

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université Mustapha Stambouli - Mascara

جامعة مصطفى اسطبولي - معسكر

Faculté des Sciences Exactes
Département de Mathématiques

THÈSE DE DOCTORAT EN SCIENCES

Spécialité : Mathématiques Appliquées

Présentée par : **BOUREGAA Senouci**

Titre de la thèse : **Fonctions Convexes et les Barycentres dans des Variétés Différentielles**

Devant le jury composé de:

Président : Mr. **KACIMI Bouazza**, Professeur, Université Mustapha Stambouli – **Mascara**

Directeur de thèse : Mr. **SEGRES Abdelkader**, Professeur, Université Mustapha Stambouli – **Mascara**

Examineur : Mr. **ZAGANE Abdelkader**, Maître de Conférences, Université Mustapha Stambouli – **Mascara**

Examineur : Mr. **ZAHAF Mohammed brahim**, Professeur, Université Abou Bekr Belkaid **Tlemcen**

Examineur : Mr. **MESSIRDI Bachir**, Maître de Conférences, Université Abou Bekr Belkaid – **Tlemcen**

Examineur : Mr. **AZZOUZ Abdelhalim**, Professeur, Université Molay Tahar – **Saida**

Année Universitaire : 2025 / 2026

Dédicace

À ma mère , Toi qui as été le pilier silencieux de ma vie, ton amour inconditionnel, ta patience sans limites et tes invocations du fond du cœur m'ont accompagné dans les moments les plus sombres comme dans les plus lumineux. Ce travail t'appartient autant qu'à moi.

À la mémoire de mon père Ton absence physique n'a jamais effacé ta présence dans mon cœur. Tes paroles résonnent encore en moi comme un écho apaisant, et ton exemple de droiture, de travail et d'humilité continue de me guider. Que Dieu t'accorde Sa miséricorde et te pardonne.

À mon épouse , Ton soutien discret, ta compréhension dans mes absences, ton calme face à mes inquiétudes. Tu as porté avec moi le poids de cette aventure, sans jamais faiblir. Mille mercis ne suffiraient pas. Merci pour ta patience infinie , tes mots d'encouragement quand le doute m'assaillait, pour ton sacrifice discret derrière chaque page écrite.

À mes beaux parents

Je leur adresse ma gratitude la plus sincère pour leur sagesse et leur gentillesse, ils ont su créer un climat de paix et de respect qui m'a profondément touché. leur présence discrète mais précieuse tout au long de ce chemin.

À mes enfants,

Vous êtes ma source d'inspiration et de motivation. Vos sourires, vos regards innocents, vos gestes pleins d'amour m'ont rappelé chaque jour pourquoi il était important de persévérer. Ce travail est aussi un message pour vous et surtout Alaa Marwa : que chaque effort sincère finit toujours par porter ses fruits.

À ma famille,

Votre amour inconditionnel a été mon ancrage dans les tempêtes académiques. Merci d'avoir supporté mes absences, mes obsessions et mes doutes avec une patience et une bienveillance qui dépassent l'entendement.

Remerciements

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont contribué, directement ou indirectement, à la réalisation de ce travail doctoral.

À mon directeur de thèse, le Professeur [SEGRESS Abdelkader],

je tiens à exprimer toute ma reconnaissance pour votre encadrement exceptionnel, votre disponibilité et vos conseils avisés tout au long de ce parcours. Votre expertise, votre rigueur scientifique et votre soutien constant ont été déterminants pour l'aboutissement de cette recherche.

À Monsieur le Professeur [KACIMI Bouazza], président du jury,

je vous remercie sincèrement d'avoir accepté de présider cette soutenance. Vos remarques constructives et votre vision éclairée ont grandement contribué à l'amélioration de ce travail.

Je souhaite adresser mes remerciements particuliers aux membres du jury :

- **Professeur [AZZOUZ Abdelhalim]** pour ses suggestions pertinentes et son regard expert.
- **Professeur [ZAHAF Mohammed brahim]** pour ses commentaires constructifs et sa bienveillance.
- **Docteur [ZAGANE Abdelkader]** pour ses précieuses recommandations.
- **Docteur [MESSIRDI Bachir]** pour son temps et ses conseils avisés.

Je tiens également à exprimer ma reconnaissance à mon professeur **Michel EMERY**, qui m'a généreusement ouvert son cœur, transmis son savoir et guidé avec bienveillance mes premiers pas, essentiels et décisifs, dans ce domaine de recherche, ainsi qu'à mon professeur **Marc ARNAUDON** pour son aide précieuse.

Je n'oublie pas mes collègues des laboratoires de **Mascara** et de **Tlemcen**, pour leur collaboration, leurs échanges enrichissants et l'esprit de soutien qu'ils ont su instaurer.

Enfin, je remercie chaleureusement mes amis et ma famille pour leur soutien inconditionnel, leur patience et leurs encouragements constants, qui m'ont accompagné tout au long de cette aventure scientifique.

Introduction

La généralisation du concept d'espérance ou de moyenne pour des variables aléatoires à valeurs dans des espaces non linéaires constitue un domaine de recherche fondamental à l'interface entre la géométrie différentielle et la statistique. Dans les espaces euclidiens, la moyenne arithmétique joue un rôle central, à la fois comme mesure de tendance centrale et comme estimateur optimal au sens des moindres carrés. Cependant, lorsque l'on quitte le cadre linéaire pour des espaces courbes, comme les espaces métriques ou les variétés riemanniennes, la notion de moyenne doit être redéfinie pour tenir compte de la géométrie de l'espace.

Les premières tentatives de généralisation de la moyenne apparaissent avec M. Fréchet dès 1948 [33], qui introduisit la notion de *p-moyenne* d'une variable aléatoire X à valeurs dans un espace métrique (M, d) comme le point minimisant l'espérance de la distance à la puissance p :

$$x_p = \arg \min_{x \in M} \mathbb{E}[d^p(x, X)].$$

Les cas particuliers $p = 1$ et $p = 2$ correspondent respectivement à la *médiane* et à la *moyenne* de Fréchet. Cette approche a ouvert la voie à une généralisation naturelle de la moyenne dans les espaces non euclidiens. Dans ce cadre, la moyenne — ou plus précisément le *barycentre* — n'est plus nécessairement unique, car la convexité géodésique n'est pas toujours garantie .

La première construction rigoureuse de la moyenne sur les variétés riemanniennes revient à H. Karcher [48], qui définit la *moyenne riemannienne* comme le minimum de la fonctionnelle d'énergie géodésique :

$$F_\mu(x) = \int_M d^2(x, p) \mu(dp),$$

où d est la distance riemannienne et μ une mesure de probabilité sur M . Sous des conditions de courbure bornée et de rayon d'injectivité suffisant, F_μ est strictement convexe dans une boule géodésique $B(a, \rho)$ et admet un unique minimum $b(\mu)$, appelé *centre de Karcher* ou *barycentre riemannien*. En particulier, si Δ est une borne supérieure des courbures sectionnelles et $\text{inj}(a)$ le rayon d'injectivité en a , l'unicité est assurée pour :

$$\rho < \min \left(\frac{\pi}{4\sqrt{\Delta}}, \frac{\text{inj}(a)}{2} \right).$$

W. S. Kendall [51] a raffiné ce résultat en précisant les conditions garantissant l'unicité de la moyenne locale. Par la suite, M. Emery et G. Mokobodzki [31] ont étendu ces notions en introduisant les *barycentres exponentiels* et *convexes*, reliant la géométrie différentielle à la théorie des martingales continues.

L'étude des barycentres sur les variétés riemanniennes a trouvé des applications majeures dans le développement de la *statistique sur les variétés*. Des travaux pionniers de P. T. Fletcher, S. Joshi et S. M. Pizer [32] ont montré comment définir et estimer les moyennes et médianes intrinsèques sur des variétés pour l'analyse de formes et la modélisation de données géométriques. De même, X. Pennec [56] a proposé un cadre général de *statistiques riemanniennes* dans les espaces de matrices symétriques définies positives (SPD), ouvrant la voie à de nombreuses applications en traitement du signal, en vision par ordinateur et en apprentissage statistique.

Les barycentres jouent un rôle crucial dans la définition de la variance géométrique, des méthodes d'estimation robustes et des algorithmes d'optimisation géodésique. Ils permettent de généraliser des concepts fondamentaux comme la moyenne empirique, la régression linéaire ou encore l'analyse en composantes principales à des espaces courbes.

Outre les variétés riemanniennes abstraites, les *variétés matricielles* constituent des espaces géométriques concrets sur lesquels se développent aujourd'hui la majorité des applications statistiques et numériques.

1. La variété de Stiefel: La **variété de Stiefel** $\text{St}(p, n)$, introduite par Eduard Stiefel en 1935, est l'ensemble des matrices $X \in \mathbb{R}^{n \times p}$ dont les colonnes forment une base orthonormée d'un sous-espace de dimension p . Elle est définie par :

$$\text{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^\top X = I_p\}.$$

Cet espace intervient dans la factorisation QR, la réduction de dimension et les méthodes d'estimation sous contraintes orthogonales. Sa structure riemannienne a été formalisée par Absil, Mahony et Sepulchre [01], permettant la définition de géodésiques et de barycentres intrinsèques utilisés dans les algorithmes d'optimisation sur les variétés.

2. La variété de Grassmann: La **variété de Grassmann** $\text{Gr}(p, n)$, introduite historiquement par Hermann Grassmann en 1844 [52], représente l'ensemble des sous-espaces vectoriels de dimension p dans \mathbb{R}^n . Elle s'identifie comme un quotient de la variété de Stiefel : $\text{Gr}(p, n) = \text{St}(p, n)/O(p)$. Grâce à sa structure riemannienne canonique, on peut y définir des moyennes et barycentres géodésiques entre sous-espaces. Ces barycentres sont utilisés en apprentissage de sous-espaces [55], en reconnaissance de formes et dans l'analyse de trajectoires de données sur des espaces projectifs.

3. La variété des matrices SPD: La **variété des matrices symétriques définies positives** $\text{SPD}(n)$ a connu un essor majeur dans les années 2000, notamment grâce aux travaux de Pennec [56], Arsigny et al. [06]. Elle est définie par :

$$\text{SPD}(n) = \{X \in \mathbb{R}^{n \times n} : X = X^\top, X \succ 0\}.$$

Sous la *métrique d'invariance affine* :

$$g_X(U, V) = \text{Tr}(X^{-1}UX^{-1}V),$$

la distance géodésique entre $A, B \in \text{SPD}(n)$ s'écrit :

$$d(A, B) = \|\log(A^{-1/2}BA^{-1/2})\|_F,$$

et le barycentre (ou moyenne géométrique) d'un ensemble $\{A_i\}_{i=1}^N$ de matrices SPD est la solution unique de :

$$\sum_{i=1}^N \log(X^{-1/2}A_iX^{-1/2}) = 0.$$

Cette moyenne, dite *moyenne de Karcher sur SPD*, joue un rôle essentiel dans l'analyse de matrices de covariance, les filtres de Kalman riemanniens et les méthodes de classification sur les espaces de matrices .

L'objectif principal de ce travail est d'approfondir l'étude des *barycentres* et *p-moyennes de Fréchet* sur les variétés riemanniennes et matricielles, en mettant l'accent sur leurs conditions d'existence et d'unicité, leurs propriétés de convexité et leurs applications statistiques. Nous nous intéressons particulièrement aux barycentres sur les variétés de Stiefel, de Grassmann et sur la variété SPD, qui constituent des modèles géométriques adaptés aux données matricielles rencontrées dans la pratique.

La méthodologie adoptée repose sur une double approche :

- **Théorique** : fondée sur la géométrie riemannienne, la convexité géodésique et l'analyse variationnelle ;
- **Numérique** : développement d'algorithmes géodésiques et stochastiques pour le calcul des barycentres sur les variétés matricielles.

Ce manuscrit est organisé comme suit :

- **Chapitre 1** : rappels de géométrie différentielle et riemannienne
- **Chapitre 2** : cadre statistiques sur les variétés ; étude des conditions d'existence et d'unicité des barycentres et moyennes de Karcher, représentation des barycentres dans les variétés matricielles (Stiefel, Grassmann, SPD) et de leurs propriétés géométriques, extension des notions de moyenne à la statistique géométrique.
- **Chapitre 3** : développement et expérimentation d'algorithmes géodésiques et stochastiques pour des applications concrètes (économétrie financière , covariance, apprentissage).

Table des matières

1	Notions fondamentales de la géométrie différentielle	1
1.1	Définitions et exemples de variétés différentiables	1
1.1.1	Définitions	1
1.1.2	Exemples	3
1.1.3	Espaces tangents	6
1.1.4	Propriétés et exemples des espaces tangents	8
1.1.5	Flots des champs de vecteurs	10
1.2	Métriques riemanniennes	11
1.2.1	Définitions et exemples	11
1.2.2	Exemples de métriques riemanniennes	13
1.2.3	Application aux sous-variétés	13
1.2.4	Cas des sous-variétés	14
1.2.5	Exemples des métriques de variétés riemanniennes	14
1.3	Connexions affines et connexion de Levi-Civita	15
1.3.1	Connexions	15
1.3.2	Propriétés locales	15
1.3.3	La connexion de Levi-Civita	18
1.3.4	Interprétation	18
1.3.5	Exemples	20
1.4	Courbure	21
1.4.1	Tenseur Courbure	22
1.4.2	Propriétés des Courbures sectionnelles	23
1.5	Groupes de Lie, algèbres de Lie et sous-groupes de Lie	27
1.5.1	Groupe de Lie	27
1.5.2	Propriétés des champs de vecteurs invariants à gauche	29
1.5.3	L'application exponentielle	32
1.6	Action de groupe et espaces homogènes	35
1.6.1	Action de groupe	36
1.6.2	Variété quotient	37
1.7	fonctions convexes dans les variétés	41

1.7.1	Ensembles et fonctions convexes dans les espaces vectoriels	42
1.7.2	Ensembles et fonctions géodésiquement convexes	43
1.7.3	Sous-niveaux et intersections convexes	45
1.7.4	Fonctions différentiables géodésiquement convexes	48
2	Optimisations et Calculs statistiques dans les Variétés différentielles	53
2.1	Optimisation dans les variétés	53
2.1.1	Convexité géodésique forte et gradients Lipschitziens	53
2.2	Applications	57
2.2.1	Convexité et Sous variétés des réels positifs	57
2.2.2	Convexité et la variété des matrices symétriques définies positives	60
2.3	Calculs statistiques dans les Variétés	63
2.3.1	La moyenne de Fréchet	65
2.3.2	Généralisation de l'Analyse en composantes principales (ACP) dans les variétés	70
2.3.3	Distributions de probabilité	72
2.3.4	Echantillonnage	73
2.3.5	Régression géodésique	75
2.4	Calculs des moyennes de Fréchet dans des variétés matricielles	76
2.4.1	Moyennes des Matrices de Covariance	76
2.4.2	Calcul de moyenne de Fréchet dans la variété de Stiefel	80
2.4.3	Calcul de moyenne de Fréchet dans la variété de Grassmann	85
2.5	barycentres et convexité dans les variétés différentielles	91
2.5.1	Définitions des barycentres et Exemples	91
2.5.2	les barycentres géodesiques dans les variétés	93
	Conclusion	99
3	Calculs explicites des barycentres dans les variétés et applications	101
3.1	Calculs des barycentres géodésiques dans les variétés à courbures constantes	101
3.1.1	Trigonométrie Sphérique et Hyperbolique	101
3.1.2	Barycentres géodésiques des triangles	102
3.1.3	Estimation de rayon des barycentres	106
3.2	Barycentres de Matrices de Covariance et Entropie Riemannienne pour l'Analyse du Risque Structurel sur les Marchés Financiers	111
3.2.1	Prologue :	111
3.2.2	Développement : Méthodologie, résultats et discussion	112
3.2.3	Epilogue :	130

Références 132

Chapitre 1

Notions fondamentales de la géométrie différentielle

Ce chapitre a pour objectif de poser les fondements théoriques nécessaires à la bonne compréhension de la suite de ce travail. Nous y présentons les définitions et principaux théorèmes qui seront utilisés dans la suite de ces chapitres.

1.1 Définitions et exemples de variétés différentiables

1.1.1 Définitions

Définition 1 ► Atlas

Soit M un espace topologique. Un *atlas* \mathcal{A} de classe C^k à valeurs dans \mathbb{R}^d est un ensemble de couples (U, φ) , appelés *cartes*, où $\varphi : U \rightarrow V$ est un homéomorphisme entre les ouverts $U \subset M$ et $V \subset \mathbb{R}^d$, tel que :

1. $M = \bigcup_{(U, \varphi)} U$
2. Pour toute paire de cartes (U, φ) et (U', φ') dans \mathcal{A} , l'*application de transition*

$$\varphi' \circ \varphi^{-1} : \varphi(U \cap U') \rightarrow \varphi'(U \cap U')$$

est un *difféomorphisme de classe C^k* .

Deux atlas sont C^k -compatibles si leur union reste un atlas de classe C^k . La compatibilité définit une relation d'équivalence, et on considérera toujours la classe d'équivalence d'un atlas. Il existe un unique atlas *maximal* (pour l'inclusion) contenant un atlas donné.

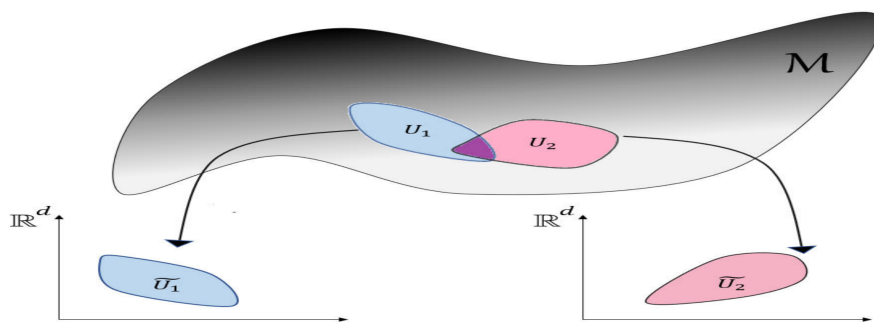


FIGURE 1.1 – Illustration d'un atlas : la variété M est recouverte par des ouverts U_i et envoyée par un homéomorphisme φ vers des ouverts $\tilde{U}_i \subset \mathbb{R}^d$.

Remarque 1

Les applications de transition étant définies entre ouverts de \mathbb{R}^d , les notions usuelles de différentiabilité s'appliquent, permettant d'étendre ces concepts aux variétés.

Définition 2 ► Variété différentielle

Une *variété différentielle* de classe C^k et de dimension d est un espace topologique M , séparé (Hausdorff) et à base dénombrable, muni d'un atlas maximal \mathcal{A} de classe C^k à valeurs dans \mathbb{R}^d .

On appelle parfois l'atlas de M sa *structure différentielle*. Cette définition est dite *intrinsèque* car les cartes sont définies sur M via un plongement dans un espace ambiant. Une carte (U, φ) définit des *coordonnées locales* (x^1, \dots, x^d) sur U , données par $x^i = pr_i(\varphi(x))$, où pr_i est la projection sur la $i^{\text{ème}}$ coordonnée de \mathbb{R}^d .

1.1.2 Exemples

Exemple 1 ► Hypersphère

La variété la plus simple que nous étudierons est l'hypersphère, ou sphère de dimension d . Elle correspond à l'ensemble des vecteurs de norme égal à 1 dans \mathbb{R}^{d+1} :

$$S^d = \{x \in \mathbb{R}^{d+1} \mid \|x\|^2 = 1\}.$$

En particulier :

1. En dimension $d = 1$, cela correspond au cercle,
2. En dimension $d = 2$, c'est la sphère usuelle.

Ces deux cas sont couramment utilisés pour représenter des angles et des directions dans l'espace, et apparaissent ainsi dans le domaine des statistiques directionnelles.

Exemple 2 ► Espace hyperbolique

L'hyperboloïde à deux nappes est défini par :

$$H^d = \{x \in \mathbb{R}^{d+1} \mid -x_0^2 + \sum_{i=1}^d x_i^2 = -1\}.$$

Un modèle de géométrie hyperbolique, utilisé pour les données hiérarchiques.

Exemple 3 ► Groupe spécial orthogonal

$$SO(n) = \{R \in M_n(\mathbb{R}) \mid R^\top R = I_n, \det(R) = 1\}.$$

$SO(n)$ est une variété plongée de dimension $\frac{n(n-1)}{2}$.

Exemple 4 ► Variété produit

Soient M et N deux variétés munies d'atlas respectifs $(U_i, \varphi_i)_{i \in I}$ et $(V_j, \psi_j)_{j \in J}$.

Pour tout $(i, j) \in I \times J$, on définit :

$$\phi_{ij} : \begin{cases} U_i \times V_j & \longrightarrow & \varphi_i(U_i) \times \psi_j(V_j) \\ (x, y) & \longmapsto & (\varphi_i(x), \psi_j(y)) \end{cases}$$

Alors $(U_i \times V_j, \varphi_{ij})_{(i,j) \in I \times J}$ forme un atlas pour l'espace produit $M \times N$. Cet atlas est indépendant du choix des atlas initiaux (dans leur classe d'équivalence) et permet de munir $M \times N$ d'une structure de variété.

La construction précédente munit $M \times N$ d'une structure de variété, appelée *variété produit* de M et N . Sa dimension est donnée par :

$$\dim(M \times N) = \dim M + \dim N.$$

Exemple 5 ► Variété de Grassmanne

Les variétés grassmanniennes généralisent l'espace projectif.

Pour $1 \leq m \leq N - 1$, on définit :

$$Gr(m, \mathbb{R}^N) = \left\{ \text{sous-espaces vectoriels } V \text{ de dimension } m \text{ de } \mathbb{R}^N \right\}.$$

On observe que $Gr(1, \mathbb{R}^N) = \mathbb{P}^{N-1}(\mathbb{R})$. Un sous-espace V est déterminé par une base de m vecteurs linéairement indépendants, représentée par une matrice $m \times N$ de rang m . Deux matrices M et M' définissent le même sous-espace si et seulement si $M' = gM$ pour $g \in GL(m, \mathbb{R})$: c'est une relation d'équivalence. Ainsi :

$$Gr(m, \mathbb{R}^N) = \{[M] : M \text{ est une matrice } m \times N \text{ de rang } m\},$$

où $[M]$ désigne la classe d'équivalence de M .

Construction de l'atlas :

Pour une suite d'indices $I = \{i_1 < \dots < i_m\} \subseteq \{1, \dots, N\}$, notons M_I la sous-matrice $m \times m$ formée des colonnes i_1, \dots, i_m . L'atlas $\mathcal{A} = \{(U_I, \varphi_I)\}$ est défini par :

$$U_I = \{[M] \in Gr(m, \mathbb{R}^N) : \det(M_I) \neq 0\},$$
$$\varphi_I : U_I \rightarrow \mathbb{R}^{m(N-m)}, \quad \varphi_I([M]) = (M_I^{-1}M)_{I^c},$$

où $(M_I^{-1}M)_{I^c}$ est la sous-matrice $m \times (N - m)$ obtenue en supprimant les colonnes indexées par I . Le changement de cartes entre φ_I et φ_J est donné par :

$$\varphi_J \circ \varphi_I^{-1}(M_{I^c}) = (M_I^{-1}M)_{J^c}.$$

Ces applications sont lisses car définies par des fonctions rationnelles (les dénominateurs $\det(M_I)$ et $\det(M_J)$ sont non nuls sur les domaines d'intersection). La dimension de $Gr(m, \mathbb{R}^N)$ est donc $m(N - m)$.

Structure différentiable de la variété de Grassmanne :

Soit \mathcal{M} une matrice $m \times N$ telle que M_I soit la matrice identité. Le changement de cartes est donné par des fonctions rationnelles dont les dénominateurs ne s'annulent pas (car $\det(M_J) \neq 0$). Les ouverts U_I recouvrent $Gr(m, \mathbb{R}^N)$, ce qui définit une structure de variété C^∞ de dimension $m(N - m)$.

Cas particulier La variété $Gr(2, \mathbb{R}^4)$, espace des plans vectoriels dans \mathbb{R}^4 , est de dimension 4. Son atlas contient six cartes correspondant aux mineurs 2×2 non nuls. Considérons $I = \{1, 2\}$ et $J = \{1, 3\}$. Pour $[M] \in U_I \cap U_J$, il existe des représentants uniques :

$$M = \begin{pmatrix} 1 & 0 & x_1 & x_2 \\ 0 & 1 & x_3 & x_4 \end{pmatrix}, \quad M' = \begin{pmatrix} 1 & y_1 & 0 & y_2 \\ 0 & y_3 & 1 & y_4 \end{pmatrix},$$

tels que $[M] = [M']$. Le changement de cartes $\varphi_J \circ \varphi_I^{-1}$ se calcule par :

$$\begin{aligned} \begin{pmatrix} 1 & y_1 & 0 & y_2 \\ 0 & y_3 & 1 & y_4 \end{pmatrix} &= \begin{pmatrix} 1 & x_1 \\ 0 & x_3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & x_1 & x_2 \\ 0 & 1 & x_3 & x_4 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -\frac{x_1}{x_3} & 0 & x_2 - \frac{x_1 x_4}{x_3} \\ 0 & \frac{1}{x_3} & 1 & \frac{x_4}{x_3} \end{pmatrix}. \end{aligned}$$

Les coordonnées (y_1, y_2, y_3, y_4) sont donc des fonctions rationnelles lisses de (x_1, x_2, x_3, x_4) , avec $x_3 \neq 0$ sur $\varphi_I(U_I \cap U_J)$. Ceci confirme la régularité C^∞ du changement de cartes.

1.1.3 Espaces tangents

Nous définissons d'abord les *courbes lisses* sur une variété, en utilisant une paramétrisation locale de M et la notion de fonction lisse de \mathbb{R} dans \mathbb{R}^d .

Définition 3 ► Application lisse

Soient M, Q deux variétés de dimensions $d_1, d_2 \in \mathbb{N}$ plongées dans \mathbb{R}^N . Une application $f : M \rightarrow Q$ est dite *lisse* si pour tout $p \in M$, il existe des paramétrisations $\varphi : V_1 \rightarrow U_1$ de M au point p et $\psi : V_2 \rightarrow U_2$ de Q au point $f(p)$ telles que

$f(U_1) \subseteq U_2$ et

$$\psi^{-1} \circ f \circ \varphi : V_1 \rightarrow V_2 \quad \text{est lisse.}$$

Définition 4 ► Courbe lisse

Soit M une variété de dimension d dans \mathbb{R}^N . Une *courbe lisse* γ dans M est une fonction $\gamma : I \rightarrow M$ où I est un intervalle ouvert de \mathbb{R} , telle que pour tout $t \in I$, en posant $p = \gamma(t)$, il existe une paramétrisation locale $\varphi : V \rightarrow U$ de M au point p et $\epsilon > 0$ tels que $\varphi^{-1} \circ \gamma : (t - \epsilon, t + \epsilon) \rightarrow \mathbb{R}^d$ soit lisse.

Cette définition s'étend aux courbes définies sur un intervalle fermé $I = [a, b]$ en exigeant que γ soit la restriction d'une courbe lisse définie sur un intervalle ouvert contenant $[a, b]$.

Définition 4 ► Vecteur tangent

Soient $d \leq N \in \mathbb{N}$, M une variété plongée dans \mathbb{R}^N de dimension d et $p \in M$. Un vecteur $v \in \mathbb{R}^N$ est *tangent* à M en p s'il existe un intervalle ouvert I centré en 0 et une courbe $\gamma : I \rightarrow M$ tels que

$$\gamma(0) = p \quad \text{et} \quad \dot{\gamma}(0) = v.$$

On note $T_p M$ l'ensemble des vecteurs tangents en p .

Théorème 1 [39]

Soit M une variété dans \mathbb{R}^N de dimension d .

1. **(Difféomorphisme local)** Si $U, V \subset \mathbb{R}^N$ sont deux voisinages ouverts respectifs de p et 0 dans \mathbb{R}^N et $f : U \rightarrow V$ est un difféomorphisme vérifiant $f(p) = 0$ et $f(U \cap M) = V \cap (\mathbb{R}^d \times \{0\})$, alors

$$T_p M = df_p^{-1}(\mathbb{R}^d \times \{0\}).$$

2. **(Paramétrisation locale)** Si $U \subseteq \mathbb{R}^N$ est un voisinage ouvert de $p \in M$, $V \subseteq \mathbb{R}^d$ un voisinage ouvert de 0 et $f : V \rightarrow \mathbb{R}^N$ est une fonction lisse telle que $f(0) = p$, f est un homéomorphisme entre V et $U \cap M$, et f est une immersion en 0, alors

$$T_p M = \text{Im } df_0.$$

3. **(Fonction implicite locale)** Si $U \subseteq \mathbb{R}^N$ est un voisinage ouvert de $p \in M$ et $f : U \rightarrow \mathbb{R}^{N-d}$ est une application lisse qui est une submersion en p , avec $U \cap M = f^{-1}(\{0\})$, alors

$$T_p M = \ker df_p.$$

4. **(Graphe local)** Si $U \subseteq \mathbb{R}^N$ est un voisinage ouvert de $p \in M$, $V \subseteq \mathbb{R}^d$ un voisinage de 0 et $f : V \rightarrow \mathbb{R}^{N-d}$ une application lisse telle que $U \cap M = \text{graph}(f)$ et $p = (0, f(0))$, alors

$$T_p M = \text{Im}\{v \mapsto (v, df_0(v))\}.$$

1.1.4 Propriétés et exemples des espaces tangents

D'après (1) on voit que T_pM est un sous-espace vectoriel de \mathbb{R}^N de dimension d . Les espaces tangents fournissent ainsi des linéarisations locales de la variété, propriété qui sera utile pour manipuler des données sur les variétés. Le théorème précédent permet de calculer les espaces tangents des variétés classiques vues précédemment.

Exemple 6 ► Espace tangent de Hypersphère

Rappelons que l'hypersphère est la variété plongée définie par $S^d = f^{-1}(0)$ où $f : x \mapsto \|x\|^2 - 1$. Ceci correspond au cas (3) du Théorème 1, donc pour tout $x \in S^d$:

$$T_x S^d = \{v \in \mathbb{R}^{d+1} \mid \langle x, v \rangle = 0\}.$$

Exemple 7 ► Espace tangent de Espace hyperbolique

De même, comme l'espace hyperbolique est défini par $H^d = f^{-1}(0)$ où $f : x \mapsto -x_0^2 + \sum_{i=1}^d x_i^2 + 1$, on obtient pour tout $x \in H^d$:

$$T_x H^d = \{v \in \mathbb{R}^{d+1} \mid -x_0 v_0 + \sum_{i=1}^d x_i v_i = 0\}.$$

Exemple 8 ► Groupe spécial orthogonal

Rappelons que $SO(n) = f^{-1}(0)$ avec $f : A \mapsto A^\top A - I_n$, et pour tout $R \in SO(n)$, $H \in M_n(\mathbb{R})$ on a $df_R(H) = R^\top H + H^\top R$. Ainsi, pour tout $R \in SO(n)$:

$$T_R SO(n) = \{H \in M_n(\mathbb{R}) \mid R^\top H + H^\top R = 0\}.$$

Notons le cas particulier $R = I_n$, alors $T_{I_n} SO(n) = \text{Skew}(n)$, l'ensemble des matrices antisymétriques de taille n .

Exemple 9 ► Espace tangent de la Variété de Stiefel

l'ensemble des matrices X rectangulaires vérifiant $X^\top X = I$ forment la variété $St(k, n)$

Et de manière analogue, pour tout $U \in St(k, n)$:

$$T_U St(k, n) = \{H \in M_{n,k}(\mathbb{R}) \mid U^\top H + H^\top U = 0\}.$$

Définition 6 ► Différentielle

Soient M, Q deux variétés de dimensions $d_1, d_2 \in \mathbb{N}$ plongées dans \mathbb{R}^N et $f : M \rightarrow Q$ une application lisse. Pour tout $p \in M$ et tout $v \in T_p M$, soit γ une courbe lisse passant par p telle que $\dot{\gamma}(0) = v$. On définit :

$$df_p(v) = (f \circ \gamma)'(0).$$

Cette définition ne dépend pas du choix de la courbe γ et l'application $df_p : T_p M \rightarrow T_{f(p)} Q$ est appelée *différentielle* ou *application tangente* de f en p . C'est une application linéaire entre espaces tangents.

Remarque 2 ► Généralisation et propriétés

Cette définition généralise la différentielle des fonctions définies sur \mathbb{R}^{N_1} à valeurs dans \mathbb{R}^{N_2} au cas des variétés. Elle coïncide avec la différentielle usuelle lorsque $M = \mathbb{R}^d$, d'où l'usage de la même notation df_p .

L'ensemble des applications lisses à valeurs réelles sur M , noté $C^\infty(M) \triangleq C^\infty(M, \mathbb{R})$, est particulièrement utile. C'est un \mathbb{R} -espace vectoriel de dimension infinie et une algèbre pour la multiplication ponctuelle.

Définition 7 ► Fibré tangent

On considère l'ensemble de tous les espaces tangents en tous points :

$$TM = \bigsqcup_{x \in M} \{x\} \times T_x M = \{(x, v) \mid x \in M, v \in T_x M\}.$$

muni de sa projection naturelle :

$$\pi : \begin{cases} TM & \longrightarrow & M \\ (x, v) & \longmapsto & x \end{cases}.$$

Cet espace est appelé *fibré tangent* de M .

On peut montrer que si M est une variété de classe C^{k+1} et dimension d , alors TM est lui-même une variété dans $\mathbb{R}^N \times \mathbb{R}^N$, de classe C^k et dimension $2d$.

Définition 8 ► Champs de vecteurs

Le fibré tangent est le domaine de définition de la différentielle des fonctions lisses :

$$f : M \rightarrow Q, \quad df : TM \rightarrow TQ$$

C'est aussi l'espace dans lequel prennent valeur les *champs de vecteurs* : un champ de vecteurs X est une assignation lisse d'un vecteur tangent à chaque point de la variété, c'est-à-dire $X : M \rightarrow TM$ telle que $\forall p \in M, \pi \circ X(p) = p$. On notera $X(p)$ par X_p par commodité.

L'ensemble de tous les champs de vecteurs est noté $\Gamma(TM)$. Muni de l'addition et multiplication par un scalaire ponctuelles, $\Gamma(TM)$ forme un espace vectoriel. La multiplication par une fonction lisse est également définie ponctuellement : pour tout $f \in C^\infty(M)$ et $X \in \Gamma(TM)$, fX est le champ de vecteurs tel que :

$$\forall p \in M, (fX)_p = f(p)X_p.$$

1.1.5 Flots des champs de vecteurs

Les champs de vecteurs peuvent être vus comme des générateurs infinitésimaux d'applications locales appelées flots, qui renseignent sur les propriétés globales de la variété.

Définition 9 ► Courbe intégrale

Soit $X \in \Gamma(M)$ un champ de vecteurs et $p_0 \in M$. Une *courbe intégrale* pour X avec condition initiale p_0 est une courbe $\gamma : I \rightarrow M$ telle que :

$$\forall t \in I, \quad \dot{\gamma}(t) = X_{\gamma(t)} \quad \text{et} \quad \gamma(0) = p_0,$$

où $I = (a, b) \subseteq \mathbb{R}$ est un intervalle ouvert contenant 0.

Une courbe intégrale est donc une courbe dont le vecteur vitesse $\dot{\gamma}(t)$ coïncide avec X en tout point de la courbe. Une famille de telles courbes est appelée un flot :

Définition 10 ► Flot local

Soit $X \in \Gamma(M)$ et $p_0 \in M$. Un *flot local* de X en p_0 est une application $\phi : I \times U \rightarrow M$ où :

1. I est un intervalle ouvert contenant 0
2. U est un ouvert de M contenant p_0

telle que pour tout $p \in U$, la courbe $t \mapsto \phi(t, p)$ soit une courbe intégrale de X partant de p .

(Existence et propriétés) Grâce à la théorie des équations différentielles ordinaires (EDO), on peut montrer que pour tout champ de vecteurs, il existe un flot local défini autour de chaque point. Si deux flots sont définis sur des domaines qui se recouvrent, ils coïncident sur l'intersection.

Pour $t \in I$, on note $\phi_t : x \mapsto \phi(t, x)$. On a :

1. ϕ_0 est l'application identité
2. Pour $t \neq 0$, ϕ_t est une application localement définie sur M

Proposition 1

Soit X un champ de vecteurs lisse sur M , $p_0 \in M$ et $\phi : I \times U \rightarrow M$ le flot local de X en p_0 . Pour tous $s, t \in I$ et $x \in U$:

1. Si $\phi_s(x) \in U$ et $t + s \in I$, alors $\phi_t \circ \phi_s(x) = \phi_{t+s}(x)$
2. ϕ_t est un difféomorphisme local
3. ϕ_t préserve X , c'est-à-dire $\forall t \in I, \forall x \in U, d(\phi_t)_x(X_x) = X_{\phi_t(x)}$

Définition 11 ► Completude

On dit que X est *complet* si le domaine de définition de son flot $(t, x) \mapsto \phi_t(x)$ est $\mathbb{R} \times M$ tout entier. Si tous les champs de vecteurs de M sont complets, on dit que M est complète.

Dans ce cas, ϕ_t est un difféomorphisme de M , et $(\phi_t)_{t \in \mathbb{R}}$ forme un *groupe à un paramètre* de $\text{Diff}(M)$.

1.2 Métriques riemanniennes

1.2.1 Définitions et exemples

Nous introduisons maintenant une nouvelle structure sur une variété différentiable : la métrique riemannienne, qui permet de définir la longueur d'une courbe,

une fonction distance, une forme volume,

Définition 12 ► Métrique riemannienne

Soit M une variété différentiable de dimension d . Une *métrique riemannienne* sur M (ou sur TM) est une famille $(\langle \cdot, \cdot \rangle_p)_{p \in M}$ de produits scalaires sur $T_p M$. Chaque espace tangent $T_p M$ est muni d'un produit scalaire $\langle \cdot, \cdot \rangle_p$ qui dépend de manière lisse de p . Plus formellement, pour toute carte φ, U et tout repère (X_1, \dots, X_n) sur U , les applications

$$p \mapsto \langle X_i(p), X_j(p) \rangle_p \quad \text{pour } 1 \leq i, j \leq n,$$

sont lisses. Le couple $(M, \langle \cdot, \cdot \rangle)$ est appelé *variété riemannienne*.

Notations alternatives Une métrique est souvent notée $g = (g_p)_{p \in M}$, où g_p est la matrice symétrique définie positive représentant le produit scalaire dans une carte :

$$g_{ij}(p) = \langle (\partial_i)_p, (\partial_j)_p \rangle_p.$$

On peut aussi définir une métrique par la notation $g = f(dx_1, \dots, dx_n)$ où f est la forme quadratique associée à g , et dx_i représentent les coordonnées vectorielles (comme formes linéaires). Par exemple, la métrique euclidienne usuelle s'écrit $g = \sum_{i=1}^n dx_i^2$.

Théorème 2 ► Existence

Toute variété différentiable admet une métrique riemannienne.

Proposition 2 ► Norme induite

Une métrique riemannienne définit une norme sur TM en chaque espace tangent :

$$\forall x \in M, \forall v \in T_x M, \quad \|v\|_x = \sqrt{g_x(v, v)} = \sqrt{\langle v, v \rangle_x}. \quad (1.1)$$

1.2.2 Exemples de métriques riemanniennes

Exemple 10 ► Métrique euclidienne

Soit $M = \mathbb{R}^d$ l'espace vectoriel standard de dimension d , qui est trivialement une variété différentiable. Considérons son produit scalaire standard défini pour tous $x, y \in \mathbb{R}^d$ par :

$$\langle x, y \rangle_2 = \sum_{i=1}^d x_i y_i = x^\top y.$$

Comme $T_x \mathbb{R}^d = \mathbb{R}^d$, cela définit une métrique sur \mathbb{R}^d , appelée *métrique euclidienne*.

Exemple 11 ► Métrique produit

Soient (M, g) et (M', g') deux variétés riemanniennes. Rappelons que , le produit cartésien $M \times M'$ est une variété. C'est également une variété riemannienne. En effet, on définit la *métrique produit* $g \oplus g'$ comme l'application qui à tout $(x, x') \in M \times M'$ et à tout $(v, v'), (w, w') \in T_x M \times T_{x'} M'$ associe :

$$g \oplus g'_{(x, x')}((v, v'), (w, w')) = g(v, w) + g'(v', w'). \quad (1.2)$$

1.2.3 Application aux sous-variétés

Soient (N, g) une variété riemannienne, M une variété différentiable et $f : M \rightarrow N$ une

Métrique image réciproque : Soit $f : M \rightarrow N$ une application différentiable. On définit la *métrique image réciproque* (f^*g) sur M par :

$$(f^*g)_x : \begin{cases} T_x M \times T_x M & \longrightarrow & \mathbb{R} \\ (v, w) & \longmapsto & g_{f(x)}(df_x v, df_x w) \end{cases}.$$

Si f est une immersion et si $(f^*g)_x$ est non-dégénérée et de signature constante pour tout $x \in M$, alors (M, f^*g) est une variété riemannienne.

Définition 13 ► Isométrie

Soient (M, g) et (M', g') deux variétés riemanniennes et $f : M \rightarrow M'$. L'application f est appelée *isométrie* si c'est une bijection et si $f^*g' = g$.

Lorsque $M = M'$ et $g = g'$, on note $\text{Isom}(M)$ l'ensemble des isométries de M . Myers et Steenrod (1939) ont montré que c'est un groupe de Lie agissant de manière lisse sur M [39].

1.2.4 Cas des sous-variétés

Considérons $M \subseteq N = \mathbb{R}^d$ une sous-variété de N , et $f = i$ l'inclusion canonique. La métrique g est appelée *métrique d'embedding* et i^*g sa restriction à M .

Exemple 12 ► Hypersphère

L'hypersphère $S^d \subset \mathbb{R}^{d+1}$ munie de la restriction de la métrique euclidienne ambiante est une variété riemannienne. On appelle cette métrique la *métrique sphérique standard*.

Exemple 13 ► Forme bilinéaire de Lorentz

Définissons la forme bilinéaire de Lorentz sur \mathbb{R}^{d+1} . C'est la forme bilinéaire canonique de signature $(1, d)$, c'est-à-dire pour tous $x, y \in \mathbb{R}^{d+1}$:

$$\langle x, y \rangle_{\mathcal{L}} = -x_0y_0 + \sum_{i=1}^d x_iy_i \quad (1.3)$$

On note $\|\cdot\|_{\mathcal{L}}$ la forme quadratique associée.

1.2.5 Exemples des métriques de variétés riemanniennes

Exemple 14 ► Espace hyperbolique

En utilisant la forme de Lorentz, rappelons que $H^d \subset \mathbb{R}^{d+1}$ est l'ensemble des points vérifiant $\|x\|_{\mathcal{L}} = -1$. Considérons maintenant le sous-ensemble ouvert H_+^d de H^d :

$$H_+^d = \left\{ x \in \mathbb{R}^{d+1} \mid x_0 > 0, \|x\|_{\mathcal{L}} = -1 \right\}. \quad (1.4)$$

Pour tout $x \in H_+^d$ et deux vecteurs tangents $v, w \in T_xM$, d'après l'Exemple 7, cela signifie que $\langle v, x \rangle_{\mathcal{L}} = \langle w, x \rangle_{\mathcal{L}} = 0$. Comme $\langle \cdot, \cdot \rangle_{\mathcal{L}}$ est définie négative sur $\mathbb{R}x$ et de signature $(1, d)$, elle est définie positive sur l'orthogonal de $\mathbb{R}x$ pour la métrique de Lorentz, c'est-à-dire sur T_xM . On en conclut que $(H_+^d, \langle \cdot, \cdot \rangle_{\mathcal{L}})$ est une variété riemannienne.

Exemple 15 ► Métrique de Frobenius

L'analogie de la métrique euclidienne sur les espaces matriciels est le produit scalaire *de Frobenius* défini par :

$$\forall A, B \in M_{m,n}(\mathbb{R}), \quad \langle A, B \rangle_F = \text{tr}(A^\top B) = \sum_{i,j} A_{ij} B_{ij}. \quad (1.5)$$

où tr désigne l'opérateur trace.

Munis de cette métrique (ou de sa restriction), les variétés $M_n(\mathbb{R})$, $GL(n)$, $SO(n)$ et $SE(n)$ sont des variétés riemanniennes.

1.3 Connexions affines et connexion de Levi-Civita

Une connexion est une structure additionnelle qui peut être définie indépendamment d'une métrique riemannienne. Elle permet de comparer les espaces tangents d'un point à un autre en définissant la notion de parallélisme (transport parallèle).

1.3.1 Connexions

Définition 14 ► Connexion

Soit M une variété différentiable. Une *connexion* sur M est une application \mathbb{R} -bilineaire $\nabla : \Gamma(TM) \times \Gamma(TM) \rightarrow \Gamma(TM)$ vérifiant pour tous $X, Y \in \Gamma(TM)$ et toute $f \in C^\infty(M)$:

1. (Linéarité) : $\nabla_{fX} Y = f \nabla_X Y$
2. (Règle de Leibniz) : $\nabla_X (fY) = X(f)Y + f \nabla_X Y$

Le champ de vecteurs $\nabla_X Y$ est appelé *dérivée covariante* de Y par rapport à X .

1.3.2 Propriétés locales

En coordonnées locales (x^1, \dots, x^d) , les *symboles de Christoffel* caractérisent la connexion. Dans le repère local $(\partial_i)_i$, on définit $(\Gamma_{ij}^k)_{ijk}$ par :

$$\nabla_{\partial_i}(\partial_j) = \Gamma_{ij}^k \partial_k,$$

où nous utilisons la convention de sommation d'Einstein (somme implicite sur l'indice répété k).

Pour deux champs $X = X^i \partial_i$ et $Y = Y^i \partial_i$, la dérivée covariante s'exprime par :

$$\nabla_X Y = X^i \left(\frac{\partial Y^j}{\partial x^i} \partial_j + Y^j \Gamma_{ij}^k \partial_k \right). \quad (1.6)$$

Définition 15 ► Transport parallèle et géodésiques

Considérons un champ de vecteurs X le long d'une courbe $\gamma : [a, b] \rightarrow M$. Il existe une *dérivée covariante* le long de γ coïncidant avec $(\nabla_{\dot{\gamma}(t)} X)_{\gamma(t)}$. Nous pouvons alors définir la notion centrale de *transport parallèle* : un champ X le long de γ est dit parallèle si $\nabla_{\dot{\gamma}} X = 0$.

Définition 16 ► Champ de vecteurs parallèle

Soit M une variété différentiable munie d'une connexion ∇ . Pour toute courbe $\gamma : [a, b] \rightarrow M$, un champ de vecteurs X le long de γ est dit *parallèle* si :

$$\nabla_{\dot{\gamma}(t)} X(t) = 0. \quad (1.7)$$

En coordonnées locales, avec les symboles de Christoffel, cette condition s'écrit :

$$\dot{X}^k(t) + \Gamma_{ij}^k X^i(t) \dot{\gamma}^j(t) = 0. \quad (1.8)$$

Proposition 3 ► Existence et unicité

Pour toute courbe C^1 $\gamma : [a, b] \rightarrow M$ et tout vecteur $v \in T_{\gamma(t)}M$, il existe un unique champ parallèle X le long de γ tel que $X(t) = v$.

Le vecteur $X(s)$ est appelé *transport parallèle* de v le long de γ de t à s , noté $X(s) = \Pi_{\gamma, t}^s v$. L'application $\Pi_{\gamma, t}^s$ est un isomorphisme linéaire entre $T_{\gamma(t)}M$ et $T_{\gamma(s)}M$.

Définition 17 ► Géodésique

Une courbe $\gamma : [a, b] \rightarrow M$ est *autoparallèle* (ou *géodésique*) si :

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = 0. \quad (1.9)$$

En coordonnées locales, l'équation géodésique devient :

$$\ddot{\gamma}^k(t) + \Gamma_{ij}^k \dot{\gamma}^i(t) \dot{\gamma}^j(t) = 0. \quad (1.10)$$

Propriétés des géodésiques : Pour tout $(x, v) \in TM$, il existe un intervalle

maximal $I_{x,v} \subseteq \mathbb{R}$ tel que $\gamma_{x,v} : I_{x,v} \rightarrow M$ soit l'unique géodésique vérifiant :

$$\gamma_{x,v}(0) = x \quad \text{et} \quad \dot{\gamma}_{x,v}(0) = v.$$

De plus, par homogénéité de (1.10), pour tout $s > 0$:

$$I_{x,sv} = \frac{1}{s}I_{x,v} \quad \text{et} \quad \gamma_{x,sv}(t) = \gamma_{x,v}(st).$$

L'ensemble des vecteurs $v \in T_x M$ tels que $1 \in I_{x,v}$ est ouvert, non vide et contient 0.

Définition 18 ► Application exponentielle

Soit ∇ une connexion sur une variété différentiable M . L'application $(x, v) \mapsto \gamma_{x,v}(1)$ définie sur l'ouvert $\{(x, v) \in TM, 1 \in I_{x,v}\}$ à valeurs dans M est appelée *application exponentielle* de ∇ . On dit que M est *géodésiquement complète* si cette application est définie sur tout TM .

Pour tout $x \in M$, on note $\text{Exp}_x : v \in T_x M \mapsto \gamma_{x,v}(1)$ l'application exponentielle en x .

Proposition 4

Soit ∇ une connexion sur M et $x \in M$:

1. La différentielle de Exp_x en 0 est l'identité
2. $(x, v) \mapsto (x, \text{Exp}_x(v))$ est un difféomorphisme lisse d'un voisinage ouvert de la section nulle vers un voisinage ouvert de la diagonale de $M \times M$

Remarque 3

Bien que nous ayons introduit de nombreux concepts, les exemples manquent car les connexions et leurs symboles de Christoffel sont rarement explicites, sauf lorsqu'elles sont compatibles avec une métrique.

1.3.3 La connexion de Levi-Civita

Théorème 3 ► Théorème fondamental de la géométrie riemannienne

Soit (M, g) une variété riemannienne. Il existe une unique connexion vérifiant pour tous $X, Y, Z \in \Gamma(TM)$:

$$1. \text{ (Sans torsion)} \nabla_X Y - \nabla_Y X = [X, Y] \quad (1.11)$$

$$2. \text{ (Compatibilité)} X(\langle Y, Z \rangle) = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle \quad (1.12)$$

Cette connexion, appelée *connexion de Levi-Civita*, est déterminée par la formule de Koszul :

$$\begin{aligned} 2\langle \nabla_X Y, Z \rangle &= X(\langle Y, Z \rangle) + Y(\langle X, Z \rangle) - Z(\langle X, Y \rangle) \\ &\quad - \langle Y, [X, Z] \rangle - \langle X, [Y, Z] \rangle + \langle Z, [Y, X] \rangle. \end{aligned} \quad (1.13)$$

1.3.4 Interprétation

La condition (1.11) assure l'unicité (condition de torsion nulle). La condition (1.12) généralise la règle de Leibniz au produit scalaire $\langle Y, Z \rangle$, signifiant que la métrique est parallèle pour la connexion.

Extension aux tenseurs : Pour un tenseur d'ordre $(2, 0)$ comme g , la dérivée covariante s'étend via la règle de Leibniz généralisée :

$$(\nabla_X g)(Y, Z) = X(\langle Y, Z \rangle) - \langle \nabla_X Y, Z \rangle - \langle Y, \nabla_X Z \rangle \quad (1.14)$$

La combinaison avec (1.12) donne $\nabla g = 0$.

Symboles de Christoffel : En coordonnées locales (x^1, \dots, x^n) , la condition de torsion nulle et le théorème de Schwarz impliquent :

$$\Gamma_{ij}^k = \Gamma_{ji}^k$$

La formule de Koszul devient :

$$2g_{kl}\Gamma_{ij}^k = \partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij}$$

D'où l'expression explicite :

$$\Gamma_{ij}^k = \frac{1}{2}g^{lk}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij}) \quad (1.15)$$

Proposition 5 ► Sous-variétés

Soient M une variété riemannienne et $N \subset M$ une sous-variété munie de la métrique induite. Les connexions de Levi-Civita vérifient :

$$\nabla_X^N Y = (\nabla_X^M Y)^\parallel \quad (1.16)$$

où \parallel désigne la projection orthogonale sur $T_p N$.

Application pratique : Pour les variétés plongées dans \mathbb{R}^N , la connexion coïncide avec la dérivée directionnelle usuelle. Le calcul se réduit à une projection sur les espaces tangents.

Conservation de la métrique : La compatibilité implique que le transport parallèle préserve la métrique.

Proposition 6 ► Isométrie du transport parallèle

Soit $\gamma : I \rightarrow M$ une courbe lisse et $s, t \in I$. Le transport parallèle $\Pi_{\gamma,s}^t : T_{\gamma(s)}M \rightarrow T_{\gamma(t)}M$ pour la connexion de Levi-Civita est une isométrie :

$$\forall v, w \in T_{\gamma(s)}M, \quad \langle \Pi_{\gamma,s}^t v, \Pi_{\gamma,s}^t w \rangle_{\gamma(t)} = \langle v, w \rangle_{\gamma(s)}.$$

Proposition 7 ► Invariance par isométrie

Soit (M, g) une variété riemannienne avec connexion de Levi-Civita ∇ . Pour toute isométrie $f \in \text{Isom}(M)$ et tout $X, Y \in \Gamma(TM)$, on a :

$$df \nabla_X Y = \nabla_{df X} df Y \quad (1.17)$$

$$\text{Exp}_{f(x)} \circ df_x = f \circ \text{Exp}_x \quad (1.18)$$

Définition 19 ► Applications Exp et Log

1. L'*exponentielle riemannienne* est l'application exponentielle de la connexion de Levi-Civita.
2. Le *rayon d'injectivité* $\text{inj}_M(x)$ est le plus grand $\epsilon > 0$ tel que Exp soit un difféomorphisme sur la boule ouverte de rayon ϵ dans $T_x M$.
3. Le *logarithme riemannien* est l'inverse de Exp sur le domaine d'injectivité $\text{Inj}(x) \subseteq T_x M$.

Propriétés des géodésiques :

1. L'énergie $E(t) = \frac{1}{2} \|\dot{\gamma}(t)\|^2$ est constante le long des géodésiques.
2. Les géodésiques sont paramétrées à vitesse constante.

3. Pour une sous-variété $N \subset M$, les géodésiques de N sont les courbes de M dont l'accélération est normale à N .

1.3.5 Exemples

Exemple 16 ► Hypersphère

Pour tout $x \in S^d \subset \mathbb{R}^{d+1}$ et $v \in T_x S^d$, la courbe géodésique est donnée par :

$$\gamma : t \mapsto \cos(t\|v\|)x + \sin(t\|v\|)\frac{v}{\|v\|}. \quad (1.19)$$

L'accélération $\ddot{\gamma}(t) = -\|v\|^2\gamma(t)$ est normale à S^d , confirmant qu'il s'agit bien de la géodésique.

Applications Exponentielle et Logarithmique

1. L'application Exp en x :

$$\text{Exp}_x(v) = \cos(\|v\|)x + \sin(\|v\|)\frac{v}{\|v\|}$$

est un difféomorphisme pour $\|v\| < \pi$.

2. Le logarithme riemannien pour $y \neq \pm x$:

$$\text{Log}_x(y) = \arccos(\langle y, x \rangle) \frac{y - \langle y, x \rangle x}{\|y - \langle y, x \rangle x\|} \quad (1.20)$$

Longueur et énergie des courbes Pour une courbe C^1 $\gamma : [a, b] \rightarrow M$:

$$L(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt, \quad E(\gamma) = \int_a^b \|\dot{\gamma}(t)\|^2 dt \quad (1.21)$$

1. La longueur est invariante par reparamétrisation
2. L'énergie dépend de la paramétrisation

Définition 20 ► Distance riemannienne

Pour (M, g) une variété riemannienne connexe :

$$d(x, y) = \inf\{L(\gamma) \mid \gamma \text{ joignant } x \text{ à } y\} \quad (1.22)$$

1. Cette distance induit la topologie de M
2. Une courbe est dite *minimisante* si $L(\gamma) = d(x, y)$

Proposition 8 ► Caractérisation des géodésiques

Pour une variété riemannienne (M, g) et une courbe $\gamma : [a, b] \rightarrow M$ de classe C^1 , les assertions suivantes sont équivalentes :

1. γ est une géodésique
2. γ est paramétrée à vitesse constante et localement minimisante
3. γ minimise localement l'énergie

Distance riemannienne : Pour $y \in \text{Exp}_x(\text{Inj}(x))$, on a :

$$d(x, y) = \|\text{Log}_x(y)\|$$

Définition 21

Un espace métrique est *complet* si toute suite de Cauchy converge.

Exemple 17

Le demi-plan $P = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$:

1. Avec la métrique euclidienne : non géodésiquement complet
2. Avec $g(x, y) = \frac{dx^2 + dy^2}{y^2}$: géodésiquement complet

Théorème 4 ► Hopf-Rinow [39]

Si (M, g) est géodésiquement complet, alors :

1. Deux points quelconques sont joints par une géodésique minimisante
2. Les notions de complétude métrique et géodésique coïncident

Pour une variété riemannienne connexe (M, g) de distance d , les assertions suivantes sont équivalentes :

1. La variété (M, g) est géodésiquement complète
2. Les parties fermées bornées de (M, d) sont compactes
3. L'espace métrique (M, d) est complet

1.4 Courbure

Dans un espace euclidien, un champ constant est parallèle le long de toute courbe. Dans une variété riemannienne en général, le transport parallèle dépend de la courbe suivie, et il peut ne pas exister de champs parallèles le long de toutes les courbes, même localement. On peut étudier l'effet du transport parallèle le long de petites

courbes fermées. Considérons un point $x \in M$ et les courbes fermées dont les vitesses tangentes en ce point engendrent un sous-espace de dimension deux.

1.4.1 Tenseur Courbure

Lemme 1

Soit $p \in \mathbb{N}$ et $A : \Gamma(TM)^p \rightarrow \Gamma(TM)$ une application $C^\infty(M)$ -multilinéaire, c'est-à-dire :

$$\forall f_1, \dots, f_p \in C^\infty(M), \forall X_1, \dots, X_p \in \Gamma(TM), \quad A(f_1 X_1, \dots, f_p X_p) = f_1 \dots f_p A(X_1, \dots, X_p).$$

Alors, pour tout $x \in M$, la valeur de $A(X_1, \dots, X_p)$ en x ne dépend que des valeurs des X_i en x .

Des calculs simples utilisant la règle de Leibniz à partir de la définition d'une connexion (2) montrent que la définition suivante est une application C^∞ -multilinéaire.

Définition 22 ► Tenseur de courbure

Soit (M, ∇) une variété équipée d'une connexion.

Le tenseur de courbure est défini comme une application de $\Gamma(TM)^3$ vers $\Gamma(TM)$ par :

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z \quad (1.23)$$

Le tenseur de courbure d'une variété riemannienne est la courbure de sa connexion de Levi-Civita.

Dans la définition ci-dessus, les arguments X, Y, Z sont des champs de vecteurs, mais le lemme 1 permet d'écrire, pour tout $x \in M$, R_x comme une application définie sur des vecteurs tangents $u, v, w \in T_x M$.

Exemple 18 ► Espace euclidien

Dans un espace euclidien \mathbb{R}^d muni du produit scalaire canonique, la connexion coïncide avec la dérivée directionnelle, qui commute. Par conséquent, $R = 0$.

Les propriétés suivantes aident à appréhender la complexité de ce tenseur.

Proposition 9 ► propriétés des courbures [39]

Soit (M, g) une variété riemannienne et R son tenseur de courbure. Les propriétés suivantes sont vérifiées :

1. (**Antisymétrie**) L'application $(X, Y) \in \Gamma(TM)^2 \mapsto R(X, Y)$ est antisymétrique.
2. (**Identité de Bianchi**) Pour tout $X, Y, Z \in \Gamma(TM)$, on a :

$$R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0. \quad (1.24)$$

3. (**Deuxième identité de Bianchi**) Pour tout $X, Y, Z \in \Gamma(TM)$, on a :

$$\nabla_X R(Y, Z) + \nabla_Y R(Z, X) + \nabla_Z R(X, Y) = 0. \quad (1.25)$$

4. Pour toute isométrie $f \in \text{Isom}(M)$ et pour tout $X, Y, Z \in \Gamma(TM)$,

$$R(dfX, dfY)(dfZ) = df(R(X, Y)Z). \quad (1.26)$$

5. Pour tout $X, Y \in \Gamma(TM)$, le champ d'applications linéaires $R(X, Y)$ est antisymétrique par rapport à la métrique g , c'est-à-dire :

$$\forall W, Z \in \Gamma(TM), \quad g(R(X, Y)Z, W) = -g(Z, R(X, Y)W). \quad (1.27)$$

6. Pour tout $X, Y, Z, W \in \Gamma(TM)$, on a :

$$g(R(X, Y)Z, W) = g(R(Z, W)X, Y) = g(R(W, Z)Y, X). \quad (1.28)$$

La compatibilité de la métrique (équations (1.11), (1.12)) est nécessaire pour les deux dernières assertions. Cela permet de définir la *courbure sectionnelle* pour tout $x \in M$ et $u, v \in T_x M$ tels que u et v ne sont pas colinéaires :

$$\kappa_x(u, v) = \frac{\langle R_x(u, v)v, u \rangle}{\|u\|^2\|v\|^2 - \langle u, v \rangle^2} \quad (1.29)$$

1.4.2 Propriétés des Courbures sectionnelles

D'après les propriétés ci-dessus, la valeur de κ_x ne dépend que du plan engendré par (u, v) , c'est-à-dire que pour toute combinaison linéaire non nulle $\alpha u + \beta v$, on a $\kappa(\alpha u + \beta v, v) = \kappa(u, v)$. En réalité, cette quantité suffit à caractériser la courbure de M . En effet, on peut prouver, par des calculs purement algébriques le résultat suivant [39] :

Lemme 2

Soient deux applications quadrilinéaires A, B définies sur un espace vectoriel V vérifiant toutes deux les propriétés suivantes pour tout $u, v, w, z \in V$:

$$A(u, v, w, z) = -A(v, u, w, z) = -A(v, u, z, w)$$

$$A(u, v, w, z) + A(u, w, z, v) + A(u, z, v, w) = 0$$

et coïncidant pour toute paire de variables $u, v \in V$ au sens où $A(u, v, u, v) = B(u, v, u, v)$. Alors $A = B$.

Ce résultat s'applique à l'application $(u, v, w, z) \mapsto g(R(w, z)v, u)$, ce qui permet de démontrer le théorème suivant.

Théorème 5 ► courbure sectionnelle

La courbure sectionnelle détermine le tenseur de courbure.

En particulier, si pour tout $x \in M$, κ prend une valeur constante κ_x sur tous les plans de $T_x M$, alors on peut calculer la courbure à cette constante près. En effet, d'après la définition (22), pour tout $u, v \in T_x M$ on a :

$$\langle R(u, v)v, u \rangle = \kappa_x (\|u\|^2 \|v\|^2 - \langle u, v \rangle^2)$$

Considérons les applications $A : (u, v, w, z) \mapsto \langle R(u, v)w, z \rangle$ et B définie par :

$$B(u, v, w, z) = \kappa_x (\langle u, w \rangle \langle v, z \rangle - \langle u, z \rangle \langle v, w \rangle)$$

Il est clair que les hypothèses du Lemme 2, sont vérifiées pour A et B , donc on conclut :

$$\langle R(u, v)w, z \rangle = \kappa_x (\langle u, w \rangle \langle v, z \rangle - \langle u, z \rangle \langle v, w \rangle) \quad (1.30)$$

Par exemple, l'équation (1.23) est toujours valable en dimension $d = 2$, car il n'y a qu'un seul plan dans chaque espace tangent. Nous énonçons maintenant le résultat surprenant de F. Schur qui donne une condition suffisante pour que la formule (1.23) soit vérifiée.

Théorème 6 ► Schur

Soit (M, g) une variété riemannienne connexe de dimension d . Si $d \geq 3$ et si la courbure sectionnelle κ ne dépend pas du plan mais seulement du point $x \in M$, alors κ est constante (c'est-à-dire qu'elle ne dépend même pas de x).

Exemple 19 ► Hypersphère et hyperboloïde

Rappelons que le groupe des rotations $SO(d+1)$ agit sur l'hypersphère S^d , et supposons que $d \geq 2$. Le stabilisateur de cette action en tout point $x \in S^d$ est isomorphe à $SO(d)$, dont l'action sur $T_x S^d$ est transitive sur les 2-plans, c'est-à-dire que tout 2-plan peut être transformé en n'importe quel autre 2-plan par une rotation. la courbure sectionnelle est préservé

d'un plan à l'autre par les rotations, elle est donc constante sur tout $T_x M$; notons cette valeur κ_x . Si $d \geq 3$, on peut utiliser le Théorème 6 pour conclure qu'elle est constante sur S^d . Bien sûr, elle est également constante pour $d = 2$, car on peut utiliser une isométrie pour transporter le tenseur de courbure d'un point à un autre.

Définition 23 ► Courbure constante

Soit (M, g) une variété riemannienne. On dit que (M, g) a une courbure constante (resp. négative, resp. positive) si sa courbure sectionnelle est constante (resp. négative, resp. positive) en tout point.

En réalité, les espaces à courbure constante (complets simplement connexes) sont tous isométriques à l'un des trois modèles suivants : l'espace euclidien (plat), l'hypersphère (courbure positive) ou l'espace hyperbolique (courbure négative) (voir par exemple Lafontaine, Gallot et al. 2004, Théorème 3.82) [38]. Ces trois exemples décrivent ainsi toute la classe des espaces à courbure constante (à revêtement et isométrie près).

Pour conclure cette section, on explique comment la courbure modifie la correspondance entre les distances dans les espaces tangents et les distances riemanniennes sur la variété. Considérons une variété riemannienne (M, g) et deux vecteurs tangents orthogonaux u, v en un point $x \in M$. Dans l'espace vectoriel $T_x M$, la distance (induite par la métrique en x) entre rv et rw pour un certain $r > 0$ est bien sûr $r\sqrt{2}$. En projetant ces vecteurs sur la variété via l'application exponentielle, la distance riemannienne entre $\text{Exp}_x(rv)$ et $\text{Exp}_x(rw)$ peut augmenter ou diminuer par rapport à la distance dans l'espace tangent. La courbure est l'outil fondamental qui permet de quantifier ces variations, et le signe de la courbure sectionnelle indique si les géodésiques s'accroissent ou s'éloignent.

Théorème 7 ► Distance géodesiques

Soit (M, g) une variété riemannienne, R son tenseur de courbure, κ sa courbure sectionnelle, $x \in M$ et γ_v, γ_w deux géodésiques partant de x avec vitesses initiales $v, w \in T_x M$. Alors pour $r \rightarrow 0$:

$$d(\gamma_v(r), \gamma_w(r))^2 = r^2 \|v - w\|^2 - \frac{r^4}{3} \langle R(v, w)w, v \rangle + O(r^5) \quad (1.31)$$

De plus, si v, w sont orthonormés :

$$d(\gamma_v(r), \gamma_w(r)) = \sqrt{2}r \left(1 - \frac{\kappa(v, w)}{12} r^2 \right) + O(r^4) \quad (1.32)$$

Nous comprenons désormais l'importance du signe de la courbure sectionnelle : si $\kappa > 0$, les géodésiques se rapprochent, tandis que si $\kappa < 0$, elles s'éloignent. En utilisant ce théorème et les formules des géodésiques sur l'hypersphère de l'Exemple 19, on peut calculer un développement de Taylor de $d(\gamma_v, \gamma_w)$ et identifier les coefficients pour trouver que $\kappa = 1$. De même, pour l'espace hyperbolique, $\kappa = -1$. On établit des développements de Taylor similaires en utilisant le tenseur de courbure pour caractériser le défaut de parallélisme lors de la construction de parallélogrammes géodésiques.

Enfin, nous énonçons deux théorèmes importants qui révèlent comment le signe de la courbure influence la géométrie.

Théorème 8 ► de Cartan-Hadamard

Soit (M, g) une variété riemannienne complète connexe à courbure sectionnelle non positive. Alors l'application exponentielle est un revêtement riemannien, c'est-à-dire un revêtement qui est une isométrie locale. Ceci implique que si M est simplement connexe, alors M est difféomorphe à \mathbb{R}^d , et deux points quelconques sont joints par une unique géodésique minimisante.

Remarque 4 ► Dimension infinie

Ce théorème s'étend aux variétés riemanniennes hilbertiennes de dimension infinie, c'est-à-dire dont les atlas sont à valeurs dans un espace de Hilbert.

À l'inverse :

Cas particulier du théorème de Myers 2.3.9: Soit (M, g) une variété riemannienne complète connexe dont la courbure sectionnelle est minorée par $\kappa > \frac{1}{r^2}$ pour un certain $r > 0$. Alors le diamètre de M (i.e. la plus grande distance entre points de M) est majoré par πr et M est compacte.[39].

1.5 Groupes de Lie, algèbres de Lie et sous-groupes de Lie

Nous définissons d'abord les notions de groupes de Lie, sous-groupes et algèbres de Lie, puis l'application exponentielle des groupes de Lie, et enfin nous introduisons les actions lisses de groupes de Lie et les espaces homogènes. Les métriques que l'on peut définir sur les groupes de Lie.

1.5.1 Groupe de Lie

Définition 24 ► Groupe de Lie

Un *groupe de Lie* est un groupe (G, \cdot) tel que G est également une variété différentielle de dimension finie, et où les structures de groupe et différentielle sont compatibles, au sens où la loi de groupe \cdot et l'application inverse $g \mapsto g^{-1}$ sont lisses.

Nous omettons généralement \cdot dans la notation et écrivons simplement la composition du groupe comme une multiplication. Soit e l'élément neutre, ou identité, de G .

Pour tout $g \in G$, nous définissons les applications de *translation à gauche* et à *droite* L_g et R_g par :

$$L_g : h \in G \mapsto gh \in G \quad \text{et} \quad R_g : h \in G \mapsto hg^{-1} \in G. \quad (1.33)$$

Par définition d'un groupe de Lie, L_g et R_g sont des difféomorphismes de G et leur différentielle est un isomorphisme linéaire des espaces tangents. Cela signifie que tous les espaces tangents de G sont identiques à $T_e G$, et qu'il existe une manière canonique de transporter les vecteurs de $T_g G$ vers $T_e G$ via $L_{g^{-1}}$ pour tout $g \in G$.

Ce fait, qui pourrait être reformulé comme "le fibré tangent d'un groupe de Lie est trivial", implique que TG est difféomorphe au produit $G \times T_e G$. Cette propriété est fondamentale pour caractériser la géométrie d'un groupe de Lie et très pratique pour l'implémentation. De plus, les applications $g \mapsto L_g$ et $g \mapsto R_g$ sont des homomorphismes de groupes entre (G, \cdot) et $(\text{Diff}(G), \circ)$, le groupe des difféomorphismes de G avec la composition d'applications comme loi de groupe.

Définition 25 ► Champ de vecteur invariant

Un champ de vecteurs $X \in \Gamma(TG)$ est dit *invariant à gauche* si $\forall g, h \in G$:

$$X_{gh} = dL_g X_h.$$

Soit $\mathcal{L}(G)$ l'ensemble de tous les champs de vecteurs invariants à gauche.

Exemple 20 ► Groupe linéaire général

Le groupe linéaire général réel $GL(n)$ joue un rôle fondamental . Il est défini comme l'ensemble des matrices inversibles de taille n :

$$GL(n) = \{A \in M_n(\mathbb{R}) \mid \det(A) \neq 0\},$$

et comme un ouvert de l'espace vectoriel $M_n(\mathbb{R}) \simeq \mathbb{R}^{n^2}$, c'est une variété différentielle.

La loi de groupe est la multiplication matricielle, qui peut s'écrire comme un polynôme des coefficients des matrices et est donc lisse. Grâce à la formule de Cramer $A^{-1} = \det(A)^{-1} \text{Com}(A)^T$, où Com désigne la matrice formée par des cofacteurs, l'application inversion est également lisse. Par conséquent, $GL(n)$ est un groupe de Lie.

Définition 26 ► Algèbre de Lie

Une *algèbre de Lie* réelle est un espace vectoriel réel \mathfrak{g} muni d'une application bilinéaire $[\cdot, \cdot]$ vérifiant :

1. (**Antisymétrie**) $\forall x, y \in \mathfrak{g}, \quad [x, y] = -[y, x],$
2. (**Identité de Jacobi**) $\forall x, y, z \in \mathfrak{g}, \quad [x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0.$

Exemple 21 ► Champs de vecteurs

L'espace des champs de vecteurs $\Gamma(TM)$ d'une variété M , muni du crochet de Lie des champs de vecteurs forme une algèbre de Lie de dimension infinie.

Exemple 22 ► Algèbre matricielle

L'algèbre des matrices carrées $M_n(\mathbb{R})$ munie du commutateur :

$$\forall A, B \in M_n(\mathbb{R}), \quad [A, B] = AB - BA \tag{1.34}$$

est une algèbre de Lie.

1.5.2 Propriétés des champs de vecteurs invariants à gauche

Théorème 9 ► champs de vecteurs invariants

Soit G un groupe de Lie de dimension finie d .

1. L'espace des champs de vecteurs invariants à gauche $\mathcal{L}(G)$ d'un groupe de Lie G est une sous-algèbre de $\Gamma(TG)$. Cela signifie que le crochet de deux champs de vecteurs invariants à gauche est également invariant à gauche.
2. En tant qu'espace vectoriel, $\mathcal{L}(G)$ est isomorphe à T_eG , l'espace tangent à l'identité de G . Ceci implique que $\mathcal{L}(G)$ est de dimension finie.

En effet, l'application $X \in \mathcal{L}(G) \mapsto X_e \in T_eG$ est linéaire et admet pour inverse $x \mapsto \tilde{x}$ où \tilde{x} est le champ de vecteurs invariant à gauche défini par :

$$\tilde{x} : g \in G \longmapsto dL_g x \in T_gG.$$

On peut ainsi définir le crochet sur $T_eG : [x, y] \triangleq [\tilde{x}, \tilde{y}]_e$, ce qui fait de T_eG une algèbre de Lie isomorphe (en tant qu'algèbres de Lie) à $\mathcal{L}(G)$. On appelle T_eG l'*algèbre de Lie* de G et on la note $\mathfrak{g} = T_eG$.

Heureusement, une représentation linéaire de G permet de calculer le crochet de Lie sans manipuler directement les champs de vecteurs. En effet, définissons pour tout $g \in G$ l'application de conjugaison $C_g : h \mapsto ghg^{-1}$, et sa différentielle à l'identité, appelée représentation adjointe de G :

$$\text{Ad}_g : \mathfrak{g} \rightarrow \mathfrak{g}.$$

La différentielle de $g \mapsto \text{Ad}_g$, notée ad , est appelée représentation adjointe de \mathfrak{g} , et on peut montrer qu'elle coïncide avec le crochet de Lie de \mathfrak{g} :

$$\text{ad}_x(y) = [x, y].$$

Exemple 23 ► Algèbre de Lie linéaire générale

$GL(n)$ est un groupe de Lie dont la structure différentiable provient de son plongement dans $M_n(\mathbb{R})$. Ainsi, d'après ce qui précède, son espace tangent à l'identité est l'espace matriciel tout entier, et en tant qu'algèbre de Lie de $GL(n)$ il est noté $\mathfrak{gl}(n)$. Calculons explicitement Ad et ad et vérifions que le crochet de $\mathfrak{gl}(n)$ coïncide avec le commutateur défini dans l'Exemple 22.

L'application de conjugaison est la restriction à $GL(n)$ de l'application linéaire $h \mapsto ghg^{-1}$ de $M_n(\mathbb{R})$ pour tout $g \in GL(n)$. Sa différentielle est donc pour tout $X \in M_n(\mathbb{R})$:

$$\text{Ad}_g(X) = gXg^{-1}. \quad (1.35)$$

Considérons maintenant une courbe $c : (-\epsilon, \epsilon) \rightarrow GL(n)$ telle que $c(0) = I_n$ et $c'(0) = X \in M_n(\mathbb{R})$. Pour tout $Y \in M_n(\mathbb{R})$, on a :

$$\text{ad}_X Y = \left. \frac{d}{dt} \right|_{t=0} (c(t)Yc(t)^{-1}) = XY - YX \quad (1.36)$$

En résumé :

1. L'algèbre de Lie d'un groupe de Lie est son espace tangent à l'identité $T_e G$ ou, de manière équivalente, l'algèbre des champs de vecteurs invariants à gauche.
2. Le crochet de Lie sur $T_e G$ est défini via la représentation adjointe et coïncide avec celui des champs de vecteurs invariants à gauche.
3. Dans le cas des algèbres de Lie linéaires (c'est-à-dire des sous-algèbres de $M_n(\mathbb{R})$), le crochet de Lie coïncide avec le commutateur matriciel.

Remarque 5 ► champs de vecteurs invariants à droite

Notons que les champs de vecteurs invariants à droite peuvent être définis de manière analogue aux champs invariants à gauche, et leur ensemble forme une algèbre de Lie. Cependant, le crochet de Lie des champs invariants à droite coïncide avec l'opposé de la représentation adjointe. Pour des raisons pratiques et historiques, les champs invariants à droite sont utilisés en dimension infinie tandis que les invariants à gauche sont utilisés en dimension finie.

Comme tous les exemples suivants seront des sous-groupes du groupe linéaire général, nous énonçons un résultat de von Neumann et Cartan qui donne une condition nécessaire et suffisante pour qu'un sous-groupe d'un groupe de Lie G soit un sous-groupe de Lie plongé, c'est-à-dire un groupe de Lie dont la structure différentielle est compatible avec celle de G , et tel que l'inclusion soit une application lisse [37].

Théorème 10 ► de Cartan-von Neumann

Soit G un groupe de Lie. Tout sous-groupe fermé $H \subset G$ est un sous-groupe de Lie de G . Réciproquement, tout sous-groupe de Lie de G est fermé.

On peut également montrer que l'algèbre de Lie d'un sous-groupe H de G est une sous-algèbre de Lie de \mathfrak{g} .

Exemple 24 ► Groupe spécial orthogonal

Rappelons que le groupe spécial orthogonal est l'ensemble des matrices orthogonales de déterminant positif. Il est clair qu'il est stable par multiplication, c'est donc un sous-groupe de $GL(n)$.

De plus, $SO(n) = f^{-1}(I_n) \cap \det^{-1}(1)$, où $f : A \mapsto A^\top A$ et \det sont des applications continues. $SO(n)$ est donc fermé dans $M_n(\mathbb{R})$, et est par conséquent un sous-groupe de Lie de $GL(n)$.

Son algèbre de Lie est :

$$\mathfrak{so}(n) \triangleq T_{I_n} SO(n) = \{A \in M_n(\mathbb{R}) \mid A^\top + A = 0\} = \text{Skew}(n). \quad (1.37)$$

Exemple 25 ► Groupe spécial euclidien

Nous introduisons maintenant un groupe présent dans les applications et qui sera utilisé . Ce groupe est défini comme l'ensemble des isométries directes - ou transformations rigides - de \mathbb{R}^n , c'est-à-dire les transformations linéaires de l'espace affine \mathbb{R}^n qui préservent son produit scalaire canonique. Une telle transformation ρ peut se décomposer en une partie rotation et une partie translation : $\rho(x) = Rx + u$, où $R \in SO(n)$ et $x, u \in \mathbb{R}^n$. Définissons :

$$SE(n) = \{(R, u) \mid R \in SO(n), u \in \mathbb{R}^n\}. \quad (1.38)$$

La composition de deux isométries $\rho = (R, u)$, $\rho' = (R', u')$ reste une isométrie :

$$\rho \circ \rho'(x) = RR'x + Ru' + u,$$

et peut s'écrire $\rho \circ \rho' = (RR', Ru' + u)$. Ceci suggère la représentation de $SE(n)$ en coordonnées *homogènes* :

$$\rho = \begin{pmatrix} R & u \\ 0 & 1 \end{pmatrix} \in GL(n+1) \quad (1.39)$$

La composition des isométries correspond alors à la multiplication matricielle, et $SE(n)$ devient ainsi un sous-groupe de Lie de $GL(n+1)$. Son algèbre de Lie est :

$$\mathfrak{se}(n) = \left\{ \begin{pmatrix} S & v \\ 0 & 0 \end{pmatrix} \mid S \in \text{Skew}(n), v \in \mathbb{R}^n \right\}. \quad (1.40)$$

Ainsi, sa dimension est $\frac{n(n+1)}{2}$. La submersion utilisée est l'application :

$$\begin{pmatrix} R & u \\ v^\top & c \end{pmatrix} \in GL_+(n+1) \mapsto (R^\top R, v, c) \in M_n(\mathbb{R}) \times \mathbb{R}^n \times \mathbb{R}, \quad (1.41)$$

de sorte que $SE(n) = f^{-1}\{(I_n, 0, 1)\}$.

1.5.3 L'application exponentielle

Nous étudions maintenant le flot des champs de vecteurs invariants à gauche d'un groupe de Lie. Cela permet d'établir une correspondance canonique entre l'algèbre de Lie et le groupe de Lie. Tout d'abord, l'invariance à gauche se traduit par une propriété de commutation du flot avec l'application de translation à gauche :

Proposition 10

Soit G un groupe de Lie et $X \in \mathcal{L}(G)$. Alors X est complet, et si $(\phi_t)_t$ est son flot, alors pour tout $t \in \mathbb{R}$ et $g, g' \in G$ on a :

$$\phi_t(gg') = g\phi_t(g'), \quad \text{c'est-à-dire} \quad \phi_t \circ L_g = L_g \circ \phi_t. \quad (1.42)$$

Définition 27 ► Application exponentielle

On appelle *application exponentielle*, et on note $\exp : \mathfrak{g} \rightarrow G$, l'application définie par $x \mapsto \phi_1(e)$ où $(\phi_t)_t$ est le flot du champ de vecteurs invariant à gauche \tilde{x} .

Remarque 6 ► types de l'exp

Il s'agit de la seconde définition d'application exponentielle que nous rencontrons, la première étant l'exponentielle riemannienne définie par le flot géodésique d'une métrique. Nous désignerons celle définie canoniquement sur un groupe de Lie comme l'*exponentielle de groupe* et utiliserons la notation minuscule \exp .

Exemple 26 ► Cas linéaire général

Le cas fondamental est bien sûr à nouveau celui du groupe linéaire général. Rappelons que son algèbre de Lie $\mathfrak{gl}(n) = M_n(\mathbb{R})$ est l'ensemble des matrices carrées, et la loi de groupe est la multiplication matricielle (linéaire). Par conséquent, un champ de vecteurs invariant à gauche \tilde{X} associé à $X \in \mathfrak{gl}(n)$ est défini par :

$$g \mapsto gX$$

Soit γ la courbe intégrale partant de l'identité I_n . Cela signifie que γ est solution de l'équation différentielle ordinaire définie sur \mathbb{R} avec condition initiale $\gamma(0) = I_n$:

$$\gamma'(t) = \gamma(t)X$$

On sait bien que l'unique solution est l'exponentielle matricielle, définie par la série :

$$e^{tX} = \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k.$$

Ainsi, $\exp(X) = e^X$.

Nous énonçons maintenant quelques propriétés fondamentales de l'application exponentielle (par exemple Gallier et Quaintance 2020 [37], Proposition 18.6-18.7-18.13).

Proposition 11 ► Propriétés de l'application exponentielle

Pour tout groupe de Lie G , l'application exponentielle $\exp : \mathfrak{g} \rightarrow G$ est lisse et un difféomorphisme local en 0.

L'inverse de l'exponentielle de groupe, définie localement, est appelée application *logarithme* et est à valeurs dans $\mathfrak{g} = T_e G$. Elle permet ainsi de transformer des données définies sur G en des données dans son algèbre de Lie \mathfrak{g} , qui est un espace vectoriel.

Remarque 7 ► Dimension infinie

La Proposition 11 n'est pas vraie en dimension infinie. Par exemple, si $G = \text{Diff}(M)$, l'application \exp n'est même pas surjective dans aucun voisinage de l'identité [12].

Proposition 12

Soient G, H des groupes de Lie, et $f : G \rightarrow H$ un homomorphisme de groupes de Lie. Alors :

$$f \circ \exp = \exp \circ df_e \quad (1.43)$$

En particulier, si $G = H$ et $f = C_g$, nous avons :

$$\exp(t \text{AD}_g(u)) = g \exp(tu) g^{-1} = C_g(\exp(tu)) \quad (1.44)$$

En utilisant $f = L_g$, ou de manière équivalente $f = R_g$, nous définissons l'application exponentielle en tout point $g \in G$ de sorte qu'elle vérifie les propriétés ci-dessus. Pour tout $v \in T_g G$:

$$\exp_g(v) = L_g \circ \exp \circ (dL_g)^{-1} = R_{g^{-1}} \circ \exp \circ dR_g.$$

où la différentiation est effectuée en e , et ainsi l'application exponentielle du groupe est maintenant comprise comme \exp_e .

De plus, nous pouvons reconstruire toute la composante connexe de l'identité à partir de l'algèbre de Lie :

Théorème 11

Si G est un groupe de Lie et G_0 la composante connexe de e , alors G_0 est engendrée par $\exp(\mathfrak{g})$.

En particulier (Gallier et Quaintance 2020, Théorème 1.6 et 1.12 .[37]), nous avons le résultat suivant :

Pour $SO(n)$ et $SE(n)$:

- L'application exponentielle $\exp : \mathfrak{so}(n) \rightarrow SO(n)$ est surjective,
- L'application exponentielle $\exp : \mathfrak{se}(n) \rightarrow SE(n)$ est surjective.

Notons qu'en appliquant la propriété de commutation à l'application d'inclusion, les applications exponentielles de $SE(n)$ et $SO(n)$ coïncident avec la restriction de l'application exponentielle de $GL(n)$, c'est-à-dire l'exponentielle matricielle.

Enfin, rappelons que les flots des champs de vecteurs complets sont des sous-groupes à un paramètre de $\text{Diff}(M)$. Définissons maintenant un sous-groupe à un paramètre de G comme un homomorphisme de groupes de Lie $t \in \mathbb{R} \mapsto \gamma_t \in G$, noté $(\gamma_t)_{t \in \mathbb{R}}$. Alors, par un changement de variable dans l'équation du flot, il est clair que $t \mapsto \exp(tX)$ est un sous-groupe à un paramètre pour tout $X \in \mathfrak{g}$. En fait, tous les sous-groupes à un paramètre sont de cette forme et les applications :

$$\theta : (\gamma_t)_{t \in \mathbb{R}} \mapsto \left. \frac{d}{dt} \right|_{t=0} \gamma_t; \quad \theta' : X \mapsto (\exp(tX))_{t \in \mathbb{R}}$$

sont inverses l'une de l'autre, entre \mathfrak{g} et l'ensemble des sous-groupes à un paramètre.

1.6 Action de groupe et espaces homogènes

Nous définissons la notion d'action de groupe, qui sera fondamentale pour modéliser les groupes comme générateurs de transformations sur une variété .

1.6.1 Action de groupe

Définition 28 ► Action de groupe

Étant donné un ensemble M et un groupe G , une *action à gauche* de G sur M est une fonction $\triangleright : G \times M \rightarrow M$, telle que :

1. $\forall g, h \in G, \forall x \in M, \quad g \triangleright (h \triangleright x) = (gh) \triangleright x,$
2. $\forall x \in M, \quad e \triangleright x = x.$

Si de plus M est une variété différentielle et G est un groupe de Lie, on dit que cette action est lisse si l'application \triangleright est lisse de la variété produit $G \times M$ vers M .

Dans ce cas, pour tout $g \in G$, $x \mapsto g \triangleright x$ est un difféomorphisme de M , dont l'inverse est $x \mapsto g^{-1} \triangleright x$. On appelle cette application *translation à gauche* par g et on la note L_g par analogie avec la loi du groupe.

Remarque 8

Notons que l'application $g \mapsto L_g$ est un homomorphisme de groupes entre G et $\text{Diff}(M)$. En fait, si cette application est injective, on dit que l'action est fidèle et on peut voir G comme un sous-groupe immergé de $\text{Diff}(M)$.

On dit qu'une action est

1. *libre* si pour tout $g \in G$ et $x \in M$, si $g \triangleright x = x$, alors $g = e$;
2. *fidèle* si pour tout $g \in G$, si $\forall x \in M, g \triangleright x = x$ alors $g = e$;
3. *transitive* si pour tout $x, y \in M$, il existe $g \in G$ tel que $y = g \triangleright x$;
4. *propre* si pour tout ensemble compact $K \subset M$, l'ensemble $\{g \in G \mid K \cap gK \neq \emptyset\}$ est compact. C'est toujours le cas si G est compact.

Notons qu'une action libre signifie que $G_x = e$ pour tout $x \in M$. En revanche, si une action est fidèle, alors l'application $x \mapsto g \triangleright x$ est différente de l'identité de M pour tout $g \in G$. Ces notions jouent un rôle essentiel dans la détermination de la géométrie de M .

Pour tout $x \in M$, on définit l'*orbite* de x comme $[x] = G \triangleright x = \{g \triangleright x \mid g \in G\}$. les orbites sont les classes d'équivalence de la relation suivante :

$$x \sim y \iff \exists g \in G, y = g \triangleright x \iff y \in g \triangleright x.$$

Définition 29 ► Espace quotient

L'ensemble des orbites est noté M/G et est appelé *quotient* de M par G . Cet ensemble est obtenu en identifiant tous les points d'une même orbite. On définit la *projection canonique*

$$\pi : \begin{cases} M & \longrightarrow M/G \\ x & \longmapsto [x] \end{cases}$$

1.6.2 Variété quotient

Nous avons les conditions suffisantes suivantes sur l'action pour garantir qu'un espace quotient est bien une variété différentiable (Lafontaine, Gallot et al. 2004, Théorème 1.95).

Théorème 12 ► Variété quotient

Soit M une variété différentiable, G un groupe de Lie avec une action lisse à gauche sur M qui est libre et propre. Alors la projection canonique (29) est une submersion et il existe une unique structure différentielle sur M/G avec cette propriété.

Ce cas se présente par exemple dans la variété de Stiefel .

Nous nous concentrons maintenant sur un cas particulier de l'action d'un sous-groupe de G sur G .

En tout point $x \in M$, on peut définir le sous-groupe *d'isotropie*, ou *stabilisateur* de x comme

$$G_x = \{g \in G \mid g \triangleright x = x\}.$$

Alors G_x est un sous-groupe de Lie de G . Nous considérons l'action de G_x sur G , ou plus généralement d'un sous-groupe de Lie H de G , évidemment définie par la loi de groupe. Nous avons (Gallier et Quaintance 2020, Corollaire 22.10 . [37]) :

Théorème 13 ► action des sous-groupe de lie

L'action d'un sous-groupe de Lie H d'un groupe de Lie G sur G est libre et propre.

Le Théorème précédent s'applique donc à l'action à droite $(g, h) \in G \times H \mapsto gh \in G$ et G/H est une variété différentiable telle que $\pi : G \rightarrow G/H$ est une submersion. Les orbites de cette action à droite sont les classes à gauche $\{gH \mid g \in G\}$.

Notons que G agit sur G/H par $g_1 \triangleright g_2 H = g_1 g_2 H$, et il est clair que cette action est transitive. En fait, toutes les actions à gauche transitives conduisent à un espace quotient d'orbites provenant d'une action à droite.

Définition 30 ► Espace homogène

On dit que M est homogène s'il existe une action lisse transitive d'un groupe de Lie G sur M .

Remarque 9

Notons qu'il peut exister plusieurs groupes de Lie différents avec une action transitive sur un même espace homogène M .

Tous les espaces homogènes correspondent (à un difféomorphisme près) à des espaces quotients de la forme G/H , comme énoncé dans le résultat suivant (Gallier et Quaintance 2020, Théorème 22.13 . [37]).

Théorème 14 ► quotient homogène

Soit G un groupe de Lie connexe agissant de manière lisse et transitive sur une variété différentiable M , de sorte que M soit homogène. Alors pour tout $x \in M$, en posant $H = G_x$, l'application

$$\begin{cases} G/H & \longrightarrow M \\ gH & \longmapsto g \triangleright x \end{cases}$$

est un difféomorphisme, donc $M \simeq G/H$.

Notons que le choix du point de référence x ci-dessus n'a pas d'importance, car pour deux choix différents x et y , l'action étant transitive, il existe $g \in G$ tel que $y = g \triangleright x$.

$y = g \triangleright x$, et les stabilisateurs G_x et $G_y = G_{g \circ x} = g G_x g^{-1}$ sont conjugués, donc isomorphes.

Nous illustrons toutes ces notions avec la sphère, la variété de Stiefel et la variété des matrices symétriques définies positives (SPD).

Exemple 27 ► Hypersphère

L'hypersphère plongée dans \mathbb{R}^{d+1} peut être vue comme un espace homogène en considérant l'action du groupe spécial orthogonal $SO(d+1)$ de l'espace ambiant. Pour $R \in SO(d+1)$, $x \in S^d$:

$$R \triangleright x = Rx \in S^d \quad (1.45)$$

En effet, la sphère est stable par l'action des matrices de rotation, et cette action est transitive. Considérons un pôle $x_0 = (1, 0, \dots, 0) \in S^d$. Son stabilisateur est l'ensemble des rotations dont l'axe est x_0 , c'est-à-dire :

$$H = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} \mid R \in SO(d) \right\} \simeq SO(d).$$

Alors, d'après le Théorème 14, $S^d = SO(d+1)/SO(d)$. Intuitivement, cela correspond à identifier toutes les rotations qui partagent le même axe, où les axes de rotation sont décrits par un vecteur unitaire, c'est-à-dire un point sur la sphère.

Exemple 28 ► Variété de Stiefel

Rappelons que la variété de Stiefel $St(k, n)$ est l'ensemble des k -repères orthonormés de \mathbb{R}^n . Le groupe spécial orthogonal $SO(n)$ agit donc sur $St(k, n)$ par :

$$R \triangleright (u_1, \dots, u_k) = (Ru_1, \dots, Ru_k)$$

Il est clair que cette action est transitive. De plus, comme un repère peut être représenté par une matrice orthogonale de taille $n \times k$, l'action ci-dessus correspond à une multiplication matricielle et est donc lisse.

Pour tout $U \in St(k, n)$, le stabilisateur de U est :

$$H = \left\{ \begin{pmatrix} I_k & 0 \\ 0 & R \end{pmatrix} \mid R \in SO(n-k) \right\} \simeq SO(n-k). \quad (1.46)$$

Alors, d'après le Théorème 14, $St(k, n) = SO(n)/SO(n-k)$. Intuitivement, cela correspond au fait qu'un k -repère peut être complété en une base orthonormée de \mathbb{R}^n , c'est-à-dire une matrice dans $SO(n)$, et les matrices dans H envoient un tel complètement vers tous les autres compléments. Ainsi, quotienter $SO(n)$ par $SO(n-k)$ revient à identifier les bases orthonormées qui coïncident sur les k premiers vecteurs.

Exemple 29 ► Variété de Grassmann

Définissons la variété de Grassmann $Gr(n, k)$ comme l'ensemble des sous-espaces de \mathbb{R}^n de dimension k . À tout sous-espace k -dimensionnel \mathcal{U} est associé un unique projecteur orthogonal, c'est-à-dire une application linéaire p définie sur \mathbb{R}^n vérifiant $p \circ p = p$, $\text{Im}(p) = \mathcal{U}$ et $\text{ker}(p) = \mathcal{U}^\perp$. Tout projecteur de ce type est représenté par une matrice symétrique P de taille n , de rang k et telle que $P^2 = P$. Nous adoptons donc la représentation de $Gr(n, k)$ comme l'ensemble :

$$Gr(n, k) = \{P \in \text{Sym}(n) \mid P^2 = P \text{ et } \text{rank}(P) = k\}. \quad (1.47)$$

Intuitivement, tout sous-espace de dimension k peut être représenté par une classe d'équivalence de bases orthonormées, de sorte que la variété de Grassmann est un espace quotient du groupe orthogonal $O(n)$. Pour considérer un groupe connexe, prenons $G = SO(n)$, et définissons l'action de G sur $Gr(k, n)$ comme un changement de base : $Q \triangleright P = QPQ^\top$. Il est clair que tout projecteur de rang k peut être représenté par une matrice de la forme :

$$P_k = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}.$$

Cela signifie précisément que l'action est transitive, et $Gr(k, n)$ est l'orbite de P_k . Soit H son stabilisateur, c'est-à-dire l'ensemble :

$$H = \left\{ \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix} \mid Q \in SO(k), R \in SO(n-k) \right\} \simeq SO(k) \times SO(n-k). \quad (1.48)$$

Ainsi, $Gr(k, n) = SO(n)/(SO(k) \times SO(n-k))$ est un espace homogène avec projection canonique $\pi : Q \in SO(n) \mapsto QP_kQ^\top$ et de dimension $k(n-k)$. Bien sûr, cette variété peut également être décrite comme un quotient de $St(k, n)$ par $SO(p)$, où l'action est donnée par multiplication à droite et la projection est identifiée à $U \in St(k, n) \mapsto UU^\top \in Gr(k, n)$. Ce quotient peut être plus pratique que celui de $SO(n)$ lorsque $n \gg k$.

Exemple 30 ► Matrices SPD

Rappelons qu'une matrice symétrique Σ est *définie positive* si elle est inversible et $\forall x \in \mathbb{R}^n, x^\top \Sigma x \geq 0$. L'ensemble $SPD(n)$ de ces matrices est un ouvert de l'espace vectoriel $\text{Sym}(n)$ des matrices symétriques, c'est donc une variété plongée et l'immersion canonique $\text{id} : SPD(n) \rightarrow \text{Sym}(n)$ définit une carte globale. De plus, pour toute matrice $\Sigma \in SPD(n)$, l'espace tangent $T_\Sigma SPD(n)$ s'identifie à $\text{Sym}(n)$.

Définissons maintenant l'action de $GL(n)$ sur $SPD(n)$ par :

$$\triangleright : (A, \Sigma) \in GL(n) \times SPD(n) \longmapsto A \Sigma A^\top. \quad (1.49)$$

Cette action est parfois appelée action par *congruence*. Elle est lisse et transitive. En effet, soient $\Sigma = PDP^\top$, $\Sigma' = Q\Delta Q^\top \in SPD(n)$ deux matrices SPD avec leur décomposition en valeurs propres donnée par le théorème spectral, alors $A = Q\Delta^{-1/2}D^{-1/2}P^\top$ est telle que $A \triangleright \Sigma = \Sigma'$. La variété des matrices SPD est donc un espace homogène.

Enfin, le groupe d'isotropie de la matrice identité I_n est le groupe orthogonal $O(n) \subset GL(n)$:

$$G_{I_n} = \{A \in GL(n) \mid A \triangleright I_n = I_n\} = \{A \in GL(n) \mid AA^\top = I_n\} = O(n). \quad (1.50)$$

C'est bien un sous-groupe fermé de $GL(n)$, donc son action à droite sur $GL(n)$ est libre et propre (Théorème 13), et par le Théorème 14, $SPD(n)$ est isomorphe à l'espace des orbites $GL(n)/O(n)$. La projection canonique

$$\pi : \begin{cases} GL(n) & \longrightarrow SPD(n) \\ A & \longmapsto AA^\top \end{cases}$$

est une submersion.

1.7 fonctions convexes dans les variétés

L'étude de la convexité sur les variétés riemanniennes — appelée **convexité géodésique** — dans le contexte de l'optimisation, a suscité un vif intérêt dès les années 1970. D'excellentes références sur le sujet incluent notamment les ouvrages d'Udriste [60] et de Boumal [12].

Il existe plusieurs façons de définir les ensembles géodésiquement convexes. Cette diversité s'explique par le fait que les exigences de convexité varient selon le contexte

d'application. Nous adoptons ici la définition de Rapcsák [12], que nous relierons à deux autres définitions.

1.7.1 Ensembles et fonctions convexes dans les espaces vectoriels

Rappelons qu'un sous-ensemble S d'un espace vectoriel E est dit **convexe** si, pour tous $x, y \in S$, le segment de droite

$$t \mapsto (1 - t)x + ty, \quad t \in [0, 1],$$

est entièrement contenu dans S .

De plus :

Définition 31

une fonction $f : S \rightarrow \mathbb{R}$ est dite **convexe** si S est convexe et si, pour tout $x, y \in S$,

$$\forall t \in [0, 1], \quad f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y).$$

De même, f est **strictement convexe** si, pour tout $x \neq y$,

$$\forall t \in (0, 1), \quad f((1 - t)x + ty) < (1 - t)f(x) + tf(y).$$

Si E est un espace euclidien muni de la norme $\|\cdot\|$, on dit que f est **μ -fortement convexe** pour un certain $\mu > 0$ si la fonction

$$x \mapsto f(x) - \frac{\mu}{2}\|x\|^2$$

est convexe. De manière équivalente, pour tout $x, y \in S$ et $t \in [0, 1]$, on a :

$$f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y) - \frac{t(1 - t)\mu}{2}\|x - y\|^2. \quad (1.51)$$

En optimisation, l'intérêt principal des fonctions convexes réside dans le fait que leurs **minimiseurs locaux** (s'ils existent) sont également **minimiseurs globaux**.

1.7.2 Ensembles et fonctions géodésiquement convexes

Définition 32 ► Ensemble géodésiquement convexe

Un sous-ensemble S d'une variété riemannienne \mathcal{M} est dit **géodésiquement convexe** si, pour tout $x, y \in S$, il existe un segment géodésique $c : [0, 1] \rightarrow \mathcal{M}$ tel que :

$$c(0) = x, \quad c(1) = y, \quad \text{et} \quad c(t) \in S \quad \forall t \in [0, 1].$$

Dans cette définition, c est une **géodésique de la variété \mathcal{M}** , et non nécessairement de S .

Si \mathcal{M} est un espace euclidien, alors un sous-ensemble est convexe au sens habituel si et seulement si il est géodésiquement convexe. En effet, la seule géodésique reliant deux points x et y (à reparamétrisation près) est :

$$c(t) = (1 - t)x + ty.$$

Théorème 15

Toute variété riemannienne connexe et complète est géodésiquement convexe.

Cela inclut notamment :

- les sphères ;
- la variété de Stiefel $\text{St}(n, p)$ pour $p < n$;
- et le groupe des rotations $\text{SO}(n)$.

Soit S un sous-ensemble d'une variété \mathcal{M} . On dit qu'une courbe c sur \mathcal{M} **relie** x à y dans S si :

$$c(0) = x, \quad c(1) = y, \quad \text{et} \quad c(t) \in S \quad \forall t \in [0, 1].$$

Dans un ensemble géodésiquement convexe S , tout couple de points $x, y \in S$ est relié dans S par au moins une géodésique c .

En composant une fonction $f : S \rightarrow \mathbb{R}$ avec une telle géodésique, on obtient une fonction réelle sur $[0, 1]$. Si toutes ces fonctions composées sont convexes au sens usuel, alors f est dite **convexe au sens géométrique**. Notons que f n'a pas besoin d'être continue, ni même différentiable.

Définition 33 ► Fonction géodésiquement convexe

Une fonction $f : S \rightarrow \mathbb{R}$ est dite **géodésiquement convexe** (ou **strictement géodésiquement convexe**) si :

1. S est géodésiquement convexe ;
2. pour toute géodésique $c : [0, 1] \rightarrow \mathcal{M}$ telle que $c([0, 1]) \subseteq S$ et $c(0) \neq c(1)$, la fonction composée

$$f \circ c : [0, 1] \rightarrow \mathbb{R}$$

est (strictement) convexe au sens usuel.

Dans cette définition, la structure riemannienne est celle de \mathcal{M} , pour laquelle $S \subseteq \mathcal{M}$ est géodésiquement convexe et les courbes c sont des géodésiques de \mathcal{M} .

Définition 34 ► Convexité géodésique forte

Une fonction $f : S \rightarrow \mathbb{R}$ est dite **géodésiquement μ -fortement convexe** pour un certain $\mu > 0$ si :

- l'ensemble S est géodésiquement convexe ;
- pour chaque segment géodésique $c : [0, 1] \rightarrow \mathcal{M}$ dont l'image est contenue dans S , on a :

$$f(c(t)) \leq (1-t)f(c(0)) + tf(c(1)) - \frac{t(1-t)\mu}{2}L(c)^2,$$

où $L(c) = \|c'(0)\|_{c(0)}$ désigne la longueur du segment géodésique.

Cette condition est équivalente à la forte convexité usuelle de la fonction composée $f \circ c : [0, 1] \rightarrow \mathbb{R}$, qui est alors $\mu L(c)^2$ -fortement convexe.

La forte convexité géodésique implique évidemment la **convexité géodésique stricte**.

Comme dans le cas de la convexité standard dans les espaces linéaires, la convexité géodésique garantit que tout **minimiseur local**, s'il existe, est également un **minimiseur global**.

Théorème 16 ► Tout minimum local est global

Si $f : S \rightarrow \mathbb{R}$ est géodésiquement convexe, alors tout minimiseur local de f est un minimiseur global.

Preuve: Supposons, par contradiction, qu'un point $x \in S$ soit un minimum local mais non un minimum global. Il existe alors un point $y \in S$ tel que $f(y) < f(x)$.

Comme il existe une géodésique c reliant x à y dans S , pour tout $t \in (0, 1]$ on a :

$$f(c(t)) \leq (1-t)f(x) + tf(y) = f(x) + t(f(y) - f(x)) < f(x),$$

ce qui contredit l'hypothèse selon laquelle x est un minimum local. Ainsi, tout minimum local est nécessairement un minimum global. \square

La **convexité stricte** assure, quant à elle, l'unicité du minimum, lorsque celui-ci existe.

Théorème 17 ► Unicité du minimum

Si $f : S \rightarrow \mathbb{R}$ est géodésiquement strictement convexe, alors elle admet au plus un minimum local, qui est nécessairement global.

Preuve: D'après le Théorème 16, tout minimum local est global. Supposons, par contradiction, qu'il existe deux minimums globaux distincts x et y tels que :

$$f(x) = f(y) = f^*.$$

Il existe alors une géodésique c reliant x à y dans S . Pour tout $t \in (0, 1)$, la stricte convexité implique :

$$f(c(t)) < (1-t)f(x) + tf(y) = f^*,$$

ce qui contredit le caractère global des minimums x et y . \square

1.7.3 Sous-niveaux et intersections convexes

Les ensembles

$$S_i = \{x \in S : f_i(x) \leq \alpha_i\}.$$

de sous-niveau d'une fonction géodésiquement convexe sont eux-mêmes géodésiquement convexes. De plus, l'intersection de plusieurs de ces ensembles reste géodésiquement convexe. Cependant, l'intersection arbitraire de sous-ensembles géodésiquement convexes n'est pas nécessairement géodésiquement convexe [12].

Théorème 18 ► Convexité des sous-niveaux

Soit $i \in I$ un indice appartenant à une collection quelconque de fonctions géodésiquement convexes $f_i : S \rightarrow \mathbb{R}$, et soit $\alpha_i \in \mathbb{R}$. On définit les ensembles de sous-niveau :

$$S_i = \{x \in S : f_i(x) \leq \alpha_i\}.$$

Alors chaque S_i est géodésiquement convexe, et leur intersection l'est également.

La réciproque n'est pas vraie : une fonction définie sur un ensemble géodésiquement convexe est dite géodésiquement quasiconvexe si tous ses sous-niveaux sont géodésiquement convexes [12].

Continuité et convexité sur les variétés compactes

Une **variété riemannienne connexe et complète** est elle-même un **ensemble géodésiquement convexe**. Son intérieur est la variété entière. [12]

Théorème 19 ► Continuité des fonctions géodésiquement convexes

Si $f : S \rightarrow \mathbb{R}$ est géodésiquement convexe, alors f est continue à l'intérieur de S .

Les variétés compactes sont complètes, et toute fonction continue sur un ensemble compact atteint son maximum. En considérant les géodésiques passant par un point maximisant f , on déduit :

Théorème 20 ► Constante sur les variétés compactes

Si \mathcal{M} est une variété riemannienne connexe et compacte, et si $f : \mathcal{M} \rightarrow \mathbb{R}$ est géodésiquement convexe, alors f est constante.

Ces résultats montrent que, sur les **variétés compactes**, la convexité géodésique n'est intéressante que lorsqu'elle est restreinte à un **sous-ensemble** d'une composante connexe.

En général, lorsqu'une fonction géodésiquement convexe admet un maximum (unique ou multiple), celui-ci se situe typiquement sur la **frontière du domaine géodésiquement convexe** :

La définition permissive introduite précédemment permet d'établir la plupart des résultats d'optimisation souhaités, mais elle présente certains **inconvenients**.

Par exemple, l'intersection de deux ensembles géodésiquement convexes peut ne pas être géodésiquement convexe.

Exemple 31

Considérons la variété riemannienne :

$$\mathcal{M} = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\},$$

c'est-à-dire le **cercle unité** vu comme sous-variété riemannienne de \mathbb{R}^2 . Définissons :

$$S_1 = \{x \in \mathcal{M} : x_1 \geq 0\}, \quad S_2 = \{x \in \mathcal{M} : x_1 \leq 0\}.$$

Les ensembles S_1 et S_2 sont clairement géodésiquement convexes, mais leur intersection :

$$S_1 \cap S_2 = \{(0, 1), (0, -1)\}$$

ne l'est pas.

Définition 35 ► Convexité géodésique totale

Un sous-ensemble S d'une variété riemannienne \mathcal{M} est dit **géodésiquement totalement convexe** si, pour tout $x, y \in S$, il existe au moins une géodésique $c : [0, 1] \rightarrow \mathcal{M}$ telle que :

$$c(0) = x, \quad c(1) = y,$$

et pour tout tel segment géodésique, on a :

$$c(t) \in S, \quad \forall t \in [0, 1].$$

Définition 36 ► Convexité géodésique forte

Un sous-ensemble S d'une variété riemannienne \mathcal{M} est dit **géodésiquement fortement convexe** si, pour tout $x, y \in S$, il existe une unique géodésique minimisante $c : [0, 1] \rightarrow \mathcal{M}$ telle que :

$$c(0) = x, \quad c(1) = y, \quad \text{et} \quad c(t) \in S \quad \forall t \in [0, 1].$$

Si un ensemble S est géodésiquement **totalement convexe** ou **fortement convexe**, alors il est nécessairement **géodésiquement convexe**.

Cependant :

1. la réciproque n'est pas vraie dans les deux cas ;

2. la convexité totale n'implique pas la convexité forte, et inversement.

Ces trois notions de convexité coïncident uniquement dans certains cas particuliers de variétés riemanniennes.

Théorème 21 ► Équivalence sur les variétés de Cartan–Hadamard

Supposons que \mathcal{M} soit une variété riemannienne complète telle que chaque paire de points $x, y \in \mathcal{M}$ soit reliée par une unique géodésique. Alors, les trois notions suivantes sont équivalentes :

1. convexité géodésique ;
2. convexité géodésique totale ;
3. convexité géodésique forte.

Dans ce cas, un sous-ensemble $S \subseteq \mathcal{M}$ est géodésiquement convexe si et seulement si, pour tout $x, y \in S$, la géodésique reliant x et y reste entièrement dans S .

Preuve: Par hypothèse, la variété \mathcal{M} est connexe et complète. Le Théorème 15 garantit que toute paire de points $x, y \in \mathcal{M}$ peut être reliée par une géodésique minimisante. De plus, aucune autre géodésique ne relie x et y . Ainsi, la géodésique reliant x à y est à la fois unique et minimisante, ce qui prouve le résultat. \square

Pour tout point x d'une variété riemannienne, il existe un rayon positif $r_0 > 0$ tel que toute **boule géodésique** de rayon $r \leq r_0$, centrée en x , soit **géodésiquement fortement convexe** [12].

La notion de **fonction géodésiquement convexe** $f : S \rightarrow \mathbb{R}$, telle que définie à la Définition 33 s'étend naturellement au cas où S est un ensemble **géodésiquement totalement** ou **fortement convexe**. Dans ces cas, on continue d'appeler ces fonctions **géodésiquement convexes**.

1.7.4 Fonctions différentiables géodésiquement convexes

Pour les fonctions possédant un **gradient** ou un **hessien**, la convexité géodésique peut être caractérisée à l'aide d'inégalités différentielles évaluées en un point de base.

Tout segment géodésique $c : [0, 1] \rightarrow \mathcal{M}$ admet une extension unique à un intervalle ouvert maximal contenant $[0, 1]$. C'est dans ce cadre que l'on définit rigoureusement les dérivées $c'(0)$ et $c'(1)$.

Théorème 22 ► Caractérisation par le gradient

Soit S un ensemble géodésiquement convexe d'une variété riemannienne \mathcal{M} , et soit $f : \mathcal{M} \rightarrow \mathbb{R}$ une fonction différentiable dans un voisinage de S . Alors, la restriction $f|_S : S \rightarrow \mathbb{R}$ est géodésiquement convexe si et seulement si, pour tout segment géodésique $c : [0, 1] \rightarrow \mathcal{M}$ contenu dans S , en notant $x = c(0)$, on a :

$$\forall t \in [0, 1], \quad f(c(t)) \geq f(x) + t \langle \text{grad } f(x), c'(0) \rangle_x. \quad (1.52)$$

De plus, $f|_S$ est géodésiquement μ -fortement convexe (pour un certain $\mu > 0$) si et seulement si :

$$\forall t \in [0, 1], \quad f(c(t)) \geq f(x) + t \langle \text{grad } f(x), c'(0) \rangle_x + \frac{t^2 \mu}{2} L(c)^2. \quad (1.53)$$

Enfin, $f|_S$ est géodésiquement strictement convexe si et seulement si, pour tout c tel que $c'(0) \neq 0$,

$$\forall t \in (0, 1], \quad f(c(t)) > f(x) + t \langle \text{grad } f(x), c'(0) \rangle_x. \quad (1.54)$$

Preuve : Par définition, $f|_S$ est (strictement) géodésiquement convexe si et seulement si, pour tout $x, y \in S$ et pour toute géodésique c reliant x à y dans S , la fonction composée $f \circ c$ est (strictement) convexe de $[0, 1]$ vers \mathbb{R} .

En prolongeant légèrement le domaine de c , on voit que $f \circ c$ est différentiable sur un intervalle ouvert contenant $[0, 1]$.

Tout d'abord, $f \circ c$ est convexe si et seulement si, pour tout $s, t \in [0, 1]$,

$$f(c(t)) \geq f(c(s)) + (t - s)(f \circ c)'(s).$$

Puisque f est différentiable dans un voisinage de S , on a :

$$(f \circ c)'(s) = Df(c(s))[c'(s)] = \langle \text{grad } f(c(s)), c'(s) \rangle_{c(s)}.$$

En combinant ces deux relations et en prenant $s = 0$, on conclut que si $f|_S$ est géodésiquement convexe, alors l'inégalité (1.52) est vérifiée. Réciproquement, si (1.52) est satisfaite, alors (par reparamétrisation de c) on en déduit que $f \circ c$ est convexe pour toute géodésique admissible, et donc que $f|_S$ est géodésiquement convexe.

La démonstration pour la forte convexité est analogue.

Enfin, en supposant $c'(0) \neq 0$, on a que $f \circ c$ est strictement convexe si et

seulement si, pour tout $s, t \in [0, 1]$ distincts :

$$f(c(t)) > f(c(s)) + (t - s)(f \circ c)'(s).$$

En utilisant la différentiabilité de f et en prenant $s = 0$, on obtient que $f \circ c$ est strictement convexe si et seulement si l'inégalité (1.54) est vérifiée. La conclusion suit comme dans la première partie. \square

Théorème 23 ► Caractérisation du minimum global

Si f est différentiable et géodésiquement convexe sur un ouvert géodésiquement convexe, alors :

$$x \text{ est un minimum global de } f \iff \text{grad } f(x) = 0.$$

Preuve: Si $\text{grad } f(x) = 0$, alors le Théorème 22 montre que $f(x) \leq f(y)$ pour tout y dans le domaine de f . (Cette direction ne nécessite pas que le domaine soit ouvert.)

Réciproquement, puisque le domaine de f est ouvert, il constitue une sous-variété ouverte de \mathcal{M} . On peut alors appliquer la Proposition 17, qui affirme que si x est un minimum global, alors :

$$\text{grad } f(x) = 0. \square$$

Théorème 24 ► Caractérisation par la Hessienne

Soit $f : S \rightarrow \mathbb{R}$ une fonction deux fois différentiable sur un ensemble ouvert géodésiquement convexe S . Alors :

1. f est **géodésiquement convexe** si et seulement si

$$\text{Hess } f(x) \succeq 0;$$

2. f est **géodésiquement μ -fortement convexe** si et seulement si

$$\text{Hess } f(x) \succeq \mu \text{Id};$$

3. f est **géodésiquement strictement convexe** si (mais pas seulement si)

$$\text{Hess } f(x) \succ 0,$$

le tout étant entendu pour tout $x \in S$.

Preuve: De manière analogue à la démonstration du Théorème 22, on commence par remarquer que f est géodésiquement convexe si et seulement si la fonction composée

$$f \circ c : [0, 1] \rightarrow \mathbb{R}$$

est convexe pour tout segment géodésique $c : [0, 1] \rightarrow \mathcal{M}$ dont l'image est contenue dans S . En appliquant le Lemme 11.1, cette condition équivaut à :

$$\forall t \in (0, 1), \quad (f \circ c)''(t) \geq 0.$$

Or, comme f est deux fois différentiable sur S , on a :

$$(f \circ c)''(t) = \frac{d}{dt} \langle \text{grad } f(c(t)), c'(t) \rangle_{c(t)} = \langle \text{Hess } f(c(t))[c'(t)], c'(t) \rangle_{c(t)},$$

où l'on a utilisé le fait que $c''(t) = 0$ puisque c est une géodésique.

Ainsi, si $\text{Hess } f(x)$ est semi-définie positive pour tout $x \in S$, alors :

$$(f \circ c)''(t) \geq 0 \quad \forall t \in (0, 1),$$

ce qui entraîne que f est géodésiquement convexe.

Réciproquement, si f est géodésiquement convexe, alors :

$$\langle \text{Hess } f(c(0))[c'(0)], c'(0) \rangle_{c(0)} \geq 0$$

pour toute géodésique admissible c (en particulier pour $t = 0$). Puisque S est ouvert dans \mathcal{M} , pour tout $x \in S$ et pour tout $v \in T_x \mathcal{M}$ suffisamment petit, la géodésique c définie par :

$$c(0) = x, \quad c'(0) = v$$

reste dans S pour $t \in [0, 1]$. On obtient donc :

$$\langle \text{Hess } f(x)[v], v \rangle_x \geq 0,$$

ce qui prouve que $\text{Hess } f(x)$ est semi-définie positive en tout point $x \in S$.

La même démonstration s'applique pour **la forte convexité**, en utilisant ou en montrant que :

$$(f \circ c)''(t) \geq \mu L(c)^2$$

pour toute géodésique admissible c et tout $t \in [0, 1]$, en rappelant que :

$$L(c) = \|c'(t)\|_{c(t)}$$

puisque c est une géodésique définie sur $[0, 1]$.

Enfin, si $\text{Hess } f(x)$ est **définie positive** en tout $x \in S$, alors :

$$(f \circ c)''(t) > 0 \quad \text{dès que} \quad c'(0) \neq 0,$$

ce qui garantit que f est **géodésiquement strictement convexe**. □

La réciproque n'est pas vraie, tout comme dans le cas euclidien. En effet, prenons $f(x) = x^4$ sur $S = (-1, 1) \subset \mathbb{R}$: la fonction est strictement convexe sans que sa Hessienne soit partout définie positive.

Théorème 25 ► Cas d'une variété compacte [12]

Soit \mathcal{M} une variété riemannienne compacte. Si $f : \mathcal{M} \rightarrow \mathbb{R}$ possède une Hessienne semi-définie positive en tout point, alors, d'après le Théorème 24, f est géodésiquement convexe sur chaque composante connexe de \mathcal{M} .

Or, comme \mathcal{M} est compacte, le Théorème 23 implique que f est constante sur chaque composante connexe. Ainsi, la Hessienne de f est en réalité identiquement nulle sur toute la variété.

Théorème 26 ► Convexité géodésique et métrique adaptée [12]

Soit $f : \mathcal{M} \rightarrow \mathbb{R}$ une fonction différentiable définie sur une variété \mathcal{M} . Il est naturel de se demander s'il existe une métrique riemannienne sur \mathcal{M} telle que f soit géodésiquement convexe.

Si une telle métrique existe, alors :

1. le domaine de f est géodésiquement convexe pour cette métrique ;
2. les points critiques de f sont nécessairement ses minimiseurs globaux .

Chapitre 2

Optimisations et Calculs statistiques dans les Variétés différentielles

2.1 Optimisation dans les variétés

les notions de **point critique** et de **minimum global** sont indépendantes de la métrique riemannienne. Ainsi, toute fonction f possédant un point critique sous-optimal ne peut être rendue géodésiquement convexe par aucune métrique riemannienne.

De même, si f (même non différentiable) possède un minimum local sous-optimal, cela reste vrai [12]. Toutefois, il peut exister des métriques qui rendent f géodésiquement convexe sur certaines sous-variétés de \mathcal{M} .

2.1.1 Convexité géodésique forte et gradients Lipschitziens

Si S est un ensemble géodésiquement convexe et que $x, y \in S$, il existe un segment géodésique dans S reliant x à y . En termes d'application exponentielle, cela signifie qu'il existe un vecteur tangent $v \in T_x\mathcal{M}$ tel que la courbe :

$$c(t) = \text{Exp}_x(tv)$$

reste entièrement contenue dans S pour tout $t \in [0, 1]$, avec :

$$c(0) = x, \quad c(1) = y.$$

De plus, la longueur de ce segment géodésique satisfait :

$$\text{dist}(x, y) \leq L(c) = \|v\|_x,$$

D'après ce qui précède, si $f : \mathcal{M} \rightarrow \mathbb{R}$ est différentiable dans un voisinage de S et si $f|_S$ est géodésiquement convexe, alors pour tout $x \in S$ et pour tout $v \in T_x\mathcal{M}$ tel que $c(t) = \text{Exp}_x(tv)$ reste dans S pour $t \in [0, 1]$, on a :

$$\forall t \in [0, 1], \quad f(\text{Exp}_x(tv)) \geq f(x) + t\langle \text{grad } f(x), v \rangle_x. \quad (2.1)$$

L'inégalité est stricte pour tout $t \in (0, 1]$ si $f|_S$ est géodésiquement strictement convexe. De plus, si $f|_S$ est géodésiquement μ -fortement convexe, alors :

$$\forall t \in [0, 1], \quad f(\text{Exp}_x(tv)) \geq f(x) + t\langle \text{grad } f(x), v \rangle_x + \frac{t^2\mu}{2}\|v\|_x^2. \quad (2.2)$$

Ces inégalités se comparent à celle obtenue lorsque le gradient de f est L -Lipschitzien :

$$\forall t \in [0, 1], \quad f(\text{Exp}_x(tv)) \leq f(x) + t\langle \text{grad } f(x), v \rangle_x + \frac{t^2L}{2}\|v\|_x^2. \quad (2.3)$$

Lorsque les inégalités (2.2) et (2.3) sont simultanément vérifiées, elles permettent d'obtenir des résultats particulièrement puissants en optimisation riemannienne, notamment des bornes de convergence pour les algorithmes de descente de gradient géodésique.

Théorème 27 ► Existence et unicité de minimum

Soit S un ensemble non vide, fermé et géodésiquement convexe dans une variété complète \mathcal{M} . On suppose que $f : \mathcal{M} \rightarrow \mathbb{R}$ est différentiable dans un voisinage de S . Si $f|_S$ est géodésiquement μ -fortement convexe avec $\mu > 0$, alors les ensembles de sous-niveau de $f|_S$ sont compacts et $f|_S$ admet un unique minimiseur global.

Preuve: Soit $x_0 \in S$ quelconque. Considérons l'ensemble de sous-niveau :

$$S_0 = \{x \in S \mid f(x) \leq f(x_0)\}.$$

Puisque f est continue autour de S et que S est fermé, S_0 est fermé. Comme \mathcal{M} est complète, il suffit de montrer que S_0 est borné dans \mathcal{M} .

Supposons le contraire : il existerait une suite $(x_k)_{k \geq 1} \subset S_0$ telle que :

$$\lim_{k \rightarrow \infty} \text{dist}(x_0, x_k) = +\infty.$$

Chaque $x_k \in S$, et S étant géodésiquement convexe, il existe pour chaque k un

vecteur $v_k \in T_{x_0}\mathcal{M}$ tel que la géodésique :

$$c_k(t) = \text{Exp}_{x_0}(tv_k)$$

reste dans S pour $t \in [0, 1]$ et vérifie $c_k(1) = x_k$.

D'après l'inégalité (2.2), on a :

$$f(x_k) \geq f(x_0) + \langle \text{grad } f(x_0), v_k \rangle_{x_0} + \frac{\mu}{2} \|v_k\|_{x_0}^2.$$

Comme $\text{dist}(x_0, x_k) \rightarrow +\infty$ et que $\text{dist}(x_0, x_k) \leq L(c_k) = \|v_k\|_{x_0}$, on en déduit que $\|v_k\|_{x_0} \rightarrow +\infty$.

De plus, pour tout k ,

$$f(x_k) \geq f(x_0) - \|\text{grad } f(x_0)\|_{x_0} \|v_k\|_{x_0} + \frac{\mu}{2} \|v_k\|_{x_0}^2.$$

Le membre de droite tend vers $+\infty$ lorsque $k \rightarrow \infty$, donc $f(x_k) \rightarrow +\infty$. Cela contredit le fait que $f(x_k) \leq f(x_0)$ pour tout k . Ainsi, S_0 est compact.

Comme $f|_{S_0}$ est continue, elle atteint son minimum en un point $x^* \in S_0$. Pour tout $x \in S$, on a :

— soit $x \notin S_0$, et alors $f(x) > f(x_0) \geq f(x^*)$; — soit $x \in S_0$, et alors $f(x) \geq f(x^*)$.

Dans les deux cas, x^* est bien un minimiseur global de $f|_S$. Enfin, la forte convexité géodésique impliquant la convexité géodésique stricte, le Théorème 17 garantit que ce minimiseur est unique. \square

Théorème 28 ► Inégalité de Polyak–Łojasiewicz

Soit S un ensemble non vide, fermé et géodésiquement convexe dans une variété complète \mathcal{M} . On suppose que $f : \mathcal{M} \rightarrow \mathbb{R}$ est différentiable dans un voisinage de S . Si $f|_S$ est géodésiquement μ -fortement convexe avec $\mu > 0$, alors f satisfait l'inégalité de Polyak–Łojasiewicz :

$$\forall x \in S, \quad f(x) - f(x^*) \leq \frac{1}{2\mu} \|\text{grad } f(x)\|_x^2, \quad (2.4)$$

où x^* désigne le minimiseur unique de $f|_S$.

Cette inégalité fournit une borne directe de l'écart de sous-optimalité $f(x) - f(x^*)$ en fonction de la norme du gradient en x . Ainsi, la forte convexité géodésique implique non seulement l'existence et l'unicité du minimum, mais aussi une convergence quadratique du gradient vers zéro à proximité de l'optimum — une propriété

essentielle en optimisation riemannienne.

Les deux résultats précédents fournissent le contexte nécessaire pour étudier une version simple de la descente de gradient riemannienne appliquée à une fonction géodésiquement fortement convexe et dont le gradient est Lipschitz-continu.

Théorème 29

Soit $f : \mathcal{M} \rightarrow \mathbb{R}$ une fonction différentiable et géodésiquement convexe sur une variété complète \mathcal{M} . Étant donné $x_0 \in \mathcal{M}$, considérons l'ensemble de sous-niveau :

$$S_0 = \{x \in \mathcal{M} : f(x) \leq f(x_0)\}.$$

On suppose que f possède un gradient L -Lipschitz-continu sur un voisinage de S_0 et que $f|_{S_0}$ est géodésiquement μ -fortement convexe avec $\mu > 0$. Considérons la descente de gradient avec rétraction exponentielle et pas constant $1/L$, initialisée en x_0 :

$$x_{k+1} = \text{Exp}_{x_k} \left(-\frac{1}{L} \text{grad } f(x_k) \right), \quad k = 0, 1, 2, \dots$$

Alors f admet un unique minimiseur x^* et les itérés x_k convergent vers celui-ci au moins linéairement. Plus précisément, en posant $\kappa = L/\mu \geq 1$ (le nombre de condition de $f|_{S_0}$), la suite entière reste dans S_0 et :

$$f(x_k) - f(x^*) \leq \left(1 - \frac{1}{\kappa}\right)^k (f(x_0) - f(x^*)) \quad (2.5)$$

et :

$$\text{dist}(x_k, x^*) \leq \sqrt{1 - \frac{1}{\kappa}} \sqrt{\kappa} \text{dist}(x_0, x^*) \quad (2.6)$$

pour tout $k \geq 0$. (Remarquons que $\sqrt{1 - \frac{1}{\kappa}} \leq 1 - \frac{1}{2\kappa}$)

Preuve Par construction, S_0 est non vide. Il est aussi fermé puisque f est continue, et géodésiquement convexe puisque f l'est. Ainsi, le théorème 27 garantit que $f|_{S_0}$ possède un minimiseur unique $x^* \in S_0$, qui est aussi le minimiseur unique de f sur \mathcal{M} , car $x \notin S_0 \Rightarrow f(x) > f(x_0) \geq f(x^*)$.

Ensuite, raisonnons par induction sur le fait que tous les x_k appartiennent à S_0 . Évidemment, $x_0 \in S_0$. Supposons $x_k \in S_0$. On sait que $\text{grad } f$ est L -Lipschitz-continu sur un voisinage U de S_0 . Considérons la courbe $c(t) = \text{Exp}_{x_k}(-t \text{grad } f(x_k))$. Remarquons que $c(0) \in U$. Soit I le plus grand intervalle autour de $t = 0$ tel que $c(t) \in U$ pour tout $t \in I$.

Cet intervalle est ouvert puisque $c^{-1}(U)$ est ouvert. Alors, la continuité Lipschit-

zienne du gradient fournit :

$$\forall t \in I, \quad f(c(t)) \leq f(x_k) - t \left(1 - \frac{tL}{2}\right) \|\text{grad } f(x_k)\|_{x_k}^2.$$

Nous voulons montrer que I contient l'intervalle $[0, 2/L]$. Pour ce faire, notons $\bar{t} = \sup I$ le premier (positif) instant tel que $c(\bar{t})$ quitte U . Si \bar{t} est infini, il n'y a rien à démontrer. Supposons \bar{t} fini. L'inégalité ci-dessus est valable pour tout $0 \leq t < \bar{t}$. Par continuité, elle doit aussi être vraie pour $t = \bar{t}$.

Pour obtenir une contradiction, supposons que $\bar{t} < 2/L$. Alors :

$$f(c(\bar{t})) \leq f(x_k) \leq f(x_0)$$

□

2.2 Applications

2.2.1 Convexité et Sous variétés des réels positifs

Notons \mathbb{R}^n l'espace euclidien muni du produit scalaire standard $\langle u, v \rangle = u^T v$. L'orthant positif est défini par :

$$\mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_1, \dots, x_n > 0\}.$$

(C'est un sous-ensemble convexe de \mathbb{R}^n au sens usuel. Étant un ensemble ouvert, il constitue également une sous-variété ouverte de \mathbb{R}^n . Ses espaces tangents sont tous naturellement identifiés à \mathbb{R}^n .)

On peut faire de \mathbb{R}_+^n une sous-variété riemannienne de \mathbb{R}^n en utilisant la métrique euclidienne. Dans ce cas, la convexité géodésique sur cette variété est équivalente à la convexité au sens classique, ce qui n'est pas particulièrement intéressant. De plus, cette variété n'est pas complète : ses géodésiques sont les droites de \mathbb{R}^n , qui cessent d'exister dès qu'elles sortent de \mathbb{R}_+^n .

Nous pouvons cependant associer \mathbb{R}_+^n d'une métrique riemannienne différente de manière à la rendre complète. Cela conduit à une notion distincte de **convexité géodésique** sur \mathbb{R}_+^n . L'idée clé consiste à établir un **difféomorphisme** entre \mathbb{R}^n et \mathbb{R}_+^n , puis à **transporter la géométrie riemannienne** de \mathbb{R}^n vers \mathbb{R}_+^n par ce difféomorphisme.

À cet effet, considérons l'application :

$$\varphi : \mathbb{R}_+^n \rightarrow \mathbb{R}^n, \quad \varphi(x) = \log(x) = (\log x_1, \dots, \log x_n)^T.$$

(C'est un **difféomorphisme** entre les variétés \mathbb{R}_+^n et \mathbb{R}^n , car elle est **lisse** et son inverse

$$\varphi^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}_+^n, \quad \varphi^{-1}(y) = \exp(y) = (e^{y_1}, \dots, e^{y_n})^T, \quad (2.7)$$

est également lisse.

Notons aussi les expressions suivantes pour le **différentiel** de φ en $x \in \mathbb{R}_+^n$ et pour son inverse (les deux étant des applications de \mathbb{R}^n dans \mathbb{R}^n) :

$$D\varphi(x)[u] = \left(\frac{u_1}{x_1}, \dots, \frac{u_n}{x_n} \right)^\top, \quad (D\varphi(x))^{-1}[z] = (x_1 z_1, \dots, x_n z_n)^\top.$$

Ces formules seront utiles par la suite.

Équipés de ce difféomorphisme, nous pouvons définir une **métrique riemannienne** $\langle \cdot, \cdot \rangle_+$ sur \mathbb{R}_+^n comme suit : $D\varphi(x)$ est une application linéaire inversible de $T_x \mathbb{R}_+^n$ vers $T_{\varphi(x)} \mathbb{R}^n$, et nous définissons le produit scalaire sur $T_x \mathbb{R}_+^n$ de manière à faire de cette application une **isométrie**, c'est-à-dire :

$$\langle u, v \rangle_x^+ \triangleq \langle D\varphi(x)[u], D\varphi(x)[v] \rangle = \sum_{i=1}^n \frac{u_i v_i}{x_i^2}. \quad (2.8)$$

(Remarquez que la métrique en x est donnée par le **hessien euclidien** de la fonction barrière logarithmique $x \mapsto -\sum_{i=1}^n \log(x_i)$.)

Comme \mathbb{R}_+^n possède maintenant deux géométries riemanniennes distinctes, nous notons :

$$\mathcal{M} = (\mathbb{R}_+^n, \langle \cdot, \cdot \rangle_+) \quad (2.9)$$

la **variété riemannienne** obtenue avec la métrique transportée. Les géodésiques de \mathcal{M} sont exactement les **images des géodésiques de \mathbb{R}^n** via φ^{-1} . Autrement dit, toutes les géodésiques de \mathcal{M} sont de la forme :

$$c(t) = \varphi^{-1}(y + tz) = \exp(y + tz) = (e^{y_1 + tz_1}, \dots, e^{y_n + tz_n}), \quad (2.10)$$

pour certains $y, z \in \mathbb{R}^n$.

Ces courbes sont définies pour tout $t \in \mathbb{R}$, d'où le fait que \mathcal{M} est **complète**. (*Intuitivement, à mesure que l'on s'approche de la frontière absente de \mathbb{R}_+^n , c'est-à-dire lorsque certaines composantes x_i tendent vers zéro, le facteur d'échelle $1/x_i^2$ de la métrique déforme les longueurs, rendant la frontière infiniment éloignée.*)

De plus, pour tout couple de points $x, x' \in \mathcal{M}$, il existe une **géodésique unique**

$c : [0, 1] \rightarrow \mathcal{M}$ (nécessairement minimisante) qui les relie :

$$c(t) = \exp(\log(x) + t(\log(x') - \log(x))). \quad (2.11)$$

Nous sommes désormais en bonne position pour étudier la **convexité géodésique**

Théorème 30

Un ensemble $S \subseteq \mathbb{R}_+^n$ est **géodésiquement convexe** sur \mathcal{M} si et seulement si $C = \log(S)$ est convexe dans \mathbb{R}^n .

Preuve Supposons que S soit géodésiquement convexe. Pour tout couple $y, y' \in C$, posons $x = \varphi^{-1}(y)$ et $x' = \varphi^{-1}(y')$ les points correspondants dans S . Puisque S est géodésiquement convexe, la géodésique (2.11) est incluse dans S pour tout $t \in [0, 1]$. Ainsi, C contient :

$$\varphi(c(t)) = \log(x) + t(\log(x') - \log(x)) = y + t(y' - y), \quad t \in [0, 1],$$

c'est-à-dire le **segment de droite reliant y à y'** . Donc, C est convexe. La preuve réciproque est similaire. □

Proposition 13

Soit S un ensemble géodésiquement convexe sur \mathcal{M} . Alors, une fonction $f : S \rightarrow \mathbb{R}$ est **géodésiquement (strictement) convexe** sur \mathcal{M} si et seulement si la fonction

$$g : \log(S) \rightarrow \mathbb{R}, \quad y \mapsto g(y) = f(\exp(y))$$

est (strictement) convexe dans \mathbb{R}^n .

Preuve: Par définition, une fonction f est **géodésiquement convexe** si et seulement si, pour tout $x, x' \in S$ et pour tout $t \in [0, 1]$, on a :

$$f(c(t)) \leq (1-t)f(x) + tf(x') = (1-t)g(y) + tg(y'),$$

où $x = \exp(y)$, $x' = \exp(y')$, et $c(t)$ est la géodésique définie de manière unique.

On conclut en observant que :

$$f(c(t)) = f(\exp(\log(x) + t(\log(x') - \log(x)))) = g((1-t)y + ty').$$

(L'argument est identique pour la **convexité géodésique stricte**.) □

2.2.2 Convexité et la variété des matrices symétriques définies positives

Considérons l'ensemble des matrices **symétriques définies positives** de taille n :

$$\text{Sym}(n)_+ = \{X \in \text{Sym}(n) : X > 0\}.$$

L'ensemble des matrices symétriques définies positives, noté $\text{Sym}(n)_+$, est un **ensemble convexe** dans l'espace euclidien $\text{Sym}(n)$ des matrices symétriques de taille n , muni du produit scalaire :

$$\langle U, V \rangle = \text{Tr}(U^\top V) = \text{Tr}(UV).$$

C'est une **sous-variété ouverte**, dont les espaces tangents s'identifient naturellement à $\text{Sym}(n)$.

Par analogie avec \mathbb{R}_+^n , nous souhaitons doter $\text{Sym}(n)_+$ d'une **structure riemannienne**, idéalement complète. Il existe au moins deux façons de procéder. Dans les deux cas, pour $n = 1$, on retrouve exactement la géométrie riemannienne construite précédemment pour \mathbb{R}_+^1 .

La métrique Log-Euclidienne

Une première approche consiste à construire un **difféomorphisme** entre $\text{Sym}(n)_+$ et une variété complète, de la même manière que la fonction logarithme fournit un difféomorphisme entre \mathbb{R}_+^n et \mathbb{R}^n .

On définit donc :

$$\varphi : \text{Sym}(n)_+ \rightarrow \text{Sym}(n), \quad \varphi(X) = \log(X), \quad (2.12)$$

où $\log(X)$ désigne le **logarithme matriciel principal**. Son inverse est l'exponentielle **matricielle** :

$$\varphi^{-1}(Y) = \exp(Y).$$

Les deux applications sont **lisses** sur leurs domaines respectifs, donc φ est bien un **difféomorphisme**.

En s'appuyant sur cette observation, on peut **transporter la métrique euclidienne** de $\text{Sym}(n)$ vers $\text{Sym}(n)_+$ (métrique par pullback), en définissant le produit scalaire suivant sur $T_X \text{Sym}(n)_+ = \text{Sym}(n)$:

$$\langle U, V \rangle_X^{\text{log}} \triangleq \langle D \log(X)[U], D \log(X)[V] \rangle. \quad (2.13)$$

C'est la **1. métrique Log-Euclidienne**, étudiée en détail par Arsigny et al. [6]. Pour les mêmes raisons que dans la section précédente, on peut décrire aisément ses

géodésiques et la convexité géodésique correspondante [6] : - La **géodésique unique (et minimisante)** reliant $X, X' \in \text{Sym}(n)_+$ pour la métrique Log-Euclidienne est :

$$c(t) = \exp(\log(X) + t(\log(X') - \log(X))). \quad (2.14)$$

- Un **ensemble** $S \subseteq \text{Sym}(n)_+$ est **géodésiquement convexe** pour cette métrique si et seulement si $\log(S)$ est convexe dans $\text{Sym}(n)$. - Pour un tel ensemble S , une **fonction** $f : S \rightarrow \mathbb{R}$ est **géodésiquement (strictement) convexe** si et seulement si $f \circ \exp$ est (strictement) convexe sur $\text{Sym}(n)$.

2. La métrique affine invariante

Une autre — et dans bien des cas plus courante — façon de définir une métrique sur $\text{Sym}(n)_+$ est la **métrique affine invariante**. Sur l'espace tangent $T_X \text{Sym}(n)_+$, elle est définie par :

$$\langle U, V \rangle_X^{\text{aff}} = \langle X^{-1/2} U X^{-1/2}, X^{-1/2} V X^{-1/2} \rangle = \text{Tr}(X^{-1} U X^{-1} V). \quad (2.15)$$

L'expression centrale garantit que les arguments du produit scalaire sont bien des **matrices symétriques**. La métrique en X coïncide avec le **hessien de la fonction barrière logarithmique**

$$X \mapsto -\log(\det(X)).$$

Cette métrique tire son nom de la propriété suivante : pour toute matrice inversible $M \in \mathbb{R}^{n \times n}$, la matrice $M X M^T$ est **définie positive**, et l'on a :

$$\langle M U M^T, M V M^T \rangle_{M X M^T}^{\text{aff}} = \langle U, V \rangle_X^{\text{aff}}. \quad (2.16)$$

Une conséquence concrète de la propriété d'invariance affine est que, si $c : [0, 1] \rightarrow \text{Sym}(n)_+$ est une courbe lisse, alors la **longueur** de c est égale à celle de la courbe $t \mapsto$

$M c(t) M^T$, car leurs vitesses sont identiques pour tout t . De même, la longueur de la courbe $t \mapsto c(t)^{-1}$ est la même que celle de c .

On peut montrer que la géodésique vérifiant $c(0) = X$ et $c'(0) = V$ est donnée par [12] :

$$\text{Exp}_X(tV) = c(t) = X^{1/2} \exp(t X^{-1/2} V X^{-1/2}) X^{1/2}. \quad (2.17)$$

Cette expression est bien définie pour tout t , ce qui montre que la variété est **complète**. De plus, cette variété est une **variété de Cartan-Hadamard**, ce qui la rend particulièrement adaptée aux applications d'**optimisation géodésiquement convexe**.

Pour s'assurer que $c(1) = X'$ (une autre matrice définie positive), on choisit V

comme suit :

$$\text{Log}_X(X') = X^{1/2} \log\left(X^{-1/2} X' X^{-1/2}\right) X^{1/2}. \quad (2.18)$$

Cette expression fournit la **vitesse initiale** en X de la **géodésique unique** reliant X et X' .

Il en découle que la **distance géodésique** est donnée par :

$$\text{dist}(X, X')^2 = \langle \text{Log}_X(X'), \text{Log}_X(X') \rangle_X^{\text{aff}} = \|\log(X^{-1/2} X' X^{-1/2})\|_F^2, \quad (2.19)$$

où $\|\cdot\|_F$ désigne la **norme de Frobenius**. Avec un certain soin, il est possible d'exprimer Exp, Log et dist sans faire intervenir de racines carrées matricielles, mais il est impossible d'éviter complètement les inverses, exponentielles et logarithmes de matrices.

Gradients et Hessiennes sur $\text{Sym}(n)_+$

Pour résoudre des problèmes d'optimisation sur $\text{Sym}(n)_+$, il est utile de pouvoir calculer les **gradients** et **hessiennes**.

Soit $\bar{f} : \text{Sym}(n) \rightarrow \mathbb{R}$ une fonction définie sur l'espace des matrices symétriques muni de la métrique euclidienne classique de $\mathbb{R}^{n \times n}$. On suppose que f est **lisse** sur l'ensemble ouvert $\text{Sym}(n)_+$, et que :

$$f = \bar{f}|_{\text{Sym}(n)_+}$$

désigne formellement la **restriction** de \bar{f} à la variété des matrices définies positives, équipée de la **métrique affine invariante**.

Les gradients et hessiennes de f et \bar{f} sont reliés par les formules suivantes, pour tout $(X, V) \in T\text{Sym}(n)_+$:

$$\text{grad } f(X) = X \text{ grad } \bar{f}(X) X, \quad (2.20)$$

$$\text{Hess } f(X)[V] = X \text{ Hess } \bar{f}(X)[V] X + \frac{1}{2} (V \text{ grad } \bar{f}(X) X + X \text{ grad } \bar{f}(X) V). \quad (2.21)$$

Si \bar{f} est définie sur tout $\mathbb{R}^{n \times n}$, il faut alors remplacer $\text{grad } \bar{f}(X)$ et $\text{Hess } \bar{f}(X)[V]$ par leurs **parties symétriques**. Ces formules sont issues de [12], où l'on trouve également les expressions de la **connexion riemannienne** et du **transport parallèle** sur $\text{Sym}(n)_+$.

Soit $\bar{f} : \text{Sym}(n) \rightarrow \mathbb{R}$ définie par :

$$\bar{f}(X) = \log(\det(X)),$$

et soit $f = \bar{f}|_{\text{Sym}(n)_+}$ sa restriction à l'ensemble des matrices définies positives, munies de la métrique affine invariante.

D'une part, on sait que :

$$\text{grad } \bar{f}(X) = X^{-1} \quad \text{et} \quad \text{Hess } \bar{f}(X)[V] = -X^{-1}VX^{-1}.$$

Il s'ensuit, d'après (2.20), que :

$$\text{grad } f(X) = X, \quad \text{Hess } f(X)[V] = 0,$$

pour tout $(X, V) \in \text{TSym}(n)_+$. En particulier, la Hessienne $\text{Hess } f(X)$ est à la fois **positive** et **négative**, ce qui signifie que f est **géodésiquement linéaire** sur $\text{Sym}(n)_+$.

Sur une variété de Cartan-Hadamard, pour tout point y , la fonction

$$f(x) = \frac{1}{2} \text{dist}(x, y)^2$$

est géodésiquement 1-fortement convexe sur toute la variété [12]. En particulier, toute boule géodésique centrée en y est géodésiquement convexe, car il s'agit d'un sous-niveau de f . Des informations plus détaillées sur la **hessienne de la distance** et de la **distance au carré** sur des variétés complètes à courbure bornée peuvent être trouvées dans [12].

2.3 Calculs statistiques dans les Variétés

Cette section propose une revue des statistiques de base pour les données sur les variétés riemanniennes, incluant des généralisations des concepts de moyenne, d'analyse en composantes principales (ACP) et de régression. Les définitions de ces statistiques dans l'espace euclidien reposent en partie sur les opérations d'espace vectoriel dans \mathbb{R}^d . Par exemple, la moyenne arithmétique est définie à l'aide de l'addition vectorielle et de la multiplication scalaire. La difficulté inhérente à la définition des statistiques sur les variétés riemanniennes générales est l'absence d'opérations d'espace vectoriel dans ces espaces.

Une approche pour analyser les données sur les variétés passe par la géométrie. Comme une variété riemannienne possède une métrique de distance, nous pouvons envisager l'ajustement de modèles comme un problème de moindres carrés, c'est-à-dire la minimisation de la somme des distances au carré entre les données et le modèle. Alors que les problèmes de moindres carrés dans l'espace euclidien ont souvent des solutions analytiques (par exemple, la régression linéaire), leur résolution sur les variétés riemanniennes nécessite généralement une optimisation itérative.

Pour l'ACP et la régression, une complication supplémentaire survient : les modèles sous-jacents dans \mathbb{R}^d sont définis comme des sous-espaces linéaires, qui

n'existent pas sur les variétés riemanniennes. Dans ces cas, les géodésiques offrent une généralisation naturelle des droites. Ainsi, la généralisation naturelle de la régression linéaire aux variétés est *la régression géodésique*, où une courbe géodésique est ajustée à des données associées à une variable explicative réelle. Dans le cas de l'ACP, la première composante principale peut être remplacée par une *géodésique principale* qui s'ajuste le mieux aux données en une dimension. Les composantes principales d'ordre supérieur sont définies comme des *sous-espaces géodésiques principaux*, générés comme l'image sous l'application exponentielle de sous-espaces linéaires d'un espace tangent.

Outre la perspective géométrique, une autre approche pour définir les statistiques sur les variétés passe par les probabilités. Dans les statistiques euclidiennes traditionnelles, l'ajustement par moindres carrés équivaut au maximum de vraisemblance.

L'estimation du maximum de vraisemblance sous l'hypothèse d'une distribution gaussienne des erreurs est également possible sur les variétés grâce à la définition d'une *loi normale riemannienne*. Nous montrons comment cette distribution fournit un cadre unificateur pour l'interprétation probabiliste de plusieurs modèles de données sur les variétés, incluant la moyenne de Fréchet, la régression géodésique et l'analyse géodésique principale.

Dans ce chapitre, soit $y_1, \dots, y_N \in M$ un ensemble de données sur une variété riemannienne. D'un point de vue statistique, nous considérerons ces données comme issues d'un échantillon aléatoire, c'est-à-dire des réalisations de N variables aléatoires indépendantes et identiquement distribuées (i.i.d.). Cependant, nous analyserons souvent les données d'un point de vue purement géométrique, sans référence explicite aux variables aléatoires.

2.3.1 La moyenne de Fréchet

Pour les données euclidiennes, la moyenne empirique est l'estimateur par défaut du centre d'un ensemble de données. C'est la statistique la plus simple à définir, mais aussi la plus importante. La moyenne empirique d'un ensemble de points $y_1, \dots, y_N \in \mathbb{R}^d$ est donnée par leur moyenne arithmétique :

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (2.22)$$

Cette définition de la moyenne dépend des opérations d'espace vectoriel de l'espace euclidien. En général, une variété riemannienne n'est pas un espace vectoriel, et cette définition n'est pas directement applicable. Pour des données sur une variété plongée dans l'espace euclidien, $\mathcal{M} \subset \mathbb{R}^d$, nous pourrions calculer la moyenne linéaire en utilisant les opérations vectorielles de l'espace ambiant \mathbb{R}^d . Cependant, le point moyen résultant pourrait ne pas appartenir à \mathcal{M} . Les deux exemples suivants illustrent comment la moyenne arithmétique de données sur une variété plongée peut échouer à rester sur la variété.

Exemple 32 ► Moyenne linéaire pour la sphère

La sphère 2D est naturellement plongée dans \mathbb{R}^3 comme l'ensemble des vecteurs unitaires, c'est-à-dire $S^2 \equiv \{y \in \mathbb{R}^3 : \|y\| = 1\}$. Pour des points $y_1, \dots, y_N \in S^2$, leur moyenne linéaire dans \mathbb{R}^3 , \bar{y} , n'appartiendra généralement pas à S^2 . Prenons par exemple $y_1 = (1, 0, 0)$ et $y_2 = (0, 1, 0)$. Leur moyenne $\bar{y} = (0.5, 0.5, 0)$ a pour norme $\|\bar{y}\| = \sqrt{2}/2$, et ne se trouve donc pas sur S^2 .

Exemple 33 ► Moyenne linéaire pour $GL(k)$

L'espace des matrices $k \times k$ à déterminant non nul forme un groupe de Lie appelé groupe linéaire général $GL(k)$.

c' est un sous-ensemble ouvert connexe de $M(k, k)$. Cependant, la moyenne calculée avec les opérations vectorielles usuelles de $M(k, k) \equiv \mathbb{R}^{k \times k}$ ne préserve pas la non-dégénérescence de $GL(k)$. Prenons par exemple les deux matrices :

$$y_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad y_2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

Les matrices y_1 et y_2 ont un déterminant égal à 1 et appartiennent donc à $GL(2)$. Leur moyenne linéaire $\bar{y} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ a un déterminant nul.

Caractérisations alternatives de la moyenne

La formule (2.22) n'étant pas définie pour les variétés riemanniennes générales, on peut se demander s'il existe des propriétés de la moyenne euclidienne généralisables aux variétés. En espace euclidien, la moyenne peut être dérivée de trois principes distincts :

1. **Algébrique** : La moyenne arithmétique est l'unique point où la somme des résidus est nulle :

$$(y_1 - \bar{y}) + \cdots + (y_N - \bar{y}) = 0 \quad (2.23)$$

Cette définition utilise exclusivement les propriétés d'espace vectoriel de \mathbb{R}^d .

2. **Géométrique** : C'est le *centroïde des moindres carrés* des données, minimisant la somme des distances au carré :

$$\bar{y} = \arg \min_{y \in \mathbb{R}^d} \sum_{i=1}^N \|y_i - y\|^2 \quad (2.24)$$

3. **Probabiliste** : Si les y_i sont des réalisations de variables aléatoires normales multivariées i.i.d. $Y_i \sim N(\mu, \Sigma)$, alors \bar{y} est l'estimateur du maximum de vraisemblance pour μ :

$$\bar{y} = \arg \max_{\mu \in \mathbb{R}^d} \prod_{i=1}^N p(y_i; \mu, \Sigma) \quad (2.25)$$

où $p(\cdot; \mu, \Sigma)$ est la densité de probabilité de la loi normale multivariée $N(\mu, \Sigma)$.

La caractérisation algébrique de la moyenne ne se généralise pas aux variétés riemanniennes, car elle dépend de la structure d'espace vectoriel. En revanche, les approches géométrique et probabiliste peuvent être adaptées. Cette section explore la caractérisation géométrique de la moyenne sur une variété riemannienne, concept introduit par Maurice Fréchet.

La généralisation de la moyenne empirique euclidienne aux variétés riemanniennes, introduite par Fréchet [18], est connue sous le nom de **moyenne de Fréchet**. Considérons un ensemble de données y_1, \dots, y_N sur une variété riemannienne M . La caractérisation géométrique s'écrit :

$$\bar{y} = \arg \min_{y \in M} \sum_{i=1}^N d(y, y_i)^2,$$

où d désigne la distance géodésique sur M . Fréchet a étendu ce concept à un cadre probabiliste général.

Existence et unicité

Théorème 31 ► Existence de la moyenne de Fréchet

Soit M un espace métrique complet. Alors la moyenne de Fréchet de tout ensemble fini de points $y_1, \dots, y_N \in M$ existe.

Définissons la fonction de coût :

$$F(y) = \sum_{j=1}^N d(y, y_j)^2 \quad (2.26)$$

La complétude de M garantit qu'une suite minimisante converge vers un point de M . Ce point réalise le minimum global (non nécessairement unique).

Preuve de théorème

Soit $r = \max_{i,j} d(y_i, y_j)$. Définissons $K = \cup_{i=1}^N \bar{B}_r(y_i)$, où $\bar{B}_r(y_i)$ désigne la boule métrique fermée de rayon r centrée en y_i . Par la complétude de X , K est un ensemble fermé, de diamètre borné par $2r$, donc compact. La restriction de la fonction de somme des carrés des distances $F(y) = \sum_{j=1}^N d(y, y_j)^2$ à K atteint un minimum dans K . Pour un point $y \in X \setminus K$, on a $F(y) > Nr^2$, ce qui dépasse la valeur minimale dans K puisque $F(y_i) \leq Nr^2$ pour tout y_i . \square

Exemple 34 ► Non-unicité de la moyenne de Fréchet

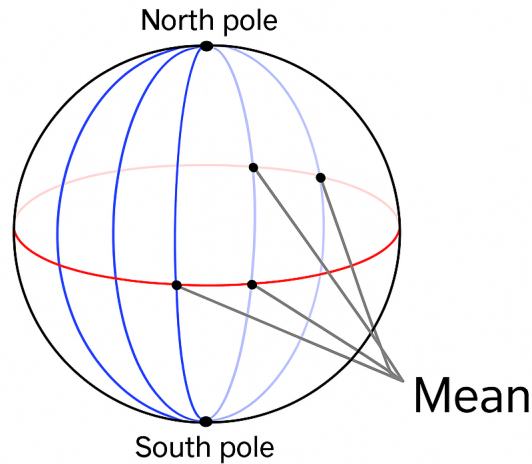
La moyenne de Fréchet peut ne pas être unique. Un exemple classique est sur S^2 :

Pour les points de données $y_1 = (0, 0, 1)$ (pôle nord) et $y_2 = (0, 0, -1)$ (pôle sud) sur S^2 , la moyenne de Fréchet est l'équateur entier :

$$\bar{y} = \{(\cos \theta, \sin \theta, 0) \mid \theta \in [0, 2\pi)\}. \quad (2.27)$$

Exemple 35 ► Non-existence de la moyenne de Fréchet

Le plan \mathbb{R}^2 privé de l'origine $(0, 0)$ constitue une variété non complète nommé **Plan épointé**. Pour des points symétriques comme $y_1 = (1, 0)$ et $y_2 = (-1, 0)$, la moyenne de Fréchet n'existe pas.



Moyenne de Fréchet sur variété lisse S^2

FIGURE 2.1 – Moyenne de Fréchet sur la variété lisse S^2

Théorème 32 ► Unicité de la moyenne de Fréchet

Soit M une variété riemannienne complète de courbure sectionnelle $\leq \Delta$, et $\text{inj}(M)$ son rayon d'injectivité. Si les données $\{y_i\}_{i=1}^N$ sont contenues dans une boule géodésique de rayon :

$$r = \begin{cases} \frac{1}{2} \min \left\{ \text{inj}(M), \frac{\pi}{\sqrt{\Delta}} \right\} & \text{si } \Delta > 0, \\ \frac{1}{2} \text{inj}(M) & \text{si } \Delta \leq 0, \end{cases} \quad (2.28)$$

alors la moyenne de Fréchet \bar{y} est unique.

Exemple 36 ► Variétés à courbure constante en 2D

- $\Delta = 0$: **Plan euclidien** \mathbb{R}^2 Le rayon d'injectivité est infini ; courbure sectionnelle 0. Tous les ensembles de données ont une moyenne de Fréchet unique.
- $\Delta = 1$: **Sphère** S^2 Rayon d'injectivité π ; courbure sectionnelle 1. Les données dans un hémisphère ouvert ($r = \frac{\pi}{2}$) ont une moyenne unique.
- $\Delta = -1$: **Plan hyperbolique** \mathbb{H}^2 Le rayon d'injectivité est infini ; courbure sectionnelle -1 . Ici, $r = \frac{1}{2} \cdot \infty = \infty$. Ainsi, *tout* ensemble de données dans \mathbb{H}^2 a une moyenne de Fréchet unique, en raison de l'absence de points conjugués et de l'effet « d'étalement » de la courbure négative.

Par définition, la moyenne de Fréchet doit être un point critique de la fonction de somme des distances au carré, vérifiant :

$$\sum_{i=1}^n \text{Log}_{\bar{x}}(x_i) = 0. \quad (2.29)$$

Une stratégie simple pour estimer cette moyenne consiste à utiliser un *algorithme itératif de point fixe* jusqu'à satisfaction de (2.29):

À chaque itération t , une mise à jour est effectuée.

La mise à jour le long d'une géodésique s'écrit :

$$x^{t+1} = \text{Exp}_{x^t} \left(\gamma \sum_{i=1}^n \text{Log}_{x^t}(x_i) \right),$$

où γ est un *pas d'apprentissage*. Les méthodes d'ordre supérieur utilisent des pas adaptatifs, des approximations de la hessienne ou des informations sur la courbure pour améliorer les performances.

Généralisations statistiques sur les variétés

La moyenne de Fréchet \bar{x} est essentielle pour définir de nombreux outils statistiques, comme la matrice de covariance empirique :

$$\Sigma = \frac{1}{n} \sum_{i=1}^n \text{Log}_{\bar{x}} x_i \otimes \text{Log}_{\bar{x}} x_i. \quad (2.30)$$

Cette approche permet de paramétrer des densités à partir de leur moyenne et covariance, comme dans Chevallier et Guigui 2020. Ce modèle offre un schéma simple d'estimation de densité.

On peut remarquer que l'application logarithme riemannien dans l'espace euclidien s'écrit $\text{Log}_y x = (x - y)$. La covariance de y devient alors :

$$\text{Cov}(y) = E \left[\left(\text{Log}_{E[y]} y \right) \left(\text{Log}_{E[y]} y \right)^T \right] \quad (2.31)$$

Cette équation se généralise à une variété riemannienne en remplaçant l'espérance euclidienne $E[y]$ par l'espérance de Fréchet. Pour un échantillon $y_1, \dots, y_n \in M$, la matrice de covariance empirique est donnée par :

$$S = \sum_{i=1}^n \left(\text{Log}_{\bar{y}} y_i \right) \left(\text{Log}_{\bar{y}} y_i \right)^T, \quad (2.32)$$

2.3.2 Généralisation de l'Analyse en composantes principales (ACP) dans les variétés

La matrice de covariance quantifie la variabilité des données multidimensionnelles, mais son interprétation directe reste difficile. Une décomposition plus utile est proposée par l'analyse en composantes principales (ACP), introduite par Pearson, Hotelling. (Voir l'ouvrage [55]) pour une revue détaillée. Les objectifs de l'ACP sont :

1. Paramétriser efficacement la variabilité des données,
2. Réduire la dimensionnalité des paramètres.

Dans cette section, nous rappelons l'ACP pour des données euclidiennes $y_1, \dots, y_N \in \mathbb{R}^d$ de moyenne \bar{y} , avant de généraliser aux variétés.

Plusieurs approches équivalentes décrivent l'ACP. La formulation choisie ici facilite la généralisation aux variétés riemanniennes. L'objectif est de trouver une suite de sous-espaces linéaires emboîtés V_1, \dots, V_d passant par la moyenne, qui approximent au mieux les données. Une méthode consiste à minimiser la somme des carrés des résidus :

$$v_k = \arg \min_{\|v\|=1} \sum_{i=1}^N \|y_i^k - \langle y_i^k, v \rangle v\|^2, \quad (2.33)$$

où $V_k = \text{span}(v_1, \dots, v_k)$ et les v_i sont orthonormaux.

Les termes y_i^k sont définis de manière récursive par :

$$y_i^1 = y_i - \bar{y},$$

$$y_i^k = y_i^{k-1} - \langle y_i^{k-1}, v_{k-1} \rangle v_{k-1}.$$

Autrement dit, y_i^k s'obtient en retirant de $(y_i - \bar{y})$ les contributions des directions précédentes v_1, \dots, v_{k-1} . Ainsi, y_i^k représente la projection de $(y_i - \bar{y})$ sur le sous-espace orthogonal à V_{k-1} .

Formulation duale de l'ACP : Une seconde définition de l'ACP consiste à maximiser la variance totale des données projetées. La variance totale pour un ensemble de points y_1, \dots, y_N est donnée par :

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N \|y_i - \bar{y}\|^2. \quad (2.34)$$

Les sous-espaces $V_k = \text{span}(v_1, \dots, v_k)$ sont alors déterminés par les vecteurs :

$$v_k = \arg \max_{\|v\|=1} \sum_{i=1}^N \langle y_i^k, v \rangle^2, \quad (2.35)$$

où les y_i^k sont définis comme précédemment. Les deux approches (minimisation des résidus et maximisation de la variance) sont équivalentes grâce au théorème de Pythagore [55].

Calcul pratique des composantes principales

La procédure se décompose comme suit :

1. Calculer la moyenne empirique :

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

2. Estimer la matrice de covariance (version sans biais) :

$$S = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})(y_i - \bar{y})^\top. \quad (2.36)$$

Propriétés de la matrice de covariance La matrice S est :

- Symétrique : $S = S^\top$,
- Semi-définie positive : $x^\top S x \geq 0 \quad \forall x \in \mathbb{R}^d$.

Ses valeurs propres $\lambda_1, \dots, \lambda_d$ sont réelles et non négatives, ordonnées par $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$. Les vecteurs propres associés v_1, \dots, v_d définissent les directions principales. En cas de multiplicités des valeurs propres, les vecteurs propres correspondants ne sont pas uniques.

Généralisations de l'ACP

L'Analyse en Composantes Principales (ACP) vise à trouver des sous-espaces maximisant la variance des projections. En géométrie riemannienne, cette équivalence n'est plus valide (le théorème de Pythagore ne s'applique pas).

L'*Analyse en Composantes Principales Géodésiques (PGA)* minimise l'erreur de reconstruction en projetant les données sur une sous-variété géodésique (Fletcher, Lu et al., 2004). La première composante est une géodésique issue de la moyenne. La projection d'un point x sur une géodésique γ de vitesse initiale v en \bar{x} est définie par :

$$\pi_{\bar{x},v}(x) = \arg \min_{t \in \mathbb{R}} d^2(x, \text{Exp}_{\bar{x}}(tv)).$$

Sous certaines conditions (données regroupées dans un voisinage convexe), cette projection existe. Une variante proche est l'ACP géodésique (Huckemann, Hotz et al.).

La minimisation de l'erreur de reconstruction globale s'effectue par descente de

gradient, similairement au calcul de la moyenne de Fréchet, où le gradient de l'exponentielle est calculé par différentiation automatique.

Étant donné un jeu de données x_1, \dots, x_n , on cherche la vitesse initiale v^* d'une géodésique minimisant :

$$v^* = \arg \min_{v \in T_{\bar{x}} M} \sum_{i=1}^n d^2(x_i, \pi_{\bar{x}, v}(x_i)). \quad (2.37)$$

Le calcul du gradient de cette fonction objectif est complexe car il implique n problèmes de minimisation. Cependant, Ablin, Peyre et al. 2020 montrent que le gradient peut être calculé efficacement via la différentiation automatique de l'approximation par descente de gradient. Ceci est illustré dans l'exemple de la sphère, où les applications Exp et Log admettent des solutions explicites. Des défis persistent lorsque le calcul du Log est numériquement coûteux.

2.3.3 Distributions de probabilité

Étant donné une mesure de probabilité, on peut définir des variables aléatoires à valeurs dans une variété M comme suit.

Définition 37

Soit $(\Omega, \mathcal{B}(\Omega), \Pr)$ un espace probabilisé, où $\mathcal{B}(\Omega)$ est la σ -algèbre de Borel sur Ω et \Pr une mesure de masse unitaire. Une *variable aléatoire* sur la variété riemannienne M est une fonction borélienne $X : \Omega \rightarrow M$.

Une métrique riemannienne offre un cadre naturel pour définir des distributions de probabilité sur les variétés, car elle induit une mesure de volume.

Définition 38 ► Forme volume riemannienne

Une variété (pseudo)-riemannienne orientée (M, g) possède une *forme volume* naturelle :

$$d\text{Vol}(x) = \sqrt{\det(g(x))} dx.$$

Cette forme volume permet de définir des intégrales sur M , c'est-à-dire une mesure. Ainsi, certaines distributions de probabilité peuvent être exprimées via des densités par rapport à cette mesure.

Définition 39

La variable aléatoire X admet une *densité* f si pour tout $\mathcal{X} \in \mathcal{B}(M)$:

$$\Pr(X \in \mathcal{X}) = \int_{\mathcal{X}} f(y) dv(y) \quad \text{et} \quad \Pr(M) = 1.$$

Exemple 37 ► Loi uniforme

Soit M une variété riemannienne compacte, par exemple l'hypersphère, le groupe spécial orthogonal ou la variété de Grassmann. Une loi uniforme sur M a pour densité :

$$f(y) = \frac{1}{\text{Vol}(M)},$$

où $\text{Vol}(M)$ désigne le volume total de M .

Exemple 38 ► Distribution gaussienne

Soit M un espace symétrique riemannien. Une distribution gaussienne de paramètres moyenne-précision (μ, Γ) est définie par la densité :

$$f(x) = \alpha(\Gamma, \mu) \exp\left(-\log_{\mu}(x)^{\top} \Gamma \log_{\mu}(x)\right).$$

C'est la distribution maximisant l'entropie (Pennec, 2006). Dans le cas isotrope, à courbure non-positive, complet et simplement connexe, l'estimateur du maximum de vraisemblance coïncide avec la moyenne de Fréchet [55].

2.3.4 Echantillonnage

Une tâche courante en statistique consiste à générer des échantillons à partir d'une distribution de probabilité donnée sur une variété. Cela peut être motivé par des problèmes d'inférence où les paramètres des distributions a posteriori sont contraints, par des tests d'adéquation pour les familles exponentielles, ou pour générer des données synthétiques afin de valider des algorithmes d'apprentissage [55].

Cependant, cela ne se réduit pas à un échantillonnage classique même lorsqu'une paramétrisation de la variété est disponible. La courbure de l'espace déforme généralement les densités, comme illustré ci-dessous.

Exemple 39 ► Loi uniforme sur la sphère

Considérons la sphère unité $S^2 \subset \mathbb{R}^3$, paramétrée en coordonnées sphériques depuis le pôle nord $e_0 = (0, 0, 1)$:

$$x = \sin(\phi) \cos(\theta), \quad y = \sin(\phi) \sin(\theta), \quad z = \cos(\phi).$$

L'élément de volume (ici l'aire) est :

$$d \text{Vol}(\theta, \phi) = \sin(\phi) d\theta d\phi.$$

Une tentative naïve d'échantillonnage uniforme consisterait à tirer θ uniformément dans $[0, 2\pi)$ et ϕ uniformément dans $[0, \pi)$.

Cependant, près des pôles ($\phi = 0$ et $\phi = \pi$), le facteur $\sin(\phi)$ réduit la densité par rapport à la mesure euclidienne. On observe alors un regroupement non uniforme des points autour des pôles.

En réalité, si Y est un vecteur gaussien dans \mathbb{R}^d de moyenne 0 et de covariance identité, alors $X = \frac{Y}{\|Y\|}$ suit une loi uniforme sur S^d .

Exemple 40 ► Groupe spécial orthogonal

Soit $SO(n)$ le groupe des matrices orthogonales de taille $n \times n$ à déterminant unitaire. Soit Y une matrice $n \times n$ dont les coefficients suivent indépendamment une loi normale standard. La décomposition QR de Y donne $Y = XR$, où $X \in SO(n)$ et R est triangulaire supérieure. Alors X est uniformément distribué sur $SO(n)$ par rapport à la mesure de Haar, qui coïncide avec la mesure riemannienne [55].

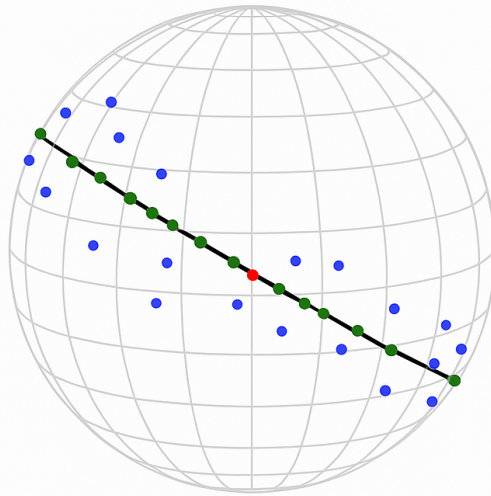
Exemple 41 ► Variété de Stiefel

La variété de Stiefel est l'ensemble des k -repères orthonormés dans \mathbb{R}^n , représenté par les matrices $\{U \in \mathbb{R}^{n \times k} \mid U^\top U = I_k\}$. Soit Z une matrice à coefficients i.i.d. suivant une loi normale standard. Alors $X = Z(Z^\top Z)^{-1/2}$ est uniformément distribué [55]. Notons que X est un facteur de la décomposition polaire de Z , représentant l'*orientation* de Z .

Exemple 42 ► Variété de Grassmann

La variété de Grassmann est l'ensemble des sous-espaces de dimension k de \mathbb{R}^n , représentée par les matrices de projection $n \times n$ symétriques de rang k vérifiant $P^2 = P$. Soit Z une matrice à coefficients i.i.d. suivant une loi normale standard. Alors $X = Z(Z^\top Z)^{-1/2}Z$ est uniformément distribué [55].

2.3.5 Régression géodésique



Régression géodésique sur la variété lisse S^2

La régression linéaire peut être généralisée en *régression géodésique* sur les variétés riemanniennes en résolvant un problème de moindres carrés. Étant donné des points cibles $y_1, \dots, y_n \in M$ et des données $t_1, \dots, t_n \in \mathbb{R}$, on cherche la géodésique qui approxime le mieux les données :

$$\min_{(p,v) \in TM} \sum_{i=1}^n d^2(\text{Exp}_p(t_i v), y_i), \quad (2.38)$$

où d et Exp sont respectivement la distance riemannienne et l'application exponentielle. Dans le cas euclidien, ce problème coïncide avec la régression linéaire classique. Cependant, il n'existe pas de solution explicite en général, et le problème doit être résolu par optimisation.

Pour différencier la fonction objectif de (2.38), il faut composer les gradients

de la distance au carré avec ceux de l'application exponentielle. Le gradient de la distance au carré est proportionnel au logarithme riemannien, comme mentionné précédemment. Le gradient de l'application Exp est généralement calculé via les champs de Jacobi.

Pour résoudre le problème d'optimisation, on utilise souvent une descente de gradient riemannienne.

2.4 Calculs des moyennes de Fréchet dans des variétés matricielles

2.4.1 Moyennes des Matrices de Covariance

Étant symétriques et définies positives, les matrices de covariance sont situées dans un espace courbe qui est muni d'une géométrie particulière dite Riemannienne. La prise en compte de cette géométrie est nécessaire afin d'adapter les outils classiques de la géométrie Euclidienne aux descripteurs de covariance.

Définitions et Notations

Soient les espaces matriciels suivants :

1. $S(n) = \{S \in M(n), S^\top = S\}$: espace des matrices symétriques $n \times n$
2. $P(n) = \{P \in S(n), \mathbf{u}^\top P \mathbf{u} > 0, \forall \mathbf{u} \in \mathbb{R}^n\}$: ensemble des matrices symétriques définies positives
3. $GL(n)$: groupe des matrices inversibles $n \times n$

Pour une matrice $P \in P(n)$, sa décomposition en valeurs propres s'écrit :

$$P = U \text{diag}(\sigma_1, \dots, \sigma_n) U^\top$$

avec $\sigma_1 > \sigma_2 > \dots > \sigma_n > 0$ (valeurs propres) et U (vecteurs propres). On définit alors :

$$\text{Exponentielle matricielle : } \exp(P) = U \text{diag}(e^{\sigma_1}, \dots, e^{\sigma_n}) U^\top$$

$$\text{Logarithme matriciel : } \log(P) = U \text{diag}(\ln \sigma_1, \dots, \ln \sigma_n) U^\top$$

$$\text{Norme de Frobenius : } \|A\|_F^2 = \text{Tr}(AA^\top) = \sum_{i,j} |A_{i,j}|^2$$

Variété Riemannienne, plan tangent et métrique Riemannienne

Une *variété Riemannienne* M est un espace topologique localement assimilable à un espace Euclidien, muni d'une métrique g . L'espace $P(n)$ forme une telle variété.

En chaque point $P \in M$ (une matrice de covariance), le *plan tangent* $T_P M$ est un espace Euclidien inclus dans $S(n)$. Ses éléments sont des vecteurs tangents représentant les dérivées directionnelles des courbes lisses passant par P .

La métrique Riemannienne naturelle sur $P(n)$ est donnée par :

$$g_P(X, Y) = \text{Tr}(P^{-1}XP^{-1}Y), \quad \forall X, Y \in T_P M$$

Cette structure permet de généraliser les algorithmes de classification (e.g., K-moyennes) à la géométrie des matrices de covariance.

La métrique Riemannienne de la variété M associe à chaque paire de vecteurs tangents $S_1, S_2 \in \mathbb{S}(n)$ un produit scalaire $\langle \cdot, \cdot \rangle_p$. Pour $S_1, S_2 \in T_p M$, ce produit scalaire s'exprime par :

$$\langle S_1, S_2 \rangle_p = \text{Tr}(S_1 P^{-1} S_2 P^{-1}) \quad (2.39)$$

La norme induite sur $T_p M$ est donnée par :

$$\|S\|_P^2 = \langle S, S \rangle_p = \text{Tr}(S P^{-1} S P^{-1}).$$

On remarque que cette norme coïncide avec la norme de Frobenius lorsque $P = I$:

$$\langle S, S \rangle_I = \|S\|_F^2.$$

Propriété d'invariance : Pour toute matrice inversible $W \in GL(n)$,

$$\langle W S_1 W^\top, W S_2 W^\top \rangle_{W P W^\top} = \langle S_1, S_2 \rangle_p \quad (2.40)$$

Géodésique, application exponentiel et application logarithmique

Une géodésique Γ reliant P_1 et P_2 sur M est une courbe lisse minimisant la longueur fonctionnelle :

$$L(\alpha) = \int_0^1 \left\| \frac{d\alpha(t)}{dt} \right\|_{\alpha(t)} dt,$$

où $\alpha(t)$ est paramétrisée par $t \in [0, 1]$ avec $\alpha(0) = P_1$ et $\alpha(1) = P_2$. La géodésique s'écrit :

$$\Gamma = \arg \min_{\alpha} L(\alpha),$$

et est définie explicitement par :

$$\Gamma : [0, 1] \rightarrow \mathbb{P}(n), \quad t \mapsto \Gamma(t).$$

La distance Riemannienne entre deux points P_1 et P_2 est donnée par la longueur de la géodésique :

$$\delta_R(P_1, P_2) = L(\Gamma) = \int_0^1 \left\| \frac{d\Gamma(t)}{dt} \right\|_{\Gamma(t)} dt = \int_0^1 \sqrt{\text{Tr} \left(\frac{d\Gamma(t)}{dt} \Gamma^{-1}(t) \frac{d\Gamma(t)}{dt} \Gamma^{-1}(t) \right)} dt \quad (2.41)$$

Cette longueur est invariante par transformation congruente :

$$L(\Gamma) = L(W\Gamma W^\top) \quad \forall W \in GL(n).$$

Le *application exponentiel* en un point $P \in M$ associe un vecteur tangent $S_i \in T_P M$ à un point $P_t \in M$:

$$\text{Exp}_P(S_i) = P_t = P^{1/2} \exp \left(P^{-1/2} S_i P^{-1/2} \right) P^{1/2} \quad (2.42)$$

L'opération inverse, le *application logarithmique*, s'écrit :

$$\text{Log}_P(P_t) = S_t = P^{1/2} \log \left(P^{-1/2} P_t P^{-1/2} \right) P^{1/2} \quad (2.43)$$

Ces applications permettent de naviguer entre la variété M et son plan tangent $T_P M$,

Distance géodésique

Distance Riemannienne La distance Riemannienne δ_R entre deux matrices P_1 et P_2 sur M s'obtient via l'intégrale de l'équation (2.41). La géodésique $\Gamma(t)$ reliant P_1 et P_2 s'écrit :

$$\Gamma(t) = P_1^{1/2} \exp \left(t P_1^{-1/2} S_2 P_1^{-1/2} \right) P_1^{1/2} \quad (2.44)$$

où $S_2 \in T_{P_1} M$ est le vecteur tangent donné par le mapping logarithmique :

$$S_2 = \log_{P_1}(P_2) = P_1^{1/2} \log \left(P_1^{-1/2} P_2 P_1^{-1/2} \right) P_1^{1/2} \quad (2.45)$$

En substituant (2.45) dans (2.44), la géodésique devient :

$$\Gamma(t) = P_1^{1/2} \left(P_1^{-1/2} P_2 P_1^{-1/2} \right)^t P_1^{1/2} \quad (2.46)$$

La distance Riemannienne finale s'exprime alors par :

$$\delta_R(P_1, P_2) = \left\| \log(P_1^{-1}P_2) \right\|_F = \left[\sum_{i=1}^n \log^2 \lambda_i \right]^{1/2} \quad (2.47)$$

où $\{\lambda_i\}$ sont les valeurs propres de $P_1^{-1}P_2$.

Distance Log-Euclidienne

La distance Log-Euclidienne δ_{LE} est définie par :

$$\delta_{LE}(P_1, P_2) = \left\| \log(P_1) - \log(P_2) \right\|_F = \delta_E(\log(P_1), \log(P_2)) \quad (2.48)$$

où δ_E est la distance Euclidienne :

$$\delta_E(P_1, P_2) = \left\| P_1 - P_2 \right\|_F = \sqrt{\text{Tr}((P_1 - P_2)(P_1 - P_2)^\top)}.$$

Elle vérifie la relation :

$$\delta_R(P_1, P_2) = \delta_{LE} \left(I_n, P_1^{-\frac{1}{2}} P_2 P_1^{-\frac{1}{2}} \right) \quad (2.49)$$

où I_n est la matrice identité $n \times n$.

Propriétés additionnelles de δ_{LE} :

— **Invariance par multiplication à gauche/droite** : $\forall P_1, P_2, P_3 \in \mathbb{P}(n)$,

$$\delta_{LE}(P_3 P_1, P_3 P_2) = \delta_{LE}(P_1 P_3, P_2 P_3) = \delta_{LE}(P_1, P_2);$$

— **Invariance par inversion** : $\delta_{LE}(P_1^{-1}, P_2^{-1}) = \delta_{LE}(P_1, P_2)$.

Moyenne de Fréchet dans l'espace des matrices de covariance

Moyenne Riemannienne

La moyenne Riemannienne J_R , ou barycentre Riemannien, est définie comme le point $\in \mathbb{P}(n)$ minimisant :

$$J_R = \arg \min_{P \in \mathbb{P}(n)} \sum_{i=1}^N \delta_R^2(P, P_i),$$

où $\{P_i\}_{i=1}^N$ sont les matrices de covariance. Cette formulation généralise la moyenne arithmétique à la géométrie Riemannienne (Pennec 2006).

La moyenne Riemannienne J_R est définie comme solution du problème d'optimi-

sation :

$$J_R = \arg \min_{P \in \mathbb{P}(n)} \sum_{i=1}^I \delta_R^2(P, A_i) \quad (2.50)$$

où $\{A_i\}_{i=1}^I$ sont des matrices de covariance dans $\mathbb{P}(n)$, et δ_R est la distance Riemannienne.

Cas particuliers :

— **Matrices(1×1) :** Pour $\{x_i > 0\}_{i=1}^I$, la moyenne Riemannienne devient :

$$J_R = \exp \left(\frac{1}{I} \sum_{i=1}^I \log x_i \right) \quad (2.51)$$

— **Deux matrices($n \times n$) :** La moyenne géométrique binaire $A_1 \# A_2$ s'écrit :

$$A_1 \# A_2 = A_1^{1/2} \left(A_1^{-1/2} A_2 A_1^{-1/2} \right)^{1/2} A_1^{1/2} \quad (2.52)$$

2.4.2 Calcul de moyenne de Fréchet dans la variété de Stiefel

La variété de Stiefel $\text{St}(n, p)$ est l'ensemble des matrices $X \in \mathbb{R}^{n \times p}$ dont les colonnes sont orthonormées :

$$\text{St}(n, p) = \{X \in \mathbb{R}^{n \times p} \mid X^\top X = I_p\}.$$

Elle représente les bases orthonormées des sous-espaces vectoriels de dimension p dans \mathbb{R}^n . Chaque point X possède un plan tangent défini par :

$$T_X \text{St}(n, p) = \{\xi \in \mathbb{R}^{n \times p} \mid X^\top \xi + \xi^\top X = 0\}.$$

La projection orthogonale d'une matrice Z sur ce plan tangent est [11] :

$$P_X(Z) = Z - X \cdot \text{sym}(X^\top Z), \quad \text{sym}(A) = \frac{1}{2}(A + A^\top).$$

Définition 40 ► barycentre

Le barycentre (ou moyenne de Fréchet) d'un ensemble $\{X_k\}_{k=1}^N \subset \text{St}(n, p)$ est défini comme :

$$G = \arg \min_{X \in \text{St}(n, p)} \frac{1}{N} \sum_{k=1}^N d^2(X, X_k),$$

où $d(\cdot, \cdot)$ désigne la distance géodésique riemannienne.

Dans $\text{St}(n, p)$, la carte logarithme $\log_X(Y)$ n'a pas d'expression analytique simple. On introduit donc la notion de rétraction pour approcher la carte exponentielle.

Rétractions et barycentre

Définition 41 ► Rétraction

Une rétraction est une application $R_X : T_X \text{St}(n, p) \rightarrow \text{St}(n, p)$ vérifiant :

$$R_X(0) = X, \quad \left. \frac{d}{dt} R_X(t\xi) \right|_{t=0} = \xi.$$

Autrement dit, $R_X(\xi) = X + \xi + o(\|\xi\|)$: elle approxime localement l'exponentielle riemannienne.

On définit le **R-barycentre** par l'équation de point fixe :

$$G^{(i+1)} = R_{G^{(i)}} \left(\frac{1}{N} \sum_{k=1}^N R_{G^{(i)}}^{-1}(X_k) \right).$$

Trois rétractions principales sont couramment utilisées [11] sur $\text{St}(n, p)$: QR, polaire (**UF**) et orthographique.

Méthode 1 : Rétraction polaire (UF)

$$R_X^{\text{pol}}(\xi) = (X + \xi) \left[(X + \xi)^T (X + \xi) \right]^{-1/2}.$$

Cette opération correspond à la normalisation orthogonale des colonnes de $X + \xi$, identique au calcul du facteur orthogonal U dans la décomposition polaire :

$$\text{uf}(A) = A(A^T A)^{-1/2}.$$

Barycentre associé

Le barycentre est obtenu analytiquement par :

$$X^* = \text{uf} \left(\frac{1}{N} \sum_{k=1}^N X_k \right).$$

Méthode 2 : Rétraction QR

$$R_X^{\text{QR}}(\xi) = \text{qf}(X + \xi),$$

où $\text{qf}(A)$ désigne le facteur Q de la décomposition QR de $A = QR$.

Interprétation

On orthonormalise simplement les colonnes de $X + \xi$, ce qui ramène le point sur la variété. C'est une rétraction rapide, mais moins fidèle pour les grands déplacements.

Méthode 3 : Rétraction orthographique

$$R_X^{\text{orth}}(\xi) = (X + \xi) \left(I_p + \frac{1}{2} X^T \xi \right)^{-1}.$$

Rétraction inverse:

$$(R_X^{\text{orth}})^{-1}(Q) = Q - \frac{1}{2}X(Q^\top X + X^\top Q).$$

Propriétés algorithmiques et remarques

La quantité X^* obtenue par cette procédure est parfois désignée dans la littérature sous les termes de *moyenne de Fréchet extrinsèque projetée* ou de *R-barycentre associé à une rétraction*. Elle ne coïncide *pas en général* avec la *moyenne de Fréchet intrinsèque* définie par :

$$G = \arg \min_{X \in \text{St}(n,p)} \frac{1}{N} \sum_{k=1}^N d^2(X, X_k),$$

où $d(\cdot, \cdot)$ désigne la distance géodésique riemannienne sur $\text{St}(n, p)$. Cette dernière nécessite en effet la résolution d'un problème d'optimisation non linéaire sur la variété, sans qu'il existe d'expression analytique fermée en toute généralité [11].

Néanmoins, la construction présentée constitue une *approximation légitime* du barycentre intrinsèque lorsque les données sont suffisamment proches les unes des autres. Dans le cas général, elle fournit une solution approchée efficace.

Exemples de calcul du barycentre dans la variété de stiefel

Exemple 43 ► calcul du barycentre de trois matrices dans $St(3, 2)$

En considérant trois matrices orthonormées correspondant à des rotations dans le plan (x, y) d'angles 0° , 15° et -20° .

Chaque matrice est de la forme :

$$X(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \\ 0 & 0 \end{pmatrix}.$$

On prend :

$$\theta_1 = 0^\circ, \quad \theta_2 = 15^\circ, \quad \theta_3 = -20^\circ, \quad \text{et donc} \quad X_i = X(\theta_i).$$

$$X_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X_2 \approx \begin{pmatrix} 0.965926 & 0.258819 \\ -0.258819 & 0.965926 \\ 0 & 0 \end{pmatrix}, \quad X_3 \approx \begin{pmatrix} 0.939693 & -0.342020 \\ 0.342020 & 0.939693 \\ 0 & 0 \end{pmatrix}.$$

Chaque X_i vérifie $X_i^T X_i = I_2$, donc $X_i \in St(3, 2)$.

Moyenne extrinsèque

On calcule la moyenne arithmétique (dans $\mathbb{R}^{3 \times 2}$) :

$$M = \frac{1}{3}(X_1 + X_2 + X_3) = \begin{pmatrix} 0.968540 & -0.027734 \\ 0.027734 & 0.968540 \\ 0 & 0 \end{pmatrix}.$$

Cette matrice ne vérifie pas exactement la contrainte $M^T M = I_2$:

$$M^T M \approx 0.938838 I_2.$$

On doit donc la ramener sur la variété $St(3, 2)$ par une rétraction R_X , selon la méthode choisie.

Méthode 1 : rétraction QR

On effectue la décomposition QR de M :

$$M = QR, \quad \text{où} \quad Q^T Q = I_2.$$

Le facteur Q fournit le barycentre :

$$X_{\text{QR}}^* = Q \approx \begin{pmatrix} 0.999590 & -0.028623 \\ 0.028623 & 0.999590 \\ 0 & 0 \end{pmatrix}.$$

Exemple 44 ► Suite

Méthode 2 : rétraction polaire (UF)

La rétraction polaire est donnée par :

$$R_X^{\text{pol}}(\xi) = (X + \xi)[(X + \xi)^T(X + \xi)]^{-1/2}.$$

En posant $A = M$, on obtient :

$$X_{\text{pol}}^* = M(M^T M)^{-1/2}.$$

Calculons :

$$(M^T M)^{-1/2} = \frac{1}{\sqrt{0.938838}} I_2 \approx 1.0326 I_2,$$

d'où :

$$X_{\text{pol}}^* = 1.0326 M = \begin{pmatrix} 0.999590 & -0.028623 \\ 0.028623 & 0.999590 \\ 0 & 0 \end{pmatrix}, \quad X_{\text{pol}}^* X_{\text{pol}}^* = I_2.$$

Méthode 3 : rétraction orthographique

La rétraction orthographique est définie par :

$$R_X^{\text{orth}}(\xi) = (X + \xi) \left(I + \frac{1}{2} X^T \xi \right)^{-1}.$$

Dans le cas du barycentre, elle conduit à la même expression que la rétraction polaire, via le facteur orthogonal de la décomposition polaire :

$$X_{\text{orth}}^* = \text{uf}(M) = X_{\text{pol}}^*.$$

Ainsi :

$$X_{\text{orth}}^* = \begin{pmatrix} 0.999590 & -0.028623 \\ 0.028623 & 0.999590 \\ 0 & 0 \end{pmatrix}.$$

Vérification de l'orthogonalité

On vérifie :

$$X_{\text{pol}}^{*\top} X_{\text{pol}}^* = I_2, \quad X_{\text{QR}}^{*\top} X_{\text{QR}}^* = I_2.$$

Les trois méthodes produisent donc des matrices orthonormées appartenant à $St(3, 2)$.

Exemple 45 ► suite

Interprétation géométrique

Les matrices X_1, X_2, X_3 représentent trois bases orthonormées définissant des sous-espaces proches (rotations d'environ $\pm 20^\circ$ autour de l'origine). Leur barycentre correspond à une rotation moyenne :

$$\theta^* \approx 1.64^\circ.$$

C'est donc le « centre géodésique » des trois bases sur la variété de Stiefel.

Résumé des résultats

Méthode	Formule du barycentre	Résultat numérique
QR	Q de la décomposition QR de M	$\begin{pmatrix} 0.99959 & -0.02862 \\ 0.02862 & 0.99959 \\ 0 & 0 \end{pmatrix}$
Polaire (UF)	$M(M^\top M)^{-1/2}$	$\begin{pmatrix} 0.99959 & -0.02862 \\ 0.02862 & 0.99959 \\ 0 & 0 \end{pmatrix}$
Orthographique	$\text{uf}(M)$	$\begin{pmatrix} 0.99959 & -0.02862 \\ 0.02862 & 0.99959 \\ 0 & 0 \end{pmatrix}$

Dans ce cas simple, les trois méthodes (QR, polaire, orthographique) conduisent à un barycentre identique au premier ordre, car les matrices initiales sont proches entre elles. La rétraction polaire offre toutefois la meilleure fidélité géométrique, tandis que la rétraction orthographique fournit une formule analytique directe :

$$X^* = \text{uf}\left(\frac{1}{3}(X_1 + X_2 + X_3)\right).$$

2.4.3 Calcul de moyenne de Fréchet dans la variété de Grassmann

La variété de Grassmann $\text{Gr}(p, n)$ représente l'ensemble de tous les sous-espaces vectoriels de dimension p de \mathbb{R}^n :

$$\text{Gr}(p, n) = \{\text{span}(X) \mid X \in \mathbb{R}^{n \times p}, X^\top X = I_p\}.$$

Elle peut être vue comme le quotient de la variété de Stiefel :

$$\text{Gr}(p, n) = \text{St}(p, n)/O(p),$$

ce qui signifie que deux matrices X et Y de $\text{St}(p, n)$ représentent le même point de Grassmann si elles engendrent le même sous-espace, i.e. $Y = XQ$ avec $Q \in O(p)$.

L'espace tangent en un point $X \in \text{St}(p, n)$, vu comme représentant de $[X] \in \text{Gr}(p, n)$, est donné par :

$$T_X \text{Gr}(p, n) = \{\xi \in \mathbb{R}^{n \times p} \mid X^\top \xi = 0\}.$$

La projection orthogonale d'une matrice $Z \in \mathbb{R}^{n \times p}$ sur cet espace tangent est :

$$P_X(Z) = (I - XX^\top)Z.$$

La distance géodésique entre deux sous-espaces $[X]$ et $[Y]$ dépend des angles principaux θ_i entre les colonnes de X et Y :

$$d([X], [Y]) = \|\Theta\|_2 = \sqrt{\sum_{i=1}^p \theta_i^2},$$

où $\cos(\theta_i)$ sont les valeurs singulières de $X^\top Y$.

Définition du barycentre sur $Gr(p, n)$

Le barycentre riemannien (ou moyenne de Fréchet) d'un ensemble de sous-espaces $\{[X_k]\}_{k=1}^N$ est défini par :

$$G = \arg \min_{[X] \in Gr(p, n)} \frac{1}{N} \sum_{k=1}^N d^2([X], [X_k]).$$

Ce problème n'admet pas toujours de solution explicite, mais plusieurs approximations pratiques existent. La plus utilisée est la **moyenne extrinsèque par projection orthogonale**.

Méthodologie pratique : moyenne de projection (extrinsèque)

Chaque sous-espace $[X_k]$ est représenté par sa **matrice de projection** :

$$P_k = X_k X_k^\top.$$

Ces matrices appartiennent à l'espace euclidien symétrique des matrices $n \times n$, satisfaisant :

$$P_k = P_k^\top, \quad P_k^2 = P_k, \quad \text{tr}(P_k) = p.$$

Étapes du calcul du barycentre :

1. **Représentation des sous-espaces** : calculer les projecteurs $P_k = X_k X_k^\top$.

2. **Moyenne euclidienne** :

$$\bar{P} = \frac{1}{N} \sum_{k=1}^N P_k.$$

3. **Projection sur la variété $Gr(p, n)$** : on diagonalise $\bar{P} = U \Lambda U^\top$, puis on prend les p premiers vecteurs propres :

$$X^* = U_{[:,1:p]}.$$

4. Le sous-espace moyen (barycentre) est alors :

$$G = \text{span}(U_{[:,1:p]}).$$

Conclusion

Le calcul du barycentre sur la variété de Grassmann repose donc sur la moyenne des projecteurs :

$$X^* = \text{vecteurs propres de plus grande énergie de } \bar{P} = \frac{1}{N} \sum_k X_k X_k^\top.$$

Cette approche conserve la structure géométrique de la variété tout en offrant une expression explicite et efficace.

Exemple 46 ► calcul du barycentre de deux droites dans $Gr(1,3)$:
 \mathbb{R}^3

On considère deux sous-espaces $[x_1]$ et $[x_2]$ définis par :

$$x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} \cos(30^\circ) \\ \sin(30^\circ) \\ 0 \end{pmatrix}.$$

Étape 1 — Projecteurs :

$$P_1 = x_1 x_1^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad P_2 = x_2 x_2^\top = \begin{pmatrix} 0.75 & 0.433 & 0 \\ 0.433 & 0.25 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Étape 2 — Moyenne euclidienne :

$$\bar{P} = \frac{P_1 + P_2}{2} = \begin{pmatrix} 0.875 & 0.2165 & 0 \\ 0.2165 & 0.125 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Étape 3 — Diagonalisation de \bar{P} : On trouve la plus grande valeur propre $\lambda_1 \approx 0.895$, et le vecteur propre associé :

$$v_1 = \begin{pmatrix} \cos(15^\circ) \\ \sin(15^\circ) \\ 0 \end{pmatrix}.$$

Étape 4 — Barycentre :

$$[x^*] = \text{span}(v_1), \quad \text{soit une droite d'angle moyen } \theta^* = 15^\circ.$$

Le barycentre correspond à la **direction médiane** entre les deux sous-espaces, c'est-à-dire à la moyenne géodésique sur la variété $Gr(1,3)$.

Exemple 47 ► calcul du barycentre de deux plans dans $Gr(2, 3)$

La variété $Gr(2, 3)$ représente l'ensemble des plans vectoriels passant par l'origine dans \mathbb{R}^3 . Chaque plan peut être représenté par une matrice $X \in \mathbb{R}^{3 \times 2}$ de colonnes orthonormées ($X^T X = I_2$). On considère deux plans $[X_1]$ et $[X_2]$, chacun défini par une base orthonormée. L'objectif est de calculer leur barycentre sur $Gr(2, 3)$.

Définition des deux plans

On définit :

$$X_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} \cos(30^\circ) & 0 \\ 0 & 1 \\ \sin(30^\circ) & 0 \end{pmatrix}.$$

Le premier plan $[X_1]$ est le plan (x, y) , le second $[X_2]$ est un plan incliné d'un angle 30° autour de l'axe y .

Étape 1 — Matrices de projection

Chaque plan est représenté par sa matrice de projection :

$$P_k = X_k X_k^T.$$

Ainsi :

$$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0.75 & 0 & 0.433 \\ 0 & 1 & 0 \\ 0.433 & 0 & 0.25 \end{pmatrix}.$$

Étape 2 — Moyenne euclidienne des projecteurs

On calcule :

$$\bar{P} = \frac{1}{2}(P_1 + P_2) = \begin{pmatrix} 0.875 & 0 & 0.2165 \\ 0 & 1 & 0 \\ 0.2165 & 0 & 0.125 \end{pmatrix}.$$

Cette matrice \bar{P} est symétrique mais n'est pas idempotente ($\bar{P}^2 \neq \bar{P}$), donc elle n'appartient pas exactement à la variété.

Exemple 48 ► suite de l'exemple

Étape 3 — Projection sur la variété $Gr(2, 3)$

On diagonalise \bar{P} :

$$\bar{P} = U\Lambda U^T,$$

puis on prend les deux vecteurs propres associés aux plus grandes valeurs propres :

$$X^* = U_{[:,1:2]}.$$

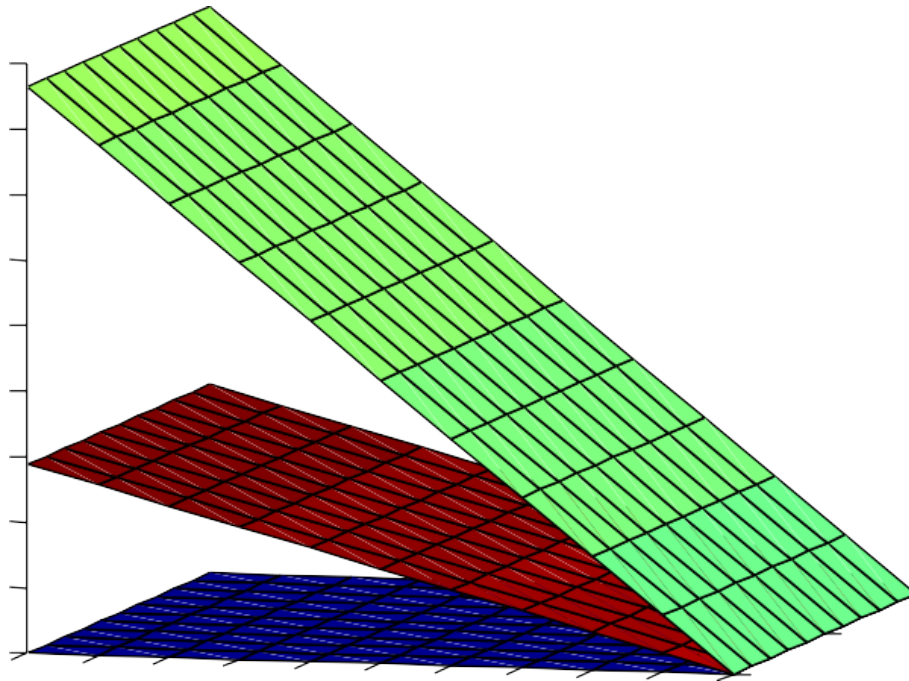


FIGURE 2.2 – plan rouge : plan moyen des deux autres plans vert et bleu

2.5 barycentres et convexité dans les variétés différentielles

Cette section est consacrée aux contributions fondamentales de M. Emery, Mokobodzki et M. Arnaudon à la théorie des barycentres convexes dans les variétés, ainsi qu'à l'étude des liens entre ces constructions et les courbure, en particulier celles permettant d'obtenir des estimations précises de la taille des barycentres. À la suite d'échanges scientifiques particulièrement enrichissants avec M. Emery (Strasbourg) et M. Arnaudon (Poitiers), nous avons entrepris d'élaborer un cadre visant à étendre et approfondir ces résultats, soit en proposant de nouvelles applications, soit en relâchant certaines hypothèses structurelles qui interviennent dans les formulations classiques des théorèmes.

Notre objectif est d'examiner l'applicabilité de ces méthodes à des variétés explicitement structurées, telles que les variétés matricielles, où l'on peut effectuer des calculs géométriques et analytiques détaillés. Dans cette perspective, et afin d'alléger la présentation, nous nous limitons au rappel des notions et résultats essentiels issus des travaux précités, tels qu'ils apparaissent dans les articles [4],[31], dont notre approche s'inspire directement.

Dans une variété munie d'une connexion et telle que deux points quelconques sont joints par une géodésique et une seule, on peut définir le barycentre de toute probabilité μ portée par au plus deux points : si $\mu = (1 - t)\varepsilon_x + t\varepsilon_y$, son barycentre est bien sûr $\gamma(t)$ où γ est la géodésique telle que $\gamma(0) = x$ et $\gamma(1) = y$.

On travaille dans une variété connexe V réelle, sans bord, C^∞ , sans torsion de dimension finie ; nous la supposons munie d'une connexion C^∞ .

2.5.1 Définitions des barycentres et Exemples

Définition 42 ► barycentre convexe

Soit μ une probabilité sur V . Le barycentre de μ est l'ensemble $b(\mu)$ des points x de V tels que $f(x) \leq \mu(f)$ pour toute fonction f convexe bornée sur V .

Puisque les fonctions convexes sont continues, $b(\mu)$ est fermé. Il possède aussi une propriété de convexité faible : son intersection avec toute géodésique est un intervalle (fermé) de celle-ci. Dualement, pour x fixé, l'ensemble des lois μ telles que $x \in b(\mu)$ est convexe et étroitement fermé.

Exemple 49

Si V est un ouvert convexe borné d'un espace affine (muni de la connexion plate), alors le barycentre $b(\mu)$ est égal à $\{b_a(\mu)\}$, où $b_a(\mu)$ désigne le barycentre de μ pour la structure affine. En effet, pour f affine, l'inégalité s'applique à f et à $-f$ et est donc une égalité, d'où $b(\mu) \subset \{b_a(\mu)\}$; l'inclusion inverse, qui n'est autre que l'inégalité de Jensen, vient de ce que toute fonction convexe est un supremum de fonctions affines. En revanche, si V est un ouvert borné non convexe d'un espace affine, on a encore $b(\mu) \subset \{b_a(\mu)\}$, et que $b(\mu)$ peut être vide même si $b_a(\mu) \in V$ et V est connexe.

Exemple 50

Si V est compacte, ou si V est un espace vectoriel muni de la connexion plate, $b(\mu) = V$ quelle que soit μ car toutes les fonctions convexes bornées sont constantes.

Exemple 51

Si V est la demi-droite $]0, \infty[$ munie de la connexion plate, $b(\mu)$ est l'intervalle $[g, \infty[$, où $g = \int x \mu(dx) \leq \infty$ est le barycentre de μ au sens usuel.

Exemple 52

Si μ est portée par deux points, donc de la forme $(1-t)\varepsilon_x + t\varepsilon_y$, et si γ est une géodésique telle que $\gamma(0) = x$ et $\gamma(1) = y$, le point $\gamma(t)$ est dans $b(\mu)$.

Proposition 14 ► Associativité des barycentres

Soient (A, \mathcal{A}, π) un espace probablisé, $(\mu_\alpha)_{\alpha \in A}$ une famille de probabilités sur V , dépendant mesurablement de α et $(x_\alpha)_{\alpha \in A}$ une v. a. dans V telle que $x_\alpha \in b(\mu_\alpha)$ pour chaque α . Alors

$$b\left(\int \varepsilon_{x_\alpha} \pi(d\alpha)\right) \subset b\left(\int \mu_\alpha \pi(d\alpha)\right).$$

Preuve : Si f est convexe, bornée et si $x \in b(\int \varepsilon_{x_\alpha} \pi(d\alpha))$,

$$f(x) \leq \int f(x_\alpha) \pi(d\alpha) \leq \int \mu_\alpha(f) \pi(d\alpha) = \int f d\mu.$$

□

Corollaire 1

langage clair : tout point du barycentre d'une famille de barycentres de $\mu = \int \mu_\alpha \pi(d\alpha)$ est lui-même dans le barycentre de μ .

La réciproque de cette propriété d'associativité est fausse.

Exemple 53 ► contre exemple

En géométrie sphérique, si A', B' et C' sont les milieux des côtés d'un triangle ABC , le milieu M du segment géodésique AA' n'est pas sur la géodésique $B'C'$. On a donc $B' \in b\left(\frac{1}{2}\varepsilon_A + \frac{1}{2}\varepsilon_C\right)$ et $C' \in b\left(\frac{1}{2}\varepsilon_A + \frac{1}{2}\varepsilon_B\right)$, mais M , bien que dans le barycentre de la mesure $\frac{1}{2}\varepsilon_A + \frac{1}{4}\varepsilon_B + \frac{1}{4}\varepsilon_C$, n'est cependant pas dans celui de $\frac{1}{2}\varepsilon_{B'} + \frac{1}{2}\varepsilon_{C'}$.

Définition 43 ► Barycentre comme moyenne conditionnelle

En termes probabilistes : Si une v. a. X , définie sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$, prend ses valeurs dans V , on appellera espérance de X , et on notera $\mathbb{E}[X]$, le barycentre de la loi de X , c'est-à-dire l'ensemble de tous les points x de V tels que $f(x) \leq \mathbb{E}[f(X)]$ pour toute f convexe bornée. Si de plus \mathcal{G} est une sous-tribu de \mathcal{F} , on appellera espérance conditionnelle de X sachant \mathcal{G} , et on notera $\mathbb{E}[X|\mathcal{G}]$, l'ensemble de toutes les v. a. Y mesurables pour \mathcal{G} et à valeurs dans V telles que $f(Y) \leq \mathbb{E}[f(X)|\mathcal{G}]$ pour toute f convexe bornée.

2.5.2 les barycentres géodesiques dans les variétés

La construction des barycentres géodesiques consiste à remplacer deux points par leur barycentre pris sur la géodésique qui les joint, procédure que l'on itère en remontant un arbre dyadique depuis les extrémités des branches vers la racine ; ceci permet de définir le barycentre de 2^n points de V , pas nécessairement tous distincts, affectés de poids dont la somme est égale à 1, et pris dans un certain ordre.

Définition 44 ► Barycentre géodésique

l'espace canonique $W^n = \{0, 1\}^n$, sur lequel le processus des coordonnées sera noté $(\omega_i)_{1 \leq i \leq n}$; si p est une probabilité sur W^n , pour $\epsilon \in \{0, 1\}$ on notera p_ϵ la probabilité sur W^{n-1} égale à la loi conditionnelle de $(\omega_2, \dots, \omega_n)$ sachant que $\omega_1 = \epsilon$; si y est une v. a. définie sur W^n , on définit des v. a. y_0 et y_1 sur W^{n-1} par $y_\epsilon(\epsilon_1, \dots, \epsilon_{n-1}) = y(\epsilon, \epsilon_1, \dots, \epsilon_{n-1})$. Le barycentre géodésique itéré $\beta(y, p)$ (ou $\beta_n(y, p)$ si l'on veut préciser le nombre d'itérations) par

- Pour $n = 0$, $\beta_0(y, 1)$ est le singleton $\{x\}$, où $x \in V$ est la valeur de la v.a. constante y .

Pour $n > 0$, $\beta_n(y, p)$ est l'ensemble des points de V de la forme $\gamma(p(\omega_1 = 1))$, où $\gamma : [0, 1] \rightarrow V$ est une géodésique telle que :

- $\gamma(0) \in \beta_{n-1}(y_0, p_0)$
- $\gamma(1) \in \beta_{n-1}(y_1, p_1)$

Pour tout n , $\beta_n(y, p) \subset b(y(p))$, où $y(p) = p \circ y^{-1}$.

Par récurrence :

- Pour $n = 0$: $\beta_0(y, 1) = \{x\}$ avec $y \equiv x$, donc $y(p) = \varepsilon_x$ et $x \in b(\varepsilon_x)$.
- Pour $x \in \beta_n(y, p)$, on a $x = \gamma(p(\omega_1 = 1))$ avec :

$$f(x) \leq p(\omega_1 = 0)E_0[f \circ y_0] + p(\omega_1 = 1)E_1[f \circ y_1] = E[f \circ y] = y(p)(f)$$

donc $x \in b(y(p))$.

Remarque 10

Dans les bons cas, l'ensemble $b(\mu)$ contient un point remarquable, qui s'appellera barycentre exponentiel de μ , que l'on peut définir à l'aide d'une carte normale, ou de l'application exponentielle; c'est une nouvelle appellation de la moyenne de Karcher.

Dans toute la suite, nous supposons V convexe au sens très fort suivant : deux points quelconques sont joints par une géodésique et une seule, qui dépend de façon C^∞ des deux points. Pour $x \in V$, si l'on désigne par $T_x V$ l'espace vectoriel tangent en x à V et par U_x l'ouvert étoilé de $T_x V$ formé des vitesses initiales $\dot{\gamma}(0)$ des géodésiques γ telles que $\gamma(1)$ soit défini, l'application exponentielle $\exp_x : U_x \rightarrow V$ qui envoie $\dot{\gamma}(0)$ sur $\gamma(1)$ est un difféomorphisme de U_x sur V .

Définition 45 ► Barycentres exponentiels

Si μ est une probabilité sur V , on appelle *barycentres exponentiels* de μ les points x de V tels que la probabilité $\tilde{\mu}_x$ sur $T_x V$, image de μ par \exp_x^{-1} , soit centrée dans l'espace vectoriel $T_x V$ (c'est-à-dire que toutes les formes linéaires sur $T_x V$ sont intégrables pour $\tilde{\mu}_x$ et d'intégrale nulle).

En d'autres termes, x est un barycentre exponentiel de μ si et seulement si l'intégrale vectorielle $\int_{z \in T_x V} z \tilde{\mu}_x(dz)$ existe dans $T_x V$ et est nulle ; lorsqu'elle existe, nous noterons $v^\mu(x)$ cette intégrale.

Remarque 11

Le barycentre de la mesure μ est, en général, est un sous-ensemble de V , tandis que les barycentres exponentiels sont des éléments de V . Cette différence de nature justifie la distinction de notation et sera précisée dans la proposition suivante. Par ailleurs, si V est contenu dans un domaine géodésiquement convexe suffisamment petit, le barycentre exponentiel est unique. Un résultat de ce type est établi dans le théorème de Kendall, qui garantit l'unicité de la moyenne de Fréchet sous des hypothèses sur le rayon d'injectivité de la variété (théorème 32).

Proposition 15 ► Barycentre exponentiel est un barycentre convexe

Soit μ une probabilité sur V . Tout barycentre exponentiel de μ est dans $b(\mu)$.

Preuve: Soient x un barycentre exponentiel de μ et f une fonction convexe bornée. Par le corollaire 1 de [30], il existe une fonction linéaire h sur $T_x V$ telle que, pour tout $y \in V$, on ait

$$f(y) - f(x) \geq h \exp_x^{-1}(y).$$

Intégrons par rapport à μ :

$$\mu(f) - f(x) \geq \int h \exp_x^{-1}(y) \mu(dy).$$

Puisque par hypothèse x est un barycentre exponentiel de μ , l'intégrale est nulle, d'où la proposition. \square

Proposition 16

On suppose V riemannienne (munie de la connexion canonique) ; soit μ une probabilité telle que, pour un point x de V , l'intégrale $f(x) = \int_V \text{dist}(x, y)^2 \mu(dy)$ soit finie. Les barycentres exponentiels de μ sont alors les points critiques de la fonction f .

Preuve : Soit ϕ une fonction C^∞ sur V . Sa différentielle $d\phi(x)$ est une forme linéaire sur $T_x V$. Sur un voisinage U de x_0 , la majoration

$$\begin{aligned} |d\phi(x) \circ \exp_x^{-1}(y)| &\leq \|d\phi(x)\| \|\exp_x^{-1}(y)\| \\ &= \|d\phi(x)\| \text{dist}(x, y) \leq \sup_U \|d\phi\| (\text{diam } U + \text{dist}(x_0, y)) \end{aligned}$$

établit l'existence, et, par convergence dominée, la continuité par rapport à x de l'intégrale

$$\mu(d\phi(x) \circ \exp_x^{-1}) = v^\mu(x)\phi;$$

donc le champ de vecteurs v^μ existe et est continu .

Pour y fixé, le gradient au point x de la fonction $\text{dist}(x, y)^2$ est $-2 \exp_x^{-1}(y)$; en intégrant en y , on trouve

$$\text{grad } f(x) = -2 \int_{y \in V} \exp_x^{-1}(y) \mu(dy) = -2v^\mu(x)$$

Ainsi, f est de classe C^1 , de gradient $-2v^\mu$ et ses points critiques sont ceux où v^μ s'annule, d'où la proposition. \square

Les critères d'existence et d'unicité du barycentre exponentiel, empruntés à Kendall [50].

Proposition 17 ► Existence du barycentre exponentiel

Toute probabilité sur V portée par un ensemble relativement compact de la forme $\{\phi < 0\}$, où ϕ est une fonction convexe de classe C^2 , admet (au moins) un barycentre exponentiel.

Preuve : Choisissons sur V une structure riemannienne arbitraire (indépendamment de la connexion). Soit f une fonction C^∞ . Pour x et y dans le compact $K = \{\phi \leq 0\}$, la majoration

$$|df \circ \exp_x^{-1}(y)| = |\langle df(x), \exp_x^{-1}(y) \rangle| \leq \|df(x)\| \|\exp_x^{-1}(y)\| \leq c(K)$$

montre, après intégration en y , que la fonction $v^\mu f(x) = \mu(df \circ \exp_x^{-1})$ est bien définie et continue sur K .

Pour $x \in \partial K$ et $y \in K$, on a $\phi(x) = 0$ et $\phi(y) \leq 0$, donc ϕ est négative sur l'arc géodésique γ joignant x à y ; en différentiant au point x , on obtient

$$\langle \exp_x^{-1}(y), d\phi(x) \rangle = \left. \frac{d}{dt} \right|_{t=0} (\phi \circ \gamma)(t) \leq 0,$$

et en intégrant en y , on trouve $v^\mu \phi(x) \leq 0$, ce qui montre que le champ de vecteurs

continu v^μ est rentrant au bord de K .

Il existe un difféomorphisme transformant K et K' en une boule de \mathbb{R}^d et son intérieur (passer par \exp_x^{-1}), où z est intérieur à K , puis par un difféomorphisme entre U_z et \mathbb{R}^d). Donc le champ de vecteurs v^μ s'annule en au moins un point de K (et même de K' , en raison de la convexité de ϕ). \square

Proposition 18 ► Unicité du barycentre exponentiel

S'il existe sur $V \times V$ une fonction convexe, bornée, nulle sur la diagonale Δ et strictement positive hors de Δ , toute probabilité sur V a au plus un barycentre exponentiel.

La connexion dont est munie $V \times V$ dans cet énoncé est bien entendu la connexion produit, caractérisée (à la torsion près, mais peu importe) par ses géodésiques, qui sont les courbes $(\gamma_1(t), \gamma_2(t))$ où chaque γ_i est une géodésique de V ; la sous-variété Δ de $V \times V$ est totalement géodésique pour cette connexion (cette condition est évidemment nécessaire à l'existence de la fonction convexe sus-mentionnée).

Lemme 3 ► Barycentre exponentiel pour les mesures produits

Si μ est une probabilité sur $V \times V$ ayant pour marges ν et π , et si x (respectivement y) est un barycentre exponentiel de ν (respectivement π) dans V , alors (x, y) est un barycentre exponentiel de μ dans $V \times V$.

Preuve de lemme : Soient ξ et η dans V ; il leur correspond des vecteurs $X \in T_x V$ et $Y \in T_y V$ tels que $\exp_x X = \xi$ et $\exp_y Y = \eta$. La courbe dans $V \times V$ définie par

$$\gamma(t) = (\exp_x tX, \exp_y tY)$$

est une géodésique telle que $\gamma(1) = (\xi, \eta)$ et $\dot{\gamma}(0) = (X, Y)$; donc

$$\exp_{(x,y)}^{-1}(\xi, \eta) = (X, Y).$$

Toute forme linéaire h sur $T_{(x,y)}(V \times V)$ vérifie $h(X, Y) = f(X) + g(Y)$ pour des formes linéaires f et g sur $T_x V$ et $T_y V$ respectivement. On en tire l'égalité

$$h \exp_{(x,y)}^{-1}(\xi, \eta) = h(X, Y) = f(X) + g(Y) = f \exp_x^{-1}(\xi) + g \exp_y^{-1}(\eta),$$

qui, puisque ν et π sont les marges de μ , s'intègre en

$$\int_{V \times V} h \exp_{(x,y)}^{-1} d\mu = \int_V f \exp_x^{-1} d\nu + \int_V g \exp_y^{-1} d\pi.$$

Ceci est nul car x et y sont des barycentres exponentiels de ν et π . \square

Démonstration de l'unicité: Soient x et y deux barycentres exponentiels de la même probabilité μ sur V , et soit $\tilde{\mu}$ la probabilité sur Δ dont les deux projections sur V sont égales à μ . Le lemme qui précède entraîne que (x, y) est un barycentre exponentiel de $\tilde{\mu}$, donc, par la proposition 2, un point du barycentre de $\tilde{\mu}$ dans $V \times V$. On a donc

$$\phi(x, y) \leq \tilde{\mu}(\phi),$$

où ϕ est une fonction convexe bornée sur $V \times V$ nulle sur Δ et positive ailleurs (une telle fonction existe par hypothèse). Puisque $\tilde{\mu}$ est portée par Δ , ceci implique $\phi(x, y) \leq 0$, donc $x = y$. □

On distingue principalement deux types de barycentres : le barycentre exponentiel (ou de Karcher) et le barycentre convexe. Le barycentre exponentiel d'une mesure de probabilité μ sur une variété riemannienne M est défini comme le point $e \in M$ qui vérifie l'équation

$$\int \log_e(y) d\mu(y) = 0,$$

où \log_e désigne l'application logarithmique au point e . Cette définition généralise la condition classique du barycentre euclidien où la somme des vecteurs position est nulle. Et pour estimer la taille de barycentre convexe, on introduit la notion de géométrie convexe :

Définition 46 ►

Une variété munie d'une connexion W' est dite à **géométrie convexe** s'il existe une fonction convexe de classe C^3 , $\psi : W' \times W' \rightarrow \mathbb{R}$, positive et s'annulant exactement sur la diagonale, telle que pour tout $a \in W'$, la fonction $\psi_a = \psi(a, \cdot)$ ait un Hessian strictement positif $\nabla d\psi_a$ sur $W' \setminus \{a\}$.

Un sous-ensemble compact V' d'une variété à connexion est dit à géométrie convexe s'il possède un voisinage ouvert à géométrie convexe.

Dans la suite, nous supposons que V à géométrie convexe. Le résultat suivant a été démontré dans [4],

Proposition 19 ► Estimation de rayon de barycentre convexe

Sous les conditions ci-dessus, pour toute distance riemannienne δ sur V , il existe une constante C telle que pour toute mesure de probabilité μ sur V et tout $b \in b(\mu)$, l'inégalité suivante est vérifiée :

$$|b(\mu)| \leq C \int_V \delta^3(b, y) \mu(dy),$$

où $|b(\mu)|$ désigne le diamètre de $b(\mu)$ par rapport à la distance δ .

Conclusion du chapitre

Dans ce chapitre, nous avons établi le cadre géométrique et statistique nécessaire pour généraliser les méthodes classiques d'analyse euclidienne vers les variétés riemanniennes, en mettant l'accent sur les notions de *convexité*, de *barycentres* et de *méthodes d'optimisation géométrique*.

Nous avons d'abord établi la résolution de problèmes statistiques sur une variété nécessite la présence de la *convexité géodésique* et ces propriétés pour garantir l'existence et l'unicité des minima.

La *moyenne de Fréchet*, définie comme minimiseur de la somme des distances géodésiques au carré, constitue une première généralisation naturelle de la moyenne euclidienne. Toutefois, son calcul explicite souvent est difficile selon la géométrie de la variété. Pour pallier ces limitations, la *moyenne de Karcher*, définie comme un point critique de l'énergie quadratique, est introduite comme *barycentre exponentiel*, c'est-à-dire comme un point où l'intégrale du logarithme riemannien s'annule.

Nous avons ensuite présenté une généralisation plus large : le *barycentre convexe* au sens d'Émery et Mokobodzki. Cette approche, indépendante de la métrique mais fondée sur les fonctions convexes et la connexion, permet de définir un barycentre même lorsque la distance géodésique ou l'application exponentielle ne fournissent pas une structure suffisamment stable. Nous avons rappelé que tout barycentre exponentiel est un barycentre convexe, sans que la réciproque ne soit vraie, faisant de ce dernier la notion la plus abstraite et la plus générale.

D'un point de vue appliqué, nous avons étudié plusieurs variétés matricielles essentielles en traitement du signal, automatique, électronique et économétrie : la variété des matrices symétriques définies positives (SPD), la variété de Stiefel et la variété de Grassmann. Pour chacune d'elles, nous avons décrit la géométrie (métriques, espaces tangents, géodésiques), puis présenté les stratégies numériques pour calculer les barycentres : moyenne riemannienne dans les variétés SPD, barycentres sur Stiefel via des rétractions (QR, polaire, orthographique), et moyenne des projecteurs pour Grassmann.

Ces moyennes généralisées constituent les éléments fondamentaux de nombreuses méthodes statistiques sur les variétés, telles que la *covariance empirique riemannienne*, l'*analyse en composantes principales géodésique (PGA)* ou la *régression géodésique*. Elles permettent notamment la réduction de dimension et l'extraction de directions principales directement dans l'espace intrinsèque, sans recourir à des approximations linéaires.

Ainsi, la combinaison de la convexité, des géodésiques, des barycentres (Fréchet, Karcher, convexes), des applications exponentielle et logarithmique ainsi que de l'optimisation riemannienne forme un cadre cohérent et puissant pour l'analyse statistique de données situées dans des espaces non-euclidiens. Ce chapitre pose les fondations théoriques et algorithmiques nécessaires aux développements ultérieurs, notamment en matière de réduction de dimension, d'analyse géométrique des données et de modélisation intrinsèque sur les variétés.

Chapitre 3

Calculs explicites des barycentres dans les variétés et applications

3.1 Calculs des barycentres géodésiques dans les variétés à courbures constantes

3.1.1 Trigonométrie Sphérique et Hyperbolique

Notre travail vise à établir une estimation systématique et propriétés des barycentres géodésiques dans les géométries sphérique et hyperbolique. Nous examinerons notamment les développements asymptotiques lorsque les côtés des triangles tendent vers zéro, les relations entre les différents types de barycentres, et l'influence de la courbure sur les propriétés géométriques.

Les géométries sphérique et hyperbolique représentent les deux types fondamentaux de géométries à courbure constante non-nulle. La sphère de dimension 2, notée S^2 , est l'exemple canonique d'espace à courbure positive constante, tandis que le plan hyperbolique H^2 illustre le cas de courbure négative constante. Ces deux géométries partagent avec l'espace euclidien la propriété d'isotropie et d'homogénéité, mais diffèrent radicalement dans leurs propriétés métriques.

La métrique sphérique en coordonnées polaires s'écrit $ds^2 = dr^2 + \sin^2(r) d\theta^2$, où $r \in [0, \pi]$ et $\theta \in [0, 2\pi]$. Cette expression révèle que la circonférence d'un cercle de rayon r sur la sphère est $2\pi \sin(r)$, croissant plus lentement qu'en géométrie euclidienne. À l'inverse, la métrique hyperbolique dans le modèle du disque de Poincaré s'écrit $ds^2 = 4(dx^2 + dy^2)/(1 - x^2 - y^2)^2$, conduisant à une croissance exponentielle des circonférences.

Les formules trigonométriques fondamentales diffèrent significativement selon la géométrie considérée. Pour un triangle sphérique de côtés a, b, c et d'angles opposés α, β, γ , la loi des cosinus sphérique s'écrit :

$$\cos c = \cos a \cos b + \sin a \sin b \cos \gamma.$$

Cette formule généralise la loi euclidienne et incorpore les effets de courbure positive.

Dans le cas hyperbolique, la loi des cosinus devient :

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos \gamma.$$

Ces différences structurelles se répercutent directement sur les propriétés des barycentres géodésiques.

3.1.2 Barycentres géodésiques des triangles

Triangles Sphériques

Considérons un triangle équilatéral sphérique ABC de côté s . En raison de la symétrie sphérique, le barycentre exponentiel de la mesure uniforme $\nu_s = \frac{1}{3}(\delta_A + \delta_B + \delta_C)$ coïncide avec le centre de symétrie O du triangle. Cette propriété fondamentale découle de l'annulation des vecteurs logarithmes par symétrie diédrale.

La relation entre le côté s et la distance $\rho = d(O, A)$ est donnée par la loi des cosinus sphérique appliquée au triangle isocèle OAB :

$$\cos s = \cos^2 \rho + \sin^2 \rho \cos \left(\frac{2\pi}{3} \right).$$

En résolvant cette équation, on obtient le développement asymptotique :

$$\rho(s) = \frac{s}{\sqrt{3}} + \frac{\cos^2(\pi/3)}{48 \sin^3(\pi/3)} s^3 + O(s^5) = \frac{s}{\sqrt{3}} + \frac{s^3}{16\sqrt{3}} + O(s^5).$$

Construction des barycentres convexes Pour prouver la première inclusion dans la Proposition 3.1, considérons un triangle équilatéral ABC sur la sphère avec des côtés de longueur $s \rightarrow 0$. Définissons :

- I : Milieu de (A, B)
- J : Barycentre de B et C avec les poids $\frac{2}{3}, \frac{1}{3}$
- K : Barycentre de A et C avec les poids $\frac{2}{3}, \frac{1}{3}$
- L : Milieu de (J, K)
- M : Barycentre de I et C avec les poids $\frac{2}{3}, \frac{1}{3}$

Soit $\nu_s = \frac{1}{3}(\delta_A + \delta_B + \delta_C)$. Les points L et M sont des barycentres convexes de ν_s . Soit N le barycentre exponentiel de ν_s . Tous se trouvent sur la géodésique IC .

Lemme 4

Lorsque $s \rightarrow 0$:

$$d(C, L) = \frac{s}{\sqrt{3}} - \frac{\sqrt{3}}{2^3 \cdot 3^4} s^3 + \mathcal{O}(s^5),$$

$$d(C, M) = \frac{s}{\sqrt{3}} + \frac{\sqrt{3}}{2^3 \cdot 3^2} s^3 + \mathcal{O}(s^5),$$

$$d(C, N) = \frac{s}{\sqrt{3}} + \frac{\sqrt{3}}{2^3 \cdot 3^3} s^3 + \mathcal{O}(s^5).$$

Preuve. Pour un triangle ABC sur l'hémisphère avec un angle $\beta = \angle CAB$, on utilise la loi des cosinus sphérique :

$$\cos BC = \cos AB \cos AC + \sin AB \sin AC \cos \beta. \quad (3.1)$$

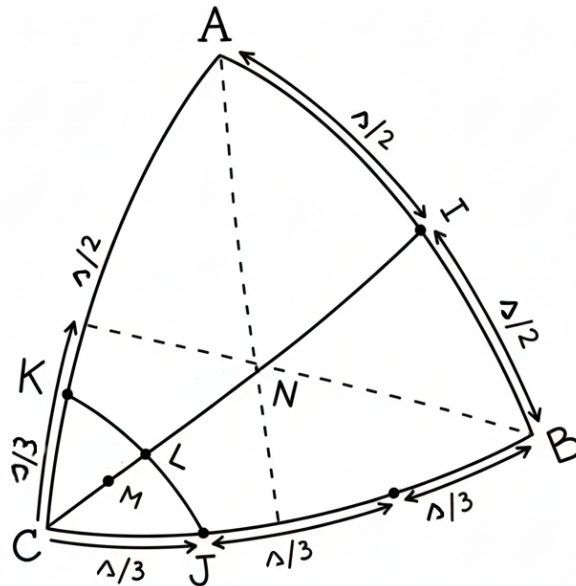


FIGURE 3.1 – Figure du triangle sphérique.

Soit $\alpha = \angle ACB$. En appliquant (3.1) au triangle ACB :

$$\cos s = \cos^2 s + \sin^2 s \cos \alpha \implies \cos \alpha = \frac{1}{8} - \frac{1}{8} s^2 + \mathcal{O}(s^4).$$

Soit $b = d(J, K)$. Pour le triangle JCK :

$$\cos(2b) = \cos^2 \frac{2s}{3} + \sin^2 \frac{2s}{3} \cos \alpha \implies \cos b = 1 - \frac{1}{18} s^2 - \frac{7}{2^2 \cdot 3^5} s^4 + \mathcal{O}(s^6).$$

Pour le triangle KLC :

$$\cos \frac{2s}{3} = \cos d(C, L) \cos b \implies d(C, L) = \frac{s}{\sqrt{3}} - \frac{\sqrt{3}}{2^3 \cdot 3^4} s^3 + \mathcal{O}(s^5).$$

Pour $d(C, M) = \frac{2}{3}d(C, I)$, en utilisant le triangle AIC :

$$\cos s = \cos \frac{s}{2} \cos d(C, I) \implies d(C, M) = \frac{s}{\sqrt{3}} + \frac{\sqrt{3}}{2^3 \cdot 3^2} s^3 + \mathcal{O}(s^5).$$

Pour $d(C, N)$, considérons le triangle ANC avec $\angle ANC = \frac{2\pi}{3}$:

$$\cos s = \cos^2 d(C, N) + \sin^2 d(C, N) \cos \frac{2\pi}{3} \implies d(C, N) = \frac{s}{\sqrt{3}} + \frac{\sqrt{3}}{2^3 \cdot 3^3} s^3 + \mathcal{O}(s^5).$$

$$\text{Ainsi, } d(N, M) = \frac{\sqrt{3}}{2^3 \cdot 3^3} s^3 + \mathcal{O}(s^5).$$

□

Triangle Hyperbolique

Les barycentres géodésiques s'étendent naturellement à l'espace hyperbolique, mais avec des modifications importantes dues au changement de signe de la courbure. Pour un triangle équilatéral hyperbolique ABC de côté s , le barycentre exponentiel N est toujours défini comme l'unique solution de l'équation de Karcher.

La construction géométrique des barycentres convexes suit le même schéma que dans le cas sphérique, mais les développements asymptotiques présentent des signes opposés pour les termes d'ordre s^3 :

Lemme 5

Lorsque $s \rightarrow 0$:

$$d(C, L) = \frac{s}{\sqrt{3}} + \frac{\sqrt{3}}{2^3 \cdot 3^4} s^3 + \mathcal{O}(s^5),$$

$$d(C, M) = \frac{s}{\sqrt{3}} - \frac{\sqrt{3}}{2^3 \cdot 3^2} s^3 + \mathcal{O}(s^5),$$

$$d(C, N) = \frac{s}{\sqrt{3}} - \frac{\sqrt{3}}{2^3 \cdot 3^3} s^3 + \mathcal{O}(s^5).$$

Pour un triangle hyperbolique de côtés a, b, c et d'angle γ opposé à c , on a :

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos \gamma$$

Cette formule diffère de son analogue sphérique par le signe moins devant le dernier terme, reflétant la courbure négative. Pour un triangle équilatéral ($a = b =$

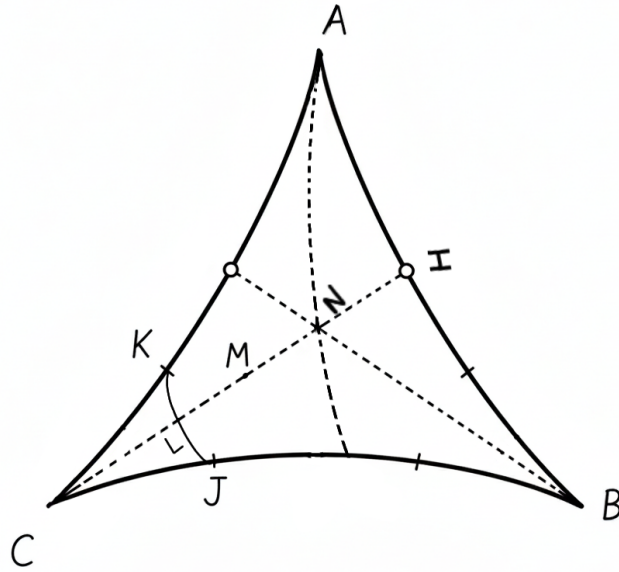


FIGURE 3.2 – Figure du triangle hyperbolique.

$c = s$), cette relation devient :

$$\cosh s = \cosh^2 \rho - \sinh^2 \rho \cos \left(\frac{2\pi}{3} \right) = \cosh^2 \rho + \frac{1}{2} \sinh^2 \rho$$

En résolvant pour $\rho = d(O, A)$, on obtient le développement :

$$\rho(s) = \frac{s}{\sqrt{3}} - \frac{\cos^2(\pi/3)}{48 \sin^3(\pi/3)} \cdot s^3 + O(s^5) = \frac{s}{\sqrt{3}} - \frac{s^3}{16\sqrt{3}} + O(s^5)$$

Le reste de la démonstration c'est comme ce qui précède .

Extension aux Polygones Réguliers

Les résultats obtenus pour les triangles équilatéraux s'étendent aux polygones réguliers géodésiques. Pour un polygone régulier à $n \geq 3$ sommets sur la sphère, le barycentre exponentiel de la mesure uniforme sur les sommets coïncide avec le centre de symétrie O . La relation entre le côté s et la distance $\rho = d(O, \text{sommet})$ est donnée par :

$$\cos s = \cos^2 \rho + \sin^2 \rho \cos \left(\frac{2\pi}{n} \right)$$

Le développement asymptotique devient :

$$\rho(s) = \frac{s}{2 \sin(\pi/n)} + \frac{\cos^2(\pi/n)}{48 \sin^3(\pi/n)} s^3 + O(s^5)$$

L'ensemble des barycentres convexes contient toujours un voisinage d'ordre s^3 autour du barycentre exponentiel.

3.1.3 Estimation de rayon des barycentres

Minoration de rayon des barycentres dans le cas sphérique

Le changement de signe des termes d'ordre s^3 est un indicateur directe de l'opposition entre courbure positive et négative. Cette différence affecte profondément la répartition des barycentres convexes autour du barycentre exponentiel

L'étude des voisinages des barycentres révèle comment la courbure influence la structure locale des ensembles de barycentres convexes noté $b(\mu)$.

Proposition 20(cas sphérique)

Si $c \in [0, \frac{\sqrt{3}}{2^3 \cdot 3^3}]$, il existe $s_0 > 0$ tel que pour tout $s \in (0, s_0)$, $b(\nu_s)$ contient $B(N, cs^3)$. et on a

$$B(N, cs^3) \subset b(\nu_s)$$

la constantes c dépend de la courbure. Pour la sphère, $c < \sqrt{3}/(8\pi)$, tandis que pour l'espace hyperbolique, les constantes sont généralement plus grandes en valeur absolue, reflétant la plus grande "dispersion" des géodésiques.

Preuve:

Construction récursive des mesures: Définissons μ_r comme la limite de mesures sur des polygones à $2^n \times 3$ côtés. Soit $\nu^0 = \nu_s$ et définissons de manière récursive :

$$\nu^n = \frac{1}{2} \left(\nu^{n-1} + \rho_{\frac{\pi}{3 \cdot 2^{n-1}}}(\nu^{n-1}) \right), \quad n \in \mathbb{N}^*$$

où ρ_θ est la rotation autour de o d'un angle θ .

Définissons les points M^n récursivement :

$$\overline{oM^{\hat{n}}} = \frac{1}{2} \left(\overline{oM^{n-1}} + \rho_{\frac{\pi}{3 \cdot 2^{n-1}}} \left(\overline{oM^{n-1}} \right) \right).$$

Par récurrence, $M^n \in b(\nu^n)$. En identifiant $T_o W \cong \mathbb{C}$ avec $o \mapsto 0$ et $M^0 \mapsto z_0 \in \mathbb{R}^+$, le vecteur $\overline{oM^{\hat{n}}}$ correspond à $z_n \in \mathbb{C}$ vérifiant :

$$z_n = \frac{1}{2} \left(z_{n-1} + e^{i \frac{\pi}{3 \cdot 2^{n-1}}} z_{n-1} \right).$$

Analyse de convergence et estimation du rayon: Cette relation de récurrence permet de démontrer par récurrence que pour tout $n \in \mathbb{N}$:

$$z_n = z_0 \frac{1}{2^n} \sum_{j=0}^{2^n-1} e^{\frac{-j\pi}{3 \cdot 2^n}},$$

ce qui se simplifie en :

$$z_n = z_0 \frac{1}{2^n} \frac{1 - e^{-\pi j \pi / 3}}{1 - e^{\frac{\pi}{3 \cdot 2^n}}}.$$

En remplaçant le module de $1 - e^{\frac{\pi}{3 \cdot 2^n}}$ par $\frac{\pi}{3 \cdot 2^n}$, on trouve que le module de z_n converge vers $\frac{3}{2^n} |z_0|$ lorsque $n \rightarrow \infty$. Comme ν^n converge faiblement vers μ , tout point limite de la suite $(M^n)_{n \in \mathbb{N}}$ appartient à $b(\mu)$. Ceci implique que le rayon de $b(\mu)$ est au moins $\frac{3}{2^n} |z_0|$.

Un équivalent asymptotique pour le module de z_0 lorsque $r \rightarrow 0$ est $\frac{\sqrt{3}}{2^2 \cdot 3^2} s^3$, ou de manière équivalente $\frac{r^3}{12}$. L'équivalent de $\frac{3}{2^n} |z_0|$ devient $\frac{\sqrt{3}}{8\pi} r^3$, ce qui complète la preuve de la proposition. \square

Majoration de rayon des barycentres:

Nous considérons ici une boule géodésique de S^2 , centrée en o . En coordonnées polaires (r, θ) , la métrique vaut :

$$ds^2 = dr^2 + \sin^2 r d\theta^2.$$

Soit μ_r la mesure uniforme sur le cercle géodésique de rayon r . L'objectif est de déterminer l'ordre de grandeur du rayon de $b(\mu_r)$.

Calcul de la Hessienne: Le calcul montre que les seuls symboles de Christoffel non nuls sont :

$$\Gamma_{\theta\theta}^r = -\sin r \cos r, \quad \Gamma_{r\theta}^\theta = \cot r.$$

La Hessienne d'une fonction du type $f(r, \theta) = r \cos \theta + u(r)$ peut alors s'écrire sous forme matricielle explicite.

Une fonction géodésiquement convexe: Le choix

$$u'(r) = r(\sin r \cos r - 1)$$

donne une fonction dont la trace et le déterminant de la Hessienne sont positifs pour r petit. Il en résulte une borne supérieure du rayon de $b(\mu_r)$.

Hessienne et convexité dans le cas hyperbolique: Les symboles de Christoffel

$$\Gamma_{\theta\theta}^r = -\sinh r \cosh r, \quad \Gamma_{r\theta}^\theta = \coth r$$

permettent de réaliser un calcul analogue à celui du cas sphérique. Le choix d'une fonction radiale $v(r)$ telle que

$$v'(r) = r(\sinh r \cosh r - 1)$$

conduit à une fonction convexe du type $r \cos \theta + v(r)$.

Proposition 21(cas sphérique)

Soit C une constante positive telle que $C > \frac{2}{9}$, alors il existe $r_0 \in [0, R[$ tel que pour tout $r \in]0, r_0[$, l'inclusion suivante est vérifiée :

$$b(\mu_r) \subset B(o, Cr^3).$$

nous construirons des fonctions convexes presque affines près de l'origine.

En coordonnées polaires, la métrique s'écrit $ds^2 = dr^2 + \sin^2 r d\theta^2$. On en déduit les symboles de Christoffel de la connexion, qui sont tous nuls sauf les trois suivants :

$$\Gamma_{\theta\sigma}^\sigma = -\cos r \sin r, \quad \Gamma_{r\theta}^\theta = \Gamma_{\theta r}^\theta = \cot r.$$

Nous calculons ensuite les hessiens :

$$\text{Hess } r = \cos r \sin r d\theta \otimes d\theta, \quad \text{Hess } \theta = -\cot r(dr \otimes d\theta + d\theta \otimes dr),$$

et

$$\text{Hess}(r \cos \theta) = \sin \theta(r \cot r - 1)(dr \otimes d\theta + d\theta \otimes dr) + \cos \theta(\cos r \sin r - r)d\theta \otimes d\theta.$$

Lemme 6

Si $R \in [0, \frac{2}{5}]$ est suffisamment petit, alors la fonction associant à (r, θ) la valeur $r \cos \theta + \int_0^r \left(\frac{s}{\sin s \cos s} - 1 \right) ds$ est convexe sur $\widehat{W} = B(o, R)$.

Corollaire 2

Soit C un réel strictement supérieur à $\frac{2}{9}$. Il existe $r_0 \in [0, R[$ tel que la fonction associant à (r, θ) la valeur $r \cos \theta + Cr^3$ est convexe sur $B(o, r_0)$.

Le corollaire s'obtient en remarquant, à la fin de la preuve du lemme, que la fonction $f : r \mapsto \int_0^r \left(\frac{s}{\sin s \cos s} - 1 \right) ds$ est asymptotiquement équivalente à $\frac{2}{9}r^3$ près de 0.

Preuve du lemme La matrice hessienne de $r \cos \theta + f(r)$ est donnée par :

$$\begin{pmatrix} f''(r) & \sin \theta (r \cot r - 1) \\ \sin \theta (r \cot r - 1) & \cos \theta (\cos r \sin r - r) + f'(r) \cos r \sin r \end{pmatrix}.$$

La fonction $r \cos \theta + f(r)$ est convexe si et seulement si cette matrice est définie positive, ce qui se produit lorsque la trace et le déterminant de la matrice sont positifs. Cela se traduit par deux inégalités :

Condition sur la trace :

$$f''(r) + \cos \theta (\cos r \sin r - r) + f'(r) \cos r \sin r \geq 0, \quad (3.2)$$

qui se simplifie (en utilisant $f'(r) = \frac{r}{\sin r \cos r} - 1$) en :

$$f''(r) + (r - \cos r \sin r)(1 - \cos \theta) \geq 0.$$

Le terme de droite est positif. À partir de l'expression $f''(r) = \frac{2 \sin 2r - 4r \cos 2r}{\sin^2 2r}$, on déduit que $f''(r)$ est asymptotiquement équivalent à $\frac{4r}{3}$ près de 0. Pour r suffisamment petit, $f''(r)$ est positif. Ainsi, l'inégalité (1) est vérifiée.

Condition sur le déterminant : En substituant $f'(r) = \frac{r}{\sin r \cos r} - 1$ dans l'inégalité (2), on obtient :

$$f''(r) [(r - \cos r \sin r)(1 - \cos \theta)] - \sin^2 \theta (r \cot r - 1)^2 \geq 0. \quad (3.3)$$

Analyse asymptotique et estimations affinées

Un équivalent de $f''(r) [(r - \cos r \sin r)(1 - \cos \theta)]$ près de $r = 0$ est $\frac{r^4}{9} \cdot 8(1 - \cos \theta)$, et un équivalent de $\sin^2 \theta (r \cot r - 1)^2$ est $\frac{r^4}{9} \sin^2 \theta$. La différence entre ces termes est :

$$\frac{r^4}{9} \sin^2 \theta \left(16 - 4 \cos^2 \frac{\theta}{2} \right).$$

Ceci confirme que l'inégalité (2) est valide près de $r = 0$.

Pour montrer $f(r) \sim \frac{2}{3}r^3$ lorsque $r \rightarrow 0$, on observe que $f'(r) \sim \frac{2}{3}r^2$, puis on intègre.

Lemme 7

Soit $C > \frac{2}{3}$. Il existe $R \in \left[0, \frac{\pi}{2}\right]$ tel que pour toute mesure de probabilité μ sur $W = B(o, R)$ et tout $b \in b(\mu)$:

$$\delta(e_\mu, b) \leq C \int_V \delta^3(e_\mu, y) \mu(dy),$$

où δ est la distance riemannienne sur W , et e_μ est le barycentre exponentiel de μ .

Preuve du Lemme: Soit $C > \frac{2}{3}$, et choisissons r_0 tel que $r \cos \theta + Cr^3$ soit convexe sur $B(o, r_0)$. Posons $R = \frac{r_0}{2}$. Soit λ une forme linéaire sur $T_{e_\mu}W$ avec $\|\lambda\| = 1$. La fonction :

$$x \mapsto \lambda(\overline{e_\mu x}) + C\|\overline{e_\mu x}\|^3$$

est convexe sur W . Pour tout $b \in b(\mu)$, ceci implique :

$$\lambda(\overline{e_\mu b}) + C\|\overline{e_\mu b}\|^3 \leq \int_W \left(\lambda(\overline{e_\mu x}) + C\|\overline{e_\mu x}\|^3 \right) \mu(dx).$$

Puisque $\int_W \lambda(\overline{e_\mu x}) \mu(dx) = 0$, on obtient :

$$\lambda(\overline{e_\mu b}) \leq C \int_W \|\overline{e_\mu x}\|^3 \mu(dx).$$

Valable pour tout λ , ceci donne :

$$\delta(e_\mu, b) \leq C \int_W \|\overline{e_\mu x}\|^3 \mu(dx).$$

Pour $\mu = \mu_r$ (Proposition 19), puisque $e_{\mu_r} = o$, ceci fournit la seconde inclusion dans la Proposition (21).

Conclusion

Cette étude comparative a mis en lumière les propriétés fondamentales des barycentres géodésiques des triangles sphériques et hyperboliques. Les développements asymptotiques précis que nous avons établis révèlent comment la courbure influence la structure locale des ensembles de barycentres convexes.

La différence de signe du terme en s^3 dans les cas sphérique et hyperbolique illustre de façon marquée l'opposition entre courbure positive et négative.

Les résultats obtenus s'étendent naturellement aux polygones réguliers et ouvrent des perspectives intéressantes pour les applications pratiques et les développements algorithmiques.

3.2 Barycentres de Matrices de Covariance et Entropie Riemannienne pour l'Analyse du Risque Structurel sur les Marchés Financiers

Dans cette section, nous rassemblons les définitions et résultats déjà employés dans les chapitres précédents. Son objectif principal est d'établir clairement les notations et de constituer une référence pour les résultats développés de mon article. L'essentiel du contenu présenté ci-après s'inspire largement de notre article : Barycenters of Covariance Matrices and Riemannian Entropy for Analyzing Structural Risk in Financial Markets.

3.2.1 Prologue :

L'évaluation précise du risque systémique constitue un défi majeur au sein de la finance quantitative contemporaine. Les approches classiques, généralement fondées sur des méthodes euclidiennes linéaires, se révèlent insuffisantes pour caractériser la complexité structurelle des interdépendances financières. En particulier, ces méthodes ignorent la nature géométrique des matrices de covariance, qui résident sur la variété Riemannienne des matrices symétriques définies positives (SPD). Cette limite entraîne une incapacité à détecter les modifications profondes de la topologie de dépendance, notamment lors des épisodes de crises globales.

La crise de la COVID-19 a constitué un choc systémique exogène d'une ampleur exceptionnelle, générant une synchronisation inédite des marchés, une explosion de l'incertitude structurelle et une rupture des régimes de dépendance historiques. Cette transition brutale a révélé la fragilité des indicateurs usuels de volatilité ou de corrélation et a mis en évidence la nécessité d'outils géométriquement cohérents pour appréhender la dynamique du risque.

La présente étude introduit un cadre géométrique fondé sur :

- le barycentre Riemannien des matrices de covariance, permettant de caractériser la tendance centrale dans l'espace SPD ;
- l'entropie Riemannienne, mesurant la dispersion structurelle et l'incertitude non linéaire du système financier.

L'objectif principal de l'article est de développer et d'examiner une nouvelle approche géométrique pour détecter et analyser les chocs structurels dans les marchés

financiers en se basant sur les barycentres dans les variétés des matrices de covariances. Nous avons représenté des nouveaux concepts :

- Capturer la tendance centrale des covariances de marché dans le temps en localisant leur barycentre dans la variété riemannienne.
- Caractériser le désordre structurel et l'incertitude du marché en utilisant une mesure d'entropie riemannienne .
- Montrer que cette approximation à double action est supérieure aux approches basées sur la volatilité pour dériver une indication rapide et précise des ruptures structurelles ou des chocs systémiques, notamment durant la pandémie de COVID-19.

- **Données** : L'analyse utilise les log-rendements hebdomadaires de 17 indices boursiers internationaux sur une période allant de janvier 2016 à janvier 2023.

- **Modélisation de la Covariance** : Chaque matrice de covariance est définie comme un point de la variété SPD.

- **Calcul du Barycentre** : Le barycentre riemannien des matrices de covariance est calculé au fil du temps pour représenter l'évolution de la structure de dépendance centrale pour des indices boursiers.

- **Mesure de l'Entropie** : La mesure de l'entropie riemannienne est calculée par la variance de Fréchet pour mesurer le désordre et la dispersion des matrices de covariance autour du barycentre.

- **Détection des Points de Changement** : Des procédures statistiques sont appliquées à chaque trajectoire du barycentre et de l'entropie pour isoler des changements de régime distincts associés à des événements de marché notables.

3.2.2 Développement : Méthodologie, résultats et discussion

Cadre méthodologique

Les matrices de covariance sont des objets fondamentaux en statistique et en apprentissage automatique. Elles appartiennent à l'espace des matrices symétriques définies positives (noté \mathcal{S}^+ ou SPD). L'analyse de leur variabilité nécessite le choix d'une métrique adaptée, car l'espace SPD n'est pas un espace vectoriel euclidien.

Parmi les approches existantes, la métrique riemannienne affine-invariante (AIRM) occupe une place centrale. Elle confère à l'espace SPD une structure de variété riemannienne à courbure non positive, ce qui garantit des propriétés théoriques remarquables telles que l'invariance par transformations linéaires inversibles et l'existence d'une unique géodésique entre deux points quelconques.

Cependant, cette richesse géométrique a un coût calculatoire élevé. En particulier, le calcul de la moyenne géométrique (ou moyenne de Fréchet) nécessite une procédure itérative — typiquement une descente de gradient riemannien — faisant intervenir à chaque itération des calculs d'exponentielle et de logarithme matriciels. Ce coût devient rédhibitoire pour de grandes séries de données ou des matrices de grande dimension. De plus, la convergence de ces algorithmes dépend de l'initialisation choisie.

Métrique affine-invariante riemannienne (AIRM)

Distance géodésique

Pour deux matrices $P, Q \in SPD$, la distance géodésique sous AIRM est donnée par :

$$\text{dist}_R(P, Q) = \left\| \log \left(P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right) \right\|_F \quad (3.4)$$

où $\| \cdot \|_F$ désigne la norme de Frobenius.

Moyenne géométrique (moyenne de Fréchet)

La moyenne riemannienne d'un ensemble $\{C_i\}_{i=1}^k \subset SPD$ est définie comme l'unique minimiseur de la somme des distances au carré :

$$\bar{C} = \arg \min_{X \in SPD} \sum_{i=1}^k \text{dist}_R^2(X, C_i) \quad (3.5)$$

Gradient de la fonction coût

La fonction coût à minimiser est $\rho(X) = \frac{1}{2k} \sum_{i=1}^k \text{dist}_R^2(X, C_i)$. Son gradient riemannien s'écrit :

$$\nabla \rho(X) = -\frac{1}{k} \sum_{i=1}^k \log_X(C_i) \quad (3.6)$$

où $\log_X(\cdot)$ est l'application logarithme riemannienne au point X .

Algorithme de descente de gradient riemannien

L'itération de mise à jour est donnée par :

$$\mu_{t+1} = \exp_{\mu_t} \left(\frac{1}{k} \sum_{i=1}^k \log_{\mu_t}(C_i) \right) \quad (3.7)$$

En développant à l'aide des applications exponentielle et logarithme matricielles, on obtient :

$$\mu_{t+1} = \mu_t^{\frac{1}{2}} \exp \left(\frac{1}{k} \sum_{i=1}^k \log \left(\mu_t^{-\frac{1}{2}} C_i \mu_t^{-\frac{1}{2}} \right) \right) \mu_t^{\frac{1}{2}} \quad (3.8)$$

Cette approche garantit des propriétés géométriques remarquables (invariance affine, courbure non positive) mais reste coûteuse en calcul en raison des itérations et des calculs d'exponentielles/logarithmes matriciels à chaque étape.

Métrie log-euclidienne (LEM)

Distance log-euclidienne

Pour deux matrices $A, B \in SPD$, la distance log-euclidienne est définie par

$$\text{dist}_{LE}(A, B) = \|\log(A) - \log(B)\|_F \quad (3.9)$$

Moyenne log-euclidienne

La moyenne log-euclidienne d'un ensemble $\{C_i\}_{i=1}^k$ est explicite :

$$\mu_{LE} = \exp \left(\frac{1}{k} \sum_{i=1}^k \log(C_i) \right) \quad (3.10)$$

Entropie log-euclidienne

Pour un ensemble de matrices $\{\Sigma_i\}_{i=1}^N$ avec moyenne $\bar{\Sigma}_{LE}$, l'entropie log-euclidienne est :

$$H_{LE} = \frac{1}{N} \sum_{i=1}^N \|\log(\Sigma_i) - \log(\bar{\Sigma}_{LE})\|_F^2 \quad (3.11)$$

Cette quantité généralise la variance euclidienne aux données évoluant dans l'espace SPD.

Choix de la métrique log-euclidienne en pratique

Dans cette étude, nous retenons la **métrique log-euclidienne** pour les raisons suivantes :

- **Simplicité algorithmique** : les équations (3.9), (3.10) et (3.11) sont directes.
- **Absence d'itérations** : pas de problème de convergence ni d'initialisation.
- **Interprétation intuitive** : l'entropie (3.11) est une véritable variance dans l'espace logarithmique.

L'entropie log-euclidienne H_{LE} permet de quantifier la dispersion des matrices autour de leur moyenne. Une **faible entropie** indique une forte concentration, donc une bonne stabilité temporelle ou structurelle. Une **entropie élevée** révèle une hétérogénéité importante, signe d'instabilité potentielle du système.

Méthodes

— Revue de la Littérature

L'étude du comportement non linéaire sur les marchés financiers a été initiée par Hsieh (1995), qui a reconnu l'existence de dynamiques non linéaires et ses implications pour l'efficacité des marchés, traçant ainsi la voie vers des méthodes plus complexes pour étudier le risque financier. À partir de ce point, la recherche a commencé à adopter l'angle de la théorie de l'information, de l'entropie et de leur utilisation pour déterminer le désordre et l'inefficacité des marchés. Par exemple, Pele et al. (2017) ont utilisé l'entropie comme mesure du risque de marché. Vozna (2017) a présenté les bases théoriques de l'entropie dans l'analyse économique. Oprean et al. (2017) ont proposé une nouvelle méthodologie pour mesurer l'efficacité du marché dans un paradigme dynamique non linéaire complexe. De même, Benedetto et al. (2021) et Brouty & Garcin (2022) ont utilisé des méthodologies basées sur l'entropie pour étudier le flux d'information et l'efficacité des marchés. Des études plus récentes ont adopté la mesure de l'efficacité en examinant l'efficacité variable dans le temps et les crises financières Patra & Hiremath (2022), Papla & Siedleck (2024), Alkan & Süsay Alkan (2024), Vu et al. (2024) ont utilisé l'entropie pour mesurer l'efficacité pendant la pandémie et Ozcan et al. (2023) ont combiné l'entropie avec l'analyse relationnelle grise pour évaluer l'effet de la pandémie de COVID-19 sur la performance financière.

Parallèlement, l'utilisation de la géométrie riemannienne s'est également étendue au domaine de la finance. Huang et al. (2017) ont utilisé l'apprentissage de variétés non linéaires pour créer des signaux d'alerte précoce de ruptures structurelles sur les marchés, tandis que Huang et al. (2019) ont proposé une nouvelle méthode pour estimer la fragilité du système financier en utilisant la courbure des variétés. De plus, Abanto-Valle et al. (2021) ont adapté la méthodologie sophistiquée de Monte Carlo Hamiltonien sur les variétés riemanniennes pour modéliser la volatilité stochastique sur les marchés latino-américains, tandis que Ahn et al. (2021) ont introduit le concept de modélisation riemannienne dans la dynamique collective et comportementale.

En outre, Shu et al. (2025) ont récemment proposé un cadre complet qui combine l'entropie et la géométrie riemannienne pour comprendre les processus de transmission du risque sur les marchés boursiers mondiaux, établissant un lien entre les stratégies computationnelles et géométriques pour l'analyse financière.

Suivant cette trajectoire bibliographique, la présente étude se distingue par son originalité à utiliser simultanément le cadre entropique et riemannien

pour étudier les marchés financiers.

— Description des Données

Cette recherche empirique a utilisé un ensemble de données de 17 indices boursiers importants, sélectionnés comme représentatifs de différentes zones géographiques et économiques. Les indices et les zones des pays sont :

- **Europe** CAC 40 (France), DAX (Allemagne), Euro Stoxx 50 (Zone Euro), PSI 20 (Portugal), AEX (Pays-Bas) et FTSE 100 (Royaume-Uni) ;
- **Amérique** Dow Jones Industrial Average, NASDAQ Composite, S&P/TSX Composite (Canada), Bovespa Index (Brésil) et S&P/BMV IPC Index (Mexique) ;
- **Asie et Moyen-Orient** Nikkei 225 (Japon), FTSE China A50 (Chine), KOSPI (Corée du Sud), Nifty 50 (Inde) et Tadawul All Share (Arabie Saoudite) ; Océanie - S&P/ASX 200 (Australie).

Les données consistaient en des prix de clôture hebdomadaires, sur une période allant du 1er janvier 2016 au 1er janvier 2023. Cette période capture le comportement du marché caractérisé par la pré-pandémie de COVID-19, la pandémie de COVID-19 et la reprise du marché post-pandémie (Zaremba et al., 2020). Au total, chaque index compte 365 observations hebdomadaires. Les données ont été collectées sur [<https://www.investing.com/indices>]. Une fréquence hebdomadaire a été choisie dans le but d'équilibrer les mouvements significatifs du comportement du marché et de prévenir le bruit haute fréquence qui pourrait affecter la validité et la robustesse de l'estimation de la matrice de covariance. Pour chaque indice $S_i(t)$, le rendement logarithmique est défini comme :

$$r_i(t) = \log(S_i(t)) - \log(S_i(t-1)),$$

où $r_i(t)$ représente le rendement composé continûment de l'indice i au temps t . Le vecteur des rendements au temps t est alors :

$$R(t) = [r_1(t), r_2(t), \dots, r_{17}(t)]^\top, \quad (3.12)$$

— Construction de la Série Temporelle de Matrices de Covariance

Afin d'examiner la dépendance variable dans le temps entre les marchés boursiers, nous avons dérivé une série temporelle à valeurs matricielles à partir des données en utilisant une approche par fenêtre glissante Feng, Y., Zhang, Y., Wang, Y. (2023). Chaque fenêtre temporelle glissante a une longueur de 26 jours, ce qui correspond approximativement à un mois. Pour chaque point dans le temps t , nous considérons la sous-matrice multivariée :

$$\mathcal{R}_t = [R(t - 25), R(t - 24), \dots, R(t)].$$

La matrice de covariance empirique est calculée comme :

$$\Sigma_t = \frac{1}{25} \sum_{k=0}^{25} (R(t - k) - \bar{R}_t) (R(t - k) - \bar{R}_t)^\top, \quad (3.13)$$

où \bar{R}_t désigne la moyenne locale des rendements dans la fenêtre. Chaque Σ_t est symétrique définie positive (SPD), c'est-à-dire $\Sigma_t \in \mathcal{S}_{++}^{17}$, formant une séquence :

$$\{\Sigma_t\}_{t=1}^T, \quad ,$$

qui constitue une série temporelle de matrices de covariance évoluant sur une variété riemannienne (Pennee, 2006).

La métrique adoptée pour caractériser les distances :

Métrique Log-Euclidienne :

$$\delta_{LE}(\Sigma_1, \Sigma_2) = \|\log(\Sigma_1) - \log(\Sigma_2)\|_F,$$

— Entropie Riemannienne et Incertitude du Marché

Pour la mesure de l'incertitude financière, nous utilisons le concept d'entropie riemannienne qui est lié à la variance généralisée de la matrice de covariance (Jeuris et al., 2012). Par conséquent, l'entropie riemannienne peut s'écrire :

$$H(\Sigma_1, \dots, \Sigma_N) = \frac{1}{N} \sum_{i=1}^N \delta_R^2(\Sigma_i, \bar{\Sigma}) \quad (3.14)$$

Cette entropie mesure la dispersion du risque de marché, ce qui nous permet de suivre l'évolution des processus d'incertitude dans le temps.

La progression de $H_R(\Sigma_t)$ et des barycentres Σ_W^* , indique trois phases d'incertitude :

- **Pré-COVID (2016-2019)**.
- **COVID (2020-2021)** .
- **Post-COVID (2022-2023)** .

Résultats et Discussion

— Matrice de Covariance pour la Période Pré-COVID

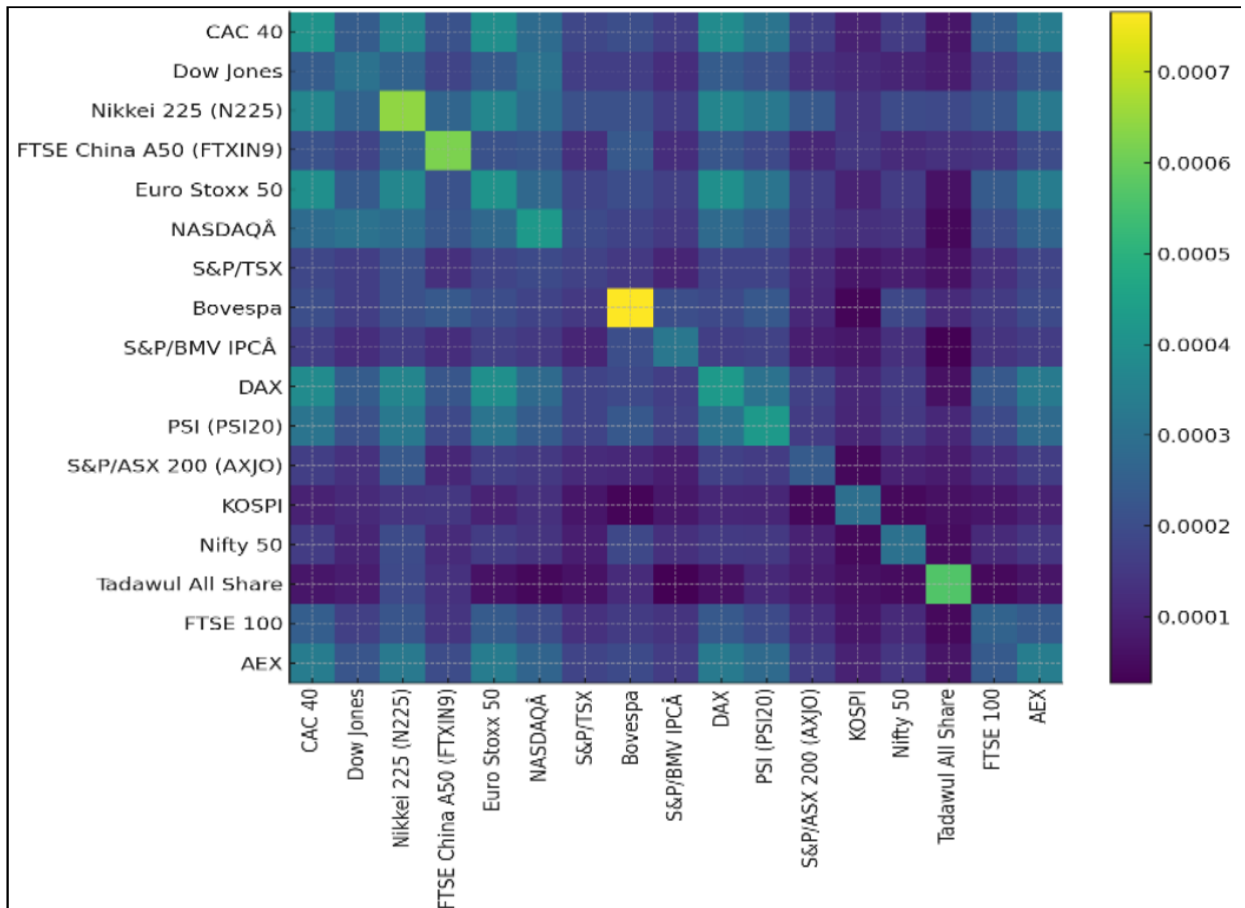


FIGURE 3.3 – Matrice de covariance avant la période COVID (2020-2021)

La matrice de covariance présentée dans la Figure (1) montre les interrelations entre les rendements hebdomadaires de certains des principaux indices boursiers mondiaux durant la période pré-COVID-19 (2016-2019). Les cellules de la matrice démontrent la valeur de covariance entre deux marchés financiers, tandis que les dégradés de couleur montrent la force et la direction de la co-mouvement.

Les couleurs claires (jaune, vert clair) indiquent une covariance plus élevée, signifiant que ces deux marchés évoluent ensemble un peu plus que la

plupart des autres marchés. Ceci est principalement visible dans les zones géographiquement et économiquement intégrées, durant les années ayant précédé la COVID-19. Un exemple de ceci inclut les marchés nord-américains (Dow Jones, NASDAQ, S&P/TSX) et les marchés européens (CAC 40, DAX, FTSE 100, Euro Stoxx 50) qui tendent à démontrer une forte intégration financière régionale et économique internationale.

Les valeurs de couleur sombres (bleu et violet) représentent des valeurs de covariance très faibles ou presque nulles, suggérant que les marchés évoluent largement de manière indépendante les uns des autres, ce qui est plus probable si les marchés sont géographiquement éloignés ou économiquement segmentés. Cela a été observé, par exemple, entre les indices asiatiques (Nikkei, KOSPI, Nifty 50) et le marché du Golfe (Tadawul).

D'un point de vue économique, le motif global de couleur montre que la structure du système financier mondial était relativement diversifiée et segmentée avant la crise de la COVID-19.

La force régionale était forte, mais la force interrégionale restait faible, ce qui suggérait un degré élevé de diversification de la finance mondiale et, par conséquent, un potentiel limité de contagion. Les couleurs claires le long de la diagonale principale représentent la variance propre de chaque marché, représentant la volatilité interne des marchés à l'activité intense tels que le NASDAQ et le Bovespa. La prédominance des couleurs sombres en dehors de la diagonale montre l'indépendance de la plupart des marchés et indique que le risque systémique global était faible. Globalement, à la fois d'un point de vue mathématique et économique, la matrice de covariance indique qu'avant la crise de la COVID-19, les marchés financiers mondiaux étaient modérément connectés, avec des liens forts principalement au sein de clusters économiques régionaux.

— Matrice de Covariance Durant la Période COVID

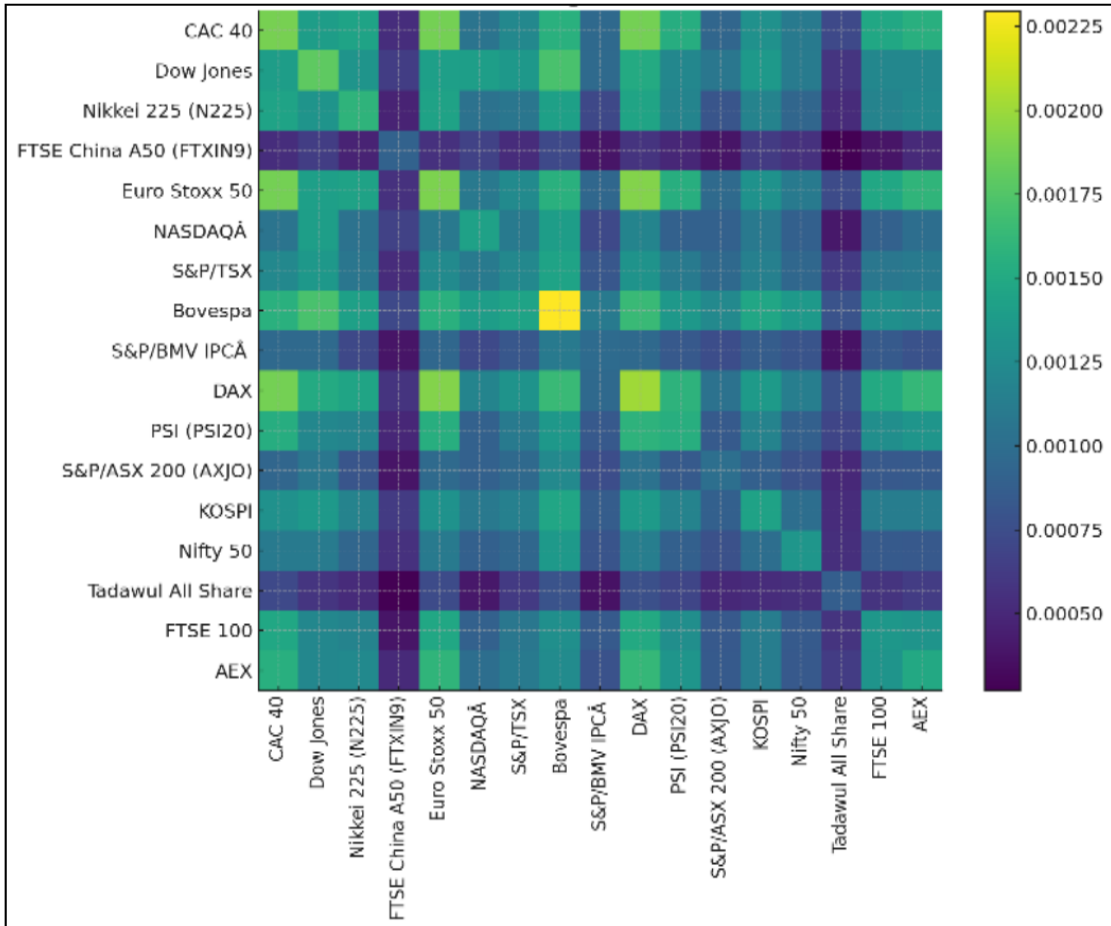


FIGURE 3.4 – Matrice de covariance durant la période COVID (2020-2021)

Cette Figure montre la matrice de covariance pour la période de la COVID-19, de janvier 2020 à juin 2021. Par sa nature mathématique, la matrice indique une augmentation substantielle des valeurs de covariance par rapport à avant la pandémie – en témoigne le plus grand nombre de couleurs claires par rapport aux couleurs sombres – ce qui indique de plus grands co-mouvements entre les marchés financiers. Les zones claires signifient des valeurs plus élevées en termes de métriques de covariance, indiquant que les rendements des marchés étaient plus synchronisés, impliquant une moindre diversification et

un niveau plus élevé d'interdépendance systémique.

Les caractéristiques en couleur claire indiquent une forte interconnexion avec les principaux indices, qui incluaient le Dow Jones, le NASDAQ, l'Euro Stoxx 50, le DAX et le Nikkei 225, car ils se déplaçaient de manière synchrone, signalant le choc systémique (sanitaire ou économique) affectant le monde, ainsi que les politiques monétaires et fiscales coordonnées au niveau mondial. D'autre part, les caractéristiques sombres qui émergeaient des marchés moins alignés ou des marchés émergents (Tadawul et S&P/BMV IPC)

suggèrent des covariances plus faibles, mais toujours positives, donc toujours liées aux effets globaux négatifs. Même pour ces marchés émergents, bien que moins intégrés dans les grands marchés financiers mondiaux, l'impact et la sensibilité aux effets de contagion étaient toujours présents.

D'un point de vue économique, la construction colorimétrique de la Figure représente une période d'interdépendance financière mondiale accrue et de risque systémique. La crise de la COVID-19 a servi de choc externe commun qui a synchronisé les comportements des marchés à travers les régions, réduisant ainsi l'unicité régionale et transformant le système financier mondial en une structure plus uniforme et interconnectée. Cette synchronisation accrue a diminué la capacité des portefeuilles à atteindre un certain niveau de diversification à travers les marchés internationaux, car les marchés ont réagi presque à l'unisson aux développements sanitaires, aux évolutions politiques et aux annonces de confinement ou de réouverture.

Globalement, la matrice de covariance pour la période de la COVID-19 montre clairement une augmentation de la synchronisation financière mondiale, ce qui signifie des valeurs de covariance plus élevées entre les marchés, tout en affichant également une augmentation significative de l'interconnectivité systémique et du risque global résultant d'un choc pandémique.

— Matrice de Covariance pour la Période Post-COVID

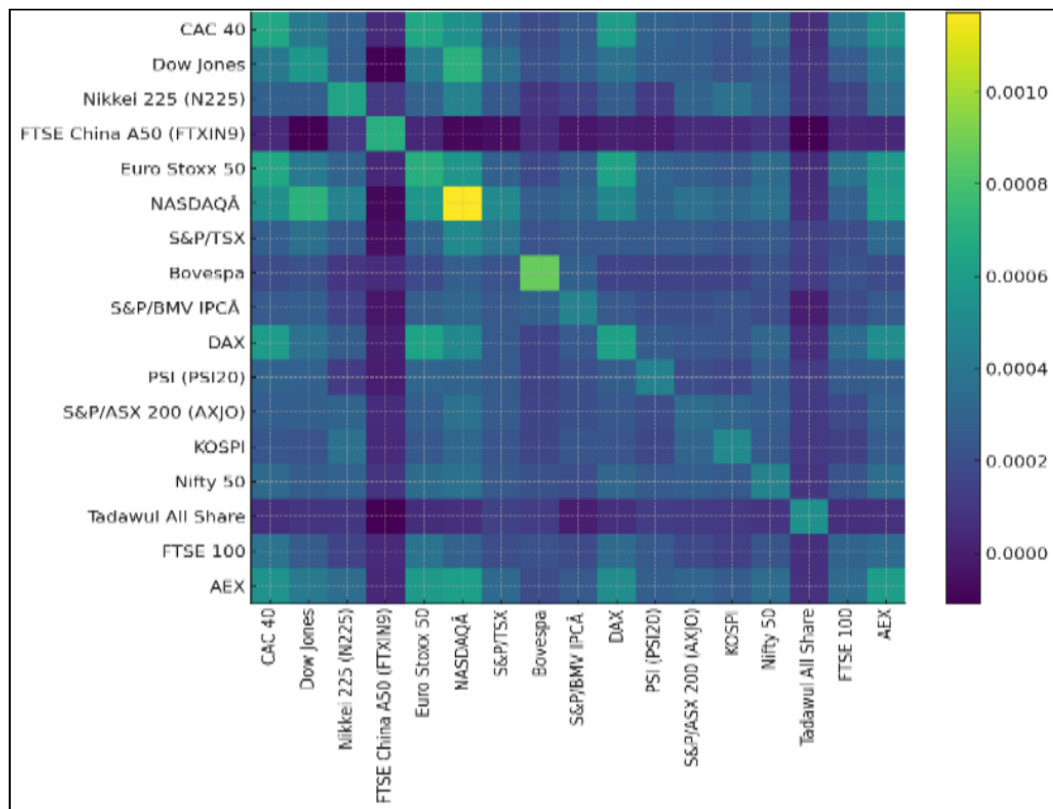


FIGURE 3.5 – Matrice de covariance après la période COVID (2020-2021)

La Figure présente la matrice de covariance dans la phase post-crise de la COVID-19, approximativement de la seconde moitié de 2021 jusqu'à fin 2023. En termes simples, d'un point de vue mathématique, la matrice identifie des mesures de covariance plus basses lorsque l'on compare la période post-crise à la période pandémique, en raison de la présence de couleurs plus sombres qui dénotent généralement des covariances plus faibles que les couleurs claires. Cela traduit toujours l'existence de moins de co-mouvements parmi les marchés financiers mondiaux dans l'ensemble, ainsi qu'un retour partiel vers des

marchés plus hétérogènes et plus indépendants après la pandémie, qui, dans certains pays, ont présenté un comportement de synchronisation élevé durant la crise pandémique.

Les zones plus sombres notables de la matrice reflètent des covariances plus faibles dans les co-mouvements entre des marchés géographiquement éloignés ou structurellement différents, comme le Nikkei 225, le Bovespa et le Tadawul, probablement en raison des différences dans le calendrier de reprise et les réponses de relance mises en œuvre par les pays, impactant les marchés individuels. Les couleurs claires de la matrice, particulièrement parmi les principales économies occidentales plus notables comme le Dow Jones, le NASDAQ, l'Euro Stoxx 50 et le DAX, indiquent qu'un certain niveau de co-mouvement entre les économies interconnectées persiste.

En termes économiques, la Figure montre la phase de réallocation financière mondiale où les marchés sont retournés à un niveau de différenciation régionale et d'indépendance dans leurs réactions aux chocs économiques. Ceci est différent de la période pandémique où les chocs déplaçaient les marchés de manière uniforme. Après la pandémie, les marchés ont commencé à évoluer principalement en fonction des fondamentaux nationaux tels que les variations des taux d'intérêt, l'inflation et la vitesse de la reprise économique. Cela montre que le risque systémique diminue et, à mesure que la corrélation entre les marchés décline, la réémergence de la diversification de portefeuille.

La matrice de covariance post-COVID a empêché le système financier mondial de rester à un niveau élevé de synchronisation et de risque systémique partagé, le ramenant vers un environnement qui devient plus stable et différencié. Mathématiquement, on observe des valeurs de covariance plus faibles et une plus grande surface de couleurs sombres, tandis que sur le plan économique, une reprise graduelle et un rétablissement de l'équilibre financier à travers les marchés mondiaux sont en cours.

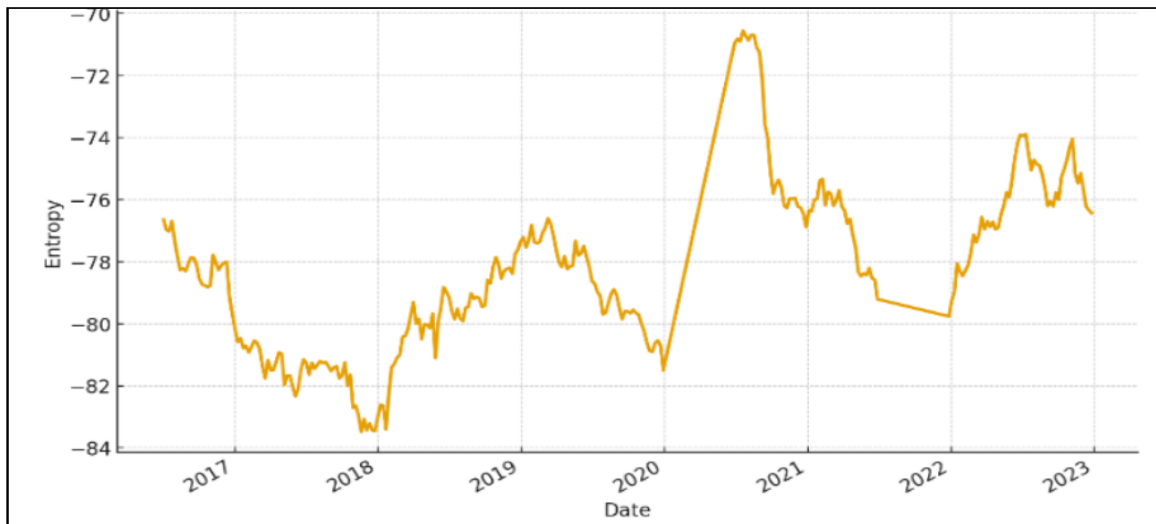


Fig : représentation graphique de l'entropie

FIGURE 3.6 – Evolution de l'entropie

La Figure montre la courbe d'Entropie Riemannienne (RE) représentant la progression temporelle de l'incertitude structurelle dans le système financier mondial sur la période 2016-2023. L'exploration de la courbe d'Entropie Riemannienne (RE) a identifié trois phases qui mettent en évidence le processus de changement dans la complexité et l'interdépendance des marchés.

Dans la phase pré-crise (2016-2019), les valeurs de l'entropie sont restées inférieures à une plage de -82 à -76, témoignant d'un niveau élevé de certitude et de stabilité structurelle pour les marchés financiers. Au cours de

cette période, les relations entre les indices de marché étaient cohérentes et prévisibles. Collectivement, un environnement financier illustrant un ordre systémique et une homogénéité était apparent.

Vers la fin 2019 et le début 2020, la mesure de l'entropie est retombée à environ -79, ce qui pourrait superficiellement suggérer une plus grande certitude ressentie par les marchés en raison de flux financiers hautement synchronisés. Néanmoins, l'état d'ordre et de certitude de cette entropie à -79 est précaire et reflète une période de compression structurelle et une diminution de la variété des comportements de marché. Cette baisse de l'entropie illustre également que le système avait absorbé un certain degré de résilience et était de plus en plus exposé aux chocs systémiques, faisant partie de l'alerte structurelle précoce avant le déclenchement de la crise de la COVID-19.

L'augmentation de l'entropie observée au début de la pandémie (2020-2021), atteignant environ -70, constitue une forte hausse de l'incertitude structurelle. L'incertitude structurelle augmente alors que les arrangements structurels précédents se décomposent et que le système financier mondial devient hautement entremêlé et synchronisé. L'augmentation de l'entropie a de nouveau atteint un pic alors que les structures passaient d'un état d'ordre relatif à un état d'incertitude structurelle maximale et de désordre systémique.

Après mi-2021, les valeurs de l'entropie ont commencé à diminuer graduellement et se sont stabilisées autour de -74, indiquant un répit par rapport à l'incertitude et l'établissement d'un nouvel équilibre. Cependant, l'équilibre post-crise est toujours plus complexe et plus incertain que celui observé avant 2019, ce qui suggère que le système financier s'est adapté à un nouveau niveau de base structurel d'incertitude.

Dans ce contexte, l'Entropie Riemannienne se présente comme un outil significatif pour évaluer la dynamique interne des systèmes financiers complexes. Elle va au-delà de son usage conventionnel en tant que mesure statistique de la variabilité et devient une métrique structurelle fondamentale de l'incertitude, au sein des réseaux financiers. Elle retrace les transitions sur le chemin du chaos avant, pendant et depuis les périodes de crise. Elle capture comment les marchés financiers évoluent d'états stables, à travers le chaos, et finalement vers un rééquilibrage de la complexité. Des niveaux élevés d'entropie sont indicatifs d'un désordre croissant et d'une cohérence systémique. Des niveaux plus bas, en revanche, reflètent une prévisibilité et une structure croissantes, qu'elles soient réelles ou une stabilité perçue de manière erronée. L'Entropie Riemannienne peut servir de structure d'alerte précoce, car elle peut évaluer l'épuisement de la certitude et de la résilience systémique, et fournit un cadre géométrique solide pour analyser l'équilibre précaire entre

l'ordre et l'incertitude qui constitue l'architecture financière mondiale.

— **Analyse et Interprétation de la Distance Riemannienne**

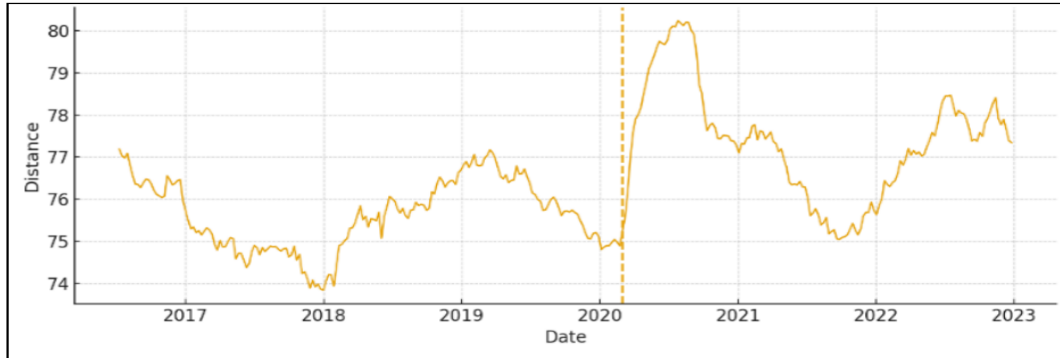


Fig : représentation graphique des distance entre la moyenne de Fréchet et les matrices de covariance.

FIGURE 3.7 – fonction distance entre le barycentre et les matrices de covariances des données

Cette figure explique l'évolution de la Distance Riemannienne entre la matrice de covariance variable dans le temps des marchés financiers et la matrice de référence pré-COVID, qui sert d'indicateur géométrique de dissimilarité structurelle par rapport à la forme de la matrice de covariance et non par rapport à l'échelle globale. Une distance plus petite indique une plus grande stabilité structurelle et un certain niveau de certitude systémique, tandis qu'une distance plus grande implique un certain degré de divergence structurelle et une incertitude accrue dans le réseau financier.

La Figure indique que généralement la Distance Riemannienne était étroitement comprise entre 74 et 77 durant la période pré-COVID (2016-

2019) et représente une période de stabilité structurelle et de cohésion entre les marchés. Au début de l'année 2020, la Distance Riemannienne a augmenté brusquement pour atteindre environ 80, indiquant une déviation géométrique considérable par rapport à la base de référence pré-crise. Cette flambée capture le choc systémique créé par la COVID-19 et la perturbation des structures de corrélation typiques sur les marchés financiers, induisant plutôt de nouveaux modèles de co-mouvement principalement dus à l'incertitude mondiale avec des réponses de marché corrélées.

Après ce pic, la distance a lentement diminué pour atteindre environ 75-77 vers mi-2021, ce qui indique une certaine reprise structurelle et une reconfiguration plus complexe des dépendances de marché. Durant les années post-crise (2022-2023), la distance a de nouveau augmenté pour atteindre environ 78, ce qui indique que le système financier avait atteint un nouvel équilibre structurel assez distinct du cadre structurel d'avant la pandémie – et qui était beaucoup plus complexe et interconnecté.

Globalement, la Figure montre clairement que la Distance Riemannienne capture très efficacement la dynamique géométrique de la transformation structurelle au sein des marchés financiers mondiaux – conduisant le lecteur d'une période de stabilité pré-crise, à travers une dysfonction systémique, vers un nouvel équilibre structurel caractérisé par une instabilité structurelle considérable.

3.2.3 Epilogue :

Cette étude fournit une évaluation explicite et complète de la transformation structurelle du système financier mondial durant les périodes pré-, pendant et post-COVID-19. En utilisant des outils de géométrie riemannienne spécifiquement la variété des matrices de covariance.

Les résultats montrent une trajectoire progressive : avant 2020, le système était régionalement stable avec une diversification financière, illustrée par une connectivité plus faible et un risque systémique plus bas, puis, largement en raison de la pandémie, l'interconnectivité des marchés mondiaux est passée à un réseau mondial plus aligné et hautement connecté, caractérisé par un risque systémique accru et des opportunités de diversification de portefeuille réduites. Dans la période post-pandémique, le système a commencé à retourner vers un nouveau équilibre.

En conclusion, nous avons développé et appliqué un nouveau cadre géométrique qui unifie l'entropie riemannienne avec les barycentres des matrices de covariance. L'analyse indique que la pandémie de COVID-19 a modifié à la fois les niveaux de risque et la géométrie structurelle des marchés financiers. La méthodologie crée une manière formelle et intuitive, adaptée aux institutions, pour analyser les systèmes financiers dynamiques, et représente une contribution précieuse à la finance mathématique. La méthodologie a amélioré la détection précoce des changements de régime structurels et de la contagion financière.

Bibliographie

- [1] Absil, P.-A., Mahony, R. et Sepulchre, R. (2007). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press.
- [2] Afsari, B. (2011). Riemannian L^p center of mass : Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, **139**, 655–673.
- [3] Longhena, A. (2025). *Hyperbolic representations of brain networks and clinical applications*. Thèse de doctorat, Université de la Sorbonne.
- [4] Arnaudon, M. (1995). Barycentres convexes et approximations des martingales continues dans les variétés. *Séminaire de Probabilités XXIX*, Lecture Notes in Mathematics **1613**, 70–85.
- [5] Arnaudon, M. (1994). Espérances conditionnelles et C-martingales dans les variétés. *Séminaire de Probabilités (Strasbourg)*, **28**, 300–311.
- [6] Arsigny, V., Commowick, O., Pennec, X. et Ayache, N. (2006). A log-Euclidean framework for statistics on diffeomorphisms. In Larsen, R., Nielsen, M. et Sporing, J. (éds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, Lecture Notes in Computer Science. Springer, 924–931.
- [7] Bačák, M. (2014). *Computing medians and means in Hadamard spaces*. arXiv :1210.2145.
- [8] Bhattacharya, R. et Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Annals of Statistics*, **33**, 1225–1259.
- [9] Bhattacharya, R. et Lin, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proceedings of the American Mathematical Society*, **145**(1), 413–428.
- [10] Bonnabel, S. (2013). Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control*, **58**(9), 2217–2229. (voir aussi arXiv :1111.5280)
- [11] Bouchard, F., Laurent, N., Said, S. et Le Bihan, N. (2025). *Beyond R-barycenters : an effective averaging method on Stiefel and Grassmann manifolds*. Prépublication arXiv.

- [12] Boumal, N. (2023). *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, prépublication.
- [13] Boumal, N., Mishra, B., Absil, P.-A. et Sepulchre, R. (2014). Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, **15**, 1455–1459.
- [14] Bouregaa, S., Bakbak, L. I., Segres, A. et Boudiab, A. (2025). Barycenters of covariance matrices and Riemannian entropy for analyzing structural risk in financial markets. *Journal of Information Systems Engineering and Management*, **10**(61s), 825–838.
- [15] Buss, S. R. et Fillmore, J. P. (2001). Spherical averages and applications to spherical splines and interpolation. *ACM Transactions on Graphics*, **20**, 95–126.
- [16] Calissano, A., Feragen, A. et Vantini, S. (2023). Populations of unlabelled networks : Graph space geometry and generalized geodesic principal components. *Biometrika*, asad024.
- [17] Calissano, A., Maignant, E. et Pennec, X. (2023). Towards quotient barycentric subspaces. In Nielsen, F. et Barbaresco, F. (éds.), *Geometric Science of Information*, Lecture Notes in Computer Science. Springer, 366–374.
- [18] Chakraborty, S., Ghosh, A. et Vemuri, B. (2019). Statistics on the compact Stiefel manifold : Theory and applications. *Annals of Statistics*.
- [19] Costa, S. I. R., Santos, S. A. et Strapasson, J. E. (2015). Fisher information distance : A geometrical reading. *Discrete Applied Mathematics*, **197**, 59–69.
- [20] Dryden, I. L. et Mardia, K. V. (2016). *Statistical Shape Analysis, with Applications in R*. John Wiley & Sons.
- [21] Dryden, I. L. et Marron, J. S. (2021). *Object Oriented Data Analysis*. Chapman & Hall/CRC.
- [22] Dryden, I. L., Hensman, J., Hicks, J. et al. (2019). Principal nested shape space analysis of molecular dynamics data. *Annals of Applied Statistics*, **13**(4), 2213–2234.
- [23] Damon, J. et Marron, J. S. (2013). Backwards principal component analysis and principal nested relations. *Journal of Mathematical Imaging and Vision*, **50**, 107–114.
- [24] Dichio, V. et De Vico Fallani, F. (2023). Statistical models of complex brain networks : A maximum entropy approach. arXiv :2209.05829.
- [25] Doss, S. (1949). Sur la moyenne d’un élément aléatoire dans un espace distancié. *Bulletin des Sciences Mathématiques*, **73**, 48–72.
- [26] Duncan, T. E. (1976). Stochastic integrals in Riemannian manifolds. *Journal of Multivariate Analysis*, **6**, 397–413.

- [27] Dryden, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Annals of Statistics*, **33**, 1643–1665.
- [28] Eltzner, B. et Huckemann, S. (2019). A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Annals of Statistics*, **47**(6), 3360–3381.
- [29] Emery, M. et Meyer, P. A. (1989). *Stochastic Calculus in Manifolds*. Springer, Berlin.
- [30] Émery, Michel and Zheng, Wei-An (1984). Fonctions convexes et semimartingales dans une variété. *Séminaire de probabilités (Strasbourg)*, Springer-Verlag **18**, 501–518.
- [31] Emery, M. et Mokobodzki, G. (1991). Sur le barycentre d’une probabilité dans une variété. *Séminaire de Probabilités XXV*, Lecture Notes in Mathematics **1485**, 220–223.
- [32] Fletcher, P. T., Lu, C., Pizer, S. M. et Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, **23**(8), 995–1005.
- [33] Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l’Institut Henri Poincaré*, **10**, 215–310.
- [34] Feydy, J. (2020). *Analyse de données géométriques, au delà des convolutions*. Thèse de doctorat, Université Paris-Saclay.
- [35] Fletcher, T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision*, **105**(2), 171–185.
- [36] Fletcher, P. T. (2013). Statistics on manifolds : Fréchet mean and geodesic regression. Lecture notes.
- [37] Gallier, J. et Quaintance, J. (2020). *Differential Geometry and Lie Groups : A Computational Perspective*. Springer, Geometry and Computing 12.
- [38] Gallot, S., Hulin, D. et Lafontaine, J. (2004). *Riemannian Geometry*. Springer-Verlag, Berlin–Heidelberg.
- [39] Guigui Nicolas, A. S. (2021). *Estimation statistique dans les variétés riemanniennes : implémentation et application à l’étude des déformations cardiaques*. Thèse de doctorat, Université Côte d’Azur.
- [40] Guigui, N., Maignant, E., Trouvé, A. et Pennec, X. (2021). Parallel transport on Kendall shape spaces. In Nielsen, F. et Barbaresco, F. (éds.), *Geometric Science of Information*, Lecture Notes in Computer Science. Springer, 103–110.
- [41] Groisser, D. (2004). Newton’s method, zeroes of vector fields, and the Riemannian center of mass. *Advances in Applied Mathematics*, **33**, 95–135.

- [42] Gordon, W. B. (1972). Convex functions and harmonic maps. *Proceedings of the American Mathematical Society*, **33**(2), 127–132.
- [43] Gawlik E. S. and M. Leok,(2018) *Interpolation on Symmetric Spaces via the Generalized Polar Decomposition*, Foundations of Computational Mathematics, vol. 18, no. 3, pp. 757–788, June 2018. doi : 10.1007/s10208-017-9353-0.
- [44] Herer, W. (1986). Espérance mathématique au sens de Doss d’une variable aléatoire à valeurs dans un espace métrique. *Comptes Rendus de l’Académie des Sciences de Paris, Série I*, **302**, 131–134.
- [45] Herer, W. (1988). Espérance mathématique d’une variable aléatoire à valeurs dans un espace métrique à courbure négative. *Comptes Rendus de l’Académie des Sciences de Paris, Série I*, **306**, 681–684.
- [46] Huckemann, S., Hotz, T. et Munk, A. (2010). Intrinsic shape analysis : Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica*, **20**, 1–58.
- [47] Hüper, K., Helmke, U. et Herzberg, S. (2010). On the computation of means on Grassmann manifolds. In *Proceedings of the Mathematical Theory of Networks and Systems (MTNS)*.
- [48] Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, **30**, 509–541.
- [49] Karcher, H. (2014). Riemannian center of mass and so called Karcher mean. arXiv :1407.2087.
- [50] Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proceedings of the London Mathematical Society*, **61**(3), 371–406.
- [51] Kendall, W. S. (1991). Convexity and the hemisphere. *Journal of the London Mathematical Society*, **43**(2), 223–261.
- [52] Kaneko, T., Fiori, S. et Tanaka, T. (2013). Empirical arithmetic averaging over the compact Stiefel manifold. *IEEE Transactions on Signal Processing*, **61**(4), 883–894.
- [53] Kaneko, T. et Tanaka, T. (2011). A method to compute averages over the compact Stiefel manifold. In *International Conference Proceedings*.
- [54] Maignant, E., Trouvé, A. et Pennec, X. (2023). Riemannian locally linear embedding with application to Kendall shape spaces. In Nielsen, F. et Barbaresco, F. (éds.), *Geometric Science of Information*, Lecture Notes in Computer Science. Springer, 12–20.
- [55] Pennec, X., Sommer, S. et Fletcher, T. (éds.) (2020). *Riemannian Geometric Statistics in Medical Image Analysis*. Academic Press.

- [56] Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds : Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, **25**, 127–154.
- [57] Pennec, X. (2015). Barycentric subspaces and affine spans in manifolds. In Nielsen, F. et Barbaresco, F. (éds.), *Geometric Science of Information*, Lecture Notes in Computer Science **9389**, 12–21. Springer.
- [58] Pennec, X. et Arsigny, V. (2013). Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups. In *Matrix Information Geometry*, 123–166. Springer.
- [59] Picard, J. (1994). Barycentre et martingales sur une variété. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, **30**(4), 647–702.
- [60] Udrişte, C. (1994). *Convex Functions and Optimization Methods on Riemannian Manifolds*. Kluwer Academic Publishers.
- [61] Wilson, B. et Leimeister, M. (2018). Gradient descent in hyperbolic space. arXiv :1805.08207.
- [62] Yang, L. (2011). *Medians of probability measures in Riemannian manifolds and applications to radar target detection*. Thèse de doctorat, Université de Poitiers.

Summary (English)

This thesis investigates barycenters and Fréchet means on Riemannian manifolds, aiming to generalize the classical notion of mean to data lying on nonlinear geometric spaces. Using convex and geodesically convex functions, it introduces the concept of convex barycenters within this geometric framework.

The theoretical results are applied to important matrix manifolds, including the Stiefel manifold, the Grassmann manifold, and the manifold of symmetric positive definite (SPD) matrices, which are widely used in statistical analysis and geometric data processing.

The thesis also develops numerical methods and geodesic algorithms for efficiently computing Riemannian barycenters, relying on the intrinsic geometry of manifolds.

A key application concerns financial risk analysis. Covariance matrices, representing dependencies between financial assets, are modeled as points on the SPD manifold. The Riemannian barycenter provides a robust estimation of the global market structure, enabling the assessment of structural risk and the analysis of financial market dynamics.

Résumé (Français)

Cette thèse étudie les barycentres et les moyennes de Fréchet sur les variétés riemanniennes, dans le but de généraliser la notion classique de moyenne aux données situées sur des espaces géométriques non linéaires. En s'appuyant sur les propriétés des fonctions convexes et géodésiquement convexes, elle introduit la notion de barycentre convexe dans ce cadre.

Les résultats théoriques sont appliqués à des variétés matricielles importantes, notamment la variété de Stiefel, la variété de Grassmann et la variété des matrices symétriques définies positives (SPD), largement utilisées en analyse statistique et en traitement de données géométriques.

La thèse propose également des méthodes numériques et des algorithmes géodésiques permettant de calculer efficacement les barycentres riemanniens, en exploitant la structure géométrique des variétés.

Une application majeure concerne l'analyse du risque sur les marchés financiers. Les matrices de covariance, représentant les dépendances entre actifs, sont modélisées comme des points sur la variété SPD. Le barycentre riemannien fournit alors une estimation robuste de la structure globale du marché, permettant d'évaluer le risque structurel, d'analyser l'évolution des corrélations et de mieux comprendre la dynamique des marchés financiers.

Mots-clés : barycentre riemannien, variété des matrices SPD, moyenne de Fréchet, analyse du risque financier, géométrie riemannienne

Keywords: Riemannian barycenter, SPD manifold, Fréchet mean, financial risk analysis, Riemannian geometry