l . There exists such an m if:

- $\text{card}(\{i, j, k, l\}) = 4$, i.e. i, j, k, l are distinct,
- $\text{card}(\{i, j, k, l\}) = 3$,
- $\text{card}(\{i, j, k, l\}) = 2$ and three of them are equal.

Thus we are left with $\langle X|X \rangle = \sum_{i < j} 4a_{ij,ij} X_{ij}^2 + \sum_{i,j} a_{ii,jj} X_{ii} X_{jj}$. Then, we get the expression (4.4) by denoting $\alpha_{ij} = 2a_{ij,ij}$ and $S_{ij} = a_{ii,jj} = S_{ji}$. Since the quadratic form splits into two quadratic forms defined on supplementary vector spaces (off-diagonal and diagonal terms), it is positive definite if and only if these two quadratic forms are positive definite, i.e. $\alpha_{ij} > 0$ for all $i \neq j$ and S is positive definite. Conversely, Equation (4.4) clearly defines $\mathcal{D}^\pm(n)$ -invariant inner products.

(b) A $\mathfrak{S}^\pm(n)$ -invariant inner product on symmetric matrices is $\mathcal{D}^\pm(n)$ -invariant so it is of the form of Equation (4.4). Since it is invariant under permutations, we have $\alpha_{ij} = \alpha_{kl} =: \alpha$ and $S_{ij} = S_{kl} =: \beta$ for all $i \neq j$ and $k \neq l$ and $S_{ii} = S_{jj} =: \gamma$ for all i, j . Under these notations, Equation (4.4) becomes Equation (4.5). Since $S = (\gamma - \beta) I_n + \beta \mathbf{1}\mathbf{1}^\top$, then $S \in \text{Sym}^+(n)$ if and only if $\gamma - \beta > 0$ and $\gamma - \beta + n\beta > 0$ as expected. Conversely, Equation (4.5) clearly defines $\mathfrak{S}^\pm(n)$ -invariant inner products. □

11.3.2 Proof of affine-invariant curvature

We prove that the sectional curvature of the affine-invariant metric satisfies $\kappa \in [-1/2\alpha; 0]$. More precisely, the Riemann and sectional curvatures are:

- $R_\Sigma(X, Y, Z, T) = \frac{\alpha}{2} (X \Sigma^{-1} Y \Sigma^{-1} (Z \Sigma^{-1} T - T \Sigma^{-1} Z) \Sigma^{-1})$,
- $\kappa_\Sigma(\Sigma^{1/2} E_{ii}^\beta \Sigma^{1/2}, \Sigma^{1/2} E_{ij}^\beta \Sigma^{1/2}) = -1/4\alpha$ for $i \neq j$,
- $\kappa_\Sigma(\Sigma^{1/2} E_{ij}^\beta \Sigma^{1/2}, \Sigma^{1/2} E_{ik}^\beta \Sigma^{1/2}) = -1/8\alpha$ for $i \neq j \neq k \neq i$,

where $E_{ij}^\beta = E_{ij} - \frac{1-p}{np} \delta_{ij} I_n$. Other terms are null.

Proof of sectional curvature in Table 4.5. Firstly, we compute the sectional curvature of the affine-invariant metrics for $\beta = 0$ at $\Sigma \in \text{Sym}^+(n)$ in the orthonormal basis $(\Sigma^{1/2} E_{ij} \Sigma^{1/2})_{1 \leq i < j \leq n}$,

$\text{Id}_{\text{GL}(n)}$ on $\text{GL}(n)$:

$$\begin{aligned}
(\nabla_{X^h}^G Y^h)|_A &= \partial_{X_A^h}(Y^0 \circ \pi)A + Y_{\pi(A)}^0 X_A^h \\
&= (\partial_{X_{\pi(A)}} Y^0)A + Y_{\pi(A)}^0 X_{\pi(A)}^0 A, \\
(\nabla_X Y)|_{AA^\top} &= d_A \pi((\nabla_{X^h}^G Y^h)|_A) \\
&= AA^\top (\partial_{X_{\pi(A)}} Y^0) + (\partial_{X_{\pi(A)}} Y^0) AA^\top \\
&\quad + AA^\top X_{\pi(A)}^0 Y_{\pi(A)}^0 + Y_{\pi(A)}^0 X_{\pi(A)}^0 AA^\top, \\
(\partial_X Y)|_\Sigma &= \Sigma (\partial_{X_\Sigma} Y^0) + (\partial_{X_\Sigma} Y^0) \Sigma + X_\Sigma Y_\Sigma^0 + Y_\Sigma^0 X_\Sigma \\
&= \Sigma (\partial_{X_\Sigma} Y^0) + (\partial_{X_\Sigma} Y^0) \Sigma + \Sigma X_\Sigma^0 Y_\Sigma^0 + Y_\Sigma^0 X_\Sigma^0 \Sigma \\
&\quad + X_\Sigma^0 \Sigma Y_\Sigma^0 + Y_\Sigma^0 \Sigma X_\Sigma^0 \\
&= (\nabla_X Y)|_\Sigma + X_\Sigma^0 \Sigma Y_\Sigma^0 + Y_\Sigma^0 \Sigma X_\Sigma^0.
\end{aligned}$$

Finally, we find $\nabla_X Y = \partial_X Y - (X^0 \Sigma Y^0 + Y^0 \Sigma X^0)$ as expected. \square

11.3.4 Proof of Bures-Wasserstein curvature

We prove that the Riemann curvature of the Bures-Wasserstein metric is $R_\Sigma(X, Y, X, Y) = \frac{3}{2} \sum_{i,j} \frac{d_i d_j}{d_i + d_j} [X^{0'}, Y^{0'}]_{ij}^2$ for all $X, Y \in T_\Sigma \text{Sym}^+(n)$ at $\Sigma \in \text{Sym}^+(n)$, where $[V, W] = VW - WV$ is the Lie bracket of matrices.

Proof of curvature in Table 4.7. Let $X, Y \in T_\Sigma \text{Sym}^+(n)$ be tangent vectors at $\Sigma \in \text{Sym}^+(n)$. We would like to compute the sectional curvature $\kappa(X, Y) = \frac{R(X, Y, X, Y)}{\|X\|^2 \|Y\|^2 - \langle X, Y \rangle^2}$, i.e. $R(X, Y, X, Y)$. Let $X^h, Y^h \in \mathcal{H}_{\Sigma^{1/2}}$ be the horizontal lifts of X, Y at $\Sigma^{1/2}$ and $X^0, Y^0 \in \text{Sym}(n)$ defined as explained above. We extend X^h, Y^h into vector fields by $X_A^h := X^0 A$ and $Y_A^h := Y^0 A$. We do so because the formula we use to compute the curvature is based on a Lie bracket and can only be computed with fields. As the curvature is a tensor, it only depends on the values of X and Y at Σ so the way we extend the fields does not influence the result (but it simplifies the computation).

A first strategy to compute the curvature is to use the Levi-Civita connection via the definition $R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z$. It is tedious but doable. Another one consists in using the relation between the curvatures of the quotient metric (here, Bures-Wasserstein) and the original metric (here, Frobenius) found in [O'Neill, 1966], formula {4}. According to this formula, since the Euclidean metric is flat, the formula is $R_\Sigma(X, Y, X, Y) = \frac{3}{4} \|\text{ver}[[X^h, Y^h]]_{\Sigma^{1/2}}\|^2$ where $\text{ver} : X^v + X^h \in T\text{GL}(n) \mapsto X^v \in \mathcal{V}$ is the vertical projection and $[[\cdot, \cdot]]$ denotes the Lie bracket on vector fields of $\text{GL}(n)$, which must be distinguished from the matrix Lie bracket $[V, W] = VW - WV$. Note that the right term only depends on $X_{\Sigma^{1/2}}^h$ and $Y_{\Sigma^{1/2}}^h$ because if $f : \text{GL}^+(n) \rightarrow \mathbb{R}$ is a map, then $\text{ver}[[fX^h, Y^h]]_{\Sigma^{1/2}} = f(\Sigma^{1/2}) \text{ver}[[X^h, Y^h]]_{\Sigma^{1/2}} + d_{\Sigma^{1/2}} f(Y^h) \underbrace{\text{ver}(X^h)}_0$.

The rest of the proof consists in computing $\text{ver}[[X^h, Y^h]] = [[X^h, Y^h]] - \text{hor}[[X^h, Y^h]]$. On the one hand, $[[X^h, Y^h]]_A = Y^0 X_A^h - X^0 Y_A^h = -[X^0, Y^0]A$. On the other hand, let $Z_A^h := \text{hor}[[X^h, Y^h]]_A =: Z_{AA^\top}^0 A \in \mathcal{H}_A$. Now, we can fix $\Sigma \in \text{Sym}^+(n)$ and $A = \Sigma^{1/2}$. We take a spectral decomposition $\Sigma = PDP^\top$ and we denote with a prime all the previous

matrices taken in the basis P of eigenvectors of Σ , e.g. $X^{0'} = P^\top X^0 P$. Then:

$$\begin{aligned}
Z_\Sigma &:= d_{\Sigma^{1/2}} \pi(\llbracket X^h, Y^h \rrbracket_{\Sigma^{1/2}}) = \Sigma[X^0, Y^0] - [X^0, Y^0]\Sigma, \\
Z_{\Sigma^{1/2}}^h &= (d_{\Sigma^{1/2}} \pi|_{\mathcal{H}_{\Sigma^{1/2}}})^{-1}(d_{\Sigma^{1/2}} \pi(\llbracket X^h, Y^h \rrbracket_{\Sigma^{1/2}})) \\
&= (d_{\Sigma^{1/2}} \pi|_{\mathcal{H}_{\Sigma^{1/2}}})^{-1}(Z_\Sigma), \\
[Z_\Sigma^{0'}]_{ij} &= \frac{1}{d_i + d_j} (D[X^{0'}, Y^{0'}] - [X^{0'}, Y^{0'}]D)_{ij} \\
&= \frac{d_i - d_j}{d_i + d_j} [X^{0'}, Y^{0'}]_{ij}, \\
[Z_{\Sigma^{1/2}}^{h'}]_{ij} &= \sqrt{d_j} [Z_\Sigma^{0'}]_{ij} = \sqrt{d_j} \frac{d_i - d_j}{d_i + d_j} [X^{0'}, Y^{0'}]_{ij}, \\
(\text{ver} \llbracket X^h, Y^h \rrbracket_{\Sigma^{1/2}})'_{ij} &= (\llbracket X^h, Y^h \rrbracket_{\Sigma^{1/2}})'_{ij} - [Z_{\Sigma^{1/2}}^{h'}]_{ij} \\
&= -[X^{0'}, Y^{0'}]_{ij} \sqrt{d_j} - \sqrt{d_j} \frac{d_i - d_j}{d_i + d_j} [X^{0'}, Y^{0'}]_{ij} \\
&= -\frac{2d_i \sqrt{d_j}}{d_i + d_j} [X^{0'}, Y^{0'}]_{ij}, \\
R_\Sigma(X, Y, X, Y) &= \frac{3}{4} \|(\text{ver} \llbracket X^h, Y^h \rrbracket_{\Sigma^{1/2}})'\|^2 \\
&= 3 \sum_{i,j} \frac{d_i^2 d_j}{(d_i + d_j)^2} [X^{0'}, Y^{0'}]_{ij}^2 \\
&= \frac{3}{2} \sum_{i,j} \frac{d_i d_j}{d_i + d_j} [X^{0'}, Y^{0'}]_{ij}^2,
\end{aligned}$$

where $P^\top X P = D X^{0'} + X^{0'} D$ and $P^\top Y P = D Y^{0'} + Y^{0'} D$. □

11.3.5 Proof of Bures-Wasserstein geodesic parallel transport between commuting matrices

We prove that along a geodesic between commuting matrices $\Sigma = P D P^\top$ and $\Lambda = P \Delta P^\top \in \text{Sym}^+(n)$, the parallel transport of the Bures-Wasserstein metric writes:

$$\Pi_{\Sigma \rightarrow \Lambda} : \begin{cases} T_\Sigma \text{Sym}^+(n) & \longrightarrow & T_\Lambda \text{Sym}^+(n) \\ X & \longmapsto & P \left[\sqrt{\frac{\delta_i + \delta_j}{d_i + d_j}} [P^\top X P]_{ij} \right] P^\top \end{cases} \cdot$$

Proof of geodesic parallel transport between commuting matrices in Table 4.7. We want to prove that the geodesic parallel transport of the Bures-Wasserstein metric between two commuting matrices is $\Pi_{\Sigma \rightarrow \Lambda} X = P \left[\sqrt{\frac{\delta_i + \delta_j}{d_i + d_j}} [P^\top X P]_{ij} \right]_{i,j} P^\top$ where $\Sigma = P D P^\top$ and $\Lambda = P \Delta P^\top \in \text{Sym}^+(n)$. The geodesic parallel transport is $O(n)$ -invariant so we only need to prove that $[\Pi_{D \rightarrow \Delta} X]_{ij} = \sqrt{\frac{\delta_i + \delta_j}{d_i + d_j}} X_{ij}$. The geodesic from D to Δ is $\gamma(t) = ((1-t)\sqrt{D} + t\sqrt{\Delta})^2$. Let us define $X(t) = \left[\sqrt{\frac{((1-t)d_i + t\delta_i)^2 + ((1-t)d_j + t\delta_j)^2}{d_i + d_j}} X_{ij} \right]_{i,j}$ and let us check that $\nabla_{\dot{\gamma}} X = 0$. We

compute:

$$\begin{aligned}
[X^0(t)]_{ij} &= \frac{[X(t)]_{ij}}{\gamma_i(t) + \gamma_j(t)} = \frac{1}{\sqrt{d_i + d_j} \sqrt{((1-t)d_i + t\delta_i)^2 + ((1-t)d_j + t\delta_j)^2}} X_{ij}, \\
\dot{\gamma}(t) &= 2(\sqrt{\Delta} - \sqrt{D})((1-t)\sqrt{D} + t\sqrt{\Delta}), \\
\dot{\gamma}^0(t) &= \frac{1}{2}\dot{\gamma}(t)\gamma^{-1}(t) = \frac{1}{2}\gamma^{-1}(t)\dot{\gamma}(t) = (\sqrt{\Delta} - \sqrt{D})((1-t)\sqrt{D} + t\sqrt{\Delta})^{-1}, \\
[\dot{X}(t)]_{ij} &= \frac{2(\sqrt{\delta_i} - \sqrt{d_i})((1-t)d_i + t\delta_i) + 2(\sqrt{\delta_j} - \sqrt{d_j})((1-t)d_j + t\delta_j)}{2\sqrt{d_i + d_j} \sqrt{((1-t)d_i + t\delta_i)^2 + ((1-t)d_j + t\delta_j)^2}} X_{ij} \\
&= (\sqrt{\delta_i} - \sqrt{d_i})((1-t)d_i + t\delta_i)[X^0(t)]_{ij} \\
&\quad + [X^0(t)]_{ij}(\sqrt{\delta_j} - \sqrt{d_j})((1-t)d_j + t\delta_j) \\
&= [\dot{\gamma}^0(t)\gamma(t)X^0(t) + X^0(t)\gamma(t)\dot{\gamma}^0(t)]_{ij}, \\
\nabla_{\dot{\gamma}(t)}X &= \dot{X}(t) - (\dot{\gamma}^0(t)\gamma(t)X^0(t) + X^0(t)\gamma(t)\dot{\gamma}^0(t)) = 0.
\end{aligned}$$

So the geodesic parallel transport from $\Sigma = PDP^\top$ to $\Lambda = P\Delta P^\top$ is $\Pi_{\Sigma \rightarrow \Lambda}X = P \left[\sqrt{\frac{\delta_i + \delta_j}{d_i + d_j}} [P^\top X P]_{ij} \right]_{i,j} P^\top$. \square

11.3.6 Proof of Proposition 4.12

Proposition 4.12 (Parallel transport equation of Bures-Wasserstein metric) Let $\gamma(t)$ the geodesic between $\gamma(0) = \Sigma$ and $\gamma(1) = \Lambda$, and a vector $X \in T_\Sigma \text{Sym}^+(n)$. We denote $\gamma^h(t) = (1-t)\Sigma^{1/2} + t\Sigma^{-1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}$ the horizontal lift of the geodesic γ . The two following statements are equivalent.

- (i) The vector field $X(t)$ defined along $\gamma(t)$ is the parallel transport of X .
- (ii) $X(t) = \gamma(t)X^0(t) + X^0(t)\gamma(t)$ where $X^0(t)$ is a curve in $\text{Sym}(n)$ satisfying the following ODE:

$$\gamma(t)\dot{X}^0(t) + \dot{X}^0(t)\gamma(t) + \gamma^h(t)\dot{\gamma}^{h\top}X^0(t) + X^0(t)\dot{\gamma}^h\gamma^h(t)^\top = 0.$$

Proof of Proposition 4.12 (Parallel transport equation of Bures-Wasserstein metric). The geodesic parallel transport equation is $\nabla_{\dot{\gamma}(t)}X = 0$ along the geodesic $\gamma(t) = \gamma^h(t)\gamma^h(t)^\top$ between Σ and $\Lambda \in \text{Sym}^+(n)$, where $\gamma^h(t) = (1-t)\Sigma^{1/2} + t\Sigma^{-1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}$. For a vector field $X(t)$ on $\text{Sym}^+(n)$ defined along $\gamma(t)$, we can define the horizontal lift $X^h(t) = X^0(t)\gamma^h(t) \in \mathcal{H}_{\gamma^h(t)}$ where $X^0(t)$ is defined by $X(t) = \gamma(t)X^0(t) + X^0(t)\gamma(t)$. We are going to prove that $X(t)$ is the geodesic parallel transport of $X \in T_\Sigma \text{Sym}^+(n)$ if and only if $X^0(t)$ satisfies the following ODE:

$$\gamma(t)\dot{X}^0(t) + \dot{X}^0(t)\gamma(t) + \gamma^h(t)\dot{\gamma}^{h\top}X^0(t) + X^0(t)\dot{\gamma}^h\gamma^h(t)^\top = 0. \quad (11.21)$$

To rewrite the geodesic parallel transport equation $\nabla_{\dot{\gamma}(t)}X = 0$, we need to compute the following derivatives:

$$\begin{aligned}
\dot{X}(t) &= \gamma(t)\dot{X}^0(t) + \dot{X}^0(t)\gamma(t) + \dot{\gamma}(t)X^0(t) + X^0(t)\dot{\gamma}(t), \\
\dot{\gamma}(t) &= \dot{\gamma}^h\gamma^h(t)^\top + \gamma^h(t)\dot{\gamma}^{h\top} \text{ where } \dot{\gamma}^h = \dot{\gamma}^0(t)\gamma^h(t).
\end{aligned}$$

Now, we simply rewrite the equation:

$$\begin{aligned}
\nabla_{\dot{\gamma}(t)} X &= 0 \iff \dot{X}(t) - (\dot{\gamma}^0(t)\gamma(t)X^0(t) + X^0(t)\gamma(t)\dot{\gamma}^0(t)) = 0 \\
&\iff \gamma(t)\dot{X}^0(t) + \dot{X}^0(t)\gamma(t) \\
&\quad + (\dot{\gamma}(t) - \dot{\gamma}^h\gamma^h(t)^\top)X^0(t) + X^0(t)(\dot{\gamma}(t) - \gamma^h(t)\dot{\gamma}^h)^\top = 0 \\
&\iff \gamma(t)\dot{X}^0(t) + \dot{X}^0(t)\gamma(t) + \gamma^h(t)\dot{\gamma}^h X^0(t) + X^0(t)\dot{\gamma}^h\gamma^h(t)^\top = 0.
\end{aligned}$$

□

11.4 Proofs of Chapter 5

In this section, we prove the results of Chapter 5: Geometry of Mixed-Euclidean metrics.

11.4.1 Proof of Theorem 5.9

Theorem 5.9 (Sufficient condition for power-Wasserstein to be mean kernel) The power-Wasserstein metric of parameter $p \leq 1$ is a mean kernel metric.

Proof of Theorem 5.9 (Sufficient condition for power-Wasserstein to be mean kernel). Let us show that if $p \leq 1$, then the function $m(x, y) = \left[(x^p + y^p) \left(p \frac{x-y}{x^p - y^p} \right)^2 \right]^{\frac{1}{2-p}}$ is non-decreasing in x (and by symmetry in y). If we factorize by y and define a new variable $t = x/y$, we have to study the function $f(t) = F(t)^{\frac{1}{2-p}}$ where $F(t) = (t^p + 1) \left(p \frac{t-1}{t^p-1} \right)^2$.

First, let us prove that F is non-decreasing if and only if $p \leq 1$. We denote $f_0(x) = p \frac{x-1}{x^p-1} > 0$ and $g_0(x) = x^p f_0(x)^2$ so that $F(x) = f_0(x)^2 + g_0(x)$. Note that f_0 is non-decreasing if and only if $p \leq 1$. We also introduce $h_0(x) = (x^p + 1)(x - 1)^2$ so that $F(x) = p^2 \frac{h_0(x)}{(x^p-1)^2}$.

1. If $p - 1 > 0$, then $F'(0) = -2p^2 < 0$ so F' cannot be positive around 0 because it is smooth at 0. So F is not non-decreasing.
2. If $p \in]0, 1]$, then F is non-decreasing as product of three non-decreasing positive functions.
3. We assume that $p < 0$. Let us prove separately that F is increasing on $(0, 1)$ and on $(1, \infty)$. As F is continuous (at 1), it will prove that F is increasing on $(0, \infty)$.

(a) Let us prove that F is increasing on $(0, 1)$. We only need to prove that g_0 is increasing on $(0, 1)$. We successively derive:

$$\begin{aligned}
\cdot g'_0(x) &= p^2 \underbrace{\frac{x-1}{(x^p-1)^3} x^{p-1}}_{<0} \underbrace{((2-p)x^{p+1} + px^p - (p+2)x + p)}_{g_1(x)}, \\
\cdot g'_1(x) &= (2-p)(p+1)x^p + p^2 x^{p-1} - (p+2), \\
\cdot g''_1(x) &= px^{p-2}((2-p)(p+1)x + p(p-1)).
\end{aligned}$$

We can notice that $g_1(1) = 0$ and $g'_1(1) = 0$. So if we prove that $g''_1 < 0$ on $(0, 1)$, then g'_1 is decreasing thus positive, so g_1 is increasing thus negative, and finally g'_0 is positive so g_0 increases.

- i. If $p + 1 = 0$, then g''_1 has the sign of $p^2(p - 1) < 0$,
- ii. if $p + 1 > 0$, then g''_1 is positive before $x_0 := \frac{p(p-1)}{(p+1)(p-2)} < 0$ and negative after,
- iii. if $p + 1 < 0$, then g''_1 is negative before $x_0 := \frac{p(p-1)}{(p+1)(p-2)} > 1$ and positive after.

So we proved that F is increasing on $(0, 1)$.

- (b) Let us prove that F is increasing on $(1, \infty)$. As $x \mapsto \frac{1}{(x^p-1)^2}$ is increasing, we only need to prove that h_0 is increasing on $(1, \infty)$ so that F is increasing on $(1, \infty)$ as product of two positive increasing functions. We derive successively:

- $h'_0(x) = (x - 1)h_1(x)$ with $h_1(x) = (p + 2)x^p - px^{p-1} + 2$,
- $h'_1(x) = px^{p-2}((p + 2)x - (p - 1))$.

We need to prove that $h_1 > 0$ on $(1, \infty)$. As before, we distinguish the cases:

- i. If $p + 2 < 0$, then h'_1 is negative before $x_0 := \frac{p-1}{p+2} > 1$ and positive after. As $h_1(x_0) = 2 - x_0^{p-1} > 1$, we have $h_1 > 0$ on $(1, \infty)$.
- ii. If $p + 2 \geq 0$, then h'_1 is negative on $(1, \infty)$. As $\lim_{x \rightarrow \infty} h_1(x) = 2$, we have $h_1 > 0$ on $(1, \infty)$.

So we proved that F is increasing on $(1, \infty)$ and therefore on $(0, \infty)$.

Finally, we proved that F is non-decreasing if and only if $p \leq 1$. As $f = \text{pow}_{\frac{1}{2-p}} \circ F$, we can assert that if $p \leq 1$, then f is non-decreasing, as expected. \square

11.4.2 Proof of Theorem 5.19

Theorem 5.19 (Sufficient condition for a balanced bilinear form to be a metric) Let $\langle \cdot | \cdot \rangle = \text{Frob}$ be the Frobenius inner product. Let g^+, g^- be deformed-Euclidean metrics respectively associated to univariate diffeomorphisms u and v . Then the balanced bilinear form g^0 is a metric.

Proof of Theorem 5.19 (Sufficient condition for a balanced bilinear form to be a metric). Let $u, v : \text{Sym}^+(n) \rightarrow \text{Sym}(n)$ be two univariate diffeomorphisms onto the respective image of $\text{Sym}^+(n)$ by u and v . Let g^+, g^- be the respective deformed-Euclidean metrics by u and v . Hence for all $\Sigma \in \text{Sym}^+(n)$ and $X, Y \in T_\Sigma \text{Sym}^+(n)$:

$$\begin{aligned} g_\Sigma^+(X, X) &= \text{tr}(d_\Sigma u(X)^2), \\ g_\Sigma^-(Y, Y) &= \text{tr}(d_\Sigma v(Y)^2). \end{aligned}$$

Hence the flat parallel transports Π^+ and Π^- do not depend on the curve, they are simply given by the differentials of u and v :

$$\begin{aligned} \Pi_{\Sigma \rightarrow \Lambda}^+ X &= (d_\Lambda u)^{-1}(d_\Sigma u(X)), \\ \Pi_{\Sigma \rightarrow \Lambda}^- Y &= (d_\Lambda v)^{-1}(d_\Sigma v(Y)), \end{aligned}$$

where $d_\Sigma u(X) = P d_D u(P^\top X P) P^\top$ and $[d_D u(P^\top X P)]_{ij} = u^{[1]}(d_i, d_j)[P^\top X P]_{ij}$ given an eigenvalue decomposition of $\Sigma = P D P^\top$ with $D = \text{diag}(d_1, \dots, d_n) \in \text{Diag}^+(n)$ and $P \in O(n)$. Note that $d_{I_n} u(X) = u'(1)X$ so $d_{I_n} u = u'(1)\text{Id}$. The same is valid for v . Therefore, since the parallel transport is $O(n)$ -equivariant, the balanced bilinear form g^0 is defined by:

$$\begin{aligned} g_\Sigma^0(X, Y) &= \text{tr}((\Pi_{\Sigma \rightarrow I_n}^+ X)(\Pi_{\Sigma \rightarrow I_n}^- Y)) \\ &= \text{tr}(P(\Pi_{D \rightarrow I_n}^+ P^\top X P)P^\top P(\Pi_{D \rightarrow I_n}^- P^\top Y P)P^\top) \\ &= \frac{1}{u'(1)v'(1)} \text{tr}(d_D u(P^\top X P)d_D v(P^\top Y P)) \\ &= \frac{1}{u'(1)v'(1)} \sum_{i,j} u^{[1]}(d_i, d_j)v^{[1]}(d_i, d_j)[P^\top X P]_{ij}[P^\top Y P]_{ij}. \end{aligned}$$

First, g^0 is symmetric. Second, since $u : (0, \infty) \rightarrow \mathbb{R}$ is a diffeomorphism, either $u' > 0$ or $u' < 0$ and by the mean value theorem, the sign of $u^{[1]}$ is the sign of u' . Hence $\frac{u^{[1]}(d_i, d_j)}{u'(1)} > 0$ and similarly $\frac{v^{[1]}(d_i, d_j)}{v'(1)} > 0$ so the coefficients of the quadratic form $g_\Sigma^0(X, X)$ are positive. So the balanced bilinear form g^0 is a Riemannian metric. \square

11.4.3 Proof of Theorem 5.25

Theorem 5.25 (Curvature of Mixed-Euclidean metrics) Let $u, v : \text{Sym}^+(n) \rightarrow \text{Sym}^+(n)$ be two univariate diffeomorphisms. We define the univariate diffeomorphism $w = v \circ u^{-1}$ so that $u : (\text{Sym}^+(n), g^{\text{ME}(u,v)}) \rightarrow (\text{Sym}^+(n), \frac{w'(1)}{u'(1)v'(1)} g^{\text{ME}(\text{Id}, w)})$ is an isometry. For $\Sigma = P D P^\top \in \text{Sym}^+(n)$, we denote $X = P X' P^\top \in T_\Sigma \text{Sym}^+(n)$ and analogously for $Y, Z, T \in T_\Sigma \text{Sym}^+(n)$, we denote $u_{ij} = u^{[1]}(d_i, d_j)$, $u_{ijk} = u^{[2]}(d_i, d_j, d_k)$ and analogously for v, w . We denote $m_{ij} = w^{[1]}(u(d_i), u(d_j)) = \frac{v_{ij}}{u_{ij}}$ and $m_{ijk} = w^{[2]}(u(d_i), u(d_j), u(d_k))$. Then the curvature of the mixed-Euclidean metric $g^{\text{ME}(u,v)}$ is:

$$\begin{aligned} R_\Sigma^{\text{ME}(u,v)}(X, Y, Z, T) &= \frac{1}{u'(1)v'(1)} \sum_{i,j,k,l} \rho_{ijkl} (X'_{ij} Y'_{jk} Z'_{kl} T'_{li} - Y'_{ij} X'_{jk} Z'_{kl} T'_{li} \\ &\quad + X'_{ij} Z'_{jk} Y'_{kl} T'_{li} - Y'_{ij} Z'_{jk} X'_{kl} T'_{li}), \end{aligned}$$

where $\rho_{ijkl} = \frac{m_{ijl} m_{jlk}}{2m_{jl}} u_{ij} u_{jk} u_{kl} u_{li} = \frac{1}{2u_{ij} v_{jl}} (u_{ij} v_{ijl} - v_{ij} u_{ijl})(u_{jk} v_{jkl} - v_{jk} u_{jkl})$ is symmetric in $i \leftrightarrow k$, in $j \leftrightarrow l$ and in $u \leftrightarrow v$. In particular, at $\Sigma = I_n$, the curvature is:

$$R_{I_n}^{\text{ME}(u,v)}(X, Y, Z, T) = \frac{1}{4} \left[\left(\ln \left| \frac{v'}{u'} \right| \right)' (1) \right]^2 R_{I_n}^A(X, Y, Z, T),$$

where A stands for the affine-invariant metric (Formula 5.8). Therefore, the sectional curvature of the mixed-Euclidean metric at I_n takes non-positive values. In particular, for mixed-power-Euclidean metrics $\text{MPE}(\alpha, \beta)$ with $\alpha^2 \neq \beta^2$ (thus excluding log-Euclidean, power-Euclidean and power-affine metrics), since $\kappa_{\lambda\Sigma}^{\text{MPE}(\alpha, \beta)}(X, Y) = \lambda^{-(\alpha+\beta)} \times \kappa_\Sigma^{\text{MPE}(\alpha, \beta)}(X, Y)$ for all $\lambda > 0$, the lower bound of the sectional curvature is $-\infty$.

Proof of Theorem 5.25 (Curvature of Mixed-Euclidean metrics). We compute the curvature of the metric $g^{\text{ME}(u,v)}$ for univariate diffeomorphisms $u, v : \text{Sym}^+(n) \rightarrow \text{Sym}^+(n)$. Since

$d_\Sigma v(X) = d_{u(\Sigma)} w(d_\Sigma u(X))$ with $w = v \circ u^{-1}$, the map u is an isometry between $g^{\text{ME}(u,v)}$ and $c g^{\text{ME}(\text{Id},w)}$ with $c = \frac{w'(1)}{u'(1)v'(1)}$. So it suffices to compute the curvature of $g^{\text{ME}(\text{Id},w)}$ and to conclude by pullback and scaling.

Let $\Sigma = PDP^\top \in \text{Sym}^+(n)$. We denote $u_{ij} = u^{[1]}(d_i, d_j)$, $u_{ijk} = u^{[2]}(d_i, d_j, d_k)$ and analogously for v and w .

The curvature of $g := g^{\text{ME}(\text{Id},w)}$ can be computed the same way as shown in [Michor et al., 2000] for the metric $\text{BKM} = \text{MPE}(1, 0) = \text{ME}(\text{Id}, \log)$. Following [Michor et al., 2000], we introduce $G_\Sigma(X) = d_\Sigma w(X)$ and $\Gamma_\Sigma(X, Y)$ such that:

$$g_\Sigma(X, Y) = \frac{1}{w'(1)} \text{tr}(d_\Sigma w(X)Y) = \frac{1}{w'(1)} \text{tr}(G_\Sigma(X)Y),$$

$$\nabla_{X_\Sigma} Y = d_\Sigma Y(X) + \Gamma_\Sigma(X, Y),$$

where ∇ is the Levi-Civita connection of g . Note that $G_\Sigma : \text{Sym}(n) \rightarrow \text{Sym}(n)$ is a linear isomorphism and Γ_Σ is symmetric. According to Lemma 5.3, $[G_D(X)]_{ij} = w_{ij} X_{ij}$. Then the Riemann curvature tensors are defined by $R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z$ and $R(X, Y, Z, T) = -g(R(X, Y)Z, T)$ so we can write R in function of Γ and $d\Gamma$ or in function of G and dG [Michor et al., 2000]:

$$\begin{aligned} R(X, Y)Z &= d\Gamma(X)(Y, Z) - d\Gamma(Y)(X, Z) + \Gamma(X, \Gamma(Y, Z)) - \Gamma(Y, \Gamma(X, Z)) \\ &= -\frac{1}{4} G^{-1}(dG(X)(G^{-1}(dG(Y)(Z)))) + \frac{1}{4} G^{-1}(dG(Y)(G^{-1}(dG(X)(Z))))), \\ R(X, Y, Z, T) &= \frac{1}{4w'(1)} \text{tr} [(dG(X)(G^{-1}(dG(Y)(Z))) - dG(Y)(G^{-1}(dG(X)(Z)))) T]. \end{aligned}$$

So we only need to express $d_\Sigma G(X)(Y) = H_\Sigma w(X, Y)$. Lemma 5.3 gives $H_\Sigma w(X, Y) = P H_D w(X', Y') P^\top$ and $[H_D(X', Y')]_{ij} = \sum_k w_{ijk} (X'_{ik} Y'_{jk} + X'_{jk} Y'_{ik})$. Hence:

$$\begin{aligned} \text{tr}(d_\Sigma G(X)(G_\Sigma^{-1}(d_\Sigma G(Y)(Z)))T) &= \text{tr}(d_D G(X')(G_D^{-1}(d_D G(Y')(Z'))T') \\ &= \sum_{i,j} [d_D G(X')(G_D^{-1}(d_D G(Y')(Z')))]_{ij} T'_{ij} \\ &= \sum_{i,j,k} w_{ijk} (X'_{ik} [G_D^{-1}(d_D G(Y')(Z'))]_{jk} + X'_{jk} [G_D^{-1}(d_D G(Y')(Z'))]_{ik}) T'_{ij} \\ &= 2 \sum_{i,j,k} w_{ijk} X'_{ik} [G_D^{-1}(d_D G(Y')(Z'))]_{jk} T'_{ij} \\ &= 2 \sum_{i,j,k} \frac{w_{ijk}}{w_{jk}} X'_{ik} [d_D G(Y')(Z')]_{jk} T'_{ij} \\ &= 2 \sum_{i,j,k,l} \frac{w_{ijk} w_{jkl}}{w_{jk}} X'_{ik} (Y'_{jl} Z'_{kl} + Y'_{kl} Z'_{jl}) T'_{ij}. \end{aligned}$$

Therefore:

$$\begin{aligned} R(X, Y, Z, T) &= \frac{1}{w'(1)} \sum_{i,j,k,l} \frac{w_{ijk} w_{jkl}}{2w_{jk}} (X'_{ik} Y'_{kl} Z'_{lj} T'_{ji} + X'_{ik} Z'_{kl} Y'_{lj} T'_{ji} \\ &\quad - Y'_{ik} X'_{kl} Z'_{lj} T'_{ji} - Y'_{ik} Z'_{kl} X'_{lj} T'_{ji}). \end{aligned}$$

To get the curvature of $g^{\text{ME}(u,v)}$, we scale this formula by $c = \frac{w'(1)}{u'(1)v'(1)}$ and we pull it back via the map u . The coefficients $w_{ij} = w^{[1]}(d_i, d_j)$ and $w_{ijk} = w^{[2]}(d_i, d_j, d_k)$ are replaced by the coefficients $m_{ij} = w^{[1]}(u(d_i), u(d_j)) = v_{ij}/u_{ij}$ and $m_{ijk} = w^{[2]}(u(d_i), u(d_j), u(d_k))$. The vectors $X = [X_{ij}]_{i,j}$ are replaced by the vectors $d_\Sigma u(X) = [u_{ij}X_{ij}]_{i,j}$. Hence the curvature of $\text{ME}(u, v)$ writes (modulo a permutation of indexes $l \rightarrow k \rightarrow j \rightarrow l$):

$$R_\Sigma^{\text{ME}(u,v)}(X, Y, Z, T) = \frac{1}{u'(1)v'(1)} \sum_{i,j,k,l} \rho_{ijkl} (X'_{ij}Y'_{jk}Z'_{kl}T'_{li} + X'_{ij}Z'_{jk}Y'_{kl}T'_{li} - Y'_{ij}X'_{jk}Z'_{kl}T'_{li} - Y'_{ij}Z'_{jk}X'_{kl}T'_{li}), \quad (11.22)$$

where $\rho_{ijkl} = \frac{m_{ijl}m_{jlk}}{2m_{jl}} u_{ij}u_{jk}u_{kl}u_{li}$. This expression is symmetric in $i \leftrightarrow k$ and $j \leftrightarrow l$ but it does not look really symmetric in $u \leftrightarrow v$. Let us check that it is though. If $d_j \neq d_l$, we can write:

$$\begin{aligned} \rho_{ijkl} &= \frac{1}{2} \frac{u_{jl}}{v_{jl}} \frac{1}{u(d_j) - u(d_l)} (m_{ij} - m_{il}) \frac{1}{u(d_j) - u(d_l)} (m_{kj} - m_{kl}) u_{ij}u_{jk}u_{kl}u_{li} \\ &= \frac{1}{2(d_j - d_l)^2} \frac{1}{u_{jl}v_{jl}} (v_{ij}u_{il} - u_{ij}v_{il})(v_{kj}u_{kl} - u_{kj}v_{kl}) \\ &= \frac{(v_{ij}(u_{il} - u_{ij}) - u_{ij}(v_{il} - v_{ij}))(v_{kj}(u_{kl} - u_{kj}) - u_{kj}(v_{kl} - v_{kj}))}{2(d_j - d_l)^2 u_{jl}v_{jl}} \\ &= \frac{1}{2u_{jl}v_{jl}} (u_{ij}v_{ijl} - v_{ij}u_{ijl})(u_{jk}v_{jkl} - v_{jk}u_{jkl}). \end{aligned}$$

Since the two expressions of ρ_{ijkl} are continuous in (d_i, d_j, d_k, d_l) , they also coincide when $d_j = d_l$. The last expression is clearly symmetric in $u \leftrightarrow v$.

The curvature at $\Sigma = D = I_n$ follows from the following computations:

$$\begin{aligned}
u_{ij} &= u^{[1]}(d_i, d_j) = u'(1), \\
m_{ij} &= \frac{v_{ij}}{u_{ij}} = \frac{v'(1)}{u'(1)}, \\
m_{ijk} &= w^{[2]}(u(d_i), u(d_j), u(d_k)) = \frac{1}{2}w''(u(1)), \\
w' &= (v' \circ u^{-1}) \times (u^{-1})' = \frac{v \circ u^{-1}}{u' \circ u^{-1}}, \\
w' \circ u &= \frac{v'}{u'}, \\
(w'' \circ u) \times u' &= \frac{v''u' - v'u''}{u'^2}, \\
m_{ijk} &= \left(\frac{v''u' - v'u''}{2u'^3} \right) (1), \\
\rho_{ijkl} &= \frac{m_{ijl}m_{jlk}}{2m_{jl}} u_{ij}u_{jk}u_{kl}u_{li} \\
&= \frac{u'(1)}{2v'(1)} \left(\frac{v''(1)u'(1) - u'(1)v''(1)}{2u'(1)^3} \right)^2 u'(1)^4 \\
&= \frac{1}{8u'(1)v'(1)} (v''(1)u'(1) - v'(1)u''(1))^2 \\
&= \frac{1}{8u'(1)v'(1)} u'(1)^2 v'(1)^2 \left(\frac{v''(1)}{v'(1)} - \frac{u''(1)}{u'(1)} \right)^2 \\
&= \frac{1}{8} u'(1)v'(1) ((\ln |v'|)' - (\ln |u'|)')^2 (1) \\
&= \frac{1}{8} u'(1)v'(1) \left[\left(\ln \left| \frac{v'}{u'} \right| \right)' (1) \right]^2.
\end{aligned}$$

Hence, according to Formula (11.22), the curvature at I_n writes:

$$\begin{aligned}
R_{I_n}^{\text{ME}(u,v)}(X, Y, Z, T) &= \frac{1}{8} \left[\left(\ln \left| \frac{v'}{u'} \right| \right)' (1) \right]^2 \text{tr}(XYZT + XZYT - YXZT - YZXT) \\
&= \frac{1}{8} \left[\left(\ln \left| \frac{v'}{u'} \right| \right)' (1) \right]^2 \text{tr}(XYZT - YXZT), \tag{11.23}
\end{aligned}$$

because the second and fourth terms cancel. Recognizing the curvature of the affine-invariant $R_{I_n}^A(X, Y, Z, T) = \frac{1}{2}\text{tr}(XYZT - YXZT)$ metric [Skovgaard, 1984, Pennec et al., 2020, Thanwerdas and Pennec, 2022b], we can finally write:

$$R_{I_n}^{\text{ME}(u,v)}(X, Y, Z, T) = \frac{1}{4} \left[\left(\ln \left| \frac{v'}{u'} \right| \right)' (1) \right]^2 R_{I_n}^A(X, Y, Z, T). \tag{11.24}$$

□

11.4.4 Proof of Theorem 5.26

Theorem 5.26 (Riemannian operations of MPE metrics) Let $\alpha, \beta \in \mathbb{R}$ such that $\alpha + \beta \neq 0$, thus excluding log-Euclidean and power-affine metrics. Table 5.3 summarizes the formulae of the geodesics, the logarithm map and the distance in the particular case where $\Sigma, \Lambda \in \text{Sym}^+(n)$ and $V \in T_\Sigma \text{Sym}^+(n)$ **commute**. They essentially reduce to the formulae of the α_0 -power-Euclidean metric with $\alpha_0 = \frac{\alpha + \beta}{2}$. These formulae are generally not valid for non-commuting matrices.

Geodesics	$\gamma_{(\Sigma, V)}(t) = (\Sigma^{\alpha_0} + t d_\Sigma \text{pow}_{\alpha_0}(V))^{1/\alpha_0}$
Logarithm	$\text{Log}_\Sigma(\Lambda) = (d_\Sigma \text{pow}_{\alpha_0})^{-1}(\Lambda^{\alpha_0} - \Sigma^{\alpha_0})$
Distance	$d(\Sigma, \Lambda) = \frac{1}{\alpha_0} \ \Lambda^{\alpha_0} - \Sigma^{\alpha_0}\ _{\text{Frob}}$

Table 5.3: Riemannian operations of Mixed-Power-Euclidean metrics for commuting matrices

Proof of Theorem 5.26 (Riemannian operations of MPE metrics). We compute the geodesics, the logarithm map and the distance between commuting matrices. We show that the geodesics of the Mixed-Power-Euclidean metrics $\text{MPE}(\alpha, \beta)$ with $\alpha + \beta \neq 0$ when the base point $\Sigma \in \text{Sym}^+(n)$ and the initial tangent vector $X \in T_\Sigma \text{Sym}^+(n)$ commute is $\gamma(t) = (\Sigma^{\alpha_0} + t \alpha_0 \Sigma^{\alpha_0-1} X)^{1/\alpha_0}$ where $\alpha_0 = \frac{\alpha + \beta}{2} \neq 0$. Once this is shown, the formulae of the logarithm and the distance are obvious so we omit the proofs. As the metric is $O(n)$ -invariant, we can assume that Σ and X are diagonal matrices.

First, we assume that $\alpha, \beta \neq 0$. As $\text{MPE}(\alpha, \beta)$ is a balanced metric, the Levi-Civita connection is $\nabla^{\text{MPE}(\alpha, \beta)} = \frac{1}{2}(\text{pow}_\alpha^* \nabla^E + \text{pow}_\beta^* \nabla^E)$ where ∇^E is the Euclidean connection on symmetric matrices. Since for any curve γ on $\text{Sym}^+(n)$, we have:

$$\begin{aligned} (\text{pow}_\alpha^* \nabla^E)_{\gamma'(t)} \gamma' &= (d_{\gamma(t)} \text{pow}_\alpha)^{-1} (\nabla_{(d_{\gamma(t)} \text{pow}_\alpha)(\gamma'(t))}^E d \text{pow}_\alpha(\gamma')) \\ &= (d_{\gamma(t)} \text{pow}_\alpha)^{-1} (\nabla_{(\gamma^\alpha)'(t)}^E (\gamma^\alpha)') \\ &= (d_{\gamma(t)} \text{pow}_\alpha)^{-1} ((\gamma^\alpha)''(t)), \end{aligned}$$

the geodesic equation $\nabla_{\gamma'}^{\text{MPE}(\alpha, \beta)} \gamma' = 0$ rewrites:

$$(d_{\gamma(t)} \text{pow}_\alpha)^{-1} ((\gamma^\alpha)''(t)) + (d_{\gamma(t)} \text{pow}_\beta)^{-1} ((\gamma^\beta)''(t)) = 0.$$

We compute:

$$\begin{aligned} \gamma(t)^\alpha &= \Sigma^\alpha (I_n + t \alpha_0 \Sigma^{-1} X)^{\frac{\alpha}{\alpha_0}}, \\ (\gamma^\alpha)'(t) &= \alpha \Sigma^{\alpha-1} X (I_n + t \alpha_0 \Sigma^{-1} X)^{\frac{\alpha}{\alpha_0}-1}, \\ (\gamma^\alpha)''(t) &= \alpha(\alpha - \alpha_0) \Sigma^{\alpha-2} X^2 (I_n + t \alpha_0 \Sigma^{-1} X)^{\frac{\alpha}{\alpha_0}-2}, \\ (d_{\gamma(t)} \text{pow}_\alpha)^{-1} ((\gamma^\alpha)''(t)) &= \frac{1}{\alpha} \gamma(t)^{1-\alpha} (\gamma^\alpha)''(t) \\ &= \frac{\alpha - \beta}{2} \Sigma^{-1} X^2 (I_n + t \alpha_0 \Sigma^{-1} X)^{\frac{1}{\alpha_0}-2}. \end{aligned} \tag{11.25}$$

As this expression is skew-symmetric in (α, β) , the curve γ satisfies the geodesic equation.

Second, we assume that $\alpha \neq 0$ and $\beta = 0$. Similarly, the Levi-Civita connection is $\nabla^{\text{MPE}(\alpha, \beta)} = \frac{1}{2}(\text{pow}_\alpha^* \nabla^E + \log^* \nabla^E)$. Hence the geodesic equation is analogously $(d_{\gamma(t)} \text{pow}_\alpha)^{-1}((\gamma^\alpha)''(t)) + (d_{\gamma(t)} \log)^{-1}((\log \gamma)''(t)) = 0$. Thus:

$$\begin{aligned} \log \gamma(t) &= \log \Sigma + \frac{1}{\alpha_0} \log(I_n + t \alpha_0 \Sigma^{-1} X), \\ (\log \gamma)'(t) &= \Sigma^{-1} X (I_n + t \alpha_0 \Sigma^{-1} X)^{-1}, \\ (\log \gamma)''(t) &= -\alpha_0 \Sigma^{-2} X^2 (I_n + t \alpha_0 \Sigma^{-1} X)^{-2}, \\ (d_{\gamma(t)} \log)^{-1}((\log \gamma)''(t)) &= \gamma(t) (\log \gamma)''(t) \\ &= -\frac{\alpha}{2} \Sigma^{-1} X^2 (I_n + t \alpha_0 \Sigma^{-1} X)^{\frac{1}{\alpha_0} - 2}. \end{aligned}$$

This expression cancels Equation (11.25) with $\beta = 0$ so the curve γ is the geodesic. \square

11.5 Proofs of Chapter 6

In this section, we prove the results of Chapter 6: Geometry of Quotient-affine metrics.

11.5.1 Proof of Theorem 6.2

Theorem 6.2 (Vertical and horizontal distributions and projections) The vertical distribution is given by $\mathcal{V}_\Sigma = \Sigma \bullet \psi(\mathbb{R}^n)$ and the horizontal distribution is given by $\mathcal{H}_\Sigma = \mathcal{S}_{\Sigma^{-1}}(\text{Hol}(n))$. The vertical projection is:

$$\text{ver} : V \in T_\Sigma \text{Sym}^+(n) \mapsto \Sigma \bullet \psi((I_n + A(\Sigma))^{-1} \text{Diag}(\Sigma^{-1} V) \mathbf{1}) \in \mathcal{V}_\Sigma.$$

Then, the horizontal projection is simply $\text{hor}(V) = V - \text{ver}(V)$.

Proof of Theorem 6.2 (Vertical and horizontal distributions and projections). We denote $\Delta_\Sigma = \text{Diag}(\Sigma)^{1/2}$. Using Section 6.2.1, we have:

$$\begin{aligned} d_\Sigma \pi(V) &= \Delta_\Sigma^{-1} \left[V - \frac{1}{2} (\Delta_\Sigma^{-2} \text{Diag}(V) \Sigma + \Sigma \text{Diag}(V) \Delta_\Sigma^{-2}) \right] \Delta_\Sigma^{-1}, \\ V \in \mathcal{V}_\Sigma &\iff \forall i, j \in \{1, \dots, n\}, V_{ij} = \Sigma_{ij} \times \frac{1}{2} \left(\frac{V_{ii}}{\Sigma_{ii}} + \frac{V_{jj}}{\Sigma_{jj}} \right) \iff V \in \Sigma \bullet \psi(\mathbb{R}^n), \\ W \in \mathcal{H}_\Sigma &\iff \forall V \in \mathcal{V}_\Sigma, \text{tr}(\Sigma^{-1} V \Sigma^{-1} W) = 0, \\ &\iff \forall \mu \in \mathbb{R}^n, \sum_{ijkl} [\Sigma^{-1}]_{ij} \Sigma_{jk} (\mu_j + \mu_k) [\Sigma^{-1}]_{kl} W_{li} = 0, \\ &\iff \forall D \in \text{Diag}(n), \text{tr}(\Sigma^{-1} D W + D \Sigma^{-1} W) = 0, \\ &\iff (\Sigma^{-1} W + W \Sigma^{-1}) \in \text{Hol}(n) \iff W \in \mathcal{S}_{\Sigma^{-1}}(\text{Hol}(n)). \end{aligned}$$

Thus, we have computed the horizontal space for the affine-invariant metric $\alpha = 1$ and $\beta = 0$. It is still valid for all $\alpha > 0$ and $\beta > -\alpha/n$ since the latter is included in the former (because $\text{tr}(\Sigma^{-1} W) = 0$) and they have the same dimension so they are equal. Now we compute the vertical projection. Let $V \in T_\Sigma \text{Sym}^+(n)$ and let $\mu \in \mathbb{R}^n$ and $W \in \mathcal{H}_\Sigma$ such that $V = \Sigma \bullet \psi(\mu) + W$. We are looking for μ . Since $\Sigma^{-1} V = \Sigma^{-1} (\Sigma \bullet \psi(\mu)) + \Sigma^{-1} W$, we have $[\Sigma^{-1} V]_{ii} = \sum_j [\Sigma^{-1}]_{ij} \Sigma_{ij} (\mu_i + \mu_j) = \mu_i + [A(\Sigma) \mu]_i$ so $\text{Diag}(\Sigma^{-1} V) \mathbf{1} = (I_n + A(\Sigma)) \mu$, hence $\mu = (I_n + A(\Sigma))^{-1} \text{Diag}(\Sigma^{-1} V) \mathbf{1}$. \square

11.5.2 Proof of Theorem 6.3

Theorem 6.3 (Horizontal lift) Let $\Sigma \in \text{Sym}^+(n)$ and $C = \text{Cor}(\Sigma) \in \text{Cor}^+(n)$. The horizontal lift at Σ of $X \in T_C \text{Cor}^+(n)$ is $X^\# = \text{hor}(\Delta_\Sigma X \Delta_\Sigma)$ with $\Delta_\Sigma = \text{Diag}(\Sigma)^{1/2}$. In particular, the horizontal lift at $C \in \text{Sym}^+(n)$ is $X^\# = \text{hor}(X)$.

Proof of Theorem 6.3 (Horizontal lift). The horizontal lift is $X^\# = \text{hor}(\Delta_\Sigma X \Delta_\Sigma)$ since $d_\Sigma \pi(\Delta_\Sigma X \Delta_\Sigma) = \Delta_\Sigma^{-1} [V - \frac{1}{2}(\Delta_\Sigma^{-2} \text{Diag}(V) \Sigma + \Sigma \text{Diag}(V) \Delta_\Sigma^{-2})] \Delta_\Sigma^{-1} = X$, where $V = \Delta_\Sigma X \Delta_\Sigma$. \square

11.5.3 Proof of Theorem 6.4

Theorem 6.4 (Expression of quotient-affine metrics) For all $C \in \text{Cor}^+(n)$ and $X \in T_C \text{Cor}^+(n)$, $g_C^{\alpha, \beta}(X, X) = \alpha g_C^{\text{QA}}(X, X)$ (independent from β) where:

$$g_C^{\text{QA}}(X, X) = \text{tr}((C^{-1}X)^2) - 2\mathbf{1}^\top \text{Diag}(C^{-1}X)(I_n + A(C))^{-1} \text{Diag}(C^{-1}X)\mathbf{1}.$$

Proof of Theorem 6.4 (Expression of quotient-affine metrics). We use the definition of the quotient metric.

$$g_C^{\alpha, \beta}(X, X) = G_C^{\alpha, \beta}(\text{hor}(X), \text{hor}(X)) = G_C^{\alpha, \beta}(X, X) - G_C^{\alpha, \beta}(\text{ver}(X), \text{ver}(X)).$$

First, we observe that $\text{tr}(C^{-1}X) = \text{tr}(C^{-1}\text{ver}(X))$ since $\text{tr}(C^{-1}\text{hor}(X)) = 0$ so the terms in β vanish and $g_C^{\alpha, \beta}(X, X) = \alpha(\text{tr}(C^{-1}XC^{-1}X) - \text{tr}(C^{-1}\text{ver}(X)C^{-1}\text{ver}(X)))$ does not depend on β .

We denote $\text{ver}(X) = C \bullet \psi(\mu)$ with $\mu = (I_n + A(C))^{-1} \text{Diag}(C^{-1}X)\mathbf{1}$. Then:

$$\begin{aligned} \text{tr}(C^{-1}\text{ver}(X)C^{-1}\text{ver}(X)) &= \sum_{ijkl} [C^{-1}]_{ij} C_{jk} (\mu_j + \mu_k) [C^{-1}]_{kl} C_{li} (\mu_l + \mu_i), \\ &= 2 \sum_{ij} (\delta_{ij} + [C^{-1}]_{ij} C_{ij}) \mu_i \mu_j, \\ &= 2\mu^\top (I_n + A(C)) \mu, \\ &= 2\mathbf{1}^\top \text{Diag}(C^{-1}X)(I_n + A(C))^{-1} \text{Diag}(C^{-1}X)\mathbf{1}. \end{aligned}$$

\square

11.5.4 Proof of Theorem 6.5

Theorem 6.5 (Geodesics of quotient-affine metrics) The geodesic from $C \in \text{Cor}^+(n)$ with initial tangent vector $X \in T_C \text{Cor}^+(n)$ is:

$$\forall t \in \mathbb{R}, \gamma_{(C, X)}^{\text{QA}}(t) = \text{Exp}_C^{\text{QA}}(tX) = \pi(C^{1/2} \exp(t C^{-1/2} \text{hor}(X) C^{-1/2}) C^{1/2}).$$

In particular, the quotient-affine metric is geodesically complete.

Proof of Theorem 6.5 (Geodesics of quotient-affine metrics). The geodesics are the the projections of horizontal geodesics [Gallot et al., 2004] so they are given by $\text{Exp}_C^{\text{QA}}(X) = \pi(\text{Exp}_C^G(X^\#)) = \pi(C^{1/2} \exp(C^{-1/2} \text{hor}(X) C^{-1/2}) C^{1/2})$. \square

11.5.5 Proof of Theorem 6.6

Theorem 6.6 (Existence of a logarithm) For all $C_1, C_2 \in \text{Cor}^+(n)$, there exists $X \in T_C \text{Cor}^+(n)$ such that $\text{Exp}_{C_1}^{\text{QA}}(X) = C_2$.

Proof of Theorem 6.6 (Existence of a logarithm). We prove the coercivity of $f : D \in \text{Diag}^+(n) \mapsto \text{tr}(\log(\Sigma D \Lambda D \Sigma)^2)$. Since $f(D) = \text{tr}(\log(\Sigma' D^{-1} \Lambda' D^{-1} \Sigma')^2)$ with $\Sigma' = \Sigma^{-1}$ and $\Lambda' = \Lambda^{-1}$, it suffices to prove that if $d_1 \rightarrow 0$, then $f(D) \rightarrow +\infty$. Indeed, if $d_1 \rightarrow 0$, then $[D \Lambda D]_{11} \rightarrow 0$ so $D \Lambda D$ tends to a singular matrix and so does $\Sigma D \Lambda D \Sigma$. Hence $f(D) \rightarrow +\infty$. \square

11.5.6 Proof of Theorem 6.7

Theorem 6.7 (Levi-Civita connection and sectional curvature of quotient-affine metrics) The Levi-Civita connection of quotient-affine metrics is:

$$\begin{aligned} (\nabla_X^{\text{QA}} Y)|_C &= d_C Y(X) + \text{sym}[\text{Diag}(X^\#)Y^\# + \text{Diag}(Y^\#)X^\# + \text{Diag}(X^\# C^{-1} Y^\#)C \\ &\quad - X^\# C^{-1} Y^\# - \frac{1}{2} \text{Diag}(X^\#)C \text{Diag}(Y^\#) - \frac{3}{2} \text{Diag}(X^\#) \text{Diag}(Y^\#)C]. \end{aligned}$$

The curvature of quotient-affine metrics is:

$$\begin{aligned} \kappa_C^{\text{QA}}(X, Y) &= \kappa_C^G(X^\#, Y^\#) + \frac{3 G_C(\text{ver}[X^\#, Y^\#], \text{ver}[X^\#, Y^\#])}{4 g_C(X, X)g_C(Y, Y) - g_C(X, Y)^2}, \\ &= \frac{2 \text{tr}((C^{-1} X^\# C^{-1} Y^\# - C^{-1} Y^\# C^{-1} X^\#)^2) + 3 \mathbf{1}^\top D(I_n + A(C))^{-1} D \mathbf{1}}{8(g_C(X, X)g_C(Y, Y) - g_C(X, Y)^2)} \end{aligned}$$

where $[V, W] = dW(V) - dV(W)$ is the Lie bracket on $\text{Sym}^+(n)$ and $D = D(X, Y) - D(Y, X)$ with $D(X, Y) = \text{Diag}(C^{-1} \text{Diag}(X^\#)Y^\# - C^{-1} Y^\# C^{-1} \text{Diag}(X^\#)C)$. There is a slight abuse of notation because $\text{ver}[X^\#, Y^\#]$ induces that $X^\#$ and $Y^\#$ are vector fields. Indeed here, they are horizontal vector fields extending the horizontal lifts at C .

Proof of Theorem 6.7 (Levi-Civita connection and sectional curvature of quotient-affine metrics). According O'Neill's equations of submersions, the Levi-Civita connection of quotient-affine metrics is given by:

$$\begin{aligned} (\nabla_X^{\text{QA}} Y)|_C &= d_\Sigma \pi((\nabla_{X^\#}^G Y^\#)|_C), \\ &= d_C \pi(d_C Y^\#(X^\#) - \frac{1}{2}(X^\# C^{-1} Y^\# + Y^\# C^{-1} X^\#)), \\ &= d_C Y^\#(X^\#) - \frac{1}{2}(\text{Diag}(d_C Y^\#(X^\#))C + C \text{Diag}(d_C Y^\#(X^\#))) \\ &\quad - \frac{1}{2}(X^\# C^{-1} Y^\# + Y^\# C^{-1} X^\# - \text{Diag}(X^\# C^{-1} Y^\#)C - C \text{Diag}(X^\# C^{-1} Y^\#)). \end{aligned} \tag{11.26}$$

Since $Y_{\pi(\Sigma)} = \Delta_\Sigma^{-1} Y_\Sigma^\# \Delta_\Sigma^{-1} - \text{sym}(\Delta_\Sigma^{-3} \text{Diag}(Y_\Sigma^\#) \Sigma \Delta_\Sigma^{-1})$ where $\Delta_\Sigma = \text{Diag}(\Sigma)^{1/2}$, then:

$$\begin{aligned} d_{\pi(\Sigma)} Y(X) &= d_\Sigma(Y \circ \pi)(X^\#), \\ &= \Delta_\Sigma^{-1}(d_\Sigma Y^\#(X^\#)) \Delta_\Sigma^{-1} - \text{sym}(\Delta_\Sigma^{-3} \text{Diag}(d_\Sigma Y^\#(X^\#)) \Sigma \Delta_\Sigma^{-1}) \\ &\quad - \text{sym}[\Delta_\Sigma^{-3} \text{Diag}(X^\#) Y^\# \Delta_\Sigma^{-1} - \frac{3}{2} \Delta_\Sigma^{-5} \text{Diag}(X^\#) \text{Diag}(Y^\#) \Sigma \Delta_\Sigma^{-1} \\ &\quad + \Delta_\Sigma^{-3} \text{Diag}(Y^\#) X^\# \Delta_\Sigma^{-1} - \frac{1}{2} \Delta_\Sigma^{-3} \text{Diag}(Y^\#) \Sigma \text{Diag}(X^\#) \Delta_\Sigma^{-3}]. \end{aligned} \tag{11.27}$$

Since line (11.26) is line (11.27) at $\Sigma = C$, we finally get:

$$\begin{aligned} (\nabla_X^{\text{QA}} Y)|_C &= d_C Y(X) + \text{sym}[\text{Diag}(X^\#)Y^\# + \text{Diag}(Y^\#)X^\# + \text{Diag}(X^\#C^{-1}Y^\#)C \\ &\quad - X^\#C^{-1}Y^\# - \frac{1}{2}\text{Diag}(X^\#)C\text{Diag}(Y^\#) - \frac{3}{2}\text{Diag}(X^\#)\text{Diag}(Y^\#)C]. \end{aligned}$$

The curvature in formula (6.10) directly comes from the fundamental equations of submersions [O'Neill, 1966]. The curvature of the affine-invariant metric comes from [Skovgaard, 1984]. Hence we only have to compute $G_C(\text{ver}[X^\#, Y^\#], \text{ver}[X^\#, Y^\#])$ where G is the affine-invariant metric, $[\cdot, \cdot]$ is the Lie bracket on the manifold of SPD matrices (it is not the matrix commutator), $C \in \text{Cor}^+(n)$ and $X^\#, Y^\#$ are horizontal vector fields on $\text{Sym}^+(n)$ extending $X_C^\#$ and $Y_C^\#$ respectively, where $X, Y \in T_C \text{Cor}^+(n)$ are tangent vectors at C . For example, we can consider that X, Y are constant vector fields on $\text{Cor}^+(n)$ and simply define $X^\#, Y^\#$ as their horizontal lifts everywhere:

$$\begin{aligned} Y_\Sigma^\# &= \text{hor}(\Delta_\Sigma Y \Delta_\Sigma), \\ &= \Delta_\Sigma Y \Delta_\Sigma - \text{ver}(\Delta_\Sigma Y \Delta_\Sigma), \\ &= \Delta_\Sigma Y \Delta_\Sigma - 2\Sigma \bullet \psi((I_n + \Sigma \bullet \Sigma^{-1})^{-1} \text{Diag}(\Sigma^{-1} \Delta_\Sigma Y \Delta_\Sigma) \mathbf{1}), \\ &= \Delta_\Sigma Y \Delta_\Sigma - 2\Sigma \bullet \psi((I_n + \Sigma \bullet \Sigma^{-1})^{-1} \text{Diag}(\pi(\Sigma)^{-1} Y) \mathbf{1}), \end{aligned}$$

with $\Delta_\Sigma = \text{Diag}(\Sigma)^{1/2}$ and $\psi(\mu) = \mu \mathbf{1}^\top + \mathbf{1} \mu^\top$ for $\mu \in \mathbb{R}^n$. Now we can compute $[X^\#, Y^\#] = \partial_{X^\#} Y^\# - \partial_{Y^\#} X^\#$ and evaluate it at C . We have:

$$\begin{aligned} \partial_{X^\#} Y^\# &= \frac{1}{2}(\Delta_\Sigma^{-1} \text{Diag}(X^\#) Y \Delta_\Sigma + \Delta_\Sigma Y \text{Diag}(X^\#) \Delta_\Sigma^{-1}) \\ &\quad - X^\# \bullet \psi((I_n + \Sigma \bullet \Sigma^{-1})^{-1} \text{Diag}(\Sigma^{-1} Y) \mathbf{1}) + v_\Sigma(X, Y), \end{aligned}$$

where $v_\Sigma(X, Y) \in \mathcal{V}_\Sigma$ is a short notation for the following expression:

$$\begin{aligned} &\Sigma \bullet \psi[(I_n + \Sigma \bullet \Sigma^{-1})^{-1} (X^\# \bullet \Sigma^{-1} - \Sigma \bullet \Sigma^{-1} X^\# \Sigma^{-1}) (I_n + \Sigma \bullet \Sigma^{-1})^{-1} \text{Diag}(\Sigma^{-1} Y) \mathbf{1} \\ &\quad + (I_n + \Sigma \bullet \Sigma^{-1})^{-1} \text{Diag}(\pi(\Sigma)^{-1} X \pi(\Sigma)^{-1} Y) \mathbf{1}] \end{aligned}$$

We evaluate it at $C \in \text{Cor}^+(n)$ so that $\Delta_C = I_n$. We denote $B = (I_n + C \bullet C^{-1})^{-1}$ to simplify the notations. Note that the last term is symmetric in X and Y so we can define $v_C^0(X, Y) = C \bullet \psi[B(X^\# \bullet C^{-1} - C \bullet C^{-1} X^\# C^{-1}) B \text{Diag}(C^{-1} Y) \mathbf{1}]$. Then:

$$[X^\#, Y^\#] = \frac{1}{2}(\text{Diag}(X^\#)Y + Y \text{Diag}(X^\#) - \text{Diag}(Y^\#)X - X \text{Diag}(Y^\#)), \quad (11.28)$$

$$- (X^\# \bullet \psi(B \text{Diag}(C^{-1} Y) \mathbf{1}) - Y^\# \bullet \psi(B \text{Diag}(C^{-1} X) \mathbf{1})), \quad (11.29)$$

$$+ v_C^0(X, Y) - v_C^0(Y, X). \quad (11.30)$$

Fortunately, the computation of $\text{ver}[X^\#, Y^\#] = C \bullet \psi(B \text{Diag}(C^{-1} [X^\#, Y^\#]) \mathbf{1})$ brings some simplifications. We proceed line by line.

1. To simplify line (11.28), we plug $Y = Y^\# - \frac{1}{2}(\text{Diag}(Y^\#)C + C \text{Diag}(Y^\#))$. We proceed term by term.

(a)

$$\begin{aligned} \text{Diag}(C^{-1} \text{Diag}(X^\#) Y) &= \text{Diag}(C^{-1} \text{Diag}(X^\#) Y^\#) \\ &\quad - \frac{1}{2} \text{Diag}(C^{-1} \text{Diag}(X^\#) \text{Diag}(Y^\#) C) - \frac{1}{2} \text{Diag}(C^{-1} \text{Diag}(X^\#) C \text{Diag}(Y^\#)) \end{aligned} \quad (11.31)$$

The second term is symmetric in X and Y .

(b)

$$\begin{aligned} \text{Diag}(C^{-1}Y\text{Diag}(X^\#)) &= \text{Diag}(C^{-1}Y^\#\text{Diag}(X^\#)) \\ &\quad - \frac{1}{2}\text{Diag}(C^{-1}\text{Diag}(Y^\#)C\text{Diag}(X^\#)) - \frac{1}{2}\text{Diag}(Y^\#)\text{Diag}(X^\#) \end{aligned} \quad (11.32)$$

The first term is null, the third term is symmetric in X and Y .

The expression composed of the third term of Equation (11.31) plus the second term of Equation (11.32) is symmetric in X and Y . Hence, the vertical projection of line (11.28) reduces to the vertical projection of $\frac{1}{2}(\text{Diag}(X^\#)Y^\# - \text{Diag}(Y^\#)X^\#)$.

2. To simplify line (11.29), we can show by computing coordinate by coordinate that $\text{Diag}(C^{-1}(X^\# \bullet \psi(B\text{Diag}(C^{-1}Y)\mathbf{1})))\mathbf{1} = (C^{-1} \bullet X^\#)B\text{Diag}(C^{-1}Y)\mathbf{1}$. Hence, line (11.29) cancels the first term of v^0 .
3. Finally, a nicer expression of the second term of v^0 can be obtained by noticing that $\text{Diag}(Y^\#)\mathbf{1} = -2B\text{Diag}(C^{-1}Y)\mathbf{1}$. After simple calculations, we get $(C \bullet C^{-1}X^\#C^{-1})B\text{Diag}(C^{-1}Y)\mathbf{1} = -\frac{1}{2}\text{Diag}(C\text{Diag}(X^\#)C^{-1}Y^\#C^{-1})\mathbf{1}$.

To summarize, we have $\text{ver}[X^\#, Y^\#] = C \bullet \psi(\frac{1}{2}BD\mathbf{1})$ with $D = D(X, Y) - D(Y, X) \in \text{Diag}(n)$ and $D(X, Y) = \text{Diag}(C^{-1}\text{Diag}(X^\#)Y^\# - C^{-1}Y^\#C^{-1}\text{Diag}(X^\#)C)$. Finally, since $G_C(C \bullet \psi(\mu), C \bullet \psi(\mu)) = 2\mu^\top B\mu$ for any vector $\mu \in \mathbb{R}^n$, we get $G_C(\text{ver}[X^\#, Y^\#], \text{ver}[X^\#, Y^\#]) = \frac{1}{2}\mathbf{1}^\top DBD\mathbf{1}$, as expected. \square

11.6 Proofs of Chapter 7

In this section, we prove the results of Chapter 7: Theoretically and computationally convenient Cholesky-based geometries.

11.6.1 Proof of Theorem 7.1

Theorem 7.1 (Bounds of curvature) The sectional curvature of the quotient-affine metric takes positive and negative values. It is bounded from below and unbounded from above.

Proof appendix of Theorem 7.1 (Bounds of curvature). First of all, $\kappa_C(X, Y) \geq \kappa_C^{\text{AI}}(X, Y) \geq -\frac{1}{2}$ so the curvature is bounded from below. Second, at $C = I_n$, $X^\# = X$ and $Y^\# = Y$ so $\text{Diag}(X^\#) = \text{Diag}(Y^\#) = 0$ and $\mu = 0$. Hence, $\kappa_{I_n}(X, Y) = \kappa_{I_n}^{\text{AI}}(X, Y) \leq 0$ and for example $\kappa_{I_n}(E_{ij}, E_{ik}) = -\frac{1}{8} < 0$ with $i \neq j \neq k \neq i \in \{1, \dots, n\}$ [Thanwerdas and Pennec, 2022b]. So the curvature takes negative values. Third, let $X = \mathbf{1}\mathbf{1}^\top - I_n$ and $Y = \mu\mathbf{1}^\top + \mathbf{1}\mu^\top - 2\text{diag}(\mu)$ with $\text{sum}(\mu) = \mathbf{1}^\top \mu = 0$ where $\mu \in \mathbb{R}^n$. Let $C = (1 - \rho)I_n + \rho\mathbf{1}\mathbf{1}^\top \in \text{Cor}^+(n)$ for $\rho \in (-\frac{1}{n-1}, 1)$. Let us show that $\kappa_C(X, Y)$ tends to $+\infty$ when $\rho \rightarrow -\frac{1}{n-1}$, which proves that the curvature is not bounded from above.

The symmetric matrix $\mathbf{1}\mathbf{1}^\top$ has two eigenvalues: 0 with multiplicity $n-1$ and n with multiplicity 1. Since $(\mathbf{1}\mathbf{1}^\top)^2 = n\mathbf{1}\mathbf{1}^\top$, the minimal polynomial is $P_{\mathbf{1}\mathbf{1}^\top}(x) = x(x-n)$ for $x \in \mathbb{R}$. For all $\alpha, \beta \in \mathbb{R}$, the symmetric matrix $\Sigma = \alpha I_n + \beta\mathbf{1}\mathbf{1}^\top$ has minimal polynomial $P_\Sigma(x) = (x-\alpha)(x-(\alpha+n\beta))$ for $x \in \mathbb{R}$, which is of degree 2. Hence Σ is positive definite if and only if $\alpha > 0$ and $\alpha + n\beta > 0$. In this case, its inverse is a polynomial in $\mathbf{1}\mathbf{1}^\top$ of degree 1. More precisely, $\Sigma^{-1} = \alpha' I_n + \beta'\mathbf{1}\mathbf{1}^\top$ with $\alpha' = \frac{1}{\alpha}$ and $\alpha' + n\beta' = \frac{1}{\alpha+n\beta}$, i.e. $\beta' = -\frac{\beta}{\alpha(\alpha+n\beta)}$. Note that $\alpha\beta' + \beta\alpha' + n\beta\beta' = 0$.

Moreover, for all $i \neq j \in \{1, \dots, n\}$, $[\Sigma \bullet \Sigma^{-1}]_{ii} = (\alpha + \beta)(\alpha' + \beta')$ and $[\Sigma \bullet \Sigma^{-1}]_{ij} = \beta\beta'$. Therefore, $I_n + \Sigma \bullet \Sigma^{-1} = AI_n + B\mathbf{1}\mathbf{1}^\top$ with $A = 1 + (\alpha + \beta)(\alpha' + \beta') - \beta\beta' = 2 + \alpha\beta' + \beta\alpha' = \frac{2\alpha(\alpha+n\beta)+n\beta^2}{\alpha(\alpha+n\beta)}$ and $B = \beta\beta'$. Note that $A + nB = 2$. And $(I_n + \Sigma \bullet \Sigma^{-1})^{-1} = A'I_n + B'\mathbf{1}\mathbf{1}^\top$ with $A' = \frac{1}{A}$ and $B' = -\frac{B}{2A}$.

We compute $\kappa_C(X, Y)$ where $C = \alpha I_n + \beta\mathbf{1}\mathbf{1}^\top \in \text{Cor}^+(n)$ with $\alpha + \beta = 1$.

$$\begin{aligned}
C^{-1}X &= (\alpha'I_n + \beta'\mathbf{1}\mathbf{1}^\top)(\mathbf{1}\mathbf{1}^\top - I_n) \\
&= -\alpha'I_n + (\alpha' + (n-1)\beta')\mathbf{1}\mathbf{1}^\top \\
&= -\frac{1}{\alpha}I_n + \frac{1}{\alpha(\alpha+n\beta)}\mathbf{1}\mathbf{1}^\top, \\
\text{Diag}(C^{-1}X) &= (n-1)\beta'I_n, \\
(I_n + C \bullet C^{-1})^{-1}\text{Diag}(C^{-1}X)\mathbf{1} &= (n-1)\beta'(A'I_n + B'\mathbf{1}\mathbf{1}^\top)\mathbf{1} \\
&= \frac{n-1}{2}\beta'\mathbf{1}, \\
X^\# &= X - (n-1)\beta'C, \\
\text{Diag}(X^\#) &= -(n-1)\beta'I_n, \\
D(X, Y) &= \text{Diag}(C^{-1}\text{Diag}(X^\#)Y^\# - C^{-1}Y^\#C^{-1}\text{Diag}(X^\#)C) \\
&= 0, \\
C^{-1}Y &= (\alpha'I_n + \beta'\mathbf{1}\mathbf{1}^\top)(\mu\mathbf{1}^\top + \mathbf{1}\mu^\top - 2\text{diag}(\mu)) \\
&= \alpha'\mu\mathbf{1}^\top + (\alpha' + (n-2)\beta')\mathbf{1}\mu^\top - 2\alpha'\text{diag}(\mu), \\
\text{Diag}(C^{-1}Y) &= (n-2)\beta'\text{diag}(\mu), \\
(I_n + C \bullet C^{-1})^{-1}\text{Diag}(C^{-1}Y)\mathbf{1} &= (n-2)\beta'(A'I_n + B'\mathbf{1}\mathbf{1}^\top)\mu \\
&= (n-2)\beta'A'\mu, \\
Y^\# &= Y - (n-2)\beta'A'(\text{diag}(\mu)C + C\text{diag}(\mu)), \\
\text{Diag}(Y^\#) &= -2(n-2)\beta'A'\text{diag}(\mu), \\
C^{-1}\text{Diag}(Y^\#)X^\# &= -2(n-2)\beta'A'(\alpha'I_n + \beta'\mathbf{1}\mathbf{1}^\top)\text{diag}(\mu)(\mathbf{1}\mathbf{1}^\top - I_n) \\
&= -2(n-2)\beta'A'(-\alpha'\text{diag}(\mu) + \alpha'\mu\mathbf{1}^\top - \beta'\mathbf{1}\mu^\top),
\end{aligned}$$

$$\begin{aligned}
&C^{-1}X^\#C^{-1}\text{Diag}(Y^\#)C \\
&= -2(n-2)\beta'A'(-\alpha'I_n + (\alpha' + (n-1)\beta')\mathbf{1}\mathbf{1}^\top)(\alpha'I_n + \beta'\mathbf{1}\mathbf{1}^\top)\text{diag}(\mu)(\alpha I_n + \beta\mathbf{1}\mathbf{1}^\top) \\
&= -2(n-2)\beta'A'(-\alpha'I_n + (\alpha' + (n-1)\beta')\mathbf{1}\mathbf{1}^\top)(\text{diag}(\mu) + \alpha'\beta\mu\mathbf{1}^\top + \alpha\beta'\mathbf{1}\mu^\top) \\
&= -2(n-2)\beta'A'[-\alpha'(\text{diag}(\mu) + \alpha'\beta\mu\mathbf{1}^\top + \alpha\beta'\mathbf{1}\mu^\top) + (\alpha' + (n-1)\beta')(1 + n\alpha\beta')\mathbf{1}\mu^\top],
\end{aligned}$$

$$\begin{aligned}
&C^{-1}\text{Diag}(Y^\#)X^\# - C^{-1}X^\#C^{-1}\text{Diag}(Y^\#)C \\
&= -2(n-2)\beta'A'[\alpha'(1 + \alpha'\beta)\mu\mathbf{1}^\top - (\alpha' + (n-1)\beta')(1 + n\alpha\beta')\mathbf{1}\mu^\top],
\end{aligned}$$

$$\begin{aligned}
D(Y, X) &= -2(n-2)\beta' A' \left(\frac{1}{\alpha^2} - \frac{1+(n-1)\alpha\beta'}{\alpha+n\beta} \right) \text{diag}(\mu) \\
&= -2(n-2)\beta' A' \left(\frac{1}{\alpha^2} - \frac{1}{(\alpha+n\beta)^2} \right) \text{diag}(\mu) \\
&= 2n(n-2) \frac{\beta^2}{2\alpha(\alpha+n\beta) + n\beta^2} \frac{2\alpha+n\beta}{\alpha^2(\alpha+n\beta)^2} \text{diag}(\mu),
\end{aligned} \tag{11.33}$$

$$\begin{aligned}
\mathbf{1}^\top D(I_n + C \bullet C^{-1})^{-1} D \mathbf{1} &= A' \left(2n(n-2) \frac{\beta^2}{2\alpha(\alpha+n\beta) + n\beta^2} \frac{2\alpha+n\beta}{\alpha^2(\alpha+n\beta)^2} \right)^2 \|\mu\|^2 \\
&= 4n^2(n-2)^2 \frac{\beta^4(2\alpha+n\beta)^2}{[\alpha(\alpha+n\beta)(2\alpha(\alpha+n\beta) + n\beta^2)]^3} \|\mu\|^2,
\end{aligned}$$

where we used $\alpha + \beta = 1$ from Equation (11.33).

Now, we compute $g_C^{\text{QA}}(X, X)$, $g_C^{\text{QA}}(Y, Y)$ and $g_C^{\text{QA}}(X, Y)$.

$$\begin{aligned}
g_C(X, X) &= \text{tr}((C^{-1}X)^2) - 2\mathbf{1}^\top \text{Diag}(C^{-1}X)(I_n + C \bullet C^{-1})^{-1} \text{Diag}(C^{-1}X) \mathbf{1} \\
&= \text{tr} \left((-\alpha' I_n + (\alpha' + (n-1)\beta') \mathbf{1}\mathbf{1}^\top)^2 \right) - 2(n-1)^2 \beta'^2 \text{sum}(A' I_n + B' \mathbf{1}\mathbf{1}^\top) \\
&= \text{tr}(\alpha'^2 I_n + (n(\alpha' + (n-1)\beta')^2 - 2\alpha'(\alpha' + (n-1)\beta')) \mathbf{1}\mathbf{1}^\top) \\
&\quad - 2(n-1)^2 \beta'^2 n(A' + nB') \\
&= n((n-1)\alpha'^2 + 2(n-1)^2 \alpha' \beta' + n(n-1)^2 \beta'^2) - n(n-1)^2 \beta'^2 \\
&= \frac{n(n-1)}{\alpha^2(\alpha+n\beta)^2} ((\alpha+n\beta)^2 - 2(n-1)\beta(\alpha+n\beta) + (n-1)^2 \beta^2) \\
&= \frac{n(n-1)}{\alpha^2(\alpha+n\beta)^2}, \\
g_C(Y, Y) &= \text{tr}((C^{-1}Y)^2) - 2\mathbf{1}^\top \text{Diag}(C^{-1}Y)(I_n + C \bullet C^{-1})^{-1} \text{Diag}(C^{-1}Y) \mathbf{1} \\
&= \text{tr} \left((\alpha' \mu \mathbf{1}^\top + (\alpha' + (n-2)\beta') \mathbf{1} \mu^\top - 2\alpha' \text{diag}(\mu))^2 \right) \\
&\quad - 2(n-2)^2 \beta'^2 \mu^\top (A' I_n + B' \mathbf{1}\mathbf{1}^\top) \mu \\
&= \text{tr}(4\alpha'^2 \text{diag}(\mu)^2 + \alpha'(\alpha' + (n-2)\beta')(n\mu \mu^\top + \|\mu\|^2 \mathbf{1}\mathbf{1}^\top)) \\
&\quad - 2\alpha' \text{tr}((2\alpha' + (n-2)\beta') \mu \mu^\top + \alpha'(\mu \bullet \mu) \mathbf{1}^\top + (\alpha' + (n-2)\beta') \mathbf{1}(\mu \bullet \mu)^\top)) \\
&\quad - 2(n-2)^2 \beta'^2 A' \|\mu\|^2 \\
&= \|\mu\|^2 (4\alpha'^2 + 2n\alpha'(\alpha' + (n-2)\beta') - 4\alpha'(2\alpha' + (n-2)\beta') - 2(n-2)^2 \beta'^2 A') \\
&= \|\mu\|^2 \left(2(n-2) \underbrace{\alpha'(\alpha' + (n-2)\beta')}_{\frac{1+\beta}{\alpha^2(\alpha+n\beta)}} - 2(n-2)^2 \underbrace{\beta'^2 A'}_{\frac{\beta^2}{\alpha(\alpha+n\beta)(2\alpha(\alpha+n\beta) + n\beta^2)}} \right) \\
&= \frac{2(n-2)\|\mu\|^2}{\alpha^2(\alpha+n\beta)(2\alpha(\alpha+n\beta) + n\beta^2)} \underbrace{((1+\beta)(2\alpha(\alpha+n\beta) + n\beta^2) - (n-2)\alpha\beta^2)}_{2\alpha(1+\beta)(\alpha+n\beta) + 2\beta^2(\alpha+n\beta) = 2(\alpha+n\beta)} \\
&= \frac{4(n-2)\|\mu\|^2}{\alpha^2(2\alpha(\alpha+n\beta) + n\beta^2)},
\end{aligned}$$

$$\begin{aligned}
& g_C(X, Y) \\
&= \text{tr}(C^{-1}XC^{-1}Y) - 2\mathbf{1}^\top \text{Diag}(C^{-1}X)(I_n + C \bullet C^{-1})^{-1} \text{Diag}(C^{-1}Y)\mathbf{1} \\
&= \text{tr}((-\alpha'I_n + (\alpha' + (n-1)\beta')\mathbf{1}\mathbf{1}^\top)(\alpha'\mu\mathbf{1}^\top + (\alpha' + (n-2)\beta')\mathbf{1}\mu^\top - 2\alpha'\text{diag}(\mu))) \\
&\quad - 2(n-1)(n-2)\beta'^2\mathbf{1}^\top(A'I_n + \beta'\mathbf{1}\mathbf{1}^\top)\mu \\
&= \text{constant} \times \text{sum}(\mu) = 0.
\end{aligned}$$

Finally:

$$\frac{3}{8} \frac{\mathbf{1}^\top D(I_n + C \bullet C^{-1})^{-1} D\mathbf{1}}{g_C(X, X)g_C(Y, Y) - g_C(X, Y)^2} = \frac{3n(n-2)}{8(n-1)} \frac{\alpha\beta^4}{\alpha + n\beta} \left(\frac{2\alpha + n\beta}{2\alpha(\alpha + n\beta) + n\beta^2} \right)^2.$$

When $\beta \rightarrow -\frac{1}{n-1}$ (and $\alpha = 1 - \beta \rightarrow \frac{n}{n-1}$), we have $\alpha + n\beta \rightarrow 0$ and we have $\alpha\beta^4 \left(\frac{2\alpha + n\beta}{2\alpha(\alpha + n\beta) + n\beta^2} \right)^2 \rightarrow \frac{n}{(n-1)^3}$ so this quantity tends to $+\infty$. Finally with $C = (1 - \rho)I_n + \rho\mathbf{1}\mathbf{1}^\top \in \text{Cor}^+(n)$ with $\rho \in (-\frac{1}{n-1}; 1)$, $X = I_n - \mathbf{1}\mathbf{1}^\top \in T_C\text{Cor}^+(n)$ and $Y = \mu\mathbf{1}^\top + \mathbf{1}\mu^\top - 2\text{diag}(\mu) \in T_C\text{Cor}^+(n)$, we have:

$$\kappa_C(X, Y) \geq -\frac{1}{2} + \frac{3}{8} \frac{\mathbf{1}^\top D(I_n + C \bullet C^{-1})^{-1} D\mathbf{1}}{g_C(X, X)g_C(Y, Y) - g_C(X, Y)^2} \xrightarrow{\rho \rightarrow -\frac{1}{n-1}} +\infty. \quad (11.34)$$

This proves that the quotient-affine sectional curvature is not bounded from above. \square

11.6.2 Proof of Theorem 7.2

Theorem 7.2 (Characterization of affine-invariant metrics) Let g be a Riemannian metric on SPD matrices. The following statements are equivalent:

1. g is $\text{GL}(n)$ -invariant,
2. g is $\text{GL}^+(n)$ -invariant,
3. g is $\text{SL}(n)$ -invariant and \mathbb{R}^+ -invariant,
4. g is $\text{SO}(n)$ -invariant and $\text{Diag}^+(n)$ -invariant,
5. g is $\mathfrak{S}(n)$ -invariant and $\text{LT}^+(n)$ -invariant.

Proof of Theorem 7.2 (Characterization of affine-invariant metrics). The first statement clearly implies the others. To prove $2 \implies 1$, we need to take a general $\text{GL}^+(n)$ -invariant metric on SPD matrices and prove that it is $\text{GL}(n)$ -invariant. It amounts to prove that $\text{SO}(n)$ -invariant inner products on symmetric matrices are $\text{O}(n)$ -invariant, which is well known so $2 \implies 1$. To prove $3 \implies 2$; $4 \implies 2$ and $5 \implies 1$, it suffices to prove that the pairs of groups respectively generate $\text{GL}^+(n)$, $\text{GL}^+(n)$ and $\text{GL}(n)$.

The group generated by $\text{SL}(n)$ and \mathbb{R}^+ is $\text{GL}^+(n)$ so $3 \implies 2$. The group generated by $\text{SO}(n)$ and $\text{Diag}^+(n)$ is also $\text{GL}^+(n)$ (it is clearly included in $\text{GL}^+(n)$ and conversely it contains $\text{Sym}^+(n)$ by the spectral theorem and $\text{GL}^+(n)$ by polar decomposition) so $4 \implies 2$.

To prove that $5 \implies 1$, let us show that the group generated by $\mathfrak{S}(n)$ and $\text{LT}^+(n)$ is $\text{GL}(n)$. First, the LU decomposition exists for any square matrix modulo a permutation. More precisely, for all $A \in \text{GL}(n)$, there exists a permutation $\sigma \in \mathfrak{S}(n)$, a lower triangular

matrix $L \in \text{LT}^*(n)$ and an upper triangular matrix $U \in \text{UT}^+(n)$ with $\text{Diag}(U) = I_n$ such that $A = P_\sigma LU$. If we permute the rows and columns of U according to the permutation $\sigma_0 : k \mapsto n + 1 - k$, then $P_{\sigma_0} U P_{\sigma_0}^\top$ is lower triangular with ones on the diagonal so it is in $\text{LT}^+(n)$. Hence the group generated by $\mathfrak{S}(n)$ and $\text{LT}^*(n)$ is $\text{GL}(n)$. Since $\text{LT}^*(n)$ is generated by $\text{LT}^+(n)$ and $\text{Diag}^*(n)$, it suffices to prove that $\text{Diag}^*(n)$ is generated by $\mathfrak{S}(n)$ and $\text{LT}^+(n)$. Since $\text{Diag}^*(n)$ is generated by $\text{Diag}^+(n)$ and matrices of the form $\text{diag}(\pm 1, \dots, \pm 1)$ and since $\text{Diag}^+(n) \subset \text{LT}^+(n)$, it suffices to prove that matrices of the form $\text{diag}(\pm 1, \dots, \pm 1)$ are generated by $\mathfrak{S}(n)$ and $\text{LT}^+(n)$. By matrix product and permutations, it suffices to prove that $\text{diag}(-1, 1, \dots, 1)$ is generated by $\mathfrak{S}(n)$ and $\text{LT}^+(n)$. The following product of matrices:

$$\mathfrak{S}(2) \ni \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{array}{c} \nearrow \\ \\ \searrow \end{array} \begin{array}{c} \overbrace{\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}}^{\in \text{LT}^+(2)} \\ \downarrow \\ \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \end{array} \begin{array}{c} \nearrow \\ \\ \searrow \end{array} \begin{array}{c} \overbrace{\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}}^{\in \text{UT}^+(2)} \\ \downarrow \\ \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix} \end{array} \begin{array}{c} \nearrow \\ \\ \searrow \end{array} \begin{array}{c} \overbrace{\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}}^{\in \text{LT}^+(2)} \\ \downarrow = \\ \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \quad (11.35)$$

can be generalized in dimension n by adding a diagonal block I_{n-2} . Hence, the matrix $\text{diag}(-1, 1, \dots, 1)$ is generated by $\mathfrak{S}(n)$ and $\text{LT}^+(n)$ so $\text{LT}^*(n)$ as well and finally $\text{GL}(n)$ entirely. \square

11.6.3 Proof of Theorem 7.4

Theorem 7.4 (Lie-Cholesky are $\text{LT}^+(n)$ -invariant metrics) A Riemannian metric on $\text{Sym}^+(n)$ is a Lie-Cholesky metric if and only if it is $\text{LT}^+(n)$ -invariant.

Proof of Theorem 7.4 (Lie-Cholesky are $\text{LT}^+(n)$ -invariant metrics). Let g be a metric on $\text{Sym}^+(n)$. Let $\Sigma \in \text{Sym}^+(n)$, $V \in T_\Sigma \text{Sym}^+(n)$, $L = \text{Chol}(\Sigma)$ and $Z = d_\Sigma \text{Chol}(X)$. Note that Equation (7.10) rewrites $L^{-1}Z = d_{I_n} \text{Chol}(L^{-1}VL^{-\top})$. Thus, by definition of the pushforward, we have the following equalities:

$$(\text{Chol}_*g)_L(Z, Z) = g_\Sigma(V, V) \quad (11.36)$$

$$(\text{Chol}_*g)_{I_n}(L^{-1}Z, L^{-1}Z) = g_{I_n}(L^{-1}VL^{-\top}, L^{-1}VL^{-\top}) \quad (11.37)$$

Hence, g is a Lie-Cholesky metric if and only if Chol_*g is left-invariant if and only if the left terms are equal if and only if the right terms are equal if and only if g is invariant under the action of $\text{LT}^+(n)$. \square

11.6.4 Proof of Lemma 7.8

Lemma 7.8 (Operator ad^* for Lie-Cholesky metrics) Let $\text{LC}(A)$ be a Lie-Cholesky metric characterized by $A \in \text{Sym}^+(\frac{n(n+1)}{2})$. Then for $X, Y \in \text{LT}(n)$, $\text{vec}(\text{ad}^*(X)(Y)) = A(I_n \otimes X^\top - X \otimes I_n)A^{-1}\text{vec}(Y)$.

Proof of Lemma 7.8 (Operator ad^ for Lie-Cholesky metrics).* In the matrix Lie group $\text{LT}(n)$, the Lie bracket is the commutator $[X, Z] = XZ - ZX$. Therefore:

$$\begin{aligned} \text{ad}^*(X)(Y) \cdot Z &= Y \cdot [X, Z] = \text{vec}(Y)^\top A \text{vec}(XZ - ZX) \\ &= \text{vec}(Y)^\top A(I_n \otimes X - X^\top \otimes I_n) \text{vec}(Z) \\ &= [\text{vec}(Y)^\top A(I_n \otimes X - X^\top \otimes I_n) A^{-1}] A \text{vec}(Z) \\ \text{vec}(\text{ad}^*(X)(Y)) &= A^{-1}(I_n \otimes X^\top - X \otimes I_n) A \text{vec}(Y). \end{aligned}$$

□

11.6.5 Proof of Lemma 7.10

Lemma 7.10 (Vertical distribution, horizontal distribution) The vertical and horizontal distributions associated to the quotient-Lie-Cholesky associated to f are $\mathcal{V}_L = \text{Diag}(n)L$ and $\mathcal{H}_L^{\text{LC}(f)} = Lf^{-1}(L^{-1}d_{LL^\top} \text{Chol}(\mathcal{S}_{(LL^\top)^{-1}}(\text{Hol}(n))))$.

Proof of Lemma 7.10 (Vertical distribution, horizontal distribution). The vertical space is the tangent space of the fiber $\text{Diag}^+(n)L$ or the kernel of the differential of the submersion $d_L\pi(Z) = \text{Diag}(L)^{-1}(Z - \text{Diag}(L^{-1}Z)L)$. Hence $\mathcal{V}_L = \text{Diag}(n)L$. Moreover, we know that when $f = \text{Id}_{\text{LT}(n)}$, the horizontal space is $\mathcal{H}_L^{\text{AI}} = d_{LL^\top} \text{Chol}(\mathcal{S}_{(LL^\top)^{-1}}(\text{Hol}(n)))$. Indeed, it is the pushforward by the Cholesky map of the horizontal space on SPD matrices (cf. Section 7.2.2). In other words, for all $Z \in \mathcal{H}_L^{\text{AI}}$, for all $Z' \in \mathcal{V}_L$, we have $\langle L^{-1}Z | L^{-1}Z' \rangle = 0$. Therefore, the horizontal space of the metric $\text{LC}(f)$ is:

$$\begin{aligned} \mathcal{H}_L^{\text{LC}(f)} &= \{Z \in \text{LT}(n) | \forall Z' \in \mathcal{V}_L, \langle f(L^{-1}Z) | L^{-1}Z' \rangle = 0\} \\ &= \{Z \in \text{LT}(n) | \forall Z' \in \mathcal{V}_L, \langle L^{-1}Lf(L^{-1}Z) | L^{-1}Z' \rangle = 0\} \\ &= \{Z \in \text{LT}(n) | Lf(L^{-1}Z) \in \mathcal{H}_L^{\text{AI}}\} \\ &= Lf^{-1}(L^{-1}\mathcal{H}_L^{\text{AI}}). \end{aligned}$$

□

11.6.6 Proof of Lemma 7.11

Lemma 7.11 (Horizontal lift, Riemannian metric, exponential map) Given the horizontal projection $\text{hor}_L : T_L\text{LT}^+(n) \rightarrow \mathcal{H}_L$ and the exponential map of the Lie-Cholesky metric $\text{Exp}^{\text{LC}} : \text{LT}(n) \rightarrow \text{LT}^+(n)$, we have for all $L \in \text{LT}^+(n)$, for all $\Gamma = \text{Diag}(L)^{-1}L \in \text{LT}^1(n)$, for all $\xi \in T_\Gamma\text{LT}^1(n) \simeq \text{LT}^0(n)$:

- (Horizontal lift) $\xi_L^\# = \text{hor}_L(\text{Diag}(L)\xi)$,
- (Riemannian metric) $g_\Gamma^{\text{QLC}}(\xi, \xi) = \langle f(\Gamma^{-1}\text{hor}_\Gamma(\xi)) | \Gamma^{-1}\text{hor}_\Gamma(\xi) \rangle$,
- (Exponential map) $\text{Exp}_\Gamma^{\text{QLC}}(t\xi) = \text{Diag}(\text{Exp}_\Gamma^{\text{LC}}(t \text{hor}(\xi)))^{-1} \text{Exp}_\Gamma^{\text{LC}}(t \text{hor}(\xi))$.

In particular, quotient-Lie-Cholesky metrics are geodesically complete.

Proof of Lemma 7.11 (Horizontal lift, Riemannian metric, exponential map). For all $\xi \in \text{LT}^0(n)$, we have $d_L \pi(\text{Diag}(L)\xi) = \text{Diag}(L)^{-1}(\text{Diag}(L)\xi - \text{Diag}(\xi)L) = \xi$. Hence $\xi_L^\# = \text{hor}_L(\text{Diag}(L)\xi)$. In particular, $\xi_\Gamma^\# = \text{hor}_\Gamma(\xi)$. Then the metric and the exponential map simply write $g_\Gamma^{\text{QLC}}(\xi, \xi) = g_\Gamma^{\text{LC}}(\xi_\Gamma^\#, \xi_\Gamma^\#)$ and $\text{Exp}_\Gamma^{\text{QLC}}(t\xi) = \pi(\text{Exp}_\Gamma^{\text{LC}}(t\xi_\Gamma^\#))$. \square

11.6.7 Proof of Theorem 7.14

Theorem 7.14 (Symmetric space structure) The manifold of full-rank correlation matrices $\text{Cor}^+(n)$ equipped with a poly-hyperbolic-Cholesky metric is a Riemannian symmetric space of non-positive sectional curvature bounded by $[a, 0]$ with $a = -\frac{1}{\min_{i \geq 2} \alpha_i}$. For $n \geq 3$, it is not of constant curvature. The canonical PHC metric writes for all $C \in \text{Cor}^+(n)$ and $X \in T_C \text{Cor}^+(n) \simeq \text{Hol}(n)$:

$$g_C^{\text{CPHC}}(X, X) = \|\text{Diag}(L)^{-1} L \text{Low}_S(L^{-1} X L^{-\top})\|^2,$$

where $L = \text{Chol}(C) \in \mathcal{L}$. The square distance between C and $C' = \phi(L')$ writes:

$$d_{\text{CPHC}}(C, C')^2 = \sum_{i=2}^n \text{arccosh}(-Q(L_i^\top, L_i'^\top))^2,$$

where L_i, L_i' are the i -th rows of L, L' respectively.

Proof of Theorem 7.14 (Symmetric space structure). The product of Riemannian symmetric spaces is a Riemannian symmetric space. The product of manifolds with sectional curvature bounded by $[a, b]$ with $a \leq 0 \leq b$ has its sectional curvature bounded by $[a, b]$. (This is also valid for $(-\infty, b]$ and $[a, +\infty)$.) Let $k \geq 2$ such that $\alpha_k = \min_{i \geq 2} \alpha_i$. The values $a = -\frac{1}{\alpha_k}$ and $b = 0$ are clearly reached for $n \geq 3$ by bivectors (X, Y) and (X, Z) respectively, where $X = (0, \dots, 0, X_k, 0, \dots, 0)$, $Y = (0, \dots, 0, Y_k, 0, \dots, 0)$ and $Z = (Z_1, 0, \dots, 0) \in T(\text{HS}^1 \times \dots \times \text{HS}^{n-1})$. To express the Riemannian metric, we take the pullback of the Riemannian metric on the hyperboloid \mathbb{H}^n defined by $g_x(v, v) = \sum_{k=1}^n v_k^2 - v_{n+1}^2$ by the diffeomorphism $\varphi^{\text{SH}} : (x_1, \dots, x_{n+1}) \in \text{HS}^n \mapsto \frac{1}{x_{n+1}}(x_1, \dots, x_n, 1) \in \mathbb{H}^n$. We compute for all $x \in \text{HS}^n$ and all $v \in T_x \text{HS}^n$, using $\sum_{k=1}^{n+1} x_k^2 = 1$ and $\sum_{k=1}^{n+1} x_k v_k = 0$:

$$\begin{aligned} d_x \varphi^{\text{SH}}(v) &= \frac{1}{x_{n+1}} \left(v_1 - x_1 \frac{v_{n+1}}{x_{n+1}}, \dots, v_n - x_n \frac{v_{n+1}}{x_{n+1}}, -\frac{v_{n+1}}{x_{n+1}} \right), \\ g_x^{\text{HS}}(v, v) &= g_{\varphi(x)}^{\mathbb{H}}(d_x \varphi^{\text{SH}}(v), d_x \varphi^{\text{SH}}(v)) \\ &= \frac{1}{x_{n+1}^2} \left(\sum_{k=1}^n \left(v_k - x_k \frac{v_{n+1}}{x_{n+1}} \right)^2 - \frac{v_{n+1}^2}{x_{n+1}^2} \right) \\ &= \frac{1}{x_{n+1}^2} \left(\sum_{k=1}^n \left(v_k^2 - 2x_k v_k \frac{v_{n+1}}{x_{n+1}} + x_k^2 \frac{v_{n+1}^2}{x_{n+1}^2} \right) - \frac{v_{n+1}^2}{x_{n+1}^2} \right) \\ &= \frac{1}{x_{n+1}^2} \left(\sum_{k=1}^n v_k^2 + 2v_{n+1}^2 + (1 - x_{n+1}^2) \frac{v_{n+1}^2}{x_{n+1}^2} - \frac{v_{n+1}^2}{x_{n+1}^2} \right) \\ &= \frac{\|v\|^2}{x_{n+1}^2}. \end{aligned}$$

Hence, for all $C \in \text{Cor}^+(n)$, $X \in \text{Hol}(n)$, $L = \text{Chol}(C) \in \mathcal{L}$ and $Y = d_C \text{Chol}(X) = L \text{Lows}(L^{-1}XL^{-\top})$ with $Y_{11} = \frac{1}{2}L_{11}[L^{-1}XL^{-\top}]_{11} = \frac{X_{11}}{2L_{11}} = 0$, we have:

$$\begin{aligned} g_C^{\text{PHC}}(X, X) &= g_L^{\text{HS}^1 \times \dots \times \text{HS}^{n-1}}(Y, Y) \\ &= \sum_{i=2}^n \frac{\|Y_{i\bullet}\|^2}{L_{ii}^2} = \sum_{i=1}^n \|\text{Diag}(L)_{ii}^{-1} Y_{i\bullet}\|^2 \\ &= \|\text{Diag}(L)^{-1} Y\|^2 = \|\text{Diag}(L)^{-1} L \text{Lows}(L^{-1}XL^{-\top})\|^2. \end{aligned}$$

Note that the general PHC metric writes $g_C^{\text{PHC}}(X, X) = \sum_{i=2}^n \|\text{Diag}(L)_{ii}^{-1} \alpha_i Y_{i\bullet}\|^2$ and the general PHC distance writes $d(C, C')^2 = \sum_{i=2}^n \alpha_i \arccos(-Q(L_i^\top, L'_i{}^\top))^2$. \square

11.6.8 Proof of Theorem 7.19

Theorem 7.19 (Group operations) The group operations associated to the Lie-Cholesky group structure on full-rank correlation matrices are, for all $C, C', C_i \in \text{Cor}^+(n)$, $X \in T_C \text{Cor}^+(n) \simeq \text{Hol}(n)$, $t \in \mathbb{R}$:

- (Exponential map) $\text{Exp}_C(tX) = \Theta^{-1}(\Theta(C) \exp(t \Theta(C)^{-1} d_C \Theta(X)))$,
- (Logarithm map) $\text{Log}_C(C') = (d_C \Theta)^{-1}(\Theta(C) \log(\Theta(C)^{-1} \Theta(C')))$,
- (Geodesic) $\gamma_{C \rightarrow C'}(t) = \Theta^{-1}(\Theta(C)(\Theta(C)^{-1} \Theta(C'))^t)$,
- (Group mean) Unique, characterized by $\sum_{i=1}^k \log(\Theta(\bar{C})^{-1} \Theta(C_i)) = 0$.

Proof of Theorem 7.19 (Group operations). The exponential map, logarithm map and geodesics are pullbacks by Θ of corresponding operations in $\text{LT}^1(n)$, which are for all $\Gamma, \Gamma' \in \text{LT}^+(n)$, $\xi \in T_\Gamma \text{LT}^+(n) \simeq \text{LT}(n)$ and $t \in \mathbb{R}$:

- (Exponential map) $\text{Exp}_\Gamma(t\xi) = \Gamma \exp(t \Gamma^{-1} \xi)$,
- (Logarithm map) $\text{Log}_\Gamma(\Gamma') = \Gamma \log(\Gamma^{-1} \Gamma')$,
- (Geodesic) $\gamma_{\Gamma \rightarrow \Gamma'}(t) = \text{Exp}_\Gamma(t \Gamma^{-1} \text{Log}_\Gamma(\Gamma')) = \Gamma(\Gamma^{-1} \Gamma')^t$.

Since the Lie algebra $\text{LT}^0(n)$ is nilpotent, the group mean $\bar{\Gamma}$ of the finite sample $\Gamma_1, \dots, \Gamma_k \in \text{LT}^1(n)$ is unique [Buser and Karcher, 1981, Example 8.1.8]. It is characterized by $0 = \sum_{i=1}^k \text{Log}_{\bar{\Gamma}}(\Gamma_i) = \bar{\Gamma} \sum_{i=1}^k \log(\bar{\Gamma}^{-1} \Gamma_i)$, which is equivalent to $\sum_{i=1}^k \log(\bar{\Gamma}^{-1} \Gamma_i) = 0$. \square

11.6.9 Proof of Theorem 7.20

Theorem 7.20 (Geodesics in dimension 2) Let $C_1 = C(\rho_1), C_2 = C(\rho_2) \in \text{Cor}^+(2)$ with $\rho_1, \rho_2 \in (-1, 1)$.

1. Quotient-affine metrics and poly-hyperbolic-Cholesky metrics coincide (up to a scaling factor). The geodesic between C_1 and C_2 is $C(\rho(t))$ for $t \in \mathbb{R}$ where:

$$\rho(t) = \frac{\rho_1 \cosh(\lambda t) + \sinh(\lambda t)}{\rho_1 \sinh(\lambda t) + \cosh(\lambda t)},$$

where $\lambda = \log \sqrt{\frac{1+\rho_2}{1-\rho_2}} - \log \sqrt{\frac{1+\rho_1}{1-\rho_1}}$ is known as the difference of the Fisher transformation of the correlation coefficients ρ_1 and ρ_2 .

2. Euclidean-Cholesky and log-Euclidean-Cholesky metrics coincide. The geodesic between C_1 and C_2 is $C(\rho(t))$ for $t \in \mathbb{R}$ where:

$$\rho(t) = \frac{F(t)}{\sqrt{1 + F(t)^2}},$$

where $F(t) = (1-t)\frac{\rho_1}{\sqrt{1-\rho_1^2}} + t\frac{\rho_2}{\sqrt{1-\rho_2^2}}$. This geodesic also coincides with the Lie-Cholesky group geodesic of Section 7.4.4.

Proof of Theorem 7.20. 1. The formula of the geodesic is known for the quotient-affine metrics in dimension 2. Hence it suffices to show that the quotient-affine metrics and the poly-hyperbolic-Cholesky metrics coincide up to a scaling factor. Let $C = C(\rho)$ and $X = \begin{pmatrix} 0 & x \\ x & 0 \end{pmatrix} \in T_C \text{Cor}^+(2)$. We compute the quotient-affine metric $g_C^{\text{QA}}(X, X) = \text{tr}(C^{-1}XC^{-1}X) - 2 \text{sum}(D(I_n + C \bullet C^{-1})^{-1}D)$ and the canonical PHC metric $g_C^{\text{CPHC}}(X, X) = \|\text{Diag}(L)^{-1}L \text{Low}_S(L^{-1}XL^{-\top})\|^2$ where $D = \text{Diag}(C^{-1}X)$ and $L = \text{Chol}(C)$.

$$C^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix},$$

$$C^{-1}X = \frac{x}{1-\rho^2} \begin{pmatrix} -\rho & 1 \\ 1 & -\rho \end{pmatrix},$$

$$C^{-1}XC^{-1}X = \frac{x^2}{(1-\rho^2)^2} \begin{pmatrix} 1+\rho^2 & -2\rho \\ -2\rho & 1+\rho^2 \end{pmatrix},$$

$$\text{tr}(C^{-1}XC^{-1}X) = \frac{2(1+\rho^2)}{(1-\rho^2)^2}x^2,$$

$$I_n + C \bullet C^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 2-\rho^2 & -\rho^2 \\ -\rho^2 & 2-\rho^2 \end{pmatrix},$$

$$(I_n + C \bullet C^{-1})^{-1} = \frac{1}{4} \begin{pmatrix} 2-\rho^2 & \rho^2 \\ \rho^2 & 2-\rho^2 \end{pmatrix},$$

$$D = \text{Diag}(C^{-1}X) = -\frac{\rho x}{1-\rho^2}I_2,$$

$$\text{sum}(D(I_n + C \bullet C^{-1})^{-1}D) = \frac{\rho^2 x^2}{(1-\rho^2)^2},$$

$$g_C(X, X) = \frac{2x^2}{(1-\rho^2)^2},$$

$$\begin{aligned}
L &= \text{Chol}(C) = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}, \\
L^{-1}XL^{-\top} &= \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} \begin{pmatrix} 0 & x \\ x & 0 \end{pmatrix} \begin{pmatrix} 1 & -\frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} \\
&= \begin{pmatrix} 0 & x \\ \frac{x}{\sqrt{1-\rho^2}} & -\frac{\rho x}{\sqrt{1-\rho^2}} \end{pmatrix} \begin{pmatrix} 1 & -\frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} \\
&= \begin{pmatrix} 0 & \frac{x}{\sqrt{1-\rho^2}} \\ \frac{x}{\sqrt{1-\rho^2}} & -\frac{2\rho x}{1-\rho^2} \end{pmatrix},
\end{aligned}$$

$$\begin{aligned}
\text{Diag}(L)^{-1}L \text{Low}_S(L^{-1}XL^{-\top}) &= \begin{pmatrix} 1 & 0 \\ \frac{\rho}{\sqrt{1-\rho^2}} & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \frac{x}{\sqrt{1-\rho^2}} & -\frac{\rho x}{1-\rho^2} \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ \frac{x}{\sqrt{1-\rho^2}} & -\frac{\rho x}{1-\rho^2} \end{pmatrix}, \\
g_C^{\text{CPHC}}(X, X) &= \left(\frac{1}{1-\rho^2} + \frac{\rho^2}{(1-\rho^2)^2} \right) x^2 \\
&= \frac{x^2}{(1-\rho^2)^2}.
\end{aligned}$$

2. Euclidean-Cholesky and log-Euclidean-Cholesky metrics coincide in dimension 2 because $\exp(\xi) = I_n + \xi$ and $\log(\Gamma) = \Gamma - I_2$ for $\xi \in \text{LT}^0(2)$ and $\Gamma \in \text{LT}^1(2)$. Thus their common Riemannian exponential and logarithm in $\text{LT}^1(2)$ are simply $\text{Exp}_\Gamma(\xi) = \Gamma + \xi$ and $\text{Log}_\Gamma(\Gamma') = \Gamma' - \Gamma$. On the other hand, the group exponential in $\text{LT}^1(2)$ is $\text{Exp}_\Gamma^{\text{LT}^1(2)}(\xi) = \Gamma \exp(\Gamma^{-1}\xi) = \Gamma(I_n + \Gamma^{-1}\xi) = \Gamma + \xi = \text{Exp}_\Gamma(\xi)$. Hence, the group geodesics coincide with the (log-)Euclidean-Cholesky geodesics. Let us compute them.

$$\begin{aligned}
\Gamma_1 &:= \text{Diag}(\text{Chol}(C_1))^{-1}\text{Chol}(C_1) = \begin{pmatrix} 1 & 0 \\ \frac{\rho_1}{\sqrt{1-\rho_1^2}} & 1 \end{pmatrix}, \\
\Gamma_2 &:= \text{Diag}(\text{Chol}(C_2))^{-1}\text{Chol}(C_2) = \begin{pmatrix} 1 & 0 \\ \frac{\rho_2}{\sqrt{1-\rho_2^2}} & 1 \end{pmatrix}, \\
C(t) &= \Theta^{-1}((1-t)\Gamma_1 + t\Gamma_2) = \Theta^{-1} \begin{pmatrix} 1 & 0 \\ F(t) & 1 \end{pmatrix} \\
&= \text{Cor} \begin{pmatrix} 1 & F(t) \\ F(t) & 1 + F(t)^2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{F(t)}{\sqrt{1+F(t)^2}} \\ \frac{F(t)}{\sqrt{1+F(t)^2}} & 1 \end{pmatrix}.
\end{aligned}$$

We can also compute the (log-)Euclidean-Cholesky metric in dimension 2.

$$\begin{aligned}
g_C(X, X) &= \|d_C\Theta(X)\|^2 = \|(\Theta \circ C)'(\rho)\|^2 \\
&= f'(\rho)^2 x^2,
\end{aligned}$$

where $f(\rho) = \frac{\rho}{\sqrt{1-\rho^2}}$. So $f'(\rho) = \frac{\sqrt{1-\rho^2} + \frac{\rho^2}{\sqrt{1-\rho^2}}}{1-\rho^2} = \frac{1}{(1-\rho^2)^{3/2}}$ and:

$$g_C(X, X) = \frac{x^2}{(1-\rho^2)^3}.$$

□

11.7 Proofs of Chapter 8

In this section, we prove the results of Chapter 8: Permutation-invariant Log-Euclidean metrics.

11.7.1 Proof of Theorem 8.5

Theorem 8.5 (Incompatibility between cor-inversion and off-log bijection) Let $n \geq 3$. There exists $C \in \text{Cor}^+(n)$, such that $\text{Log}(\mathcal{I}(C)) \neq -\text{Log}(C)$. Otherwise said, the following diagram does not commute.

$$\begin{array}{ccc} \text{Cor}^+(n) & \xrightarrow{\mathcal{I}} & \text{Cor}^+(n) \\ \downarrow \text{Log} & & \downarrow \text{Log} \\ \text{Hol}(n) & \xrightarrow{-\text{Id}} & \text{Hol}(n) \end{array}$$

Proof of Theorem 8.5 (Incompatibility between cor-inversion and off-log bijection). It is easy to see it numerically. For a formal proof, one can look for a matrix $C \in \text{Cor}^+(3)$ such that $\log(C)$ and $\log(\mathcal{I}(C))$ are easy to compute manually. We propose the following example, with $x = \frac{1}{\sqrt{7}}$ and $a = \frac{x}{\sqrt{1-x^2}} = \frac{1}{\sqrt{6}}$:

$$C = \begin{pmatrix} 1 & x & -x \\ x & 1 & 0 \\ -x & 0 & 1 \end{pmatrix} = PDP^\top \text{ with } P = \frac{1}{2} \begin{pmatrix} 0 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & 1 & -1 \\ \sqrt{2} & -1 & 1 \end{pmatrix}, D = \text{diag}(1, 1 + \frac{\sqrt{2}}{\sqrt{7}}, 1 - \frac{\sqrt{2}}{\sqrt{7}}),$$

$$\mathcal{I}(C) = \begin{pmatrix} 1 & -a & a \\ -a & 1 & -a^2 \\ a & -a^2 & 1 \end{pmatrix} = Q\Delta Q^\top \text{ with } Q = \frac{1}{\sqrt{14}} \begin{pmatrix} 0 & \sqrt{6} & 2\sqrt{2} \\ \sqrt{7} & -2 & \sqrt{3} \\ \sqrt{7} & 2 & -\sqrt{3} \end{pmatrix}, \Delta = \frac{1}{6} \text{diag}(5, 10, 3).$$

Thus:

$$\begin{aligned} [\text{Log}(C) + \text{Log}(\mathcal{I}(C))]_{12} &= [\log(C) + \log(\mathcal{I}(C))]_{12} \\ &= [P \log(D) P^\top + Q \log(\Delta) P^\top]_{12} \\ &= \sum_{k=1}^3 (\ln(D_{kk}) P_{1k} P_{2k} + \ln(\Delta_{kk}) Q_{1k} Q_{2k}) \\ &= \frac{1}{2\sqrt{2}} \ln \left(\frac{\sqrt{7} + \sqrt{2}}{\sqrt{7} - \sqrt{2}} \right) + \frac{\sqrt{6}}{7} \ln \left(\frac{3}{10} \right) > 0. \end{aligned}$$

For $n \geq 4$, it suffices to take the block diagonal matrix $\text{Diag}(C, I_{n-3})$.

□

11.7.2 Proof of Theorem 8.6

Theorem 8.6 (Log = Off \circ log is a diffeomorphism) The off-log bijection $\text{Log} : \text{Cor}^+(n) \rightarrow \text{Hol}(n)$ is a smooth diffeomorphism. We give the differentials of Log and Exp in function of the differentials of the symmetric matrix logarithm and exponential maps log and exp. For all $C \in \text{Cor}^+(n)$ and $S, X, Y \in \text{Hol}(n)$:

$$\begin{aligned} d_C \text{Log}(X) &= \text{Off}(d_C \log(X)), \\ d_S \text{Exp}(Y) &= d_{\log(\text{Exp}(S))} \exp(Y - \text{Diag}(\text{Exp}(S)Y)). \end{aligned}$$

Proof of Theorem 8.6 (Log = Off \circ log is a diffeomorphism). It suffices to show that \mathcal{D} is smooth. We use the implicit function theorem with the smooth map $\Phi : (S, D) \in \text{Hol}(n) \times \text{Diag}(n) \mapsto \text{Diag}(\exp(D + S)) - I_n \in \text{Diag}(n)$ which is such that $D = \mathcal{D}(S)$ if and only if $\Phi(S, D) = 0$. We want to show that for $(S, D) \in \Phi^{-1}(0)$, the differential of $\Phi_S : D \in \text{Diag}(n) \mapsto \Phi(S, D) \in \text{Diag}(n)$ is invertible. We introduce the intermediate smooth map $\varphi_S : D \in \text{Diag}(n) \mapsto D - \log \text{Diag} \exp(D + S) \in \text{Diag}(n)$ as in [Archakov and Hansen, 2021] so that $\Phi_S(D) = \exp(D - \varphi_S(D)) - I_n$. It is proved in the Appendix of [Archakov and Hansen, 2021] that for all $D \in \text{Diag}(n)$, the linear map $d_D \varphi_S : \text{Diag}(n) \rightarrow \text{Diag}(n)$ has its eigenvalues in $[0, 1) \subset \mathbb{R}$. For all $(S, D) \in \Phi^{-1}(0)$ and all $\Delta \in \text{Diag}(n)$, we have:

$$\begin{aligned} d_D \Phi_S(\Delta) &= \underbrace{\exp(D - \varphi_S(D))}_{I_n} (\Delta - d_D \varphi_S(\Delta)) \\ &= (\text{Id} - d_D \varphi_S)(\Delta). \end{aligned}$$

Therefore, the linear map $d_D \Phi_S = \text{Id} - d_D \varphi_S$ has its eigenvalues in $(0, 1]$ so it is invertible. Hence, the implicit function $\mathcal{D} : \text{Sym}(n) \rightarrow \text{Diag}(n)$ is smooth so $\pi = \exp \circ (\mathcal{D} + \text{Id}_{\text{Sym}(n)}) : \text{Sym}(n) \rightarrow \text{Cor}^+(n)$ is smooth. Then $\text{Exp} = \pi|_{\text{Hol}(n)} : \text{Hol}(n) \rightarrow \text{Cor}^+(n)$ is smooth and Log is a smooth diffeomorphism.

Since $\text{Log} = \text{Off} \circ \log$ and Off is linear, the differential of Log is clear. Let us compute the differential of f , \mathcal{D} and Exp. Let $S \in \text{Sym}(n)$, $C = \text{Exp}(S) \in \text{Cor}^+(n)$, $D, \Delta \in \text{Diag}(n)$, $X \in \text{Hol}(n)$ and $Y = d_C \text{Log}(X) \in \text{Hol}(n)$.

$$\begin{aligned} d_{(S,D)} \Phi(Y, \Delta) &= \text{Diag}(\exp(D + S)(\Delta + Y)) \\ &= \text{Diag}(\exp(D + S))\Delta + \text{Diag}(\exp(D + S)Y), \\ d_{(S,\mathcal{D}(S))} \Phi(Y, \Delta) &= \Delta + \text{Diag}(CY), \\ \underbrace{d_{(S,\mathcal{D}(S))} \Phi(Y, d_S \mathcal{D}(Y))}_{=d_S(\Phi \circ (\text{Id}, \mathcal{D}))(Y)=0} &= d_S \mathcal{D}(Y) + \text{Diag}(CY), \\ d_S \mathcal{D}(Y) &= -\text{Diag}(CY) = -\text{Diag}(\text{Exp}(S)Y), \\ d_S \text{Exp}(Y) &= d_{\mathcal{D}(S)+S} \exp(d_S \mathcal{D}(Y) + Y) \\ &= d_{\log(C)} \exp(Y - \text{Diag}(\text{Exp}(S)Y)). \end{aligned}$$

□

11.7.3 Proof of Theorem 8.13

Theorem 8.13 (Existence and equivalence of conjectures) We define the smooth map $f : \Delta \in \text{Diag}^+(n) \mapsto d^{\text{AI}}(I_n, \Delta C \Delta)^2 = \text{tr}(\log(\Delta C \Delta)^2)$. It gives the affine-invariant squared distance

between I_n and all points of the fiber $\text{Cor}^{-1}(C) = \{\Delta C \Delta \in \text{Sym}^+(n) | \Delta \in \text{Diag}^+(n)\}$.

1. The smooth map f has a global minimizer.
2. For all $\Sigma \in \text{Sym}^+(n)$, there exists $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta \Sigma \Delta) \in \text{Hol}(n)$.
3. The following conjectures are equivalent for all $C \in \text{Cor}^+(n)$.
 - (i) There exists a unique $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta C \Delta) \in \text{Hol}(n)$ (Conjecture 8.11).
 - (ii) There exists a unique $X \in \text{Hol}(n)$ such that $\text{Exp}_{I_n}^{\text{QA}}(X) = C$.
 - (iii) There exists a unique local minimizer of the smooth map f , which is actually the global minimizer ensured by statement 1.
4. The previous conjectures imply the uniqueness of the quotient-affine logarithm at I_n .

Proof of Theorem 8.13 (Existence and equivalence of conjectures).

1. The smooth map f has a global minimizer because it is coercive (Theorem 6.6).
2. Hence, there exists $\Delta \in \text{Diag}^+(n)$ such that $f(\Delta) = \min f$. In other words, $\Delta C \Delta$ is “in optimal position” to I_n [Huckemann et al., 2010, Definition 2.3]. Thus [Huckemann et al., 2010, Theorem 2.4], the geodesic from I_n to $\Delta C \Delta$ is horizontal, i.e. $\text{Log}_{I_n}^{\text{AI}}(\Delta C \Delta) \in \mathcal{H}_{I_n}^{\text{QA}}$, i.e. $\log(\Delta C \Delta) \in \text{Hol}(n)$. For $\Sigma = \text{Diag}(\Sigma)^{1/2} C \text{Diag}(\Sigma)^{1/2}$, it suffices to take $\Delta \text{Diag}(\Sigma)^{-1/2} \in \text{Diag}^+(n)$.
3. Since $\text{Exp}_{I_n}^{\text{QA}} = \text{Cor} \circ \text{exp}$, we clearly have (i) \iff (ii) because $X = \log(\Delta C \Delta)$ and $\Delta = \text{Diag}(\text{exp}(X))^{1/2}$. To prove (i) \iff (iii), let us compute the differential and the Hessian of f . Let $\Delta \in \text{Diag}^+(n)$ and $D, D' \in \text{Diag}(n)$. We denote $E = D \Delta^{-1} \in \text{Diag}(n)$ and $A = \Delta C \Delta = P B P^\top \in \text{Sym}^+(n)$ with $P \in \text{O}(n)$ and $B \in \text{Diag}^+(n)$.

$$\begin{aligned}
 d_\Delta f(D) &= 2 \text{tr}(\log(\Delta C \Delta) d_{\Delta C \Delta} \log((D C \Delta + \Delta C D))) \\
 &= 2 \text{tr}(\log(A) d_A \log(EA + AE^\top)) \\
 &= 2 \text{tr}(\log(B) d_B \log(P^\top E P B + B P^\top E^\top P)) \\
 &= 2 \text{tr}(\log(B) B^{-1} (P^\top E P B + B P^\top E^\top P)) \\
 &= 4 \text{tr}(\log(A) E) \\
 &= 4 \text{tr}(\log(\Delta C \Delta) D \Delta^{-1}), \\
 d_\Delta f = 0 &\iff \text{Diag}(\log(\Delta C \Delta)) = 0 \iff \log(\Delta C \Delta) \in \text{Hol}(n), \\
 H_\Delta f(D, D') &= 4 \text{tr}(d_A \log(EA + AE^\top) D' \Delta^{-1} - \log(A) D' \Delta^{-1} D \Delta^{-1}) \\
 &= 4 \text{tr}(2 \Delta^{-2} D D' - \text{Diag}(\log(\Delta C \Delta)) \Delta^{-2} D D'), \\
 H_\Delta f(D, D) &= 4 \text{tr}((2I_n - \text{Diag}(\log(\Delta C \Delta))) \Delta^{-2} D^2).
 \end{aligned}$$

Hence, if $\Delta \in \text{Diag}^+(n)$ is such that $\log(\Delta C \Delta) \in \text{Hol}(n)$, then the Hessian of f at Δ is positive definite so f has a local minimum at Δ . Thus, if proposition (iii) is true, then Δ has to be the global minimizer so it is unique. Conversely, if f has a local minimum at $\Delta \in \text{Diag}^+(n)$, then $d_\Delta f = 0$ so $\log(\Delta C \Delta) \in \text{Hol}(n)$. Thus, if proposition (i) is true, then Δ is unique. Therefore, assertions (i) and (iii) are equivalent.

4. A quotient-affine logarithm of $C \in \text{Cor}^+(n)$ at I_n is a tangent vector $X \in \text{Hol}(n)$ of minimal length such that $\text{Exp}_{I_n}^{\text{QA}}(X) = \text{Cor}(\exp(X)) = C$. Otherwise said, it is a tangent vector $X = \log(\Delta C \Delta)$ where Δ minimizes $f(\Delta) = \text{tr}(X^2) = \|X\|^2$. Thus the uniqueness in the conjectures of statement 3 imply the uniqueness of the quotient-affine logarithm at I_n . □

11.7.4 Proof of Theorem 8.17

Theorem 8.17 (Properties of the log-scaling bijection) The log-scaling bijection satisfies the following properties.

1. (Equivariance) Log^* and Exp^* are equivariant under permutations.
2. (Equicorrelation) For all $\rho \in (-\frac{1}{n-1}, 1)$, $\text{Log}^*(C(\rho)) = \frac{1}{n} \ln \left(\frac{1+(n-1)\rho}{1-\rho} \right) (\mathbf{1}\mathbf{1}^\top - nI_n)$. In dimension $n = 2$, $\text{Log}^*(C(\rho)) = \begin{pmatrix} -F(\rho) & F(\rho) \\ F(\rho) & -F(\rho) \end{pmatrix}$ where $F(\rho) = \frac{1}{2} \log(\frac{1+\rho}{1-\rho}) \in \mathbb{R}$ is the Fisher transformation of the correlation coefficient $\rho \in (-1, 1)$.
3. (Block equicorrelation matrix) If C is a block equicorrelation matrix of signature $I = \{i_1, \dots, i_p\}$, then $\text{Log}^*(C)$ is a block symmetric matrix with null row sum of signature I with diagonal blocks of the form $(\alpha_j - \beta_j)I_{i_j} + \beta_j \mathbf{1}_{i_j} \mathbf{1}_{i_j}^\top$ and off-diagonal blocks of the form $\beta_{jk} \mathbf{1}_{i_j} \mathbf{1}_{i_k}^\top$.
4. (Generalization) For all $x \in (\mathbb{R}^+)^n$, for all $\Sigma \in \text{Sym}^+(n)$, there exists a unique $\Delta \in \text{Diag}^+(n)$ such that $\log(\Delta \Sigma \Delta)x = 0$.

Proof of Theorem 8.17 (Properties of the log-scaling bijection). 1. This is clear.

2. Let $C = C(\rho)$. The result is clear for $\rho = 0$ so we assume that $\rho \neq 0$. One easily checks that $\Delta = \sqrt{a}I_n$ with $a = \frac{1}{1+(n-1)\rho}$ satisfies $\Sigma := \Delta C \Delta = aC = (a-b)I_n + b\mathbf{1}\mathbf{1}^\top \in \text{Row}_\mathbf{1}^+(n)$ with $b = a\rho = \frac{\rho}{1+(n-1)\rho}$. Indeed, $a + (n-1)b = 1$. Since $\log(\Sigma)$ is a polynomial in Σ , there exists $(\alpha, \beta) \in \mathbb{R}^2$ such that $\log(\Sigma) = (\alpha - \beta)I_n + \beta\mathbf{1}\mathbf{1}^\top$ and $\alpha + (n-1)\beta = 0$. Moreover, $\text{eig}(\Sigma) = \{1; a-b\}$ with $a-b \neq 1$ and $\text{eig}(\log(\Sigma)) = \{0; \alpha - \beta\}$ so $\alpha - \beta = \ln(a-b) = -\ln\left(\frac{1+(n-1)\rho}{1-\rho}\right)$. Therefore, $n\beta = -(\alpha - \beta) = \ln\left(\frac{1+(n-1)\rho}{1-\rho}\right)$ and:

$$\text{Log}^*(C) = (\alpha - \beta)I_n + \beta\mathbf{1}\mathbf{1}^\top = \beta(\mathbf{1}\mathbf{1}^\top - nI_n) = \frac{1}{n} \ln \left(\frac{1+(n-1)\rho}{1-\rho} \right) (\mathbf{1}\mathbf{1}^\top - nI_n).$$

3. If C is a block equicorrelation matrix of signature $I = \{i_1, \dots, i_p\}$, it is clear that the matrix Δ is a block diagonal matrix of signature I with scalar blocks because the sums of all the rows belonging to the same interval $[i_j+1; i_{j+1}]$ are equal. The matrix product preserves the signature and the form of the blocks so the logarithm as well.
4. Similarly to Theorem 8.14, one can prove that given $x \in (\mathbb{R}^+)^n$, the map $\exp : \{S \in \text{Sym}(n) | Sx = 0\} \rightarrow \{\Sigma \in \text{Sym}^+(n) | \Sigma x = x\}$ is a diffeomorphism. We denote $X = \text{diag}(x) \in \text{Diag}^+(n)$. Therefore, for all $\Delta \in \text{Diag}^+(n)$, $\log(\Delta \Sigma \Delta)x = 0$ if and only if $\Delta \Sigma \Delta x = x$ if and only if $X \Delta \Sigma \Delta X \mathbf{1} = X^2 \mathbf{1}$ if and only if $X \Delta \in \text{Diag}^+(n)$ scales Σ onto

an SPD matrix with row sums prescribed by $X^2\mathbf{1}$. Thus the existence and uniqueness is ensured by [Marshall and Olkin, 1968]. \square

11.7.5 Proof of Theorem 8.19

Theorem 8.19 ($\text{Exp}^* = \text{Cor} \circ \text{exp}$ is a diffeomorphism) The log-scaling bijection $\text{Log}^* : \text{Cor}^+(n) \rightarrow \text{Row}_0(n)$ is a smooth diffeomorphism. We give the differentials of Log^* and Exp^* in function of the differentials of the symmetric matrix logarithm and exponential maps \log and \exp . For all $C \in \text{Cor}^+(n)$, $S, Y \in \text{Row}_0(n)$ and $X \in \text{Hol}(n)$ such that $\Sigma = \mathcal{D}^*(C) \star C = \exp(S)$:

$$d_S \text{Exp}^*(Y) = \Delta^{-1} \left[d_S \exp(Y) - \frac{1}{2}(\Delta^{-2} \text{Diag}(d_S \exp(Y)) \Sigma + \Sigma \text{Diag}(d_S \exp(Y)) \Delta^{-2}) \right] \Delta^{-1},$$

$$d_C \text{Log}^*(X) = d_\Sigma \log \left(\Delta X \Delta + \frac{1}{2}(X^0 \Sigma + \Sigma X^0) \right),$$

where $\Delta = \text{Diag}(\Sigma)^{1/2}$ and $X^0 = -2 \text{diag}((I_n + \Sigma)^{-1} \Delta X \Delta \mathbf{1})$.

Proof of Theorem 8.19 ($\text{Exp}^* = \text{Cor} \circ \text{exp}$ is a diffeomorphism). It suffices to show that \mathcal{D}^* is smooth. We apply the implicit function theorem to the smooth function $\Phi^* : (\Sigma, \Delta) \in \text{Sym}^+(n) \times \text{Diag}^+(n) \mapsto \Delta \Sigma \Delta \mathbf{1} - \mathbf{1} \in (\mathbb{R}^+)^n$ which satisfies $\Delta = \mathcal{D}^*(\Sigma)$ if and only if $\Phi^*(\Sigma, \Delta) = 0$. Let us prove that for all $(\Sigma, \Delta) \in (\Phi^*)^{-1}(0)$, the differential of the partial function $\Phi_\Sigma^* : \Delta \in \text{Diag}^+(n) \mapsto \Phi^*(\Sigma, \Delta) \in (\mathbb{R}^+)^n$ is invertible. In the direction $D \in T_\Delta \text{Diag}^+(n) = \text{Diag}(n)$:

$$\begin{aligned} d_\Delta \Phi_\Sigma^*(D) &= D \Sigma \Delta \mathbf{1} + \Delta \Sigma D \mathbf{1} \\ &= D \Delta^{-1} \mathbf{1} + \Delta \Sigma D \mathbf{1} \\ &= \Delta(\Delta^{-2} + \Sigma) D \mathbf{1}. \end{aligned}$$

Since $\Delta(\Delta^{-2} + \Sigma) \in \text{GL}(n)$, the differential is invertible so \mathcal{D}^* is smooth.

Since $\text{Exp}^* = \text{Cor} \circ \text{exp}$, we have $d_S \text{Exp}^*(Y) = d_\Sigma \text{Cor}(d_S \exp(Y))$ with $\Sigma = \exp(S)$. Using $d_\Sigma \text{Cor}(Z) = \Delta^{-1} \left[Z - \frac{1}{2}(\Delta^{-2} \text{Diag}(Z) \Sigma + \Sigma \text{Diag}(Z) \Delta^{-2}) \right] \Delta^{-1}$ with $\Delta = \text{Diag}(\Sigma)^{1/2}$ and $Z = d_S \exp(Y) \in T_\Sigma \text{Row}_1^+(n) = \text{Row}_0(n)$, we get the expected result. Now we want to invert the relation $X = d_S \text{Exp}^*(Y)$ to get $Y = d_\Sigma \text{Log}^*(X)$. We use the intermediate matrices $\Sigma \in \text{Row}_1^+(n)$ and $Z \in \text{Row}_0(n)$ and the relations $X = d_\Sigma \text{Cor}(Z)$ and $Y = d_\Sigma \log(Z)$.

$$\begin{aligned} \Delta X \Delta &= Z - \frac{1}{2}(\Delta^{-2} \text{Diag}(Z) \Sigma + \Sigma \text{Diag}(Z) \Delta^{-2}), \\ \Delta X \Delta \mathbf{1} &= -\frac{1}{2}(\Delta^{-2} \text{Diag}(Z) \Sigma \mathbf{1} + \Sigma \text{Diag}(Z) \Delta^{-2} \mathbf{1}) \\ &= -\frac{1}{2}(I_n + \Sigma) \Delta^{-2} \text{Diag}(Z) \mathbf{1}, \\ \Delta^{-2} \text{Diag}(Z) \mathbf{1} &= -2(I_n + \Sigma)^{-1} \Delta X \Delta \mathbf{1}, \\ \Delta^{-2} \text{Diag}(Z) &= X^0, \\ Z &= \Delta X \Delta + \frac{1}{2}(X^0 \Sigma + \Sigma X^0), \end{aligned}$$

which allows to conclude with $d_\Sigma \text{Log}^*(X) = Y = d_\Sigma \log(Z)$. \square

11.7.6 Proof of Theorem 8.20

Theorem 8.20 (Permutation-invariant inner products on $\text{Row}_0(n)$) For $n \geq 4$, permutation-invariant inner products on $\text{Row}_0(n)$ are the symmetric bilinear forms associated to the following positive definite quadratic forms q^* defined for $Y \in \text{Row}_0(n)$:

$$q^*(Y) = \alpha \text{tr}(Y^2) + \delta \text{tr}(\text{Diag}(Y)^2) + \zeta \text{tr}(Y)^2, \quad (11.38)$$

with $\alpha > 0$, $n\alpha + (n-2)\delta > 0$ and $n\alpha + (n-1)(\delta + n\zeta) > 0$. For $n = 3$, the permutation-invariant inner products have the same form with $\alpha = 0$. For $n = 2$, they have the same form with $\alpha = \delta = 0$.

Proof of Theorem 8.20 (Permutation-invariant inner products on $\text{Row}_0(n)$). We rely on the characterization of permutation-invariant inner products on $\text{Hol}(n)$ and on the equivariant isomorphism $\Theta : \begin{cases} \text{Hol}(n) & \longrightarrow & \text{Row}_0(n) \\ X & \longmapsto & Y = AXA^\top \end{cases}$ with $A = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ found in [Kurata and Bapat, 2016], whose inverse isomorphism is given by $X = \Theta^{-1}(Y) = Y - (\mu\mathbf{1}^\top + \mathbf{1}\mu^\top)$ with $\mu = \frac{1}{2}\text{Diag}(Y)\mathbf{1}$. Indeed, let $q^* : \text{Row}_0(n) \longrightarrow \mathbb{R}$ be a permutation-invariant quadratic form. Then q^* is positive definite if and only if $q^* \circ \Theta : \text{Hol}(n) \longrightarrow \mathbb{R}$ is a permutation-invariant quadratic form on $\text{Hol}(n)$. Hence, by Theorem 8.7, q^* is of the form $q^*(Y) = \alpha \text{tr}(\Theta^{-1}(Y)^2) + \beta \text{Sum}(\Theta^{-1}(Y)^2) + \gamma \text{Sum}(\Theta^{-1}(Y))^2$ with $\min(\alpha, 2\alpha + (n-2)\beta, \alpha + (n-1)(\beta + n\gamma)) > 0$. We compute X^2 with $X = \Theta^{-1}(Y)$, and the three terms:

$$\begin{aligned} X^2 &= Y^2 + \frac{1}{2}\text{tr}(Y)(\mu\mathbf{1}^\top + \mathbf{1}\mu^\top) + n\mu\mu^\top + \|\mu\|^2\mathbf{1}\mathbf{1}^\top - Y\mu\mathbf{1}^\top - \mathbf{1}\mu^\top Y, \\ \text{tr}(X^2) &= \text{tr}(Y^2) + \frac{1}{2}\text{tr}(Y)^2 + \frac{n}{2}\text{tr}(\text{Diag}(Y)^2), \\ \text{Sum}(X^2) &= \frac{n}{2}\text{tr}(Y)^2 + \frac{n}{4}\text{tr}(Y)^2 + \frac{n^2}{4}\text{tr}(\text{Diag}(Y)^2), \\ \text{Sum}(X) &= -n \text{tr}(Y), \\ \text{Sum}(X)^2 &= n^2 \text{tr}(Y)^2. \end{aligned}$$

Hence $g(Y, Y) = \alpha \text{tr}(Y^2) + \delta \text{tr}(\text{Diag}(Y)^2) + \zeta \text{tr}(Y)^2$ with $\delta = \frac{n}{2}(\alpha + \frac{n}{2}\beta)$ and $\zeta = \frac{\alpha}{2} + \frac{3n}{4}\beta + n^2\gamma$. The inverse relations between coefficients are $\beta = \frac{4}{n^2}\delta - \frac{2}{n}\alpha$ and $\gamma = \frac{1}{n^2}(\zeta - \frac{3}{n}\delta + \alpha)$. Thus $2\alpha + (n-2)\beta = \frac{4}{n^2}(n\alpha + (n-2)\delta)$ and $\alpha + (n-1)(\beta + n\gamma) = \frac{1}{n^2}(n\alpha + (n-1)(\delta + n\zeta))$, which gives the expected positivity condition. \square

11.7.7 Proof of Theorem 8.23

Theorem 8.23 (Coincidence of the metrics in dimension 2) In dimension 2, up to a positive scaling factor, the quotient-affine metric, the off-log metric and the log-scaled metric coincide. We recall that the Fisher transformation is the increasing map $F : \rho \in (-1, 1) \longmapsto$

$\frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right) \in \mathbb{R}^+$. Let $C = C(\rho)$ and $X = \begin{pmatrix} 0 & x \\ x & 0 \end{pmatrix}$ with $\rho \in (-1, 1)$ and $x \in \mathbb{R}$. Then:

1. (Metric) $g_C(X, X) = \frac{x^2}{(1-\rho^2)^2}$ (up to a scaling factor $\alpha > 0$),

2. (Geodesic) $\gamma(t) = C(\rho(t))$ where $\rho(t) = \frac{\rho_1 \cosh(\lambda t) + \sinh(\lambda t)}{\rho_1 \sinh(\lambda t) + \cosh(\lambda t)}$ with $\lambda = F(\rho_2) - F(\rho_1)$ is monotonic (increasing if $\rho_1 < \rho_2$, decreasing if $\rho_1 > \rho_2$, constant if $\rho_1 = \rho_2$),
3. (Distance) $d(C_1, C_2) = |\lambda| = |F(\rho_2) - F(\rho_1)|$ (up to a scaling factor $\sqrt{\alpha}$).

Proof of Theorem 8.23 (Coincidence of the metrics in dimension 2). It suffices to use the 2nd statement of Theorems 8.4 and 8.17 and the formulae of distances of the off-log metric and the log-scaled metric. For the former, $d(C_1, C_2) = q(F(\rho_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - F(\rho_1) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}) \propto \lambda^2$ and similarly for the ladder. Up to a multiplicative constant, these distances are equal to the quotient-affine distance in dimension 2 [Thanwerdas and Penne, 2021]. Therefore, the Riemannian metrics coincide up to a constant and the geodesics coincide. The formulae can be found in [Thanwerdas and Penne, 2022c]. \square

11.7.8 Proof of Theorem 8.24

Theorem 8.24 (Optimization problem to compute the scaling) $\arg \min_{\Omega} f \in \mathcal{M}$ and $\Omega = \text{conv}(\mathcal{M})$. Thus x^* is the global minimizer of the convex map f on the closed convex set Ω .

Proof of Theorem 8.24 (Optimization problem to compute the scaling). For all $x \in \Omega$, the vector $x_0 = \frac{1}{\alpha}x \in \mathcal{M}$ with $\alpha = (\prod_{i=1}^n x_i)^{1/n} \geq 1$ satisfies $f(x_0) = x_0^\top \Sigma x_0 = \frac{1}{\alpha^2} x^\top \Sigma x \leq f(x)$. Moreover, if $\alpha > 1$, then $f(x_0) < f(x)$. Hence if x minimizes f on Ω , then $f(x) \leq f(x_0) \leq f(x)$ so $f(x_0) = f(x)$ so $\alpha = 1$ and $x = x_0 \in \mathcal{M}$. Otherwise said, $\arg \min_{\Omega} f = \arg \min_{\mathcal{M}} f \in \mathcal{M}$.

Let us show that $\Omega = \text{conv}(\mathcal{M})$. Let $x, y \in \Omega$ and $t \in [0, 1]$. Then since \log is concave, $\sum_{i=1}^n \log((1-t)x_i + ty_i) \geq (1-t) \sum_{i=1}^n \log(x_i) + t \sum_{i=1}^n \log(y_i) \geq 0$ so $(1-t)x + ty \in \Omega$. Hence Ω is convex and it contains \mathcal{M} so $\text{conv}(\mathcal{M}) \subseteq \Omega$.

Let $x \in \Omega \setminus \mathcal{M}$. Let $F : a \in (-x_2, x_1) \mapsto \log(x_1 - a) + \log(x_2 + a) + \sum_{i=3}^n \log(x_i) \in \mathbb{R}$. Note that $F(0) > 0$. Then $F'(a) = \frac{1}{x_2 + a} - \frac{1}{x_1 - a} = \frac{x_1 - x_2 - 2a}{(x_1 - a)(x_2 + a)}$. Hence with $a_0 = \frac{x_1 - x_2}{2} \in (-x_2, x_1)$, $F'(a)$ is positive when $a < a_0$ and negative when $a > a_0$. Thus F increases on $(-x_2, a_0)$ and decreases on (a_0, x_1) with $F(a_0) \geq F(0) > 0$, $F \xrightarrow{a \rightarrow -x_2} -\infty$ and $F \xrightarrow{a \rightarrow x_1} -\infty$. So, since F is continuous, there exist $a_2 \in (-x_2; a_0)$ and $a_1 \in (a_0; x_1)$ such that $F(a_1) = F(a_2) = 0$. Note that $a_2 < 0 < a_1$ because $F(0) > 0$. Therefore, $x = (1-t)y + tz$ with $y = (x_1 - a_1, x_2 + a_1, x_3, \dots, x_n) \in \mathcal{M}$, $z = (x_1 - a_2, x_2 + a_2, x_3, \dots, x_n) \in \mathcal{M}$ and $t = \frac{a_1}{a_1 - a_2} \in (0, 1)$. Finally, $\Omega \subseteq \text{conv}(\mathcal{M})$ so $\Omega = \text{conv}(\mathcal{M})$. \square

11.7.9 Proof of Theorem 8.32

Theorem 8.32 (Stopping criterion) $\forall z \in \mathbb{R}^n \setminus \Omega, \forall \alpha > 1, \|\Pi_{\Omega_\alpha} z - \Pi_{\Omega} z\| \leq (\alpha^{1/n} - 1) \|\Pi_{\Omega} z\|$.

Proof of Theorem 8.32. Let $z \in \mathbb{R}^n \setminus \Omega$. We denote $x = \Pi_{\Omega} z$ and $a \in \mathbb{R}^+$ such that $x - z = a \frac{1}{x}$. Thus it is equivalent to prove that for all $x \in \mathcal{M}$ and $a \in \mathbb{R}^+$, $\|\Pi_{\Omega_\alpha}(x - \frac{a}{x}) - x\| \leq (\alpha^{1/n} - 1) \|x\|$.

Hence, we define the smooth maps:

$$\begin{aligned} z &: a \in \mathbb{R}^+ \mapsto x - \frac{a}{x} \in \mathbb{R}^n \setminus \Omega, \\ y &: a \in \mathbb{R}^+ \mapsto \Pi_{\Omega_\alpha}(z(a)) \in \mathcal{M}_\alpha, \\ D &: a \in \mathbb{R}^+ \mapsto \|y(a) - x\|^2 = \sum_{i=1}^n (y_i(a) - x_i)^2 \in \mathbb{R}^+. \end{aligned}$$

We want to prove that for all $a \in \mathbb{R}^+$, $D(a) \leq (\alpha^{1/n} - 1)^2 \|x\|^2$. We proceed in two steps. Firstly, we prove that $\lim_{a \rightarrow +\infty} D(a) = (\alpha^{1/n} - 1)^2 \|x\|^2$. Secondly, we prove that D is increasing.

First step: we compute the limit of D at $+\infty$.

We recall that $x \in \mathcal{M}$, $z(a) = x - \frac{a}{x}$, $y(a) = \Pi_{\Omega_\alpha} z(a)$ and $D(a) = \|y(a) - x\|^2$. We denote $u = \frac{1}{\|x\|} \frac{x}{\|x\|} \in \mathbb{S}^{n-1}$.

Let us prove that y is bounded. Let $y_0 = \alpha^{1/n} x$. We consider the line $\mathcal{D} = y_0 + \mathbb{R} \frac{1}{y_0}$ and the infinite right circular cylinder \mathcal{C} centered on \mathcal{D} and passing through x . More precisely, denoting $R = \|y_0 - x - \langle y_0 - x | u \rangle u\|$, the cylinder is $\mathcal{C} = \{p \in \mathbb{R}^n \mid d(p, \mathcal{D}) \leq R\}$. Since the line $\mathcal{D}' = x + \mathbb{R} \frac{1}{x}$ is parallel to \mathcal{D} and since $x \in \mathcal{C}$, we have $\mathcal{D}' \subset \mathcal{C}$. For $z(a) \in \mathcal{D}'$, we define $p = \Pi_{\mathcal{D}} z(a) = z_0 + y_0 - x - \langle y_0 - x | u \rangle u \in \mathcal{D}$. Then:

$$\|y(a) - y_0\| = \|\Pi_{\Omega_\alpha}(z(a)) - \Pi_{\Omega_\alpha}(p)\| \leq \|z(a) - p\| = R.$$

So y is bounded.

Let us prove that $\lim_{a \rightarrow +\infty} y(a) = y_0$. We denote $v(a) = \frac{y(a) - x - \langle y(a) - x | u \rangle u}{\|y(a) - x - \langle y(a) - x | u \rangle u\|} \in \mathbb{S}^{n-1}$ so that $u \perp v(a)$ and the triangle $\mathcal{T}(a)$ with vertices $x, y(a), z(a)$ is in the plane $(x, u, v(a))$. In $\mathcal{T}(a)$, since $x = \Pi_{\Omega}(z(a))$ and $y(a) \in \Omega$, we have $\langle x - z(a) | x - y(a) \rangle \leq 0$, i.e. $\cos(\widehat{z(a)xy(a)}) \leq 0$, i.e. $\widehat{z(a)xy(a)} \in [\frac{\pi}{2}, \pi)$. Hence the angles $\theta(a) = \widehat{xz(a)y(a)}$ and $\varphi(a) = \widehat{xy(a)z(a)}$ are in $(0, \frac{\pi}{2}]$. Moreover, the law of sines ensures that $\frac{\|y(a) - x\|}{\sin \theta(a)} = \frac{\|z(a) - x\|}{\sin \varphi(a)} \geq \|z(a) - x\| \xrightarrow{a \rightarrow +\infty} +\infty$. Since $\|y(a) - x\|$ is bounded, $\lim_{a \rightarrow +\infty} \sin \theta(a) = 0$ so $\lim_{a \rightarrow +\infty} \theta(a) = 0$. Hence, since v is bounded and $\theta(a) \xrightarrow{a \rightarrow +\infty} 0$, we have:

$$\frac{1}{\left\| \frac{1}{y(a)} \right\|} \frac{1}{y(a)} = \frac{y(a) - z(a)}{\|y(a) - z(a)\|} = \cos(\theta(a))u + \sin(\theta(a))v(a) \xrightarrow{a \rightarrow +\infty} u.$$

Now, let $(a_k)_{k \in \mathbb{N}} \subset (\mathbb{R}^+)^{\mathbb{N}}$ be a sequence such that $\lim_{k \rightarrow \infty} a_k = +\infty$. Let $y_k = y(a_k)$. Let Y be a subsequential limit of $(y_k)_{k \in \mathbb{N}}$. Then $\frac{1}{Y} \parallel u$ so $Y \parallel x$. Moreover $Y \in \mathcal{M}_\alpha$ since \mathcal{M}_α is closed so $Y = \alpha^{1/n} x = y_0$. Otherwise said, $(y_k)_{k \in \mathbb{N}}$ is a bounded sequence with only one subsequential limit y_0 so $(y(a_k))_{k \in \mathbb{N}}$ converges to y_0 . This being valid for any sequence $(a_k)_{k \in \mathbb{N}}$ tending to $+\infty$, we proved that $y(a)$ tends to $y_0 = \alpha^{1/n} x$ when a tends to $+\infty$. Thus $\lim_{a \rightarrow +\infty} D(a) = (\alpha^{1/n} - 1)^2 \|x\|^2$.

Second step: we prove that D is increasing.

Since $y = y(a)$ is characterized by $y(a) - z(a) \parallel \frac{1}{y(a)}$ and $\prod_{i=1}^n y_i(a) = \alpha$, it is implicitly

defined by the following equations:

$$\begin{aligned} \forall i \in \{1, \dots, n-1\}, y_i(y_i - z_i(a)) &= y_{i+1}(y_{i+1} - z_{i+1}(a)), \\ \prod_{i=1}^n y_i &= \alpha. \end{aligned}$$

Let us apply the implicit function theorem to the smooth map $G : (a, y) \in \mathbb{R}^+ \times \mathbb{R}^n \mapsto (y_2(y_2 - z_2(a)) - y_1(y_1 - z_1(a)), \dots, y_n(y_n - z_n(a)) - y_{n-1}(y_{n-1} - z_{n-1}(a)), \sum_{i=1}^n \ln(y_i) - \ln(\alpha))$.

$$\begin{aligned} \frac{\partial G}{\partial y_1} &= (z_1(a) - 2y_1, 0, \dots, 0, \frac{1}{y_1}), \\ \forall i \in \{2, \dots, n-1\}, \frac{\partial G}{\partial y_i} &= (0, \dots, 0, 2y_i - z_i(a), z_i(a) - 2y_i, 0, \dots, 0, \frac{1}{y_i}), \\ \frac{\partial G}{\partial y_n} &= (0, \dots, 0, 2y_n - z_n(a), \frac{1}{y_n}), \end{aligned}$$

$$\det \left(\frac{\partial G}{\partial y} \right) = \begin{vmatrix} u_1 & 0 & 0 & \cdots & 0 & \frac{1}{y_1} \\ -u_2 & u_2 & 0 & \cdots & 0 & \frac{1}{y_2} \\ 0 & -u_3 & u_3 & \cdots & 0 & \frac{1}{y_3} \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & -u_{n-1} & u_{n-1} & \frac{1}{y_{n-1}} \\ 0 & \cdots & & 0 & -u_n & \frac{1}{y_n} \end{vmatrix} = \begin{vmatrix} u_1 & 0 & 0 & \cdots & 0 & \frac{1}{y_1} \\ 0 & u_2 & 0 & \cdots & 0 & \frac{1}{y_2} \\ 0 & 0 & u_3 & \cdots & 0 & \frac{1}{y_3} \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 0 & u_{n-1} & \frac{1}{y_{n-1}} \\ -u_n & \cdots & & -u_n & -u_n & \frac{1}{y_n} \end{vmatrix}$$

$$= \begin{vmatrix} u_1 & 0 & 0 & \cdots & 0 & \frac{1}{y_1} \\ 0 & u_2 & 0 & \cdots & 0 & \frac{1}{y_2} \\ 0 & 0 & u_3 & \cdots & 0 & \frac{1}{y_3} \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 0 & u_{n-1} & \frac{1}{y_{n-1}} \\ 0 & \cdots & & 0 & 0 & \frac{1}{y_n} + u_n \sum_{i=1}^{n-1} \frac{1}{u_i y_i} \end{vmatrix} = \underbrace{\prod_{i=1}^n u_i}_{\neq 0} \underbrace{\sum_{j=1}^n \frac{1}{u_j y_j}}_{< 0} \neq 0.$$

where $u_i = z_i - 2y_i$. So by the implicit function theorem, $y : \mathbb{R}^+ \rightarrow \mathbb{R}^n$ is smooth. We introduce an intermediate map $b : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ defined by $z_i(a) = x_i - \frac{a}{x_i} = y_i(a) - \frac{b(a)}{y_i(a)}$. It does not depend on i since $y - z \parallel \frac{1}{y}$. It is smooth since it writes $b = y_i(y_i - z_i)$.

Let us prove that $D : a \in \mathbb{R}^+ \mapsto \|y(a) - x\|^2 = \sum_{i=1}^n (y_i(a) - x_i)^2 \in \mathbb{R}^+$ is increasing.

$$\begin{aligned}
D'(a) &= 2 \sum_{i=1}^n (y_i(a) - x_i) y_i'(a), \\
-\frac{1}{x_i} &= \left(1 + \frac{b(a)}{y_i(a)^2}\right) y_i'(a) - \frac{b'(a)}{y_i(a)}, \\
-\frac{y_i(a)}{x_i} &= \left(y_i(a) + \frac{b(a)}{y_i(a)}\right) y_i'(a) - b'(a), \\
-\frac{1}{n} \sum_{i=1}^n \frac{y_i(a)}{x_i} &= \frac{1}{n} \sum_{i=1}^n y_i(a) y_i'(a) - b'(a), \\
\frac{1}{n} \sum_{j=1}^n \frac{y_j(a)}{x_j} - \frac{y_i(a)}{x_i} &= \left(1 + \frac{b(a)}{y_i(a)^2}\right) y_i(a) y_i'(a) - \frac{1}{n} \sum_{i=1}^n y_i(a) y_i'(a), \\
u(a) &= M(a)v(a),
\end{aligned}$$

where $u(a) = \left(\frac{1}{n} \sum_{j=1}^n \frac{y_j(a)}{x_j}\right) \mathbf{1} - \frac{y(a)}{x}$, $v(a) = y(a)y'(a)$ (product of vectors are Hadamard products, i.e. coordinate-wise) and:

$$M(a) = \begin{pmatrix} \left(1 + \frac{b(a)}{y_1(a)^2}\right) - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \left(1 + \frac{b(a)}{y_2(a)^2}\right) - \frac{1}{n} & & \vdots \\ \vdots & & \ddots & -\frac{1}{n} \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \left(1 + \frac{b(a)}{y_n(a)^2}\right) - \frac{1}{n} \end{pmatrix}.$$

Thus we can get $y'(a)$ by inverting $M(a)$. Let us compute the inverse with the formula $M^{-1} = \frac{1}{\det(M)} \text{com}(M)^\top$ where $\text{com}(M)$ is the comatrix of M . First, one can show that the matrix with diagonal $\text{Diag}(m_1, \dots, m_n)$ and off-diagonal terms all equal to $\mu \notin \{m_1, \dots, m_n\}$, then its determinant is $\det(M) = \prod_{i=1}^n (m_i - \mu) \left(1 + \mu \sum_{i=1}^n \frac{1}{m_i - \mu}\right)$. This also gives the diagonal cofactors $[\text{com}(M)]_{ii} = \prod_{j \neq i} (m_j - \mu) \left(1 + \mu \sum_{j \neq i} \frac{1}{m_j - \mu}\right)$. The off-diagonal cofactors are, up to a sign, the determinant of a matrix of the same type, where $m_i = \mu$, which is $\mu \prod_{j \neq i} (m_j - \mu)$. Hence the off-diagonal cofactors for $i < j$ are $[\text{com}(M)]_{ij} = [\text{com}(M)]_{ji} = \underbrace{(-1)^{i+j} (-1)^{j-1-i}}_{-\mu} \mu \prod_{j \neq i} (m_j - \mu)$ because the $(j-1)$ -th column has to be put in i -th position (by inverting consecutive columns) so that we are in the previous situation. This requires $j-1-i$ transpositions. Finally:

$$\begin{aligned}
M^{-1} &= \frac{1}{\det(M)} \text{com}(M) \\
&= \text{Diag}(c_1, \dots, c_n) + \frac{1}{n\delta} [c_i c_j]_{1 \leq i, j \leq n},
\end{aligned}$$

where $c_i = c_i(a) = \frac{1}{m_i - \mu} = \frac{1}{1 + \frac{b(a)}{y_i(a)^2}}$ and $\delta = \delta(a) = 1 - \frac{1}{n} \sum_{i=1}^n c_i(a) > 0$. Now it suffices to compute $D'(a)$ and show that it is positive. We remove the dependence in a for readability.

We denote $s = \frac{1}{n} \sum_{j=1}^n \frac{y_j}{x_j}$.

$$\begin{aligned}
\frac{\delta}{2} D'(a) &= \delta \sum_{i=1}^n (y_i - x_i) y'_i \\
&= \delta \left(\frac{y - x}{y} \right)^\top v \\
&= \delta \left(\frac{y - x}{y} \right)^\top M^{-1} u \\
&= \delta \sum_{i=1}^n c_i \left(1 - \frac{x_i}{y_i}\right) \left(s - \frac{y_i}{x_i}\right) + \frac{1}{n} \sum_{i,j} c_i c_j \left(1 - \frac{x_i}{y_i}\right) \left(s - \frac{y_j}{x_j}\right) \\
&= \sum_{i=1}^n c_i \left(1 - \frac{x_i}{y_i}\right) \left(s - \frac{y_i}{x_i}\right) - \frac{1}{n} \sum_{j=1}^n c_j \sum_{i=1}^n c_i \left(1 - \frac{x_i}{y_i}\right) \left(s - \frac{y_i}{x_i}\right) + \frac{1}{n} \sum_{i,j} c_i c_j \left(1 - \frac{x_i}{y_i}\right) \left(s - \frac{y_j}{x_j}\right) \\
&= \sum_i c_i \left(1 - \frac{x_i}{y_i}\right) \left(s - \frac{y_i}{x_i}\right) + \frac{1}{n} \sum_{i,j} c_i c_j \left(1 - \frac{x_i}{y_i}\right) \left(\frac{x_i}{y_i} - \frac{y_j}{x_j}\right) \\
&= \sum_i c_i \left(1 - \frac{x_i}{y_i}\right) \frac{1}{n} \sum_j \left(\frac{y_j}{x_j} - \frac{y_i}{x_i}\right) - \sum_i c_i \left(1 - \frac{x_i}{y_i}\right) \frac{1}{n} \sum_j c_j \left(\frac{y_j}{x_j} - \frac{y_i}{x_i}\right) \\
&= \frac{1}{n} \sum_{i,j} c_i (1 - c_j) \left(1 - \frac{x_i}{y_i}\right) \left(\frac{y_j}{x_j} - \frac{y_i}{x_i}\right) \\
&= \frac{1}{2n} \sum_{i,j} \left(\frac{y_j}{x_j} - \frac{y_i}{x_i}\right) (1 - c_i) (1 - c_j) \left[\frac{c_i \left(1 - \frac{x_i}{y_i}\right)}{1 - c_i} - \frac{c_j \left(1 - \frac{x_j}{y_j}\right)}{1 - c_j} \right] \\
&= \frac{1}{2nb} \sum_{i,j} \underbrace{(1 - c_i)(1 - c_j)}_{\geq 0} \left(\frac{y_j}{x_j} - \frac{y_i}{x_i}\right) [y_i(y_i - x_i) - y_j(y_j - x_j)]
\end{aligned}$$

Since $b = y_i(y_i - z_i) = y_i(y_i - x_i) + a \frac{y_i}{x_i}$ and $a > 0$, we have $y_i(y_i - x_i) - y_j(y_j - x_j) = a \left(\frac{y_i}{x_j} - \frac{y_i}{x_i}\right)$. Finally:

$$D'(a) = \frac{a}{nb\delta} \sum_{i,j} (1 - c_i)(1 - c_j) \left(\frac{y_j}{x_j} - \frac{y_i}{x_i}\right)^2 > 0 \quad (11.39)$$

The inequality is strict because the sum is null if and only if x and y are colinear, i.e. $y = \alpha^{1/n} x$, which corresponds to $a = +\infty$.

So D is increasing so $D \leq \lim_{a \rightarrow +\infty} D(a) = (\alpha^{1/n} - 1)^2 \|x\|^2$. \square

11.8 Proofs of Chapter 9

In this section, we prove the results of Chapter 9: Bures-Wasserstein stratified geometry of covariance matrices.

11.8.1 Proof of Lemma 9.20

Lemma 9.20 (Euclidean and Bures-Wasserstein topologies coincide) The Euclidean distance d^E and the Bures-Wasserstein distance d^{BW} define the same topology on $\text{Cov}(n)$.

Proof of Lemma 9.20 (Euclidean and Bures-Wasserstein topologies coincide). The map $\pi : X \in \text{Mat}(n) \mapsto XX^\top \in (\text{Cov}(n), d^E)$ is continuous so the quotient topology, i.e. the topology induced by the Bures-Wasserstein distance, is finer than the Euclidean topology. Conversely, let \mathcal{U} be an open set for the Bures-Wasserstein distance. Let $\Sigma \in \mathcal{U}$. Let $\varepsilon > 0$ such that the Bures-Wasserstein ball $\mathcal{B}^{\text{BW}}(\Sigma, \varepsilon)$ is included in \mathcal{U} . The set $\mathcal{V} = \text{pow}_2(\mathcal{B}^E(\Sigma^{1/2}, \varepsilon))$ is open for the Euclidean distance because the map $\text{pow}_2 : \Sigma \mapsto \Sigma^2$ is a homeomorphism of $(\text{Cov}(n), d^E)$. Moreover, if $\Lambda \in \mathcal{V}$, then $d^{\text{BW}}(\Sigma, \Lambda) \leq d^E(\Sigma^{1/2}, \Lambda^{1/2}) \leq \varepsilon$ so $\Lambda \in \mathcal{B}^{\text{BW}}(\Sigma, \varepsilon) \subseteq \mathcal{U}$. So $\mathcal{V} \subseteq \mathcal{U}$ is a Euclidean neighborhood of Σ , so \mathcal{U} is open for the Euclidean distance. Therefore the two topologies coincide. \square

11.8.2 Proof of Theorem 9.22

Theorem 9.22 (Bures-Wasserstein geodesics on $\text{Sym}^+(n)$) Let $\Sigma \in \text{Sym}^+(n)$.

1. (Exponential map) [Malagò et al., 2018] For all $V \in T_\Sigma \text{Sym}^+(n) \equiv \text{Sym}(n)$, the geodesic from Σ with initial speed V writes $\gamma_{(\Sigma, V)}(t) = \Sigma + tV + t^2 \mathcal{S}_\Sigma(V) \Sigma \mathcal{S}_\Sigma(V) \in \text{Sym}^+(n)$.
2. (Definition interval) Let $\lambda_{\max} = \max \text{sp}(\mathcal{S}_\Sigma(V))$ and $\lambda_{\min} = \min \text{sp}(\mathcal{S}_\Sigma(V))$. The definition interval of the geodesic $\gamma_{(\Sigma, V)}$ is the interval $I_{\Sigma, V}$ defined by:

- $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, -\frac{1}{\lambda_{\min}})$ if $\lambda_{\min} < 0 < \lambda_{\max}$,
- $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_{\min}})$ if $\lambda_{\min} < 0$ and $\lambda_{\max} \leq 0$,
- $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, +\infty)$ if $\lambda_{\min} \geq 0$ and $\lambda_{\max} > 0$,
- $I_{\Sigma, V} = \mathbb{R}$ if $\lambda_{\min} = \lambda_{\max} = 0$ (which only happens for $V = 0$).

3. (Cut time) The cut time is $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$ if $\lambda_{\min} < 0$ or $+\infty$ otherwise. The geodesic $\gamma_{(\Sigma, V)} : I_{\Sigma, V} \rightarrow \mathcal{M}$ is even minimizing on $I_{\Sigma, V}$.
4. (Logarithm map) For all $\Lambda \in \text{Sym}^+(n)$, there exists a unique preimage $V \in \text{Pre}_\Sigma(\Lambda)$. It writes $V = 2 \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}) - 2\Sigma$, where $\text{sym}(M) = \frac{1}{2}(M + M^\top)$. The geodesic joining Σ to Λ writes:

$$\gamma_{\Sigma \rightarrow \Lambda}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}).$$

Moreover, it is a logarithm: $V \in \mathcal{L}og_x(y)$. Thus the logarithm map is defined on $\mathcal{U}_\Sigma = \text{Sym}^+(n)$ and it writes $\text{Log}_\Sigma : \Lambda \in \text{Sym}^+(n) \mapsto 2 \text{sym}(\Sigma^{1/2}(\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} \Sigma^{-1/2}) - 2\Sigma \in T_\Sigma \text{Sym}^+(n)$.

Proof of Theorem 9.22 (Bures-Wasserstein geodesics on $\text{Sym}^+(n)$). We prove statement 3 in the end because it requires statement 4.

1. (Exponential map) The expression of the exponential map comes from [Malagò et al., 2018].
2. (Definition domain) The domain $I_{\Sigma, V}$ is described in [Malagò et al., 2018] as the connected component of 0 in $J_{\Sigma, V} = \{t \in \mathbb{R} | I_n + t \mathcal{S}_\Sigma(V) \in \text{Sym}^+(n)\}$. Since $t \in J_{\Sigma, V}$ if and only if $0 \notin \{1 + t\lambda | \lambda \in \text{sp}(\mathcal{S}_\Sigma(V))\}$ if and only if $t \notin \{-\frac{1}{\lambda} | \lambda \in \text{sp}(\mathcal{S}_\Sigma(V))\}$, we have $\max(-\infty, 0) \cap \{-\frac{1}{\lambda} | \lambda \in \text{sp}(\mathcal{S}_\Sigma(V))\} = -\frac{1}{\lambda_{\max}}$ if $\lambda_{\max} > 0$ and $\min[0, +\infty) \cap \{-\frac{1}{\lambda} | \lambda \in \text{sp}(\mathcal{S}_\Sigma(V))\} = -\frac{1}{\lambda_{\min}}$ if $\lambda_{\min} < 0$. Therefore, we have the following cases:

- if $\lambda_{\min} < 0 < \lambda_{\max}$, then $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, -\frac{1}{\lambda_{\min}})$,
 - if $\lambda_{\min} < 0$ and $\lambda_{\max} \leq 0$, then $(-\infty, 0] \subseteq J_{\Sigma, V}$ so $I_{\Sigma, V} = (-\infty, -\frac{1}{\lambda_{\min}})$,
 - if $\lambda_{\min} \geq 0$ and $\lambda_{\max} > 0$, then $[0, +\infty) \subseteq J_{\Sigma, V}$ so $I_{\Sigma, V} = (-\frac{1}{\lambda_{\max}}, +\infty)$,
 - if $\lambda_{\min} \geq 0$ and $\lambda_{\max} \leq 0$, which means $\lambda_{\min} = \lambda_{\max} = 0$, then $V = 0$ and $I_{\Sigma, V} = J_{\Sigma, V} = \mathbb{R}$.
4. (Logarithm map) The existence of a preimage $V \in \mathcal{P}re_{\Sigma}(V)$ and even a logarithm $V \in \mathcal{L}og_{\Sigma}(V)$ (because it satisfies $\|V\| = d(\Sigma, \Lambda)$) is due to [Bhatia et al., 2019]. The geodesic joining Σ to Λ (Equation 9.3) is derived in [Bhatia et al., 2019] and it suffices to derive the expression at $t = 0$ to compute $V = \dot{\gamma}_{\Sigma \rightarrow \Lambda}(0) = 2 \text{sym}(\Sigma^{1/2}(\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2}\Sigma^{-1/2}) - 2\Sigma$.

The uniqueness of the preimage comes from [Massart and Absil, 2020, Proposition 4.4] on $\text{Sym}^+(n, k)$ applied to $k = n$. Indeed, it is stated that there exists a unique $W \in \mathcal{H}_{\Sigma^{1/2}}$ such that:

- (a) for all $t \in [0, 1]$, $\Sigma^{1/2}(\Sigma^{1/2} + tW) \in \text{GL}(n)$,
- (b) $(\Sigma^{1/2} + W)(\Sigma^{1/2} + W)^{\top} = \Lambda$.

Therefore, there exists a unique $V = d_{\Sigma^{1/2}}\pi(W) = \Sigma^{1/2}W^{\top} + W\Sigma^{1/2} \in T_{\Sigma}\text{Sym}^+(n)$ (and $W = V_{\Sigma^{1/2}}^{\#}$) such that $1 \in I_{\Sigma, V}$ (for all $t \in [0, 1]$, $\text{Exp}_{\Sigma}(tV) \in \text{Sym}^+(n)$) and $\text{Exp}_{\Sigma}(V) = \Lambda$, i.e. $V \in \mathcal{P}re_{\Sigma}(\Lambda)$. Thus the logarithm map Log_{Σ} is defined on $\mathcal{U}_{\Sigma} = \text{Sym}^+(n)$.

3. (Cut time) Let us prove that $\gamma_{(\Sigma, V)}$ is minimizing on $I_{\Sigma, V}$. This will prove in particular that $t_{\text{cut}}(\Sigma, V) = \sup I_{\Sigma, V}$. Let $t, t' \in I_{\Sigma, V}$, $t < 0 < t'$, let $\Lambda = \text{Exp}_{\Sigma}(tV)$ and $\Lambda' = \text{Exp}_{\Sigma}(t'V)$. Changing the base point of the geodesic, we have $\Lambda' = \text{Exp}_{\Lambda}((t' - t)V')$ with $V' = -\dot{\gamma}_{(\Sigma, V)}(t) \in T_{\Lambda}\text{Sym}^+(n)$. Since for all $s \in [0, 1]$, $(1 - s)t + st' \in I_{\Sigma, V}$ and $\text{Exp}_{\Lambda}(s(t' - t)V') = \text{Exp}_{\Sigma}(((1 - s)t + st')V) \in \text{Sym}^+(n)$, we have $1 \in I_{\Lambda, (t' - t)V'}$ so $(t' - t)V' \in \mathcal{P}re_{\Lambda}(\Lambda')$. By uniqueness of the preimage of Λ' from Λ , $\text{Log}_{\Lambda}(\Lambda') = (t' - t)V'$ and $\gamma_{\Lambda, (t' - t)V'}$ is minimizing on $[0, 1]$. Equivalently, $\gamma_{(\Sigma, V)}$ is minimizing on $[t, t']$ so it is minimizing on $I_{\Sigma, V}$.

□

11.8.3 Proof of Theorem 9.25

Theorem 9.25 (Horizontal lift, tangent space, metric) Let $\Sigma \in \text{Sym}^+(n, k)$, let $X \in \mathbb{R}_*^{n \times k}$ such that $\Sigma = XX^{\top} \in \text{Sym}^+(n, k)$ and let $V \in T_{\Sigma}\text{Sym}^+(n, k)$. Let $\Sigma = UDU^{\top}$ be a singular value decomposition with $D \in \text{Diag}^+(k)$ and $U \in \text{St}(n, k)$. We denote $S = S_{\Sigma, V} = U\mathcal{S}_D(U^{\top}VU)U^{\top}$, where $\mathcal{S}_A(B)$ denotes the unique solution Z of the Sylvester equation $AZ + ZA = B$. Note that $S_{\Sigma, V}$ and $(I_n - UU^{\top})$ are independent from the chosen decomposition.

1. (Tangent space) $T_{\Sigma}\text{Sym}^+(n, k) = \{V \in \text{Sym}(n) \mid X_{\perp}^{\top}VX_{\perp} = 0\}$,
2. (Horizontal lift) $V_X^{\#} = X(X^{\top}X)^{-1}\mathcal{S}_{X^{\top}X}(X^{\top}VX) + \underbrace{(I_n - X(X^{\top}X)^{-1}X^{\top})}_{X_{\perp}X_{\perp}^{\top}}VX(X^{\top}X)^{-1}$,

3. (Bures-Wasserstein metric) $g_{\Sigma}^{BW(n,k)}(V, V) = \text{tr}(S_{\Sigma, V} \Sigma S_{\Sigma, V} + V \Sigma^{\dagger} V (I_n - U U^{\top}))$.

Proof of Theorem 9.25 (Horizontal lift, tangent space, metric). 1&2. We prove the expression of the tangent space and the horizontal lift together. Let $V \in T_{\Sigma} \text{Sym}^+(n, k)$. The horizontal lift is defined by:

- (lift) $V = d_X \pi(V^{\#}) = X(V^{\#})^{\top} + V^{\#} X^{\top}$,
- (horizontal) $V^{\#} = X(X^{\top} X)^{-1} F + X_{\perp} K$ where $F \in \text{Sym}(r)$ and $K \in \text{Mat}(n-r, r)$.

When we plug the second equality in the first one and we multiply by X^{\top} on the left and X on the right, since $X^{\top} X_{\perp} = 0$, we get immediately $X^{\top} V X = X^{\top} X F + F X^{\top} X$ so $F = \mathcal{S}_{X^{\top} X}(X^{\top} V X)$. By multiplying by X_{\perp}^{\top} on the left instead, we get $X_{\perp}^{\top} V X = K X^{\top} X$ so $K = X_{\perp}^{\top} V X (X^{\top} X)^{-1}$. Since the matrix $(X(X^{\top} X)^{-1/2}; X_{\perp})$ is orthogonal, we have $X(X^{\top} X)^{-1} X^{\top} + X_{\perp} X_{\perp}^{\top} = I_n$.

We compute $d_X \pi(V^{\#})$ to check that it is equal to V :

$$\begin{aligned} d_X \pi(V^{\#}) &= X[(X^{\top} X)^{-1} F + F(X^{\top} X)^{-1}] X^{\top} + 2\text{sym}(X_{\perp} X_{\perp}^{\top} V X (X^{\top} X)^{-1} X^{\top}) \\ &= X(X^{\top} X)^{-1} X^{\top} V X (X^{\top} X)^{-1} X^{\top} + 2\text{sym}(X_{\perp} X_{\perp}^{\top} V (I_n - X_{\perp} X_{\perp}^{\top})) \\ &= (I_n - X_{\perp} X_{\perp}^{\top}) V (I_n - X_{\perp} X_{\perp}^{\top}) + X_{\perp} X_{\perp}^{\top} V + V X_{\perp} X_{\perp}^{\top} - 2X_{\perp} X_{\perp}^{\top} V X_{\perp} X_{\perp}^{\top} \\ &= V - X_{\perp} X_{\perp}^{\top} V X_{\perp} X_{\perp}^{\top}. \end{aligned}$$

Thus, $X_{\perp} X_{\perp}^{\top} V X_{\perp} X_{\perp}^{\top} = 0$ so $X_{\perp}^{\top} V X_{\perp} = 0$. Conversely, if $X_{\perp}^{\top} V X_{\perp} = 0$, then V is the image by $d_X \pi$ of a horizontal vector so $V \in T_{\Sigma} \text{Sym}^+(n, k)$. Hence $T_{\Sigma} \text{Sym}^+(n, k) = \{V \in \text{Sym}(n) \mid X_{\perp}^{\top} V X_{\perp} = 0\}$.

3. The quotient metric is defined by $g_{\Sigma}^{BW(n,k)}(V, V) = \text{tr}(V_X^{\#} (V_X^{\#})^{\top})$ so we only need to compute $V^{\#} (V^{\#})^{\top}$ for any X , for example $X = U D^{1/2}$, and its trace.

$$\begin{aligned} V^{\#} &= U D^{1/2} D^{-1} \mathcal{S}_D(D^{1/2} U^{\top} V U D^{1/2}) + (I_n - U D^{1/2} D^{-1} D^{1/2} U^{\top}) V U D^{1/2} D^{-1} \\ &= U \mathcal{S}_D(U^{\top} V U) D^{1/2} + (I_n - U U^{\top}) V U D^{-1/2} \\ &= S U D^{1/2} + (I_n - U U^{\top}) V U D^{-1/2}, \\ V^{\#} (V^{\#})^{\top} &= S U D U^{\top} S + S V (I_n - U U^{\top}) + (I_n - U U^{\top}) V S + (I_n - U U^{\top}) V \Sigma^{\dagger} V (I_n - U U^{\top}). \end{aligned}$$

$$\text{Hence } \text{tr}(V^{\#} (V^{\#})^{\top}) = \text{tr}(S \Sigma S + V \Sigma^{\dagger} V (I_n - U U^{\top})).$$

□

11.8.4 Proof of Theorem 9.26

Theorem 9.26 (Bures-Wasserstein geodesics on $\text{Sym}^+(n, k)$) Let $\Sigma, \Lambda \in \text{Sym}^+(n, k)$ and $X, Y \in \mathbb{R}_*^{n \times k}$ such that $XX^{\top} = \Sigma$ and $YY^{\top} = \Lambda$. Let $U \in \text{St}(n, k)$ and $D \in \text{Diag}^+(k)$ such that $\Sigma = U D U^{\top}$.

1. (Exponential map) For all $V \in T_{\Sigma} \text{Sym}^+(n, k)$, the geodesic from Σ with initial speed V is $\gamma_{(\Sigma, V)} : t \in I_{\Sigma, V} \mapsto \Sigma + tV + t^2 W_{\Sigma, V}$, where $W_{\Sigma, V} = S_{\Sigma, V} \Sigma S_{\Sigma, V} + S_{\Sigma, V} V (I_n - U U^{\top}) + (I_n - U U^{\top}) V S_{\Sigma, V} + (I_n - U U^{\top}) V \Sigma^{\dagger} V (I_n - U U^{\top})$ and $S_{\Sigma, V} = U \mathcal{S}_D(U^{\top} V U) U^{\top}$.

2. (Definition interval) Let $F_{X,V}^0 = \mathcal{S}_{X^\top X}((X^\top X)^{-1/2} X^\top V X (X^\top X)^{-1/2})$ and $M_{X,V}^0 = (X^\top X)^{-3/2} X^\top V (I_n - X(X^\top X)^{-1} X^\top) V X (X^\top X)^{-3/2} \in \text{Sym}(n)$. Let $\mathcal{E}_{\Sigma,V} = \{\lambda \in \text{sp}(F_{X,V}^0) \mid \ker(\lambda I_k - F_{X,V}^0) \cap \ker(M_{X,V}^0) \neq \{0\}\} \subseteq \text{sp}(F_{X,V}^0) = \text{sp}(S_{\Sigma,V})$. If $\mathcal{E}_{\Sigma,V}$ is non-empty, then let $\lambda_+ = \max \mathcal{E}_{\Sigma,V}$ and $\lambda_- = \min \mathcal{E}_{\Sigma,V}$. The definition interval of the geodesic $\gamma_{(\Sigma,V)}$ is the interval $I_{\Sigma,V}$ defined by:

- $I_{\Sigma,V} = (-\frac{1}{\lambda_+}, -\frac{1}{\lambda_-})$ if $\lambda_- < 0 < \lambda_+$,
- $I_{\Sigma,V} = (-\infty, -\frac{1}{\lambda_-})$ if $\lambda_- < 0$ and $\lambda_+ \leq 0$,
- $I_{\Sigma,V} = (-\frac{1}{\lambda_+}, +\infty)$ if $\lambda_- \geq 0$ and $\lambda_+ > 0$,
- $I_{\Sigma,V} = \mathbb{R}$ if $\mathcal{E}_{\Sigma,V}$ is empty.

Applying this to $X = UD^{1/2}$ without loss of generality, $F_{X,V}^0 = \mathcal{S}_D(U^\top V U)$ and $M_{X,V}^0 = D^{-1} U^\top V (I_n - U U^\top) V U D^{-1}$ which is a bit more tractable to compute $\mathcal{E}_{\Sigma,V}$.

3. (Cut time) Let $\lambda_{\max} = \max \text{sp}(F_{X,V}^0)$ and $\lambda_{\min} = \min \text{sp}(F_{X,V}^0)$. Note that if $\mathcal{E}_{\Sigma,V} \neq \emptyset$, then we have $(\lambda_-, \lambda_+) \subseteq (\lambda_{\min}, \lambda_{\max})$. The cut time is $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$ if $\lambda_{\min} < 0$ or $+\infty$ otherwise. Symmetrically, we have $t_{\text{cut}}(\Sigma, -V) = \frac{1}{\lambda_{\max}}$ if $\lambda_{\max} > 0$ or $+\infty$ otherwise.
4. (Preimages) We define the indexing set $\mathcal{I}_{X,Y}^{\text{Pre}}$ by:

$$\begin{aligned} \mathcal{I}_{X,Y}^{\text{Pre}} = \{R \in O(n) \mid H := X^\top Y R^\top \in \text{Sym}(n) \text{ and} \\ \forall \mu < 0, \ker(\mu I_k - (X^\top X)^{-1/2} H (X^\top X)^{-1/2}) \\ \cap \ker(\mu^2 I_k - (X^\top X)^{-1/2} R Y^\top Y R^\top (X^\top X)^{-1/2}) = \{0\}\}. \end{aligned}$$

For $R \in \mathcal{I}_{X,Y}^{\text{Pre}}$, we denote $H = H_{X,Y,R} = X^\top Y R^\top$ so that $X^\top Y = H R$.

Then, the map $R \in \mathcal{I}_{X,Y}^{\text{Pre}} \mapsto V = 2 \text{sym}(X R Y^\top) - 2\Sigma \in \mathcal{P}re_\Sigma(\Lambda)$ is a bijection whose inverse is $V \in \mathcal{P}re_\Sigma(\Lambda) \mapsto R = (Y^\top Y)^{-1} Y^\top (X + V_X^\#) \in \mathcal{I}_{X,Y}^{\text{Pre}}$.

The geodesic joining Σ to Λ parametrized by $R \in \mathcal{I}_{X,Y}^{\text{Pre}}$ writes:

$$\forall t \in [0, 1], \gamma_{\Sigma \rightarrow \Lambda}^R(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(X R Y^\top). \quad (11.40)$$

5. (Logarithms) Let $\mathcal{I}_{X,Y}^{\text{Log}} = \{R \in O(n) \mid H_{X,Y,R} = X^\top Y R^\top \in \text{Cov}(n)\} = \{R \in O(n) \mid H_{X,Y,R} = (X^\top \Lambda X)^{1/2}\} = \{R \in O(n) \mid X^\top Y = (X^\top \Lambda X)^{1/2} R\} \subseteq \mathcal{I}_{X,Y}^{\text{Pre}}$.

Then, the map $R \in \mathcal{I}_{X,Y}^{\text{Log}} \mapsto V = 2 \text{sym}(X R Y^\top) - 2\Sigma \in \mathcal{L}og_\Sigma(\Lambda)$ is a bijection whose inverse is $V \in \mathcal{L}og_\Sigma(\Lambda) \mapsto R = (Y^\top Y)^{-1} Y^\top (X + V_X^\#) \in \mathcal{I}_{X,Y}^{\text{Log}}$.

6. (Logarithm map) Let $r = \text{rk}(\Sigma \Lambda) = \text{rk}(X^\top Y) = \text{rk}(H)$.

- (a) If $r = k$, then there exists a unique logarithm of Λ from Σ . In this case, the minimizing geodesic joining Σ to Λ writes:

$$\gamma_{\Sigma \rightarrow \Lambda}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2} ((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda). \quad (11.41)$$

- (b) If $r = k - 1$, then there exist exactly two logarithms of Λ from Σ .
- (c) If $r < k - 1$, then there is an infinity of logarithms of Λ from Σ .

Therefore, the logarithm map is defined on $\mathcal{U}_\Sigma = \{\Lambda \in \text{Sym}^+(n, k) | \text{rk}(\Sigma\Lambda) = k\}$ and it writes $\text{Log}_\Sigma : \Lambda \in \mathcal{U}_\Sigma \mapsto 2 \text{sym}(\Sigma^{1/2}((\Sigma^{1/2}\Lambda\Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2}\Lambda) - 2\Sigma \in T_\Sigma \text{Sym}^+(n, k)$.

Proof of Theorem 9.26 (Bures-Wasserstein geodesics on $\text{Sym}^+(n, k)$). We prove statement 3 in the end because it requires statements 4 and 5.

- (Exponential map) The exponential map is simply $\text{Exp}_\Sigma(tV) = \pi(\text{Exp}_X(tV_X^\#)) = (X + tV_X^\#)(X + tV_X^\#)^\top = \Sigma + tV + V_X^\#(V_X^\#)^\top$. The matrix $W = V_X^\#(V_X^\#)^\top$ was already computed in the proof of Theorem 9.25.
- (Definition domain) As in $\text{Sym}^+(n)$, let us first determine $J_{\Sigma, V} = \{t \in \mathbb{R} | \text{rk}(\Sigma + tV + t^2W_{\Sigma, V}) = k\}$. According to [Massart and Absil, 2020, Proposition 3.2] applied to $tV_X^\# = t[X(X^\top X)^{-1}F_{X, V} + X_\perp K_{X, V}]$ with $F = F_{X, V} = \mathcal{S}_{X^\top X}(X^\top V X)$ and $K = K_{X, V} = X_\perp^\top V X(X^\top X)^{-1}$, we have $t \in J_{\Sigma, V}$ if and only if $\ker(I_k + t(X^\top X)^{-1}F) \cap \ker(K) = \{0\}$. Denoting $\mathcal{E}_{\Sigma, V} = \{\lambda \in \text{sp}((X^\top X)^{-1}F) | \ker(\lambda I_k - (X^\top X)^{-1}F) \cap \ker(K) \neq \{0\}\}$, it is clear that $J_{\Sigma, V} = \mathbb{R} \setminus \{-\frac{1}{\lambda} | \lambda \in \mathcal{E}_{\Sigma, V}\}$. Then $I_{\Sigma, V}$ is the connected component of 0 in $J_{\Sigma, V}$. Its computation is analogous to the one in the proof of Theorem 9.22.

To get a condition on the kernels of symmetric matrices in $\mathcal{E}_{\Sigma, V}$, it suffices to note that the condition rewrites $\ker(\lambda I_k - (X^\top X)^{-1/2}F(X^\top X)^{-1/2}) \cap \ker(K(X^\top X)^{-1/2}) \neq \{0\}$ and $\ker(K(X^\top X)^{-1/2}) = \ker((X^\top X)^{-1/2}K^\top K(X^\top X)^{-1/2})$ with $(X^\top X)^{-1/2}K^\top K(X^\top X)^{-1/2} = (X^\top X)^{-3/2}X^\top V X_\perp X_\perp^\top V X(X^\top X)^{-3/2} = (X^\top X)^{-3/2}X^\top V(I_n - X(X^\top X)^{-1}X^\top)V X(X^\top X)^{-3/2}$.

Note that $\mathcal{E}_{\Sigma, V}$ is independent from the choice of X because $F_{XR, V} = R^\top F_{X, V} R$ and $K_{XR, V} = R^\top K_{X, V} R$ for all $R \in O(n)$ so the condition does not depend on X .

- (Preimages) In [Massart and Absil, 2020, Propositions 4.4 & 4.5], the solutions of the equation $\text{Exp}_\Sigma(V) = \Lambda$ are computed in a wider set than the definition domain \mathcal{D}_Σ of the exponential map because they did not need more to characterize the logarithms. They are actually computed in the set $\{V \in T_\Sigma \text{Sym}^+(n, k) | \Sigma + V + W_{\Sigma, V} \in \text{Sym}^+(n, k)\}$. However, the geodesic $\gamma_{(\Sigma, V)}$ may leave the manifold $\text{Sym}^+(n, k)$ before reaching Λ . Therefore, we complete their work with the additional condition $1 \in I_{\Sigma, V}$ to characterize the preimages of Λ from Σ .

From [Massart and Absil, 2020], we know that preimages V necessarily satisfy $V_X^\# = YR^\top - X$ with $X^\top Y = HR$, $H \in \text{Sym}(n)$, $R \in O(n)$. Thus, $V = XRY^\top + YR^\top X^\top - 2XX^\top$ so $X^\top V X = HX^\top X + X^\top XH - 2(X^\top X)^2$ so $F = \mathcal{S}_{X^\top X}(X^\top V X) = H - X^\top X$. Moreover, $K = X_\perp^\top V X(X^\top X)^{-1} = X_\perp^\top YR^\top$. Denoting $A = (X^\top X)^{-1/2}H(X^\top X)^{-1/2}$ and $B = (X^\top X)^{-1/2}RY^\top YR^\top(X^\top X)^{-1/2}$, we have $(X^\top X)^{-1/2}F(X^\top X)^{-1/2} = A - I_k$ and $(X^\top X)^{-1/2}K^\top K(X^\top X)^{-1/2} = (X^\top X)^{-1/2}RY^\top X_\perp X_\perp^\top YR^\top(X^\top X)^{-1/2} = B - A^2$.

We can now compute $\mathcal{E}_{\Sigma,V}$ and $I_{\Sigma,V}$. For all $\lambda \in \mathbb{R}$, for all $Z \in \mathbb{R}^n$:

$$\begin{aligned} Z \in \ker(\lambda I_k - S_{X,V}^0) \cap \ker(M_{X,V}^0) &\iff Z \in \ker((\lambda + 1)I_k - A) \cap \ker(B - A^2) \\ &\iff AZ = (\lambda + 1)Z \text{ and } BZ = A^2Z \\ &\iff AZ = (\lambda + 1)Z \text{ and } BZ = (\lambda + 1)^2Z \\ &\iff Z \in \ker((\lambda + 1)I_k - A) \cap \ker((\lambda + 1)^2I_k - B). \end{aligned}$$

Therefore:

$$\begin{aligned} \mathcal{E}_{\Sigma,V} &= \{\lambda \in \text{sp}(A - I_k) \mid \ker((\lambda + 1)I_k - A) \cap \ker((\lambda + 1)^2I_k - B) \neq \{0\}\} \\ &= \{\mu - 1 \in \text{sp}(A - I_k) \mid \ker(\mu I_k - A) \cap \ker(\mu^2 I_k - B) \neq \{0\}\}. \end{aligned}$$

Thus, denoting $\lambda_- = \min \mathcal{E}_{\Sigma,V}$, the condition $1 \in I_{\Sigma,V}$ rewrites:

$$\begin{aligned} 1 \in I_{\Sigma,V} &\iff \lambda_- \geq 0 \text{ or } -\frac{1}{\lambda_-} \geq 1 \iff \lambda_- \geq -1 \\ &\iff \forall \mu \in \text{sp}(A), \ker(\mu I_k - A) \cap \ker(\mu^2 I_k - B) \neq \{0\} \implies \mu \geq 0 \\ &\iff \forall \mu < 0, \ker(\mu I_k - A) \cap \ker(\mu^2 I_k - B) = \{0\}. \end{aligned}$$

To conclude, with the notations of statement 4, $V = 2 \text{sym}(XRY^\top) - 2\Sigma \in \mathcal{Pre}_\Sigma(\Lambda)$ if and only if $R \in \mathcal{I}_{X,Y}^{\text{Pre}}$.

5. (Logarithms) In [Massart and Absil, 2020], it is stated that the shortest vectors $V = d\pi(V_X^\#)$ with $V_X^\# = YR^\top - X$ are those for which (H, R) is a polar decomposition of $X^\top Y$, i.e. $H \geq 0$. Therefore, we necessarily have $H = (X^\top \Lambda X)^{1/2}$. From Definition 9.19, they even satisfy $\|V\| = \|V^\#\| = d^{\text{BW}}(\Sigma, \Lambda)$. Moreover, if $H \geq 0$, the condition $1 \in I_{\Sigma,V}$ is automatically satisfied, as stated in [Massart and Absil, 2020, Corollary 3.3 (5)]. So with the notations of statement 5, the logarithms are indexed by $\mathcal{I}_{X,Y}^{\text{Log}}$.
6. (Logarithm map) It is clear that $\text{rk}(\Sigma\Lambda) \leq \text{rk}(X^\top Y)$ since $\Sigma\Lambda = X(X^\top Y)Y^\top$. We also have $X^\top Y = (X^\top X)^{-1}X^\top(\Sigma\Lambda)Y(Y^\top Y)^{-1}$ so $\text{rk}(X^\top Y) \leq \text{rk}(\Sigma\Lambda)$. Finally, $\text{rk}(\Sigma\Lambda) = \text{rk}(X^\top Y) = \text{rk}(H)$. We denote it $r = \text{rk}(\Sigma\Lambda)$.

- (a) As stated in [Massart and Absil, 2020], if $r = k$, then there exists a unique logarithm of Σ from Λ . Moreover, we can compute an explicit expression. Indeed, $R = H^{-1}X^\top Y$ so $XRY^\top = XH^{-1}X^\top \Lambda = X(X^\top \Lambda X)^{-1/2}X^\top \Lambda$. Since the choice of X is free, let us take $X = UD^{1/2}$ where $\Sigma = UDU^\top$ with $U \in \text{St}(n, k)$ and $D \in \text{Diag}^+(k)$. Therefore:

$$\begin{aligned} XRY^\top &= UD^{1/2}(D^{1/2}U^\top \Lambda UD^{1/2})^{-1/2}D^{1/2}U^\top \Lambda \\ &= UD^{1/2}U^\top U((D^{1/2}U^\top \Lambda UD^{1/2})^{1/2})^{-1}U^\top UD^{1/2}U^\top \Lambda \\ &= \Sigma^{1/2}(U(D^{1/2}U^\top \Lambda UD^{1/2})^{1/2}U^\top)^\dagger \Sigma^{1/2} \Lambda \\ &= \Sigma^{1/2}((UD^{1/2}U^\top \Lambda UD^{1/2}U^\top)^{1/2})^\dagger \Sigma^{1/2} \Lambda \\ &= \Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda. \end{aligned}$$

So the unique minimizing geodesic joining Σ to Λ writes:

$$\gamma_{(\Sigma, \Lambda)}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda).$$

- (b) If $r = k - 1$, without loss of generality, let us assume that $X^\top Y \in \text{Diag}(k)$, $X^\top Y = \text{Diag}(d_1, \dots, d_{k-1}, 0)$. Then $H \geq 0$ and $H^2 = (X^\top Y)^2$ imposes that $H = \text{Diag}(|d_1|, \dots, |d_{k-1}|, 0)$. Therefore, there are only two matrices $R_+, R_- \in O(n)$ defined by $R_\pm = \text{Diag}(\text{sgn}(d_1), \dots, \text{sgn}(d_{k-1}), \pm 1)$ such that $X^\top Y = HR_\pm$. Thus there are exactly two logarithms of Λ from Σ .
- (c) If $r < k - 1$, similarly we can assume without loss of generality that $X^\top Y = \text{Diag}(d_1, \dots, d_r, 0, \dots, 0)$. Then $H = \text{Diag}(|d_1|, \dots, |d_r|, 0, \dots, 0)$ and $R = \text{Diag}(\varepsilon, R_0)$ is a block-diagonal matrix with $\varepsilon = \text{Diag}(\text{sgn}(d_1), \dots, \text{sgn}(d_r)) \in \text{Diag}(r)$ and $R_0 \in O(k - r)$. Thus there is an infinity of logarithms of Λ from Σ .

Thus the logarithm map is defined on $\mathcal{U}_\Sigma = \{\Lambda \in \text{Sym}^+(n, k) \mid \text{rk}(\Sigma\Lambda) = k\}$, as stated in [Massart and Absil, 2020].

3. (Cut time) Let $t \in I_{\Sigma, V} \cap \mathbb{R}_+$. Let $\Lambda = \gamma_{(\Sigma, V)}(t)$, $Y \in \mathbb{R}_*^{n \times k}$ such that $YY^\top = \Lambda$, $(H, R) \in \text{Sym}(k) \times O(k)$ such that $X^\top Y = HR$ and $V_X^\# = YR^\top - X$. Then, $X^\top X + tX^\top V_X^\# = X^\top YR^\top = H$. Besides, $X^\top X + tX^\top V^\# = (X^\top X)^{1/2}(I_n + tS_{X, V}^0)(X^\top X)^{1/2}$. Let $\lambda_{\min} = \min \text{sp}(S_{X, V}^0) = \min \text{sp}(S_{\Sigma, V})$. Therefore, $\gamma_{(\Sigma, V)}$ is minimizing on $[0, t]$ if and only if $H \geq 0$ if and only if $I_n + tS_{X, V}^0 \geq 0$ if and only if $1 + t\lambda_{\min} \geq 0$. If $\lambda_{\min} \geq 0$, the condition is empty so $t_{\text{cut}}(\Sigma, V) = +\infty$. If $\lambda_{\min} < 0$, the condition writes $t \leq -\frac{1}{\lambda_{\min}}$ so $t_{\text{cut}}(\Sigma, V) = -\frac{1}{\lambda_{\min}}$.

□

11.8.5 Proof of Theorem 9.31

Theorem 9.31 (Bures-Wasserstein minimizing geodesic segments in $\text{Cov}(n)$) Let $\Sigma, \Lambda \in \text{Cov}(n)$ with $\text{rk}(\Sigma) = k$ and $\text{rk}(\Lambda) = l$. Let $X, Y \in \mathbb{R}^{n \times n}$ such that $XX^\top = \Sigma$ and $YY^\top = \Lambda$. The two following statements are equivalent:

- (i) the curve $\gamma : [0, 1] \rightarrow \text{Cov}(n)$ is a minimizing geodesic segment from Σ to Λ ,
- (ii) there exists $R \in O(n)$ such that $H_{X, Y, R} := X^\top YR^\top \in \text{Cov}(n)$ and for all $t \in [0, 1]$, $\gamma(t) = \gamma_{\Sigma \rightarrow \Lambda}^R(t) = (1 - t)^2 \Sigma + t^2 \Lambda + 2t(1 - t) \text{sym}(XRY^\top)$.

Moreover, $H_{X, Y, R} = (X^\top \Lambda X)^{1/2}$ and the minimizing geodesic $\gamma_{\Sigma \rightarrow \Lambda}^R$ is of constant rank $p \geq \max(k, l)$ on $(0, 1)$.

Proof of Theorem 9.31 (Bures-Wasserstein minimizing geodesic segments in $\text{Cov}(n)$).

(Necessity) Let $\gamma : [0, 1] \rightarrow \text{Cov}(n)$ be a minimizing geodesic segment from $\Sigma = \gamma(0)$ to $\Lambda = \gamma(1)$. Let $p = \max_{t \in [0, 1]} \text{rk}(\gamma(t))$. By Lemma 9.30, γ is of constant rank $p \geq \max(k, l)$ on $(0, 1)$. In other words, $\gamma|_{(0, 1)} : (0, 1) \rightarrow \text{Sym}^+(n, p)$ is a minimizing geodesic of $\text{Sym}^+(n, p)$. Let $c_0 : (0, 1) \rightarrow \mathbb{R}_*^{n \times p}$ be a horizontal lift of $\gamma|_{(0, 1)}$. Necessarily, $c_0(t) = (1 - t)X_0 + tY_0$ with $X_0, Y_0 \in \mathbb{R}^{n \times p}$ with $X_0 X_0^\top = \Sigma$ and $Y_0 Y_0^\top = \Lambda$ since $(X_0, \Sigma) = \lim_{t \rightarrow 0} (c_0(t), \gamma(t))$ and $(Y_0, \Lambda) = \lim_{t \rightarrow 1} (c_0(t), \gamma(t))$. Let us show that $X_0^\top Y_0 \in \text{Cov}(p)$.

For all $[a, b] \subset (0, 1)$, the tangent vectors $V_{c_0(a)}^\# = (b - a)(Y_0 - X_0) \in T_{c_0(a)} \mathbb{R}_*^{n \times p} = \mathbb{R}^{n \times p}$ and $V = c_0(a)(V_{c_0(a)}^\#)^\top + V_{c_0(a)}^\# c_0(a)^\top \in \mathcal{L}og_{\gamma(a)}(\gamma(b)) \subset T_{\gamma(a)} \text{Sym}^+(n, p)$ uniquely determine

a pair of matrices $R_{a,b} \in \mathcal{I}_{c_0(a),c_0(b)}^{\text{Log}}$ and $H_{a,b} = H_{c_0(a),c_0(b),R_{a,b}} = c_0(a)^\top c_0(b) \in \text{Cov}(p)$. We compute $H_{a,b}$:

$$\begin{aligned} H_{a,b} &= c_0(a)^\top c_0(b) = [(1-a)X_0 + aY_0]^\top [(1-b)X_0 + bY_0] \\ &= (1-a)(1-b)X_0^\top X_0 + abY_0^\top Y_0 + (1-a)bX_0^\top Y_0 + a(1-b)Y_0^\top X_0. \end{aligned}$$

Therefore, $\lim_{\substack{a \rightarrow 0 \\ b \rightarrow 1}} H_{a,b} = X_0^\top Y_0$ so $X_0^\top Y_0 \in \text{Cov}(p)$.

Since $[X_0 \ 0][X_0 \ 0]^\top = \Sigma$ and $[Y_0 \ 0][Y_0 \ 0]^\top = \Lambda$, there exist $P, Q \in O(n)$ such that $X = [X_0 \ 0]P$ and $Y = [Y_0 \ 0]Q$. Thus the curve $c : t \in [0, 1] \mapsto (1-t)X + tYR^\top \in \text{Mat}(n)$ with $R = P^\top Q \in O(n)$ satisfies $\gamma(t) = c(t)c(t)^\top$ for all $t \in [0, 1]$. Indeed, it is equal to $c_0(t)c_0(t)^\top$ on $(0, 1)$ and the equality is clear for $t \in \{0, 1\}$. Moreover, $H_{X,Y,R} = X^\top YR^\top = U^\top \text{Diag}(X_0^\top Y_0, 0)U \in \text{Cov}(n)$ so $H_{X,Y,R} = ((X^\top Y)(X^\top Y)^\top)^{1/2} = (X^\top \Lambda X)^{1/2}$. Finally, $\gamma(t) = c(t)c(t)^\top = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(XRY^\top)$.

(Sufficiency) Let $\gamma_{(\Sigma, \Lambda)}^R(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(XRY^\top)$ with $H = H_{X,Y,R} = X^\top YR^\top \in \text{Cov}(n)$ and let us prove that it is a minimizing geodesic segment. We define $W = YR^\top - X$ and $c(t) = X + tW = (1-t)X + tYR^\top$ for $t \in [0, 1]$. The curve c is a geodesic of $\text{Mat}(n)$ such that $c(t)c(t)^\top = \gamma_{(\Sigma, \Lambda)}^R(t)$ for all $t \in [0, 1]$. Moreover, $L(c) = \|W\| = \text{tr}(XX^\top + YY^\top - 2X^\top YR^\top)^{1/2} = \text{tr}(\Sigma + \Lambda - 2H)^{1/2}$. Let $Q \in O(n)$ such that $X = \Sigma^{1/2}Q$. Since $H \geq 0$, $H = (X^\top \Lambda X)^{1/2} = Q^\top (\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2} Q$. Therefore, $L(c) = \|W\| = \|YR^\top - X\| = \text{tr}(\Sigma + \Lambda - 2H)^{1/2} = d^{\text{BW}}(\Sigma, \Lambda)$. In other words, $c : [0, 1] \rightarrow \text{Cov}(n)$ is a minimizing curve between two registered points X and YR^\top so by Lemma 9.13, its projection $\gamma : [0, 1] \rightarrow \text{Cov}(n)$ is a minimizing curve and $L(\gamma) = L(c) = d^{\text{BW}}(\Sigma, \Lambda)$.

By Lemma 9.30 again, γ has constant rank $p \geq \max(k, l)$ on $(0, 1)$ so $\gamma|_{(0,1)} : (0, 1) \rightarrow \text{Sym}^+(n, p)$ is a minimizing curve of $\text{Sym}^+(n, p)$. Since $c|_{(0,1)}$ has constant speed, so does $\gamma|_{(0,1)}$. By continuity of the length, γ has constant speed on $[0, 1]$ so $\gamma : [0, 1] \rightarrow \text{Cov}(n)$ is a minimizing geodesic segment. \square

11.8.6 Proof of Lemma 9.33

Lemma 9.33 (Elementary algebra) Let $\Sigma, \Lambda \in \text{Cov}(n)$ with $\text{rk}(\Sigma) = k$ and $\text{rk}(\Lambda) = l$. Let $r = \text{rk}(\Sigma\Lambda)$.

1. For all $X, Y \in \text{Mat}(n)$ such that $XX^\top = \Sigma$ and $YY^\top = \Lambda$, $r = \text{rk}(X^\top Y)$.
2. We have $l - r \leq n - k$.

Proof of Lemma 9.33 (Elementary algebra). 1. Let $X_0 \in \mathbb{R}_*^{n \times k}$ and $Y_0 \in \mathbb{R}_*^{n \times l}$ such that $X_0 X_0^\top = \Sigma$ and $Y_0 Y_0^\top = \Lambda$. Thus there exist $P, Q \in O(n)$ such that $X = [X_0 \ 0]P$ and $Y = [Y_0 \ 0]Q$. Since $\Sigma\Lambda = X(X^\top Y)Y^\top$, we have $r \leq \text{rk}(X^\top Y) = \text{rk}(X_0^\top Y_0)$. Since $X_0^\top Y_0 = (X_0^\top X_0)^{-1} X_0^\top \Sigma \Lambda Y_0 (Y_0^\top Y_0)^{-1}$, we have $\text{rk}(X_0^\top Y_0) \leq \text{rk}(\Sigma\Lambda)$. Finally, $r = \text{rk}(X_0^\top Y_0) = \text{rk}(X^\top Y)$.

2. Let $f, g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be linear endomorphisms respectively represented by Σ and Λ in the canonical basis. From the rank-nullity theorem applied to $f|_{\text{im}(g)} : \text{im}(g) \rightarrow \mathbb{R}^n$, the restriction of f to $\text{im}(g)$, and since $\text{im}(f|_{\text{im}(g)}) = \text{im}(f \circ g)$ and $\ker(f|_{\text{im}(g)}) \subseteq \ker(f)$, we

have $\text{rk}(g) = \text{rk}(f|_{\text{im}(g)}) + \dim \ker(f|_{\text{im}(g)}) \leq \text{rk}(f \circ g) + \dim(\ker f) = \text{rk}(f \circ g) + n - \text{rk}(f)$.
 This writes $l - r \leq n - k$. □

11.8.7 Proof of Theorem 9.34

Theorem 9.34 (Number of Bures-Wasserstein minimizing geodesic segments in $\text{Cov}(n)$)
 Let $\Sigma, \Lambda \in \text{Cov}(n)$ with $\text{rk}(\Sigma) = k$ and $\text{rk}(\Lambda) = l$. We assume that $k \geq l$ without loss of generality. We denote $r = \text{rk}(\Sigma\Lambda)$. We have $l - r \leq n - k$.

1. There exists a bijection between the set of minimizing geodesics from Σ to Λ and the closed unit ball of $\mathbb{R}^{(k-r) \times (l-r)}$ for the spectral norm $\bar{\mathcal{B}}_S(0, 1) = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid \|R_0\|_S \leq 1\} = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid 0 \leq R_0^\top R_0 \leq I_{l-r}\}$. More precisely, this bijection is given by:

$$\gamma_{\Sigma \rightarrow \Lambda}^{R_0}(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(X_r Y_r^\top + X_{k-r} R_0 Y_{l-r}^\top),$$

where $X = (X_r \ X_{k-r} \ 0) \in \mathbb{R}^{n \times n}$ and $Y = (Y_r \ Y_{l-r} \ 0) \in \mathbb{R}^{n \times n}$ are such that $XX^\top = \Sigma$, $YY^\top = \Lambda$ and $X^\top Y = \text{Diag}(D_r, 0)$, and $R = \begin{pmatrix} I_r & 0 & 0 \\ 0 & R_0 & * \\ 0 & * & * \end{pmatrix} \in O(n)$, with $X_r, Y_r \in \mathbb{R}^{n \times r}$, $X_{k-r} \in \mathbb{R}^{n \times (k-r)}$, $Y_{l-r} \in \mathbb{R}^{n \times (l-r)}$, $D_r \in \text{Diag}^+(r)$, $R_0 \in \bar{\mathcal{B}}_S(0, 1)$.

2. The minimizing geodesic is unique if and only if $r = l$. This includes the cases $k = n$.
3. There are infinitely many minimizing geodesics if and only if $r < l$.
4. The minimizing geodesics corresponding to the choices $R_0 \in \text{St}(k-r, l-r)$ (including the empty matrix if $r = l$) have rank exactly k on $[0, 1)$ (on $[0, 1]$ if $l = k$). Note that $\text{St}(k-r, l-r)$ is included in the unit sphere $\mathcal{S}_S(0, 1) = \{R_0 \in \mathbb{R}^{(k-r) \times (l-r)} \mid \|R_0\|_S = 1\}$.
5. The minimizing geodesic corresponding to the choice $R_0 = 0$ (or the empty matrix if $r = l$) writes for all $t \in [0, 1]$:

$$\gamma_{\Sigma \rightarrow \Lambda}^0(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda).$$

If $r = l$, it has rank exactly k on $[0, 1)$.

The number of minimizing geodesic segments in $\text{Sym}^+(n, k)$ and in $\text{Cov}(n)$ is summarized in Table 9.2 with $n \geq k \geq l \geq r$.

$\Sigma \in$	$\Lambda \in$	$r = \text{rk}(\Sigma\Lambda)$	Number of minimizing geodesics	
			in $\text{Sym}^+(n, k)$	in $\text{Cov}(n)$
$\text{Sym}^+(n)$	$\text{Sym}^+(n)$	n	1	1
$\text{Sym}^+(n)$	$\text{Sym}^+(n, k)$	k	1	1
$\text{Sym}^+(n, k)$	$\text{Sym}^+(n, k)$	k	1	1
		$k - 1$	2	∞
		$< k - 1$	∞	∞
$\text{Sym}^+(n, k)$	$\text{Sym}^+(n, l)$	l	1	1
		$< l$	∞	∞

Table 9.2: Number of Bures-Wasserstein minimizing geodesic segments ($n \geq k \geq l \geq r$)

Proof of Theorem 9.34 (Number of Bures-Wasserstein minimizing geodesic segments in $\text{Cov}(n)$). 1.

Without loss of generality, let us choose $X, Y \in \text{Mat}(n)$ with an convenient form. We choose $X_0 \in \mathbb{R}_*^{n \times k}$ and $Y_0 \in \mathbb{R}_*^{n \times l}$ as in the proof of the previous lemma. Given a singular value decomposition of $X_0^\top Y_0 = U_k D V_l^\top$ with $U_k \in O(k)$, $V_l \in O(l)$ and $D = \text{Diag}(D_r, 0)$ with $D_r \in \text{Diag}^+(r)$, we define $P = \text{Diag}(U_k, I_{n-k}) \in O(n)$ and $Q = \text{Diag}(V_l, I_{n-l}) \in O(n)$. We choose $X = [X_0 \ 0]P = [X_k \ 0]$ with $X_k = X_0 U_k \in \mathbb{R}_*^{n \times k}$ and $Y = [Y_0 \ 0]Q = [Y_l \ 0]$ with $Y_l = Y_0 V_l \in \mathbb{R}_*^{n \times l}$. Therefore we have $XX^\top = \Sigma$, $YY^\top = \Lambda$, $X = [X_k \ 0]$, $Y = [Y_l \ 0]$ and $X^\top Y = \text{Diag}(D_r, 0)$. We denote $X = [X_r \ X_{k-r} \ 0]$ and $Y = [Y_r \ Y_{l-r} \ 0]$ with $X_r, Y_r \in \mathbb{R}_*^{n \times r}$, $X_{k-r} \in \mathbb{R}_*^{n \times (k-r)}$ and $Y_{l-r} \in \mathbb{R}_*^{n \times (l-r)}$.

Necessarily, $H = (X^\top Y Y^\top X)^{1/2} = \text{Diag}(D_r, 0)$. The possible $R \in O(n)$ such that $HR = X^\top Y$ are $R = \text{Diag}(I_r, R_{n-r})$ with $R_{n-r} = [R_{l-r} \ R_{n-l}] \in O(n-r)$ where $R_{l-r} \in \text{St}(n-r, l-r)$ and $R_{n-l} \in \text{St}(n-r, n-l)$. We denote $R_{l-r} = \begin{pmatrix} R_0 \\ R_1 \end{pmatrix}$ with $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$ and $R_1 \in \mathbb{R}^{(n-k) \times (l-r)}$ with $R_0^\top R_0 + R_1^\top R_1 = I_{l-r}$. Note that both R_0 and R_1 have more rows than columns.

A simple calculus gives $XRY^\top = X_r Y_r^\top + X_{k-r} R_0 Y_{l-r}^\top$. Given R, R' satisfying $HR = HR' = X^\top Y$, we have $XRY^\top = XR'Y^\top$ if and only if $R_0 = R'_0$ since $X_{k-r}^\top X_{k-r}$ and $Y_{l-r}^\top Y_{l-r}$ are invertible. We even have $\text{sym}(XRY^\top) = \text{sym}(XR'Y^\top)$ if and only if $R_0 = R'_0$. Indeed, if $\text{sym}(XRY^\top) = \text{sym}(XR'Y^\top)$, then $X_{k-r}(R_0 - R'_0)Y_{l-r}^\top = Y_{l-r}(R'_0 - R_0)^\top X_{k-r}^\top$. Since $X_{k-r}^\top Y_{l-r} = 0$, it suffices to multiply on the left by X_{k-r}^\top and on the right by Y_{l-r} to conclude that $R_0 = R'_0$. Thus there is a bijection between minimizing geodesic segments and submatrices $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$ of $R_{l-r} = \begin{pmatrix} R_0 \\ R_1 \end{pmatrix} \in \text{St}(n-r, l-r)$.

Since R_1 has more rows than columns, any $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$ such that $R_0^\top R_0 \leq I_{l-r}$ can be completed by an appropriate $R_1 = \begin{pmatrix} (I_{l-r} - R_0^\top R_0)^{1/2} \\ \mathbf{0}_{n-k-(l-r), l-r} \end{pmatrix}$.

Therefore, the minimizing geodesic segments are in bijection with the matrices $R_0 \in \mathbb{R}^{(k-r) \times (l-r)}$ such that $R_0^\top R_0 \leq I_{l-r}$, that is the closed unit ball for the spectral norm $\bar{\mathcal{B}}_S(0, 1)$.

2. When $r = l$, the component Y_{l-r} of Y is the empty matrix. In other words, the dependence of the minimizing geodesic on R_0 vanishes so the minimizing geodesic is unique. In particular when $k = n$, $r = \text{rk}(\Sigma\Lambda) = \text{rk}(\Lambda) = l$.

3. On the contrary, when $r < l$ (thus $n > k \geq l$), there is an infinite number of convenient

R_0 's. For example, $R_0 = \begin{pmatrix} \cos \theta & \mathbf{0}_{1, l-r-1} \\ \mathbf{0}_{k-r-1, 1} & \mathbf{0}_{k-r-1, l-r-1} \end{pmatrix}$ and $R_1 = \begin{pmatrix} \sin \theta & \mathbf{0}_{1, l-r-1} \\ \mathbf{0}_{l-r-1, 1} & I_{l-r-1} \\ \mathbf{0}_{n-k-(l-r), 1} & \mathbf{0}_{n-k-(l-r), l-r-1} \end{pmatrix}$.

4. Since R_0 has more rows than columns, R_1 may be null which means that $R_0 \in \text{St}(k-r, l-r)$, that is $R_0^\top R_0 = I_{l-r}$. A simple calculus shows that $YR^\top = [Y_r \ Y_{l-r} R_0^\top \ \mathbf{0}_{n-k}]$. Therefore, the curve $c(t) = (1-t)X + tYR^\top$ has its $n-k$ columns identically null so it has rank less than k . But it also has rank at least k because $\text{rk}(\Sigma) = \text{rk}(X) = k$. So c and $\gamma_{\Sigma \rightarrow \Lambda}^{R_0}$ are of rank exactly k on $[0, 1)$ (and on $[0, 1]$ if $l = \text{rk}(\Lambda) = k$).

5. At the other extremity, there is $R_0 = 0$ (and $R_1 \in \text{St}(n-k, l-r)$). For $r = l$, it corresponds to the empty matrix. In this case, $XY^\top = X_r Y_r^\top$. Inspired by the case of the unique geodesic in $\text{Sym}^+(n, k)$ (with $k = l = r$), we notice that $XH^\dagger X^\top Y Y^\top = XH^\dagger H Y^\top = X \text{Diag}(I_r, 0) Y^\top = X_r Y_r^\top = XY^\top$. Therefore, denoting $X = UD^{1/2}V^\top$ with $U, V \in \text{St}(n, k)$ and $D \in \text{Diag}^+(k)$, we have:

$$\begin{aligned}
XY^\top &= XH^\dagger X^\top \Lambda \\
&= UD^{1/2}V^\top ((VD^{1/2}U^\top \Lambda UD^{1/2}V^\top)^{1/2})^\dagger VD^{1/2}U^\top \\
&= UD^{1/2}((D^{1/2}U^\top \Lambda UD^{1/2})^{1/2})^\dagger D^{1/2}U^\top \\
&= UD^{1/2}U^\top ((UD^{1/2}U^\top \Lambda UD^{1/2}U^\top)^{1/2})^\dagger UD^{1/2}U^\top \\
&= \Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda.
\end{aligned}$$

Thus the minimizing geodesic writes:

$$\gamma_{\Sigma \rightarrow \Lambda}^0(t) = (1-t)^2 \Sigma + t^2 \Lambda + 2t(1-t) \text{sym}(\Sigma^{1/2}((\Sigma^{1/2} \Lambda \Sigma^{1/2})^{1/2})^\dagger \Sigma^{1/2} \Lambda).$$

When the geodesic is unique, i.e. when $r = l$, i.e. when R_0 is the empty matrix, it has rank exactly k on $[0, 1)$.

□

Bibliography

- [Absil et al., 2009] Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.
- [Afsari, 2011] Afsari, B. (2011). Riemannian Lp center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673.
- [Afsari et al., 2013] Afsari, B., Tron, R., and Vidal, R. (2013). On the Convergence of Gradient Descent for Finding the Riemannian Center of Mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260. Publisher: Society for Industrial and Applied Mathematics.
- [Aleksandrov et al., 1986] Aleksandrov, A. D., Berestovskii, V. N., and Nikolaev, I. G. (1986). Generalized Riemannian spaces. *Russian Mathematical Surveys*, 41(3):1.
- [Alekseevsky, 1997] Alekseevsky, D. (1997). Flag manifolds. *Sbornik Radova*, 11:3–35.
- [Alekseevsky et al., 2001] Alekseevsky, D., Kriegl, A., Losik, M., and Michor, P. W. (2001). The Riemannian geometry of orbit spaces. The metric, geodesics, and integrable systems. *Publicationes Mathematicae*, 62:1–30.
- [Alexander et al., 2019] Alexander, S., Kapovitch, V., and Petrunin, A. (2019). Alexandrov geometry: foundations. pages 1–288.
- [Allen-Zhu et al., 2017] Allen-Zhu, Z., Li, Y., Oliveira, R., and Wigderson, A. (2017). Much faster algorithms for matrix scaling. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 890–901, Los Alamitos, CA, USA. IEEE Computer Society.
- [Amari, 1995] Amari, S.-I. (1995). Information geometry of the EM and em algorithms for neural networks. *Neural Networks*, 8(9):1379–1408.
- [Amari, 2014] Amari, S.-I. (2014). Information geometry of positive measures and positive-definite matrices: Decomposable dually flat structure. *Entropy*, 16(4):2131–2145.
- [Amari, 2016] Amari, S.-I. (2016). *Information Geometry and Its Applications*. Springer Publishing Company, Incorporated, 1st edition.
- [Amari and Armstrong, 2014] Amari, S.-I. and Armstrong, J. (2014). Curvature of Hessian manifolds. *Differential Geometry and its Applications*, 33:1–12.

- [Amari and Nagaoka, 2000] Amari, S.-I. and Nagaoka, H. (2000). *Methods of Information Geometry*, volume 191. Oxford University Press.
- [Archakov and Hansen, 2021] Archakov, I. and Hansen, P. R. (2021). A New Parametrization of Correlation Matrices. *Econometrica*, 89(4):1699–1715.
- [Arnaudon et al., 2013] Arnaudon, M., Barbaresco, F., and Yang, L. (2013). Medians and Means in Riemannian Geometry: Existence, Uniqueness and Computation. In Nielsen, F. and Bhatia, R., editors, *Matrix Information Geometry*, pages 169–197. Springer, Berlin, Heidelberg.
- [Arnaudon et al., 2012] Arnaudon, M., Dombry, C., Phan, A., and Yang, L. (2012). Stochastic algorithms for computing means of probability measures. *Stochastic Processes and their Applications*, 122(4):1437–1455.
- [Arnaudon and Miclo, 2014] Arnaudon, M. and Miclo, L. (2014). Means in complete manifolds: uniqueness and approximation. *ESAIM: Probability and Statistics*, 18:185–206.
- [Arsigny et al., 2006] Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2006). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421.
- [Artin, 2011] Artin, M. (2011). *Algebra*. Pearson Prentice Hall, Second edition.
- [Bacák, 2014] Bacák, M. (2014). Computing Medians and Means in Hadamard Spaces. *SIAM J. Optim.*, 24(3):1542–1566.
- [Banerjee et al., 2005] Banerjee, A., Merugu, S., Dhillon, I. S., and Ghosh, J. (2005). Clustering with Bregman Divergences. *Journal of Machine Learning Research*, 6(58):1705–1749.
- [Barachant et al., 2012] Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2012). Multiclass Brain-Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928. Publisher: Institute of Electrical and Electronics Engineers.
- [Barachant et al., 2013] Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, 112:172–178.
- [Barbaresco, 2013] Barbaresco, F. (2013). Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median. In Nielsen, F. and Bhatia, R., editors, *Matrix Information Geometry*, pages 199–255. Springer, Berlin, Heidelberg.
- [Barden and Le, 2018] Barden, D. and Le, H. (2018). The logarithm map, its limits and Fréchet means in orthant spaces. *Proceedings of the London Mathematical Society*, 117(4):751–789.

- [Batchelor et al., 2005] Batchelor, P. G., Moakher, M., Atkinson, D., Calamante, F., and Connelly, A. (2005). A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53(1):221–225.
- [Besse, 1987] Besse, A. L. (1987). *Einstein Manifolds*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, Berlin, Heidelberg.
- [Bhatia, 1997] Bhatia, R. (1997). *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer New York, New York, NY.
- [Bhatia and Holbrook, 2006] Bhatia, R. and Holbrook, J. (2006). Riemannian geometry and matrix geometric means. *Linear Algebra and its Applications*, 413(2-3):594–618.
- [Bhatia et al., 2019] Bhatia, R., Jain, T., and Lim, Y. (2019). On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191.
- [Bhattacharya and Patrangenaru, 2003] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31(1):1–29. Publisher: Institute of Mathematical Statistics.
- [Bhattacharya and Patrangenaru, 2005] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds—II. *The Annals of Statistics*, 33(3):1225–1259. Publisher: Institute of Mathematical Statistics.
- [Billera et al., 2001] Billera, L. J., Holmes, S. P., and Vogtmann, K. (2001). Geometry of the Space of Phylogenetic Trees. *Advances in Applied Mathematics*, 27(4):733–767.
- [Bonnabel, 2013] Bonnabel, S. (2013). Stochastic Gradient Descent on Riemannian Manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229. Conference Name: IEEE Transactions on Automatic Control.
- [Bonnabel and Sepulchre, 2010] Bonnabel, S. and Sepulchre, R. (2010). Riemannian Metric and Geometric Mean for Positive Semidefinite Matrices of Fixed Rank. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1055–1070.
- [Boumal et al., 2014] Boumal, N., Mishra, B., Absil, P.-A., and Sepulchre, R. (2014). Manopt, a Matlab Toolbox for Optimization on Manifolds. *Journal of Machine Learning Research*, 15:1455–1459.
- [Bourbaki, 1971] Bourbaki, N. (1971). *General Topology: Chapters 1–4*. Springer, Berlin, Heidelberg.
- [Breiding et al., 2018] Breiding, P., Kozhasov, K., and Lerario, A. (2018). On the Geometry of the Set of Symmetric Matrices with Repeated Eigenvalues. *Arnold Mathematical Journal*, 4:423–443.
- [Bridson and Haefliger, 1999] Bridson, M. R. and Haefliger, A. (1999). *Metric Spaces of Non-Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, Berlin, Heidelberg.

- [Bubeck, 2015] Bubeck, S. (2015). Convex Optimization: Algorithms and Complexity. *Foundations and Trends in Machine Learning*, 8(3-4):231–357.
- [Bures, 1969] Bures, D. (1969). An extension of Kakutani’s theorem on infinite product measures to the tensor product of semifinite w^* -algebras. *Transactions of the American Mathematical Society*, 135:199–212.
- [Burg, 1975] Burg, J. P. (1975). *Maximum Entropy Spectral Analysis*. PhD thesis, Stanford University.
- [Buser and Karcher, 1981] Buser, P. and Karcher, H. (1981). *Gromov’s almost flat manifolds*. Société Mathématique de France.
- [Calissano et al., 2020] Calissano, A., Feragen, A., and Vantini, S. (2020). Populations of Unlabeled Networks: Graph Space Geometry and Geodesic Principal Components. MOX Report.
- [Chewi et al., 2020] Chewi, S., Maunu, T., Rigollet, P., and Stromme, A. J. (2020). Gradient descent algorithms for Bures-Wasserstein barycenters. In *Conference on Learning Theory*, pages 1276–1304. PMLR.
- [Cichocki and Amari, 2010] Cichocki, A. and Amari, S.-I. (2010). Families of Alpha- Beta- and Gamma- Divergences: Flexible and Robust Measures of Similarities. *Entropy*, 12(6):1532–1568.
- [Cichocki et al., 2011] Cichocki, A., Cruces, S., and Amari, S.-I. (2011). Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. *Entropy*, 13(1):134–170.
- [David, 2019] David, P. (2019). *A Riemannian Quotient Structure for Correlation Matrices with Applications to Data Science*. PhD thesis, Institute of Mathematical Sciences, Claremont Graduate University.
- [David and Gu, 2019] David, P. and Gu, W. (2019). A Riemannian structure for correlation matrices. *Operators and Matrices*, 13(3):607–627.
- [de la Fuente et al., 2004] de la Fuente, A., Bing, N., Hoeschele, I., and Mendes, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, 20(18):3565–3574.
- [do Carmo, 1992] do Carmo, M. P. (1992). *Riemannian Geometry*. Mathematics: Theory and Applications. Birkhäuser.
- [Dodero et al., 2015] Dodero, L., Minh, H. Q., San Biagio, M., Murino, V., and Sona, D. (2015). Kernel-based classification for brain connectivity graphs on the Riemannian manifold of positive definite matrices. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 42–45. IEEE.
- [Dowson and Landau, 1982] Dowson, D. and Landau, B. (1982). The Fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis*, 12(3):450–455.

- [Dryden et al., 2009] Dryden, I. L., Koloydenko, A., and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 3(3):1102–1123.
- [Dryden et al., 2010] Dryden, I. L., Pennec, X., and Peyrat, J.-M. (2010). Power Euclidean metrics for covariance matrices with application to diffusion tensor imaging. ArXiv e-prints.
- [Epskamp and Fried, 2018] Epskamp, S. and Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4):617–634. Place: US Publisher: American Psychological Association.
- [Feragen et al., 2010] Feragen, A., Lauze, F., Lo, P., de Bruijne, M., and Nielsen, M. (2010). Geometries on spaces of treelike shapes. In *Proceedings of the 10th Asian conference on Computer vision - Volume Part II, ACCV'10*, pages 160–173, Berlin, Heidelberg. Springer-Verlag.
- [Feragen and Nye, 2020] Feragen, A. and Nye, T. (2020). Chapter 8. Statistics on stratified spaces. In Pennec, X., Sommer, S., and Fletcher, T., editors, *Riemannian Geometric Statistics in Medical Image Analysis*, pages 299–342. Academic Press.
- [Fillard et al., 2007] Fillard, P., Pennec, X., Arsigny, V., and Ayache, N. (2007). Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics. *IEEE Transactions on Medical Imaging*, 26(11):1472–82.
- [Fletcher, 2013] Fletcher, P. T. (2013). Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds. *International Journal of Computer Vision*, 105(2):171–185.
- [Fletcher and Joshi, 2007] Fletcher, P. T. and Joshi, S. (2007). Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data. *Signal Processing*, 87(2):250–262.
- [Fletcher and Joshi, 2004] Fletcher, P. T. and Joshi, S. C. (2004). Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors. In *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, ECCV 2004 Workshops CVAMIA and MMBIA*, Lecture Notes in Computer Science, pages 87–98, Prague, Czech Republic.
- [Fréchet, 1948] Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, 10(4):215–310.
- [Fuster et al., 2016] Fuster, A., Dela Haije, T., Tristán-Vega, A., Plantinga, B., Westin, C.-F., and Florack, L. (2016). Adjugate Diffusion Tensors for Geodesic Tractography in White Matter. *Journal of Mathematical Imaging and Vision*, 54(1):1–14.
- [Gallier, 2008] Gallier, J. (2008). Logarithms and Square Roots of Real Matrices. ArXiv e-prints.
- [Gallot et al., 2004] Gallot, S., Hulin, D., and Lafontaine, J. (2004). *Riemannian Geometry*. Springer-Verlag Berlin Heidelberg.

- [Garba et al., 2021] Garba, M. K., Nye, T. M. W., Lueg, J., and Huckemann, S. F. (2021). Information geometry for phylogenetic trees. *Journal of Mathematical Biology*, 82(3):1–39.
- [Ginestet et al., 2017] Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., and Kolaczyk, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, pages 725–750.
- [Groetzner and Dür, 2020] Groetzner, P. and Dür, M. (2020). A factorization method for completely positive matrices. *Linear Algebra and its Applications*, 591:1–24.
- [Groisser et al., 2017] Groisser, D., Jung, S., and Schwartzman, A. (2017). Geometric foundations for scaling-rotation statistics on symmetric positive definite matrices: Minimal smooth scaling-rotation curves in low dimensions. *Electronic Journal of Statistics*, 11(1):1092 – 1159.
- [Grubišić and Pietersz, 2007] Grubišić, I. and Pietersz, R. (2007). Efficient rank reduction of correlation matrices. *Linear Algebra and its Applications*, 422(2):629–653.
- [Guigui, 2021] Guigui, N. (2021). *Estimation statistique dans les variétés Riemanniennes: implémentation et application à l'étude des déformations cardiaques*. PhD thesis, Université Côte d'Azur.
- [Guigui et al., 2021] Guigui, N., Mocerri, P., Sermesant, M., and Pennec, X. (2021). Cardiac motion modeling with parallel transport and shape splines. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1394–1397. IEEE.
- [Guigui and Pennec, 2021a] Guigui, N. and Pennec, X. (2021a). Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. *Foundations of Computational Mathematics*. Conference Name: GSI 2021 - 5th conference on Geometric Science of Information.
- [Guigui and Pennec, 2021b] Guigui, N. and Pennec, X. (2021b). A Reduced Parallel Transport Equation on Lie Groups with a Left-Invariant Metric. In Nielsen, F. and Barbaresco, F., editors, *Geometric Science of Information*, Lecture Notes in Computer Science, pages 119–126, Cham. Springer International Publishing.
- [Hà Quang, 2019] Hà Quang, M. (2019). A Unified Formulation for the Bures-Wasserstein and Log-Euclidean/Log-Hilbert-Schmidt Distances between Positive Definite Operators. In *Proceedings of GSI 2019 - 4th conference on Geometric Science of Information*, volume 11712 of *Lecture Notes in Computer Science*, pages 475–483, Toulouse, France. Springer International Publishing.
- [Hà Quang, 2022] Hà Quang, M. (2022). Alpha Procrustes metrics between positive definite operators: a unifying formulation for the Bures-Wasserstein and Log-Euclidean/Log-Hilbert-Schmidt metrics. *Linear Algebra and its Applications*, 636:25–68.
- [Hà Quang et al., 2014] Hà Quang, M., San Biagio, M., and Murino, V. (2014). Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces. *Advances in neural information processing systems*, 27.

- [Han et al., 2021] Han, A., Mishra, B., Jawanpuria, P., and Gao, J. (2021). Generalized Bures-Wasserstein geometry for positive definite matrices. *arXiv preprint arXiv:2110.10464*.
- [Helstrom, 1967] Helstrom, C. W. (1967). Minimum mean-squared error of estimates in quantum statistics. *Physics letters A*, 25(2):101–102.
- [Hiai and Petz, 2009] Hiai, F. and Petz, D. (2009). Riemannian metrics on positive definite matrices related to means. *Linear Algebra and its Applications*, 430(11):3105–3130.
- [Hiai and Petz, 2012] Hiai, F. and Petz, D. (2012). Riemannian metrics on positive definite matrices related to means. II. *Linear Algebra and its Applications*, 436(7):2117–2136.
- [Huckemann et al., 2010] Huckemann, S., Hotz, T., and Munk, A. (2010). Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica*, 20(1):1–58.
- [Idel, 2016] Idel, M. (2016). A review of matrix scaling and Sinkhorn’s normal form for matrices and positive maps.
- [Johnson and Reams, 2009] Johnson, C. R. and Reams, R. (2009). Scaling of symmetric matrices by positive diagonal congruence. *Linear and Multilinear Algebra*, 57(2):123–140.
- [Journée et al., 2010] Journée, M., Bach, F., Absil, P.-A., and Sepulchre, R. (2010). Low-Rank Optimization on the Cone of Positive Semidefinite Matrices. *SIAM Journal on Optimization*, 20(5):2327–2351.
- [Kalantari, 2005] Kalantari, B. (2005). Matrix scaling dualities in convex programming. Technical report, Department of Computer Science, Rutgers University.
- [Karcher, 1977] Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541.
- [Katō, 1995] Katō, T. (1995). *Perturbation theory for linear operators*. Classics in mathematics. Springer, Berlin.
- [Kendall, 1984] Kendall, D. G. (1984). Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121.
- [Kendall, 1990] Kendall, W. (1990). Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406.
- [Kercheval, 2008] Kercheval, A. N. (2008). On Rebonato and Jäckel’s parametrization method for finding nearest correlation matrices. *International Journal of Pure and Applied Mathematics*, 45(3):383–390.
- [Khachiyan and Kalantari, 1992] Khachiyan, L. and Kalantari, B. (1992). Diagonal Matrix Scaling and Linear Programming. *SIAM Journal on Optimization*, 2(4):668–672.

- [Kim et al., 2021] Kim, K.-R., Dryden, I. L., Le, H., and Severn, K. E. (2021). Smoothing splines on Riemannian manifolds, with applications to 3D shape space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(1):108–132.
- [Koller and Friedman, 2009] Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. Adaptive computation and machine learning. MIT Press, Cambridge, MA.
- [Kroshnin et al., 2021] Kroshnin, A., Spokoiny, V., and Suvorikova, A. (2021). Statistical inference for Bures–Wasserstein barycenters. *The Annals of Applied Probability*, 31(3):1264–1298.
- [Kupczok et al., 2008] Kupczok, A., von Haeseler, A., and Klaere, S. (2008). An exact algorithm for the geodesic distance between phylogenetic trees. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 15(6):577–591.
- [Kurata and Bapat, 2016] Kurata, H. and Bapat, R. B. (2016). Moore-Penrose inverse of a hollow symmetric matrix and a predistance matrix. *Special Matrices*, 4(1).
- [Lauritzen, 1996] Lauritzen, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series. Oxford University Press, Oxford, New York.
- [Le, 2004] Le, H. (2004). Estimation of Riemannian barycenters. *LMS J. Comput. Math.*, 7:193–200.
- [Le Brigant et al., 2021] Le Brigant, A., Preston, S. C., and Puechmorel, S. (2021). Fisher-Rao geometry of Dirichlet distributions. *Differential Geometry and its Applications*, 74:101702.
- [Lee, 2012] Lee, J. M. (2012). *Introduction to Smooth Manifolds*. Springer, New York, NY, 2nd edition.
- [Lenglet et al., 2004] Lenglet, C., Deriche, R., and Faugeras, O. (2004). Inferring White Matter Geometry from Diffusion Tensor MRI: Application to Connectivity Mapping. In Pajdla, T. and Matas, J., editors, *Computer Vision - ECCV 2004*, volume 3024 of *Lecture Notes in Computer Science*, pages 127–140, Berlin, Heidelberg. Springer.
- [Lenglet et al., 2006] Lenglet, C., Rousson, M., Deriche, R., and Faugeras, O. (2006). Statistics on the Manifold of Multivariate Normal Distributions: Theory and Application to Diffusion Tensor MRI Processing. *Journal of Mathematical Imaging and Vision*, 25(3):423–444.
- [Li et al., 2017] Li, P., Wang, Q., Zeng, H., and Zhang, L. (2017). Local Log-Euclidean Multivariate Gaussian Descriptor and Its Application to Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):803–817.
- [Lim et al., 2019] Lim, L.-H., Sepulchre, R., and Ye, K. (2019). Geometric distance between positive definite matrices of different dimensions. *IEEE Transactions on Information Theory*, 65(9):5401–5405.

- [Lin, 2019] Lin, Z. (2019). Riemannian Geometry of Symmetric Positive Definite Matrices via Cholesky Decomposition. *SIAM Journal on Matrix Analysis and Applications*, 40(4):1353–1370.
- [Liu et al., 2012] Liu, M., Vemuri, B. C., Amari, S.-I., and Nielsen, F. (2012). Shape Retrieval Using Hierarchical Total Bregman Soft Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2407–2419. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [Lorenzi et al., 2011] Lorenzi, M., Ayache, N., and Pennec, X. (2011). Schild’s Ladder for the Parallel Transport of Deformations in Time Series of Images. In Székely, G. and Hahn, H. K., editors, *Information Processing in Medical Imaging*, pages 463–474, Berlin, Heidelberg. Springer.
- [Lorenzi and Pennec, 2013] Lorenzi, M. and Pennec, X. (2013). Efficient Parallel Transport of Deformations in Time Series of Images: from Schild’s to Pole Ladder. *Journal of Mathematical Imaging and Vision*, 50(1-2):5.
- [Malagò et al., 2018] Malagò, L., Montrucchio, L., and Pistone, G. (2018). Wasserstein Riemannian geometry of Gaussian densities. *Information Geometry*, 1(2):137–179.
- [Marrelec et al., 2006] Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehericy, S., Doyon, J., and Benali, H. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage*, 32(1):228–237.
- [Marshall and Olkin, 1968] Marshall, A. W. and Olkin, I. (1968). Scaling of matrices to achieve specified row and column sums. *Numerische Mathematik*, 12(1):83–90.
- [Marti et al., 2021] Marti, G., Goubet, V., and Nielsen, F. (2021). cCorrGAN: Conditional Correlation GAN for Learning Empirical Conditional Distributions in the Elliptope. In Nielsen, F. and Barbaresco, F., editors, *Geometric Science of Information*, Lecture Notes in Computer Science, pages 613–620, Cham. Springer International Publishing.
- [Massart and Absil, 2020] Massart, E. and Absil, P.-A. (2020). Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. *SIAM Journal on Matrix Analysis and Applications*, 41(1):171–198.
- [Massart et al., 2019] Massart, E., Hendrickx, J. M., and Absil, P.-A. (2019). Curvature of the Manifold of Fixed-Rank Positive-Semidefinite Matrices Endowed with the Bures–Wasserstein Metric. In Nielsen, F. and Barbaresco, F., editors, *Proceedings of GSI 2019 - 4th conference on Geometric Science of Information*, volume 11712 of *Lecture Notes in Computer Science*, pages 739–748, Toulouse, France. Springer International Publishing.
- [Mather, 1970] Mather, J. (2012 (1970)). Notes on Topological Stability. *Bulletin of the American Mathematical Society*, 49(4):475–506.
- [Michor, 2008] Michor, P. W. (2008). *Topics in Differential Geometry*, volume 93 of *Graduate Studies in Mathematics*. American Mathematical Soc.

- [Michor et al., 2000] Michor, P. W., Petz, D., and Andai, A. (2000). The Curvature of the Bogoliubov-Kubo-Mori Scalar Product on Matrices. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 3(2):1–14.
- [Miolane et al., 2020a] Miolane, N., Guigui, N., Le Brigant, A., Mathe, J., Hou, B., Thanwerdas, Y., Heyder, S., Peltre, O., Koep, N., Zaatiti, H., Hajri, H., Cabanes, Y., Gerald, T., Chauchat, P., Shewmake, C., Brooks, D., Kainz, B., Donnat, C., Holmes, S., and Pennec, X. (2020a). Geomstats: A Python Package for Riemannian Geometry in Machine Learning. *Journal of Machine Learning Research*, 21(223):1–9.
- [Miolane et al., 2020b] Miolane, N., Guigui, N., Zaatiti, H., Shewmake, C., Hajri, H., Brooks, D., Le Brigant, A., Mathe, J., Hou, B., Thanwerdas, Y., Heyder, S., Peltre, O., Koep, N., Cabanes, Y., Gerald, T., Chauchat, P., Kainz, B., Donnat, C., Holmes, S., and Pennec, X. (2020b). Introduction to Geometric Learning in Python with Geomstats. In *SciPy 2020 - 19th Python in Science Conference*, pages 48–57, Austin, Texas, United States.
- [Moakher, 2005] Moakher, M. (2005). A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747.
- [Monk, 1959] Monk, D. (1959). The geometry of flag manifolds. *Proceedings of the London Mathematical Society*, 3(2):253–286.
- [Mostajeran et al., 2020] Mostajeran, C., Grussler, C., and Sepulchre, R. (2020). Geometric matrix midranges. *SIAM Journal on Matrix Analysis and Applications*, 41(3):1347–1368.
- [Mostajeran and Sepulchre, 2018] Mostajeran, C. and Sepulchre, R. (2018). Ordering positive definite matrices. *Information Geometry*, 1(2):287–313.
- [Nielsen, 2022] Nielsen, F. (2022). The Many Faces of Information Geometry. *NOTICES OF THE AMERICAN MATHEMATICAL SOCIETY*, 69(1).
- [Nielsen and Nock, 2009] Nielsen, F. and Nock, R. (2009). Sided and Symmetrized Bregman Centroids. *IEEE Transactions on Information Theory*, 55(6):2882–2904. Conference Name: IEEE Transactions on Information Theory.
- [Nielsen and Sun, 2019] Nielsen, F. and Sun, K. (2019). Clustering in Hilbert’s Projective Geometry: The Case Studies of the Probability Simplex and the Elliptope of Correlation Matrices. In Nielsen, F., editor, *Geometric Structures of Information*, pages 297–331. Springer International Publishing, Cham.
- [Nye, 2011] Nye, T. M. W. (2011). Principal components analysis in the space of phylogenetic trees. *The Annals of Statistics*, 39(5):2716–2739. Publisher: Institute of Mathematical Statistics.
- [Nye et al., 2017] Nye, T. M. W., Tang, X., Weyenberg, G., and Yoshida, R. (2017). Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *Biometrika*, 104(4):901–922.

- [O’Leary, 2003] O’Leary, D. (2003). Scaling symmetric positive definite matrices to prescribed row sums. *Linear Algebra and its Applications*, 370:185–191.
- [Olkin and Pukelsheim, 1982] Olkin, I. and Pukelsheim, F. (1982). The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263.
- [Ollivier et al., 2017] Ollivier, Y., Arnold, L., Auger, A., and Hansen, N. (2017). Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles. *Journal of Machine Learning Research*, 18(18):1–65.
- [O’Neill, 1966] O’Neill, B. (1966). The fundamental equations of a submersion. *Michigan Mathematical Journal*, 13(4):459–469.
- [Owen and Provan, 2011] Owen, M. and Provan, J. S. (2011). A Fast Algorithm for Computing Geodesic Distances in Tree Space. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 8:2–13.
- [Paulin, 2014] Paulin, F. (2014). Groupes et géométries. Notes de cours.
- [Peng et al., 2009] Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial Correlation Estimation by Joint Sparse Regression Models. *Journal of the American Statistical Association*, 104(486):735–746.
- [Pennec, 2006] Pennec, X. (2006). Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127.
- [Pennec, 2009] Pennec, X. (2009). Statistical Computing on Manifolds: From Riemannian Geometry to Computational Anatomy. In *Emerging Trends in Visual Computing: LIX Fall Colloquium, ETVC 2008, Palaiseau, France, November 18-20, 2008. Revised Invited Papers*, volume 5416 of *Lecture Notes in Computer Science*, pages 347–386. Springer, Berlin, Heidelberg.
- [Pennec et al., 2006] Pennec, X., Fillard, P., and Ayache, N. (2006). A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66.
- [Pennec et al., 2020] Pennec, X., Sommer, S., and Fletcher, T. (2020). *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier.
- [Petz and Toth, 1993] Petz, D. and Toth, G. (1993). The Bogoliubov inner product in quantum statistics. *Letters in Mathematical Physics*, 27(3):205–216.
- [Pinheiro and Bates, 1996] Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296.
- [Ramgoolam, 2019] Ramgoolam, S. (2019). Permutation invariant Gaussian matrix models. *Nuclear Physics B*, 945:114682.

- [Rao, 1945] Rao, C. R. (1945). Information and the Accuracy Attainable in the Estimation of Statistical Parameters. *Bulletin of Calcutta Mathematical Society*, 37:81–91.
- [Rebonato and Jaeckel, 2001] Rebonato, R. and Jaeckel, P. (2001). The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk*, 2:17–27.
- [Serre, 2010] Serre, D. (2010). *Matrices: Theory and Applications*, volume 216 of *Graduate Texts in Mathematics*. Springer New York, Lyon, second edition.
- [Severn et al., 2019] Severn, K. E., Dryden, I. L., and Preston, S. P. (2019). Manifold valued data analysis of samples of networks, with applications in corpus linguistics. *arXiv preprint arXiv:1902.08290*.
- [Shima and Yagi, 1997] Shima, H. and Yagi, K. (1997). Geometry of Hessian manifolds. *Differential Geometry and its Applications*, 7(3):277–290.
- [Siegel, 1943] Siegel, C. L. (1943). Symplectic Geometry. *American J. of Mathematics*, 65(1):1–86.
- [Skovgaard, 1984] Skovgaard, L. T. (1984). A Riemannian Geometry of the Multivariate Normal Model. *Scandinavian Journal of Statistics*, 11(4):211–223.
- [Sommer et al., 2014] Sommer, S., Lauze, F., and Nielsen, M. (2014). Optimization over Geodesics for Exact Principal Geodesic Analysis. *Advances in Computational Mathematics*, 40.
- [Sturm, 2003] Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In Auscher, P., Coulhon, T., and Grigor’yan, A., editors, *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, volume 338, pages 357–390. American Mathematical Society, Providence, Rhode Island, contemporary mathematics edition.
- [Su et al., 2012] Su, J., Dryden, I. L., Klassen, E., Le, H., and Srivastava, A. (2012). Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds. *Image and Vision Computing*, 30(6):428–442.
- [Takatsu, 2010] Takatsu, A. (2010). On Wasserstein geometry of Gaussian measures. In Kotani, M., Hino, M., and Kumagai, T., editors, *Probabilistic Approach to Geometry*, volume 57 of *Advanced Studies in Pure Mathematics*, pages 463–472, Kyoto University, Japan. Mathematical Society of Japan.
- [Takatsu, 2011] Takatsu, A. (2011). Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics*, 48(4):1005–1026.
- [Thanwerdas, 2022] Thanwerdas, Y. (2022). Permutation-invariant log-Euclidean geometries on full-rank correlation matrices. Preprint.

- [Thanwerdas and Pennec, 2019a] Thanwerdas, Y. and Pennec, X. (2019a). Exploration of Balanced Metrics on Symmetric Positive Definite Matrices. In *Proceedings of GSI 2019 - 4th conference on Geometric Science of Information*, volume 11712 of *Lecture Notes in Computer Science*, pages 484–493, Toulouse, France. Springer International Publishing.
- [Thanwerdas and Pennec, 2019b] Thanwerdas, Y. and Pennec, X. (2019b). Is affine-invariance well defined on SPD matrices? A principled continuum of metrics. In *Proceedings of GSI 2019 - 4th conference on Geometric Science of Information*, volume 11712 of *Lecture Notes in Computer Science*, pages 502–510, Toulouse, France. Springer International Publishing.
- [Thanwerdas and Pennec, 2021] Thanwerdas, Y. and Pennec, X. (2021). Geodesics and Curvature of the Quotient-Affine Metrics on Full-Rank Correlation Matrices. In *Proceedings of GSI 2021 - 5th conference on Geometric Science of Information*, volume 12829 of *Lecture Notes in Computer Science*, pages 93–102, Paris, France. Springer International Publishing.
- [Thanwerdas and Pennec, 2022a] Thanwerdas, Y. and Pennec, X. (2022a). The geometry of mixed-Euclidean metrics on symmetric positive definite matrices. *Differential Geometry and its Applications*, 81:101867.
- [Thanwerdas and Pennec, 2022b] Thanwerdas, Y. and Pennec, X. (2022b). $O(n)$ -invariant Riemannian metrics on SPD matrices. Preprint.
- [Thanwerdas and Pennec, 2022c] Thanwerdas, Y. and Pennec, X. (2022c). Theoretically and computationally convenient geometries on full-rank correlation matrices. Preprint.
- [Thom, 1969] Thom, R. (1969). Ensembles et morphismes stratifiés. *Bulletin of the American Mathematical Society*, 75(2):240–284.
- [Tropp, 2018] Tropp, J. A. (2018). Simplicial Faces of the Set of Correlation Matrices. *Discrete & Computational Geometry*, 60(2):512–529.
- [Trotman, 2020] Trotman, D. (2020). Stratification theory. In *Handbook of Geometry and Topology of Singularities I*, pages 243–273. Springer.
- [Turner et al., 2014] Turner, K., Mileyko, Y., Mukherjee, S., and Harer, J. (2014). Fréchet Means for Distributions of Persistence Diagrams. *Discrete & Computational Geometry*, 52(1):44–70.
- [van Oostrum, 2020] van Oostrum, J. (2020). Bures-Wasserstein Geometry. ArXiv e-prints, submitted.
- [Vandereycken et al., 2009] Vandereycken, B., Absil, P.-A., and Vandewalle, S. (2009). Embedded geometry of the set of symmetric positive semidefinite matrices of fixed rank. In *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, pages 389–392. ISSN: 2373-0803.

- [Vandereycken et al., 2013] Vandereycken, B., Absil, P.-A., and Vandewalle, S. (2013). A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank. *IMA Journal of Numerical Analysis*, 33(2):481–514.
- [Varoquaux et al., 2010] Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P., and Thirion, B. (2010). Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In Shen, D., Frangi, A., and Szekely, G., editors, *Medical Image Computing and Computer Added Intervention*, Lecture Notes in Computer Science, pages 200–208, Beijing, China. Tianzi Jiang, Springer.
- [Wang et al., 2004] Wang, Z., Vemuri, B. C., Chen, Y., and Mareci, T. H. (2004). A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex dwi. *IEEE Transactions on Medical Imaging*, 23(8):930–939.
- [Yang, 2010] Yang, L. (2010). Riemannian median and its estimation. *LMS Journal of Computation and Mathematics*, 13:461–479.
- [Ye and Lim, 2016] Ye, K. and Lim, L.-H. (2016). Schubert varieties and distances between subspaces of different dimensions. *SIAM Journal on Matrix Analysis and Applications*, 37(3):1176–1197.
- [Zhang et al., 2018] Zhang, Z., Su, J., Klassen, E., Le, H., and Srivastava, A. (2018). Rate-Invariant Analysis of Covariance Trajectories. *Journal of Mathematical Imaging and Vision*, 60(8):1306–1323.