

Université
de Toulouse

THÈSE

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline :

Mathématiques appliquées

École doctorale :

Mathématiques Informatique Télécommunications (MITT)

Unité de recherche :

Institut de Mathématiques de Toulouse (UMR 5219)

Présentée et soutenue par

Benjamin CHARLIER

le 8 décembre 2011

Titre :

Étude des propriétés statistiques des moyennes de Fréchet dans des modèles de déformations pour l'analyse de courbes et d'images en grande dimension

Directeurs de thèse :

Dominique BAKRY (Université Toulouse III Paul Sabatier)

Jérémy BIGOT (Université Toulouse III Paul Sabatier)

Rapporteurs :

Peter KIM (Université de Guelph)

Alain TROUVÉ (École Normale Supérieure de Cachan)

Examineurs :

Marc ARNAUDON (Université de Poitiers)

Fabrice GAMBOA (Université Toulouse III Paul Sabatier)

Remerciements

Charité bien ordonnée commence par ses directeurs de thèse. Merci Dominique et Jérémie pour avoir encadré mes divagations mathématiques durant ces quatre années et leurs avoir fait prendre la forme de ce mémoire. Je retiendrai la porte toujours (!) ouverte du bureau de Dominique ainsi que les longues heures de discussion du lundi après midi qui m'auront appris à me questionner et rester pugnace face aux difficultés du travail de recherche. Un grand merci aussi Jérémie pour m'avoir permis d'aller en Amérique latine pendant quelques mois à la découverte d'un continent et d'une langue dont j'ignorais tout jusqu'alors.

Je remercie bien sûr Peter Kim qui a bien voulu relire ce travail et traverser un océan pour venir écouter la soutenance. Un merci tout particulier à Alain Trouvé qui a relu plus qu'attentivement cette thèse et m'a suggéré de nombreuses et pertinentes améliorations ainsi que des pistes de réflexions.

Je suis reconnaissant envers Marc Arnaudon qui a accepté de faire parti du jury de cette thèse et de se déplacer sur Toulouse pour l'occasion. Merci enfin à Fabrice Gamboa qui me fais le plaisir d'examiner ce travail.

Je tiens à remercier Sébastien Gadat pour m'avoir accueilli quelques fois dans son bureau. Merci aussi à Monique Pontier sans qui je ne serai sans doute pas descendu à Toulouse. Je tiens aussi à remercier les personnes qui ont rendu ces quatre années de travail bien agréables. Je pense à Marie-Laure et Delphine que je considère plus comme des amies que des collègues et qui mériteraient de figurer au paragraphe suivant. Il y a aussi Agnès Requis, Françoise Michel, Marie-Line Domenjole toujours disponibles pour régler les formalités administratives. Je n'oublie pas les informatologues avec Christophe pour ses cours de bash... cours commencés avec Marc actuellement exilé en terre Nancéienne.

Sur une note plus personnelle :

Big up (!) à Julien, compagnon de bureau, avec qui, c'est une chose précieuse, j'ai pu discuter librement de maths en posant un nombre incroyable de questions stupides. Merci aussi d'avoir su sélectionner les meilleurs arabicas de la planète qui nous ont permis de venir à bout des matins difficiles. Une pensée pour Anne-Charline, Mélanie, Tiphaine, Badra et Mickael tous passés par ce même bureau. Il y a aussi Erwan, Matthieu, Mathieu, Maxime ($\times 2$), Laetitia, Laurent, Robin, Stefan, Jean-Luc, Jorge et les autres...

Une pensée pour toi SalVador, arrivé en même temps que moi sur Toulouse et trop vite reparti dans ton Chili natal, pour vous Jo' et Sylvia pour votre accueil sans faille toutes les fois où j'ai débarqué sans prévenir, pour toi Dominique pour la montagne, le feuilletage et le vélo, pour toi Gauthier pour

le ski, la grimpe et tout et tout, pour vous Sebagathe, Xavier, Marie, Yohan, Manue pour la colloc et toutes ces choses partagées... Et puis il y a celle avec qui le temps passe bien trop vite : merci Delphine d'être là !

Ces remerciements passent naturellement par les Pyrénées dans lesquelles j'occupe une bonne partie de mon temps. Dans la section montagne de la fac, cela serait bien long de citer tout le monde. Je me hasarde à une liste non exhaustive : Cécile, Claire, Yvan, Valérie, Joël, Jean, Médhi, Marion, Joanna, Ludovic, Lucile, Frédérique, Élise, Olivier, Stéphane, Antoine, Amine, Alexandre, Vanessa, Pierre... Toutes mes confuses aux absents.

À ceux qui sont loin, dans le plat pays, entre Orléans, Paris, Tours et Saint-Brissson-sur-Loire...

Enfin, je pense à mes parents qui, à défaut de m'avoir transmis leur orthographe, ont corrigé les nombreuses fautes de ce mémoire. À mes frères et leur famille !

Table des matières

Table des matières	iii
Table des figures	vii
Avant-propos	ix
Summary	xiii
Introduction	1
I.1 Un exemple réel : l'authentification de signatures	1
I.1.1 Présentation et description de la procédure	1
I.1.2 L'acquisition des données	2
I.1.3 Variabilité des données et déformations	3
I.1.4 Modéliser l'écriture	4
I.1.5 Contexte statistique	5
I.2 Une nouvelle classe de modèles de déformations	6
I.2.1 Définition	6
I.2.2 Exemples de déformations	7
I.2.3 Méthode d'estimation	10
I.2.4 Contributions des chapitres 1 et 2	11
I.3 Statistiques du premier ordre dans les espaces non euclidiens	12
I.3.1 Valeurs centrales dans les espaces métriques	12
I.3.2 Historique	13
I.3.3 Groupe agissant sur un ensemble	16
I.3.4 Espaces quotients et passage des structures au quotient	20
I.3.5 Sections et distances sur les sections	23
I.3.6 Exemples de distances	24
I.3.7 Contributions des chapitres 2 et 3	25
I.4 Discussion et retour sur l'exemple de l'authentification de signatures	25
1 Étude de la consistance des moyennes de Fréchet dans des modèles de déformations	29
1.1 Introduction	30
1.1.1 A statistical deformable model for curve and image analysis	30
1.1.2 Organization of the paper	33
1.2 The estimating procedure	34

1.2.1	A dissimilarity measure based on deformation operators	34
1.2.2	Previous work in mean pattern estimation and geometric variability analysis	35
1.3	Lower bounds for the estimation of the deformation parameters	35
1.3.1	Shape Invariant Model	35
1.3.2	General model	36
1.3.3	Application to the shifted curves model	36
1.4	Identifiability conditions	37
1.4.1	The shifted curves model	37
1.4.2	The general case	38
1.5	Consistent estimation in the shifted curves model	39
1.5.1	The random perturbations Z_j^n	40
1.5.2	Choice of the smoothed estimators	41
1.5.3	Consistent estimation of the random shifts	41
1.5.4	Consistent estimation of the mean pattern	42
1.5.5	A lower bound for the Fréchet mean	42
1.6	Notations and main assumptions in the general case	43
1.6.1	Smoothness of the mean pattern and the deformation operators	43
1.6.2	The preliminary smoothing step	43
1.6.3	Random perturbation of the mean pattern f by the Z_j 's	44
1.7	Consistency in the general case	45
1.7.1	Consistent estimation of the deformation parameters	45
1.7.2	Consistent estimation of the mean pattern	45
1.8	Numerical experiments for randomly shifted curves	46
1.8.1	A general gradient descent algorithm	46
1.8.2	Randomly shifted curves	47
1.9	Conclusion and perspectives	49
1.A	Proof of the results in Section 1.3	50
1.B	Proof of the results in Section 1.4	52
1.C	Proof of the results in Section 1.5	52
1.D	Proof of the results in Section 1.7	54
1.E	Technical Lemmas	57
2	Estimation consistante de la courbe moyenne dans des modèles de déformations	61
2.1	Introduction	62
2.1.1	A deformable model for statistical shape analysis	62
2.1.2	Main contributions	63
2.1.3	Organization	65
2.2	Group structure and Generalized Procrustes Analysis	65
2.2.1	The similarity group	65
2.2.2	Group action	66
2.2.3	Generalized Procrustes analysis and Kendall's shape space	68
2.3	Identifiability conditions	69
2.3.1	Random deformations	69

2.3.2	The deterministic criterion D	70
2.3.3	The constraint set	71
2.4	The estimating procedure	73
2.4.1	A dimension reduction step	73
2.4.2	Estimation of the deformation parameters	74
2.4.3	Procrustes analysis of smoothed shapes	75
2.5	Consistency results	75
2.5.1	Covariance structure of the error terms	75
2.5.2	Regularity conditions on f	76
2.5.3	Consistent estimation of the deformation parameters	76
2.5.4	Consistent estimation of the mean shape	78
2.6	Numerical experiments	78
2.6.1	Description of the data	78
2.6.2	Description of the procedure	79
2.6.3	Results: estimation of the mean pattern	81
2.A	Proofs	83
2.B	Technical Lemma	91
3	Condition nécessaire et suffisante d'existence de la moyenne de Fréchet sur le cercle	95
3.1	Introduction	96
3.1.1	Organization of the paper	97
3.2	Fréchet Mean	98
3.2.1	Previous work on the uniqueness of the Fréchet mean	98
3.2.2	Other definition of mean	100
3.3	Notations	100
3.3.1	The distance function	100
3.3.2	Normal coordinates	101
3.3.3	Probability measure on the circle	102
3.4	The Fréchet functional on the Circle	102
3.4.1	The derivative of the Fréchet functional	103
3.4.2	Local minimum of the Fréchet functional	104
3.5	A necessary and sufficient condition for the existence of the Fréchet mean	106
3.5.1	Main result	106
3.5.2	Proof of Theorem 3.8	107
3.5.3	Sufficient conditions of existence	108
3.6	Fréchet mean of an empirical measure	113
3.6.1	Consistency of the empirical Fréchet mean	113
3.6.2	Computation of the Empirical Fréchet mean	117
3.7	Conclusion	118
3.7.1	Flat Torus	118
	Conclusion	121

Bibliographie	123
----------------------	------------

Table des figures

I.1	Un stylo d'acquisition de données.	2
I.2	Action des rotations sur la sphère unité de $\mathbb{R}^3$	18
I.3	Action des changements de temps sur les fonctions de $[0, 1]$ dans $\mathbb{R}$	19
I.4	Application quotient	19
I.5	Action des homothéties du plan	20
I.6	Action des translations sur le plan	21
I.7	Action des translations sur le tore plat	22
I.8	Distances sur le cercle	25
I.9	Distance extrinsèque et distance intrinsèque	27
1.1	Identifiabilité dans les modèles de translations aléatoires	39
1.2	Simulations : exemples de données	47
1.3	Simulations : résultats dans le <i>Shape Invariant Model</i>	48
1.4	Simulations : résultats dans le modèle général avec bruit non stationnaire	48
1.5	Simulations : résultats dans le modèle général avec bruit stationnaire	49
2.1	Orbites de l'action du groupe de similarités du plan	67
2.2	Identifiabilité dans les modèles de déformations avec une action de groupe	72
2.3	Simulations : la courbe moyenne	79
2.4	Simulations : exemples de données avec le bruit blanc	80
2.5	Simulations : exemples de données avec le bruit stationnaire	80
2.6	Simulations : exemples de données avec le bruit corrélé	80
2.7	Simulations : résultats de l'estimation de la forme moyenne avec le bruit blanc	82
2.8	Simulations : résultats de l'estimation de la forme moyenne avec le bruit stationnaire	82
2.9	Simulations : résultats de l'estimation de la forme moyenne avec le bruit corrélé	82
3.1	Moyenne intrinsèque et moyenne extrinsèque	101
3.2	Points critiques de la fonctionnelle de Fréchet	103
3.3	Fonctionnelle de Fréchet de deux masses de Dirac antipodales sur le cercle	109
3.4	Exemple de distribution satisfaisant $P(\alpha, \varphi)$	110
3.5	Minima locaux de la fonctionnelle de Fréchet	117
3.6	Fonctionnelle de Fréchet sur le tore plat de dimension deux	119

Avant-propos

Présentation

Cette thèse porte sur l'analyse statistique de données sur lesquelles agissent des déformations. C'est un problème sur l'authentification de signatures manuscrites qui en a marqué le point de départ. L'apparition de capteurs électroniques miniaturisés a rendu possible l'acquisition de données à grande échelle dans des instruments de mesure de petite taille. Un stylo équipé d'accéléromètres a ainsi été mis au point, ce qui permet, lorsqu'on l'utilise, d'enregistrer un signal correspondant aux mouvements. Une signature peut alors être vue comme une courbe temporelle qui est fonction de l'accélération du stylo. En pratique, on n'observe pas de signaux continus car l'acquisition se fait de manière discrète et les données prennent la forme de matrices de grandes dimensions.

Comme l'expérience personnelle le suggère, la variabilité de la signature d'un individu est importante et dépend de multiples facteurs. Par exemple, la durée d'acquisition n'est jamais identique d'une signature à l'autre. De plus, on observe des décalages en temps avec certaines parties du geste qui sont effectuées plus ou moins rapidement. Outre ces variations en temps, c'est la forme même d'une signature qui est changeante. Le dessin est modifié par un geste un peu plus ample qu'à l'habitude, des motifs sont ajoutés ou au contraire enlevés. Comme il est classique en statistique, chaque signature sera vue comme la version perturbée d'une signature moyenne. Cet objet moyen, ou courbe moyenne, est une valeur centrale, caractéristique de l'ensemble des signatures d'un individu. Authentifier une signature revient alors à la comparer à la signature moyenne.

Pour décrire ce problème de manière rigoureuse, il convient de poser un modèle statistique. Cependant, les modèles classiques avec un seul terme d'erreur additif, comme le modèle linéaire gaussien, ne permettent pas de rendre compte des variabilités en temps et en espace décrites ci-dessus. Par contre, ces variations peuvent être interprétées comme le fait de déformations qui agissent sur la signature moyenne en plus du traditionnel terme de bruit. Les modèles statistiques de ce type sont appelés *modèles de déformations*. Ils peuvent être utilisés partout où les observations d'un phénomène physique se caractérisent par une structure commune sous-jacente modifiée par l'action de diverses transformations. C'est le cas, notamment, en imagerie médicale, en traitement du signal ou dans l'étude des phénomènes biologiques.

Au cours de ce travail, nous avons introduit une nouvelle famille générale de modèles semi-paramétriques de déformations. Une fois le modèle posé, le but est d'estimer les paramètres d'intérêts non observés qui sont de deux types, comme le suggère le terme semi-paramétrique. D'une part, nous cherchons à estimer le signal moyen qui est une fonction et donc un objet de dimension infinie.

D'autre part, et ce point de vue est moins classique, nous considérons les paramètres de déformations - de dimension finie - comme des variables aléatoires. Il s'agit alors d'estimer la valeur des déformations qui agissent sur chaque nouvelle réalisation. En effet, il peut être intéressant de discriminer les observations, non seulement en les comparant avec une courbe moyenne, mais aussi en comparant une déformation donnée à la moyenne des déformations. C'est pourquoi les deux premiers chapitres de cette thèse sont consacrés à l'étude de la consistance d'estimateurs de la courbe moyenne et des paramètres de déformations.

Plaçons nous dans le cas où l'ensemble des déformations possède une structure de groupe agissant sur un espace de fonctions donné. Deux signaux sont équivalents si une déformation permet de transformer l'un en l'autre. L'ensemble de ces classes d'équivalence définit un nouvel espace qui est appelé *espace quotient*. Sous certaines hypothèses, l'espace quotient hérite de structures de l'espace quotienté comme par exemple d'une topologie, d'une distance ou d'une métrique riemannienne. Mais, de manière générale, l'espace quotient n'est pas un espace euclidien. Lorsque l'on observe des données générées par un modèle de déformations, il peut n'être possible d'estimer le signal moyen qu'à une déformation près. En d'autres termes, on peut estimer la classe d'équivalence de la moyenne et non la moyenne elle-même. Ainsi, les estimateurs considérés prennent leurs valeurs dans des espaces quotients. Si ces espaces ne sont pas euclidiens, il faut pour étudier le comportement de ces estimateurs, préciser, dans ce nouveau cadre, les définitions des objets statistiques élémentaires. Dans ce travail, nous nous focalisons plus particulièrement sur les statistiques du premier ordre en étudiant le comportement d'estimateurs de type moyenne empirique.

De la même manière que l'on peut considérer, pour décrire un ensemble de points sur la droite réelle, la moyenne euclidienne ou la médiane, la notion de moyenne dans les espaces non euclidiens n'est pas unique. La *moyenne de Fréchet* d'une mesure généralise aux espaces métriques la définition de moyenne euclidienne. Elle peut être décrite comme le point de cet espace qui minimise la moyenne du carré de la distance aux points qui portent la masse. Les estimateurs de la courbe moyenne qui sont étudiés ici, sont des estimateurs de types moyenne de Fréchet. Par ailleurs, il est souhaitable de pouvoir mesurer des distances, car cela permet de quantifier les vitesses de convergence et ainsi de définir des intervalles de confiance. La distance apparaît donc comme la notion clé pour l'étude des modèles de déformations. Malheureusement, il n'existe pas toujours de distance canonique définie sur les espaces quotients, ou ce qui revient au même, il n'existe pas toujours de distance non triviale, entre deux signaux, qui soit invariante sous l'action d'un groupe. Ainsi, définir une distance entre les classes d'équivalences relève d'un choix qui va influencer le calcul de la moyenne. C'est ce point qui sera développé un peu plus en détail à la fin du chapitre introductif.

Enfin, étant donné un espace métrique muni d'une mesure de probabilité, la moyenne de Fréchet, quand elle existe, n'est pas nécessairement définie de manière unique. Dans le cas des variétés riemanniennes, il existe des conditions suffisantes pour assurer la bonne définition de la moyenne de Fréchet pour des mesure de probabilité à densité. L'idée est de s'assurer que la mesure de probabilité est suffisamment concentrée, mais ces conditions, qui portent sur le support de la mesure, sont très restrictives pour les applications pratiques. Au chapitre 3, nous étudions la moyenne de Fréchet sur le cercle unité muni de la distance angulaire. Bien que nous nous soyons restreint à un cas particulier, nous donnerons une condition nécessaire et suffisante d'existence et d'unicité pour des mesures de

Organisation et principales contributions

Ce mémoire comporte trois chapitres en plus du chapitre introductif et de la conclusion.

Introduction. Nous y présentons dans la partie I.1 le problème de l'authentification de signatures et le cadre mathématique qui lui est associé. Puis, dans la partie I.2, nous introduisons les modèles de déformations qui seront étudiés plus en détail aux chapitres 1 et 2. Un certain nombre d'exemples d'applications possibles de ces modèles sont aussi développés. Dans la partie I.3, nous effectuons un survol des statistiques du premier ordre dans les espaces non euclidiens, en lien avec le chapitre 3. Nous mettons l'accent sur les espaces quotientés par l'action de groupes de déformations. Nous rappelons les définitions d'une action de groupe et donnons des exemples. Nous présentons enfin diverses stratégies pour définir des distances invariantes sous l'action d'un groupe de déformations. Ces distances peuvent alors être utilisées pour calculer des moyennes dans les espaces quotients.

Les chapitres 1 et 2 sont complémentaires et portent tous deux sur l'estimation des paramètres d'intérêt dans les modèles de déformations. Le troisième chapitre est indépendant des deux premiers.

Chapitre 1. Nous y étudions les modèles de déformations généraux introduits dans la partie I.2 avec comme fil rouge l'exemple des courbes temporelles sur lesquelles agissent des translations aléatoires en temps. L'étude porte en particulier sur des estimateurs de type moyenne de Fréchet de la courbe moyenne ainsi que sur des estimateurs des paramètres de déformations. C'est l'étude de la convergence de ces estimateurs en fonction du nombre n de points de discrétisation de chaque courbe et du nombre J de courbes observées qui constitue la principale contribution de ce chapitre .

Dans la partie 1.3 nous déterminons des bornes inférieures de type Van-Tree qui permettent de montrer qu'il est impossible d'estimer de manière consistante les paramètres de déformations en augmentant seulement le nombre J d'observations et laissant le nombre n de points d'échantillonnage de chaque courbe temporelle observée fixe (Théorèmes 1.1, 1.2 et 1.3). Dans le cadre des courbes translatées aléatoirement, ceci implique que l'on ne peut pas reconstruire la courbe moyenne en augmentant seulement J et avec n qui reste fixé (Théorème 1.7).

Pour montrer des résultats (positifs) de consistances pour l'estimation des paramètres de déformations et de la courbe moyenne, nous devons préciser les conditions d'identifiabilité du modèle statistique (partie 1.4) et les hypothèses que doivent vérifier les déformations et les objets déformés (partie 1.6). Ainsi, nous pouvons donner dans la partie 1.7, des bornes supérieures sous la forme d'inégalités de concentrations qui montrent qu'il est possible, avec une double asymptotique en n et en J , d'estimer de manière consistante les paramètres de déformations (Théorèmes 1.5 et 1.9) ainsi que la courbe moyenne (Théorèmes 1.6 et 1.10). Des simulations numériques sur l'exemple des courbes translatées aléatoirement sont présentées dans la partie 1.8.

Chapitre 2. Nous y appliquons à un cas particulier les modèles de déformations très généraux qui ont été introduits au chapitre 1. Le groupe des similarités du plan agit sur les courbes planaires par rotation, translation et homothétie. Ainsi, les objets observés sont issus de l'échantillonnage d'une courbe plane bruitée sur laquelle agit une similarité. Ce type de modèle a déjà fait l'objet de nombreuses études en théorie statistique des formes.

Dans la partie 2.2, nous reformulons le problème en termes d'actions de groupes. On pourra aussi voir à ce propos la partie I.3 dans le chapitre introductif. La méthode d'estimation proposée est de type moyenne de Procruste avec une étape de lissage préalable, voir la partie 2.4. La principale contribution de ce chapitre est l'étude de l'influence du nombre de points k de l'échantillonnage et du nombre d'observations J sur l'estimation de la courbe moyenne. Nous montrons en particulier que ce type d'estimateur peut être consistant sous des hypothèses peu restrictives sur le bruit moyennant une étape de lissage (Théorème 2.1 et 2.13). Dans la partie 2.6 des simulations numériques illustrent ces résultats.

Chapitre 3. Il porte sur l'étude de la moyenne de Fréchet sur le cercle unité du plan muni de la distance de la longueur d'arc. La principale contribution de ce chapitre est l'énoncé d'une condition nécessaire et suffisante qui garantit l'existence et l'unicité de la moyenne de Fréchet sur le cercle pour des mesures de probabilité quelconques.

Rappelons que la moyenne de Fréchet est définie comme l'unique minimum global de la fonctionnelle de Fréchet, voir la partie I.3. Ainsi, pour étudier la bonne définition de la moyenne de Fréchet sur le cercle, nous calculons dans la partie 3.4 la dérivée de la fonctionnelle de Fréchet. Cela permet en outre d'étudier les points critiques de la fonctionnelle de Fréchet. Le résultat principal est énoncé dans la partie 3.5 : c'est un critère d'existence qui met en évidence le lien entre l'existence d'une moyenne de Fréchet pour une mesure et la comparaison de cette mesure avec la mesure uniforme sur le cercle (Théorème 3.8). De plus il nous permet de retrouver et d'étendre les conditions (suffisantes) classiques qui garantissent l'unicité de la moyenne de Fréchet. Par la suite, nous donnons également une condition suffisante, simple à utiliser en pratique et sans hypothèses restrictives sur le support de la mesure (Théorème 3.15).

Enfin, dans la partie 3.6, nous étudions le comportement de la moyenne de Fréchet empirique dont nous redémontrons la convergence vers la moyenne de Fréchet grâce à une inégalité de concentration. Un algorithme explicite de calcul de la moyenne de Fréchet empirique est aussi proposé.

Summary

In this thesis, we are concerned with the statistical analysis of data observed with extra nuisance deformations. In what follows, we suppose that the curves or images at hand are generated by a noisy ideal mean pattern on which act some deformations operators. The contributions of this work are twofold: we first study how one can estimate the parameters of interest in the so-called *statistical deformable models* and subsequently, we study some statistical properties of the *Fréchet means*. This work contains three chapters written in English plus an introduction and a conclusion written in French. This summary is a loosely English version of the *Avant-Propos* in which we described the organization of this thesis.

Introduction. In this chapter, we present the motivations and the statistical and geometric framework used in the last three chapters. Indeed, this work was first motivated by problems arising from the concrete application of handwritten signature authentication. Hence, in Section I.1 we present the authentication procedure and the mathematical problems that are associated with this particular problem. In Section I.2 we introduce a new class of deformable models that is studied in Chapters 1 and 2. We also give examples of possible applications for these models. In Section I.3, we discuss some properties of first order statistics in non-Euclidean spaces in relation with Chapter 3.

Chapters 1 and 2 are focused on the deformable models and complement each other. Chapter 3 is independent and is focused on the properties of the Fréchet mean.

Chapter 1. We introduce a new class of general deformable models to study high-dimensional curves or images. It is shown that an appropriate tool for statistical inference in such models is the notion of sample Fréchet means, which leads to estimators of the deformation parameters and the mean pattern. The main contribution of this paper is to study how the behavior of these estimators depends on the number n of design points and the number J of observed curves (or images).

In Section 1.3, we derive a lower bound for the quadratic risk of estimators of the deformation parameters. It shows us that it is not possible to estimate consistently the deformation parameters with n fixed and when J goes to infinity, see Theorems 1.1, 1.2 and 1.3. In the randomly shifted curves model these results imply that it is not possible to estimate consistently the mean pattern with an asymptotic in J only, see Theorem 1.7.

In Section 1.5 we derive consistency results for the Fréchet mean in the case (1.2) of randomly

shifted curves, see Theorems 1.5 and 1.6. In Section 1.6 and Section 1.7, we give general conditions to extend these results to the more general deformable model (1.5), see Theorems 1.9 and 1.10. Section 1.8 contains some numerical experiments.

Chapter 2. We consider a particular case of the general deformable models studied in Chapter 1. The similarity group of the plane acts on the planar curves by translations, rotations and scaling. Hence Chapter 2 is devoted to the study of the deformable model (2.1) associated to this framework. The main contribution of the work is to study the influence of the dimension k of the configurations and of the number J of observations on the performance on the estimation of the parameters of interest.

In Section 2.2, we describe the deformations acting on the mean pattern in term of a group action and we recall some properties on the similarity group. Then, we discuss the General Procrustes Analysis and we compare it to our approach. In Section 2.3 we discuss the identifiability conditions related to estimating procedure described in Section 2.4. The consistency results are given in Section 2.5, see Theorems 2.1 et 2.13. We also give some numerical experiments in Section 2.6.

Chapter 3. Let $(\mathbb{S}^1, d_{\mathbb{S}^1})$ be the unit circle in $\mathbb{R}^2$ endowed with the arclength distance. In this chapter, we give a sufficient and necessary condition for a general probability measure μ to admit a well defined Fréchet mean on $(\mathbb{S}^1, d_{\mathbb{S}^1})$, see Theorem 3.8. This criterion allows us to recover the already known sufficient conditions of existence and allows us to derive a new sufficient condition of existence of the Fréchet mean with no restriction on the support of the measure, see Theorem 3.15.

In Section 3.4, we give explicit expressions of the Fréchet functional and its derivative. It allows us to study the critical points of the Fréchet functional. Section 3.5 contains Theorem 3.8 and its proof. We also propose in this section a new criterion that ensures the well definiteness of the Fréchet mean. In Section 3.6, we study the convergence of the empirical Fréchet mean to the Fréchet mean, and we describe an algorithm to compute the empirical Fréchet mean.

Introduction

Ce chapitre introductif contient une synthèse des notions abordées dans ce mémoire. Il est rédigé en français et de manière assez informelle. Nous y présentons les motivations, le contexte et l'état d'esprit dans lequel les travaux des chapitres 1, 2 et 3 ont été entrepris. Nous rappelons en outre des définitions et des résultats et donnons des exemples qui complètent les chapitres suivants.

I.1. Un exemple réel : l'authentification de signatures

I.1.1. Présentation et description de la procédure

Définition. L'*authentification* de signatures consiste à mettre en place une procédure automatique visant à certifier à partir d'une signature manuscrite, que l'identité de l'individu qui signe est bien celle dont il se réclame. L'authentification diffère donc de l'*identification* (ou reconnaissance) de signatures qui, elle, consiste à retrouver, sans que l'utilisateur ait fourni d'identifiant, l'identité de cette personne à partir d'une base de données. Ainsi, le problème d'authentification apparaît plus simple que celui de l'identification car il se borne à vérifier l'appartenance d'un objet (la signature) à une sous-population (l'échantillon de signatures déjà certifiées) et non à identifier la sous-population à laquelle cet objet appartient.

Méthode. La société *Softwin*, qui est spécialisée dans la sécurité informatique et basée à Bucarest en Roumanie, a mis au point un stylo équipé de capteurs miniatures de mouvements (voir la Figure I.1). Lorsqu'on manipule le stylo, un signal résultant du mouvement est généré et transmis, pour analyse, à un ordinateur. On discrimine alors une signature, non seulement par la ressemblance du dessin sur le papier mais aussi, et surtout, par la forme caractéristique du geste réflexe effectué par la personne en train de signer. Le postulat de la méthode est que ce geste réflexe est plus stable dans le temps que le dessin de la signature, et qu'il est inimitable, même si l'on dispose d'un échantillon de la signature manuscrite.

Procédure en deux étapes. La première étape consiste à enregistrer un échantillon d'apprentissage de la signature d'une personne. Nous appellerons cette procédure phase d'enregistrement. Il est difficile de faire signer un utilisateur plus d'une dizaine de fois au moment de la phase d'enregistrement car le processus pourrait rebuter les utilisateurs. Un échantillon d'apprentissage est donc, typiquement,

composé de 5 à 10 signatures seulement. Il faut noter toutefois que cet ensemble de signatures, effectuées dans un laps de temps court, n'est pas nécessairement représentatif de toutes les variations « naturelles » du geste. En effet, d'importantes variations peuvent être dues à la fatigue, au stress, à la précipitation, *etc.* Le but de la phase d'enregistrement est alors d'extraire de l'échantillon d'apprentissage des invariants, c'est à dire des quantités stables, d'une signature à l'autre, pour un individu donné, et qui caractérisent cette signature. Ce procédé n'a pas besoin d'être instantané, c'est la procédure d'authentification qui doit l'être.

La deuxième phase, celle de l'authentification est faite à la volée. Il s'agit, étant donné une nouvelle signature, de déterminer si ce nouveau signal a bien été émis par la même personne que celle de l'échantillon d'apprentissage. Dans ce cas, l'authentification est dite *online* car elle utilise les données cinétiques du geste. Par opposition, on parle d'authentification *off-line* pour les procédures basées uniquement sur la forme du dessin, voir à ce propos [PS00]. C'est le cas, par exemple, lorsqu'on cherche à certifier l'authenticité d'un tableau à partir de la signature bien connue d'un peintre décédé.

I.1.2. L'acquisition des données

Stylo d'acquisition. L'enregistrement d'une signature se fait à l'aide d'un stylo spécial muni de deux types de capteurs : deux accéléromètres soudés sur un circuit imprimé dans le corps du stylo, ainsi que deux diodes infrarouge situées au niveau de la mine, voir la Figure I.1. Un câble relie le stylo à un ordinateur qui enregistre le signal émis par les différents capteurs. Ainsi, la phase d'enregistrement et d'authentification ne nécessite que la possession d'un stylo et un accès à un serveur informatique où sont stockés les algorithmes et les données déjà recueillies.

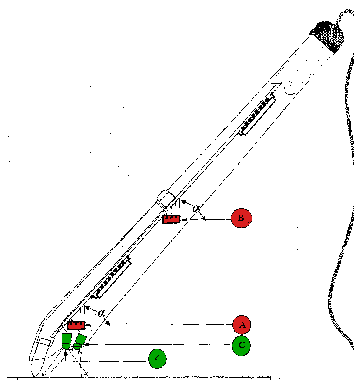


Figure I.1: Un stylo d'acquisition de données. Les accéléromètres MEMS sont en rouge et les diodes infrarouge (émetteur en 4 et récepteur en C) sont en vert. Source [MSDD06].

Mesure de l'accélération et de la position. Les accéléromètres miniatures sont des puces électroniques connues sous le vocable générique de MEMS pour *MicroElectroMechanical Systems*. Leur développement a débuté dans les années 1970, mais ce n'est que depuis le début des années 2000 qu'ils ont atteint des performances suffisantes, en terme de rapport qualité/prix, pour rendre possibles les applications grand public. Le signal enregistré par les deux accéléromètres est transmis à l'ordinateur sur quatre canaux qui correspondent à deux composantes orthogonales x_1, y_1 et x_2, y_2 de chaque MEMS. On dispose ainsi de deux courbes du plan dépendantes du temps et observées sur une grille

de discrétisation régulière. Le pas de discrétisation est de l'ordre de la milliseconde. Les signaux d'accélération sont entachés d'un bruit de fond important, dû à la grande sensibilité des capteurs qui sont capables, par exemple, de détecter les vibrations issues du contact entre la mine et les anfractuosités du papier. En fait, la plage des fréquences que peuvent mesurer les MEMS utilisés est bien plus grande que la plage de fréquences des mouvements caractéristiques de l'écriture qui, elle, est comprise entre 2 et 50 Hz. De ce fait, l'interprétation physique des courbes enregistrées via les MEMS présente quelques difficultés. Les deux MEMS enregistrent des signaux très similaires et leur disposition, à la manière de la stéréoscopie, doit permettre de calculer l'accélération des mouvements suivant la troisième composante. Néanmoins, il n'est évident que l'on puisse reconstruire le mouvement du stylo à partir de ces quatre signaux. En effet, les capteurs enregistrent, outre la force exercée par l'utilisateur, la résultante de la force de gravité sur les axes de mesure x_i et y_i , $i = 1, 2$. De plus, suivant la manière dont est tenu le stylo, les axes x_i et y_i ne sont pas nécessairement dans un plan horizontal.

Les diodes infrarouge (IR) transmettent un signal plus régulier qui donne la vitesse de la mine sur le papier. En intégrant une fois on peut donc en déduire des courbes de position. Une diode émet un rayon IR qui se reflète sur le support d'écriture et qui est capté ensuite par le récepteur - à la manière d'une télécommande de téléviseur. Le signal est enregistré suivant deux axes permettant, en principe, de reconstruire le trajet de la mine sur le papier. Mais ce signal se révèle peu utile en pratique car les changements d'orientation du stylo au cours de l'utilisation modifient l'angle d'incidence du rayon émis par la diode, ce qui déforme de manière importante le dessin enregistré. Ainsi, même si ces données ne nécessitent pas de lissage préalable, elles présentent d'importantes variations d'une signature à l'autre pour un individu donné. Il s'avère, de plus, que ce signal peut être plus facilement contrefait si l'on a accès à un modèle de la signature écrite.

I.1.3. Variabilité des données et déformations

Modèle de déformations. On suppose ici qu'une signature est une version aléatoirement perturbée d'un objet central idéal, que l'on appellera signature moyenne. Un modèle statistique traditionnel, expliquant la variabilité du phénomène uniquement par un terme d'erreur additif, ne serait pas pertinent. En effet, comment expliquer dans ce cas les différences de durée, les décalages en temps, les changements d'échelle, d'orientation ou l'aplatissement observés sur les données? On suppose donc qu'une partie des variations des observations est le fait de déformations qui agissent sur la signature moyenne qui peut, de plus, être bruitée par un terme additif aléatoire. On distinguera ici deux types de déformations qui agissent sur les courbes temporelles issues de l'acquisition d'une signature : les déformations en temps qui modifient la cinétique du geste, et les déformations en espace qui modifient la forme du dessin de la signature. Cette partie a pour but de répertorier les déformations susceptibles d'agir sur cette signature moyenne.

Les déformations en espace. Ce sont des transformations qui changent le dessin de la signature. On considère tout d'abord, les *déformations solides* qui agissent globalement sur les courbes. Elles permettent d'expliquer les variations dues aux changements de position de l'utilisateur ou de sa manière de tenir le stylo. Elles sont engendrées par les rotations, les changements d'échelle et les translations. On peut les voir comme l'action du groupe des similarités sur les courbes du plan. On peut aussi

considérer des changements d'échelle non-isotropes, c'est à dire, que l'on s'autorise un changement de taille différent dans des directions différentes. A noter qu'elles sont de dimension finie, au sens où il suffit d'un nombre fini de paramètres pour les décrire.

Naturellement, les déformations solides n'expliquent pas toute la variabilité de la forme d'une signature car elle agissent de manière globale sur les courbes. On parle de *déformations non-rigides* pour les déformations qui opèrent localement sur le signal. Typiquement, elles servent à modéliser un changement d'amplitude pour une boucle ou l'apparition d'un motif dans la signature. Comme leur action est plus localisée, ces familles de déformations sont de grandes dimensions, voire infinie dimensionnelle. Un exemple possible est la composition (à gauche) par les difféomorphismes du plan, voir la section I.3.3. Cependant l'ajout d'une boucle dans le dessin, ne peut être modélisé par des difféomorphismes qui sont, par définition, injectifs.

Les déformations en temps. Elles ne modifient pas la forme du dessin de la signature mais uniquement le rythme auquel est fait le geste. Typiquement, la longueur des signaux enregistrés varie d'une signature à l'autre. Ainsi, il y a une déformation globale qui agit sur la vitesse moyenne à laquelle est opéré le geste.

Comme pour les déformations en espace, la variabilité dans le rythme ne saurait être décrite par un simple changement d'échelle globale. Pour modéliser plus précisément les variations locales de la vitesse, il faut considérer des déformations beaucoup plus riches et qui agissent sur une partie seulement du signal. La reparamétrisation d'une courbe revient à composer (à droite) avec un difféomorphisme. L'ensemble des déformations possibles prend alors la forme de l'ensemble des fonctions réelles à valeurs réelles, croissantes, différentiables et d'inverses différentiables, qui est un espace fonctionnel de dimension infinie.

I.1.4. Modéliser l'écriture

Signature et écriture. L'authentification de signature est un problème qui peut être rapproché de la reconnaissance de l'écriture manuscrite. Avec l'apparition des moyens de numérisation dans les années 1970, la communauté informatique a commencé à s'intéresser à ce sujet connu sous le nom de *Handwriting recognition problem*. De nombreuses pistes et algorithmes ont ainsi été explorés, chacun répondant à une problématique bien précise, voir par exemple [Lor99] pour un état de l'art sur le sujet. L'utilisation de modèles aléatoires s'est ainsi répandue, avec en particulier l'utilisation de modèle utilisant les chaînes de Markov, sans que cela ait eut, pour autant, de résonance dans la communauté statistique. On peut tout de même citer [Goo91] pages 330 et 331, et [HKCF92] où le même auteur propose un modèle de déformations pour les signatures et une procédure de comparaison de données générées à partir d'une tablette graphique. En fait, l'intérêt de la communauté statistique pour les modèles de déformations s'est réellement manifesté avec le développement d'applications en imagerie médicale (*Computational anatomy*) et en biologie.

À la recherche d'invariants. Pour la reconnaissance de caractères comme les lettres ou les chiffres, [Lor99] propose diverses quantités considérées comme stables d'une observation à l'autre : le nombre de

points de rebroussement, le nombre de croisements, le nombre des points de courbure maximum. Cette méthodologie fait appel à des techniques d'extraction de points remarquables, appelés *landmarks*. Pour rester pertinente, cette étape doit être faite manuellement ce qui n'est pas envisageable dans le cas de grands échantillons. Il faut noter que, quand les données sont très similaires, on peut utiliser des techniques d'extraction automatiques de landmarks, voir par exemple [Big06a]. Malheureusement, il est pratiquement impossible de garantir la présence des même points caractéristiques sur tout un jeu de signatures d'un même individu. C'est ainsi que ce qui peut être vrai sur des lettrages d'une écriture manuelle ne se vérifie pas nécessairement sur des signatures. D'un point de vue statistique, une approche globale, comme celle proposée ci-dessous, sera en principe moins sensible aux variations locales des courbes.

Une modélisation délicate. Il semble difficile de modéliser une signature manuscrite car c'est un phénomène qui dépend fortement du contexte. Les variations de l'écriture sont dues à de multiples facteurs qui sont difficilement identifiables, [MSDD06]. Citons tout de même, l'aspect culturel (il existe des types de signatures différents de la même manière qu'il existe des alphabets différents), l'état d'esprit de l'utilisateur (la fatigue, l'énervement, le stress modifient la manière de signer), la complexité du geste (une signature est d'autant plus variable qu'elle est longue), l'évolution de la signature dans le temps (des parties seulement du signal évoluent, elle a tendance à « s'épurer »), l'acquisition des données (suivant la position de l'utilisateur, la signature peut changer de taille, d'orientation, être « aplatie », *etc*).

Le cas de l'authentification de signatures pose des difficultés particulières dans la modélisation. Un des principaux problèmes est de déterminer l'ensemble des déformations dont on veut supprimer l'effet. On souhaiterait ainsi ne pas tenir compte des variations « naturelles » du geste dues par exemple à la fatigue ou au stress. Le but est donc de définir des procédures statistiques invariantes pour un ensemble de déformations défini *a priori*. Il faut noter qu'en considérant un ensemble trop riche de déformations, on appauvrit l'ensemble possible des formes que peut prendre une signature. *A contrario*, en ne considérant que trop peu de déformations on n'arrive plus à expliquer la variabilité du phénomène.

I.1.5. Contexte statistique

Grandes dimensions. Des courbes ou des images observées avec des déformations peuvent être considérées comme des objet continus. En effet, une courbe du plan est la donnée de deux fonctions $(x(\cdot), y(\cdot))$ à valeurs réelles et définies sur un intervalle de $\mathbb{R}$. Une image en niveau de gris peut être vue comme une fonctions à valeurs réelles et définies sur un sous-ensemble du plan. Néanmoins, les données observées à l'aide de capteurs numériques sont toujours discrétisées et on considérera dans notre étude des modèles discrets. Avec les avancées techniques, la grille de discrétisation peut être très fine et les données de grandes dimensions.

Cadre non asymptotique. Comme le suggère l'exemple de l'authentification de signatures, le nombre de courbes que l'on observe peut être petit - de l'ordre de la dizaine. Ainsi, les résultats présentés au chapitre 2 et 3 sont non asymptotiques. Cela permet de donner des intervalles de confiance qui

dépendent du nombre de courbes observées et du nombre de points dans la grille de discrétisation. En faisant tendre ces paramètres vers l'infini on montre la convergence des estimateurs, ce qui permet d'étudier leur consistance.

Espaces non-euclidiens. L'action d'un groupe de déformations sur un espace de configurations induit une partition de cet espace en classes d'équivalence appelées orbites. On peut alors considérer l'espace quotient qui est formé de ces classes d'équivalence. Un exemple simple est le cercle unité du plan qui peut être vu comme le quotient du plan par l'action des changements d'échelle. Comme cet exemple nous l'apprend, ces espaces quotients perdent en général le caractère euclidien de l'espace de départ. Pour utiliser les outils statistiques élémentaires, il est alors nécessaire de préciser les définitions dans ce nouveau cadre.

1.2. Une nouvelle classe de modèles de déformations

1.2.1. Définition

Soit Ω un domaine de $\mathbb{R}^d$ avec $d = 1$ si l'on modélise des courbes temporelles et $d = 2$ ou 3 si l'on modélise des images bi ou tri-dimensionnelle. Ainsi, une courbe temporelle ou une image en niveaux de gris Y est la donnée d'une application $Y : \Omega \rightarrow \mathbb{R}$ de $L^2(\Omega)$, l'ensemble des fonctions de carré intégrable pour la mesure de Lebesgue. On suppose que les observations sont générées par le modèle de régression suivant,

$$Y_j = f_j + \varepsilon_j, \quad j = 1, \dots, J$$

où les f_j sont des fonctions aléatoires de $L^2(\Omega)$ et les ε_j sont des réalisations indépendantes et identiquement distribuées (i.i.d.) d'un processus stochastique centré ε . On suppose ici que chaque f_j est une version perturbée d'une forme moyenne $f \in L^2(\Omega)$ qui est déformée. Soit donc $(T_\theta)_{\theta \in \Theta}$ une collection d'opérateurs de $L^2(\Omega)$ dans $L^2(\Omega)$ indexée par Θ , un sous-ensemble de $\mathbb{R}^P$. Cette famille d'opérateurs est vue comme les déformations agissant sur la forme moyenne. On suppose que l'on a pour tout $j = 1, \dots, J$

$$f_j = T_{\theta_j^*}(f + Z_j)$$

où θ_j^* est une variable aléatoire à valeurs dans Θ , f est la forme moyenne et Z_j est une réalisations i.i.d. d'un processus aléatoire Z centré. Les deux termes de bruit Z et ε sont supposés indépendants et modélisent deux formes de perturbations différentes. Le terme ε est un bruit additif standard qui rend compte des erreurs de mesure faites lors de l'acquisition des données. Le terme Z est une perturbation de la forme moyenne f sur laquelle agit la déformation et qui modélise les variations de la forme moyenne au sein d'une même population. En pratique, les données acquises par des capteurs électroniques sont des données discrètes. Soit donc $(t_\ell)_{\ell=1}^n$ une grille de discrétisation non aléatoire de Ω , également appelée *design* déterministe. Ainsi, un modèle de déformations est un modèle statistique de la forme

$$Y_j^\ell = T_{\theta_j^*}(f(t_\ell) + Z_j(t_\ell)) + \varepsilon_j(t_\ell), \quad j = 1, \dots, J \text{ et } \ell = 1, \dots, n. \quad (1.1)$$

Pour alléger l'écriture on utilisera des symboles en gras pour les vecteurs. Ainsi, pour les versions discrétisées, on note $\mathbf{Y} = (Y_j^\ell)_{\ell=1}^n$, $\mathbf{f} = (f(t_\ell))_{\ell=1}^n$, $\mathbf{Z}_j = (Z_j(t_\ell))_{\ell=1}^n$ et $\boldsymbol{\varepsilon}_j = (\varepsilon_j(t_\ell))_{\ell=1}^n$ qui sont alors

des vecteurs de $\mathbb{R}^n$. Le but des deux premiers chapitres de ce mémoire est d'étudier l'estimation des paramètres non observés du modèle (I.1). C'est un modèle dit *semi-paramétrique* car les variables d'intérêt sont les paramètres de déformations θ_j^* qui sont supposés de dimension finie, et la forme moyenne f , qui appartient à un espace fonctionnel de dimension $+\infty$.

I.2.2. Exemples de déformations

Opérateurs de déformations. De nombreux exemples de déformations rentrent dans le cadre du modèle (I.1). On rappelle que l'ensemble des déformations $(T_\theta)_{\theta \in \Theta}$ est supposé pouvoir être paramétré par un ensemble de dimension finie $\Theta \subset \mathbb{R}^P$. Le cas des groupes de déformations infinis dimensionnels ne sera aborder que de manière très succincte au cours de ce chapitre. On peut classer les déformations suivant divers critères : possèdent ou non une structure de groupe, agissent de manière isométrique, *etc.* Dans ce qui suit, $L^2(\Omega)$ est l'ensemble des fonctions définies sur Ω et à valeurs réelles et telles que $\|f\|_{L^2} := \int_\Omega |f(t)|^2 dt < +\infty$. Étant donné un opérateur de déformation $T_\theta : L^2(\Omega) \longrightarrow L^2(\Omega)$, $\theta \in \Theta$, on note $\tilde{T}_\theta$ l'opérateur inverse. On a par définition

$$\tilde{T}_\theta T_\theta f = f, \quad \text{pour tout } f \in L^2(\Omega).$$

Groupes de déformations. La première grande famille de déformations que l'on considère est celle des actions de groupes, voir les définitions partie I.3.3. De manière informelle, c'est le cas lorsque la déformation d'une déformation est encore une déformation et lorsque la déformation inverse est aussi une déformation. On en donne quelques exemples.

Translations : c'est un modèle très simple de déformation qui agit sur les fonctions réelles, et qui sert d'exemple tout au long du chapitre 1 sous le nom de *shifted curves model*. Ce type de déformation a été largement étudié, voir [Røn01, GLM07, BG10a]. Soit donc $\Omega = [0, 1[$ et $f \in L^2_{per}([0, 1[)$, l'ensemble des fonctions périodiques de période 1 et de carré intégrable sur $[0, 1[$. Pour tout $\theta \in [0, 1[= \Theta$, les opérateurs de translation sont définis par

$$T_\theta f(t) = f(t - \theta), \quad \text{pour tout } t \in [0, 1[.$$

Il s'agit d'une action isométrique car $\|T_\theta f\|_{L^2} = \|f\|_{L^2}$. On peut voir $L^2_{per}([0, 1[)$ comme l'ensemble des fonctions définies sur le tore $\mathbb{R}/\mathbb{Z}$ et à valeurs réelles. Une translation est alors l'action de $(\mathbb{R}/\mathbb{Z}, +)$ sur les fonctions de $\mathbb{R}/\mathbb{Z}$ dans $\mathbb{R}$. Il s'agit donc d'une action de groupe compact et commutatif.

Transformations affines agissant à droite : Soit $\Omega \subset \mathbb{R}^d$ et $f : \Omega \longrightarrow \mathbb{R}$. On suppose que $\Theta \subset \mathbb{R}^2$ et pour tout $\theta = (a, b) \in \Theta$ on pose

$$T_\theta f(t) = f(e^a t + b), \quad \text{pour tout } t \in \mathbb{R}.$$

Ainsi, la famille d'opérateurs $(T_\theta)_{\theta \in \Theta}$ a une structure de groupe définie pour tout $\theta_1 = (a_1, b_1)$ et $\theta_2 = (a_2, b_2) \in \Theta$ par $T_{\theta_1} T_{\theta_2} = T_{\theta_2 \cdot \theta_1}$ où $\theta_2 \cdot \theta_1 = (a_1 + a_2, e^{a_2} b_1 + b_2)$. En fait ce groupe de déformations est isomorphe à $(\mathbb{R}, +) \rtimes_\varphi (\mathbb{R}, +)$ qui est un produit semi-direct associé au morphisme de groupe $\varphi : (\mathbb{R}, +) \longrightarrow (\text{Aut}(\mathbb{R}, +), \circ)$ qui est défini par $\varphi(a) = (b \mapsto e^a b)$. L'ensemble $(\text{Aut}(\mathbb{R}, +), \circ)$ désigne l'ensemble des applications linéaires de $\mathbb{R}$ dans $\mathbb{R}$ muni de la composition

◦ usuelle. Dans ce cas, les déformations ne sont pas isométriques et le groupe de déformations est non compact et non commutatif.

Transformations affines agissant à gauche : Dans l'exemple précédent, les déformations affines agissent « en temps », c'est à dire sur la variable $t \in \mathbb{R}$. On peut aussi faire agir le groupe affine sur les valeurs que prend f . Avec les mêmes notations que ci-dessus, on a

$$T_{\theta}f(t) = e^a f(t) + b, \text{ pour tout } t \in \mathbb{R}.$$

On parle d'action à gauche car on a maintenant pour tout $\theta_1 = (a_1, b_1)$ et $\theta_2 = (a_2, b_2) \in \Theta$, $T_{\theta_1}T_{\theta_2} = T_{\theta_1 \cdot \theta_2}$ avec $\theta_1 \cdot \theta_2 = (a_1 + a_2, e^{a_1}b_2 + b_1)$.

D'autres types de déformations rentrent dans ce cadre, en particulier si on suppose que f qui est définie sur $\Omega \subset \mathbb{R}^d$ est à valeurs dans $\mathbb{R}^m$, $m \in \mathbb{N}$. Citons l'exemple du groupe des déformations solides engendré par les rotations, les translations et les homothéties dans $\mathbb{R}^m$. Pour le cas $m = 2$, ce groupe est connu sous le nom de similarités du plan et pour tout $\theta = (a, \alpha, b) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^2$ on a

$$T_{\theta}f = e^a f R_{\alpha} + b \in \mathbb{R}^2, \text{ où } R_{\alpha} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

Ces déformations ont été étudiées dans la théorie statistique des formes (*Statistical shape analysis*), voir le chapitre 2. On peut également remplacer les rotations par un groupe de transformations plus riche comme par exemple le groupe spécial linéaire.

Action de groupes de Lie : Les exemples précédents peuvent être généralisés aux groupes de Lie de dimension finie capables d'agir à droite ou à gauche sur les fonctions $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$. Citons le cas particulier des fonctions définies sur un groupe de Lie compact $(\mathcal{G}, \cdot)$ de dimension d . On peut alors considérer l'ensemble $L^2(\mathcal{G})$ des applications définies sur $\mathcal{G}$ et à valeurs réelles et de carré intégrable pour la mesure de Haar du groupe $\mathcal{G}$. La mesure de Haar est la mesure sur $\mathcal{G}$ invariante pour l'action de $\mathcal{G}$ sur lui-même. Cela généralise l'invariance par translation de la mesure de Lebesgue. On a alors pour tout $\theta \in \Theta = \mathcal{G}$ un opérateur de translation à gauche défini par

$$T_{\theta}f(t) = f(\theta \cdot t), \text{ pour tout } t \in \mathcal{G}.$$

Des modèles de déformations similaires à (I.1) et qui utilisent ce cadre sont étudiés dans [BLV10].

Familles paramétriques de déformations. La structure de groupe apparaît comme une condition naturelle pour les déformations. Néanmoins, certains groupes de dimension infinie ne possèdent pas de sous-groupe fini dimensionnel non trivial. C'est par exemple le cas des difféomorphismes de la droite réelle ou du plan munis de la composition usuelle des fonctions. Il peut alors être utile de n'en considérer qu'un sous-ensemble paramétrique de dimension finie. Ce sous-ensemble n'est alors plus assez riche pour garantir la stabilité par la loi du groupe. Néanmoins, on s'assure de la stabilité par passage à l'inverse indispensable pour estimer les paramètres du modèle (I.1).

Reparamétrisation de courbes : On se place dans le cas où $\Omega = [0, 1]$ et on considère les fonctions de $f : [0, 1] \rightarrow \mathbb{R}$. Soit Θ l'ensemble des fonctions $\theta : [0, 1] \rightarrow [0, 1]$ continues, strictement croissantes, satisfaisant $\theta(0) = 0$ et $\theta(1) = 1$ et muni de la composition usuelle des fonctions.

Ainsi, les opérateurs $(T_\theta)_{\theta \in \Theta}$ agissent sur f par composition à droite d'un homéomorphisme θ , c'est à dire,

$$T_\theta f(t) = f(\theta(t)), \quad t \in [0, 1].$$

On parle d'actions en temps pour souligner qu'elle agissent sur le domaine de définition de f . Elles conservent la norme uniforme car on a $\sup_{t \in [0, 1]} |T_\theta f(t)| = \sup_{t \in [0, 1]} |f(t)|$ pour tout $\theta \in \Theta$ et toute fonction bornée $f \in L^2([0, 1])$. Naturellement, la norme L^2 de f est modifiée et on a $0 \leq \|f\|_{L^2} \leq \sup_{t \in [0, 1]} |f(t)|$. Ces déformations sont étudiées dans le domaine connu sous le nom de *dynamical time warping*, voir par exemple [WG97, RL01].

Une sous-famille paramétrique simple d'homéomorphismes de $[0, 1]$ est composée des fonctions strictement croissantes, continues, et affines par morceaux pour un nombre $P/2$ fixé de morceaux. Une telle fonction θ est caractérisée par une subdivision $0 < t_1 < \dots < t_{P/2} < 1$ de $[0, 1]$, *i.e.* les nœuds auxquelles elle change de pente, et par la valeur qu'elle prend en ces points $\theta(t_i)$, $i = 1, \dots, P/2$. Cette famille de dimension P , n'est pas stable par composition mais est stable par passage à l'inverse.

Composition à droite par des difféomorphismes : Soit Ω un sous-ensemble de $\mathbb{R}^d$ avec $d = 1, 2$ ou 3 . Un difféomorphisme θ de Ω dans lui-même agit sur les fonctions $f : \Omega \rightarrow \mathbb{R}$ par composition à droite, comme dans le cas des reparamétrisations de courbes. On note Θ l'ensemble des difféomorphismes de Ω et les opérateurs $(T_\theta)_{\theta \in \Theta}$ sont définis par

$$T_\theta f(t) = f(\theta(t)), \quad t \in \Omega.$$

La plus grande régularité des θ par rapport à l'exemple précédent permet de conserver le caractère différentiable des objets déformés. Les structures géométriques des espaces quotients associés à ce type de déformations sont étudiées depuis les années 1980, voir par exemple [CMM91]. Leur utilisation dans le champs statistique a été initié plus récemment par [Gre93]. C'est surtout en vision par ordinateur (*computer vision*) ou en traitement d'image que ce cadre s'est développé, voir un panorama des techniques dans [TY11, You10].

Il est possible de définir une famille paramétrique de dimension finie de tels opérateurs et qui est stable par passage à l'inverse. Une méthode populaire pour générer des difféomorphismes de $\Omega \subset \mathbb{R}^d$ dans Ω est de calculer le flot associé à un champ de vecteurs de Ω , voir par exemple [TY05a, BG10b]. Pour fixer les idées, on se place dans le cas où $d = 1$ et $\Omega = [0, 1]$. La méthode se généralise en dimension d quelconque. Un champ de vecteurs sur $[0, 1]$ est alors une fonction v de $[0, 1]$ dans $\mathbb{R}$ dont on suppose de plus que $v(0) = v(1) = 0$. Le flot ϕ_v^τ associé à v est défini comme la solution de l'équation différentielle ordinaire (E.D.O.),

$$\frac{d}{d\tau} \phi_v^\tau(t) = v(\phi_v^\tau(t)), \quad \text{où } t, \tau \in [0, 1].$$

La condition de nullité au bord sur v implique que pour tout $\tau \in [0, 1]$, l'application ϕ_v^τ est un difféomorphisme de $[0, 1]$ dans lui-même. On a $\phi_v^0 = Id$ et on note $\phi_v = \phi_v^1$ la solution au temps $\tau = 1$ de l'E.D.O. Étant données P fonctions $e_1, \dots, e_P$ de $[0, 1]$ dans $\mathbb{R}$ satisfaisant $e_p(0) = e_p(1) = 0$ pour tout $p = 1, \dots, P$, le difféomorphisme ϕ_θ associé à $\theta = (\alpha_1, \dots, \alpha_P) \in \mathbb{R}^P$ est $\phi_\theta = \phi_v$ avec $v = \sum_{p=1}^P \alpha_p e_p$. On a alors une famille de déformations de dimension P qui est indexée par θ et qui agit sur $L^2([0, 1])$ comme suit,

$$T_\theta f(t) = f(\phi_\theta(t)), \quad \text{pour tout } t \in [0, 1].$$

L'inverse de ϕ_{θ} est alors $\phi_{-\theta}$ mais étant donnés $\theta_1, \theta_2 \in \mathbb{R}^P$, il n'existe en général pas de $\theta_3 \in \mathbb{R}^P$ tel que $\phi_{\theta_2} \circ \phi_{\theta_1} = \phi_{\theta_3}$.

Composition à gauche par des difféomorphismes : On se place dans le cas où $\Omega = [0, 1]$ et où $f : [0, 1] \rightarrow \mathbb{R}^2$ est une courbe du plan. Soit Θ l'ensemble des difféomorphismes de $\mathbb{R}^2$ dans lui-même. La composition à gauche d'un $\theta \in \Theta$ est une déformation de la courbe $f([0, 1])$ notée :

$$T_{\theta}f(t) = \theta.f(t), \quad t \in [0, 1].$$

L'étude de ce groupe de déformations est connue dans la littérature sous le nom d'*elastic curves*, [You98]. Il faut aussi noter que ces déformations peuvent agir de manière locale contrairement aux exemples de groupes affines définis ci-dessus. Bien que ce groupe de déformations semble très riche, il ne permet pas de modéliser l'apparition d'une boucle dans la courbe. En effet, cela reviendrait à trouver une relation homéomorphe entre un cercle et un segment.

Pour définir une sous famille de $(T_{\theta})_{\theta \in \Theta}$ stable par passage à l'inverse, on peut utiliser la même méthode que pour les actions à droite des difféomorphismes, voir ci-dessus.

I.2.3. Méthode d'estimation

Le modèle de déformations (I.1) est un modèle semi-paramétrique dont on souhaite estimer les paramètres d'intérêt non observés. Ce sont les paramètres de déformations $\theta^* = (\theta_1^*, \dots, \theta_J^*) \in \mathbb{R}^{PJ}$ qui sont des variables aléatoires, et la courbe moyenne f qui est une fonction dans $L^2(\Omega)$ déterministe. On rappelle que le modèle (I.1) est un modèle discret : on observe J courbes, chacune étant évaluée au travers d'une grille de discrétisation déterministe comportant n points. On suppose aussi que la famille de déformations est connue *a priori*.

Heuristique. Pour comprendre la procédure d'estimation envisagée ici, il est bon de considérer une version sans bruit du modèle (I.1). On suppose que l'on observe, $Y_j^\ell = T_{\theta_j^*}f(t_\ell)$, $\ell = 1, \dots, n$ et $j = 1, \dots, J$, et que, pour simplifier le propos, les opérateurs inverses $\tilde{T}_{\theta}$ peuvent agir sur les versions discrétisées de $T_{\theta_j^*}f$. En pratique, il faut interpoler les données. Il est alors clair que la fonctionnelle $F_{j,n}$ définie sur $\mathbb{R}^{PJ} \times \mathbb{R}^n$ par

$$F_{j,n}(\theta, \mu) = \frac{1}{J} \sum_{j=1}^J \|\tilde{T}_{\theta_j} \mathbf{Y}_j - \mu\|^2,$$

s'annule en les vraies valeurs des paramètres, *i.e.* $\theta = \theta^*$ et $\mu = f$. Ainsi, en revenant au modèle (I.1) avec le terme de bruit et *modulo* les problèmes d'unicité, on peut définir les « estimateurs » $(\hat{\theta}, \hat{f})$ de (θ^*, f) par

$$(\hat{\theta}, \hat{f}) = \underset{\theta \in \Theta^J, \mu \in \mathbb{R}^n}{\operatorname{argmin}} F_{j,n}(\theta, \mu).$$

Comme à θ fixé, la fonctionnelle $F_{j,n}(\theta, \mu)$ atteint son minimum en $\mu = \frac{1}{J} \sum_{j=1}^J \tilde{T}_{\theta_j} \mathbf{Y}_j$ on a

$$\hat{\theta} = \underset{\theta \in \Theta^J}{\operatorname{argmin}} \frac{1}{J} \sum_{j=1}^J \|\tilde{T}_{\theta_j} \mathbf{Y}_j - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \mathbf{Y}_{j'}\|^2.$$

Il s'avère que l'on a $\frac{1}{J} \sum_{j=1}^J \|\tilde{T}_{\theta_j} \mathbf{Y}_j - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \mathbf{Y}_{j'}\|^2 = \frac{1}{2J^2} \sum_{j=1}^J \|\tilde{T}_{\theta_j} \mathbf{Y}_j - \tilde{T}_{\theta_{j'}} \mathbf{Y}_{j'}\|^2$. En d'autres termes, les $\hat{\theta}_j$ sont les déformations qui recalent au mieux les données entre elles et $\hat{f}$ est la moyenne

euclidienne des données recalées. Cette méthode a été étudiée en analyse de Procuste¹, voir [DM98]. Dans le modèle général, on utilise une étape de lissage préalable qui permet d'atténuer le niveau du bruit et permet d'assurer la convergence des estimateurs $\hat{\theta}$ et $\hat{f}$.

Étape de lissage. Avant d'estimer les paramètres du modèle (I.1), chaque observation $\mathbf{Y}_j$, $j = 1, \dots, J$ est au préalable lissée. On se donne donc un opérateur de lissage qui, à chaque vecteur $\mathbf{Y}_j$ de $\mathbb{R}^n$ associe une fonction $\hat{f}_j$ de $L^2(\Omega)$ régulière. Le niveau de régularité est donné par un paramètre réel que l'on choisit en fonction des observations. Bien entendu, dans la suite, nous définirons précisément les espaces fonctionnels considérés. Pour des raisons pratiques, nous utiliserons un lissage linéaire par des méthodes à noyaux. C'est le cas, par exemple, d'un filtre en fréquence calculé en utilisant la transformée de Fourier ou d'un lissage spline. Typiquement, la variance et le biais des estimateurs $\hat{f}_j$ de $T_{\theta_j^*} f$ dépendent du nombre n de points d'observation. Si la forme moyenne est suffisamment régulière, il est possible de garantir l'existence d'un compromis biais/variance pour les $\hat{f}_j$ qui assure la convergence de ces estimateurs quand n tend vers l'infini. L'importance de cette étape de lissage met en évidence le rôle déterminant de la dimension des objets à reconstruire, voir le chapitre 2 pour plus de détails.

Estimation des paramètres de déformations. Une fois les observations lissées, on définit

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \frac{1}{J} \sum_{j=1}^J \|\tilde{T}_{\theta_j} \hat{f}_j - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \hat{f}_{j'}\|^2$$

L'ensemble Θ sur lequel on optimise est un sous-ensemble de Θ^J . Il est choisi de façon à garantir l'identifiabilité du problème. On insiste sur le fait que l'estimateur $\hat{\theta}$ minimise aussi $\sum_{j,j'=1}^J \|\tilde{T}_{\theta_j} \hat{f}_j - \tilde{T}_{\theta_{j'}} \hat{f}_{j'}\|$. Ainsi, les $\hat{\theta}_j$ sont les déformations qui recalent au mieux les données lissées. Il est important de noter que les $\hat{\theta}$ dépendent du niveau de lissage.

Estimation de la forme moyenne. L'estimateur de la forme moyenne f est alors

$$\hat{f} = \frac{1}{J} \sum_{j=1}^J \tilde{T}_{\hat{\theta}_j} \hat{f}_j.$$

En d'autres termes, c'est la moyenne euclidienne des données lissées et recalées.

I.2.4. Contributions des chapitres 1 et 2

Les modèles de déformations décrits ci-dessus sont étudiés plus en détail dans les deux premiers chapitres de ce mémoire. Dans le chapitre 1 on étudie les modèles de type (I.1) de manière assez générale. Dans le chapitre 2 on se concentre sur le cas particulier courbes du plan sur lesquelles agissent les similarités. Dans les deux cas, l'étude porte sur le rôle des paramètres n (nombre de points de l'échantillonnage) et J (nombre de courbes observées) dans l'estimation des paramètres d'intérêt du modèle. On renvoie à l'avant-propos pour un résumé des contributions de ces deux chapitres.

1. S'écrit Προκρούστης en grec, Procuste en français et Procrustes en anglais...

I.3. Statistiques du premier ordre dans les espaces non euclidiens

Dans cette section nous utilisons des notions de géométrie différentielle et riemannienne. Pour les définitions nous nous référons à [Küh02] ou aux textes plus avancés [Hel01, Cha06]. On rappelle qu'une *distance* sur un espace $\mathcal{X}$ est une application $\mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+$ qui vérifie les trois conditions suivantes pour tout $x_1, x_2, x_3 \in \mathcal{X}$,

- i. $d(x_1, x_2) = 0$ si et seulement si $x_1 = x_2$,
- ii. $d(x_1, x_2) = d(x_2, x_1)$,
- iii. $d(x_1, x_2) \leq d(x_1, x_3) + d(x_3, x_2)$.

On peut parfois relaxer la condition i : on appelle *pseudo-distance* une application d qui vérifie les conditions ii et iii et qui s'annule sur la diagonale, *i.e.* $d(x_1, x_1) = 0$ pour tout $x_1 \in \mathcal{X}$.

I.3.1. Valeurs centrales dans les espaces métriques

Soit $\mathcal{X}$ un espace muni d'une mesure de probabilité μ . Une valeur centrale pour μ est un point de $\mathcal{X}$ rendant compte de la localisation de la distribution de probabilité. Nous proposons dans cette partie quelques définitions possibles et on dira qu'une valeur centrale est bien définie si elle existe et est unique.

Moyennes de Fréchet. Soit $(\mathcal{X}, d)$ un espace métrique muni de la distance d . Étant donnée une mesure de probabilité μ sur $(\mathcal{X}, d)$, on appelle *fonctionnelle de Fréchet* l'application définie sur $\mathcal{X}$ et à valeurs réelles

$$F_\mu(p) = \frac{1}{2} \int_{\mathcal{X}} d^2(x, p) d\mu(x), \quad p \in \mathcal{X}. \quad (I.2)$$

On dit que $m(\mu) \in \mathcal{X}$ est la *moyenne de Fréchet* de μ dans $(\mathcal{X}, d)$, si F_μ admet un unique minimiseur en $m(\mu)$, *i.e.* il existe un unique $m(\mu) \in \mathcal{X}$ tel que $F_\mu(m(\mu)) = \min_{x \in \mathcal{X}} F_\mu(x)$. Il semblerait que cette définition soit apparue pour la première fois dans [Car46] ou [Fré48].

Insistons sur le fait que la moyenne de Fréchet d'une mesure de probabilité μ dépend du choix de la distance d sur $\mathcal{X}$. Dans le cas où $\mathcal{X}$ est une variété riemannienne, [BP03] a introduit une terminologie permettant de distinguer deux cas particuliers de distance :

Moyenne intrinsèque : Soit $(M, \langle \cdot, \cdot \rangle)$ une variété munie d'un tenseur de métrique riemannien $\langle \cdot, \cdot \rangle$.

Rappelons que la métrique est la donnée d'un champ de formes bilinéaires sur le fibré tangent de M satisfaisant des hypothèses de régularité, on renvoie à [Küh02, Hel01] pour les définitions.

On note $d_{\langle \cdot, \cdot \rangle}$ la distance géodésique définie par la métrique $\langle \cdot, \cdot \rangle$. Étant donnée une mesure de probabilité μ sur M , la *moyenne intrinsèque* est la moyenne de Fréchet de μ dans $(M, d_{\langle \cdot, \cdot \rangle})$.

Moyenne extrinsèque : Soit E un espace euclidien muni d'une norme $\| \cdot \|$ et (M, g) une variété riemannienne. On appelle *plongement* de M dans E une application $\phi : M \rightarrow E$ lisse, dont la différentielle est injective en tout point, et qui définit un homéomorphisme de M dans $\phi(M)$. Cela implique, en particulier, que l'ensemble $\phi(M)$ est une sous-variété de E . Pour simplifier le propos, on suppose que le plongement ϕ est isométrique. Les théorèmes de plongement [Nas54] assurent qu'il existe toujours un tel plongement pour $E = \mathbb{R}^n$ pourvu que n soit suffisamment

grand. Étant donnée une mesure μ sur M , la *moyenne extrinsèque* est, s'il existe, l'unique minimiseur de la fonctionnelle de Fréchet définie sur M par

$$p \longrightarrow \int_M \|\phi(p) - \phi(x)\|^2 d\mu(x)$$

où $\|\cdot\|$ est la norme euclidienne de E . Autrement dit, la moyenne extrinsèque est la moyenne de Fréchet de (M, d_ϕ) , où $d_\phi(x, p) = \|\phi(x) - \phi(p)\|$, pour tout $x, p \in M$, est la distance E restreinte à $\phi(M)$. On peut aussi l'interpréter comme la projection sur M de la moyenne euclidienne de la mesure image $\mu \circ \phi^{-1}$ vue comme une mesure (singulière) dans E .

Un exemple simple de variété unidimensionnelle est le cercle unité du plan. Cet exemple est traité dans le chapitre 3 dans le paragraphe 3.2.2.

L^p -moyennes de Fréchet. Ce type de moyennes est étudié dans [Afs11]. Soit $(\mathcal{X}, d)$ un espace métrique et p un paramètre réel, $1 \leq p \leq \infty$. La fonctionnelle de Fréchet d'ordre p est

$$x \longmapsto \begin{cases} \frac{1}{p} \int_{\mathcal{X}} d^p(y, x) d\mu(y), & \text{si } 1 \leq p < \infty, \\ \sup_{y \in \text{supp } \mu} d(y, x), & \text{si } p = \infty. \end{cases} \quad (I.3)$$

où $\text{supp } \mu \subset \mathcal{X}$ est le support de la mesure μ . Si le minimiseur de (I.3) existe et est unique on l'appelle *L^p -moyenne de Fréchet*. La moyenne de Fréchet minimise par définition (I.2) et est donc la L^2 -moyenne de Fréchet. Dans le cas où $\mathcal{X} = \mathbb{R}$, il est bien connu que la moyenne euclidienne empirique est sensible à l'apparition de données extrêmes dans un échantillon. Lorsqu'on a besoin d'une valeur centrale plus robuste aux données aberrantes, aussi appelées *outliers*, il est alors d'usage de calculer la médiane. Dans la terminologie que nous venons d'introduire, la médiane dans $\mathbb{R}$ correspond à la L^1 -moyenne de Fréchet. C'est pourquoi, dans le cas où $(\mathcal{X}, d)$ est une variété riemannienne, la L^1 -moyenne de Fréchet est appelée médiane géométrique, voir [FVJ09]. Ainsi, le paramètre $1 \leq p \leq +\infty$ permet de régler la sensibilité de la moyenne aux valeurs extrêmes. On renvoie à [FVJ09] et les références citées pour une courte introduction aux statistiques robustes.

Moyennes de Fréchet non-métriques. C'est une généralisation de la moyenne de Fréchet définie en affaiblissant les hypothèses sur l'intégrande dans la fonctionnelle de Fréchet (I.2), voir [Huc10, Huc11]. Soit $\mathcal{X}$ un espace polonais et $\rho : \mathcal{X} \times \mathcal{X} \longrightarrow [0, \infty[$ une fonction continue. On dira que ρ est une fonction coût si ρ s'annule sur la diagonale, *i.e.* $\rho(x, x) = 0$, pour tout $x \in \mathcal{X}$. La ρ -moyenne de Fréchet est alors, s'il existe, l'unique élément de $\mathcal{X}$ qui minimise

$$p \longmapsto \int_{\mathcal{X}} \rho^2(x, p) d\mu(x), \quad p \in \mathcal{X}.$$

I.3.2. Historique

L'étude statistique de données à valeurs dans des espaces non euclidiens n'a cessé de se développer au cours du siècle dernier. Nous proposons ici un bref historique non exhaustif de travaux centrés sur la notion de moyenne. Voir aussi à ce propos [Pen06, Pen09, Afs11, MR93].

Espaces métriques. Les premiers travaux concernant les variables aléatoires à valeurs dans des espaces non-euclidiens remontent au premier tiers du 20^e siècle avec [Car46] et [Fré48]. Dans ces travaux on trouve, entre autre, une définition de moyenne pour les variables aléatoires à valeurs dans un espace métrique. L'idée est d'utiliser une propriété de minimisation vérifiée par la moyenne euclidienne usuelle et qui est indépendante de la structure vectorielle, *i.e.* de l'addition. Étant donnée une mesure de probabilité, « la » moyenne de Fréchet, ou de Fréchet-Cartan selon les auteurs, est l'*ensemble* des minimiseurs de la fonctionnelle (I.2).

Durant les années 1970, [Zie77] démontre une loi des grands nombres pour la moyenne de Fréchet empirique de variables aléatoires à valeurs dans des espace semi-métriques séparables et de diamètre fini. Il applique ce résultat à l'ensemble des k -uplets de points de $\mathbb{R}^2$, vus comme l'ensemble des matrices $k \times 2$ et muni d'une pseudo-métrique invariante sous l'action de certaines transformations du plan. Ce travail préfigure les travaux sur les espaces de formes de Kendall qui seront publiés près d'une décennie plus tard.

Sphères et quotients de sphères. On appelle *statistiques directionnelles* l'étude des variables aléatoires à valeurs dans la sphère unité d'un espace euclidien. Le développement de cette théorie a été motivé par des problèmes liés à la géologie ou la biologie comme on peut le voir dans [Fis93] ou [MJ00] qui présentent un bon panorama sur le sujet. La notion de moyenne utilisée dans la littérature des statistiques directionnelles est, le plus souvent, la moyenne extrinsèque. Bien que parfois instable, cette moyenne est facile à calculer et les conditions qui permettent d'assurer son existence sont bien comprises. On renvoie au chapitre 3 pour les définitions et des exemples. Notons que la moyenne intrinsèque est néanmoins utilisée : des conditions suffisantes d'existence de la moyenne intrinsèque pour des mesure finies sur la sphère unité de $\mathbb{R}^3$ sont données dans [BF01] en vue d'une application au calcul de splines sur la sphère.

C'est au début des années 1980 que se développe l'analyse statistique de formes (*Statistical shape analysis*), voir entre autres [KBCL99, DM98] et les références citées. C'est l'observation des alignements de pierres dans des sites archéologiques du sud du Royaume-Uni qui en a marqué le point de départ. L'idée est de considérer les k -uplets de points de $\mathbb{R}^n$, $n \geq 2$, modulo les effets de translations, de changements d'échelle et de rotations. Cette théorie des formes est dédiée à l'étude systématique des espaces de formes Σ_n^k introduits par D.G. Kendall dans [Ken84]. On note $\mathbb{S}_n^k$ l'ensemble des k -uplets de points de $\mathbb{R}^n$ de moyenne 0 et de norme 1. Cet espace est isomorphe à la sphère unité de dimension $(k-1)n-1$. Les effets de translations et de changements d'échelle ayant été supprimés, l'espace des formes devient alors le quotient de $\mathbb{S}_n^k$ par l'action du groupe spécial orthogonal $\mathbf{SO}(n)$, voir la section I.3.3 pour les définitions. On le note

$$\Sigma_n^k = \mathbb{S}_n^k / \mathbf{SO}(n). \quad (I.4)$$

Cet espace peut être muni de diverses distances issues par exemple des méthodes de Procuste, voir à ce propos la section I.3.5 de ce mémoire, la section 9.2 de [KBCL99], le chapitre 4 de [DM98] ou [Huc10]. Lorsque $n = 2$, Σ_2^k est l'ensemble des formes du plan. Cet espace possède une propriété remarquable : on peut le munir d'une métrique riemannienne qui le rend isomorphe à l'espace projectif complexe $PC(k-2)$. Les choses se compliquent lorsque $n > 2$, les espaces Σ_n^k ont alors des singularités et ils deviennent incompatibles avec la structure riemannienne, voir le chapitre 6 de [KBCL99].

Des modèles statistiques sur l'espaces de formes du plan Σ_2^k ont été introduits dans [Goo91]. La question de l'estimation de la forme moyenne à partir d'observations générées par ce modèle est alors soulevée, voir à ce propos le chapitre 2. Ce sont, à notre connaissance, les premiers modèles de déformations qui ont été étudiés. La consistance des estimateurs de type moyenne de Procuste empirique pour les modèles de [Goo91] est étudiée dans [Lel93, KM97]. Il s'avère que la moyenne de Procuste complète - *Full Procrustes mean* - dans Σ_2^k est une moyenne de Fréchet associée à une distance particulière. On peut alors appliquer la loi des grands nombres de [Zie77] et montrer la convergence des moyennes empiriques, voir [Le98, KBCL99]. De nombreux travaux ont suivi et montrent la convergence de moyennes de Fréchet empiriques associées à d'autres distances sur l'espace des formes [LK00, KL00, BP03]. Dans [BP05], les auteurs complètent l'étude des moyennes de Procuste empiriques en décrivant leur comportement asymptotique à l'aide d'un théorème de la limite centrale.

Lorsque $n = 3$, l'espace des formes est singulier et les méthodes développées dans [BP03, BP05] ne s'appliquent plus. Il faut attendre [Huc10] pour qu'une loi des grands nombres et un théorème de la limite centrale soient démontrés pour des moyennes de Procuste empiriques dans Σ_3^k .

Variétés riemanniennes. L'obstacle majeur à la bonne définition de la moyenne de Fréchet est qu'elle n'est pas en général définie de manière unique. Dans ce paragraphe, la moyenne de Fréchet désigne la moyenne intrinsèque d'une variété $(M, \langle \cdot, \cdot \rangle)$ donnée. Dans le cadre riemannien, les deux éléments qui jouent un rôle prépondérant dans la bonne définition de la moyenne de Fréchet sont les points de coupures (*cut locus*) et la courbure, [Kar77, Ken92]. Dès les années 1970, des conditions suffisantes pour garantir l'existence et l'unicité de la moyenne intrinsèque sur les variétés dont la courbure sectionnelle est minorée sont données dans [Kar77] et les références citées. Pour un historique de ce sujet et une extension de ces résultats on se reportera à [Afs11, Pen06]. Ces conditions garantissent que la mesure est suffisamment concentrée. Elle sont de deux sortes : une borne sur le support qui dépend de la courbure de la métrique g ou une condition de symétrie sur la densité de la distribution de probabilité, voir la section 3.2.1 pour plus de détails. Notons que ces conditions ont été appliquées aux espaces de formes Σ_2^k dans [Le01, Le04]. À partir de la fin des années 1980, c'est avec l'étude des processus stochastiques, et en particulier des martingales, définies sur des variétés, qu'est apparu le besoin de définir une notion de barycentre riemannien, [Ken90, Pic94, ÉM91, Arn95, CK99]. Suivant les propriétés désirées, la définition de ces barycentres peut varier d'un auteur à l'autre et partage parfois des propriétés communes avec la notion de moyenne de Fréchet. Dans les années 1990, des bornes de type Cramer-Rao sont utilisées pour mesurer les performances d'estimateurs de la moyenne de Fréchet sur des variétés voir par exemple [Hen91, OC95]. Au début des années 2000, une présentation unifiée des différents types de moyennes sur les variétés riemanniennes est exposée dans [BP03, BP05, Pen06]. Les auteurs y donnent une loi des grands nombres et étudient les propriétés asymptotiques des moyennes empiriques en énonçant un théorème de la limite centrale.

Citons enfin les travaux de généralisation de l'Analyse en composante principale pour des données à valeurs dans des variétés [HHM10a, HHM10b, Huc11].

Groupes de Lie. Depuis le début de la seconde moitié du 20^e siècle, de nombreux travaux se sont attachés à étendre les résultats de convergence aux variables aléatoires à valeurs dans des groupes

de Lie. La loi des grands nombres et le théorème de la limite centrale donnent des informations sur le comportement asymptotique de la moyenne empirique, *i.e.* une somme partielle renormalisée de variables aléatoires. Dans les groupes de Lie, la loi de composition du groupe prend la place de l'addition. Ainsi, dans certains cas favorables de groupes de Lie de dimension finie, il est possible d'établir des théorèmes de la limite centrale, voir par exemple [Gre63].

Dans le cadre des modèles de déformations, les déformations aléatoires sont des variables aléatoires à valeur dans des groupes de Lie, voir à ce propos [BLV10, CL09]. Pour finir, citons [Gre93, GM07] où est introduit la *Pattern theory* qui utilise des modèles infinis dimensionnels pour expliquer les variations de formes observées dans les phénomènes physiques. L'idée d'une action de groupes de dimension infinie sur un espace de configurations, lui aussi de dimension infinie, s'est largement développer depuis les vingt dernières années dans la communauté de l'imagerie médicale notamment, voir une revue dans [TY11]. Néanmoins, peu de traitements statistiques ont été réalisés dans cette direction, voir [AST07, AKT10].

1.3.3. Groupe agissant sur un ensemble

Action de groupe. Soit $(\mathcal{G}, \cdot)$ un groupe dont l'élément neutre est noté e et dont l'inverse d'un élément $g \in \mathcal{G}$ est noté g^{-1} . L'action de $(\mathcal{G}, \cdot)$ sur un ensemble $\mathcal{X}$ est la donnée d'une application

$$\begin{aligned} \mathcal{G} \times \mathcal{X} &\longrightarrow \mathcal{X} \\ (g, x) &\longmapsto g.x \end{aligned} \tag{1.5}$$

satisfaisant pour tout $x \in \mathcal{X}$ les deux conditions suivantes : $g_2.(g_1.x) = (g_2.g_1).x$ (associativité) et $e.x = x$. Il faut remarquer que l'on a abusivement utilisé le même symbole pour l'action de groupe et la loi du groupe. Ainsi à $g \in \mathcal{G}$ fixé, l'action de $\mathcal{G}$ sur $\mathcal{X}$ définit une application bijective de $\mathcal{X}$ dans $\mathcal{X}$. On peut alors considérer une action de groupe comme un morphisme entre le groupe $(\mathcal{G}, \cdot)$ et l'ensemble des bijections de $\mathcal{X}$ dans $\mathcal{X}$ muni de la composition usuelle. De manière informelle, cela signifie qu'un sous-ensemble des bijections de $\mathcal{X}$ dans $\mathcal{X}$ se comporte de la même manière que les éléments de $\mathcal{G}$. On a déjà vu dans la section I.2.2 quelques exemples d'actions de groupes finies et infinies dimensionnelles.

Dans la suite, on suppose que le groupe $\mathcal{G}$ possède une topologie ou d'une structure différentielle. On demande de plus que la structure de groupe soit compatible avec cette topologie (*resp.* la structure différentielle), c'est à dire que les opérations

$$\begin{aligned} \mathcal{G} \times \mathcal{G} &\longrightarrow \mathcal{G} & \text{et} & & \mathcal{G} &\longrightarrow \mathcal{G} \\ (g, h) &\longmapsto g.h & & & g &\longmapsto g^{-1} \end{aligned}$$

soient continues (*resp.* différentiables). On parle alors de *groupe topologique* (*resp.* de *groupe de Lie*). En outre, un groupe compact est simplement un ensemble compact pour sa topologie.

Orbite et stabilisateur. En théorie des groupes, pour caractériser un morphisme entre deux groupes, il est d'usage de définir son ensemble image et son noyau. De la même manière, étant donnée l'action d'un groupe $(\mathcal{G}, \cdot)$ sur un ensemble $\mathcal{X}$, on définit l'orbite et le stabilisateur d'un point x de $\mathcal{X}$ sous

l'action de $\mathcal{G}$. D'une part, l'*orbite* de $x \in \mathcal{X}$ est l'ensemble image de x sous l'action de $\mathcal{G}$ et on note

$$\mathcal{G}.x = \{g.x \in \mathcal{X}, g \in \mathcal{G}\}.$$

C'est donc un sous-ensemble de $\mathcal{X}$. On définit d'autre part le *stabilisateur* de x comme le sous-ensemble de $\mathcal{G}$ qui fixe x . On note alors

$$I_x = \{g \in \mathcal{G}, g.x = x\}.$$

On peut facilement voir que I_x est en fait un sous-groupe de $\mathcal{G}$ qui est parfois appelé *groupe d'isotropie* de x . Il est intéressant de noter que l'orbite d'un point « ressemble » au groupe $\mathcal{G}$. Par exemple, dans le cas d'une action de groupe de Lie compact sur une variété différentielle $\mathcal{X}$, l'orbite d'un point $x \in \mathcal{X}$ est difféomorphe au groupe $\mathcal{G}$ quotienté par le stabilisateur I_x . On renvoie à [Pal61] pour un énoncé rigoureux avec des hypothèses moins restrictives que la compacité et aux exemples ci-dessous pour une illustration.

Types d'action de groupe. Les actions de groupes peuvent jouir de nombreuses propriétés. On se contentera d'en citer quelques unes. Ainsi, une action peut être,

libre : c'est le cas lorsque le seul élément de $\mathcal{G}$ qui fixe les points de $\mathcal{X}$ est l'élément neutre e . On peut aussi dire que pour tout $x \in \mathcal{X}$, le stabilisateur de x est réduit à e . De manière équivalente,

$$\forall g \in \mathcal{G}, g \neq e, \exists x \in \mathcal{X}, g.x \neq x.$$

transitive : cela traduit le fait que deux points quelconques x et y de $\mathcal{X}$ peuvent être envoyés l'un sur l'autre par l'action de $\mathcal{G}$. Autrement dit, l'orbite d'un point est $\mathcal{X}$ tout entier. On a alors

$$\forall x, y \in \mathcal{X}, \exists g \in \mathcal{G}, g.x = y.$$

propre : c'est une condition plus technique qui sert à garantir certaines propriétés de l'espace quotient, voir la partie I.3.4. On suppose ici que $\mathcal{X}$ et $\mathcal{G}$ sont des espaces topologiques. Une action est dite propre si l'application

$$\begin{aligned} \mathcal{G} \times \mathcal{X} &\longrightarrow \mathcal{X} \times \mathcal{X} \\ (g, x) &\longmapsto (g.x, x) \end{aligned}$$

est propre. C'est à dire, si les images réciproques des compacts de $\mathcal{X} \times \mathcal{X}$ sont des compacts de $\mathcal{G} \times \mathcal{X}$. En particulier, les orbites générées par une action propre sont des fermés de $\mathcal{X}$, voir le chapitre 4 dans [Mic] pour plus de détails. Il faut enfin noter que si l'application (I.5) est propre, alors l'action est propre, voir le résultat 12-16 dans [Lee11].

isométrique : Soit $(M, \langle \cdot, \cdot \rangle)$ une variété munie de la métrique riemannienne notée $\langle \cdot, \cdot \rangle$ et sur laquelle agit un groupe $\mathcal{G}$. Soit $p \in M$, la métrique est notée, comme habituellement, $\langle V_p, W_p \rangle_p \in \mathbb{R}$ pour tous vecteurs tangents V_p et W_p dans $T_p M$ qui est l'espace tangent à M en p . Étant donnée une application $\phi : M \longrightarrow M$, on note $d_p \phi : T_p M \longrightarrow T_{\phi(p)} M$ la différentielle en p de ϕ . On dit alors que l'action de $\mathcal{G}$ est isométrique si on a quel que soit $p \in M$,

$$\langle V_p, W_p \rangle_p = \langle d_p(g.)V_p, d_p(g.)W_p \rangle_p, \quad \text{pour tout } V_p, W_p \in T_p M,$$

où l'on a noté l'application $(g.) : x \longmapsto g.x \in M$ pour tout $x \in M$. Cela implique en particulier que la distance géodésique $d_{\langle \cdot, \cdot \rangle}$ définie par la métrique $\langle \cdot, \cdot \rangle$ est équivariante pour l'action de $\mathcal{G}$, *i.e.* pour tout $x, y \in M$ et $g \in \mathcal{G}$ on a $d_{\langle \cdot, \cdot \rangle}(g.x, g.y) = d_{\langle \cdot, \cdot \rangle}(x, y)$. Pour plus de détails voir par exemple la section 2 de [HHM10b].

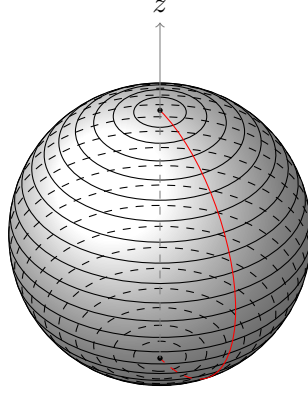


Figure I.2: Action des rotations autour de l'axe des z sur la sphère unité de $\mathbb{R}^3$. Les orbites de cette action correspondent aux cercles de latitude tracés en noir. Aux pôles nord et sud, les orbites sont réduites à un point - l'action n'est pas libre. Une section est donnée par un demi méridien, un exemple est tracé en rouge.

Exemples. Considérons la sphère unité $\mathbb{S}^2 = \{(x, y, z) \in \mathbb{R}^3, x^2 + y^2 + z^2 = 1\}$ de $\mathbb{R}^3$ sur laquelle agit $\mathbf{SO}(2, z)$, le groupe des rotations autour de l'axe z . Les orbites de cette action sont tracées sur la figure I.2. Le groupe $\mathbf{SO}(2, z)$ admet la représentation matricielle suivante,

$$\mathbf{SO}(2, z) = \left\{ \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \alpha \in \mathbb{R} \right\}.$$

C'est un groupe de Lie de dimension 1 qui est isomorphe à $\mathbf{SO}(2)$, le groupe spécial orthogonal du plan et qui peut aussi être identifié au cercle unité $\mathbb{S}^1$ du plan. Les pôles nord et sud, de coordonnées $(0, 0, 1)$ ou $(0, 0, -1)$ respectivement, sont les seuls points invariants de $\mathbb{S}^2$ par l'action de $\mathbf{SO}(2, z)$. On a $I_x = \{e\}$ pour tout $x \in \mathbb{S}^2 \setminus \{(0, 0, 1), (0, 0, -1)\}$ et $I_{(0,0,1)} = I_{(0,0,-1)} = \mathbf{SO}(2)$. Par ailleurs, les orbites de l'action de $\mathbf{SO}(2, z)$ sur $\mathbb{S}^2$ sont des cercles pour tous les points qui ne sont pas aux pôles et sont réduites à des points pour les pôles nord et sud. Elles sont ainsi difféomorphes à $\mathbb{S}^1$ ou $\mathbb{S}^1/\mathbb{S}^1 = \{e\}$. Cet exemple illustre bien le fait que les orbites ne sont pas de dimension constante. C'est précisément ce phénomène qui rend les espaces de Kendall Σ_n^k singuliers pour $n > 2$, voir à ce sujet le chapitre 6 de [KBCL99] et [Huc10].

On a déjà cité dans la section I.3.2 l'exemple de l'espace des formes où le groupe des similarités agit sur les k -uplets du plan. Cette action de groupe est traitée en détails au chapitre 2.

Considérons l'exemple des fonctions $f : [0, 1] \rightarrow \mathbb{R}$ de classe $\mathcal{C}^1$ sur lesquelles agissent les changements de temps, *i.e.* les reparamétrisations difféomorphes. On note $\mathcal{G}$ l'ensemble des difféomorphismes de $[0, 1]$ dans lui même, c'est à dire l'ensemble des fonctions $\theta : [0, 1] \rightarrow [0, 1]$ de classe $\mathcal{C}^1$ et d'inverse de classe $\mathcal{C}^1$ et satisfaisant $\theta(0) = 0$ et $\theta(1) = 1$. L'action de groupe se fait par composition à droite et tous les éléments dans l'orbite de f s'écrivent $f \circ \theta$ pour un $\theta \in \mathcal{G}$. On peut caractériser l'orbite d'une fonction dont la dérivée s'annule en un nombre fini de point $0 < t_1, \dots, t_k < 1$, $k \in \mathbb{N}$. On appelle *tableau de variation* de f le $(k+3)$ -uplet $TV(f) \in \{-1, 0, 1\}^{k+3}$ défini par

$$TV(f) = (\text{sgn}(f'(0)), \text{sgn}(f'(t_1/2)), \dots, \text{sgn}(f'((t_i + t_{i-1})/2)), \dots, \text{sgn}(f'((1 + t_k)/2)), \text{sgn}(f'(1))),$$

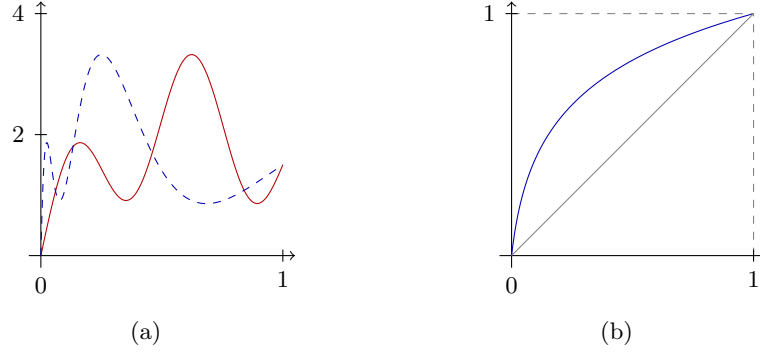


Figure I.3: (a) Deux fonctions de $[0, 1]$ dans $\mathbb{R}$ dont le tableau de variations est $(1, 1, -1, 1, -1, 1, 1)$. (b) Graphe du difféomorphisme $\theta : [0, 1] \rightarrow [0, 1]$ utilisé pour passer de la courbe rouge à la bleue pointillée.

où sgn est la fonction signe. Le tableau de variations décrit donc les changements de signe successifs de la dérivée de f , voir un exemple sur la figure I.3. L'orbite de f est alors $\{h : [0, 1] \rightarrow \mathbb{R}, h \in \mathcal{C}^1 : TV(h) = TV(f) \text{ et } h([0, 1]) = f([0, 1])\}$, c'est à dire les fonctions dont l'image et le tableau de variations sont identiques. Pour les fonctions très oscillantes, comme par exemple $x \mapsto x^2 \sin(\frac{1}{x})$, le tableau de variations peut être une suite infinie et certaines précautions doivent être prises. Le stabilisateur d'une fonction $f : [0, 1] \rightarrow \mathbb{R}$ de classe $\mathcal{C}^1$ est réduit à l'identité lorsque les zéros de la dérivée de f constituent un ensemble dénombrable. Dans le cas contraire, la fonction f est constante sur au moins un intervalle de $[a, b] \subset [0, 1]$. Le stabilisateur de f est alors inclus dans le sous-groupe de difféomorphismes qui fixe $[a, b]$ que l'on note $\mathcal{G}_{[a, b]} = \{\theta \in \Theta, \theta(a) = a \text{ et } \theta(b) = b\}$. En pratique, ce cas arrive fréquemment car il n'est pas rare que des signaux/images observés soient constants sur un sous-ensemble de leur domaine de définition. Ainsi l'action n'est pas libre car, dans ce cas, le stabilisateur n'est pas réduit à l'identité.

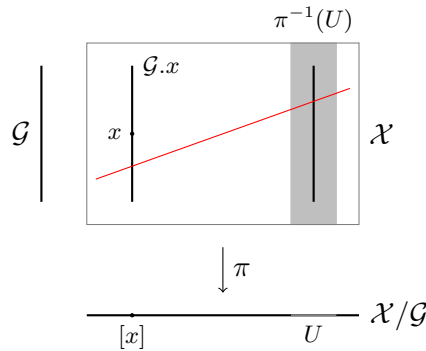


Figure I.4: Représentation de l'application quotient π . Le groupe $\mathcal{G}$ représenté à gauche agit sur $\mathcal{X}$ et l'image réciproque de $[x]$ est l'orbite $\mathcal{G}.x$ de x . Les points « au-dessus » de l'ouvert $U \subset \mathcal{X}/\mathcal{G}$ forment l'ouvert $\pi^{-1}(U)$ de $\mathcal{X}$. Une *section* correspond au choix d'un unique élément dans chaque fibre (en rouge).

I.3.4. Espaces quotients et passage des structures au quotient

Espace quotient. L'action d'un groupe $\mathcal{G}$ définit une relation d'équivalence $\sim$ (relation binaire réflexive, symétrique et transitive) sur $\mathcal{X}$. Pour tout $x, y \in \mathcal{G}$, on a $x \sim y$ si et seulement si x et y sont dans la même orbite, c'est à dire, il existe $g \in \mathcal{G}$ tel que $x = g.y$. L'ensemble des orbites définit un nouvel espace, appelé espace quotient, et noté $\mathcal{X}/\mathcal{G}$. L'application quotient

$$\begin{aligned}\pi : \mathcal{X} &\longrightarrow \mathcal{X}/\mathcal{G} \\ x &\longmapsto \mathcal{G}.x,\end{aligned}$$

est, par construction, surjective, voir la figure I.4. On notera aussi $[x] := \pi(\mathcal{G}.x)$ la classe d'équivalence de $x \in \mathcal{X}$ et, par extension un élément de $\mathcal{X}/\mathcal{G}$.

Distance quotient. On suppose que $(\mathcal{X}, d_{\mathcal{X}})$ est un espace métrique muni de la distance $d_{\mathcal{X}}$ et sur lequel agit un groupe $\mathcal{G}$. On dit que la distance $d_{\mathcal{X}}$ sur $\mathcal{X}$ est *équivariante* pour l'action de $\mathcal{G}$ si quels que soient $x, y \in \mathcal{X}$ on a

$$d_{\mathcal{X}}(x, y) = d_{\mathcal{X}}(g.x, g.y) \quad \text{pour tout } g \in \mathcal{G}.$$

Intuitivement, cela revient à dire que les orbites sont « parallèles » comme à la figure I.6. De plus, on dit que la distance $d_{\mathcal{X}}$ sur $\mathcal{X}$ est *invariante* pour $\mathcal{G}$ si c'est une pseudo-distance équivariante qui s'annule le long des orbites, *i.e.* satisfaisant

$$d_{\mathcal{X}}(x, y) = 0 \quad \text{si et seulement si} \quad y = g.x, \quad g \in \mathcal{G}.$$

Définir une distance sur $\mathcal{X}$ qui est invariante pour $\mathcal{G}$ est équivalent à définir une distance sur l'espace quotient $\mathcal{X}/\mathcal{G}$.

Lorsqu'elle est bien définie, on appelle *distance quotient* la distance $d_{\mathcal{X}/\mathcal{G}}$ sur l'espace quotient $\mathcal{X}/\mathcal{G}$ donnée par

$$d_{\mathcal{X}/\mathcal{G}}([x], [y]) = \inf_{x' \in \mathcal{G}.x, y' \in \mathcal{G}.y} d_{\mathcal{X}}(x', y'), \quad \text{pour tout } [x], [y] \in \mathcal{X}/\mathcal{G}. \quad (\text{I.6})$$

C'est donc la distance dans $\mathcal{X}$ entre les orbites de x et de y et, pour ainsi dire, le choix canonique de distance pour $\mathcal{X}/\mathcal{G}$. L'expression (I.6) étant définie comme un infimum, il est possible que $d_{\mathcal{X}/\mathcal{G}}$

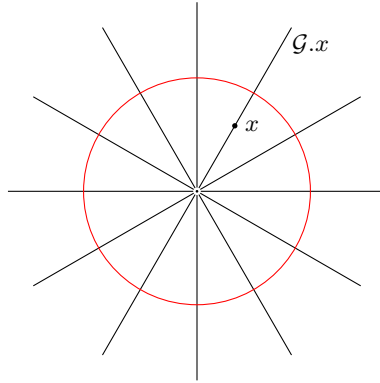


Figure I.5: Action des homothéties dans le plan. Les orbites sont les demi-droites issues de l'origine (en noir) et une section est donnée par le cercle unité (en rouge).

soit identiquement nulle, ce qui rend la définition de la distance quotient inutilisable pour nos besoins statistiques. Voir par exemple la figure I.5, où sont tracées les orbites de l'action de $(\mathbb{R}_+^*, \cdot)$ qui agit par changement d'échelle sur $\mathbb{R}^2$. Les orbites sont des demi-droites qui s'agglutinent à l'origine et la distance quotient est alors *dégénérée*.

On dispose d'un critère pour assurer la bonne définition de la distance quotient. Si la distance $d_{\mathcal{X}}$ sur $\mathcal{X}$ est *équivariante* pour l'action de $\mathcal{G}$ et que les orbites $\mathcal{G}.x$, $x \in \mathcal{X}$, sont des fermés de $\mathcal{X}$, alors (I.6) définit bien une distance sur $\mathcal{X}/\mathcal{G}$, voir la Proposition 12.1 de [You10]. Encore une fois, si le groupe $\mathcal{G}$ est compact, les orbites $\mathcal{G}.x$, $x \in \mathcal{X}$ sont nécessairement fermées dans $\mathcal{X}$ et une distance équivariante sur $\mathcal{X}$ pour l'action de $\mathcal{G}$ définira une distance quotient non nulle sur $\mathcal{X}/\mathcal{G}$, s'il existe au moins deux orbites distinctes. Cette condition de compacité n'est pas nécessaire, comme le montre l'exemple de $\mathcal{G} = (\mathbb{R}, +)$ qui agit par translation sur $\mathcal{X} = \mathbb{R}^2$ via $t + (x, y) = (x + t, y)$ pour tout $t \in \mathbb{R}$, $(x, y) \in \mathbb{R}^2$ et où l'action a été notée additivement, voir la figure I.6. Les orbites sont fermées mais ne sont pas compactes (droites parallèles) et la distance quotient est bien définie par $d_{\mathcal{X}/\mathcal{G}}([(x_1, y_1)], [(x_2, y_2)]) = |y_1 - y_2|$.

Dans les espaces de dimension infinie, comme par exemple l'espace des courbes paramétrées du plan sur lequel agissent les reparamétrisations difféomorphes, il peut être difficile de définir des distances quotients non dégénérées et, de surcroît, calculables explicitement. On donne quelques exemples dans la partie I.3.6. Il faut noter que ce sujet a fait l'objet de nombreux travaux au cours des deux dernières décennies et on peut voir à ce propos le chapitre 12 de [You10] ou [TY11] et les références citées.

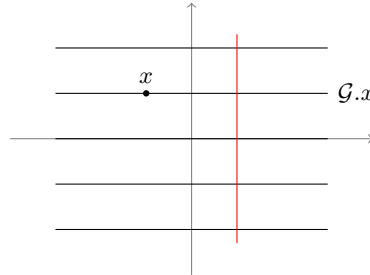


Figure I.6: Action des translations sur la première coordonnée des points du plan. Les orbites sont tracées en noir et une section est tracée en rouge.

Structures sur les espaces quotients. Dans ce paragraphe, on donne des conditions qui garantissent le passage de certaines structures de l'ensemble $\mathcal{X}$ (topologique, différentielle et métrique) à l'espace quotient $\mathcal{X}/\mathcal{G}$ au travers de l'application quotient π . Cette approche peut paraître, au premier abord, bien générale pour les besoins de ce travail. Néanmoins, comme cela est défini au paragraphe I.3.1, la généralisation des concepts de base des statistiques comme la moyenne est intimement liée à la notion de distance qui est sous-tendue par des notions topologiques et peut être pratique à utiliser dans un cadre riemannien. Malheureusement, comme nous allons le voir, même les structures les plus générales passent difficilement au quotient.

Topologie : On suppose ici que $\mathcal{X}$ et $\mathcal{G}$ sont des espaces topologiques. On appelle *topologie quotient* la plus petite topologie qui rend l'application quotient π continue. Une partie U de $\mathcal{X}/\mathcal{G}$ est un ouvert pour cette topologie si et seulement si son image réciproque $\pi^{-1}(U)$ est un ouvert de $\mathcal{X}$,

voir la figure I.4. On rappelle aussi qu'un espace métrique est un espace séparé, *i.e.* deux points distincts admettent toujours des voisinages disjoints. Par ailleurs, si un groupe $\mathcal{G}$ agit sur un espace séparé $\mathcal{X}$, la topologie quotient n'est pas nécessairement séparée. Pour illustrer ceci, on se place dans le cas où $\mathcal{X}$ est le tore $\mathbb{R}^2/\mathbb{Z}^2$. Soit $\lambda \in \mathbb{R}$ fixé, on considère l'action de groupe qui agit par translation et qui est définie par

$$t.(x, y) = (x + t, y + \lambda t), \quad \text{pour tout } (x, y) \in \mathbb{R}^2/\mathbb{Z}^2, \text{ tout } t \in \mathbb{R}.$$

Si λ est rationnel, les orbites « bouclent » et sont compactes dans $\mathbb{R}^2/\mathbb{Z}^2$, voir la figure I.7. Si λ est irrationnel, les orbites sont des sous-ensembles denses de $\mathbb{R}^2/\mathbb{Z}^2$ et on peut facilement voir que la topologie quotient n'est pas séparée. Ainsi, pour éviter les topologies quotients pathologiques, on suppose que l'action de $\mathcal{G}$ sur $\mathcal{X}$ est continue et propre. Dans ce cas, si $\mathcal{X}$ est un espace séparé alors $\mathcal{X}/\mathcal{G}$ est aussi séparé, voir la proposition 12.24 dans [Lee11]. En particulier, on peut simplifier ces hypothèses en considérant l'action d'un groupe compact $\mathcal{G}$ agissant sur un espace séparé $\mathcal{X}$. Cette action est propre et $\mathcal{X}/\mathcal{G}$ est alors un espace séparé.

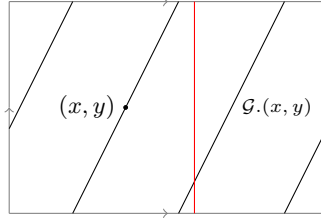


Figure I.7: Exemple d'action propre sur le tore plat de dimension 2. L'orbite de (x, y) est en noir et une section est en rouge.

Structure différentielle : On suppose maintenant que le groupe $\mathcal{G}$ possède une structure différentielle, voir [Lee00] pour un rappel des définitions. Soit donc $\mathcal{G}$ un groupe de Lie compact qui agit sur une variété différentielle $\mathcal{X}$ de manière différentiable, libre et propre. L'espace quotient $\mathcal{X}/\mathcal{G}$ admet alors une unique structure de variété différentielle de dimension $(\dim \mathcal{X} - \dim \mathcal{G})$, et l'application quotient π est une submersion. C'est l'objet de la Proposition 7.10 de [Lee00]. On peut affaiblir l'hypothèse d'action libre, au prix d'un passage au quotient par le groupe d'isotropie. Il faut alors utiliser le *Slice Theorem* de [Pal61] qui montre que les orbites $\mathcal{G}.x$ de $x \in \mathcal{X}$ sont localement difféomorphes à $\mathcal{G}/I_x$. A ce propos, voir la figure I.2 et la discussion associée ainsi que les figures I.5, I.6, I.7 et 2.1.

Structure riemannienne : On suppose que $(\mathcal{X}, g)$ est une variété riemannienne et que $\mathcal{G}$ est un groupe de Lie agissant sur $\mathcal{X}$. Comme on peut s'y attendre, l'espace quotient n'est pas nécessairement une variété riemannienne. Néanmoins si on suppose que $\mathcal{G}$ est un groupe de Lie compact qui agit proprement et isométriquement sur $\mathcal{X}$, alors $\mathcal{X}/\mathcal{G}$ peut être muni d'une métrique riemannienne, voir [HHM10b]. Les hypothèses demandées sont très restrictives en pratique. Malgré cela, l'espace des formes du plan Σ_2^k vu comme le quotient de la sphère $\mathbb{S}_2^k$ par le groupe compact $\mathbf{SO}(2)$ satisfait les hypothèses, voir (I.4). L'action est bien isométrique et l'espace quotient Σ_2^k muni de la distance de Procuste est isomorphe à l'espace projectif complexe de dimension $k - 2$, voir [KBCL99, HHM10b].

I.3.5. Sections et distances sur les sections

Comme on l'a vu au paragraphe I.3.4, les distances invariantes par l'action de $\mathcal{G}$ donnent un moyen de comparer des configurations d'un espace $\mathcal{X}$ modulo l'action d'un groupe $\mathcal{G}$. Il n'est cependant pas toujours possible de définir la distance canonique (I.6) sur l'espace quotient. Une autre définition possible de distance invariante par $\mathcal{G}$ requiert de choisir un représentant dans chaque orbite. On débute cette partie par une définition.

Section. On considère un espace $\mathcal{X}$ sur lequel agit un groupe $\mathcal{G}$. Une *section* est un sous-ensemble de $\mathcal{X}$ contenant un et un seul élément de chaque orbite. Il n'y a en général pas unicité car cela revient à choisir un « inverse » pour l'application quotient. Plus précisément, une section de $\mathcal{X}$ pour l'action de $\mathcal{G}$ est la donnée d'une application $s : \mathcal{X}/\mathcal{G} \rightarrow \mathcal{X}$ tel que $\pi \circ s = Id_{\mathcal{X}/\mathcal{G}}$. On a alors le diagramme commutatif suivant,

$$\begin{array}{ccc} & & \mathcal{X} \\ & \nearrow s & \downarrow \pi \\ \mathcal{X}/\mathcal{G} & \xrightarrow{Id} & \mathcal{X}/\mathcal{G} \end{array}$$

En pratique, on demande que s soit continue ou différentiable pour assurer une certaine régularité sur $s(\mathcal{X}/\mathcal{G})$. Par exemple, si $\mathcal{X}$ est une variété riemannienne, il est pratique de considérer des sections qui sont des sous-variétés de $\mathcal{X}$, voir par exemple [Mic]. Dans la suite on fera un abus de notation en utilisant le symbole s à la fois pour l'application et pour l'ensemble image $s(\mathcal{X}/\mathcal{G}) \subset \mathcal{X}$. Divers exemples de sections sont tracés - en rouge - dans les figures I.2, I.4, I.5, I.7 et 2.1.

Il est utile de définir g_s^x qui est l'élément de $\mathcal{G}$ amenant $x \in \mathcal{X}$ sur $s(\mathcal{X}/\mathcal{G})$, ce qui peut être noté

$$g_s^x.x = s([x]). \quad (I.7)$$

Considérons par exemple le cas de l'action des changements d'échelle dans le plan définie par $t.x = e^t x$, $t \in \mathbb{R}$, $x \in \mathbb{R}^2$ et représentée sur la figure I.5. Étant donnée une norme quelconque sur $\mathbb{R}^2$, la sphère de rayon $a > 0$ pour cette norme définit alors une section. En particulier, on pose $s = \mathbb{S}^1$, la sphère Euclidienne de rayon 1 (voir la figure I.5) on a $g_s^x = -\log(\|x\|)$ avec $\|\cdot\|$ la norme euclidienne.

Distances et fonctions coût sur les sections. Lorsque la distance quotient (I.6) sur $\mathcal{X}/\mathcal{G}$ est dégénérée, on peut définir d'autres distances $\mathcal{G}$ -invariantes en choisissant une section s des orbites. On a alors la distance suivante sur $\mathcal{X}/\mathcal{G}$,

$$d_s([x], [y]) = d_{\mathcal{X}}(s([x]), s([y])), \quad [x], [y] \in \mathcal{X}/\mathcal{G}. \quad (I.8)$$

C'est donc la distance entre les éléments de s dans $\mathcal{X}$. On peut aussi la voir comme une pseudo-distance sur $\mathcal{X}$ et on a $d_s(x, y) = d_{\mathcal{X}}(g_s^x.x, g_s^y.y)$, pour tout $x, y \in \mathcal{X}$. Dans l'exemple des homothéties qui agissent sur le plan, nous avons considéré le cercle unité $\mathbb{S}^1$ comme choix de section. La distance d_s est alors la distance de la corde, voir la figure I.8.

Lorsque $s(\mathcal{X}/\mathcal{G})$ est une sous-variété de $\mathcal{X}$, d_s définie en (I.8) est appelée distance extrinsèque associée à s , *i.e.* $d_{\mathcal{X}}$ est la distance de l'espace dans lequel $\mathcal{X}/\mathcal{G}$ est plongé via s . Dans ce cas, $s(\mathcal{X}/\mathcal{G})$

peut aussi être muni d'une métrique riemannienne $\langle \cdot, \cdot \rangle$ (voir des conditions suffisantes sur $\mathcal{X}$ et $\mathcal{G}$ dans la partie I.3.4) et que l'on note $d_{\langle \cdot, \cdot \rangle}$ la distance géodésique associée. On définit une distance intrinsèque sur $\mathcal{X}/\mathcal{G}$ par

$$d_r([x], [y]) = d_{\langle \cdot, \cdot \rangle}(s([x]), s([y])), \quad [x], [y] \in \mathcal{X}/\mathcal{G}. \quad (\text{I.9})$$

En résumé, une distance invariante pour $\mathcal{G}$ sur $\mathcal{X}$ peut être définie en deux temps. Tout d'abord, on choisit une section s , puis on considère une distance sur s vue comme un sous-ensemble de $\mathcal{X}$.

On peut relaxer les hypothèses qui font de d une distance et se contenter d'une fonction coût sur s , *i.e.* une application $\rho : \mathcal{X}/\mathcal{G} \times \mathcal{X}/\mathcal{G} \rightarrow \mathbb{R}^+$ s'annulant sur la diagonale. Étant donnée une section s , un exemple possible est

$$\rho([x], [y]) = \inf_{y' \in \mathcal{G}.y} d_{\mathcal{X}}(s([x]), y'). \quad (\text{I.10})$$

Il peut être souhaitable de conserver la symétrie $\rho([x], [y]) = \rho([y], [x])$ pour tout $[x], [y] \in \mathcal{X}/\mathcal{G}$ et pour cela on peut considérer $\tilde{\rho}([x], [y]) = \frac{1}{2}(\rho([x], [y]) + \rho([y], [x]))$.

I.3.6. Exemples de distances

Le cercle. Le premier exemple complète l'introduction du chapitre 3 et porte sur le cercle unité du plan $\mathbb{S}^1 = \{x = (x_1, x_2) \in \mathbb{R}^2, \|x\|^2 = x_1^2 + x_2^2 = 1\}$. C'est une sous-variété de $\mathbb{R}^2$ qui peut être vue comme l'image du tore plat $\mathbb{R}/2\pi\mathbb{Z}$ par le plongement isométrique $\phi : \theta \mapsto (\cos(\theta), \sin(\theta))$ défini pour tout $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Étant donnée une mesure de probabilité μ sur $\mathbb{S}^1$, on peut calculer la moyenne de Fréchet suivant différentes distances. On en donne trois exemples qui sont illustrés par la Figure I.8.

La longueur de la corde : C'est la distance définie par $d_\phi(x, y) = \|x - y\|$ pour tout $x, y \in \mathbb{S}^1$. Elle correspond à la distance (I.8) pour l'action des homothéties du plan. La moyenne de Fréchet de μ dans $(\mathbb{S}^1, d_\phi)$ est la moyenne extrinsèque associée au plongement de ϕ .

La longueur d'arc : C'est la distance définie par $d_{\mathbb{S}^1}(x, y) = 2 \arcsin(\frac{\|x - y\|}{2})$ pour tout $x, y \in \mathbb{S}^1$. C'est une distance issue d'une métrique riemannienne et qui correspond à (I.9). La moyenne de Fréchet de μ dans $(\mathbb{S}^1, d_{\mathbb{S}^1})$ est donc la moyenne intrinsèque.

Distance de Procuste : Considérons une troisième distance définie par $d_\rho(x, y) = \min_{a \in \mathbb{R}} \{\|e^a x - y\|\} = \min\{\sin(d_{\mathbb{S}^1}(x, y)), 1\}$, voir la figure I.8. Dans le cas précis de l'action des homothéties du plan, c'est une distance qui correspond à la formule (I.10). Cette distance est notamment utilisée en analyse de Procuste, voir la *full Procrustes distance* dans la partie 2.2.3 ou dans [DM98, KBCL99].

Les espaces des formes. Les espaces de Kendall Σ_2^k sont décrits en détail dans le chapitre 2 dans la partie 2.2.3. Une distance semblable à (I.10) est utilisée en analyse de forme pour comparer les *k-uplets* du plan identifiés à l'espace des matrices $\mathbb{R}^{k \times 2}$. On considère $\mathcal{X} = \mathbb{S}_2^k$, l'ensemble des configurations centrées et de norme 1 (voir la pre-shape sphere dans la partie 2.2.3) muni de la distance euclidienne de $\mathbb{R}^{k \times 2}$ et sur lequel agit le groupe $\mathcal{G}$ engendré par les homothéties et les rotations du plan. Ainsi, les différentes distances d_F (*full Procrustes distance*) et d_P (*partial Procrustes distance*) introduites dans la partie 2.2.3 sont similaires à l'exemple du cercle ci-dessus.

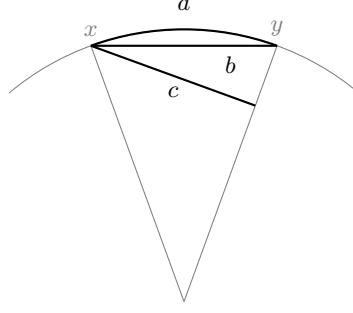


Figure I.8: Différentes distances entre deux points x et y sur le cercle $\mathbb{S}^1$. On a $d_\phi(x, y) = b$, $d_{\mathbb{S}^1}(x, y) = a$ et $d_\rho(x, y) = c$

Courbes planeaire. Considérons l'ensemble $\mathcal{X}$ des courbes différentiables du plan $f : [0, 1] \rightarrow \mathbb{R}^2$ sur lequel agit le groupe $\mathcal{G}$ des reparamétrisations par composition à droite d'un difféomorphisme de $[0, 1]$ dans lui même. L'espace $\mathcal{X}$ peut être muni de la distance $d_\infty(f, c) = \sup_{t \in [0, 1]} \{|f(t) - c(t)|\}$ pour tout $f, c \in \mathcal{X}$. Cette distance est équivariante pour l'action de $\mathcal{G}$ ainsi, la distance quotient

$$d_F(f, c) = \inf_{\theta_1, \theta_2 \in \mathcal{G}} d_\infty(f \circ \theta_1, c \circ \theta_2) \quad (\text{I.11})$$

est bien définie. La distance (I.11) est connue dans la littérature sous le nom de distance de Fréchet. Lorsque dans (I.11) on remplace la distance uniforme d_∞ par une distance de type $d_p(f, c) = \left(\int_0^1 |f(t) - c(t)|^p dt \right)^{\frac{1}{p}}$, $1 \leq p \leq +\infty$, on perd l'équivariance pour la composition par des difféomorphismes. De nombreuses autres distances invariantes sont par l'action des difféomorphismes sont décrites dans [You10, TY11].

I.3.7. Contributions des chapitres 2 et 3

Les notions abordées dans cette partie I.3 sont utilisées dans les chapitres 2 et 3. Au chapitre 2 nous appliquons le modèle de déformations (I.1) aux courbes du plan sur lesquelles agit le groupe des similarités. Cela permet de décrire en termes d'actions de groupe l'analyse de Procuste. Du point de vue des statistiques cela permet de comparer l'approche développée dans ce travail avec l'analyse de Procuste. Dans le chapitre 3, nous étudions des conditions d'existence et d'unicité de la moyenne de Fréchet sur le cercle unité $(\mathbb{S}^1, d_{\mathbb{S}^1})$ avec objectif de donner une condition qui ne borne pas le support des mesures. Un résumé et le détail des contributions figurent dans l'avant-propos et au début de ces deux chapitres.

I.4. Discussion et retour sur l'exemple de l'authentification de signatures

Rappelons que la signature manuscrite d'un individu peut être vue comme une courbe du plan (la courbe moyenne) bruitée et sur laquelle agissent des groupes de déformations en temps et en espace. Au vu des parties I.1, I.2 et I.3, l'idée est de considérer les différentes réalisations de la signature d'un individu comme appartenant à des orbites très proches (idéalement la même orbite) pour l'action de ces groupes de déformations. La procédure d'authentification reviendrait alors à

calculer une distance entre orbites et à définir un seuil pour discriminer les faux éventuels. Nous avons déjà évoqué l'exemple des courbes du plan au long de ce chapitre introductif. Nous discutons plus particulièrement ici la pertinence d'un modèle de déformations du type (I.1) et des limitations de la modélisation mathématique pour des problèmes aussi complexes que l'authentification des signatures.

Groupe de déformations Comme nous l'avons vu dans la partie I.1, la modélisation d'un processus cognitif comme la signature manuscrite est un problème complexe. On ne dispose pas de modèle physique permettant de décrire *a priori* le type de variations naturelles d'une signature d'un individu. Il est important de noter que ces variations peuvent être très importantes pour certaines signatures. Cela rend d'autant plus difficile l'isolement de ces déformations des variations extraordinaires dues au geste d'un autre individu.

En dimension finie et lorsque l'action de groupe est libre, la dimension de chaque orbite est la même que celle du groupe, voir la partie I.3.3. Par conséquent, on appauvrit l'espace quotient lorsqu'on augmente la « taille » du groupe de déformations. Ainsi, s'il est tentant de se donner un groupe de déformations de grande dimension pour expliquer les fortes variations de l'échantillon d'apprentissage, on risque de ne plus discriminer aucun faux.

Une illustration du problème du choix des déformations est donnée par l'action des changements de temps sur les courbes planaires. Il faut noter que de nombreux travaux portent sur l'étude de ce type de quotient, [TY11, You10, MM06] et les références citées. Néanmoins, l'idée de la démarche d'authentification *online* est de discriminer les signatures manuscrites sur la cinétique du geste et non sur le seul dessin sur le papier. Ainsi, quotienter l'ensemble des courbes du plan par l'action des difféomorphismes de $[0, 1]$ dans lui-même fait perdre toute l'information sur la cinétique du geste. Le groupe de déformations devient alors trop riche et il faudrait alors considérer un sous-groupe de difféomorphismes qui convienne.

Pour les structures quotient, le cas extrême se produit lorsque que l'on considère une action transitive. En effet, on a alors, par définition, une seule orbite et toute signature peut alors être vue comme la version déformée d'une autre signature. Même s'il s'agit d'un autre point de vue que celui développé dans ce travail, on peut utiliser ce cadre pour définir des distances entre les courbes. Cette approche est largement développée en analyse d'image et en *computational anatomy*, voir un panorama de ces techniques dans [TY11, You10]. L'idée est de définir une fonction d'énergie sur les déformations, et de prendre comme distance l'énergie nécessaire pour passer d'une courbe à une autre. Cette construction sous-tend une structure riemannienne pour espace des configurations. Il faut enfin noter que de telles distances peuvent être dégénérées et que l'ajout d'un terme de pénalité bien choisi permet de rendre la distance non-triviale.

Le choix des normalisations La méthode d'estimation présentée dans la partie I.2 peut être résumée comme suit : dans un premier temps, on normalise les données puis on calcule la moyenne euclidienne de ces données recalées. La normalisation peut se faire de multiples manières et cela revient à choisir une section s des orbites des courbes sous l'action des déformations, voir la partie I.3.5. La procédure de normalisation revient alors essentiellement à calculer la déformation amenant chaque configuration sur s , voir (I.7). La normalisation peut être choisie indépendamment des données observées ou, au

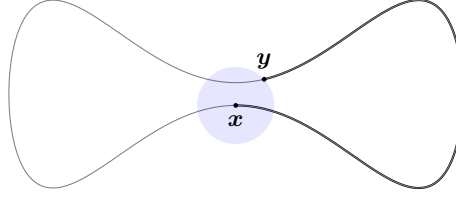


Figure I.9: Exemple de sous-variété du plan. La distance intrinsèque (riemannienne) entre x et y est égale à la longueur de la courbe noire. La distance extrinsèque (euclidienne) est égale à la longueur du rayon du disque bleu.

contraire, en dépendre (*data-driven*).

Par exemple, dans le cas des courbes du plan sur lesquelles agissent les similarités, la littérature montre que plusieurs stratégies ont été développées. L'une d'elle est connue sous le nom de coordonnées de Bookstein en analyse de Procruste. Elle revient à choisir deux points caractéristiques sur la courbe et à les ramener sur $(0, 0)$ et $(1, 0)$ respectivement. Un choix simple est de considérer $f(0)$ et $f(1)$ pour une courbe $f : [0, 1] \rightarrow \mathbb{R}^2$. Un autre exemple est donné avec la *pre-shape sphere* dans la partie 2.2.3.

Le choix des sections va fortement influencer la distance associée. En effet, dans le cas où l'on munit les sections de la distance extrinsèque (I.8), c'est la forme de la section qui influe sur la distance. Ce point est soulevé dans [DM98] où les auteurs font la comparaison entre différents types de coordonnées en analyse de forme plane. Intuitivement, dans le cas des coordonnées de Bookstein, un rôle particulier est donné à deux points. Ainsi, de petites variations sur ces points peuvent induire une grande distance sur les versions normalisées. Cela peut aussi induire artificiellement des corrélations entre les observations.

Distance entre les courbes On a défini dans la partie I.3.6 la distance de Hausdorff qui est un exemple de distance invariante pour l'action des changements de temps. Hormis le fait que cette action de groupe n'est sans doute pas adaptée à l'authentification de signatures *online*, cette distance basée sur la norme uniforme est bien trop sensible aux petites variations des courbes. Il faut alors comparer les courbes « en moyenne » en considérant des normes L^p , $1 \leq p < \infty$.

Un point de vue largement partagé dans la littérature est de considérer l'ensemble des courbes comme une variété riemannienne, [KSMJ04, TY11, You10]. On appelle distances intrinsèques ou distances riemanniennes les distances géodésiques issues d'un tenseur de métrique riemannien. Elles sont du type (I.9). De manière informelle, elles sont définies comme l'infimum des sommes de petites déformations infinitésimales qui permettent de passer d'une configuration à une autre. On a alors, en quelque sorte, un « continuum » de configurations qui permet d'aller d'un point à un autre, voir la figure I.9. Cette définition présente l'avantage de munir l'espace des configurations d'une structure géométrique mathématiquement satisfaisante et ainsi de pouvoir utiliser les résultats existants sur les variétés. En analyse d'image, on aime à considérer comme plus proches deux configurations qui diffèrent par une transformation lisse plutôt que deux configurations qui diffèrent par des variations plus chaotiques réparties le long de la courbe, le tout à distance euclidienne fixe, voir les *desirable properties* des distances dans l'introduction de [TY11]. Un désavantage est le calcul explicite de la distance. En

effet, il passe par le calcul des géodésiques qui se fait via la résolution d'équations différentielles, ce qui peut poser de nombreux problèmes numériques.

À l'opposé, on distingue les distances extrinsèques aussi parfois appelées (abusivement) distances euclidiennes. Ces distances sont du type (I.8) et correspondent à la distance euclidienne entre les éléments d'une section. Elles sont donc plus simples à calculer, mais les difficultés sont alors reportées sur le choix de la section. Néanmoins, comme nous l'apprend l'exemple des signatures, des courbes semblables à l'œil et que l'on souhaiterait considérer comme proches, peuvent différer par l'ajout de boucles ou d'autres artefacts qu'une distance intrinsèque pondérera fortement, voir la figure I.9 pour une illustration. Ainsi, il n'apparaît pas clairement que le cadre riemannien soit nécessairement le bon cadre pour l'étude d'objets observés avec des déformations.

Étude de la consistance des moyennes de Fréchet dans les modèles de déformations pour l'analyse de courbes et d'images

Ce travail a été réalisé en collaboration avec Jérémie Bigot en grande partie durant un séjour au *Centro de Modelamiento Matemático* de l'Universidad de Chile à Santiago du Chili durant l'été 2010. Nous tenons à remercier ici le CMM pour l'accueil et l'aide financière apportée pour la réalisation de ce séjour. Ce chapitre est rédigé sous la forme d'un article à paraître dans la revue *Electronic Journal of Statistics* sous le titre *On the consistency of Fréchet means in deformable models for curve and image analysis*. Ce chapitre est donc autonome du reste de ce mémoire du point de vue des notations et des définitions.

Résumé

Dans ce chapitre nous étudions l'estimation des paramètres d'intérêt des modèles de déformations introduits dans la section I.2. Ces modèles peuvent être utilisés pour l'analyse de données en grandes dimensions comme les courbes temporelles ou les images. En plus du terme de bruit additif usuel, ces modèles de déformations incluent un deuxième terme d'erreur qui modélise la variabilité individuelle en intensité des observations autour d'une forme moyenne f . Nous démontrons qu'un estimateur de type moyenne de Fréchet permet l'estimation consistante des paramètres de déformations et de la forme moyenne. La principale contribution de ce travail est l'étude de la convergence de ces estimateurs sous des hypothèses générales en fonction du nombre J de données observées et du nombre n de points dans la grille de discrétisation.

Abstract

A new class of statistical deformable models is introduced to study high-dimensional curves or images. In addition to the standard measurement error term, these deformable models include an extra error term modeling the individual variations in intensity around a mean pattern. It is shown that an appropriate tool for statistical inference in such models is the notion of sample Fréchet means, which leads to estimators of the deformation parameters and the mean pattern. The main contribution of this paper is to study how the behavior of these estimators depends on the number n of design points and the number J of observed curves (or images). Numerical experiments are given to illustrate the finite sample performances of the procedure.

Mots clés : Mean pattern estimation; Fréchet mean; Shape analysis; Deformable models; Curve registration; Image warping; Geometric variability; High-dimensional data.

1.1. Introduction**1.1.1. A statistical deformable model for curve and image analysis**

In many applications, one observes a set of curves or grayscale images which are high-dimensional data. In such settings, it is reasonable to assume that the data at hand Y_j^ℓ , denoting the ℓ -th observation for the j -th curve (or image), satisfy the following regression model:

$$Y_j^\ell = f_j(t_\ell) + \sigma \varepsilon_j^\ell, \quad j = 1, \dots, J, \quad \text{and} \quad \ell = 1, \dots, n, \quad (1.1)$$

where $f_j : \Omega \rightarrow \mathbb{R}$ are unknown regression functions (possibly random) with Ω a convex subset of $\mathbb{R}^d$, the t_ℓ 's are non-random points in Ω (deterministic design), the error terms ε_j^ℓ are i.i.d. normal variables with zero mean and variance 1, and $\sigma > 0$. In this paper, we will suppose that the f_j 's are random elements which vary around the same mean pattern. Our goal is to estimate such a mean pattern and to study the consistency of the proposed estimators in various asymptotic settings: either when both the number n of design points and the number J of curves (or images) tend to infinity, or when n (resp. J) remains fixed while J (resp. n) tends to infinity.

In many situations, data sets of curves or images exhibit a source of geometric variations in time or shape. In such settings, the usual Euclidean mean $\bar{Y}^\ell = \frac{1}{J} \sum_{j=1}^J Y_j^\ell$ in model (1.1) cannot be used to recover a meaningful mean pattern. Indeed, consider the following simple model of randomly shifted curves (with $d = 1$) which is commonly used in many applied areas such as neuroscience [TIR11] or biology [Røn01],

$$f_j(t_\ell) = f(t_\ell - \theta_j^*), \quad j = 1, \dots, J, \quad \text{and} \quad \ell = 1, \dots, n, \quad (1.2)$$

where $f : \Omega \rightarrow \mathbb{R}$ is the mean pattern of the observed curves, and the θ_j^* 's are i.i.d. random variables in $\mathbb{R}$ with density g and independent of the ε_j^ℓ 's. In model (1.2), the shifts θ_j^* represent a source of variability in time. However, in (1.2) the Euclidean mean is not a consistent estimator of the mean pattern f since by the law of large numbers

$$\lim_{J \rightarrow \infty} \bar{Y}^\ell = \lim_{J \rightarrow \infty} \frac{1}{J} \sum_{j=1}^J f(t_\ell - \theta_j^*) = \int f(t_\ell - \theta) g(\theta) d\theta \quad a.s.$$

The randomly shifted curves model (1.2) is close to the perturbation model introduced by [Goo91] in shape analysis for the study of consistent estimation of a mean pattern from a set of random planar shapes. The mean pattern to estimate in [Goo91] is called a population mean, but to stress the fact that it comes from a perturbation model [Huc10] coins the term of perturbation mean. To achieve consistency in such models, a Procrustean procedure is used in [Goo91], which leads to the statistical analysis of sample Fréchet means [Fré48] which are extensions of the usual Euclidean mean to non-linear spaces using non-Euclidean metrics. For random variables belonging to a nonlinear manifold, a well-known example is the computation of the mean of a set of planar shapes in the Kendall's shape space [Ken84] which leads to the Procrustean means studied in [Goo91]. Consistent estimation of a mean planar shape has been studied by various authors, see e.g. [Goo91, KM97, KBCL99, Le98, LK00]. A detailed study of some properties of the Fréchet mean in finite dimensional Riemannian manifolds (such as consistency and uniqueness) has been performed in [Zie77, OC95, BP03, BP05, Huc10, Huc11, Afs11].

The main goal of this paper is to introduce statistical deformable models for curve and image analysis that are analogue to Goodall's perturbation models [Goo91], and to build consistent estimators of a mean pattern in such models. Our approach is inspired by Grenander's pattern theory which considers that the curves or images f_j in model (1.1) are obtained through the deformation of a mean pattern by a Lie group action [Gre93, GM07]. In the last decade, there has been a growing interest in transformation Lie groups to model the geometric variability of images, and the study of the properties of such deformation groups is now an active field of research (see e.g. [MY01, TY05b] and references therein). There is also currently a growing interest in statistics on the use of Lie group actions to analyze geometric modes of variability of a data set [HHM10a, HHM10b].

To describe more formally geometric variability, denote by $L^2(\Omega)$ the set of square integrable real-valued functions on Ω , and by $\mathcal{P}$ an open subset of $\mathbb{R}^p$. To the set $\mathcal{P}$, we associate a parametric family of operators $(T_\theta)_{\theta \in \mathcal{P}}$ such that for each $\theta \in \mathcal{P}$ the operator $T_\theta : L^2(\Omega) \rightarrow L^2(\Omega)$ represents a geometric deformation (parametrized by θ) of a curve or an image. Examples of such deformation operators include the cases of:

Shifted curves: $T_\theta f(t) := f(t - \theta)$, with $\Omega = [0, 1]$, $f \in L^2_{per}([0, 1])$ (the space of periodic functions in $L^2([0, 1])$ with period 1) and $\mathcal{P}$ an open set of $\mathbb{R}$.

Rigid deformation of two-dimensional images: $T_\theta f(t) := f(e^a R_\alpha t - b)$, for $\theta = (a, \alpha, b) \in \mathcal{P}$, with $\Omega = \mathbb{R}^2$, $\mathcal{P} \subset \mathbb{R} \times \mathbb{R} \times \mathbb{R}^2$ where $R_\alpha = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$ is a rotation matrix in $\mathbb{R}^2$, e^a is an isotropic scaling and b a translation in $\mathbb{R}^2$.

Deformation by a Lie group action: the two above cases are examples of a Lie group action on the space $L^2(\Omega)$ (see [Hel01] for an introduction to Lie groups). More generally, assume that G is a connected Lie group of dimension p acting on Ω , meaning that for any $(g, t) \in G \times \Omega$ the action $\cdot$ of G onto Ω is such that $g \cdot t \in \Omega$. In general, G is not a linear space but can be locally parametrized by its Lie algebra $\mathcal{G} \simeq \mathbb{R}^p$ using the exponential map $\exp : \mathcal{G} \rightarrow G$. If $\mathcal{P} \subset \mathbb{R}^p$. This leads for $(\theta, f) \in \mathcal{P} \times L^2(\Omega)$ to define the deformation operators

$$T_\theta f(t) := f(\exp(\theta) \cdot t).$$

Non-rigid deformation of curves or images: assume that one can construct a family $(\psi_\theta)_{\theta \in \mathcal{P}}$ of parametric diffeomorphisms of Ω (see e.g. [BGL09]). Then, for $(\theta, f) \in \mathcal{P} \times L^2(\Omega)$, define the deformation

operators

$$T_{\theta}f(t) := f(\psi_{\theta}(t)).$$

Then, in model (1.1), we assume that the f_j 's have a certain homogeneity in structure in the sense that there exists some $f \in L^2(\Omega)$ such that

$$f_j(t) = T_{\theta_j^*}[f + Z_j](t), \quad \text{for all } t \in \Omega, \text{ and } j = 1, \dots, J, \quad (1.3)$$

where $\theta_j^* \in \mathcal{P}$, $j = 1, \dots, J$ are i.i.d. random variables (independent of the ε_j^ℓ 's) with an unknown density g with compact support Θ included in $\mathcal{P}$ satisfying:

Assumption 1.1. The density g of the θ_j^* 's is continuously differentiable on $\mathcal{P}$ and has a compact support Θ included in $\mathcal{P} \subset \mathbb{R}^p$. We assume that Θ can be written

$$\Theta = \left\{ \theta = (\theta^1, \dots, \theta^p) \in \mathbb{R}^p, \quad |\theta^{p_1}| \leq \rho, \quad 1 \leq p_1 \leq p \right\} \quad (1.4)$$

where $\rho > 0$.

The function f in model (1.3) represents the unknown mean pattern of the f_j 's. The Z_j 's are supposed to be independent of the ε_j^ℓ 's and are i.i.d. realizations of a second order centered Gaussian process Z taking its values in $L^2(\Omega)$. The Z_j 's represent the individual variations in intensity around f , while the random operators T_{θ_j} model geometric deformations in time or space. Then, if we assume that the T_{θ} 's are linear operators, equation (1.3) leads to the following *statistical deformable model* for curve or image analysis

$$Y_j^\ell = T_{\theta_j^*}f(t_\ell) + T_{\theta_j^*}Z_j(t_\ell) + \sigma\varepsilon_j^\ell, \quad j = 1, \dots, J, \text{ and } \ell = 1, \dots, n, \quad (1.5)$$

where ε_j^ℓ are i.i.d. normal variables with zero mean and variance 1. In the particular case of the shifted curves framework we also consider a slight variation of (1.5) called *small perturbation model*. It is defined by equation (1.16) and contains an error term Z_j^n with a covariance matrix which depends on the number n of point in the design. This dependence in n allows us to ensure the consistency of our estimation procedure. Indeed, even with simple example of noise Z_j , the estimation of the translation parameters θ^* and the template f is not guaranteed, see Section 1.5 for details and examples.

Model (1.5) could be also called a *perturbation model* using the terminology in [Goo91, Huc10] for shape analysis. To be more precise, let $\mathbf{Y} \in \mathbb{R}^{n \times 2}$ be a set of n points in $\mathbb{R}^2$ representing a planar shape. Define a deformation operator T_{θ} for $\theta = (a, \alpha, b) \in \Theta = \mathbb{R} \times [0, 2\pi] \times \mathbb{R}^2$ acting on $\mathbb{R}^{n \times 2}$ in the following way

$$T_{\theta}\mathbf{Y} = e^a \mathbf{Y} R_{\alpha} + \mathbf{1}_n b', \text{ where } R_{\alpha} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix},$$

and $\mathbf{1}_n = (1, \dots, 1)' \in \mathbb{R}^n$. Consistent estimation of a mean shape has been first studied in [Goo91] when a set of random shapes $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ is drawn from the following perturbation model

$$\mathbf{Y}_j = T_{\theta_j^*}(\mu + \zeta_j), \quad j = 1, \dots, J. \quad (1.6)$$

Model (1.6) is similar to the statistical deformable model (1.5), where $\mu \in \mathbb{R}^{n \times 2}$ is the unknown perturbation mean to estimate, and ζ_j are i.i.d. random vectors in $\mathbb{R}^{n \times 2}$ with zero mean. Nevertheless,

there exists major differences between our approach and the one in [Goo91]. First, in model (1.5), the deformation parameters θ_j^* are assumed to be random variables following an unknown distribution, whereas they are just nuisance parameters in model (1.6) for shape analysis, see [Goo91, KM97]. In some applications (e.g. in biomedical imaging [JDJG04]), it is of interest to reconstruct the unobserved parameters θ_j^* and to estimate their distribution. One of the main contribution of this paper is then to construct upper and lower bounds for the estimation of such deformation parameters. Moreover, in model (1.5), they are too additive error terms, whereas the model (1.6) only include the error term ζ_j . In model (1.5), the ε_j^ℓ is an additive noise modeling the errors in the measurements, while the Z_j 's model (possibly smooth) variations in intensity of the individuals around the mean pattern f .

In [KM97], the authors studied the relationship between isotropicity of the additive noise ζ_j and the convergence of Procrustean procedures to the perturbation mean μ as $J \rightarrow +\infty$. It is shown in [KM97] that, for isotropic errors, Procrustean means are consistent, but that, for non-isotropic errors, they may not converge to μ . For a recent discussion on the issues of consistency of sample Procrustes means in perturbation models and extension to non-metrical Fréchet means, we refer to [Huc10] and [Huc11]. In this paper, we carefully analyze the role of the dimension n and the number of samples J on the consistency of Procrustean means in model (1.5). To obtain consistent procedures, we show that it is not required to impose very restrictive conditions on the error terms Z_j such as isotropicity for the ζ_j in (1.6) for shape analysis. Here, the key quantity is the dimension n of the data (number of design points) which plays the central role to guarantee the converge of our estimators. This point is another major difference with the approach of statistical shape analysis [Goo91] that does not take into account the dimensionality of the shape space to analyze the consistency of Procrustean estimators.

Note that a subclass of the deformable model (1.5) is the so-called shape invariant model (SIM)

$$Y_j^\ell = T_{\theta_j^*} f(t_\ell) + \sigma \varepsilon_j^\ell, \quad j = 1, \dots, J, \quad \text{and} \quad \ell = 1, \dots, n, \quad (1.7)$$

i.e. without incorporating in (1.5) the additive terms Z_j .

The goal of this paper is twofold. First, we propose a general methodology for estimating f and the θ_j^* 's based on observations coming from model (1.5). For this purpose, we show that an appropriate tool is the notion of sample Fréchet mean of a data set [Fré48, Zie77, BP03] that has been widely studied in shape analysis [Goo91, KM97, Le98, LK00, Huc10] and more recently in biomedical imaging [JDJG04, Pen06]. Secondly, we study the consistency of the resulting estimators in various asymptotic settings: either when n and J both tend to infinity, or when n is fixed and $J \rightarrow +\infty$, or when J is fixed and $n \rightarrow +\infty$.

1.1.2. Organization of the paper

Section 1.2 contains a description of our estimating procedure and a review of previous work in mean pattern estimation. In Section 1.3, we derive a lower bound for the quadratic risk of estimators of the deformation parameters. In Section 1.4, we discuss some identifiability issues in model (1.5). In Section 1.5 we derive consistency results for the Fréchet mean in the case (1.2) of randomly shifted curves. In Section 1.6 and Section 1.7, we give general conditions to extend these results to the more general deformable model (1.5). Section 1.8 contains some numerical experiments. A small conclusion with some perspectives are given in Section 1.9. All proofs are postponed to a technical Appendix.

1.2. The estimating procedure

1.2.1. A dissimilarity measure based on deformation operators

To define a notion of sample Fréchet mean for curves or images, let us suppose that the family of deformation operators $(T_\theta)_{\theta \in \mathcal{P}}$ is invertible in the sense that there exists a family of operators $(\tilde{T}_\theta)_{\theta \in \mathcal{P}}$ such that for any $(\theta, f) \in \mathcal{P} \times L^2(\Omega)$

$$\tilde{T}_\theta f \in L^2(\Omega) \quad \text{and} \quad \tilde{T}_\theta T_\theta f = f.$$

Then, for two functions $f, h \in L^2(\Omega)$ introduce the following dissimilarity measure

$$d_T^2(h, f) = \inf_{\theta \in \mathcal{P}} \int_{\Omega} \left(\tilde{T}_\theta h(t) - f(t) \right)^2 dt.$$

If $d_T^2(h, f) = 0$ then there exists $\theta \in \mathcal{P}$ such that $f = \tilde{T}_\theta h$ meaning that the functions f and h are equal up to a geometric deformation. Note that d_T is not necessarily a distance on $L^2(\Omega)$, but it can be used to define a notion of sample Fréchet mean of data from model (1.5). For this purpose let $\mathcal{F}$ denote a subspace of $L^2(\Omega)$ and suppose that $\hat{f}_j$ are smooth functions in $\mathcal{F} \subset L^2(\Omega)$ obtained from the data Y_j^ℓ , $\ell = 1, \dots, n$ for $j = 1, \dots, J$, see Section 1.5.2 and Section 1.6.2 for precise definitions. Following the definition of a Fréchet mean in general metric space [Fré48], define an estimator of the mean pattern f as

$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \frac{1}{J} \sum_{j=1}^J d_T^2(\hat{f}_j, f). \quad (1.8)$$

Note that $\hat{f}$ falls into the category of non-metrical sample Fréchet means whose definitions and asymptotic properties are discussed in [Huc10] for random variables belonging to Riemannian manifolds. However, unlike the usual approach in shape analysis, the Fréchet mean (1.8) is based on smoothed data. In what follows, we show that smoothing is a key preliminary step to obtain the convergence of $\hat{f}$ to the mean pattern f in the deformable model (1.5). It can be easily shown that the computation of $\hat{f}$ can be done in two steps: first minimize the following criterion

$$(\hat{\theta}_1, \dots, \hat{\theta}_J) = \operatorname{argmin}_{(\theta_1, \dots, \theta_J) \in \Theta^J} M(\theta_1, \dots, \theta_J), \quad (1.9)$$

where

$$M(\theta_1, \dots, \theta_J) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \hat{f}_j(t) - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \hat{f}_{j'}(t) \right)^2 dt, \quad (1.10)$$

which gives an estimation of the deformation parameters $\theta_1^*, \dots, \theta_J^*$, and then in a second step take

$$\hat{f}(t) = \frac{1}{J} \sum_{j=1}^J \tilde{T}_{\hat{\theta}_j} \hat{f}_j(t), \quad \text{for } t \in \Omega,$$

as an estimator of the mean pattern f .

Note that this two steps procedure belongs to the category of Procrustean methods (see e.g [DM98, Goo91]). A similar approach to (1.9) has been developed by [JDJG04] in the context of biomedical images using diffeomorphic deformation operators.

1.2.2. Previous work in mean pattern estimation and geometric variability analysis

Estimating the mean pattern of a set of curves that differ by a time transformation is usually referred to as the curves registration problem, see e.g. [GK92, Big06b, RL01, WG97, LM04]. However, in these papers, studying consistent estimators of the mean pattern f as the number of curves J and design points n tend to infinity is not considered. For the SIM (1.7), a semiparametric point of view has been proposed in [GLM07] and [Vim10] to estimate non-random deformation parameters (such as shifts and amplitudes) as the number n of observations per curve grows, but with a fixed number J of curves. A generalisation of this semiparametric approach for two-dimensional images is proposed in [BGV09]. The case of image deformations by a Lie group action is also investigated in [BLV10] from a semiparametric point of view using a SIM.

In the simplest case of randomly shifted curves in a SIM, [BG10a] have studied minimax estimation of the mean pattern f by letting only the number J of curves going to infinity. Self-modelling regression (SEMR) methods proposed by [KG88] are semiparametric models where each observed curve is a parametric transformation of the same regression function. However, the SEMR approach does not incorporate a random fluctuations in intensity of the individuals around a mean pattern f through an unknown process Z_j as in model (1.5). The authors in [KG88] studied the consistency of the SEMR approach using a Procrustean algorithm. Recently, there has also been a growing interest on the development of statistical deformable models for image analysis and the construction of consistent estimators of a mean pattern, see [GM01, BGV09, BGL09, AST07, AKT10].

1.3. Lower bounds for the estimation of the deformation parameters

In this section, we derive non-asymptotic lower bounds for the quadratic risk of an arbitrary estimator of the deformation parameters under the following smoothness assumption of the mapping $(\theta, t) \mapsto T_\theta f(t)$.

Assumption 1.2. For all $\theta = (\theta^1, \dots, \theta^p) \in \mathcal{P}$, $T_\theta : L^2(\Omega) \rightarrow L^2(\Omega)$ is a linear operator such that the function $t \mapsto \partial_{\theta^{p_1}} T_\theta f(t)$ exists and belongs to $L^2(\Omega)$ for any $p_1 = 1, \dots, p$. Moreover, there exists a constant $C(\Theta, f) > 0$ such that

$$\|\partial_{\theta^{p_1}} T_\theta f\|_{L^2}^2 \leq C(\Theta, f),$$

for all $p_1 = 1, \dots, p$ and $\theta \in \Theta$.

1.3.1. Shape Invariant Model

Theorem 1.1. Consider the SIM (1.7) and suppose that Assumption 1.2 holds. Assume that g satisfies Assumption 1.1, and that $\int_\Theta \|\partial_\theta \log(g(\theta))\|^2 g(\theta) d\theta < +\infty$. Let $\hat{\theta} \in \mathcal{P}^J$ be any estimator (a measurable function of the data) of $\theta^* = (\theta_1^*, \dots, \theta_J^*)$. Then, for any $n \geq 1$ and $J \geq 1$,

$$\mathbb{E} \left[\frac{1}{J} \|\hat{\theta} - \theta^*\|_{\mathbb{R}^{pJ}}^2 \right] \geq \frac{\sigma^2 n^{-1}}{C(\Theta, f) + \sigma^2 n^{-1} \int_\Theta \|\partial_\theta \log(g(\theta))\|^2 g(\theta) d\theta}. \quad (1.11)$$

where $C(\Theta, f)$ is the constant defined in Assumption 1.2, and $\|\cdot\|_{\mathbb{R}^{pJ}}$ is the standard Euclidean norm in $\mathbb{R}^{pJ}$.

The lower bound given in inequality (1.11) does not decrease as J increases. Thus, if the number n of design points is fixed, increasing the number J of curves (or images) does not improve the quality of the estimation of the deformation parameters for any estimator $\hat{\boldsymbol{\theta}}$. Nevertheless, this lower bound is going to 0 as the dimension $n \rightarrow +\infty$.

1.3.2. General model

The main difference between the general model (1.5) and the SIM (1.7) is the extra error terms $T_{\boldsymbol{\theta}_j^*} Z_j$, $j = 1, \dots, J$. In what follows, $\mathbb{E}_{\boldsymbol{\theta}}[\cdot]$ denotes expectation conditionally to $\boldsymbol{\theta} \in \Theta^J$. Since the random processes Z_j 's are observed through the action of the random deformation operators $T_{\boldsymbol{\theta}_j^*}$ it is necessary to specify how the $T_{\boldsymbol{\theta}_j^*}$'s modify the law of the process Z_j .

Assumption 1.3. There exists a positive semi-definite symmetric $n \times n$ matrix $\boldsymbol{\Sigma}_n(\Theta)$ such that the covariance matrix of $\mathbf{Z} = [Z(t_\ell)]_{\ell=1}^n$ satisfies $\mathbb{E}_{\boldsymbol{\theta}}[\mathbf{T}_{\boldsymbol{\theta}} \mathbf{Z} (\mathbf{T}_{\boldsymbol{\theta}} \mathbf{Z})'] = \boldsymbol{\Sigma}_n(\Theta)$.

This assumption means that the law of the random process Z is somewhat invariant by the deformation operators $T_{\boldsymbol{\theta}}$. Such an hypothesis is similar to the condition given in [KM97] to ensure consistency of Fréchet mean estimators in Kendall's shape space using model similar to (1.5) with $\sigma = 0$. After a normalization step, the deformations considered in [KM97] are rotations of the plane, and the authors in [KM97] study the case where the law of the error term Z is isotropic, that is to say, invariant by the action of rotations.

Theorem 1.2. Consider the general model (1.5). Suppose that Assumption 1.2 and 1.3 hold. Assume that the density g satisfies Assumption 1.1 and that $\int_{\Theta} \|\partial_{\boldsymbol{\theta}} \log(g(\boldsymbol{\theta}))\|^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta} < +\infty$. Let $\hat{\boldsymbol{\theta}} \in \mathcal{P}^J$ be any estimator (a measurable function of the data) of $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_J^*)$. Then, for any $n \geq 1$ and $J \geq 1$, we have

$$\mathbb{E} \left[\frac{1}{J} \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_{\mathbb{R}^{pJ}}^2 \right] \geq \frac{(\sigma^2 + s_n^2(\Theta))n^{-1}}{C(\Theta, f) + (\sigma^2 + s_n^2(\Theta))n^{-1} \int_{\Theta} \|\partial_{\boldsymbol{\theta}_1} \log(g(\boldsymbol{\theta}))\|^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (1.12)$$

where $C(\Theta, f)$ is the constant defined in Assumption 1.2, and $s_n^2(\Theta)$ denotes the smallest eigenvalue of $\boldsymbol{\Sigma}_n(\Theta)$.

Again, the lower bound (1.12) does not depends on J . Thus, increasing the number J of observations does not decrease the quadratic risk of any estimator of the deformations parameters. Moreover, the lower bound (1.12) tends to zero as $n \rightarrow +\infty$ only if $\lim_{n \rightarrow +\infty} n^{-1} s_n^2(\Theta) = 0$.

1.3.3. Application to the shifted curves model

Consider the shifted curves model (1.2) with an equi-spaced design, namely

$$Y_j^\ell = f\left(\frac{\ell}{n} - \boldsymbol{\theta}_j^*\right) + Z_j\left(\frac{\ell}{n} - \boldsymbol{\theta}_j^*\right) + \sigma \varepsilon_j^\ell, \quad j = 1, \dots, J, \quad \text{and} \quad \ell = 1, \dots, n. \quad (1.13)$$

Theorem 1.3. Consider the model (1.13). Assume that f is continuously differentiable on $[0, 1]$ and that Z is a centered stationary process with value in $L_{\text{per}}^2([0, 1])$. Suppose that $\Theta = [-\rho, \rho]$ with

$\rho < \frac{1}{2}$ and $\int_{\Theta} (\partial_{\theta} \log(g(\theta)))^2 g(\theta) d\theta < +\infty$. Let $\hat{\theta} \in \mathbb{R}^J$ be any estimator of the true random shifts $\theta^* = (\theta_1^*, \dots, \theta_J^*)$, i.e. a measurable function of the data in model (1.13). Then, for any $n \geq 1$ and $J \geq 1$

$$\mathbb{E} \left[\frac{1}{J} \|\hat{\theta} - \theta^*\|_{\mathbb{R}^J}^2 \right] \geq \frac{n^{-1} \sigma^2}{\|\partial_t f\|_{\infty}^2 + n^{-1} \sigma^2 \int_{\Theta} (\partial_{\theta} \log(g(\theta)))^2 g(\theta) d\theta},$$

where $\|\partial_t f\|_{\infty} = \sup_{t \in [0,1]} \{|\partial_t f(t)|\}$ with $\partial_t f$ denoting the first derivative of f .

1.4. Identifiability conditions

1.4.1. The shifted curves model

Without any further assumptions, the randomly shifted curves models (1.13) and (1.16) below are not identifiable. Indeed, if $\theta_0 \in \Theta$ satisfies $\theta_j^* + \theta_0 \in \Theta$, $j = 1, \dots, J$, then replacing $f(\cdot)$ by $f(\cdot - \theta_0)$ and θ_j^* by $\theta_j^* + \theta_0$ does not change the formulation of models (1.13) and (1.16). Choosing identifiability conditions amounts to impose constraints on the minimization of the criterion

$$D(\theta) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(f(t - \theta_j^* + \theta_j) - \frac{1}{J} \sum_{j'=1}^J f(t - \theta_{j'}^* + \theta_{j'}) \right)^2 dt,$$

for $\theta = (\theta_1, \dots, \theta_J) \in \Theta^J$, which can be interpreted as a version without noise of the criterion (1.9) using the ideal smoothers $\hat{f}_j(\cdot) = f(\cdot - \theta_j^*)$. Obviously, the criterion $D(\theta)$ has a minimum at $\theta^* = (\theta_1^*, \dots, \theta_J^*)$ such that $D(\theta^*) = 0$, but this minimizer of D on Θ^J is clearly not unique. If the true shifts are supposed to have zero mean (i.e. $\int_{\Theta} \theta g(\theta) d\theta = 0$) it is natural to introduce the constrained set

$$\Theta_0 = \{(\theta_1, \dots, \theta_J) \in \Theta^J, \theta_1 + \dots + \theta_J = 0\}. \quad (1.14)$$

It is shown in [BG10a] Lemma 6, that if $f \in L^2([0, 1])$ is such that $\int_0^1 f(t) e^{-i2\pi t} dt \neq 0$ and if $\rho < 1/4$ (recall that $\Theta = [-\rho, \rho]$), then the criterion $D(\theta)$ has a unique minimum on Θ_0 in the sense that $D(\theta) > D(\theta_{\Theta_0}^*)$ for all $\theta \in \Theta_0$ with $\theta \neq \theta_{\Theta_0}^*$ where

$$\theta_{\Theta_0}^* = (\theta_1^* - \bar{\theta}^*, \dots, \theta_J^* - \bar{\theta}^*) \text{ with } \bar{\theta}^* = \frac{1}{J} \sum_{j=1}^J \theta_j^*. \quad (1.15)$$

Under such assumptions, we will compute estimators of the random shifts by minimizing the criterion (1.9) over the constrained set Θ_0 and not directly on Θ^J . Consistency of such constrained estimators will then be studied under the following identifiability conditions:

Assumption 1.4. The mean pattern f is such that $\int_0^1 f(t) e^{-i2\pi t} dt \neq 0$.

Assumption 1.5. The support of the density g is included in $[-\rho', \rho']$ for some $0 < \rho' \leq \frac{\rho}{2} < 1/4$ and is such that $\int_{\Theta} \theta g(\theta) d\theta = 0$.

Under such assumptions, $D(\theta)$ can be bounded from below by the quadratic function $\frac{1}{J} \|\theta - \theta_{\Theta_0}^*\|^2$ which will be an important property to derive consistent estimators.

Proposition 1.4. *Suppose that Assumptions 1.4 and 1.5 hold with $\rho < 1/16$. Then, for any $\theta = (\theta_1, \dots, \theta_J) \in \Theta_0$, one has that*

$$D(\theta) - D(\theta_{\Theta_0}^*) \geq C(f, \rho) \frac{1}{J} \|\theta - \theta_{\Theta_0}^*\|^2,$$

where $C(f, \rho) > 0$ is a constant depending only on f and ρ .

Assumption 1.5 and the condition that $\rho < 1/16$ in Proposition 1.4 mean that the support of the density g of the shifts is sufficiently small, and that the shifted curves $f_j(t) = f(t - \theta_j^*)$ are in some sense concentrated around the mean pattern f . Such an assumption of concentration of the data around the same mean pattern has been used in various papers to prove the uniqueness and the consistency of Fréchet means for random variables lying in a Riemannian manifold, see [Kar77, Le98, BP03, Afs11, Ken90].

1.4.2. The general case

In the case of general deformation operators, define for $\theta = (\theta_1, \dots, \theta_J) \in \Theta^J$ the criterion

$$D(\theta) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} T_{\theta_j^*} f(t) - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} T_{\theta_{j'}^*} f(t) \right)^2 dt.$$

Obviously, using that for all $\theta \in \Theta$, $\tilde{T}_{\theta} T_{\theta} f = f$, the criterion $D(\theta)$ has a minimum at $\theta^* = (\theta_1^*, \dots, \theta_J^*)$ such that $D(\theta^*) = 0$. However, without any further restrictions the minimizer of $D(\theta)$ is not necessarily unique on Θ^J .

Assumption 1.6. Let $\Theta \subset \Theta^J$ such that there exists a unique $\theta_{\Theta}^* \in \Theta$ satisfying $D(\theta_{\Theta}^*) = 0$.

Then, Θ is the set onto which we will carry the minimization of the criterion $M(\theta)$ (1.10). In the case of shifted curves and under Assumption 1.4 and 1.5, the only set onto which the criterion D vanishes is the line $\{\theta^* + \theta_0 \mathbf{1}_J, \theta_0 \in \mathbb{R}\} \subset \mathbb{R}^J$, where $\mathbf{1}_J = (1, \dots, 1)' \in \mathbb{R}^J$. An easy way to choose the set Θ is to take a linear subset of Θ^J , see Figure 1.1 for an illustration. By considering the subset

$$\Theta_0 = \Theta^J \cap \mathbf{1}_J^\perp = \{(\theta_1, \dots, \theta_J) \in \Theta^J, \theta_1 + \dots + \theta_J = 0\},$$

where $\mathbf{1}_J^\perp$ is the orthogonal of $\mathbf{1}_J$ in $\mathbb{R}^J$, then Assumption 1.6 is satisfied with $\theta_{\Theta_0}^*$ given in (1.15). More generally, if the deformation parameters θ_j , $j = 1, \dots, J$ are supposed to be random variables with zero mean, then optimizing $D(\theta)$ on Θ_0 is a natural choice. Another identifiability condition for shifted curves is proposed in [GLM07] and [Vim10] by taking

$$\Theta_1 = \Theta^J \cap e_1^\perp = \{(\theta_1, \dots, \theta_J) \in \Theta^J, \theta_1 = 0\}.$$

where $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^J$. In this case, $\theta_{\Theta_1}^* = (0, \theta_2^* - \theta_1^*, \dots, \theta_J^* - \theta_1^*)$. Choosing to minimize $D(\theta)$ on Θ_1 amounts to choose the first curve as a reference onto which all the others curves are aligned, meaning that the first shift θ_1^* is not random, see Figure 1.1.

Following the classical guidelines in M-estimation (see e.g. [vdV98]), a necessary condition to ensure the convergence of M-estimators such as (1.9) is that the local minima of $D(\theta)$ over Θ are well separated from the global minimum of $D(\theta)$ at $\theta = \theta_{\Theta}^*$ (satisfying $D(\theta_{\Theta}^*) = 0$). The following assumption can be interpreted in this sense.

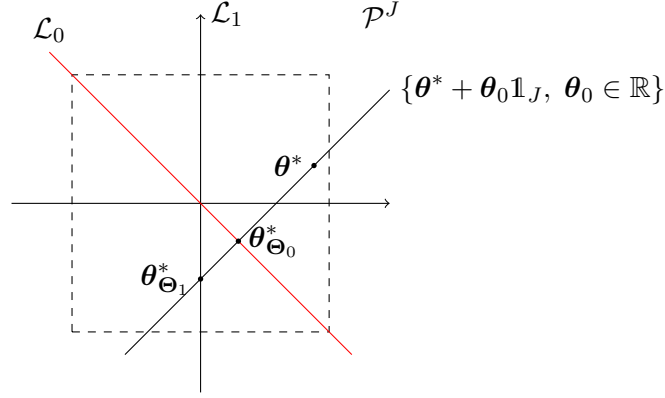


Figure 1.1: Choice of identifiability conditions for shifted curves in the case $J = 2$.

Assumption 1.7. For all $\theta \in \Theta$ we have

$$D(\theta) - D(\theta_{\Theta}^*) \geq C(\Theta, \mathcal{F}) \frac{1}{J} \|\theta - \theta_{\Theta}^*\|^2$$

for a constant $C(\Theta, \mathcal{F}) > 0$ independent of J .

In the shifted curve model, Assumption 1.7 is verified if Assumption 1.4 and 1.5 hold (see Proposition 1.4).

1.5. Consistent estimation in the shifted curves model

In this section, we give conditions to ensure the consistency of the estimators defined in Section 1.2 in the shifted curves model. To prove the convergence of the estimator $\hat{\theta}^\lambda$ to the true translation parameters θ^* we use Theorem 1.9 below. Unfortunately, we are not able to apply this result under the hypothesis used in Theorem 1.3. Recall that in Theorem 1.3, we considered model (1.13) with a perturbation Z which was a stationary Gaussian process. In what follows, the covariance function of such a Z is denoted by $R : [0, 1] \rightarrow \mathbb{R}$. For instance, when $R(t) = \cos(2\pi t)$, $t \in [0, 1]$, Assumption 1.12 is not satisfied since we have $\gamma_{\max}(\mathbb{E}_{\theta}[\mathbf{T}_{\theta} \mathbf{Z}(\mathbf{T}_{\theta} \mathbf{Z})']) = \frac{n}{2}$ and equation (1.21) does not hold. This shows that, even with simple example of stationary noise, it is not necessarily possible to estimate the deformations parameters with our procedure.

To tackle this problem, let us introduced the following shifted curves model

$$Y_j^\ell = f\left(\frac{\ell}{n} - \theta_j^*\right) + Z_j^n\left(\frac{\ell}{n} - \theta_j^*\right) + \sigma \varepsilon_n^\ell \quad (1.16)$$

Note that this model is similar to (1.13) but here, the covariance function $R^n : [0, 1] \rightarrow \mathbb{R}$ of the Gaussian random process Z^n depends on the number n of points in the design. We call such a model a *small perturbation model* - see the justification in the following subsection.

1.5.1. The random perturbations Z_j^n

In order to apply Theorem 1.9 to the shifted curve model, the random perturbations Z^n must satisfy Assumption 1.12. This amounts to control the largest eigenvalue $\gamma_{\max}(\Sigma_n)$ of the covariance matrix $\Sigma_n \in \mathbb{R}^{n \times n}$ of the centered Gaussian vector $\left[Z^n \left(\frac{\ell}{n} - \theta^* \right) \right]_{\ell=1}^n$. Actually, model (1.16) allows us to consider perturbation terms Z_j^n with a $\gamma_{\max}(\Sigma_n)$ bounded by a constant independent of n . Let us make the following assumption,

Assumption 1.8. Let W be a stationary second order Gaussian process independent of the error terms ε_j and with a continuous covariance function $R : [0, 1] \rightarrow \mathbb{R}$. The error terms Z^n is a Gaussian vector satisfying

$$Z^n = \frac{W}{\sqrt{n}}.$$

The normalization by factor $\sqrt{n}$ justifies the terminology *small perturbation* model. Note also that for all $j = 1, \dots, J$, the processes Z_j^n are independent realizations of the random process Z^n . Let us now describe some consequence of Assumption 1.8. We have for each independent realizations W_j , $j = 1, \dots, J$, of W

$$\mathbb{E}_{\theta_j^*} \left[W_j \left(\frac{\ell}{n} - \theta_j^* \right) W_j \left(\frac{\ell'}{n} - \theta_j^* \right) \right]_{\ell, \ell'=1}^n = \mathbb{E} \left[W \left(\frac{\ell}{n} \right) W \left(\frac{\ell'}{n} \right) \right]_{\ell, \ell'=1}^n = \left[R \left(\frac{|\ell - \ell'|}{n} \right) \right]_{\ell, \ell'=1}^n := \Gamma_n.$$

Remark that the $n \times n$ matrix Γ_n is a semi-positive definite matrix whose largest eigenvalue $\gamma_{\max}(\Gamma_n) \leq \text{tr} \Gamma_n = nR(0)$. For all $j = 1, \dots, J$ and $n \in \mathbb{N}$, the covariance matrix of Z_j^n is

$$\Sigma_n = \mathbb{E}_{\theta_j^*} \left[Z_j^n \left(\frac{\ell}{n} - \theta_j^* \right) Z_j^n \left(\frac{\ell'}{n} - \theta_j^* \right) \right]_{\ell, \ell'=1}^n = \frac{1}{n} \Gamma_n.$$

Thence, the largest eigenvalue $\gamma_{\max}(\Sigma_n)$ of the matrix Σ_n satisfies

$$\gamma_{\max}(\Sigma_n) \leq R(0). \quad (1.17)$$

The strictly positive constant $R(0)$ is independent of n and represents an upper bound for the variance of Z_j^n , $j = 1, \dots, J$. In particular, the bound (1.17) means that Assumption 1.12 holds when Assumption 1.8 is satisfied. Finally, note that a numerical illustration with a non-stationary error term satisfying $\gamma_{\max}(\mathbb{E}_{\theta} [\mathbf{T}_{\theta} \mathbf{Z} (\mathbf{T}_{\theta} \mathbf{Z})']) = n$ is given by Figure 1.4.

Remark: We could consider another approach to guaranty the uniform boundedness in n of $\gamma_{\max}(\Sigma_n)$. For a fixed $n \in \mathbb{N}$, let Z^n be a centered Gaussian vector of covariance matrix

$$\Sigma_n = [S(|\ell - \ell'|)]_{\ell, \ell'=1}^n,$$

where $S : \mathbb{R} \rightarrow \mathbb{R}$ is an integrable function (*i.e.* $\int_{\mathbb{R}} |S(t)| dt < +\infty$). The $n \times n$ matrix Σ_n is a Toeplitz matrix whose largest eigenvalue $\gamma_{\max}(\Sigma_n)$ is bounded uniformly in n . Indeed, the entries $S(k)$, $k = 0, \dots, n-1$ of the matrix Σ_n can be interpreted as the Fourier coefficient of a function $h : [0, 1] \rightarrow \mathbb{R}$, namely $h(u) = \sum_{k \in \mathbb{Z}} S(k) e^{2\pi i k u}$, $u \in [0, 1]$. It follows from standard results on sequence of Toeplitz matrices (see e.g. [Gra06]) that

$$\gamma_{\max}(\Sigma_n) \leq \int_0^1 |S(t)| dt.$$

1.5.2. Choice of the smoothed estimators $\hat{f}_j$

A convenient choice for the smoothing of the observed curves in (1.13) is to do low-pass Fourier filtering. Let $\hat{c}_{j,k} = \frac{1}{n} \sum_{\ell=1}^n Y_j^\ell e^{-i2\pi k \frac{\ell}{n}}$ for $k = -(n-1)/2, \dots, (n-1)/2$ (assuming for simplicity that n is odd), and define for a spectral cut-off parameter $\lambda \in \mathbb{N}$ and $t \in [0, 1]$ the linear estimators

$$\hat{f}_j^\lambda(t) = \sum_{|k| \leq \lambda} \hat{c}_{j,k} e^{i2\pi kt}.$$

Then, define the Sobolev ball $H_s(A)$ of radius $A > 0$ and regularity $s > 0$ as

$$H_s(A) = \left\{ f \in L_{per}^2([0, 1]), \sum_{k \in \mathbb{Z}} (1 + |k|^2)^s |c_k(f)|^2 < A \right\}. \quad (1.18)$$

with $c_k(f) = \int_0^1 f(t) e^{-i2\pi kt} dt$, $k \in \mathbb{Z}$ for a function $f \in L_{per}^2([0, 1])$, and take $\mathcal{F} = H_s(A)$ as the smoothness class to which the mean pattern f is supposed to belong.

1.5.3. Consistent estimation of the random shifts

Using low-pass filtering, and following the discussion in Section 1.4.1 on identifiability issues, the estimators of the random shifts $\theta_1^*, \dots, \theta_J^*$ are given by

$$\hat{\theta}^\lambda = (\hat{\theta}_1^\lambda, \dots, \hat{\theta}_J^\lambda) = \underset{(\theta_1, \dots, \theta_J) \in \Theta_0}{\operatorname{argmin}} M_\lambda(\theta_1, \dots, \theta_J). \quad (1.19)$$

where the criterion $M_\lambda(\theta) = M_\lambda(\theta_1, \dots, \theta_J)$ for $\theta \in \Theta^J$ is

$$M_\lambda(\theta) = \frac{1}{J} \sum_{j=1}^J \int_\Omega \left(\hat{f}_j^\lambda(t + \theta_j) - \frac{1}{J} \sum_{j'=1}^J \hat{f}_{j'}^\lambda(t + \theta_{j'}) \right)^2 dt$$

and Θ_0 is the constrained set defined in (1.14).

Theorem 1.5. *Consider the model (1.16) and let $\hat{\theta}^\lambda$ be the estimator defined by (1.19). Assume that $\mathcal{F} = H_s(A)$ for some $A > 0$ and $s \geq 1$, and that the error term Z^n satisfies Assumption 1.8. Suppose that Assumptions 1.4 and 1.5 hold with $\rho < 1/16$. Then, for any $\lambda \geq 1$ and $x > 0$*

$$\mathbb{P} \left(\frac{1}{J} \|\hat{\theta}^\lambda - \theta^*\|_{\mathbb{R}^J}^2 \geq C_1(\Theta, \mathcal{F}, f) A_1(x, J, n, \lambda, \sigma^2, \gamma) + A_2(x, J) \right) \leq 4e^{-x},$$

with $A_1(x, J, n, \lambda, \sigma^2, \gamma) = (\sigma^2 + \gamma) \left(\sqrt{v(x, J, n, \lambda)} + v(x, J, n, \lambda) \right) + \left(\sqrt{B(\lambda, n)} + B(\lambda, n) \right)$ and $A_2(x, J) = \left(\sqrt{\frac{2x}{J}} + \frac{x}{3J} \right)^2$, where $C_1(\Theta, \mathcal{F}, f) > 0$ is constant depending only on $\Theta, \mathcal{F}, f$, $v(x, J, n, \lambda) = \frac{2\lambda+1}{n} \left(1 + 4\frac{x}{J} + \sqrt{4\frac{x}{J}} \right)$, $B(\lambda, n) = \frac{2\lambda+1}{n} + \lambda^{-2s}$ and $\gamma = R(0)$.

First, remark that for fixed values of n and λ , then $\lim_{J \rightarrow +\infty} A_2(x, J) = 0$. The term $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ depends on the spectral cutoff λ via the bias $B(\lambda, n)$ and the variance $v(x, J, n, \lambda)$ of the estimators $\hat{f}_j$. By choosing a sequence $\lambda = \lambda_n$ such that $\lim_{n \rightarrow +\infty} \lambda_n = +\infty$ and $\lim_{n \rightarrow +\infty} \frac{\lambda_n}{n} = 0$ (tradeoff between low variance and low bias) it follows that for fixed J and $x > 0$, then $\lim_{n \rightarrow +\infty} A_1(x, J, n, \lambda_n, \sigma^2, \gamma) = 0$. However, if n remains fixed, then $\lim_{J \rightarrow +\infty} A_1(x, J, n, \lambda, \sigma^2, \gamma) > 0$.

Thus, Theorem 1.5 is consistent with the conclusions of Theorem 1.3, that is, if n is fixed, then it is not possible to estimate θ^* by letting only J grows to infinity. Hence, under the assumptions of Theorem 1.5, one can only prove the convergence in probability of $\hat{\theta}^\lambda$ to the true shifts θ^* by taking the double asymptotic $n \rightarrow +\infty$ and $J \rightarrow +\infty$, provided the smoothing parameter $\lambda = \lambda_n$ is well chosen.

1.5.4. Consistent estimation of the mean pattern

In the case of randomly shifted curves, the Fréchet mean estimator (1.8) of f is $\hat{f}^\lambda(t) = \frac{1}{J} \sum_{j=1}^J \hat{f}_j^\lambda(t + \hat{\theta}_j^\lambda)$.

Theorem 1.6. *Under the assumptions of Theorem 1.5, for any $\lambda \geq 1$ and $x > 0$*

$$\mathbb{P}\left(\|\hat{f}^\lambda - f\|_{L^2}^2 \geq C_2(\Theta, \mathcal{F}, f)A_1(x, J, n, \lambda, \sigma^2, \gamma) + C_3(\Theta, f)A_2(x, J)\right) \leq 4e^{-x},$$

where $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ and $A_2(x, J)$ are defined in Theorem 1.5, $C_2(\Theta, \mathcal{F}, f)$ and $C_3(\Theta, f)$ are positive constants depending only on $\Theta, \mathcal{F}, f$, and $\|\hat{f}^\lambda - f\|_{L^2}^2 = \int_0^1 |f^\lambda(t) - f(t)|^2 dt$.

Similar comments to those made on the consistency of the estimators of the shifts can be made. A double asymptotic in n and J is needed to show that the Fréchet mean $\hat{f}^\lambda$ converges in probability to the true mean pattern f . Moreover, if λ_n is too large (e.g. such that $\lim_{n \rightarrow +\infty} \frac{\lambda_n}{n} \neq 0$, which correspond to undersmoothing), then Theorem 1.6 cannot be used to prove that $\hat{f}^\lambda$ converges to f in probability. This illustrates the fact that, to achieve consistency, a sufficient amount of pre-smoothing is necessary before computing the Fréchet mean (1.8).

1.5.5. A lower bound for the Fréchet mean

From the results of Theorem 1.3, it is expected that the Fréchet mean $\hat{f}^\lambda$ does not converge to f in the setting n fixed and $J \rightarrow +\infty$. To support this argument, consider the following ideal estimator

$$\tilde{f}(t) = \frac{1}{J} \sum_{j=1}^J f_j(t + \hat{\theta}_j^\lambda) = \frac{1}{J} \sum_{j=1}^J f(t - \theta_j^* + \hat{\theta}_j^\lambda), \quad \text{for all } t \in [0, 1], \quad (1.20)$$

where $f_j(t) = f(t - \theta_j^*)$, $j = 1, \dots, J$. This corresponds to the case of an ideal smoothing step from the data (1.13) that would yield $\hat{f}_j = f_j$ for all $j = 1, \dots, J$. Obviously, $\tilde{f}(t)$ is not an estimator since it depends on the unobserved quantities f and θ_j^* , but we can consider it as a benchmark to analyse the converge of the Fréchet mean $\hat{f}^\lambda$ to f .

Theorem 1.7. *Suppose that the assumptions of Theorem 1.3 are satisfied with $\rho < \frac{3}{4\pi}$. Then, for any $n \geq 1$, there exists $J_0 \in \mathbb{N}$ such that $J \geq J_0$ implies*

$$\mathbb{E}[\|\tilde{f} - f\|_{L^2}] \geq C(f, \rho) \frac{n^{-1}\sigma^2}{\|\partial_t f\|_\infty^2 + n^{-1}\sigma^2 \int_\Theta (\partial_\theta \log(g(\theta)))^2 g(\theta) d\theta},$$

where the constant $C(f, \rho) > 0$ depends on f and ρ .

Hence, in the setting n fixed and $J \rightarrow +\infty$, even the ideal estimator $\tilde{f}$ does not converge to f for the expected quadratic risk. This illustrates the central role played by the dimension n of the data to obtain consistent estimators.

1.6. Notations and main assumptions in the general case

1.6.1. Smoothness of the mean pattern and the deformation operators

In this part, the notation $(\mathcal{L}_\theta)_{\theta \in \mathcal{P}}$ is used to denote either $(T_\theta)_{\theta \in \mathcal{P}}$ or their inverse $(\tilde{T}_\theta)_{\theta \in \mathcal{P}}$.

Assumption 1.9. For all $\theta \in \mathcal{P}$, $\mathcal{L}_\theta : L^2(\Omega) \rightarrow L^2(\Omega)$ is a linear operator satisfying $\mathcal{L}_\theta f \in \mathcal{F}$ for all $f \in \mathcal{F}$. There exists a constant $C(\Theta) > 0$ such that for any $f \in L^2(\Omega)$ and $\theta \in \Theta$

$$\|\mathcal{L}_\theta f\|_{L^2}^2 \leq C(\Theta) \|f\|_{L^2}^2,$$

and a constant $C(\mathcal{F}, \Theta) > 0$ such that for any $f \in \mathcal{F}$ and $\theta_1, \theta_2 \in \Theta$,

$$\|\tilde{T}_{\theta_1} f - \tilde{T}_{\theta_2} f\|_{L^2}^2 \leq C(\mathcal{F}, \Theta) \|\theta_1 - \theta_2\|^2.$$

Assumption 1.9 can be interpreted as a Lipschitz condition on the mapping $(f, \theta) \mapsto \mathcal{L}_\theta f$. The first inequality, that is $\|\mathcal{L}_\theta f\|_{L^2}^2 \leq C(\Theta) \|f\|_{L^2}^2$, means that the action of the operator $\mathcal{L}_\theta$ does not change too much the norm of f when θ varies in Θ . Such an assumption on T_θ and its inverse $\tilde{T}_\theta$ forces the optimization problem (1.9) to have non trivial solutions by avoiding the functional $M(\theta)$ in (1.10) being arbitrarily small. It can be easily checked that Assumption 1.9 is satisfied in the case (1.2) of shifted curves with $\mathcal{F} = H_s(A)$ and $s \geq 1$.

1.6.2. The preliminary smoothing step

For $j = 1, \dots, J$ the $\hat{f}_j$'s are supposed to belong to the class of linear estimators in the sense of the following definition:

Definition 1.8. Let Λ denote either $\mathbb{N}$ or $\mathbb{R}_+$ (set of smoothing parameters). To every $\lambda \in \Lambda$ is associated a non-random vector valued function $S_\lambda : \Omega \rightarrow \mathbb{R}^n$ such that for all $j = 1, \dots, J$ and all $t \in \Omega$

$$\hat{f}_j(t) = \hat{f}_j^\lambda(t) = \langle S_\lambda(t), \mathbf{Y}_j \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^n$ and $\mathbf{Y}_j = (Y_j^\ell)_{\ell=1}^n \in \mathbb{R}^n$.

Assumption 1.10. For all $\lambda \in \Lambda$ and all $\ell = 1, \dots, n$, the function $t \mapsto S_\lambda^\ell(t)$ belong to $L^2(\Omega)$, where $S_\lambda^\ell(t)$ denotes the ℓ -th component of the vector $S_\lambda(t)$. Moreover, for all $\lambda \in \Lambda$, $f \in \mathcal{F}$ and $\theta \in \Theta$, the function $t \mapsto \langle S_\lambda(t), \mathbf{T}_\theta f \rangle$ belongs to $\mathcal{F}$ where $\mathbf{T}_\theta f = (T_\theta f(t_\ell))_{\ell=1}^n$.

In the case (1.2) of randomly shifted curves with an equi-spaced design, then Assumption 1.10 holds with

$$S_\lambda(t) = \left[\frac{1}{n} \sum_{|k| \leq \lambda} e^{i2\pi k(t - \frac{\ell}{n})} \right]_{\ell=1}^n.$$

Let us now specify how the bias/variance behavior of the linear estimators $\hat{f}_j^\lambda$ depends on the smoothing parameter λ . For this, consider for some function $f \in \mathcal{F}$ the following regression model

$$Y^\ell = f(t_\ell) + \sigma \varepsilon^\ell, \ell = 1, \dots, n,$$

where the ε_ℓ 's are i.i.d normal variables with zero mean and variance 1. The performances of a linear estimator $\hat{f}^\lambda(t) = \langle S_\lambda(t), \mathbf{Y} \rangle$, where $\mathbf{Y} = (Y_\ell)_{\ell=1}^n$, can be evaluated in term of the expected quadratic risk $R_\lambda(\hat{f}^\lambda, f)$ defined by

$$R_\lambda(\hat{f}^\lambda, f) := \mathbb{E} \|\hat{f}^\lambda - f\|_{L^2}^2 = \int_{\Omega} |B_\lambda(f, t)|^2 dt + \sigma^2 \int_{\Omega} V_\lambda(t) dt,$$

where B_λ and V_λ denote the usual bias and variance of $\hat{f}^\lambda$ given by $B_\lambda(f, t) = \langle S_\lambda(t), \mathbf{f} \rangle - f(t)$ and $V_\lambda(t) = \|S_\lambda(t)\|_{\mathbb{R}^n}^2$, for $t \in \Omega$, where $\mathbf{f} = (f(t_\ell))_{\ell=1}^n$. Define also $V(\lambda) = \int_{\Omega} V_\lambda(t) dt$, and let us make the following assumption on the asymptotic behavior of the bias/variance of $\hat{f}^\lambda$:

Assumption 1.11. There exist a constant $\kappa(\mathcal{F}) > 0$ and a real-valued function $\lambda \mapsto B(\lambda)$, such that for all $f \in \mathcal{F}$,

$$\|B_\lambda(f, \cdot)\|_{L^2}^2 = \|\langle S_\lambda(\cdot), \mathbf{f} \rangle - f(\cdot)\|_{L^2}^2 \leq \kappa(\mathcal{F}) B(\lambda).$$

Moreover there exists a sequence of smoothing parameters $(\lambda_n)_{n \in \mathbb{N}} \in \Lambda^{\mathbb{N}}$ with $\lim_{n \rightarrow +\infty} \lambda_n = +\infty$ such that $\lim_{n \rightarrow +\infty} B(\lambda_n) = 0$ and $\lim_{n \rightarrow +\infty} V(\lambda_n) = 0$.

Let us illustrate Assumption 1.11 in the case of shifted curves with an equi-spaced design, and a smoothing step obtained by low-pass Fourier filtering. As in Section 1.5, take $\mathcal{F} = H_s(A)$ defined in (1.18). In this setting, $V(\lambda) = \frac{2\lambda+1}{n}$. It can be also checked that $\|B_\lambda(f, \cdot)\|_{L^2}^2 \leq C(A)B(\lambda)$ for some positive constant $C(A)$ depending only on A , and $B(\lambda) = \frac{2\lambda+1}{n} + \lambda^{-2s}$. Thus, Assumption 1.11 holds with $\lambda_n = n^{\frac{1}{2s+1}}$.

1.6.3. Random perturbation of the mean pattern f by the Z_j 's

Assumption 1.12. For any $n \geq 1$, there exists a real $\gamma_n(\Theta) > 0$ such that for any $\theta \in \Theta$

$$\gamma_{\max}(\mathbb{E}_\theta[\mathbf{T}_\theta \mathbf{Z}(\mathbf{T}_\theta \mathbf{Z})']) \leq \gamma_n(\Theta)$$

where $\mathbf{T}_\theta \mathbf{Z} = (T_\theta Z(t_\ell))_{\ell=1}^n \in \mathbb{R}^n$, and $\gamma_{\max}(A)$ denotes the largest eigenvalue of a symmetric matrix A . Moreover,

$$\lim_{n \rightarrow \infty} \gamma_n(\Theta) \sqrt{V(\lambda_n)} = 0, \tag{1.21}$$

where $V(\lambda_n)$ is the variance defined in Assumption 1.11.

Intuitively, the condition (1.21) means that the variance of the linear smoother $S_\lambda(\cdot)$ has to be asymptotically smaller than the maximal correlations (measured by $\gamma_n(\Theta)$) between $T_\theta Z(t_\ell)$ and $T_\theta Z(t_{\ell'})$ for $\ell, \ell' = 1, \dots, n$ and all $\theta \in \Theta$. In the case of randomly shifted curves with an equi-spaced design, a condition for which Assumption 1.12 holds is given by Assumption 1.8 (see the arguments in Section 1.5.1).

1.7. Consistency in the general case

1.7.1. Consistent estimation of the deformation parameters

Consider for $\lambda \in \Lambda$ the following estimator of the deformation parameters

$$\hat{\boldsymbol{\theta}}^\lambda = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} M_\lambda(\boldsymbol{\theta}),$$

where

$$M_\lambda(\boldsymbol{\theta}) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\boldsymbol{\theta}_j} \langle S_\lambda(t), \mathbf{Y}_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\boldsymbol{\theta}_{j'}} \langle S_\lambda(t), \mathbf{Y}_{j'} \rangle \right)^2 dt, \quad (1.22)$$

and Θ is the constrained set introduced in Assumption 1.6. The estimator $\hat{\boldsymbol{\theta}}^\lambda$ thus depends on the choice of Θ , and it will be shown that $\hat{\boldsymbol{\theta}}^\lambda$ is a consistent estimator of the vector $\boldsymbol{\theta}_\Theta^* \in \mathbb{R}^{pJ}$ defined in Assumption 1.6. Note that depending on the problem at hand and the choice of the constrained set Θ , it can be shown that $\boldsymbol{\theta}_\Theta^*$ is close to the true deformation parameters $\boldsymbol{\theta}^*$. For example, in the case of shifted curves, if $\Theta = \Theta_0$ defined in (1.14) and if the density g of the shifts has zero mean, then $\boldsymbol{\theta}_{\Theta_0} = (\boldsymbol{\theta}_1^* - \bar{\boldsymbol{\theta}}^*, \dots, \boldsymbol{\theta}_J^* - \bar{\boldsymbol{\theta}}^*)$ with $\bar{\boldsymbol{\theta}}^* = \frac{1}{J} \sum_{j=1}^J \boldsymbol{\theta}_j^*$ can be shown to be close to $\boldsymbol{\theta}^*$ (see Lemma 1.11 in the Appendix). This allows to show the consistency of $\hat{\boldsymbol{\theta}}^\lambda$ to $\boldsymbol{\theta}^*$ as formulated in Theorem 1.5. Therefore, the next result only bounds the distance between $\hat{\boldsymbol{\theta}}^\lambda$ and $\boldsymbol{\theta}_\Theta^*$.

Theorem 1.9. *Consider the model (1.5) and suppose that Assumptions 1.1, 1.6, 1.7 and 1.9 to 1.12 hold with $n \geq 1$ and $J \geq 2$. Then, for any $\lambda \in \Lambda$ and $x > 0$*

$$\begin{aligned} \mathbb{P} \left(\frac{1}{J} \|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_\Theta^*\|_{\mathbb{R}^{pJ}}^2 \geq C_1(\Theta, \Theta, \mathcal{F}, f) \left[(\gamma_n(\Theta) + \sigma^2) \left(\sqrt{v(x, J, \lambda)} + v(x, J, \lambda) \right) \right. \right. \\ \left. \left. + \left(\sqrt{B(\lambda)} + B(\lambda) \right) \right] \right) \leq 2e^{-x}, \end{aligned}$$

with $C_1(\Theta, \Theta, \mathcal{F}, f) > 0$, $v(x, J, \lambda) := V(\lambda) \left(1 + 4\frac{x}{J} + \sqrt{4\frac{x}{J}} \right)$.

Using Assumptions 1.11 and 1.12, it follows that $\lim_{n \rightarrow +\infty} \gamma_n(\Theta) \left(\sqrt{v(x, J, \lambda_n)} + v(x, J, \lambda_n) \right) = 0$ for any $x > 0$ and $J \geq 2$. If J remains fixed, Theorem 1.9 thus implies that $\hat{\boldsymbol{\theta}}^\lambda$ converges in probability to $\boldsymbol{\theta}_\Theta^*$ as $n \rightarrow +\infty$. To the contrary, let us fix n , and consider an asymptotic setting where only $J \rightarrow +\infty$. For any $x > 0$ and $\lambda \in \Lambda$, $\lim_{J \rightarrow +\infty} v(x, J, \lambda) = V(\lambda)$. Therefore, Theorem 1.9 cannot be used to prove that $\hat{\boldsymbol{\theta}}^\lambda$ converges to $\boldsymbol{\theta}_\Theta^*$ as $J \rightarrow +\infty$. This confirms that $\hat{\boldsymbol{\theta}}^\lambda$ is not a consistent estimator of $\boldsymbol{\theta}_\Theta^*$ (and thus of $\boldsymbol{\theta}^*$) as n remains fixed and J tends to infinity.

1.7.2. Consistent estimation of the mean pattern

Recall that the estimator $\hat{f}^\lambda$ of the mean pattern f is defined as $\hat{f}^\lambda = \frac{1}{J} \sum_{j=1}^J \tilde{T}_{\hat{\boldsymbol{\theta}}_j^\lambda} \hat{f}_j^\lambda$. We study the consistency of $\hat{f}^\lambda$ with respect to the shape function

$$f_\Theta^* := \frac{1}{J} \sum_{j=1}^J \tilde{T}_{[\boldsymbol{\theta}_\Theta^*]_j} T_{\boldsymbol{\theta}_j^*} f,$$

defined for $\boldsymbol{\theta}_{\Theta}^* = ([\boldsymbol{\theta}_{\Theta}^*]_1, \dots, [\boldsymbol{\theta}_{\Theta}^*]_J)$. Again, depending on the problem at hand and the choice of the constrained set Θ , it can be shown that f_{Θ}^* is close to the true mean pattern f . For example, in the case of shifted curves with $\Theta = \Theta_0$ defined in (1.14), then $\boldsymbol{\theta}_{\Theta_0} = (\boldsymbol{\theta}_1^* - \bar{\boldsymbol{\theta}}^*, \dots, \boldsymbol{\theta}_J^* - \bar{\boldsymbol{\theta}}^*)$ with $\bar{\boldsymbol{\theta}}^* = \frac{1}{J} \sum_{j=1}^J \boldsymbol{\theta}_j^*$. In this case $f_{\Theta_0}^*(t) := \frac{1}{J} \sum_{j=1}^J f(t - \boldsymbol{\theta}_j^* + [\boldsymbol{\theta}_{\Theta_0}^*]_j) = f(t - \bar{\boldsymbol{\theta}}^*)$. Hence, under the condition that $\int_{\Theta} \boldsymbol{\theta} g(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0$, then $\bar{\boldsymbol{\theta}}^* \approx 0$ for J sufficiently large, and thus $f_{\Theta}^*(t)$ is close to f which allows to show the consistency of $\hat{f}^\lambda$ to f as formulated in Theorem 1.6.

Theorem 1.10. *Consider the model (1.5) and suppose that Assumptions 1.1, 1.6, 1.7 and 1.9 to 1.12 hold. Then, for any $\lambda \in \Lambda$ and $x > 0$*

$$\begin{aligned} \mathbb{P}\left(\|\hat{f}^\lambda - f_{\Theta}^*\|_{L^2}^2 \geq C_2(\Theta, \Theta, \mathcal{F}, f) \left[(\gamma_n(\Theta) + \sigma^2) \left(\sqrt{v(x, J, \lambda)} + v(x, J, \lambda) \right) \right. \right. \\ \left. \left. + \left(\sqrt{B(\lambda)} + B(\lambda) \right) \right] \right) \leq 2e^{-x}, \end{aligned}$$

where $C_2(\Theta, \Theta, \mathcal{F}, f) > 0$ is a constant depending only Θ , Θ , $\mathcal{F}$, and f .

The consistency of $\hat{f}^\lambda$ to f_{Θ}^* is thus guaranteed when n goes to infinity provided the level of smoothing $\lambda = \lambda_n$ is chosen so that $\lim_{n \rightarrow +\infty} V(\lambda_n) = \lim_{n \rightarrow +\infty} B(\lambda_n) = 0$. Again, if n remains fixed and only J is let going to infinity then Theorem 1.10 cannot be used to prove the convergence of $\hat{f}^\lambda$ to f_{Θ}^* .

1.8. Numerical experiments for randomly shifted curves

1.8.1. A general gradient descent algorithm

To compute the estimator $\hat{\boldsymbol{\theta}}^\lambda$ one has to minimize the criterion $M_\lambda(\boldsymbol{\theta})$ defined in (1.22). Generally, the expression of the gradient of $M_\lambda(\boldsymbol{\theta})$ is available in a closed form, and thus a gradient descent algorithm with an adaptive step can be easily implemented. More precisely the algorithm is composed of the following steps:

Initialization: let $\boldsymbol{\theta}^0 \in \Theta$ (if $\Theta = \Theta_0$ a possible choice is $\boldsymbol{\theta}^0 = 0$), $\gamma_0 = \frac{1}{\|\nabla M_\lambda(\boldsymbol{\theta}^0)\|}$, $F(0) = M_\lambda(\boldsymbol{\theta}^0)$, and set $m = 0$.

Step 1: let $\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^m - \gamma_m \nabla M_\lambda(\boldsymbol{\theta}^m)$ and $F(m+1) = M_\lambda(\boldsymbol{\theta}^{new})$.

While $F(m+1) > F(m)$ do

$\gamma_m = \gamma_m / \kappa$, and $\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^m - \gamma_m \nabla M_\lambda(\boldsymbol{\theta}^m)$, and let $F(m+1) = M_\lambda(\boldsymbol{\theta}^{new})$.

End while

Then, take $\boldsymbol{\theta}^{m+1} = \boldsymbol{\theta}^{new}$. Set $m = m + 1$.

Step 2: if $F(m) - F(m+1) \geq \rho(F(1) - F(m+1))$ then return to Step 2, else stop the iterations, and take $\hat{\boldsymbol{\theta}}^\lambda = \boldsymbol{\theta}^{m+1}$.

In the above algorithm, $\rho > 0$ is a small stopping parameter and $\kappa > 1$ is a parameter to control the choice of the adaptive step γ_m . In practice, one does not know the compact set Θ to which the $\boldsymbol{\theta}_j^*$ are supposed to belong. So it is difficult to constraint $\boldsymbol{\theta}^m$ to belong to a specific compact set. Nevertheless, depending on the choice of the constrained set Θ , it is possible to impose that $\boldsymbol{\theta}^m$

satisfies the same constraints. For instance, if $\Theta = \Theta_0$ one can modify $\theta^m = (\theta_j^m)_{j=1}^J$ at each iteration such that $\sum_{j=1}^J \theta_j^m = 0$.

1.8.2. Randomly shifted curves

Consider the model (1.13) with random shifts θ_j having a uniform density g with compact support equal to $[-\frac{1}{5}, \frac{1}{5}]$, and $f(t) = 9 \sin(2\pi t) + 2 \cos(8\pi t)$ for $t \in [0, 1]$ as a mean pattern, see Figure 1.2(a). For the constrained set we took

$$\Theta_0 = \left\{ \theta \in \left[-\frac{1}{2}, \frac{1}{2}\right]^J, \theta_1 + \dots + \theta_J = 0 \right\}.$$

We use Fourier low pass filtering with spectral cut-off to $\lambda = 7$ which is reasonable value to reconstruct f representing a good tradeoff between bias and variance. We present some results of simulations under various assumptions of the process Z and the level σ of additive noise in the measurements.

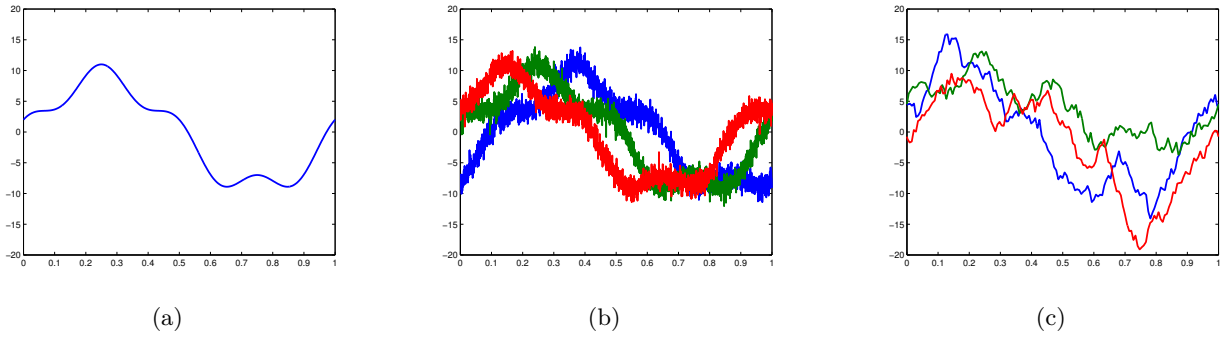


Figure 1.2: (a) mean pattern f . (b) $J = 3$ noisy curves in the SIM with $\sigma = 2$. (c) $J = 3$ noisy curves with $\sigma = 0$ and a stationary process Z with $\zeta = 4$.

Shape invariant model (SIM). The first numerical applications illustrate the role of n and J in the SIM model. Figure 1.2(b) gives a sample of the data used with $\sigma = 2$. The factors in the simulations are the number J of curves and the number of design points n . For each combination of these two factors, we simulate $M = 20$ repetitions of model (1.13). For each repetition we computed $\frac{1}{J} \|\hat{\theta}^\lambda - \theta^*\|^2$ and $\|\hat{f}^\lambda - f\|_{L^2}^2$. Boxplot of these quantities are displayed in Figure 1.3(a) and 1.3(b) respectively, for $J = 20, 40, \dots, 100$ and $n = 512$ (in gray) and $n = 1024$ (in black). As the smoothing parameter is fixed to $\lambda = 7$, increasing n simply reduces the variance of the linear smoothers $\hat{f}_j^\lambda$. Recall that the lower bound given in Theorem 1.3 shows that $\frac{1}{J} \mathbb{E}[\|\theta^* - \hat{\theta}^\lambda\|^2]$ does not decrease as J increases but should be smaller when the number of point n increases. This is exactly what we observe in Figure 1.3. Similarly, the quantity $\|\hat{f}^\lambda - f\|_{L^2}^2$ is clearly smaller with $n = 1024$ than with $n = 512$.

Complete model. We now add the terms Z_j in (1.13) to model linear variations in amplitude of the curves around the template f . First, we generate a stationary periodic Gaussian process. To do this, the covariance matrix must be a particular Toeplitz matrix. As suggested in [Gre93] one possibility is to choose

$$R(t) = \varsigma^2 \frac{e^{\phi(t-1/2)} + e^{-\phi(t-1/2)}}{e^{\phi/2} + e^{-\phi/2}},$$

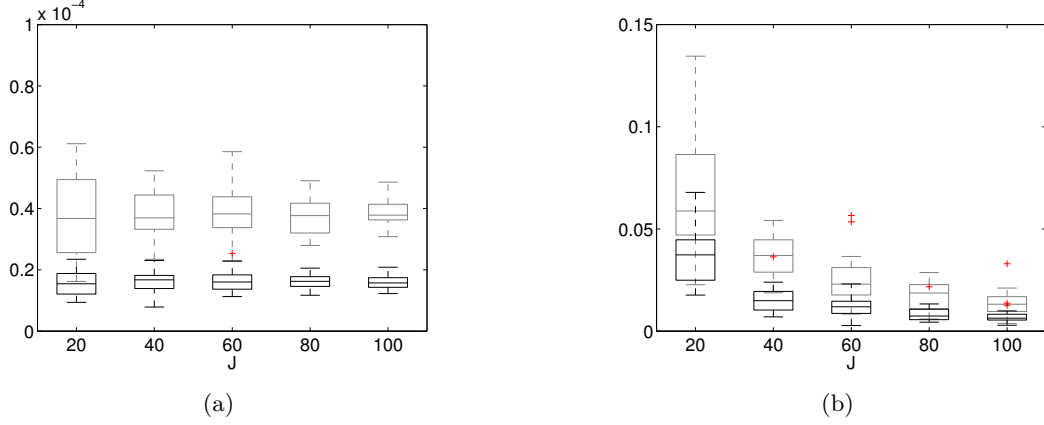


Figure 1.3: Boxplot of $\frac{1}{J}\|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ (a) and $\|\hat{f}^\lambda - f_{\Theta_0}\|_{L^2}^2$ (b) over $M = 20$ repetitions from a SIM model of shifted curves. Boxplot in gray correspond to $n = 512$, and in black to $n = 1024$.

where ϕ is a strictly positive parameter (we took $\phi = 4$) and ς a variance parameter. The level of additive noise is $\sigma = 8$, and we took $\varsigma = 4$. As an illustration, in Figure 1.2(c) we plot $f + Z_j$, $j = 1, 2, 3$ with $\varsigma = \phi = 4$. Over $M = 20$ repetitions, we have computed the values of $\frac{1}{J}\|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ and $\|\hat{f}^\lambda - f_{\Theta_0}\|_{L^2}^2$ for J is varying from 20 to 100 and $n = 512, 1024$. The results are displayed in Figure 1.4(a) and 1.4(b). We observe the same behaviors than in the simulations with the SIM model: the variance of $\frac{1}{J}\|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ does not decrease as J increases (see Figure 1.4(a)) and $\|\hat{f}^\lambda - f_{\Theta_0}\|_{L^2}^2$ has a smaller mean and variance as n increases.

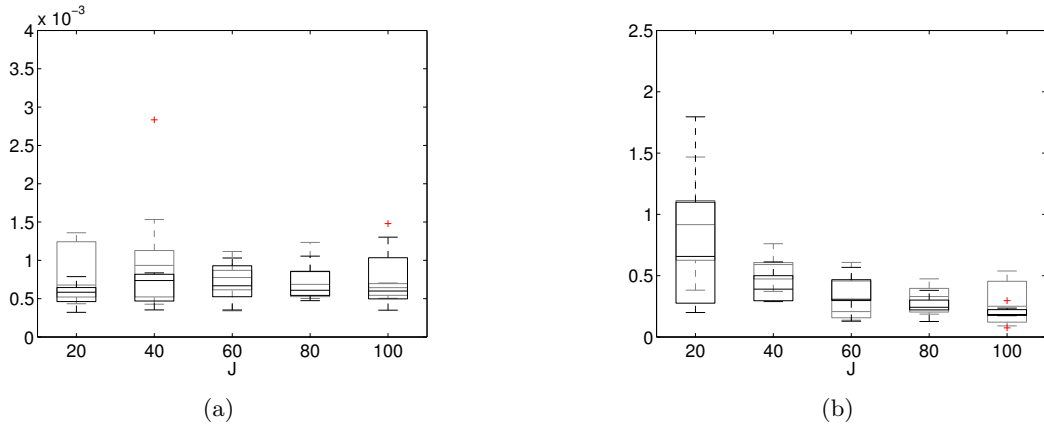


Figure 1.4: Boxplot of $\frac{1}{J}\|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ (a) and $\frac{1}{J}\|\hat{f}^\lambda - f_{\Theta_0}\|_{L^2}^2$ (b) in model (1.13) with a stationary error term Z . Boxplot in gray correspond to $n = 512$, and in black to $n = 1024$.

We finally run the same simulations with a non stationary noise $Z_j(t) = \alpha_j \psi(t)$ where ψ is a positive periodic smooth deterministic function such that $\|\psi\|_{L^2} = 1$ and $\alpha_j \sim \mathcal{N}(0, \varsigma^2)$ with $\varsigma = 4$. Note that, in this case, the sequence $\gamma_n(\Theta)$ is of order n and Assumption 1.12 is not verified. The levels of noise (σ and ς) are the same than in the stationary case in order to make things comparable. The results are presented in the same manner in Figure 1.5(a) for $\frac{1}{J}\|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ and in Figure 1.5(b) for $\|\hat{f}^\lambda - f_{\Theta_0}\|_{L^2}^2$. One can see that the results are very different. The estimators of the shifts have

a much larger mean and variance, and the variance of $\frac{1}{J}\|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ remains rather high even when n or J increases (see Figure 1.5(a)). The convergence to zero of $\|\hat{f}^\lambda - f_{\Theta_0}\|_{L_2}^2$ which was clear in the SIM case, is now not so obvious in view of the numerical results displayed in Figure 1.5(b).

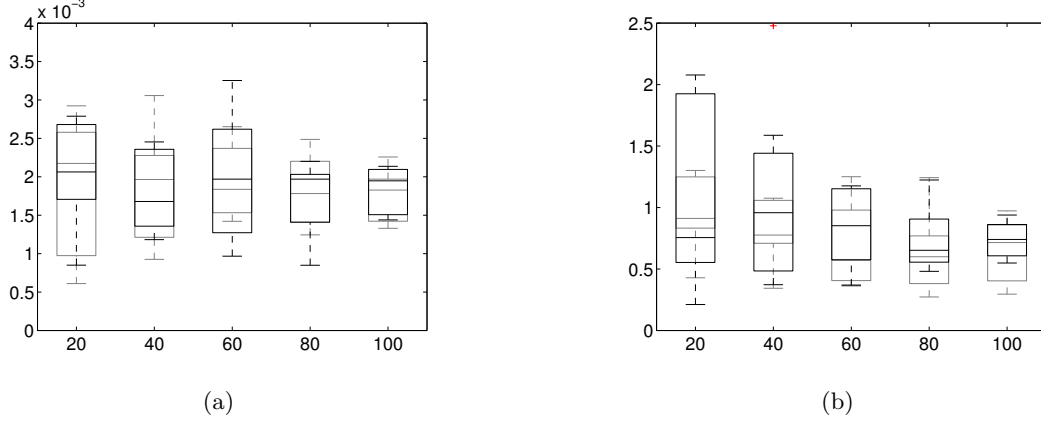


Figure 1.5: Boxplot of $\frac{1}{J}\|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ (a) and $\frac{1}{J}\|\hat{f}^\lambda - f_{\Theta_0}\|^2$ (b) in model (1.13) with a non-stationary error term Z . Boxplot in gray correspond to $n = 512$, and in black to $n = 1024$.

1.9. Conclusion and perspectives

We have proposed to use a Fréchet mean of smoothed data to estimate a mean pattern of curves or images satisfying a non-parametric regression model including random deformations. Upper and lower bounds (in probability and expectation) for the estimation of the deformation parameters and the mean pattern have been derived. Our main result is that these bounds go to zero as the dimension n of the data (the number of sample points) goes to infinity, but that an asymptotic setting only in J (the number of observed curves or images) is not sufficient to obtain consistent estimators. To sum up the above discussion, we compile the results on the convergence (in probability or in expectation) of $\hat{\theta}^\lambda$ and $\hat{f}^\lambda$ in the following diagram,

$$\begin{array}{cccc}
 \hat{\theta}^\lambda \xrightarrow{n,J} \theta^* & \hat{\theta}^\lambda \not\xrightarrow{J} \theta^* \text{ (} n \text{ fixed)} & \hat{\theta}^\lambda \xrightarrow{n} \theta_{\Theta_0}^* \text{ (} J \text{ fixed)} & \hat{\theta}^\lambda \not\xrightarrow{J} \theta_{\Theta_0}^* \text{ (} n \text{ fixed)} \\
 & \hat{f}^\lambda \xrightarrow{n,J} f & \hat{f}^\lambda \xrightarrow{n} f_{\Theta_0} \text{ (} J \text{ fixed)} &
 \end{array}$$

where a parameter (n or J) appearing under an arrow means that it is let going to infinity while the other remains fixed. A crossed arrow means that consistency is not guaranteed.

An interesting topic for future investigation would be to study the rate of convergence of such estimators and to analyze their optimality (e.g. from a minimax point of view).

1.A. Proof of the results in Section 1.3

1.A.1. Proof of Theorem 1.1

Write $\boldsymbol{\theta}_j^* = ([\theta^*]_j^1, \dots, [\theta^*]_j^p) \in \Theta$, and let $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_J) \in \mathbb{R}^{nJ}$ be the column vector of the observations generated by model (1.7). Conditionally to $\boldsymbol{\theta}^*$, $\mathbf{Y}$ is a Gaussian vector and its log-likelihood is equal to

$$\log(p(\mathbf{Y}|\boldsymbol{\theta}^*)) = -\frac{Jn}{2} \log(2\pi) - \frac{J}{2} \log(\det(\Lambda)) - \frac{1}{2} \sum_{j=1}^J (\mathbf{Y}_j - \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f})' \Lambda (\mathbf{Y}_j - \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f}), \quad (1.23)$$

where $\Lambda = \sigma^{-2} Id_n$. Therefore, we have the expected score $\mathbb{E}_{\boldsymbol{\theta}^*}[\partial_{[\theta^*]_{j_1}^{p_1}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*))] = 0$ for all $j_1 = 1, \dots, J$ and $p_1 = 1, \dots, p$ and

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}^*} [\partial_{[\theta^*]_{j_1}^{p_1}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*)) \partial_{[\theta^*]_{j_2}^{p_2}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*))] \\ = \begin{cases} 0 & \text{if } j_1 \neq j_2, \\ -[(\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f})' \Lambda (\partial_{[\theta^*]_{j_1}^{p_2}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f})]_{p_1, p_2=1}^p & \text{if } j_1 = j_2, \end{cases} \end{aligned} \quad (1.24)$$

where $\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f} = [\partial_{[\theta^*]_{j_1}^{p_1}} T_{\boldsymbol{\theta}_{j_1}^*} f(t_\ell)]_{\ell=1}^n$. Then, for each $j_1 = 1, \dots, J$ and $p_1 = 1, \dots, p$ we have

$$(\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f})' \Lambda (\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}) \leq \sigma^{-2} \|\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}\|^2 \leq C(\Theta, f) n \sigma^{-2}, \quad (1.25)$$

where the last inequality is a consequence of Assumption 1.2. From now on, $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{Y}) = (\hat{\boldsymbol{\theta}}_1(\mathbf{Y}), \dots, \hat{\boldsymbol{\theta}}_1(\mathbf{Y}))$ is an arbitrary estimator (*i.e.* any measurable function of $\mathbf{Y}$) of the true parameter $\boldsymbol{\theta}^*$. Let also

$$U = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \quad \text{and} \quad V = \left[[\partial_{[\theta^*]_1^{p_1}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*)g(\boldsymbol{\theta}^*))]_{p_1=1}^p, \dots, [\partial_{[\theta^*]_J^{p_1}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*)g(\boldsymbol{\theta}^*))]_{p_1=1}^p \right]$$

be a matrix of column vectors of $\mathbb{R}^{pJ}$. Then, Cauchy-Schwarz inequality implies

$$(\mathbb{E}[U'V])^2 \leq \mathbb{E}[U'U] \mathbb{E}[V'V]. \quad (1.26)$$

In the sequel we note $g^J(\boldsymbol{\theta}) d\boldsymbol{\theta} = g(\boldsymbol{\theta}_1) \dots g(\boldsymbol{\theta}_J) d\boldsymbol{\theta}_1 \dots d\boldsymbol{\theta}_J$. We have

$$\begin{aligned} \mathbb{E}[U'V] &= \sum_{j=1}^J \sum_{p_1=1}^p \int_{\mathbb{R}^{nJ}} \int_{\Theta^J} (\hat{\theta}_j^{p_1}(y) - [\theta]_j^{p_1}) \partial_{[\theta^*]_j^{p_1}} (p(y|\boldsymbol{\theta}) g^J(\boldsymbol{\theta})) d\boldsymbol{\theta} dy \\ &= \sum_{j=1}^J \sum_{p_1=1}^p \int_{\mathbb{R}^{nJ}} \hat{\theta}_j^{p_1}(y) \int_{\Theta^J} \partial_{[\theta^*]_j^{p_1}} (p(y|\boldsymbol{\theta}) g^J(\boldsymbol{\theta})) d\boldsymbol{\theta} dy \\ &\quad - \sum_{j=1}^J \sum_{p_1=1}^p \int_{\mathbb{R}^{nJ}} \int_{\Theta^J} [\theta]_j^{p_1} \partial_{[\theta^*]_j^{p_1}} (p(y|\boldsymbol{\theta}) g^J(\boldsymbol{\theta})) d\boldsymbol{\theta} dy \end{aligned}$$

Assumption 1.1 and the differentiability of the density g imply that for all $p_1 = 1, \dots, p$ and all $\boldsymbol{\theta} \in \Theta$ we have $\lim_{\theta^{p_1} \rightarrow \rho} g(\boldsymbol{\theta}) = 0$. Then, an integration by part together with Fubini's theorem give $\int_{\Theta^J} \partial_{[\theta^*]_j^{p_1}} (p(y|\boldsymbol{\theta}) g^J(\boldsymbol{\theta})) d\boldsymbol{\theta} = 0$. Again, with the same arguments, $\int_{\Theta^J} [\theta]_j^{p_1} \partial_{[\theta^*]_j^{p_1}} (p(y|\boldsymbol{\theta}) g^J(\boldsymbol{\theta})) d\boldsymbol{\theta} = - \int_{\Theta^J} p(y|\boldsymbol{\theta}) g^J(\boldsymbol{\theta}) d\boldsymbol{\theta}$ and thus

$$\mathbb{E}[U'V] = pJ.$$

Now, using that the expected score is zero and equation (1.24) we have

$$\begin{aligned}\mathbb{E}[V'V] &= \sum_{j=1}^J \sum_{p_1=1}^p \mathbb{E}[(\partial_{[\theta^*]_{j_1}^{p_1}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*)))^2] + \mathbb{E}[(\partial_{[\theta^*]_{j_1}^{p_1}} \log(g(\boldsymbol{\theta}^*)))^2] \\ &= \sum_{j=1}^J \sum_{p_1=1}^p \int_{\Theta} (\partial_{[\theta]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_j} \mathbf{f})' \Lambda (\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}} \mathbf{f}) g^J(\boldsymbol{\theta}) d\boldsymbol{\theta} + J \int_{\Theta} \|\partial_{\boldsymbol{\theta}_1} \log(g(\boldsymbol{\theta}_1))\|^2 g(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1.\end{aligned}$$

where $\partial_{\boldsymbol{\theta}_1} \log(g(\boldsymbol{\theta}_1)) = [\partial_{[\boldsymbol{\theta}_1]_1^1} \log(g(\boldsymbol{\theta}_1)), \dots, \partial_{[\boldsymbol{\theta}_1]_1^p} \log(g(\boldsymbol{\theta}_1))] \in \mathbb{R}^p$. Then, using inequality 1.25, it gives

$$\mathbb{E}[V'V] \leq pJnC(\Theta, f)\sigma^{-2} + J \int_{\Theta} \|\partial_{\boldsymbol{\theta}_1} \log(g(\boldsymbol{\theta}_1))\|^2 g(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1,$$

Hence, using equation (1.26) for any estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{Y})$ (see Theorem 1 in [GL95])

$$\begin{aligned}\mathbb{E}[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2] &\geq \frac{pJ}{nC(\Theta, f)\sigma^{-2} + p^{-1} \int_{\Theta} \|\partial_{\boldsymbol{\theta}_1} \log(g(\boldsymbol{\theta}_1))\|^2 g(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1} \\ &\geq \frac{\sigma^2 n^{-1} pJ}{C(\Theta, f) + n^{-1} p^{-1} \sigma^2 \int_{\Theta} \|\partial_{\boldsymbol{\theta}_1} \log(g(\boldsymbol{\theta}_1))\|^2 g(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}.\end{aligned}$$

And since $p \geq 1$, the claim in Theorem 1.1 is proved. $\square$

1.A.2. Proof of Theorem 1.2

As above, let $\mathbf{Y} \in \mathbb{R}^{nJ}$ is the column vector generated by model (1.5). Then, conditionally to $\boldsymbol{\theta}^*$, $\mathbf{Y}$ is a Gaussian vectors and Assumption 1.3 ensures that its log-likelihood has the same expression as in equation (1.23) but with

$$\Lambda = \Lambda(\Theta) = (\sigma^2 Id_n + \mathbb{E}_{\boldsymbol{\theta}^*}[\mathbf{T}_{\boldsymbol{\theta}^*} \mathbf{Z}_j (\mathbf{T}_{\boldsymbol{\theta}^*} \mathbf{Z}_j)'])^{-1} = (\sigma^2 Id_n + \Sigma_n(\Theta))^{-1}$$

As the matrix $\Sigma_n(\Theta)$ is positive semi definite with it smallest eigenvalue denoted by $s_n^2(\Theta)$ (see Assumption 1.3), the uniform bound (1.25) becomes

$$(\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f})' \Lambda(\Theta) (\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}) \leq (\sigma^2 + s_n^2(\Theta))^{-1} \|\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}\|^2 \leq C(\Theta, f)n(\sigma^2 + s_n^2(\Theta))^{-1},$$

for all $p_1 = 1, \dots, p$ and $j = 1, \dots, J$. As above the last inequality is a consequence of Assumption 1.2 and the rest of the proof is identical to the proof of Theorem 1.1. $\square$

1.A.3. Proof of Theorem 1.3

For all $\boldsymbol{\theta} \in \mathbb{R}$ the operators $T_{\boldsymbol{\theta}} f(\cdot) = f(\cdot - \boldsymbol{\theta})$ are isometric from $L^2([0, 1])$ to $L^2([0, 1])$ as a change of variable implies immediately that $\|T_{\boldsymbol{\theta}} f\|_{L^2}^2 = \|f\|_{L^2}^2$. For all continuously differentiable function f , we have $\partial_{\boldsymbol{\theta}} T_{\boldsymbol{\theta}} f(t) = -\text{sign}(\boldsymbol{\theta}) \partial_t f(t - \boldsymbol{\theta})$, where $\text{sign}(\cdot)$ is the sign function. Then, for all $\boldsymbol{\theta} \in \Theta$, $\|\partial_{\boldsymbol{\theta}} T_{\boldsymbol{\theta}} f\|_{L^2}^2 = \|\partial_t f\|_{L^2}^2 \leq \|\partial_t f\|_{\infty}^2$ and Assumption 1.2 is satisfied with $C(\Theta, f) = \|\partial_t f\|_{\infty}^2$. Finally, as the error terms Z_j 's are i.i.d stationary random process the covariance function is invariant by the action of the shifts and Assumption 1.3 is satisfied with

$$\Sigma_n(\Theta) = \mathbb{E}_{\boldsymbol{\theta}^*}[\mathbf{T}_{\boldsymbol{\theta}^*} \mathbf{Z}_j (\mathbf{T}_{\boldsymbol{\theta}^*} \mathbf{Z}_j)'] = \left[\mathbb{E} \left[Z \left(\frac{\ell}{n} \right) Z \left(\frac{\ell'}{n} \right) \right] \right]_{\ell, \ell'=1}^n = \left[R \left(\frac{|\ell - \ell'|}{n} \right) \right]_{\ell, \ell'=1}^n$$

where $R : [0, 1] \rightarrow \mathbb{R}$ is the covariance function of the random process Z . Then, the result of Theorem 1.3 follows as an application of Theorem 1.2. $\square$

1.B. Proof of the results in Section 1.4

1.B.1. Proof of Proposition 1.4

Remark that $D(\boldsymbol{\theta}) = \sum_{k \in \mathbb{Z}} |c_k^*|^2 \left(1 - \left| \frac{1}{J} \sum_{j=1}^J e^{i2\pi k(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)} \right|^2\right)$, where $c_k^* = \int_0^1 f(t) e^{-i2\pi kt} dt$. Thanks to Assumption 1.4, it follows that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$D(\boldsymbol{\theta}) \geq |c_1^*|^2 \left(1 - \left| \frac{1}{J} \sum_{j=1}^J e^{i2\pi(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)} \right|^2\right) \quad (1.27)$$

with $c_1^* \neq 0$. Then, remark that

$$\left| \frac{1}{J} \sum_{j=1}^J e^{i2\pi(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)} \right|^2 = \frac{1}{J} + \frac{2}{J^2} \sum_{j=1}^{J-1} \sum_{j'=j+1}^J \cos\left(2\pi \left((\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*)\right)\right).$$

Using a second order Taylor expansion and the mean value theorem, one has that $\cos(2\pi u) \leq 1 - C(\rho)|u|^2$ for any real u such that $|u| \leq 4\rho < 1/4$ with $C(\rho) = 2\pi^2 \cos(8\pi\rho) > 0$. Therefore, the above equality implies that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$\begin{aligned} \left| \frac{1}{J} \sum_{j=1}^J e^{i2\pi(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)} \right|^2 &\leq \frac{1}{J} + \frac{2}{J^2} \sum_{j=1}^{J-1} \sum_{j'=j+1}^J 1 - C(\rho) \left| (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*) \right|^2 \\ &\leq 1 - \frac{2}{J^2} \sum_{j=1}^{J-1} \sum_{j'=j+1}^J C(\rho) \left| (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*) \right|^2, \end{aligned}$$

since $|(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*)| \leq 4\rho < 1/4$ for all $m, q = 1, \dots, n$ by Assumption 1.5 and the hypothesis that $\rho < 1/16$. Hence, using the lower bound (1.27), it follows that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$D(\boldsymbol{\theta}) \geq C(f, \rho) \frac{1}{J^2} \sum_{j=1}^{J-1} \sum_{j'=j+1}^J \left| (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*) \right|^2 \quad (1.28)$$

with $C(f, \rho) = 2|c_1^*|^2 C(\rho)$. Now assume that $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$. Using the properties that $\sum_{j=1}^J \boldsymbol{\theta}_j = 0$ and $\sum_{j=1}^J (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) = -\sum_{j=1}^J \boldsymbol{\theta}_j^* = J\bar{\boldsymbol{\theta}}^*$, it follows from elementary algebra that

$$\frac{1}{J} \sum_{j=1}^{J-1} \sum_{j'=j+1}^J \left| (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*) \right|^2 = \sum_{j=1}^J (\boldsymbol{\theta}_j - (\boldsymbol{\theta}_j^* - \bar{\boldsymbol{\theta}}^*))^2.$$

The above equality together with the lower bound (1.28) completes the proof. $\square$

1.C. Proof of the results in Section 1.5

1.C.1. Proof of Theorem 1.5

Let us state the following lemma which is direct consequence of Bernstein's inequality for bounded random variables (see e.g. Proposition 2.9 in [Mas07]):

Lemma 1.11. *Suppose that Assumption 1.5 holds. Then, for any $x > 0$*

$$\mathbb{P}\left(\frac{1}{J} \|\boldsymbol{\theta}_{\boldsymbol{\Theta}_0}^* - \boldsymbol{\theta}^*\|^2 \geq \rho^2 \left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2\right) \leq 2e^{-x}.$$

Using the inequality $\frac{1}{J}\|\hat{\theta}^\lambda - \theta^*\|^2 \leq \frac{2}{J}\|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2 + \frac{2}{J}\|\theta_{\Theta_0}^* - \theta^*\|^2$, it follows that Theorem 1.5 is a consequence of Lemma 1.11 and Theorem 1.9. Indeed, it can be easily checked that, under the assumptions of Theorem 1.5, Assumptions 1.9 to 1.12 are satisfied in the case of randomly shifted curves with an equi-spaced design and low-pass Fourier filtering, see the various arguments given in Sections 1.5 and 1.6. The identifiability condition of Assumption 1.7 is given by Proposition 1.4. $\square$

1.C.2. Proof of Theorem 1.6

Consider the following inequality $\|\hat{f}^\lambda - f\|^2 \leq 2\|\hat{f}^\lambda - f_{\Theta_0}\|^2 + 2\|f_{\Theta_0} - f\|^2$, where $f_{\Theta_0}(t) = f(t - \bar{\theta}^*)$ and $\bar{\theta}^* = \frac{1}{J} \sum_{j=1}^J \theta_j^* \in \Theta$. As f is assumed to be in $H_s(A)$, there exists a constant $C(\Theta, f) > 0$ such that $\|f_{\Theta_0} - f\|_{L^2}^2 \leq C(\Theta, f)|\bar{\theta}^*|^2 = C(\Theta, f)\frac{1}{J}\|\theta_{\Theta_0}^* - \theta^*\|^2$. As explained in part 1.C.1 the assumptions of Theorem 1.6 are satisfied in the case of randomly shifted curves with an equi-spaced design and low-pass Fourier filtering. The result then follows from Theorem 1.10. $\square$

1.C.3. Proof of Theorem 1.7

Let $n \geq 1$. We have that

$$\mathbb{E}[\|\tilde{f} - f\|_{L^2}] = \mathbb{E}\|\tilde{f} - f_{\Theta_0} + f_{\Theta_0} - f\|_{L^2} \geq \left| \underbrace{\mathbb{E}\|\tilde{f} - f_{\Theta_0}\|_{L^2}}_{\mathbf{A}} - \underbrace{\mathbb{E}\|f_{\Theta_0} - f\|_{L^2}}_{\mathbf{B}} \right| \quad (1.29)$$

where for all $t \in [0, 1]$, $\tilde{f}(t) = \frac{1}{J} \sum_{j=1}^J f(t - \theta_j^* + \hat{\theta}_j^\lambda)$, and $f_{\Theta_0}(t) = f(t + \bar{\theta}^*)$, with $\bar{\theta}^* = \frac{1}{J} \sum_{j=1}^J \theta_j^*$. In the rest of the proof, we show that $\mathbf{A}$ is bounded from below by a quantity $C_0(f, g, n, \sigma^2, \rho) = C(f, \rho) \frac{n^{-1}\sigma^2}{\|\partial_t f\|_\infty^2 + n^{-1}\sigma^2 \int_{\Theta} (\partial_\theta \log(g(\theta)))^2}$ independent of J (this statement is made precise later) and that $\mathbf{B}$ goes to zero as J goes to infinity. Then, these two facts imply that there exists a $J_0 \in \mathbb{N}$ such that $J \geq J_0$ implies that $\mathbb{E}\|\tilde{f} - f\|_{L^2} \geq \frac{1}{2}C_0(f, g, n, \sigma^2, \rho)$, which will yield the desired result.

Lower bound on $\mathbf{A}$. Recall that $c_k^* = \int_0^1 f(t) e^{-i2\pi kt} dt$, then

$$\begin{aligned} \|\tilde{f} - f_{\Theta_0}\|_{L^2} &= \left\| \frac{1}{J} \sum_{j=1}^J f(\cdot - \theta_j^* + \hat{\theta}_j^\lambda) - f(\cdot + \bar{\theta}^*) \right\|_{L^2} \\ &= \left(\sum_{k \in \mathbb{Z}} \left| \frac{1}{J} \sum_{j=1}^J \left(e^{i2\pi k(-\theta_j^* + \hat{\theta}_j^\lambda)} - e^{i2\pi k \bar{\theta}^*} \right) c_k^* \right|^2 \right)^{\frac{1}{2}}, \\ &\geq |c_1^*| \left| \frac{1}{J} \sum_{j=1}^J (e^{i2\pi(\hat{\theta}_j^\lambda - [\theta_{\Theta_0}^*]_j)} - 1) \right|, \end{aligned}$$

where $\theta_{\Theta_0}^* = (\theta_1^* - \bar{\theta}^*, \dots, \theta_J^* - \bar{\theta}^*)$, the right hand side of the preceding inequality being positive since Assumption 1.5 ensures that $c_1^* \neq 0$ for all $j = 1, \dots, J$. Let $u_j = 2\pi(\hat{\theta}_j^\lambda - [\theta_{\Theta_0}^*]_j)$, $j = 1, \dots, J$. Since $\sum_{j=1}^J u_j = 0$ and $|u_j| \leq 4\pi\rho < 3$, $j = 1, \dots, J$ (by our assumption on ρ), Lemma 1.12 implies that

$$\|\tilde{f} - f_{\Theta_0}\|_{L^2} \geq C(f, \rho) \frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2. \quad (1.30)$$

Now, remark that $\mathbb{E}[\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_{\Theta_0}^*\|^2] \geq \mathbb{E}[\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*\|^2] - \mathbf{C}$ with $\mathbf{C} = 2\mathbb{E}[\|\bar{\boldsymbol{\theta}}^*\| \frac{1}{J} \sum_{j=1}^J \|\hat{\boldsymbol{\theta}}_j^\lambda - \boldsymbol{\theta}_j^*\|]$. By applying Theorem 1.3 we get that

$$\mathbb{E}[\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*\|^2] \geq C(f, g, n, \sigma^2), \text{ with } C(f, g, n, \sigma^2) = \frac{n^{-1}\sigma^2}{\|\partial_t f\|_\infty^2 + n^{-1}\sigma^2 \int_{\Theta} (\partial_{\boldsymbol{\theta}} \log(g(\boldsymbol{\theta})))^2}.$$

Then, remark that $\mathbf{C} \leq 4\rho\sqrt{\mathbb{E}[\|\bar{\boldsymbol{\theta}}^*\|^2]} \leq C(\rho, g)J^{-1/2}$. Hence $\mathbf{C}$ tends to 0 as J goes to infinity. Therefore, using equation (1.30), it follows that there exists $C_0(f, g, n, \sigma^2, \gamma, \rho) > 0$ and $J_1 \in \mathbb{N}$ such that $J \geq J_1$ implies that

$$\mathbf{A} = \mathbb{E}[\|\tilde{f}^\lambda - \tilde{f}\|_{L^2}] \geq C_0(f, g, n, \sigma^2, \rho). \quad (1.31)$$

Upper bound on $\mathbf{B}$. By assumption, f is continuously differentiable on $[0, 1]$ implying that $\|f_{\Theta_0} - f\|_{L^2} = \|f(\cdot + \bar{\boldsymbol{\theta}}^*) - f\|_{L^2} \leq \|\partial_t f\|_\infty \|\bar{\boldsymbol{\theta}}^*\|$. Therefore, $\mathbb{E}\|f_{\Theta_0} - f\|_{L^2} \leq \|\partial_t f\|_\infty \sqrt{\mathbb{E}[\|\bar{\boldsymbol{\theta}}^*\|^2]} \leq C(f, g)J^{-1/2}$. Hence, there exists a $J_2 \in \mathbb{N}$ such that $J \geq J_2$ implies

$$\mathbf{B} = \mathbb{E}[\|\tilde{f}_{\Theta_0} - \tilde{f}\|_{L^2}] \leq \frac{1}{2}C_0(f, g, n, \sigma^2, \rho). \quad (1.32)$$

To conclude the proof, equations (1.29), (1.31) and (1.32) imply that there exists a $J_0 \in \mathbb{N}$ such that $J \geq J_0$ implies $\mathbb{E}\|\hat{f}^\lambda - \tilde{f}\|_{L^2} \geq |\mathbf{A} - \mathbf{B}| \geq \frac{1}{2}C_0(f, g, n, \sigma^2, \rho)$. $\square$

1.D. Proof of the results in Section 1.7

1.D.1. Proof of Theorem 1.9

We explain here the main arguments of the proof of Theorem 1.9. Technical Lemmas are given in the second part of the Appendix. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) = (\theta_1^1, \dots, \theta_1^p, \dots, \theta_J^1, \dots, \theta_J^p) \in \mathbb{R}^{pJ}$ and decompose the criterion (1.22) as follows,

$$\begin{aligned} M_\lambda(\boldsymbol{\theta}) &= \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\boldsymbol{\theta}_j} \langle S_\lambda(t), \mathbf{Y}_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\boldsymbol{\theta}_{j'}} \langle S_\lambda(t), \mathbf{Y}_{j'} \rangle \right)^2 dt \\ &= D(\boldsymbol{\theta}) + [R_\lambda(\boldsymbol{\theta}) + Q_\lambda(\boldsymbol{\theta})] + [Q_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^{Z,\varepsilon}(\boldsymbol{\theta}) + Q_\lambda^\varepsilon(\boldsymbol{\theta}) + R_\lambda^\varepsilon(\boldsymbol{\theta})], \end{aligned}$$

where $D(\boldsymbol{\theta}) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\boldsymbol{\theta}_j} T_{\boldsymbol{\theta}_j^*} f(t) - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\boldsymbol{\theta}_{j'}} T_{\boldsymbol{\theta}_{j'}^*} f(t) \right)^2 dt$, the terms R_λ and Q_λ are due to the smoothing, namely,

$$\begin{aligned} Q_\lambda(\boldsymbol{\theta}) &= \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\boldsymbol{\theta}_j} B_\lambda(T_{\boldsymbol{\theta}_j^*} f, t) - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\boldsymbol{\theta}_{j'}} B_\lambda(T_{\boldsymbol{\theta}_{j'}^*} f, t) \right)^2 dt \\ R_\lambda(\boldsymbol{\theta}) &= \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\boldsymbol{\theta}_j} T_{\boldsymbol{\theta}_j^*} f(t) - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\boldsymbol{\theta}_{j'}} T_{\boldsymbol{\theta}_{j'}^*} f(t) \right) \\ &\quad \times \left(\tilde{T}_{\boldsymbol{\theta}_j} B_\lambda(T_{\boldsymbol{\theta}_j^*} f, t) - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\boldsymbol{\theta}_{j'}} B_\lambda(T_{\boldsymbol{\theta}_{j'}^*} f, t) \right) dt, \end{aligned}$$

and the others terms contain the Z_j 's and ε_j 's error terms. Let $\mathbf{T}_{\theta_j^*} \mathbf{Z}_j = (T_{\theta_j^*} Z_j(t_\ell))_{\ell=1}^n$ and $\mathbf{T}_{\theta_j^*} \mathbf{f} = (T_{\theta_j^*} f(t_\ell))_{\ell=1}^n$, then

$$\begin{aligned}
Q_\lambda^Z(\boldsymbol{\theta}) &= \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \mathbf{T}_{\theta_j^*} \mathbf{Z}_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \mathbf{T}_{\theta_{j'}^*} \mathbf{Z}_{j'} \rangle \right)^2 dt \\
R_\lambda^Z(\boldsymbol{\theta}) &= \frac{2}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \mathbf{T}_{\theta_j^*} \mathbf{f} \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \mathbf{T}_{\theta_{j'}^*} \mathbf{f} \rangle \right) \\
&\quad \times \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \mathbf{T}_{\theta_j^*} \mathbf{Z}_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \mathbf{T}_{\theta_{j'}^*} \mathbf{Z}_{j'} \rangle \right) dt, \\
R_\lambda^{Z,\varepsilon}(\boldsymbol{\theta}) &= \frac{2\sigma}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \mathbf{T}_{\theta_j^*} \mathbf{Z}_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \mathbf{T}_{\theta_{j'}^*} \mathbf{Z}_{j'} \rangle \right) \\
&\quad \times \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \varepsilon_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \varepsilon_{j'} \rangle \right) dt \\
Q_\lambda^\varepsilon(\boldsymbol{\theta}) &= \frac{\sigma^2}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \varepsilon_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \varepsilon_{j'} \rangle \right)^2 dt \\
R_\lambda^\varepsilon(\boldsymbol{\theta}) &= \frac{2\sigma}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \mathbf{T}_{\theta_j^*} \mathbf{f} \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \mathbf{T}_{\theta_{j'}^*} \mathbf{f} \rangle \right) \\
&\quad \times \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \varepsilon_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \varepsilon_{j'} \rangle \right) dt.
\end{aligned}$$

At this stage, recall that $\boldsymbol{\theta}_\Theta^* = \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} D(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}^\lambda = \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} M_\lambda(\boldsymbol{\theta})$. The proof follows a classical guideline in M-estimation: we show that the uniform (over Θ) convergence in probability of the criterion M_λ to D , yielding the convergence in probability of their argmins $\boldsymbol{\theta}_\Theta^*$ and $\hat{\boldsymbol{\theta}}^\lambda$ respectively. Assumption 1.7 ensures that there is a constant $C(\Theta, \mathcal{F}, f) > 0$ such that,

$$\frac{1}{J} \|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_\Theta^*\|^2 \leq C(\Theta, \Theta, \mathcal{F}, f) |D(\hat{\boldsymbol{\theta}}^\lambda) - D(\boldsymbol{\theta}_\Theta^*)| \quad (1.33)$$

Then, a classical inequality in M-estimation and the decomposition of $M_\lambda(\boldsymbol{\theta})$ given above yield

$$\begin{aligned}
|D(\hat{\boldsymbol{\theta}}^\lambda) - D(\boldsymbol{\theta}_\Theta^*)| &\leq 2 \sup_{\boldsymbol{\theta} \in \Theta} |D(\boldsymbol{\theta}) - M_\lambda(\boldsymbol{\theta})| \\
&= \underbrace{2 \sup_{\boldsymbol{\theta} \in \Theta} \{R_\lambda(\boldsymbol{\theta}) + Q_\lambda(\boldsymbol{\theta})\}}_{\mathbf{B}} + \underbrace{2 \sup_{\boldsymbol{\theta} \in \Theta} \{Q_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^{Z,\varepsilon}(\boldsymbol{\theta}) + Q_\lambda^\varepsilon(\boldsymbol{\theta}) + R_\lambda^\varepsilon(\boldsymbol{\theta})\}}_{\mathbf{V}}
\end{aligned} \quad (1.34)$$

The rest of the proof is devoted to control the $\mathbf{B}$ and $\mathbf{V}$ terms.

Control of B. Using Assumption 1.11 and 1.9, we have that $Q_\lambda(\boldsymbol{\theta}) \leq \frac{C(\Theta)}{J} \sum_{j=1}^J \|B_\lambda(T_{\theta_j^*} f, t)\|_{L^2}^2 \leq C(\Theta, \mathcal{F}) B(\lambda)$. Now by applying the Cauchy-Schwarz inequality, $|R_\lambda(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta} \in \Theta} \{\sqrt{D(\boldsymbol{\theta})}\} \sqrt{Q_\lambda(\boldsymbol{\theta})}$. By Assumption 1.9, there exists a constant such $\sup_{\boldsymbol{\theta} \in \Theta} \{D(\boldsymbol{\theta})\} \leq C(\Theta, \mathcal{F}, f)$ and thus

$$\mathbf{B} \leq C(\Theta, \mathcal{F}, f) (B(\lambda) + \sqrt{B(\lambda)}). \quad (1.35)$$

Control of $\mathbf{V}$. We give a control in probability of the stochastic quadratic term Q_λ^Z and Q_λ^ε . As previously, one can show that there is a constant $C(\Theta, \mathcal{F}, f) > 0$ such that,

$$\left| Q_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^{Z,\varepsilon}(\boldsymbol{\theta}) + Q_\lambda^\varepsilon(\boldsymbol{\theta}) + R_\lambda^\varepsilon(\boldsymbol{\theta}) \right| \leq C(\Theta, \mathcal{F}, f) \left(\sqrt{Q_\lambda^Z(\boldsymbol{\theta})} + Q_\lambda^Z(\boldsymbol{\theta}) + Q_\lambda^\varepsilon(\boldsymbol{\theta}) + \sqrt{Q_\lambda^\varepsilon(\boldsymbol{\theta})} \right),$$

where we have used the inequality $2ab \leq a^2 + b^2$, valid for any $a, b > 0$ to control the term $R_\lambda^{Z,\varepsilon}$. The quadratic terms Q_λ^Z and Q_λ^ε are controlled by Corollaries 1.14 and 1.15 respectively. It yields immediately to

$$\mathbb{P} \left(\mathbf{V} \geq C(\Theta, \mathcal{F}, f)(\gamma_{\max}(n) + \sigma^2)(v(x, J, \lambda) + \sqrt{v(x, J, \lambda)}) \right) \leq 2e^{-x}, \quad (1.36)$$

where $v(x, J, \lambda) = V(\lambda) \left(1 + 4\frac{x}{J} + \sqrt{4\frac{x}{J}} \right)$.

Putting together equations (1.33), (1.34), (1.35) and (1.36), we have

$$\begin{aligned} \mathbb{P} \left(\frac{1}{J} \|\boldsymbol{\theta}_\Theta^* - \hat{\boldsymbol{\theta}}^\lambda\|^2 \geq C(\Theta, \boldsymbol{\Theta}, \mathcal{F}, f)(\gamma_{\max}(n) + \sigma^2) \right. \\ \left. \times \left(\left(\sqrt{v(x, J, \lambda)} + v(x, J, \lambda) \right) + \left(B(\lambda) + \sqrt{B(\lambda)} \right) \right) \right) \leq 2e^{-x}, \end{aligned}$$

which completes the proof of Theorem 1.9. $\square$

1.D.2. Proof of Theorem 1.10

In this part, we use the notations introduced in the proof of Theorem 1.9. We have,

$$\begin{aligned} \|f_\Theta^* - \hat{f}^\lambda\|_{L^2}^2 &\leq \underbrace{\frac{2}{J} \sum_{j=1}^J \left\| \tilde{T}_{[\boldsymbol{\theta}_\Theta^*]_j} T_{\boldsymbol{\theta}_j^*} f - \tilde{T}_{[\boldsymbol{\theta}_\Theta^*]_j} \langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f} \rangle \right\|_{L^2}^2}_{\mathbf{B}'} \\ &\quad + \underbrace{\frac{2}{J} \sum_{j=1}^J \left\| \tilde{T}_{[\boldsymbol{\theta}_\Theta^*]_j} \langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f} \rangle - \tilde{T}_{\hat{\boldsymbol{\theta}}_j^\lambda} \langle S_\lambda(\cdot), \mathbf{Y}_j \rangle \right\|_{L^2}^2}_{\mathbf{V}'}. \end{aligned}$$

Again, the first term above depends on the bias, and the second term (stochastic) can be controlled in probability. Under Assumptions 1.9 and 1.11 we have that

$$\mathbf{B}' \leq \frac{C(\Theta)}{J} \sum_{j=1}^J \left\| \langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f} \rangle - T_{\boldsymbol{\theta}_j^*} f \right\|_{L^2}^2 \leq C(\Theta, \mathcal{F}) B(\lambda),$$

and

$$\begin{aligned} \mathbf{V}' &= \frac{2}{J} \sum_{j=1}^J \left\| \tilde{T}_{[\boldsymbol{\theta}_\Theta^*]_j} \langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f} \rangle - \tilde{T}_{\hat{\boldsymbol{\theta}}_j^\lambda} \langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f} \rangle + \tilde{T}_{\hat{\boldsymbol{\theta}}_j^\lambda} \langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f} \rangle - \tilde{T}_{\hat{\boldsymbol{\theta}}_j^\lambda} \langle S_\lambda(\cdot), \mathbf{Y}_j \rangle \right\|_{L^2}^2 \\ &\leq \frac{C(\Theta, \mathcal{F})}{J} \sum_{j=1}^J \left(\left\| \hat{\boldsymbol{\theta}}_j^\lambda - [\boldsymbol{\theta}_\Theta^*]_j \right\|^2 + \left\| \langle S_\lambda(\cdot), \mathbf{Y}_j - \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f} \rangle \right\|_{L^2}^2 \right), \\ &\leq C(\Theta, \mathcal{F}) \left(\frac{1}{J} \left\| \hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_\Theta^* \right\|^2 + \frac{1}{J} \sum_{j=1}^J \left\| \langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j + \boldsymbol{\varepsilon}_j \rangle \right\|_{L^2}^2 \right) \end{aligned}$$

The stochastic term $\frac{1}{J} \sum_{j=1}^J \left\| \langle S_\lambda(\cdot), \mathbf{T}_{\theta_j^*} \mathbf{Z}_j + \boldsymbol{\varepsilon}_j \rangle \right\|_{L^2}^2$ in the above inequality can be controlled using Lemma 1.13 and the arguments in the proof of Corollaries 1.14 and 1.15 to obtain that for any $x > 0$

$$\mathbb{P} \left(\frac{1}{J} \sum_{j=1}^J \left\| \langle S_\lambda(\cdot), \mathbf{T}_{\theta_j^*} \mathbf{Z}_j + \boldsymbol{\varepsilon}_j \rangle \right\|_{L^2}^2 \geq C(\Theta, \mathcal{F}, f) (\gamma_{\max}(n) + \sigma^2) \left(\sqrt{v(x, J, \lambda)} + v(x, J, \lambda) \right) \right) \leq e^{-x}.$$

Then, from Theorem 1.9 it follows that

$$\mathbb{P} \left(\mathbf{B}' + \mathbf{V}' \geq C(\Theta, \Theta, \mathcal{F}, f) \left[(\gamma_{\max}(n) + \sigma^2) \left(\sqrt{v(x, J, \lambda)} + v(x, J, \lambda) \right) + (B(\lambda) + \sqrt{B(\lambda)}) \right] \right) \leq 2e^{-x},$$

which completes the proof. $\square$

1.E. Technical Lemmas

Lemma 1.12. *Let $u = (u_1, \dots, u_J)$ such that $\sum_{j=1}^J u_j = 0$ with $|u_j| \leq \delta$ for some $0 \leq \delta < 3$ for all $j = 1, \dots, J$. Then, there exists a constant $C(\delta) > 0$ such that*

$$\left| \frac{1}{J} \sum_{j=1}^J (e^{iu_j} - 1) \right| \geq \frac{C(\delta)}{J} \|u\|^2$$

where $\|u\|^2 = u_1^2 + \dots + u_J^2$.

Proof. Let $F(u_1, \dots, u_J) = \frac{1}{J} \sum_{j=1}^J e^{iu_j}$. A Taylor expansion implies that there exists $t_j \in [-\delta, \delta]$, $j = 1, \dots, J$ such that

$$F(u_1, \dots, u_J) = 1 + \frac{i}{J} \sum_{j=1}^J u_j - \frac{1}{2J} \sum_{j=1}^J u_j^2 - \frac{i}{6J} \sum_{j=1}^J u_j^3 e^{it_j},$$

holds for all $|u_j| \leq \delta$. Now, since $\sum_{j=1}^J u_j = 0$ it follows that

$$\left| \frac{1}{J} \sum_{j=1}^J e^{iu_j} - 1 \right| = \left| -\frac{1}{2J} \sum_{j=1}^J u_j^2 - \frac{i}{6J} \sum_{j=1}^J u_j^3 e^{it_j} \right| \geq \frac{1}{2J} \left| \sum_{j=1}^J u_j^2 - \left| \frac{i}{3} \sum_{j=1}^J u_j^3 e^{it_j} \right| \right|.$$

Since $|u_j| \leq \delta$, we have that $\left| \frac{i}{3} \sum_{j=1}^J u_j^3 e^{it_j} \right| \leq \frac{\delta}{3} \sum_{j=1}^J |u_j|^2$ which finally implies that $\left| \frac{1}{J} \sum_{j=1}^J e^{iu_j} - 1 \right| \geq \frac{3-\delta}{6} \frac{1}{J} \sum_{j=1}^J u_j^2$, which proves the result by letting $C(\delta) = \frac{3-\delta}{6} > 0$ since $\delta < 3$. $\square$

Lemma 1.13. *Let $\xi_{\lambda, J}(A_1, \dots, A_J) = \frac{1}{J} \sum_{j=1}^J \|\langle S_\lambda(\cdot), A_j \boldsymbol{\varepsilon}_j \rangle\|_{L^2}^2$, where $\boldsymbol{\varepsilon}_j \sim \mathcal{N}(0, I_n)$ and the A_j 's are nonrandom non-negative $n \times n$ symmetric matrices. Then, for all $x > 0$ and all $n \geq 1$,*

$$\mathbb{P} \left(\xi_{\lambda, J}(A_1, \dots, A_J) \geq \frac{1}{J} \|\mathbf{A}\| \left(1 + 4\frac{x}{J} + \sqrt{4\frac{x}{J}} \right) \right) \leq e^{-x}.$$

where $\|\mathbf{A}\| = \sum_{j=1}^J \sum_{\ell=1}^n r_{j,\ell}$ with $r_{j,\ell}$ being the ℓ -th eigenvalue of the matrix $\mathbf{A}_j = A_j \left[\langle S_\lambda^\ell, S_\lambda^{\ell'} \rangle_{L^2} \right]_{\ell, \ell'=1}^n A_j$.

Proof. Some parts of the proof follows the arguments in [BM98] (Lemma 8, part 7.6). We have

$$\xi_{\lambda,J} = \frac{1}{J} \sum_{j=1}^J \left\| \sum_{\ell=1}^n S_{\lambda}^{\ell}(\cdot) [A \varepsilon_j]^{\ell} \right\|_{L_2}^2 = \frac{1}{J} \sum_{j=1}^J \sum_{\ell, \ell'=1}^n \langle S_{\lambda}^{\ell}, S_{\lambda}^{\ell'} \rangle_{L_2} [A_j \varepsilon_j]^{\ell} [A_j \varepsilon_j]^{\ell'} = \frac{1}{J} \sum_{j=1}^J \varepsilon_j' \mathbf{A}_j \varepsilon_j,$$

where $\mathbf{A}_j = A_j \mathbf{S}_{\lambda} A_j$ with $\mathbf{S}_{\lambda} = [\langle S_{\lambda}^{\ell}, S_{\lambda}^{\ell'} \rangle_{L_2}]_{\ell, \ell'=1}^n$. Now, denote by $r_{j,1} \geq \dots \geq r_{j,n}$ the eigenvalues of $\mathbf{A}_j$ with $r_{j,1} \geq \dots \geq r_{j,n} \geq 0$ and $r_1 = \max_{j,\ell} \{r_{j,\ell}\}$. We can write $\mathbf{A}_j = (\mathbf{S}_{\lambda}^{\frac{1}{2}} A_j)' (\mathbf{S}_{\lambda}^{\frac{1}{2}} A_j)$ and is positive semi-definite. Then, let $\tilde{\xi}_{\lambda,J} = J \xi_{\lambda,J} - J \mathbb{E} \xi_{\lambda,J} = \sum_{j=1}^J (\varepsilon_j' \mathbf{A}_j \varepsilon_j - \text{tr } \mathbf{A}_j)$. Let $\alpha > 0$, by Markov's inequality it follows that for all $u \in (0, \frac{1}{2r_1})$, $\mathbb{P}(\tilde{\xi}_{\lambda,J} \geq \alpha) = \mathbb{P}(e^{u \tilde{\xi}_{\lambda,J}} \geq e^{u \alpha}) \leq e^{-u \alpha} \prod_{j=1}^J \mathbb{E} [e^{u \varepsilon_j' \mathbf{A}_j \varepsilon_j - u \text{tr } \mathbf{A}_j}]$, since the ε_j 's are independent. The log-Laplace transform of $\tilde{\varphi}_{\lambda,j} = \varepsilon_j' \mathbf{A}_j \varepsilon_j - \text{tr } \mathbf{A}_j$ is $\log(\mathbb{E}[e^{u \tilde{\varphi}_{\lambda,j}}]) = \sum_{\ell=1}^n -u r_{j,\ell} - \frac{1}{2} \log(1 - 2u r_{j,\ell})$. We now use the inequality $-x - \frac{1}{2} \log(1 - 2x) \leq \frac{x^2}{1-2x}$ for all $0 < x < \frac{1}{2}$ which holds since $u \in (0, \frac{1}{2r_1})$. This implies that $\log(\mathbb{E}[e^{u \tilde{\varphi}_{\lambda,j}}]) \leq \sum_{\ell=1}^n \frac{u^2 r_{j,\ell}^2}{1-2u r_{j,\ell}} \leq \frac{u^2 \|r_j\|^2}{1-2u r_1}$, where $\|r_j\|^2 = r_{j,1}^2 + \dots + r_{j,n}^2$. Finally, we have

$$\mathbb{P}(\tilde{\varphi}_{\lambda,J} \geq \alpha) \leq \exp\left(-\left(u\alpha - \sum_{j=1}^J \frac{\|r_j\|^2 u^2}{1-2u r_1}\right)\right) = \exp\left(-\left(u\alpha - \frac{\|r\|^2 u^2}{1-2u r_1}\right)\right), \quad (1.37)$$

where $\|r\|^2 = \sum_{j=1}^J \sum_{\ell=1}^n r_{j,\ell}^2$. The right hand side of the above inequality achieves its minimum at $u = \frac{1}{2r_1} \left(1 - \frac{\|r\|}{\sqrt{2\alpha r_1 + \|r\|^2}}\right)$. Evaluating (1.37) at this point and using the inequality $(1+x)^{1/2} \leq 1 + \frac{x}{2}$, valid for all $x \geq -1$, one has that

$$\begin{aligned} \mathbb{P}(\tilde{\xi}_{\lambda,J} \geq \alpha) &\leq \exp\left(-\frac{\alpha^2}{2r_1 \alpha + 2\|r\|^2 + 2\|r\|^2(1+4\alpha r_1/(2\|r\|^2))^{1/2}}\right) \\ &\leq \exp\left(-\frac{\alpha^2}{4r_1 \alpha + 4\|r\|^2}\right), \end{aligned}$$

by setting $x = \frac{\alpha^2}{4r_1 \alpha + 4\|r\|^2}$. We derive the following concentration inequality for $\xi_{\lambda,J} = \frac{1}{J} \tilde{\xi}_{\lambda,J} + \frac{1}{J} \sum_{j=1}^J \text{tr}(\mathbf{A}_j)$,

$$\mathbb{P}\left(\xi_{\lambda,J} \geq \frac{1}{J} \sum_{j=1}^J \sum_{\ell=1}^n r_{j,\ell} + 4 \frac{r_1}{J} x + \frac{\|r\|}{J} \sqrt{4x}\right) \leq e^{-x}.$$

To finish the proof, remark that $\|r\|^2 = \sum_{j=1}^J \sum_{\ell=1}^n r_{j,\ell}^2 \leq \left(\sum_{j=1}^J \sum_{\ell=1}^n r_{j,\ell}\right)^2$ since all the $r_{j,\ell}$'s are positive. $\square$

Corollary 1.14. *Under Assumptions 1.9 to 1.11, there exists a constant $C(\Theta, \mathcal{F}) > 0$ such that for all $x > 0$,*

$$\mathbb{P}\left(\sup_{\theta \in \Theta} Q_{\lambda}^{\varepsilon}(\theta) \geq C(\Theta, \mathcal{F}) \sigma^2 V(\lambda) \left(1 + 4 \frac{x}{J} + \sqrt{4 \frac{x}{J}}\right)\right) \leq e^{-x}.$$

Proof. Assumption 1.9 gives the uniform bound

$$\begin{aligned} Q_{\lambda}^{\varepsilon}(\theta) &\leq \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \langle S_{\lambda}(t), \sigma \varepsilon_j \rangle\right)^2 dt \\ &\leq \frac{C(\Theta, \mathcal{F})}{J} \sum_{j=1}^J \|\langle S_{\lambda}(t), \sigma \varepsilon_j \rangle\|_{L_2}^2 = C(\Theta, \mathcal{F}) \xi_{\lambda,J}(\sigma Id_n, \dots, \sigma Id_n), \end{aligned}$$

where $\xi_{\lambda,J}(\sigma Id_n, \dots, \sigma Id_n)$ is defined in Lemma 1.13 and does not depend on $\boldsymbol{\theta}$. Thus, the result immediately follows from Lemma 1.13. $\square$

Corollary 1.15. *Under Assumptions 1.9 to 1.12, there exists a constant $C(\Theta, \mathcal{F}) > 0$ such that for all $x \geq 0$,*

$$\mathbb{P} \left(\sup_{\boldsymbol{\theta} \in \Theta} Q_{\lambda}^Z(\boldsymbol{\theta}) \geq C(\Theta, \mathcal{F}) \gamma_n(\Theta) V(\lambda) \left(1 + 4 \frac{x}{J} + \sqrt{4 \frac{x}{J}} \right) \right) \leq e^{-x}.$$

Proof. Assumption 1.9 gives the uniform bound

$$Q_{\lambda}^Z(\boldsymbol{\theta}) \leq \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\boldsymbol{\theta}_j} \left\langle S_{\lambda}(t), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j \right\rangle \right)^2 dt \leq \frac{C(\Theta, \mathcal{F})}{J} \sum_{j=1}^J \left\| \left\langle S_{\lambda}, \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j \right\rangle \right\|_{L^2}^2.$$

Hence, conditionally on $\boldsymbol{\theta}^*$ we have that $\sup_{\boldsymbol{\theta} \in \Theta^J} Q_{\lambda}^Z(\boldsymbol{\theta}) \leq C(\Theta, \mathcal{F}) \xi_{\lambda,J}(A_1, \dots, A_J)$, where the term $\xi_{\lambda,J}(A_1, \dots, A_J)$ is defined in Lemma 1.13 with $A_j = \mathbb{E}_{\boldsymbol{\theta}^*} [\mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j (\mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j)']^{\frac{1}{2}}$. Let us now give an upper bound on the largest eigenvalues of the matrices $\mathbf{A}_j = A_j \mathbf{S}_{\lambda} A_j$ with $\mathbf{S}_{\lambda} = [\langle S_{\lambda}^{\ell}, S_{\lambda}^{\ell'} \rangle_{L^2}]_{\ell, \ell'=1}^n$. Under Assumption 1.12 we have that $\text{tr}(\mathbf{A}_j) \leq \gamma_{\max}(A_j) \text{tr} \mathbf{S}_{\lambda} \leq \gamma_n(\Theta) V(\lambda)$, for all $j = 1, \dots, J$ and any $\boldsymbol{\theta}^* \in \Theta^J$. Thus, the result follows by arguing as in the proof of Lemma 1.13 and by taking expectation with respect to $\boldsymbol{\theta}^*$. $\square$

Estimation consistante de la courbe moyenne dans des modèles de déformations pour l'analyse de forme en grande dimension

Ce chapitre se situe dans la continuité du chapitre 1. Nous y appliquons à l'exemple particulier des formes du plan le modèle général de déformations présenté au chapitre précédent. Ce chapitre est rédigé en anglais et est autonome du reste de ce mémoire du point de vue des définitions et des notations.

Résumé

Le groupe des similarités agit par rotations, translations et homothéties sur les courbes planaires. On appelle configuration un ensemble de k points du plan correspondant à l'échantillonnage d'une courbe. Le modèle de déformations associé à ce cadre est défini en (2.1). Une méthode d'estimation des paramètres de déformations et de la courbe moyenne est présentée. Nous étudions ensuite le rôle de la dimension k des configurations et du nombre J de courbes observées dans l'estimation de la courbe moyenne ayant généré les données. Des simulations numériques illustrent ces résultats.

Abstract

The similarity group of the plane acts on the planar curves by translations, rotations and scaling. In this chapter, we mean by (planar) configuration a k -ads of the plane obtained by sampling a planar curve. We study the deformable model (2.1) associated to this framework. We then present an estimation procedure for the deformation parameters and the mean pattern. The main contribution of the work is to study the influence of the

dimension k of the configurations and of the number J of observations on the performance on the estimation of the parameters of interest. Some numerical simulations illustrate these results.

Mots clés : Mean pattern estimation; Procrustes analysis; Shape analysis; Deformable models; High-dimensional data.

2.1. Introduction

2.1.1. A deformable model for statistical shape analysis

Statistical analysis of planar shapes is the study of random (planar) configurations

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} \in \mathbb{R}^{k \times 2}$$

described by a set of k landmarks $Y_\ell = (Y_\ell^{(1)}, Y_\ell^{(2)}) \in \mathbb{R}^2$, $\ell = 1, \dots, k$. Since the seminal work of Kendall [Ken84], one considers that the shape of $\mathbf{Y}$ is “what remains when translations, rotations and scaling are filtered out”. More precisely, two configurations $\mathbf{Y}_1, \mathbf{Y}_2 \in \mathbb{R}^{k \times 2}$ are said to have the same shape if there exists a vector $(a, \alpha, b) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ such that

$$\mathbf{Y}_2 = e^a \mathbf{Y}_1 R_\alpha + \mathbf{1}_k \otimes b, \quad \text{with } R_\alpha = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}, \quad (2.1)$$

where $\mathbf{1}_k$ denotes the vector of $\mathbb{R}^k$ with all entries equal to one, and $\otimes$ denotes the tensor product. Note that (2.1) defines an equivalence relation which is reflexive, transitive and symmetric. In shape analysis, an important issue is the computation of a sample mean shape from a set of J random planar configurations $\mathbf{Y}_1, \dots, \mathbf{Y}_J$, and the study of its consistency as the number of samples J goes to infinity. A statistical model for shapes must include some nuisance parameters associated to the ambiguity of location, rotation and scaling. In [Goo91], consistent estimation of a mean shape is therefore considered in the following deformable model:

$$\mathbf{Y}_j = e^{\alpha_j^*} (\mathbf{f} + \boldsymbol{\zeta}_j) R_{\alpha_j^*} + \mathbf{1}_k \otimes b_j^*, \quad \text{with } R_{\alpha_j^*} = \begin{pmatrix} \cos(\alpha_j^*) & -\sin(\alpha_j^*) \\ \sin(\alpha_j^*) & \cos(\alpha_j^*) \end{pmatrix}, \text{ and } j = 1, \dots, J, \quad (2.2)$$

where the mean pattern $\mathbf{f} \in \mathbb{R}^{k \times 2}$ is an unknown configuration of k landmarks which is also called a population mean in [Goo91] or a perturbation mean in [Huc10]. The error terms $\boldsymbol{\zeta}_j \in \mathbb{R}^{k \times 2}$, $j = 1, \dots, J$ are independent copies of a random perturbation $\boldsymbol{\zeta}$ in $\mathbb{R}^{k \times 2}$ with zero expectation. For all $j = 1, \dots, J$, the deformation parameters $(a_j^*, \alpha_j^*, b_j^*) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ are independent and identically distributed (i.i.d) random scaling, rotation and translation parameters. According to [Goo91] a sample mean pattern $\hat{\mathbf{f}}$ computed from $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ is said to be consistent if, as $J \rightarrow \infty$, it has asymptotically the same shape than the mean pattern $\mathbf{f}$. Since Goodall’s proposal, the deformable model (2.2) has been highly popular in the statistical shape community, and the study of consistent procedures to estimate the shape of the mean pattern $\mathbf{f}$ using this model has been considered by

various authors [KM97, Le98, Lel93]. In this setting, sample mean patterns obtained by a Procrustes procedure have received a special attention. In particular, it is shown in [KM97, Le98] that, in the very specific case of isotropic perturbations ζ , the so-called full and partial Procrustes sample means are consistent estimators of the shape of $\mathbf{f}$. Nevertheless, these estimators can be inconsistent for non-isotropic perturbations. Therefore, it is generally the belief that consistent statistical inference based on Procrustes analysis is restricted to very limited assumptions on the distribution of the data, see also [DM98, KBCL99] for further discussions.

The aim of this paper is to show that a Procrustes sample mean can be considered as a consistent procedure even in the case of non-isotropic perturbations. To this purpose, we propose to exhibit the relation that exists between the dimensionality k of the data and the consistency of Procrustes sample means in the perturbation model (2.2). Our main result (see Theorem 2.1 below) is that when the dimensionality k is high and the mean pattern $\mathbf{f}$ is the discretization of a sufficiently smooth plane curve, it is possible to build a consistent estimator of the shape of $\mathbf{f}$ in model (2.2) under general assumptions on the perturbation ζ_j . The problem of analyzing high-dimensional 2D configurations (*i.e.* when the number of landmarks k is high) arises in the statistical study of a set of random points in $\mathbb{R}^2$ that have been sampled from planar curves. A typical example is the analysis of contours extracted from digital 2D images.

The main idea of this paper is thus to consider that the unknown mean pattern $\mathbf{f} \in \mathbb{R}^{k \times 2}$ has been obtained by sampling a planar curve $f : [0, 1] \rightarrow \mathbb{R}^2$ on an equi-spaced design, meaning that

$$\mathbf{f} = (f(\frac{\ell}{k}))_{\ell=1}^k.$$

Under appropriate smoothness assumptions on f , we use a two steps procedure to estimate $\mathbf{f}$. First, we perform a dimension reduction step by projecting the data into a low-dimensional space of $\mathbb{R}^{k \times 2}$ to eliminate the influence of the random perturbations ζ_j . Then, in a second step, we apply Procrustes analysis in this low-dimensional space to obtain a consistent estimator of $\mathbf{f}$.

2.1.2. Main contributions

To give a precise definition of our estimating procedure, introduce the following $k \times k$ matrix

$$\mathbf{A}^\lambda = \left(\frac{1}{k} \sum_{0 \leq |m| \leq \lambda} e^{i2\pi m \frac{\ell - \ell'}{k}} \right)_{\ell, \ell'=1}^k, \quad (2.3)$$

The matrix $\mathbf{A}^\lambda$ is nothing else but the smoothing matrix corresponding to a discrete Fourier low pass filter with frequency cutoff $\lambda \in \mathbb{N}$. It can be viewed as a projection matrix in a (trigonometric) sub-space $\mathcal{V}^\lambda$ of $\mathbb{R}^k$ of dimension $2\lambda + 1$. Then, we project the data on $\mathcal{V}^\lambda \times \mathcal{V}^\lambda \subset \mathbb{R}^{k \times 2}$, and we estimate the scaling, rotation and translation parameters in model (2.2) using M-estimation. In the rest of this chapter, we use a bold notation to denote a vector in $\mathbb{R}^J$ or $\mathbb{R}^{2J}$. Thus, we denote the scaling parameters by $\mathbf{a} = (a_1, \dots, a_J) \in \mathbb{R}^J$, the rotation parameters by $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J) \in \mathbb{R}^J$ and the translation parameters by $\mathbf{b} = (b_1, \dots, b_J) \in \mathbb{R}^{2J}$. Then, let us introduce the following matching functional,

$$M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{-a_j} \mathbf{A}^\lambda(\mathbf{Y}_j - \mathbf{1}_k \otimes b_j) R_{-\alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{-a_{j'}} \mathbf{A}^\lambda(\mathbf{Y}_{j'} - \mathbf{1}_k \otimes b_{j'}) R_{-\alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2, \quad (2.4)$$

with $\|\cdot\|_{\mathbb{R}^{k \times 2}}$ is the standard Euclidean norm in $\mathbb{R}^{k \times 2}$ defined by $\|\mathbf{Y}\|_{\mathbb{R}^{k \times 2}}^2 = \text{tr } \mathbf{Y}'\mathbf{Y}$ where tr denotes the trace operator and $'$ is the transpose of the matrix $\mathbf{Y} \in \mathbb{R}^{k \times 2}$. An estimator of the deformation parameters is finally given by

$$(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda) \in \underset{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \Theta_0}{\text{argmin}} M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}), \quad (2.5)$$

where $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda) = (\hat{a}_1^\lambda, \dots, \hat{a}_J^\lambda, \hat{\alpha}_1^\lambda, \dots, \hat{\alpha}_J^\lambda, \hat{b}_1^\lambda, \dots, \hat{b}_J^\lambda) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$ and the constraint set on which the minimization is carried out is

$$\Theta_0 = \left\{ (\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^J \times \mathbb{R}^{2J} : \sum_{j=1}^J a_j = 0, \sum_{j=1}^J \alpha_j = 0 \text{ and } \sum_{j=1}^J b_j = 0 \right\}, \quad (2.6)$$

where $A, \mathcal{A} > 0$ are parameters whose values will be discussed below. Finally, the mean shape $\mathbf{f}$ is estimated by the following smoothed Procrustes mean

$$\hat{\mathbf{f}}^\lambda = \frac{1}{J} \sum_{j=1}^J e^{-\hat{a}_j^\lambda} \left(\mathbf{A}^\lambda \mathbf{Y}_j - \mathbb{1}_k \otimes \hat{b}_j^\lambda \right) R_{-\hat{\alpha}_j^\lambda}. \quad (2.7)$$

To answer the question of the convergence of the estimator $\hat{\mathbf{f}}^\lambda$ to the mean pattern $\mathbf{f}$, let us precise the regularity conditions that we need on the planar curve f and on the covariance structure of the random variable ζ in $\mathbb{R}^{k \times 2}$.

Assumption 2.1. The function f belongs to a Sobolev ball $H_s(L)$ of radius $L > 0$, where $s \geq \frac{3}{2}$ is a smoothness parameter - see a precise definition in (2.21). Moreover, the $k \times 2$ matrix $\mathbf{f} = (f(\frac{\ell}{k}))_{\ell=1}^k$ is of rank two, *i.e.* there is at least two different landmarks in the k -ads composing $\mathbf{f}$.

Let us now introduce the vectorized version $\tilde{\zeta} = (\zeta_1^{(1)}, \dots, \zeta_k^{(1)}, \zeta_1^{(2)}, \dots, \zeta_k^{(2)}) \in \mathbb{R}^{2k}$ of $\zeta = (\zeta_\ell^{(1)}, \zeta_\ell^{(2)})_{\ell=1}^k \in \mathbb{R}^{k \times 2}$. Thence, the covariance structure of $\tilde{\zeta}$ can now be described by a $2k \times 2k$ covariance matrix Σ .

Assumption 2.2. The error term $\tilde{\zeta}$ is a centered Gaussian vector in $\mathbb{R}^{2k}$ of covariance matrix $\Sigma \in \mathbb{R}^{2k \times 2k}$. Let $\gamma_{\max}(k)$ be the largest eigenvalue of Σ . Then,

$$\lim_{k \rightarrow \infty} \gamma_{\max}(k) k^{-\frac{2s}{2s+1}} = 0,$$

where the parameter s is defined in Assumption 2.1.

We emphasis here that no isotropy nor invariance conditions are required on the covariance structure of ζ . As above, let $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) = (a_1^*, \dots, a_J^*, \alpha_1^*, \dots, \alpha_J^*, b_1^*, \dots, b_J^*)$. We have the following results,

Theorem 2.1. Consider model (2.2) and suppose that the random variables $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ are bounded and belong to $[-\frac{A}{2}, \frac{A}{2}]^J \times [-\frac{\mathcal{A}}{2}, \frac{\mathcal{A}}{2}]^J \times [-B, B]^{2J}$ for some $0 < A, B$ and $0 < \mathcal{A} < \frac{\pi}{4}$. If Assumptions 2.1 and 2.2 hold, and if $\lambda(k) = k^{\frac{1}{2s+1}}$, then for any $J \geq 2$ there exists $(a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ such that

$$\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}_{\Theta_0}\|_{\mathbb{R}^{k \times 2}}^2 \xrightarrow[k \rightarrow \infty]{} 0, \quad \text{in probability,} \quad (2.8)$$

where $\mathbf{f}_{\Theta_0} = e^{a_0} \mathbf{f} R_{\alpha_0} + \mathbf{1}_k \otimes b_0$. Suppose, in addition, that the random variables $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ have zero expectation in $[-\frac{A}{2}, \frac{A}{2}]^J \times [-\frac{A}{2}, \frac{A}{2}]^J \times [-B, B]^{2J}$ with $A, A < 0.1$. Then we have,

$$\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \xrightarrow[k, J \rightarrow \infty]{} 0, \quad \text{in probability.} \quad (2.9)$$

The statement (2.8) means that, under mild assumptions on the covariance structure of the error terms ζ_j 's, it is possible to consistently estimate the shape of the mean pattern $\mathbf{f}$ when the number of observations J is fixed and the number k of landmarks increases. Note also that (a_0, α_0, b_0) depends on J and is given by formula (2.15) in Section 2.3. In statement (2.9) we assume that the random deformations are concentrated around zero and have zero expectation. In a sense, it means that the deformations are centered around the identity. Then, in this case, one can consistently estimate the true mean pattern $\mathbf{f}$ when the sample size J and the number of landmarks k go to infinity.

2.1.3. Organization

The rest of the paper is organized as follows. In Section 2.2, we describe the deformations acting on the mean pattern $\mathbf{f}$ in term of a group action and we recall some properties on the similarity group. Then, we discuss the General Procrustes Analysis and we compare it to our approach. In Section 2.3 we discuss the identifiability conditions related to estimating procedure described in Section 2.4. The consistency results are given in Section 2.5. We give some numerical experiments in Section 2.6.

2.2. Group structure and Generalized Procrustes Analysis

2.2.1. The similarity group

The similarity group of the plane is the group $(\mathcal{G}, \cdot)$ generated by isotropic scaling, rotations and translations. More precisely, let us denote by $\mathbf{SO}(2)$ the special orthogonal group of $\mathbb{R}^2$ endowed with the usual composition of operators denoted by $\circ$. An element R_α of $\mathbf{SO}(2)$ can be represented by a 2×2 matrix of rotation parametrized by an angle $\alpha \in [-\pi, \pi[$ as in formula (2.1). Thence, the similarity group $\mathcal{G}$ can be defined as

$$(\mathcal{G}, \cdot) = ((\mathbb{R}, +) \times (\mathbf{SO}(2), \circ)) \ltimes_\varphi (\mathbb{R}^2, +),$$

where $\ltimes_\varphi$ is a semi-direct product associated to the group morphism

$$\begin{aligned} \varphi : (\mathbb{R}, +) \times (\mathbf{SO}(2), \circ) &\longrightarrow (Aut(\mathbb{R}^2, +), \circ) \\ (a, R_\alpha) &\longmapsto \{(b^{(1)}, b^{(2)}) \mapsto e^a(b^{(1)}, b^{(2)})R_\alpha\}, \end{aligned}$$

where $Aut(\mathbb{R}^2, +)$ is the set of linear applications from $\mathbb{R}^2$ to $\mathbb{R}^2$ endowed with the usual composition also denoted by $\circ$. Definition of the semi-direct product and discussions on groups can be found in [DF04]. The identity element in $\mathcal{G}$ is denoted e and the inverse of $g \in \mathcal{G}$ is denoted by g^{-1} . Note that we have $g.g^{-1} = g^{-1}.g = e$ for all $g \in \mathcal{G}$.

In what follows, we parametrize the group $\mathcal{G}$ by a scaling parameter $a \in \mathbb{R}$, an angle $\alpha \in [-\pi, \pi[$ and a translation $b \in \mathbb{R}^2$. We use the identification

$$\mathcal{G} \simeq \mathbb{R} \times [-\pi, \pi[\times \mathbb{R}^2$$

and in what follows, we make no difference between $g \in \mathcal{G}$ and its parametrization $(a, \alpha, b) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$. The composition law of $\mathcal{G}$ can be explicitly written thanks to this parametrization. For all $g_1 = (a_1, \alpha_1, b_1), g_2 = (a_2, \alpha_2, b_2) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ we have

$$\begin{aligned} g_1 \cdot g_2 &= (a_1, \alpha_1, b_1) \cdot (a_2, \alpha_2, b_2) = (a_1 + a_2, \alpha_1 + \alpha_2, e^{a_1} b_2 R_{\alpha_1} + b_1), \\ g_1^{-1} &= (a_1, \alpha_1, b_1)^{-1} = (-a_1, -\alpha_1, -e^{-a_1} b_1 R_{-\alpha_1}), \\ e &= (0, 0, 0). \end{aligned} \tag{2.10}$$

The similarity group $\mathcal{G}$ is a non-commutative group as we have $(a_1, \alpha_1, b_1) \cdot (a_2, \alpha_2, b_2) \neq (a_2, \alpha_2, b_2) \cdot (a_1, \alpha_1, b_1)$. The non-commutativity is due to the semi-direct product structure and affects only the translation term. Note also that $\mathcal{G}$ is not a compact group. In the following, the standard Euclidean norm in $\mathbb{R}^i$ is denoted by $\|\cdot\|_{\mathbb{R}^i}$ where $i \in \mathbb{N}$.

2.2.2. Group action

The action of $\mathcal{G}$ onto $\mathbb{R}^{k \times 2}$ is given by a mapping

$$\begin{aligned} \mathcal{G} \times \mathbb{R}^{k \times 2} &\longrightarrow \mathbb{R}^{k \times 2} \\ (g, \mathbf{x}) &\longmapsto g \cdot \mathbf{x} \end{aligned}$$

which satisfies for all $g_1, g_2 \in \mathcal{G}$ and $\mathbf{x} \in \mathbb{R}^{k \times 2}$ the following two properties: $g_1 \cdot (g_2 \cdot \mathbf{x}) = (g_1 \cdot g_2) \cdot \mathbf{x}$ and $e \cdot \mathbf{x} = \mathbf{x}$. Note that we use the same symbol “.” for the composition law of $\mathcal{G}$ and its action on $\mathbb{R}^{k \times 2}$. For $g = (a, \alpha, b) \in \mathcal{G}$ and $\mathbf{x} \in \mathbb{R}^{k \times 2}$, we define the action $g \cdot \mathbf{x}$ of $\mathcal{G}$ onto $\mathbb{R}^{k \times 2}$ as in (2.1) by

$$g \cdot \mathbf{x} := e^a \mathbf{x} R_\alpha + \mathbf{1}_k \otimes b.$$

Then, formula (2.10) immediately implies that $(g_1 \cdot g_2) \cdot \mathbf{x} = e^{a_1 + a_2} \mathbf{x} R_{\alpha_1 + \alpha_2} + \mathbf{1}_k \otimes (e^{a_1} b_2 R_{\alpha_1} + b_1)$. Let us introduce some definition that will be useful in the sequel. Let

$$\mathbf{1}_{k \times 2} = \mathbf{1}_k \otimes \mathbb{R}^2$$

be the two dimensional linear subspace of $\mathbb{R}^{k \times 2}$ consisting of degenerated configurations, *i.e.* configurations composed of k times the same landmarks. Note that this corresponds to the set of matrices in $\mathbb{R}^{k \times 2}$ that are not of full rank. The orthogonal subspace $\mathbf{1}_{k \times 2}^\perp$ is the set of centered configurations, *i.e.* configuration such that the barycenter of the k landmarks is at the origin $(0, 0)$. We have the orthogonal decomposition,

$$\mathbb{R}^{k \times 2} = \mathbf{1}_{k \times 2}^\perp \oplus \mathbf{1}_{k \times 2}.$$

In what follows, for any configuration $\mathbf{x} \in \mathbb{R}^{k \times 2}$ we write $\mathbf{x} = \mathbf{x}_0 + \bar{\mathbf{x}} \in \mathbf{1}_{k \times 2}^\perp \oplus \mathbf{1}_{k \times 2}$. We call $\mathbf{x}_0$ the centered configuration of $\mathbf{x}$ and $\bar{\mathbf{x}} = \mathbf{1}_k \otimes \left(\frac{1}{k} \sum_{\ell=1}^k x_\ell^{(1)}, \frac{1}{k} \sum_{\ell=1}^k x_\ell^{(2)} \right)$ the degenerated configuration associated to $\mathbf{x}$, see Figure 2.1.

Definition 2.2. Given a configuration $\mathbf{x}$ in $\mathbb{R}^{k \times 2}$, the *orbit* of $\mathbf{x}$ is defined as the set

$$\mathcal{G} \cdot \mathbf{x} := \{g \cdot \mathbf{x}, g \in \mathcal{G}\} \subset \mathbb{R}^{k \times 2}.$$

This set is also called the shape of $\mathbf{x}$.

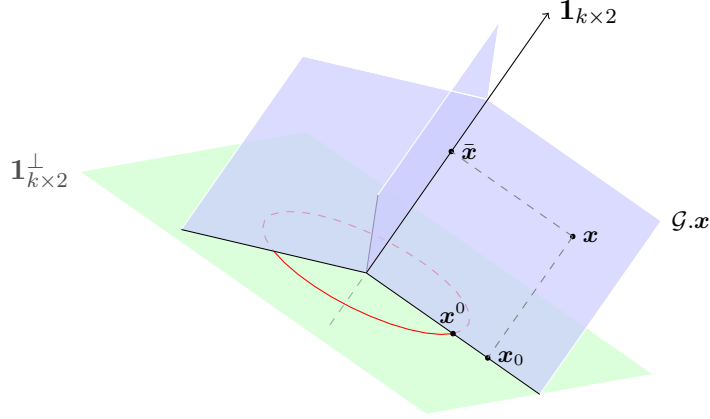


Figure 2.1: Three orbits of the action of the similarity group $\mathcal{G}$ are represented in blue. The space of centered configurations $\mathbf{1}_{k \times 2}^\perp$ is the green plane. The preshape sphere $\mathbb{S}_2^k$ is the red circle. For a particular $\mathbf{x} \in \mathbb{R}^{k \times 2}$, the centered version is $\mathbf{x}_0$ and the centered and normalized version is $\mathbf{x}^0$. The degenerated configuration associated to $\mathbf{x}$ is $\bar{\mathbf{x}}$.

Consider now a degenerated configuration $\mathbf{x} \in \mathbf{1}_{k \times 2}$. Its orbit $\mathcal{G}.\mathbf{x}$ is the entire subspace $\mathbf{1}_{k \times 2}$. It means that any two degenerated configurations have the same shape or, equivalently, that a point on the plane can be transformed onto another one by the action of $\mathcal{G}$. Note that the linear subspace $\mathbf{1}_{k \times 2}$ is stable by the action of $\mathcal{G}$, and that the action of $\mathcal{G}$ on $\mathbf{1}_{k \times 2}$ is not free, that is $g_1.\mathbf{x} = g_2.\mathbf{x}$ does not imply that $g_1 = g_2$. Now, if $\mathbf{x} \in \mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$ is a non-degenerated configuration of k landmarks, its orbit $\mathcal{G}.\mathbf{x}$ is a sub-manifold of $\mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$ of dimension $\dim(\mathcal{G}) = 4$. Remark also that the subspace $\mathbf{1}_{k \times 2}^\perp$ of $\mathbb{R}^{k \times 2}$ is stable under the action of scaling and rotations only. A loosely graphical representation of the orbits of $\mathcal{G}$ is given Figure 2.1.

Definition 2.3. Given a configuration $\mathbf{x} \in \mathbb{R}^{k \times 2}$, the stabilizer $I(\mathbf{x})$ is the closed subgroup of $\mathcal{G}$ which leaves $\mathbf{x}$ invariant, namely

$$I(\mathbf{x}) = \{g \in \mathcal{G} : g.\mathbf{x} = \mathbf{x}\}.$$

If $\mathbf{x} \in \mathbf{1}_{k \times 2}$ is a degenerated configuration, it can be written $\mathbf{x} = \mathbb{1}_k \otimes (x^{(1)}, x^{(2)})$ and its stabilizer is non trivial and is equal to $I(\mathbf{x}) = \{(a, \alpha, (x^{(1)}, x^{(2)}) - e^a(x^{(1)}, x^{(2)})R_\alpha), a \in \mathbb{R}, \alpha \in [-\pi, \pi[\}$. If $\mathbf{x} \in \mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$ is a non-degenerated configuration, its stabilizer $I(\mathbf{x})$ is reduced to the identity $\{e\}$. The action of $\mathcal{G}$ is said *free* if the stabilizer of any point is reduced to the identity. Hence, the action of $\mathcal{G}$ is free on the set of non-degenerated configurations of k -ads in $\mathbb{R}^2$.

We end this section by the following definition.

Definition 2.4. A section of the orbits of $\mathcal{G}$ is a subset of $\mathbb{R}^{k \times 2}$ containing a unique element of each orbit.

Two well-known examples of sections for the similarity group acting on $\mathbb{R}^{k \times 2}$ are the so-called Bookstein's and Kendall's coordinates defined section 2.3.2 and 2.3.4 in [DM98].

2.2.3. Generalized Procrustes analysis and Kendall's shape space

Shape space. Let $\mathbf{x} \in \mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}$ be a non-degenerated planar configuration of k landmarks. Let H be a $k \times k$ centering matrix, *i.e.* a matrix with real entries satisfying $\ker H = \mathbf{1}_k \cdot \mathbb{R}$. Typically one can take $H = Id_k - \frac{1}{k} \mathbf{1}_k \mathbf{1}_k'$. The effect of translation can be eliminated by centering the configuration $\mathbf{x}$ using the matrix H (see [DM98] for other centering methods), while the effect of isotropic scaling is removed by projecting the centered configuration on a unit sphere, which yields to the so-called pre-shape $\mathbf{x}^0$ of $\mathbf{x}$ defined as

$$\mathbf{x}^0 = \frac{H\mathbf{x}}{\|H\mathbf{x}\|_{\mathbb{R}^{k \times 2}}} \in \mathbb{R}^{k \times 2}.$$

Consider now the pre-shape sphere defined by $\mathbb{S}_2^k := \{\mathbf{x}^0, \mathbf{x} \in \mathbb{R}^{k \times 2} \setminus \mathbf{1}_{k \times 2}\}$ which can be identified to a sphere of dimension $2k - 3$. This normalization of the planar configurations amounts to choose a section for the action of the group generated by the translation and scaling in the plane, see Figure 2.1.

It now remains only the action of the rotation group on the pre-shape sphere. This action is isometric and free and the group $\mathbf{SO}(2)$ is compact, see definitions in Section I.3.3. Hence, the Kendall's shape space which is defined as the quotient of $\mathbb{S}_2^k$ by $\mathbf{SO}(2)$, namely

$$\Sigma_2^k := \mathbb{S}_2^k / \mathbf{SO}(2) = \{[\mathbf{x}^0] : \mathbf{x}^0 \in \mathbb{S}_2^k\} \text{ with } [\mathbf{x}^0] = \{\mathbf{x}^0 R_\alpha, \alpha \in [-\pi, \pi[\},$$

can be endowed with a Riemannian structure. In [Ken84], Kendall proved that a particular choice of Procrustean distance makes Σ_2^k isometric to a complex projective space. We refer [KBCL99] for a detail discussion on the geometric properties of the Shape Space.

Distances. The partial Procrustes distance is defined on the pre-shape sphere $\mathbb{S}_2^k$ as

$$d_P^2(\mathbf{x}^0, \mathbf{y}^0) = \inf_{\alpha \in [-\pi, \pi[} \|\mathbf{x}^0 - \mathbf{y}^0 R_\alpha\|_{\mathbb{R}^{k \times 2}}^2, \quad \mathbf{x}^0, \mathbf{y}^0 \in \mathbb{S}_2^k.$$

Hence, it is the (Euclidean) distance between the orbits $[\mathbf{x}^0] = \mathbf{SO}(2) \cdot \mathbf{x}^0$ and $[\mathbf{y}^0] = \mathbf{SO}(2) \cdot \mathbf{y}^0$ with $\mathbf{x}^0, \mathbf{y}^0 \in \mathbb{S}_2^k$. Let us introduce the group of transformations of the plane $\mathcal{H} \simeq \mathbb{R} \times \mathbf{SO}(2)$ generated by the scaling and rotations. The action of $h \in \mathcal{H}$ on the centered configuration $\mathbf{x}^0$ is given by $h \cdot \mathbf{x}^0 := e^\alpha \mathbf{x}^0 R_\alpha$ where $h = (a, \alpha) \in \mathbb{R} \times [-\pi, \pi[$. The Full Procrustes distance is then defined as

$$d_F^2(\mathbf{x}^0, \mathbf{y}^0) = \inf_{h \in \mathcal{H}} \|\mathbf{x}^0 - h \cdot \mathbf{y}^0\|_{\mathbb{R}^{k \times 2}}^2, \quad \mathbf{x}^0, \mathbf{y}^0 \in \mathbb{S}_2^k.$$

Consistency result. The full Procrustes sample mean $\hat{\mathbf{Y}}_F$ of $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ (see e.g. [Goo91, DM98]) is given by

$$\hat{\mathbf{Y}}_F = \operatorname{argmin}_{\mathbf{x}^0 \in \mathbb{S}_2^k} \sum_{j=1}^J d_F^2(\mathbf{Y}_j^0, \mathbf{x}^0). \quad (2.11)$$

It is sometimes defined in the equivalent way $\hat{\mathbf{Y}}_F = \frac{1}{J} \sum_{j=1}^J \hat{h}_j \cdot \mathbf{Y}_j^0$ where the generalized Procrustes parameters $\hat{h}_1, \dots, \hat{h}_J$ are the solutions over $(h_1, \dots, h_J) \in \mathcal{H}^J$ of the generalized Procrustes sum of squares

$$\begin{cases} \text{Minimize } G(h_1, \dots, h_J) = \sum_{j,j'=1}^J \|h_j \cdot \mathbf{Y}_j^0 - h_{j'} \cdot \mathbf{Y}_{j'}^0\|_{\mathbb{R}^{k \times 2}}^2 \\ \text{subject to } \left\| \frac{1}{J} \sum_{j=1}^J h_j \cdot \mathbf{Y}_j^0 \right\|_{\mathbb{R}^{k \times 2}}^2 = 1. \end{cases} \quad (2.12)$$

This method is known as Generalized Procrustes Analysis (GPA). Obviously, the mean shape $\mathbf{f}$ in model (2.2) does not necessarily belong to $\mathbb{S}_2^k$, and will generally not have the same orientation than $\hat{\mathbf{Y}}_F$. Therefore, using the Euclidean distance in $\mathbb{R}^{k \times 2}$ to compare $\hat{\mathbf{Y}}_F$ and $\mathbf{f}$ is not meaningful. Moreover, the criterion (2.11) is clearly invariant under the action of rotation on $\mathbb{R}^{(k-1) \times 2}$, meaning that $\hat{\mathbf{Y}}_F R_\alpha$ is a minimizer of (2.11) for any $\alpha \in [-\pi, \pi[$.

To study the consistency of the sample mean $\hat{\mathbf{Y}}_F$, it is classical in the literature on shape analysis to use the Kendall's shape space Σ_2^k . It has then been proved in [KM97, Le98] that, under mild assumptions on $\mathbf{f}$, $[\hat{\mathbf{Y}}_F]$ converges almost surely to $[\mathbf{f}^0]$ as $J \rightarrow +\infty$ when the covariance matrix of the random perturbation $\boldsymbol{\zeta}$ in (2.2) is isotropic (see Proposition 1 in [Le98] for a precise definition of isotropy for random variables belonging to $\mathbb{R}^{k \times 2}$). When the random perturbation are non-isotropic, [KM97, Le98] have also shown that $[\hat{\mathbf{Y}}_F]$ does not converge to $[\mathbf{f}^0]$ for some specific configurations of mean shape. Note that [KM97, Le98] also studied the convergence of the so-called partial Procrustes sample mean defined as

$$\hat{\mathbf{Y}}_P \in \operatorname{argmin}_{\mathbf{x}^0 \in \mathbb{S}_2^k} \frac{1}{J} \sum_{j=1}^J d_P^2(\mathbf{x}^0, \mathbf{Y}_j^0) = \operatorname{argmin}_{\mathbf{x}^0 \in \mathbb{S}_2^k} \min_{(\alpha_1, \dots, \alpha_J) \in [-\pi, \pi[^J} \frac{1}{J} \sum_{j=1}^J \left\| \mathbf{Y}_j^0 R_{\alpha_j} - \mathbf{x}^0 \right\|_{\mathbb{R}^{k \times 2}}^2,$$

and they arrived at the same conclusion on the consistency of $\hat{\mathbf{Y}}_P$. Therefore, it is commonly the belief that Procrustes sample means can be inconsistent when considering convergence in Σ_2^k and the asymptotic setting $J \rightarrow +\infty$.

Finally, note that one main difference between the approach developed in this chapter and the one used in the GPA is the choice of the normalization of the data. On the one hand, in GPA, the generalized Procrustes parameters $\hat{h}_1, \dots, \hat{h}_J$ are computed independently of the true parameters $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ since the data are projected on the pre-shape sphere *via* the constraint appearing in (2.12). On the other hand, to ensure the well-posedness of the problem (2.5) we choose to use the estimator $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda)$ that depends on $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ *via* the hypothesis that the deformations are centered around the identity. This yields to the constraint set Θ_0 which is defined as the intersection between a linear space and a subset of parameters set $\mathbb{R}^J \times [-\pi, \pi[^J \times \mathbb{R}^{2J}$, see Section 2.3 below.

2.3. Identifiability conditions

In general, the deformation model (2.2) does not allow to recover the true template $\mathbf{f}$, see e.g [Goo91, KM97]. Without any hypothesis on the mean pattern $\mathbf{f}$ or/and on the distribution of the deformation parameters one can only expect to estimate a representative in the orbit $\mathcal{G}.\mathbf{f}$.

2.3.1. Random deformations

Recall that in model (2.2) random deformations acting on the the mean pattern $\mathbf{f}$ are parametrized by a vector $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) = (a_1^*, \dots, a_J^*, \alpha_1^*, \dots, \alpha_J^*, b_1^*, \dots, b_J^*)$ in $\mathbb{R}^J \times [-\pi, \pi[^J \times \mathbb{R}^{2J}$. Then, let us make the following assumption,

Assumption 2.3. Let $0 < A, B$ and $0 < \mathcal{A} < \pi$ be three real numbers. The deformation parameters $(\mathbf{a}_j^*, \boldsymbol{\alpha}_j^*, \mathbf{b}_j^*)$, are i.i.d random variables with zero expectation and with values in

$$\Theta^* = \left[-\frac{A}{2}, \frac{A}{2}\right] \times \left[-\frac{\mathcal{A}}{2}, \frac{\mathcal{A}}{2}\right] \times [-B, B]^2.$$

The deformation parameters are independent of the error terms ζ_j , $j = 1, \dots, J$.

In this chapter, rather than working in $(\Theta^*)^J$ it is convenient to reorder the entries and to consider

$$\Theta^* = \left[-\frac{A}{2}, \frac{A}{2}\right]^J \times \left[-\frac{\mathcal{A}}{2}, \frac{\mathcal{A}}{2}\right]^J \times [-B, B]^{2J}.$$

Under Assumption 2.3, we have $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \Theta^*$. The compactness of Θ^* (and thus of Θ^*) is an essential condition to ensure the consistency of our procedure. Indeed, the estimation of the deformation parameters $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ and the mean pattern $\mathbf{f}$ is based on the minimization of the criterion (2.4). For example, if there is no restriction on the amplitude of the scaling parameter, the degenerate solution $a_j = -\infty$ for all $j = 1, \dots, J$ is always a zero of (2.4). Therefore, the minimization has to be performed under additional compact constraints which is the aim of Assumption 2.3.

Recall that the deformation parametrized by $(0, 0, 0)$ is the identity e . We assume that the deformation parameters have zero expectation since, in a lot of applications, it is reasonable to suppose that the deformations are “centered around the identity”.

2.3.2. The deterministic criterion D

Let $(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$ and consider the following criterion,

$$D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{1}{Jk} \sum_{j=1}^J \left\| (g_j^{-1} \cdot g_j^*) \cdot \mathbf{f} - \frac{1}{J} \sum_{j'=1}^J (g_{j'}^{-1} \cdot g_{j'}^*) \cdot \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2. \quad (2.13)$$

where $g_j = (a_j, \alpha_j, b_j)$ and $g_j^* = (a_j^*, \alpha_j^*, b_j^*)$ for all $j = 1, \dots, J$. The criterion D is a version without noise of the criterion M^λ defined at (2.4). The estimation procedure described in Section 2.1 is based on the convergence of the argmins of M^λ toward the argmin of D when k goes to infinity. As a consequence, choosing identifiability conditions amounts to fix a subset Θ_0 of $\mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$ on which D has a unique argmin. In the rest of this section, we determine the zeros of D , and then we fix a convenient constraint set Θ_0 that contains a unique point at which D vanishes.

The criterion D clearly vanishes at $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$. This minimum is not unique since easy algebra implies that

$$D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = 0 \quad \Longleftrightarrow \quad (g_j^{-1} \cdot g_j^*) \cdot \mathbf{f} = (g_{j'}^{-1} \cdot g_{j'}^*) \cdot \mathbf{f}, \quad \text{for all } j, j' = 1, \dots, J.$$

Suppose now that $\mathbf{f} \notin \mathbf{1}_{2 \times k}$ is a non-degenerated planar configuration. In Section 2.2.1, we have seen that the action of $\mathcal{G}$ on $\mathbf{f}$ is free, that is, the stabilizer $I(\mathbf{f})$ is reduced to the identity. Thus, we obtain,

$$\begin{aligned} D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = 0 & \iff g_j^{-1} \cdot g_j^* = g_{j'}^{-1} \cdot g_{j'}^*, & \text{for all } j, j' = 1, \dots, J \\ & \iff \begin{cases} a_j^* - a_j = a_{j'}^* - a_{j'}, \\ \alpha_j^* - \alpha_j = \alpha_{j'}^* - \alpha_{j'}, \\ e^{-a_j}(b_j^* - b_j)R_{-\alpha_j} = e^{-a_{j'}}(b_{j'}^* - b_{j'})R_{-\alpha_{j'}}, \end{cases} & \text{for all } j, j' = 1, \dots, J \end{aligned}$$

We have proved the following result,

Lemma 2.5. *Let $\mathbf{f} \in \mathbb{R}^{k \times 2}$ be a non-degenerated configuration of k -ads in the plane, i.e. $\mathbf{f} \notin \mathbf{1}_{2 \times k}$. Then, $D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = 0$ if and only if $(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b})$ belongs to the set*

$$(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * \mathcal{G} = \left\{ (\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0), (a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi[\times \mathbb{R}^2 \right\},$$

where $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0) = (a_1^* + a_0, \dots, a_J^* + a_0, \alpha_1^* + \alpha_0, \dots, \alpha_J^* + \alpha_0, e^{a_1^*} b_0 R_{\alpha_1^*} + b_1^*, \dots, e^{a_J^*} b_0 R_{\alpha_J^*} + b_J^*) \in \mathbb{R}^J \times [-\pi, \pi[\times \mathbb{R}^{2J}$.

Remark 2.6. Lemma 2.5 is simpler than it appears. By reordering the entries of the vector $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ there is an obvious correspondence between $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \boldsymbol{\Theta}^*$ and $(g_1^*, \dots, g_J^*) \in \mathcal{G}^J$ via the parametrization of the similarity group defined in Section 2.2.1. Hence, Lemma 2.5 tells us that the criterion D vanishes for all the vectors $(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \mathbb{R}^J \times [-\pi, \pi[\times \mathbb{R}^{2J}$ corresponding to the subset of the group $\mathcal{G}^J$ given by

$$(g_1^*, \dots, g_J^*) * \mathcal{G} = \{(g_1^* \cdot g_0, \dots, g_J^* \cdot g_0), g_0 \in \mathcal{G}\} \subset \mathcal{G}^J.$$

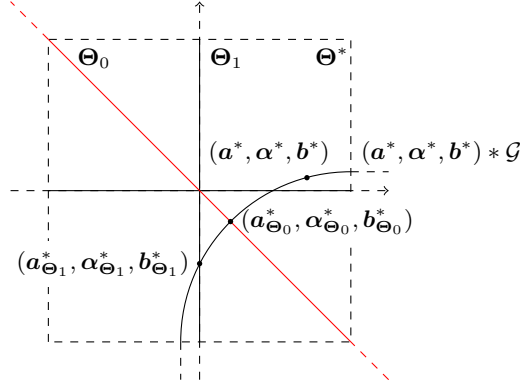
The “*” notation is nothing else than the right composition by a same $g_0 \in \mathcal{G}$ of all the entries of a $(g_1, \dots, g_J) \in \mathcal{G}^J$. Hence the subset $(g_1^*, \dots, g_J^*) * \mathcal{G}$ can be interpreted as the orbit of $(g_1^*, \dots, g_J^*) \in \mathcal{G}^J$ under the (right) action of $\mathcal{G}$. Indeed, $\mathcal{G}$ acts naturally by (right) composition on the all the coordinates of an element of $\mathcal{G}^J$.

2.3.3. The constraint set $\boldsymbol{\Theta}_0$

As we have identified the zeros of D , choosing the identifiability conditions amounts to fix a subset $\boldsymbol{\Theta}_0$ of $\mathbb{R}^J \times [-\pi, \pi[\times \mathbb{R}^{2J}$ which contains a unique element where D is zero. As $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ is a random variable, it may take its value anywhere in $\boldsymbol{\Theta}^*$. Hence, the set $\boldsymbol{\Theta}_0$ must intersect at a unique point, say $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*)$, each set $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * \mathcal{G}$. Finally, let us consider the constrained optimization problem

$$\begin{cases} \text{Minimize } D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \\ \text{Subject to } (\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \boldsymbol{\Theta}_0 \subset \mathbb{R}^J \times [-\pi, \pi[\times \mathbb{R}^{2J} \end{cases}$$

which is, by construction, well posed. It is convenient to choose $\boldsymbol{\Theta}_0$ to be of the form $\boldsymbol{\Theta}_0 = (\mathbb{R}^J \times [-\pi, \pi[\times \mathbb{R}^{2J}) \cap \mathcal{L}_0$ where $\mathcal{L}_0$ is a linear space of $\mathbb{R}^{4J}$. As said above, the linear space $\mathcal{L}_0$ is chosen so that for any $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ in $\boldsymbol{\Theta}^*$, there exists a unique point $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*)$ in $\boldsymbol{\Theta}_0$ that can be written as $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*, \mathbf{b}_{\boldsymbol{\Theta}_0}^*) = (\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0)$ for some $(a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi[\times \mathbb{R}^2$. See Lemma 2.5 for the definition of the “*” operation.


 Figure 2.2: Choice of identifiability conditions when $J = 2$.

Remark 2.7. As we have seen in Remark 2.6, the set $(a^*, \alpha^*, b^*) * \mathcal{G}$ can be interpreted as an orbit of the action of $\mathcal{G}$ on $\mathcal{G}^J$. In this terminology, the set Θ_0 can be viewed as a section of the orbits. Indeed, the section is the set of representatives $(a_{\Theta_0}^*, \alpha_{\Theta_0}^*, b_{\Theta_0}^*)$ of each orbit.

Note that we are putting together two different mathematical structures: the Euclidean structure *via* the linear constraint $\mathcal{L}$ and the group structure *via* the $*$ operation. In Chapter 1, an example of commutative group was treated and the two structures were compatible. Here, the similarity group $\mathcal{G}$ is non commutative and the linear constraint $\mathcal{L}$ have to be carefully chosen to indeed a section. See Figure 2.2 for an illustration.

We present here a choice of Θ_0 motivated by the fact that, under Assumption 2.3, the random deformation parameters have zero expectation. We consider the centered version of the estimated deformation parameter by choosing $\mathcal{L}_0 = \mathbf{1}_{4J}^\perp$, that is the orthogonal of the linear space $\mathbf{1}_{4J} = \mathbf{1}_{4J} \cdot \mathbb{R} \subset \mathbb{R}^{4J}$. Hence, the set Θ_0 defined equation (2.6) is

$$\Theta_0 = \{(a, \alpha, b) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^J \times \mathbb{R}^{2J}, (a_1 + \dots + a_J, \alpha_1 + \dots + \alpha_J, b_1 + \dots + b_J) = 0\}.$$

We have to show that for any $(a^*, \alpha^*, b^*) \in \Theta^*$ there is a $(a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ such that $(a_{\Theta_0}^*, \alpha_{\Theta_0}^*, b_{\Theta_0}^*) = (a^*, \alpha^*, b^*) * (a_0, \alpha_0, b_0) \in \Theta_0$. This amounts to solve the following equations

$$\begin{cases} a_1^* + a_0 + \dots + a_J^* + a_0 = 0, \\ \alpha_1^* + \alpha_0 + \dots + \alpha_J^* + \alpha_0 = 0, \\ b_1^* + e^{a_1^*} b_0 R_{\alpha_1^*} + \dots + b_J^* + e^{a_J^*} b_0 R_{\alpha_J^*} = 0. \end{cases} \quad (2.14)$$

After some computations, we obtain that equations (2.14) are verified if

$$(a_0, \alpha_0, b_0) = (-\bar{a}^*, -\bar{\alpha}^*, -\bar{b}^* (\overline{e^{a^*} R_{\alpha^*}})^{-1}), \quad (2.15)$$

where $\bar{a}^* = \frac{1}{J} \sum_{j=1}^J a_j^* \in \mathbb{R}$, $\bar{\alpha}^* = \frac{1}{J} \sum_{j=1}^J \alpha_j^* \in \mathbb{R}$, $\bar{b}^* = \frac{1}{J} \sum_{j=1}^J b_j^* \in \mathbb{R}^2$ and $\overline{e^{a^*} R_{\alpha^*}} = \frac{1}{J} \sum_{j=1}^J e^{a_j^*} R_{\alpha_j^*}$ is a 2×2 invertible matrix. Finally, $(a_{\Theta_0}^*, \alpha_{\Theta_0}^*, b_{\Theta_0}^*)$ is given by

$$([a_{\Theta_0}^*]_j, [\alpha_{\Theta_0}^*]_j, [b_{\Theta_0}^*]_j) = (a_j^* - \bar{a}^*, \alpha_j^* - \bar{\alpha}^*, b_j^* - e^{a_j^*} (\bar{b}^* (\overline{e^{a^*} R_{\alpha^*}})^{-1}) R_{\alpha_j^*}),$$

for $j = 1, \dots, J$. One can easily check that we have $(a_{\Theta_0}^*, \alpha_{\Theta_0}^*, b_{\Theta_0}^*) \in \Theta_0$.

Remark 2.8. Another classical approach is to fix, say the first observation as a reference. Practically, the criterion D could be optimized on the following subspace of $\mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$

$$\Theta_1 = \{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^{2J} \times \mathbb{R}^2, (a_1, \alpha_1, b_1) = 0\}.$$

and for any $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) \in \Theta^*$, the j -th coordinate of $(\mathbf{a}_{\Theta_1}^*, \boldsymbol{\alpha}_{\Theta_1}^*, \mathbf{b}_{\Theta_1}^*) = (\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0)$ is given by

$$([\mathbf{a}_{\Theta_1}^*]_j, [\boldsymbol{\alpha}_{\Theta_1}^*]_j, [\mathbf{b}_{\Theta_1}^*]_j) = (a_j^* - a_1^*, \alpha_j^* - \alpha_1^*, b_j^* - e^{a_j^* - a_1^*} b_1^* R_{\alpha_j^* - \alpha_1^*}),$$

where $(a_0, \alpha_0, b_0) = (a_1^*, \alpha_1^*, b_1^*)^{-1} = (-a_1^*, -\alpha_1^*, -e^{-a_1^*} b_1^* R_{-\alpha_1^*})$, see Figure 2.2 for an illustration.

2.4. The estimating procedure

2.4.1. A dimension reduction step

We use Fourier filtering to project the data into a low-dimensional space as follows. Assume for convenience that k is odd. For $\mathbf{x} = (\mathbf{x}'_1, \dots, \mathbf{x}'_k)' \in \mathbb{R}^{k \times 2}$, that is $\mathbf{x}_\ell = (x_\ell^{(1)}, x_\ell^{(2)}) \in \mathbb{R}^2$ for all $\ell = 1, \dots, k$, and $m = -\frac{k-1}{2}, \dots, \frac{k-1}{2}$, let

$$c_m(\mathbf{x}) = \sum_{\ell=1}^k \mathbf{x}_\ell e^{-i2\pi m \frac{\ell}{k}} = \left(\sum_{\ell=1}^k x_\ell^{(1)} e^{-i2\pi m \frac{\ell}{k}}, \sum_{\ell=1}^k x_\ell^{(2)} e^{-i2\pi m \frac{\ell}{k}} \right) \in \mathbb{C}^2$$

be the m -th (discrete) Fourier coefficient of $\mathbf{x}$. Let $\lambda \in \{1, \dots, \frac{k-1}{2}\}$ be a smoothing (or dimension reduction) parameter, and define for each $\mathbf{Y}_j$ the smoothed shapes

$$\hat{\mathbf{f}}_j^\lambda = \left(\frac{1}{k} \sum_{0 \leq |m| \leq \lambda} c_m(\mathbf{Y}_j) e^{i2\pi m \frac{\ell}{k}} \right)_{\ell=1}^k = \mathbf{A}^\lambda \mathbf{Y}_j \in \mathbb{R}^{k \times 2}.$$

Recall that the definition of $\mathbf{A}^\lambda$ was given in (2.3). The range $\mathcal{V}^\lambda$ of the linear application $\mathbf{A}^\lambda$ is a subset of $\mathbb{R}^{k \times 2}$ of dimension $2\lambda + 1$. In section 2.2.1, we have shown that the similarity group is not free on the subset $\mathbf{1}_{k \times 2}$ of degenerated configurations composed of k identical landmarks, see Section 2.2.1. That is why we are going to treat separately the subspace $\mathbf{1}_{k \times 2}$ and $\mathbf{1}_{k \times 2}^\perp$ by considering the matrices

$$\bar{A} = \frac{1}{k} \mathbf{1}_k \mathbf{1}_k' \quad \text{and} \quad A_0^\lambda = \left(\frac{1}{k} \sum_{0 < |m| \leq \lambda} e^{i2\pi m \frac{\ell - \ell'}{k}} \right)_{\ell, \ell'=1}^k. \quad (2.16)$$

Remark that $\bar{A}$ is a projection matrix on the one dimensional sub-space $\bar{\mathcal{V}} := \mathbf{1}_k \cdot \mathbb{R} = \{c \mathbf{1}_k : c \in \mathbb{R}\}$ of $\mathbb{R}^k$. The space $\bar{\mathcal{V}} \times \bar{\mathcal{V}} = \mathbf{1}_{k \times 2}$ where $\mathbf{1}_{k \times 2}$ is the space of the degenerated configurations. The matrix A_0^λ is a projection matrix in a (trigonometric) sub-space $\mathcal{V}_0^\lambda$ of dimension 2λ . Note that it is included in the linear space $\bar{\mathcal{V}}^\perp = \{x \in \mathbb{R}^k : x' \mathbf{1}_k = 0\}$. Hence, $\mathcal{V}_0^\lambda \times \mathcal{V}_0^\lambda$ is a linear subspace of $\mathbf{1}_{k \times 2}^\perp$ which is the space of the centered configurations. We have,

$$\mathbf{A}^\lambda = A_0^\lambda + \bar{A} \quad \text{and} \quad \mathcal{V}^\lambda = \mathcal{V}_0^\lambda \oplus \bar{\mathcal{V}}. \quad (2.17)$$

Thus, we can write the smoothed shape as

$$\hat{\mathbf{f}}_j^\lambda = \mathbf{A}^\lambda \mathbf{Y}_j = A_0^\lambda \mathbf{Y}_j + \bar{A} \mathbf{Y}_j \in (\bar{\mathcal{V}} \times \bar{\mathcal{V}}) \oplus (\mathcal{V}_0^\lambda \times \mathcal{V}_0^\lambda)$$

where $\bar{\mathcal{V}} \times \bar{\mathcal{V}} = \mathbf{1}_{k \times 2}$ and $\mathcal{V}_0^\lambda \times \mathcal{V}_0^\lambda \subset \mathbf{1}_{k \times 2}^\perp$. In other words, $A_0^\lambda \mathbf{Y}_j$ is the smoothed centered configuration associated to $\mathbf{Y}_j$ and $\bar{A} \mathbf{Y}_j$ is the degenerated configuration given by the Euclidean mean of the k landmarks composing $\mathbf{Y}_j$. Finally, remark that the low pass filtering and the action of similarity

group commute, that is, we have for all $g \in \mathcal{G}$ and $\mathbf{f} \in \mathbb{R}^{k \times 2}$

$$g.(\mathbf{A}^\lambda \mathbf{f}) = e^a \mathbf{A}^\lambda \mathbf{f} R_\alpha + \mathbf{1}_k \otimes b = \mathbf{A}^\lambda (e^a \mathbf{f} R_\alpha + \mathbf{1}_k \otimes b) = \mathbf{A}^\lambda (g.\mathbf{f}).$$

2.4.2. Estimation of the deformation parameters

Recall that for fixed J and k , the estimator $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda)$ of $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$ is defined by the optimization problem (2.5). Nevertheless, as it is suggested by the discussion of Sections 2.2.1 and 2.4.1, one can carry out the estimation process in two steps. First, we estimate the rotation and scaling parameters on the space $\mathcal{V}_0^\lambda \times \mathcal{V}_0^\lambda \subset \mathbf{1}_{k \times 2}^\perp$ of the centered configurations. We then use these estimators to estimate the translation parameters which act on $\bar{\mathcal{V}} \times \bar{\mathcal{V}} = \mathbf{1}_{k \times 2}$. Note that this procedure is equivalent to the optimization problem (2.5) as it is shown by Lemma 2.9 below.

Estimation of rotations and scaling. It is made using the smoothed and centered configurations. Let us introduce the following criterion,

$$M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{-a_j} A_0^\lambda \mathbf{Y}_j R_{-\alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{-a_{j'}} A_0^\lambda \mathbf{Y}_{j'} R_{-\alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2.$$

Let $\mathbf{1}_J^\perp = \{(a_1, \dots, a_J) \in \mathbb{R}^J, a_1 + \dots, a_J = 0\}$ and consider the following space,

$$\boldsymbol{\Theta}_0^{a, \alpha} = ([-A, A]^J \cap \mathbf{1}_J^\perp) \times ([-\mathcal{A}, \mathcal{A}]^J \cap \mathbf{1}_J^\perp),$$

Then, we consider the estimator of the rotation and scaling parameters given by,

$$(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) \in \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{a, \alpha}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}). \quad (2.18)$$

Estimation of translations. As we have defined an estimator of the rotation and scaling parameters, let us consider the following criterion,

$$\bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{-a_j} \bar{A}(\mathbf{Y}_j - \mathbf{1}_k \otimes b_j) R_{-\alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{-a_{j'}} \bar{A}(\mathbf{Y}_{j'} - \mathbf{1}_k \otimes b_{j'}) R_{-\alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2$$

and the following space

$$\boldsymbol{\Theta}_0^b = \{(b_1^{(1)}, \dots, b_J^{(1)}, b_1^{(2)}, \dots, b_J^{(2)}) \in \mathbb{R}^{2J}, b_1^{(1)} + \dots + b_J^{(1)} = b_1^{(2)} + \dots + b_J^{(2)} = 0\}.$$

The estimator of the translation parameters is now given by,

$$\hat{\mathbf{b}}^\lambda = \underset{\mathbf{b} \in \boldsymbol{\Theta}_0^b}{\operatorname{argmin}} \bar{M}(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \mathbf{b}). \quad (2.19)$$

We emphasize that the estimators of the translation parameters depend on the estimated rotation and scaling parameters. It is shown in the proof of Lemma 2.9 below that we have an explicit expression of $\hat{\mathbf{b}}^\lambda$ given by

$$\hat{\mathbf{b}}^\lambda = (\bar{\mathbf{Y}}_1 - e^{a_1} \bar{\mathbf{Y}} (\overline{e^a R_\alpha})^{-1} R_{\alpha_1}, \dots, \bar{\mathbf{Y}}_J - e^{a_J} \bar{\mathbf{Y}} (\overline{e^a R_\alpha})^{-1} R_{\alpha_J}) \quad (2.20)$$

where $\overline{e^a R_\alpha} = \frac{1}{J} \sum_{j=1}^J e^{a_j} R_{\alpha_j} \in \mathbb{R}^{2 \times 2}$, $\bar{\mathbf{Y}}_j = \frac{1}{k} \sum_{\ell=1}^k [\mathbf{Y}_j]_\ell \in \mathbb{R}^2$, that is $\bar{A} \mathbf{Y}_j = \mathbf{1}_k \otimes \bar{\mathbf{Y}}_j \in \mathbb{R}^{k \times 2}$ is a degenerated configuration, and $\bar{\mathbf{Y}} = \frac{1}{J} \sum_{j=1}^J \bar{\mathbf{Y}}_j \in \mathbb{R}^2$.

This two steps procedure is equivalent to the optimization problem (2.5) as we have the following decompositions $M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) + \bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b})$ and $\boldsymbol{\Theta}_0 = \boldsymbol{\Theta}_0^{a, \boldsymbol{\alpha}} \times \boldsymbol{\Theta}_0^b$. More precisely, we have,

$$\textbf{Lemma 2.9. } (\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda) \in \underset{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \boldsymbol{\Theta}_0}{\operatorname{argmin}} M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \iff \begin{cases} (\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) \in \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{a, \boldsymbol{\alpha}}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) \\ \hat{\mathbf{b}}^\lambda = \underset{\mathbf{b} \in \boldsymbol{\Theta}_0^b}{\operatorname{argmin}} \bar{M}(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \mathbf{b}). \end{cases}$$

2.4.3. Procrustes analysis of smoothed shapes

The estimator of the mean pattern $\mathbf{f}$ is defined by

$$\hat{\mathbf{f}}^\lambda = \frac{1}{J} \sum_{j=1}^J e^{-\hat{\mathbf{a}}_j^\lambda} (\mathbf{A}^\lambda \mathbf{Y}_j - \mathbf{1}_k \otimes \hat{\mathbf{b}}_j^\lambda) R_{-\hat{\boldsymbol{\alpha}}_j^\lambda},$$

where $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$ are given by (2.18) and $\hat{\mathbf{b}}^\lambda$ by (2.19).

2.5. Consistency results

In what follows, C, C_0, C_1 denote positive constants whose value may change from line to line. The notation $C(\cdot)$ specifies the dependency of C on some quantities.

2.5.1. Covariance structure of the error terms $\boldsymbol{\zeta}$

The error term $\boldsymbol{\zeta}$ that appears in model (2.2) models shape variations of the mean pattern $\mathbf{f}$. Under Assumption 2.2 the error term $\boldsymbol{\zeta}$ is a centered Gaussian random variable. Let $\boldsymbol{\zeta}^{(i)} = (\zeta_1^{(i)}, \dots, \zeta_k^{(i)})'$, for $i = 1, 2$, be the column vector containing the i -th coordinates of each $\boldsymbol{\zeta}_\ell$, $\ell = 1, \dots, k$ and recall our notations: the column vector $\tilde{\boldsymbol{\zeta}} = (\boldsymbol{\zeta}^{(1)'} , \boldsymbol{\zeta}^{(2)'})' \in \mathbb{R}^{2k}$ is the vectorized version of $\boldsymbol{\zeta} = (\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}) \in \mathbb{R}^{k \times 2}$. Thence, the covariance matrix of $\tilde{\boldsymbol{\zeta}}$ which was denoted by $\boldsymbol{\Sigma}$ is equals to

$$\boldsymbol{\Sigma} = \mathbb{E}(\tilde{\boldsymbol{\zeta}} \tilde{\boldsymbol{\zeta}}') = \mathbb{E} \begin{pmatrix} \boldsymbol{\zeta}^{(1)} \boldsymbol{\zeta}^{(1)'} & \boldsymbol{\zeta}^{(1)} \boldsymbol{\zeta}^{(2)'} \\ \boldsymbol{\zeta}^{(2)} \boldsymbol{\zeta}^{(1)'} & \boldsymbol{\zeta}^{(2)} \boldsymbol{\zeta}^{(2)'} \end{pmatrix} \in \mathbb{R}^{2k \times 2k}.$$

Under Assumption 2.2 the matrix $\boldsymbol{\Sigma}$ can be rather general as we only make an hypothesis on its largest eigenvalue $\gamma_{\max}(k)$.

Remark 2.10. We are able to describe the effect of a random similarity on the law of $\boldsymbol{\zeta}$. The similarity group acts non-linearly on the perturbed mean pattern $\mathbf{f} + \boldsymbol{\zeta}$ as follow,

$$g(\mathbf{f} + \boldsymbol{\zeta}) = e^a (\mathbf{f} + \boldsymbol{\zeta}) R_\alpha + \mathbf{1}_k \otimes b = g \cdot \mathbf{f} + e^a \boldsymbol{\zeta} R_\alpha$$

for any $g = (a, \alpha, b) \in \mathcal{G}$. Now, given a random $(a, \alpha) \in [-A, A] \times [-\mathcal{A}, \mathcal{A}]$ which is independent of $\boldsymbol{\zeta}$, we denote the vectorized version of $e^a \boldsymbol{\zeta} R_\alpha$ by $\tilde{\boldsymbol{\zeta}}_{a, \alpha} = e^a \cos(\alpha) \begin{pmatrix} \boldsymbol{\zeta}^{(1)} \\ \boldsymbol{\zeta}^{(2)} \end{pmatrix} + e^a \sin(\alpha) \begin{pmatrix} \boldsymbol{\zeta}^{(2)} \\ -\boldsymbol{\zeta}^{(1)} \end{pmatrix}$. Then,

conditionally to (a, α) the random variable $\zeta_{a, \alpha}$ is still Gaussian. In practice, as the deformation parameters (a_j^*, α_j^*) , $j = 1, \dots, J$, are random variables, the law of $\zeta_{a_j^*, \alpha_j^*}$ is unknown. To tackle this problem and to show consistency results given Section 2.5.3 and 2.5.4, we use uniform bounds of the type $\|\zeta_{a, \alpha}\|_{\mathbb{R}^{2k}}^2 \leq C(A)\|\tilde{\zeta}\|_{\mathbb{R}^{2k}}$. Nevertheless, because of the independence, the expectation of $\zeta_{a, \alpha}$ is zero and its covariance matrix is

$$\begin{aligned} \mathbb{E}(\tilde{\zeta}_{a, \alpha}(\tilde{\zeta}_{a, \alpha})') &= \mathbb{E}(e^{2a} \cos^2(\alpha)) \mathbb{E} \begin{pmatrix} \zeta^{(1)} \zeta^{(1)'} & \zeta^{(1)} \zeta^{(2)'} \\ \zeta^{(2)} \zeta^{(1)'} & \zeta^{(2)} \zeta^{(2)'} \end{pmatrix} + \mathbb{E}(e^{2a} \sin^2(\alpha)) \mathbb{E} \begin{pmatrix} \zeta^{(2)} \zeta^{(2)'} & -\zeta^{(2)} \zeta^{(1)'} \\ -\zeta^{(1)} \zeta^{(2)'} & \zeta^{(1)} \zeta^{(1)'} \end{pmatrix} \\ &\quad + \mathbb{E}(e^{2a} \cos(\alpha) \sin(\alpha)) \mathbb{E} \begin{pmatrix} \zeta^{(1)} \zeta^{(2)'} + \zeta^{(2)} \zeta^{(1)'} & \zeta^{(2)} \zeta^{(2)'} - \zeta^{(1)} \zeta^{(1)'} \\ \zeta^{(2)} \zeta^{(2)'} - \zeta^{(1)} \zeta^{(1)'} & -\zeta^{(2)} \zeta^{(1)'} - \zeta^{(1)} \zeta^{(2)'} \end{pmatrix}. \end{aligned}$$

In the particular case where $\zeta^{(1)}$ and $\zeta^{(2)}$ are i.i.d, the expectation of the cross products are zero and we have $\mathbb{E}(\tilde{\zeta}_{a, \alpha}(\tilde{\zeta}_{a, \alpha})') = \mathbb{E}(e^{2a})\Sigma$.

2.5.2. Regularity conditions on f

Let us define the following smoothness class of functions characterized in term of decay of their Fourier coefficients. Let $L > 0$ and $s \geq \frac{3}{2}$ and consider

$$H_s(L) = \left\{ f = (f^{(1)}, f^{(2)}) \in L^2([0, 1], \mathbb{R}^2), \sum_{m \in \mathbb{Z}} (1 + |m|^2)^s (|c_m(f^{(1)})|^2 + |c_m(f^{(2)})|^2) < L \right\} \quad (2.21)$$

where $c_m(f) = (c_m(f^{(1)}), c_m(f^{(2)})) = \left(\int_0^1 f^{(1)}(t) e^{-i2\pi m t} dt, \int_0^1 f^{(2)}(t) e^{-i2\pi m t} dt \right) \in \mathbb{C}^2$ is the m -th Fourier coefficient of $f = (f^{(1)}, f^{(2)}) \in L^2([0, 1], \mathbb{R}^2)$, for $m \in \mathbb{Z}$. The space $H_s(L)$ is the ball of radius $L > 0$ of a Sobolev space of regularity $s \geq \frac{3}{2}$.

Let $\mathbf{Y} = e^a(\mathbf{f} + \zeta)R_\alpha + \mathbf{1}_k \otimes b$ be a realization of model (2.2) and let $\hat{\mathbf{f}}^\lambda = \mathbf{A}^\lambda \mathbf{Y}$. Recall also that $g \cdot \mathbf{f} = e^a \mathbf{f} R_\alpha + \mathbf{1}_k \otimes b$ and consider the quadratic risk

$$\begin{aligned} \frac{1}{k} \mathbb{E}(\|\hat{\mathbf{f}}^\lambda - g \cdot \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2) &= \frac{1}{k} \mathbb{E}(\|\mathbf{A}^\lambda g \cdot \mathbf{f} - g \cdot \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2) + \frac{1}{k} \mathbb{E}(\|e^a \mathbf{A}^\lambda \zeta R_\alpha\|_{\mathbb{R}^{k \times 2}}^2) \\ &\leq \underbrace{e^{2A} \frac{1}{k} \|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2}_{\text{Bias}} + \underbrace{e^{2A} \frac{1}{k} \mathbb{E}(\|\mathbf{A}^\lambda \zeta\|_{\mathbb{R}^{k \times 2}}^2)}_{\text{Variance}}. \end{aligned}$$

If Assumption 2.1 holds, *i.e.* the mean function f belongs to $H_s(L)$, $s \geq \frac{3}{2}$ and $L > 0$, then we can use the classical bias/variance tradeoff that exists for projection estimators in Sobolev ball, see Section 1.10 of [Tsy09]. Typically, by choosing a frequency cutoff $\lambda = \lambda(k) = k^{\frac{1}{2s+1}}$, the bias term can be bounded by a term of order $k^{-\frac{2s}{2s+1}}$ and the variance term by a term of order $\gamma_{\max}(k) k^{-\frac{2s}{2s+1}}$ for $s \geq \frac{3}{2}$, see Lemma 2.21 for details.

2.5.3. Consistent estimation of the deformation parameters

Rotation and scaling. Recall that the rotation and scaling parameters are estimated separately on the smoothed and centered observations. We have the following result,

Theorem 2.11. *Consider model (2.2) and suppose that Assumptions 2.1 and 2.2 hold and that Assumption 2.3 is verified with $\max\{A, \mathcal{A}\} < 0.1$. Consider $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$ the estimators defined in equation*

(2.18). If $\lambda = k^{\frac{1}{2s+1}}$ then for all $x > 0$ we have

$$\mathbb{P}\left(\frac{1}{J}\|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2 \geq C_1(L, s, A, \mathcal{A}, \mathbf{f}_0)A_1(k, J, x) + C_2(A, \mathcal{A})A_2(J, x)\right) \leq 4e^{-x}, \quad (2.22)$$

where $C_1(L, s, A, \mathcal{A}, \mathbf{f}_0), C_2(A, \mathcal{A}) > 0$ are positive constants independent of k and J , $A_1(k, J, x) = F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x))$ with $V_1(k, J, x) = \gamma_{\max}(k)k^{\frac{-2s}{2s+1}}\left(1 + 2\sqrt{\frac{x}{J}k^{\frac{-1}{2s+1}}} + \frac{x}{J}k^{\frac{-1}{2s+1}}\right)$ and $A_2(J, x) = \left(\frac{x}{3J} + \sqrt{\frac{2x}{J}}\right)^2$, where $F(u) = u + \sqrt{u}$, for all $u \geq 0$.

Remark that a direct consequence of Theorem 2.11 is the consistency of $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$ to $(\mathbf{a}^*, \boldsymbol{\alpha}^*)$ when $k, J \rightarrow \infty$. Indeed, we have $\lim_{u \rightarrow 0} F(u) = 0$ and under Assumption 2.2 and for any fixed $J \geq 1$ and $x > 0$, the term $V_1(k, J, x)$ tends to zero as k goes to infinity. Hence for any $x > 0$ and $J \geq 1$, we have

$$\lim_{k \rightarrow \infty} A_1(k, J, x) = 0.$$

Proposition 2.14 then ensures the convergence of $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$ to $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$ as J remains fixed and $k \rightarrow \infty$. Now, for all $x > 0$, we have

$$\lim_{J \rightarrow \infty} A_2(J, x) = 0.$$

Thus a double asymptotic $\min\{k, J\} \rightarrow \infty$ ensures that $\frac{1}{J}\|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2$ tends to 0 in probability.

Translation parameters. We have the following result,

Theorem 2.12. Consider the hypothesis and notations of Theorem 2.11 and the estimator $\hat{\mathbf{b}}^\lambda$ given by formula (2.19). Then we have for all $x > 0$,

$$\mathbb{P}\left(\frac{1}{J}\|\hat{\mathbf{b}}^\lambda - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2 \geq C_3(L, s, A, \mathcal{A}, B, \mathbf{f})A_3(k, J, x) + C_4(A, \mathcal{A}, B)A_2(J, x)\right) \leq 9e^{-x}, \quad (2.23)$$

where $C_3(L, s, A, \mathcal{A}, B, \mathbf{f}), C_4(A, \mathcal{A}, B) > 0$ are positive constants independent of k and J , $A_3(k, J, x) = F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)) + V_2(k, J, x)$ with $V_2(k, J, x) = \frac{\gamma_{\max}(k)}{k}(1 + \sqrt{2\frac{x}{J}} + \frac{x}{J})$.

Similar comments to those made after Theorem 2.11 are still valid here. For any $J \in \mathbb{N}$ and $x > 0$, we have

$$\lim_{k \rightarrow \infty} A_3(k, J, x) = 0,$$

since $V_2(k, J, x)$ tends to 0 as k goes to infinity by Assumption 2.2. Then Proposition 2.16 ensures the convergence in probability of $\hat{\mathbf{b}}^\lambda$ to $\mathbf{b}_{\boldsymbol{\Theta}_0}^*$ with only an asymptotic in k . In the double asymptotic setting $k, J \rightarrow \infty$, Theorem 2.12 ensure the consistency of $\hat{\mathbf{b}}^\lambda$ to the true value $\mathbf{b}^*$ of the translation parameters.

Let us now make some comments about the hypothesis of Theorem 2.11 and 2.12. The extra condition $A, \mathcal{A} < 0.1$ to Assumption 2.3 is required to derive the concentration inequalities (2.22) and (2.23) but is not necessary to ensure the convergence in probability of $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \hat{\mathbf{b}}^\lambda)$ to $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*)$. A careful reading of the proof of Theorem 2.11 and 2.12 shows that restrictions on A and $\mathcal{A}$ are needed only in the proof of Lemmas 2.20 to determine a uniform lower bound in J of the criterion D_0 . At fixed $J \in \mathbb{N}$ and when $k \rightarrow \infty$, Lemmas 2.18 and 2.19 ensure the consistency of $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$ to $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$ and thus of $\hat{\mathbf{b}}^\lambda$ to $\mathbf{b}_{\boldsymbol{\Theta}_0}^*$ without the extra condition $A, \mathcal{A} < 0.1$ on Assumption 2.3.

2.5.4. Consistent estimation of the mean shape

Theorem 2.13. *Consider model (2.2) and suppose that Assumptions 2.1 and 2.2 hold and that Assumption 2.3 is verified with $\max\{A, \mathcal{A}\} < 0.1$. Consider the estimator $\hat{\mathbf{f}}^\lambda$ defined in (2.7) and let $\lambda = k^{\frac{1}{2s+1}}$. Then, we have for all $x > 0$,*

$$\mathbb{P}\left(\frac{1}{k}\|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f})\left(A_1(k, J, x) + A_3(k, J, x) + A_2(J, x)\right)\right) \leq 14e^{-x},$$

where $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$ is a constant independent of k and J , $A_1(k, J, x) = F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x))$ with $V_1(k, J, x) = \gamma_{\max}(k)k^{\frac{-2s}{2s+1}}\left(1 + 2\sqrt{\frac{x}{J}k^{\frac{-1}{2s+1}}} + \frac{x}{J}k^{\frac{-1}{2s+1}}\right)$, $A_3(k, J, x) = V_2(k, J, x) + F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x))$ with $V_2(k, J, x) = \frac{\gamma_{\max}(k)}{k}(1 + \sqrt{2\frac{x}{J}} + \frac{x}{J})$ and $A_2(J, x) = \left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2$.

The terms $A_1(k, J, x)$, $A_3(k, J, x)$ and $A_2(J, x)$ that appear in the statement of Theorem 2.13 are the same to those appearing in Theorems 2.11 and 2.12. As a consequence, we have $\frac{1}{k}\|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \rightarrow 0$ in probability when $\min\{k, J\} \rightarrow \infty$, see the comments after Theorems 2.11 and 2.12. Let

$$\mathbf{f}_{\Theta_0} = e^{\bar{\mathbf{a}}^*}(\mathbf{f} + \mathbb{1}_k \otimes \bar{\mathbf{b}}^*(\overline{e^{\mathbf{a}^*} R_{\alpha^*}})^{-1})R_{\bar{\alpha}^*} \quad (2.24)$$

where $\bar{\mathbf{a}}^* = \frac{1}{J} \sum_{j=1}^J \mathbf{a}_j^* \in \mathbb{R}$, $\bar{\alpha}^* = \frac{1}{J} \sum_{j=1}^J \alpha_j^* \in \mathbb{R}$, $\bar{\mathbf{b}}^* = \frac{1}{J} \sum_{j=1}^J \mathbf{b}_j^* \in \mathbb{R}^2$ and $\overline{e^{\mathbf{a}^*} R_{\alpha^*}} = \frac{1}{J} \sum_{j=1}^J e^{\mathbf{a}_j^*} R_{\alpha_j^*}$ is an invertible 2×2 matrix, see also formula (2.15). A slight modification of the proof of Theorem 2.13 gives the following inequality,

$$\mathbb{P}\left(\frac{1}{k}\|\hat{\mathbf{f}}^\lambda - \mathbf{f}_{\Theta_0}\|^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f})\left(A_1(k, J, x) + A_3(k, J, x)\right)\right) \leq 14e^{-x}.$$

Theorem 2.13 gives rates of convergences of $\hat{\mathbf{f}}^\lambda$ to $\mathbf{f}$ thanks to a concentration inequality. As in Theorems 2.11 and 2.12, there is no need to assume an extra bound on A and $\mathcal{A}$ to ensure the convergence in probability of $\hat{\mathbf{f}}^\lambda$ to $\mathbf{f}$, see the discussion following Theorems 2.11 and 2.12. In this case, we were not able to derive a rate of convergence and it yields to statements (2.8) and (2.9) in Theorem 2.1.

2.6. Numerical experiments

2.6.1. Description of the data

We make here some numerical simulations to show the effect of the dimension k and the number J of observations on the estimation of the deformation parameters and the mean pattern with data generated by model (2.2). Different types of noise are considered. For all $t \in [0, 1]$, let

$$f(t) = (10 \sin^2(\pi t) + \cos(10\pi) + 20, 2 \sin(6\pi t) - 11 \sin^2(\pi t) + 12 \exp(-25(t - 0.4)^2) + 1).$$

This curve is plotted in Figure 2.3. The deformation parameters $(a_j^*, \alpha_j^*, b_j^*)$, $j = 1, \dots, J$, are i.i.d uniform random variables taking their values in $\Theta = [-\frac{1}{4}, \frac{1}{4}] \times [-\frac{1}{2}, \frac{1}{2}] \times [-1, 1]^2$. The law of the

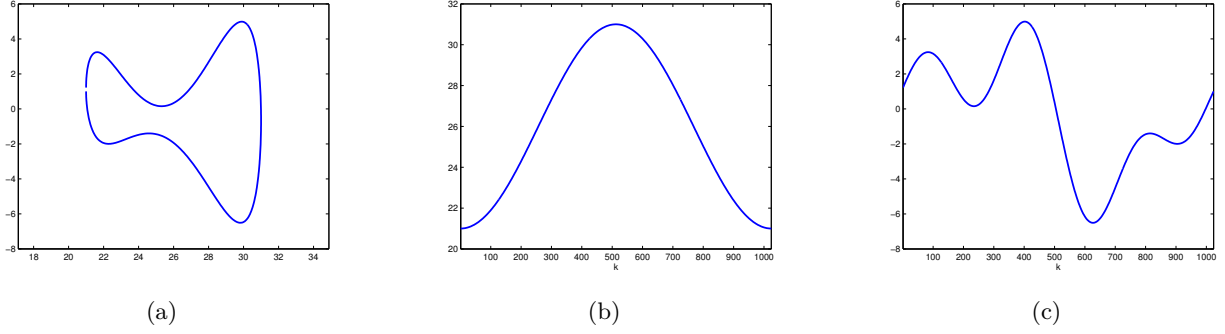


Figure 2.3: (a) Plot of the mean pattern $\mathbf{f} = (\mathbf{f}^{(1)}, \mathbf{f}^{(2)})$ used in the simulations with $k = 1024$. (b) The first coordinates $\mathbf{f}^{(1)}$. (c) The second coordinates $\mathbf{f}^{(2)}$.

deformation parameters is supposed to be unknown *a priori* and the minimization is performed on the constraint set

$$\Theta_0 = \left\{ (\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in [-1, 1]^J \times [-1, 1]^J \times [-5, 5]^{2J}, \sum_{j=1}^J a_j = 0, \sum_{j=1}^J \alpha_j = 0 \text{ and } \sum_{j=1}^J b_j = 0 \right\}.$$

Recall our notations : the error term is denoted by $\boldsymbol{\zeta} = (\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}) \in \mathbb{R}^{k \times 2}$ and the vectorized version of $\boldsymbol{\zeta}$ is denoted by $\tilde{\boldsymbol{\zeta}} \in \mathbb{R}^{2k}$. The simulations were run with three different kinds of noise.

White noise: the random variable $\tilde{\boldsymbol{\zeta}} = (\zeta_1^{(1)}, \zeta_2^{(1)}, \dots, \zeta_1^{(k)}, \zeta_2^{(k)})' \in \mathbb{R}^{2k}$ is a centered Gaussian vector of variance

$$\boldsymbol{\Sigma}_1 = 4Id_{2k}.$$

We have $\gamma_{\max}(k) = 4$ and this correspond to an isotropic Gaussian noise as in [KM97, Le98], see Figure 2.4.

Stationary noise: the random variable $\tilde{\boldsymbol{\zeta}}$ is a centered Gaussian vector of variance $\boldsymbol{\Sigma}_2$ with

$$\boldsymbol{\Sigma}_2^{\frac{1}{2}} = Id_2 \otimes \left[\frac{1}{2} \exp \left(-\frac{|\ell - \ell'|^2}{100} \right) \right]_{\ell, \ell'=1}^k$$

Hence, $\boldsymbol{\Sigma}_2$ is a Toeplitz matrix and it follows from classical matrix theory, see e.g. [HJ90], that $\gamma_{\max}(k)$ is bounded (here $\gamma_{\max}(k) \leq 80$). See Figure 2.5.

Correlated noise: the random variable $\tilde{\boldsymbol{\zeta}}$ is a centered Gaussian vector of variance

$$\boldsymbol{\Sigma}_3 = Id_2 \otimes P \text{Diag} \left(\frac{\ell^2}{2k}, \ell = 1, \dots, k \right) P'$$

where P is an arbitrary matrix in $\mathbf{SO}(k)$. Hence, in this case $\gamma_{\max}(k) = \frac{k}{4}$ and the level of noise increase with k , see Figure 2.6.

2.6.2. Description of the procedure

The estimation procedure follows the guidelines described in Section 2.4. We are testing the effect of the number J of observations and the number k of landmarks on the estimation of the parameters of interest of model (2.2). All the simulations are performed with $J = 10, 100, 500$ and

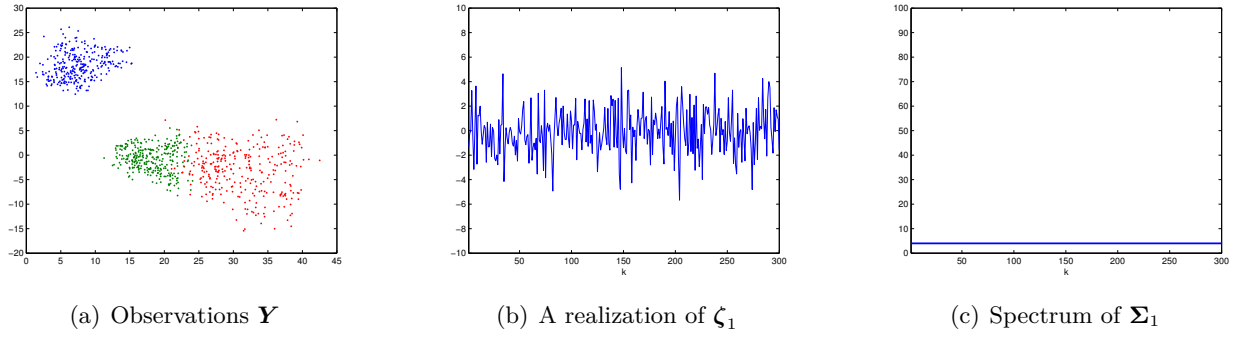


Figure 2.4: Example of data generated by model (2.2) with white noise.

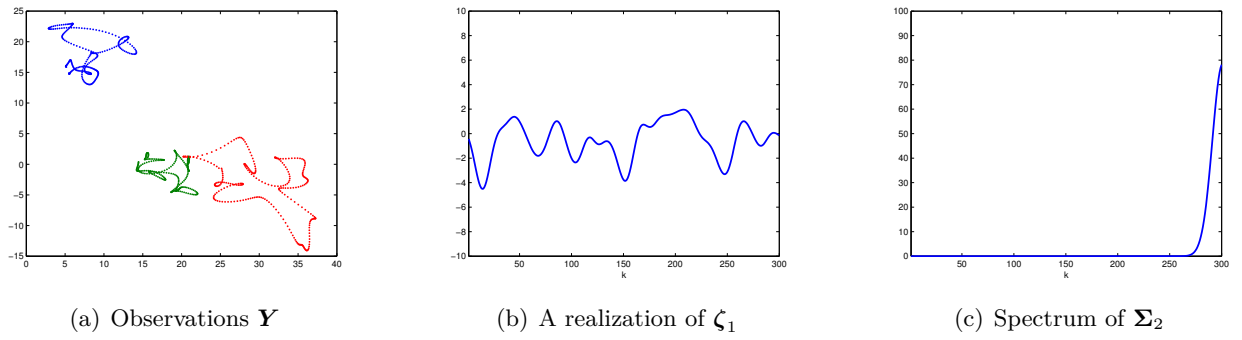


Figure 2.5: Example of data generated by model (2.2) with stationary noise.

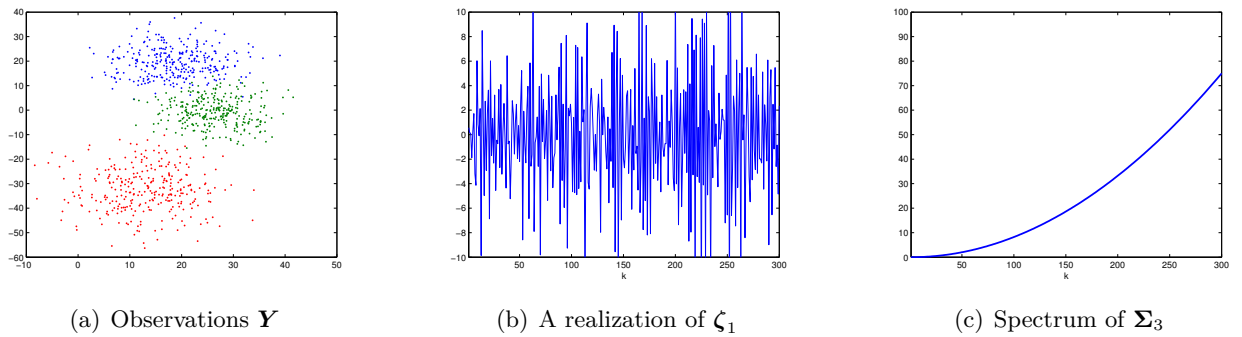


Figure 2.6: Example of observations generated by model (2.2) with the correlated noise.

$k = 20, 50, 100, 1000, 3000$ and for each combination of these two factors the simulations are performed with $M = 30$ repetitions of model (2.2).

Moreover, the estimation are done without and with the pre-smoothing step. In the former case we have $\lambda(k) = \frac{k}{2}$, that is, there is no reduction of the dimension. In the latter case, the smoothing parameter $\lambda(k)$ is fixed manually to ensure a proper reconstruction of the mean pattern f . Note that we need $\lambda > 7$ to get correct results and we took $\lambda_{20} = \lambda_{50} = \lambda_{100} = 7$, $\lambda_{1000} = 11$ and $\lambda_{3000} = 25$.

We use a quasi-Newton trust-region based algorithm to solve the optimization problems (2.18) and (2.19). The formula for the gradient is given in (2.42). All the computations are performed with *Matlab*.

2.6.3. Results: estimation of the mean pattern

For each of the 30 repetitions of model (2.2) with the possible values of k and J , we compute the quantities $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}_{\Theta_0}\|_{\mathbb{R}^{k \times 2}}^2$ where $\hat{\mathbf{f}}^\lambda$ corresponds to the smoothed Procrustes mean of the observations defined in (2.7) and, $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}_{\Theta_0}\|_{\mathbb{R}^{k \times 2}}^2$ where $\hat{\mathbf{f}}$ is the (non smoothed) Procrustes mean of the data. Recall that $\mathbf{f}_{\Theta_0}$ is defined by formula (2.24).

Boxplots of the results are given in Figures 2.7, 2.8 and 2.9 for the different kinds of error terms described in Section 2.6.1. In the figures, the abscissa represents the different values of the number k of landmarks and boxplots in red correspond to $J = 10$ observations, in green to $J = 100$ observations and in blue to $J = 500$ observations.

The estimation of the mean pattern with the white noise error term is given by Figure 2.7. In Figure 2.7(a), for a fixed k , the non-smoothed version $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2$ decreases when J increases. Moreover, the values of $\frac{1}{k} \|\hat{\mathbf{f}} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2$ remain stable when J remains fixed and k increases. Recall that this framework corresponds to the isotropic Gaussian noise described in [KM97]. The simulations seem to confirm their conclusions and show that in this framework the dimension k is not preponderant. At Figure 2.7(b), the smoothed version $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2$ decreases when J and k increase. The main difference with the non smoothed estimation is the convergence to 0 of $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2$ when J remains fixed and k increases.

At Figure 2.8, the results of the estimation of the mean pattern are plotted for the stationary noise term. Figure 2.7(a) shows us a similar behavior of the non-smoothed Procrustes mean but with non-decreasing values of $\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2$ when k increases and J remains fixed. At Figure 2.7(b), the smoothed Procrustes mean converges as k goes to infinity and the bigger J is the faster the convergence is.

The results of the estimations of the mean pattern with the correlated noise are presented Figure 2.9. The results that appear Figure 2.8(a) are quite different compared to those presented Figures 2.7(a) and 2.7(a). The estimation seems to be worst when k increases and J remains fixed. The reason is that the level of noise, measured by $\gamma_{\max}(k)$ is increasing with k . The smoothing step is efficient and the estimations presented Figure 2.8(b) have a similar behavior to those of Figures 2.7(b) and 2.7(b).

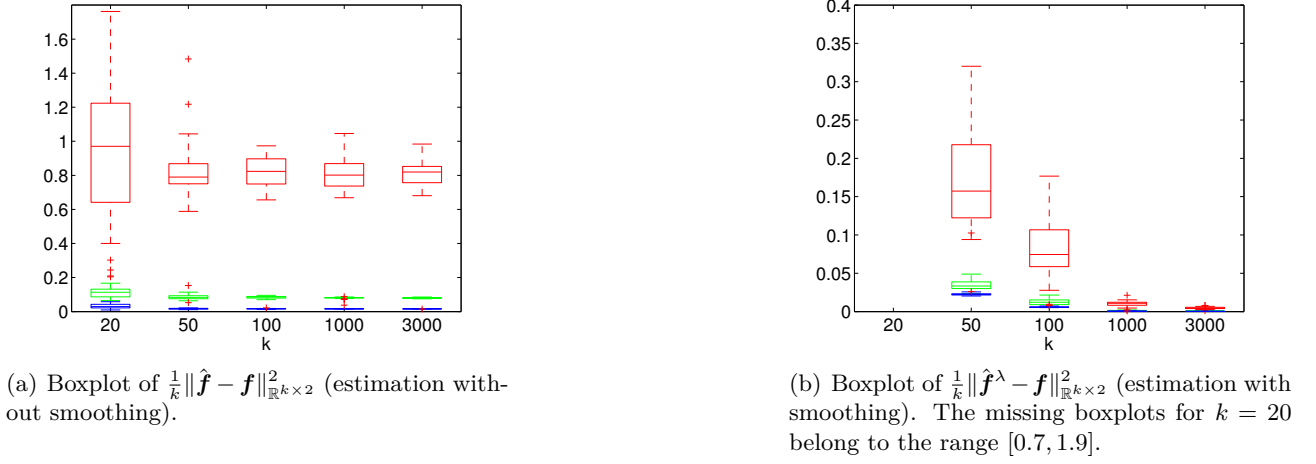


Figure 2.7: Estimation of the mean pattern with the white noise. Boxplot in red correspond to $J = 10$, in green to $J = 100$ and in blue to $J = 500$.

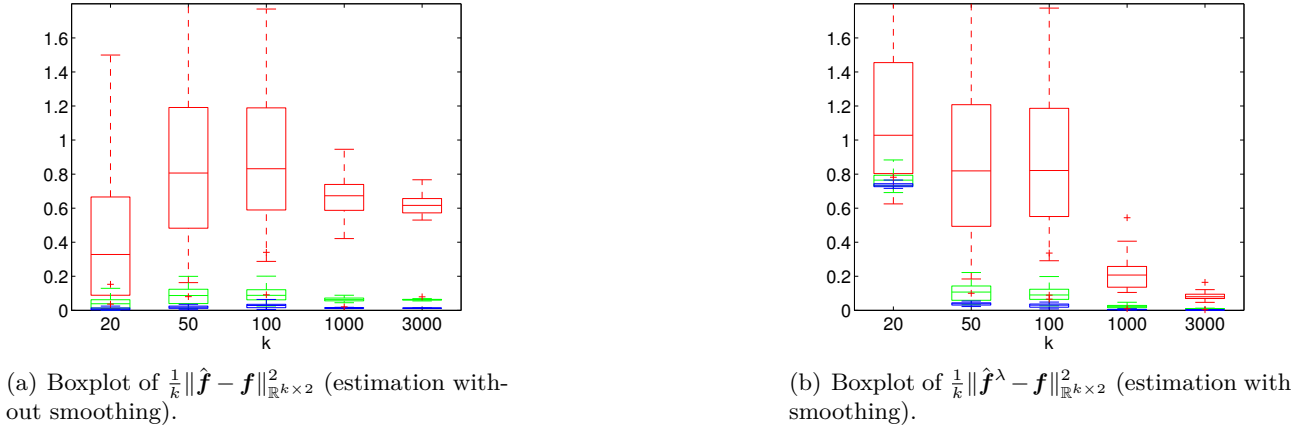


Figure 2.8: Estimation of the mean pattern with the stationary noise. Boxplot in red correspond to $J = 10$, in green to $J = 100$ and in blue to $J = 500$.

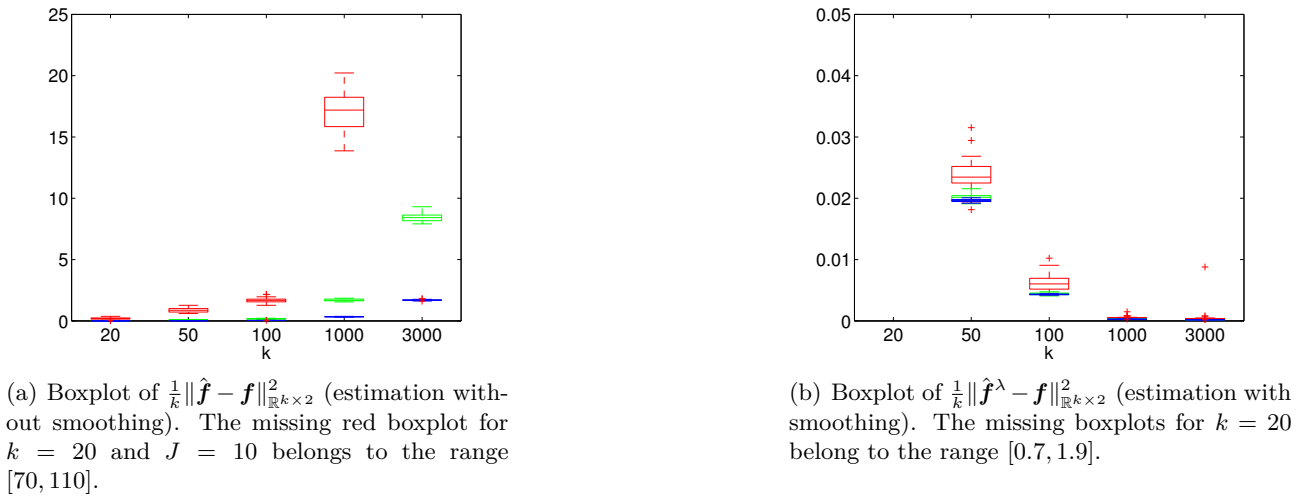


Figure 2.9: Estimation of the mean pattern with the correlated noise. Boxplot in red correspond to $J = 10$, in green to $J = 100$ and in blue to $J = 500$.

2.A. Proofs

2.A.1. Proof of Lemma 2.9

Using the decomposition (2.17), we obtain the following identity

$$M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) + \bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}). \quad (2.25)$$

Note that we have used the fact that the subspaces $\mathcal{V}_0^\lambda$ and $\bar{\mathcal{V}}$ are orthogonal and that $\bar{A}\mathbb{1}_k \otimes b_j = \mathbb{1}_k \otimes b_j$, for all $b_j \in \mathbb{R}^2$. Let us also introduce the notation $\bar{\mathbf{Y}}_j = \frac{1}{k} \sum_{\ell=1}^k [\mathbf{Y}_j]_\ell \in \mathbb{R}^2$, that is $\bar{A}\mathbf{Y}_j = \mathbb{1}_k \otimes \bar{\mathbf{Y}}_j \in \mathbb{R}^{k \times 2}$ is a degenerated configuration, and $\bar{\mathbf{Y}} = \frac{1}{J} \sum_{j=1}^J \bar{\mathbf{Y}}_j \in \mathbb{R}^2$. For a fixed $(\mathbf{a}, \boldsymbol{\alpha}) \in \mathbb{R}^J \times [-\pi, \pi]^J$, the functional $\mathbf{b} \mapsto \bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b})$ vanishes if and only if there exists a $b_0 \in \mathbb{R}^2$ such that $e^{-a_j}(\bar{A}\mathbf{Y}_j - \mathbb{1}_k \otimes b_j)R_{-\alpha_j} = \mathbb{1}_k \otimes b_0$ for all $j = 1, \dots, J$. Therefore, for this fixed $(\mathbf{a}, \boldsymbol{\alpha})$, there is a unique point $\mathbf{b} = \mathbf{b}(\mathbf{a}, \boldsymbol{\alpha}) := (\bar{\mathbf{Y}}_1 - e^{a_1}\bar{\mathbf{Y}}(e^{\mathbf{a}}R_{\boldsymbol{\alpha}})^{-1}R_{\alpha_1}, \dots, \bar{\mathbf{Y}}_J - e^{a_J}\bar{\mathbf{Y}}(e^{\mathbf{a}}R_{\boldsymbol{\alpha}})^{-1}R_{\alpha_J})$ with $\overline{e^{\mathbf{a}}R_{\boldsymbol{\alpha}}} = \frac{1}{J} \sum_{j=1}^J e^{a_j}R_{\alpha_j} \in \mathbb{R}^{2 \times 2}$ and which satisfies,

$$\mathbf{b}(\mathbf{a}, \boldsymbol{\alpha}) = \underset{\mathbf{b} \in \Theta_0^b}{\operatorname{argmin}} \bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b})$$

Thence, thanks to the decomposition (2.25) and the fact that $\bar{M}(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}(\mathbf{a}, \boldsymbol{\alpha})) = 0$, we have

$$\underset{(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) \in \Theta_0}{\operatorname{argmin}} M^\lambda(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \left(\underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}), \mathbf{b} \left(\underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) \right) \right),$$

and the claim is proved. $\square$

2.A.2. Proof of Theorem 2.11

For all $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$ and $(\mathbf{a}^*, \boldsymbol{\alpha}^*) \in [-A, A] \times [-\mathcal{A}, \mathcal{A}]$, we have the following inequality

$$\frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2 \leq \frac{2}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 + \frac{2}{J} \|(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2.$$

The proof of Theorem 2.11 is a direct consequence of Proposition 2.14 and Lemma 2.15 below which control the convergence in probability of the two terms in the right hand side in the preceding inequality.

Proposition 2.14. *Consider model (2.2) and suppose that Assumptions 2.1 and 2.2 hold and that Assumption 2.3 is verified with $A, \mathcal{A} < 0.1$. If $\lambda = \lambda(k) = k^{\frac{1}{2s+1}}$ then there exists a constant $C(L, s, A, \mathcal{A}, \mathbf{f}_0)$ such that for all $x > 0$*

$$\mathbb{P} \left(\frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \geq C(L, s, A, \mathcal{A}, \mathbf{f}_0) \left(F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)) \right) \right) \leq 2e^{-x},$$

where $V_1(k, J, x) = \gamma_{\max}(k)k^{\frac{-2s}{2s+1}} \left(1 + 2\sqrt{\frac{x}{J}k^{\frac{-1}{2s+1}}} + \frac{x}{J}k^{\frac{-1}{2s+1}} \right)$ and $F : \mathbb{R}^+ \rightarrow \mathbb{R}$, with $F(u) = u + \sqrt{u}$.

The proof of Proposition 2.14 is postponed to Section 2.A.5. The following lemma is a direct consequence of Bernstein's inequality for bounded random variables, see e.g. Proposition 2.9 in [Mas07].

Lemma 2.15. *Suppose that Assumption 2.3 holds and that the random variables (a_j^*, α_j^*) , $j = 1, \dots, J$ have zero expectation in $[-A, A] \times [-\mathcal{A}, \mathcal{A}]$. Then, for any $x > 0$, we have*

$$\mathbb{P}\left(\frac{1}{J}\|(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) - (\mathbf{a}^*, \boldsymbol{\alpha}^*)\|_{\mathbb{R}^{2J}}^2 \geq C(A, \mathcal{A})\left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2\right) \leq 4e^{-x},$$

where $C(A, \mathcal{A}) = 4 \max\{A^2, \mathcal{A}^2\}$. □

2.A.3. Proof of Theorem 2.12

The proof of Theorem 2.12 follows the same guideline as the proof of Theorem 2.11. Consider the inequality

$$\frac{1}{J}\|\hat{\mathbf{b}}^\lambda - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2 \leq \frac{2}{J}\|\hat{\mathbf{b}}^\lambda - \mathbf{b}_{\Theta_0}^*\|_{\mathbb{R}^{2J}}^2 + \frac{2}{J}\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2.$$

Theorem 2.12 is now a direct consequence of Proposition 2.16 and Lemma 2.17.

Proposition 2.16. *Under the hypothesis of Proposition 2.14, there exists a constant $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$ such that for all $x > 0$,*

$$\mathbb{P}\left(\frac{2}{J}\|\hat{\mathbf{b}}^\lambda - \mathbf{b}_{\Theta_0}^*\|_{\mathbb{R}^{2J}}^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f})\left(F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)) + V_2(k, J, x)\right)\right) \leq 5e^{-x},$$

where $V_1(k, J, x) = \gamma_{\max}(k)k^{\frac{-2s}{2s+1}}\left(1 + 2\sqrt{\frac{x}{J}k^{\frac{-1}{2s+1}}} + \frac{x}{J}k^{\frac{-1}{2s+1}}\right)$, $V_2(k, J, x) = \frac{\gamma_{\max}(k)}{k}(1 + \sqrt{2\frac{x}{J}} + \frac{x}{J})$ and $F(u) = u + \sqrt{u}$, $u \geq 0$.

The proof of Proposition 2.16 is postponed to Section 2.A.6.

Lemma 2.17. *Suppose that Assumption 2.3 holds with $\mathcal{A} < \frac{\pi}{4}$ and that the random variables $(a_j^*, \alpha_j^*, b_j^*)$, $j = 1, \dots, J$ have zero expectation in $\Theta \subset \mathbb{R}^{4J}$. For any $x > 0$, we have*

$$\mathbb{P}\left(\frac{2}{J}\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2 \geq C(A, \mathcal{A}, B)\left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2\right) \leq 4e^{-x},$$

where $C(A, \mathcal{A}, B) = 8B^2e^{4A}(\cos \mathcal{A} - \sin \mathcal{A})^{-2}$.

Proof. This result is a consequence of the Bernstein's inequality for bounded random variable. We have

$$\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}} = \sqrt{\sum_{j=1}^J \|e^{a_j^*} \bar{\mathbf{b}}^* (\overline{e^{\mathbf{a}} R_{\boldsymbol{\alpha}}})^{-1} R_{\alpha_j^*}\|_{\mathbb{R}^2}^2} \leq e^A \sqrt{J} \|\bar{\mathbf{b}}^* (\overline{e^{\mathbf{a}} R_{\boldsymbol{\alpha}}})^{-1}\|_{\mathbb{R}^2},$$

where $\overline{e^{\mathbf{a}} R_{\boldsymbol{\alpha}}} = \frac{1}{J} \sum_{j=1}^J e^{a_j^*} R_{\alpha_j^*}$ is an invertible 2×2 matrix whose smallest eigenvalue is greater than $e^{-A}(\cos \mathcal{A} - \sin \mathcal{A}) > 0$ as $\mathcal{A} < \frac{\pi}{4}$. To see this, remark that the eigenvalues of $\overline{e^{\mathbf{a}} R_{\boldsymbol{\alpha}}}$ are $\frac{1}{J} \sum_{j=1}^J e^{a_j^*} (\cos \alpha_j^* \pm i \sin \alpha_j^*)$ and we have $|\frac{1}{J} \sum_{j=1}^J e^{a_j^*} (\cos \alpha_j^* \pm i \sin \alpha_j^*)| \geq e^{-A}(\cos \mathcal{A} - \sin \mathcal{A}) > 0$. We now have

$$\frac{1}{\sqrt{J}}\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}} \leq C(A, \mathcal{A})\|\bar{\mathbf{b}}^*\|_{\mathbb{R}^2},$$

where $C(A, \mathcal{A}) = e^{2A}(\cos \mathcal{A} - \sin \mathcal{A})^{-1}$. Finally, for all $u > 0$ we have $\mathbb{P}(\frac{1}{\sqrt{J}}\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}} \geq u) \leq \mathbb{P}(C(A, \mathcal{A})\|\bar{\mathbf{b}}^*\|_{\mathbb{R}^2} \geq u)$ and a Bernstein type inequality (see e.g. Proposition 2.9 in [Mas07]) gives us $\mathbb{P}(\|\bar{\mathbf{b}}^*\|_{\mathbb{R}^2} \geq 2B(\sqrt{\frac{2x}{J}} + \frac{x}{3J})) \leq 4e^{-x}$ which yields

$$\mathbb{P}\left(\frac{1}{J}\|\mathbf{b}_{\Theta_0}^* - \mathbf{b}^*\|_{\mathbb{R}^{2J}}^2 \geq C(A, \mathcal{A}, B)\left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2\right) \leq 4e^{-x},$$

where $C(A, \mathcal{A}, B) = 4B^2e^{4A}(\cos \mathcal{A} - \sin \mathcal{A})^{-2}$. □

2.A.4. Proof of Theorem 2.13

Recall the notations introduced Section 2.2.1: for $\hat{g}_j^\lambda = (\hat{a}_j^\lambda, \hat{\alpha}_j^\lambda, \hat{b}_j^\lambda)$ we have $\hat{g}_j^\lambda \cdot \mathbf{f} = e^{\hat{a}_j^\lambda} \mathbf{f} R_{\hat{\alpha}_j^\lambda} + \hat{b}_j^\lambda$. Then, we have

$$\begin{aligned} \frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 &= \frac{1}{k} \left\| \frac{1}{J} \sum_{j=1}^J (\hat{g}_j^\lambda)^{-1} \cdot (\mathbf{A}^\lambda \mathbf{Y}_j) - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2 \\ &\leq \frac{2}{kJ} \sum_{j=1}^J \left\| \mathbf{A}^\lambda \left((\hat{g}_j^\lambda)^{-1} \cdot \mathbf{Y}_j - \mathbf{f} \right) \right\|_{\mathbb{R}^{k \times 2}}^2 + \frac{2}{k} \left\| \mathbf{A}^\lambda \mathbf{f} - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2 \\ &\leq \underbrace{\frac{4}{kJ} \sum_{j=1}^J \left\| \mathbf{A}^\lambda \left((\hat{g}_j^\lambda)^{-1} \cdot g_j^* \cdot \mathbf{f} - \mathbf{f} \right) \right\|_{\mathbb{R}^{k \times 2}}^2}_{\mathbf{V}} + \underbrace{\left\| e^{a_j^* - \hat{a}_j^\lambda} \mathbf{A}^\lambda \zeta_j R_{\alpha_j^* - \hat{\alpha}_j^\lambda} \right\|_{\mathbb{R}^{k \times 2}}^2 + \frac{2}{k} \left\| \mathbf{A}^\lambda \mathbf{f} - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2}_{\mathbf{B}}. \end{aligned}$$

The rest of the proof is devoted to control the terms $\mathbf{B}$ and $\mathbf{V}$. The term $\mathbf{B}$ is controlled by the bias of the low pass filter. According to Lemma 2.21, if Assumption 2.1 holds and by choosing the optimal frequency cutoff $\lambda = \lambda(k) = k^{\frac{1}{2s+1}}$, there exists a constant $C(L, s) > 0$ such that

$$\mathbf{B} = \frac{2}{k} \left\| \mathbf{A}^\lambda \mathbf{f} - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2 \leq C(L, s) k^{-\frac{2s}{2s+1}}. \quad (2.26)$$

The term $\mathbf{V}$ contains two expressions. To bound the first one we use Bessel's inequality and Lemma 2.24. More precisely, we have for all $j = 1, \dots, J$

$$\begin{aligned} \frac{1}{k} \left\| \mathbf{A}^\lambda ((g_j^\lambda)^{-1} \cdot g_j^* \cdot \mathbf{f} - \mathbf{f}) \right\|_{\mathbb{R}^{k \times 2}}^2 &\leq \frac{1}{k} \left\| (g_j^\lambda)^{-1} \cdot g_j^* \cdot \mathbf{f} - \mathbf{f} \right\|_{\mathbb{R}^{k \times 2}}^2 \\ &\leq C(A, \mathbf{f}) \left\| (a_j^* - \hat{a}_j^\lambda, \alpha_j^* - \hat{\alpha}_j^\lambda, e^{\hat{a}_j^\lambda} (b_j^* - \hat{b}_j^\lambda) R_{-\hat{\alpha}_j^\lambda}) \right\|_{\mathbb{R}^4}^2 \\ &\leq C(A, \mathbf{f}) \left\| (a_j^* - \hat{a}_j^\lambda, \alpha_j^* - \hat{\alpha}_j^\lambda) \right\|_{\mathbb{R}^2}^2 + C(A, \mathbf{f}) \|b_j^* - \hat{b}_j^\lambda\|_{\mathbb{R}^2}^2 \end{aligned}$$

We can now use Theorem 2.11 and 2.12 to derive the following concentration inequality,

$$\begin{aligned} \mathbb{P} \left(\frac{4}{kJ} \sum_{j=1}^J \left\| \mathbf{A}^\lambda ((g_j^\lambda)^{-1} \cdot g_j^* \cdot \mathbf{f} - \mathbf{f}) \right\|_{\mathbb{R}^{k \times 2}}^2 \right. \\ \left. \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) \left(A_1(k, J, x) + A_3(k, J, x) + A_2(J, x) \right) \right) \leq 13e^{-x} \end{aligned} \quad (2.27)$$

where $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$ is a constant independent of k and J and A_1, A_2, A_3 are defined in the statement of Theorem 2.11 and 2.12. The second term contained in $\mathbf{V}$ is treated by equation (2.34) below. Hence, formulas (2.27) and (2.34) yield

$$\mathbb{P} \left(\mathbf{V} \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) \left(A_1(k, J, x) + A_3(k, J, x) + A_2(J, x) \right) \right) \leq 14e^{-x}, \quad (2.28)$$

for some constant $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$. Putting together equations (2.26) and (2.28) gives

$$\mathbb{P} \left(\frac{1}{k} \|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) \left(A_1(k, J, x) + A_3(k, J, x) + A_2(J, x) \right) \right) \leq 14e^{-x}.$$

for some constant $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$. The proof of Theorem 2.13 is completed. $\square$

2.A.5. Proof of Proposition 2.14

The mean pattern $\mathbf{f}$ can be decomposed as $\mathbf{f} = \bar{\mathbf{f}} + \mathbf{f}_0 \in \mathbf{1}_{k \times 2} \oplus \mathbf{1}_{k \times 2}^\perp$. Then, $\mathbf{f}_0$ is the centered version of $\mathbf{f}$ and we can consider the criterion,

$$D_0(\mathbf{a}, \boldsymbol{\alpha}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} \mathbf{f}_0 R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2.$$

We now have,

$$(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) = \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}}{\operatorname{argmin}} M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) \quad \text{and} \quad (\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*) = \underset{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}}{\operatorname{argmin}} D_0(\mathbf{a}, \boldsymbol{\alpha}).$$

Then, the convergence of $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$ to $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$ is guaranteed if $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$ is uniquely defined and if there is a uniform convergence in probability of M_0^λ to D_0 , see e.g. [vdV98]. This is the aim of Lemmas 2.18 and 2.19 below.

Lemma 2.18. *Let $\mathbf{f}$ be a non-degenerated configuration in $\mathbb{R}^{k \times 2}$, i.e. $\mathbf{f} \notin \mathbf{1}_{k \times 2}$. Then, the argmin of D_0 on $\boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}$ is unique and denoted by $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*) = (a_j - \bar{\mathbf{a}}^*, \alpha_j - \bar{\boldsymbol{\alpha}}^*)$, where $\bar{\mathbf{a}}^* = \frac{1}{J} \sum_{j=1}^J a_j^*$, $\bar{\boldsymbol{\alpha}}^* = \frac{1}{J} \sum_{j=1}^J \alpha_j^*$.*

Proof. As $\mathbf{f}$ is a non-degenerated configuration, we have $\mathbf{f}_0 \neq 0$. Thus, the stabilizer of $\mathbf{f}$ is reduced to the identity, see Section 2.2.1. Then $D_0(\mathbf{a}, \boldsymbol{\alpha}) = 0$ if and only if there exists $(a_0, \alpha_0) \in \mathbb{R}^2$ such that $(\mathbf{a}, \boldsymbol{\alpha}) = (\mathbf{a}^*, \boldsymbol{\alpha}^*) * (a_0, \alpha_0) = (a_1^* + a_0, \alpha_1^* + \alpha_0, \dots, a_J^* + a_0, \alpha_J^* + \alpha_0)$. By choosing $(a_0, \alpha_0) = (-\bar{\mathbf{a}}^*, -\bar{\boldsymbol{\alpha}}^*)$ we have $\sum_{j=1}^J (a_j^*, \alpha_j^*) * (a_0, \alpha_0) = 0$. That is $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*) = (\mathbf{a}^*, \boldsymbol{\alpha}^*) * (a_0, \alpha_0) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}$. $\square$

We now show the uniform convergence in probability,

Lemma 2.19. *Suppose that Assumptions 2.1, 2.2 and 2.3 hold and let $F : \mathbb{R} \rightarrow \mathbb{R}$, with $F(u) = u + \sqrt{u}$. For any $x > 0$ we have*

$$\mathbb{P} \left(\sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}} |M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}, \boldsymbol{\alpha})| \geq C(L, s, A, \mathbf{f}_0) \left(F(k^{-\frac{2s}{2s+1}}) + F(V(k, J, x)) \right) \right) \leq 2e^{-x}$$

where $C(L, s, A, \mathbf{f}_0) = e^{2A} \max \left\{ \frac{2}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, \frac{\sqrt{2}e^A}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, 2 \right\}$ and $V(k, J, x) = 3\gamma_{\max}(k)k^{-\frac{2s}{2s+1}}(1 + F(\frac{2x}{Jk}))$.

Proof. Let us write the following decomposition,

$$M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} \mathbf{f}_0 R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2 \quad (2.29)$$

$$+ \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2 \quad (2.30)$$

$$+ \frac{2}{Jk} \sum_{j=1}^J \left\langle e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} \mathbf{f}_0 R_{\alpha_{j'}^* - \alpha_{j'}}, \right. \\ \left. e^{a_j^* - a_j} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_{j'}^* - \alpha_{j'}} \right\rangle_{\mathbb{R}^{k \times 2}} \quad (2.31)$$

$$+ \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} A_0^\lambda \boldsymbol{\zeta}_j R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} A_0^\lambda \boldsymbol{\zeta}_{j'} R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2 \quad (2.32)$$

$$+ \frac{2}{Jk} \sum_{j=1}^J \left\langle e^{a_j^* - a_j} A_0^\lambda \mathbf{f} R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} A_0^\lambda \mathbf{f} R_{\alpha_{j'}^* - \alpha_{j'}}, \right. \\ \left. e^{a_j^* - a_j} A_0^\lambda \boldsymbol{\zeta}_j R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} A_0^\lambda \boldsymbol{\zeta}_{j'} R_{\alpha_{j'}^* - \alpha_{j'}} \right\rangle_{\mathbb{R}^{k \times 2}} \quad (2.33)$$

Then, criterion M_0^λ is viewed as a perturbed version of the criterion $D_0(\mathbf{a}, \boldsymbol{\alpha}) = (2.29)$,

$$M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = D_0(\mathbf{a}, \boldsymbol{\alpha}) + \mathbf{B}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) + \mathbf{V}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha})$$

where the bias term is $\mathbf{B}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = (2.30) + (2.31)$ and the variance term is $\mathbf{V}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) = (2.32) + (2.33)$.

The bias term. We have for any $(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{a, \alpha}$,

$$(2.30) = \frac{1}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} \right\|_{\mathbb{R}^{k \times 2}}^2 - \left\| \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2 \\ \leq e^{2A} \frac{1}{k} \left\| A_0^\lambda \mathbf{f} - \mathbf{f}_0 \right\|_{\mathbb{R}^{k \times 2}}^2.$$

A double application of Cauchy-Schwarz inequality implies that,

$$|(2.31)| \leq \frac{2}{Jk} \sum_{j=1}^J \left\| e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} \mathbf{f}_0 R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}} \\ \left\| e^{a_j^* - a_j} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} - \frac{1}{J} \sum_{j'=1}^J e^{a_{j'}^* - a_{j'}} (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_{j'}^* - \alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}} \\ \leq \frac{2}{k} \left(\frac{1}{J} \sum_{j=1}^J e^{2A} \left\| \mathbf{f}_0 R_{\alpha_j^* - \alpha_j} \right\|_{\mathbb{R}^{k \times 2}}^2 \right)^{\frac{1}{2}} \left(\frac{1}{J} \sum_{j=1}^J e^{2A} \left\| (A_0^\lambda \mathbf{f} - \mathbf{f}_0) R_{\alpha_j^* - \alpha_j} \right\|_{\mathbb{R}^{k \times 2}}^2 \right)^{\frac{1}{2}} \\ = 2e^{2A} \frac{1}{k} \left\| \mathbf{f}_0 \right\|_{\mathbb{R}^{k \times 2}} \left\| A_0^\lambda \mathbf{f} - \mathbf{f}_0 \right\|_{\mathbb{R}^{k \times 2}}.$$

Finally, by using Lemma 2.21 we have,

$$\sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{a, \alpha}} |\mathbf{B}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha})| \leq C_1(L, s, A, \mathbf{f}) \left(k^{-\frac{2s}{2s+1}} + k^{-\frac{s}{2s+1}} \right).$$

where $C_1(L, s, A, \mathbf{f}) = C(L, s)e^{2A} \max \left\{ 2\frac{1}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, 1 \right\}$.

The variance term. First, the term (2.33) is by the Cauchy-Schwarz inequality controlled by $2e^{2A} \frac{1}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}} \sqrt{(2.32)}$. The term (2.32) is bounded from above by $\frac{1}{Jk} \sum_{j=1}^J e^{2A} \|A_0^\lambda \boldsymbol{\zeta}_j\|_{\mathbb{R}^{k \times 2}}^2$. To derive an upper bound in probability, note that we have the following equality in law,

$$\sum_{j=1}^J \|A_0^\lambda \boldsymbol{\zeta}_j\|_{\mathbb{R}^{k \times 2}}^2 = \boldsymbol{\xi}' \mathbf{B} \boldsymbol{\xi},$$

with $\mathbf{B} = [Id_J \otimes \boldsymbol{\Sigma}^{\frac{1}{2}}] [Id_{2J} \otimes (A_0^\lambda)' A_0^\lambda] [Id_J \otimes \boldsymbol{\Sigma}^{\frac{1}{2}}] \in \mathbb{R}^{2Jk \times 2Jk}$ and $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{2Jk})'$ is a centered Gaussian vector of variance Id_{2Jk} . Note that, without loss of generality, we suppose that $\mathbf{B}$ is a diagonal matrix, *i.e.* if needed we can make an orthogonal change of basis. In what follows, $\gamma_{\max}(A)$ is the largest eigenvalue of a matrix A . Recall that $A_0^\lambda \in \mathbb{R}^{n \times n}$ is a projection matrix of rank 2λ . This implies that the largest eigenvalue $\gamma_{\max}((A_0^\lambda)' A_0^\lambda)$ of $(A_0^\lambda)' A_0^\lambda \in \mathbb{R}^{n \times n}$ is less or equal to 1 and easy algebra yields $\text{tr}((A_0^\lambda)' A_0^\lambda) = 2\lambda$. We are now able to use the concentration inequality for quadratic form of multivariate Gaussian random variables given by Lemma 1 in [LM00]. In order to apply this result we need the following estimates: $\gamma_{\max}(\mathbf{B}) \leq \gamma_{\max}(\boldsymbol{\Sigma}) = \gamma_{\max}(k)$ and $\sqrt{\text{tr}(\mathbf{B}^2)} \leq \sqrt{\text{tr}(\mathbf{B})\gamma_{\max}(\mathbf{B})} \leq \sqrt{4J\lambda\gamma_{\max}(k)}$. Thence, we have for all $x > 0$,

$$\mathbb{P}(\boldsymbol{\xi}' \mathbf{B} \boldsymbol{\xi} \geq 2J\gamma_{\max}(k)\lambda + 4\gamma_{\max}(k)\sqrt{J\lambda x} + 2\gamma_{\max}(k)x) \leq e^{-x},$$

which yields together with the fact that $\lambda = k^{\frac{1}{2s+1}}$,

$$\mathbb{P}\left(\frac{1}{Jk} \sum_{j=1}^J e^{2A} \|A_0^\lambda \boldsymbol{\zeta}_j\|_{\mathbb{R}^{k \times 2}}^2 \geq 2e^{2A} \gamma_{\max}(k) k^{-\frac{2s}{2s+1}} \left(1 + 2\sqrt{\frac{x}{J} k^{\frac{-1}{2s+1}}} + \frac{x}{J} k^{\frac{-1}{2s+1}}\right)\right) \leq e^{-x}. \quad (2.34)$$

Hence we have,

$$\begin{aligned} & \mathbb{P}\left(\sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}^{\mathbf{a}, \boldsymbol{\alpha}}} |\mathbf{V}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) + \mathbf{B}_0^\lambda(\mathbf{a}, \boldsymbol{\alpha})| \right. \\ & \quad \left. \geq C(L, s, A, \mathbf{f}_0) \left(k^{-\frac{2s}{2s+1}} + k^{-\frac{s}{2s+1}} + V_1(k, J, x) + \sqrt{V_1(k, J, x)}\right)\right) \leq 2e^{-x}, \end{aligned}$$

where $C(L, s, A, \mathbf{f}_0) = e^{2A} \max \left\{ C(L, s) \frac{2}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, C(L, s), \frac{\sqrt{2}e^A}{\sqrt{k}} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}, 2 \right\}$ does not depend on k and J and $V_1(k, J, x) = \gamma_{\max}(k) k^{\frac{-2s}{2s+1}} \left(1 + 2\sqrt{\frac{x}{J} k^{\frac{-1}{2s+1}}} + \frac{x}{J} k^{\frac{-1}{2s+1}}\right)$. $\square$

For a fixed $J \in \mathbb{N}$, the convergence of the M-estimator $(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda)$ to $(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)$ when $k \rightarrow \infty$ is guaranteed by Lemma 2.18 and 2.19, see e.g. [vdV98]. Nevertheless, we are able to give a rate of convergence and non-asymptotic bounds in k and J by using the classical inequality,

$$|D_0(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - D_0(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)| \leq 2 \sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_0^{\mathbf{a}, \boldsymbol{\alpha}}} |D_0(\mathbf{a}, \boldsymbol{\alpha}) - M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha})|.$$

This, together with Lemma 2.20 below will prove Proposition 2.14.

Lemma 2.20. *Assume that $A, \mathcal{A} < 0.1$. There exists a constant $C(A, \mathcal{A}) > 0$ independent of J such that for all $(\mathbf{a}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}^{\mathbf{a}, \boldsymbol{\alpha}}$ we have*

$$|D_0(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)| \geq C(A, \mathcal{A}, \mathbf{f}_0) \frac{1}{J} \|(\mathbf{a} - \mathbf{a}_{\boldsymbol{\Theta}_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\boldsymbol{\Theta}_0}^*)\|_{\mathbb{R}^{2J}}^2,$$

where $C(A, \mathcal{A}, \mathbf{f}_0) = C(A, \mathcal{A}) \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2$.

Proof. By definition, given a $(\mathbf{a}^*, \boldsymbol{\alpha}^*) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^J$, the point $(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)$ is the unique minimum of D_0 on $\Theta_0^{\mathbf{a}, \boldsymbol{\alpha}} = [-A, A]^J \cap \mathbb{1}_J^\perp \times [-\mathcal{A}, \mathcal{A}]^J \cap \mathbb{1}_J^\perp$. Then, for all $(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$, there exists a $\mathbf{c} = \mathbf{c}(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$ such that the Taylor expansion of D_0 at $(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)$ can be written,

$$\begin{aligned} D_0(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) &= \frac{1}{2}(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)' [\nabla^2 D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)] (\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*) \\ &\quad + \frac{1}{6} [\nabla^3 D_0(\mathbf{c})] (\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*). \end{aligned}$$

Let $\delta = \max\{A, \mathcal{A}\}$. This, together with Lemma 2.22 and 2.23 imply that

$$\begin{aligned} D_0(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) &\geq \frac{1}{2}(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)' [\nabla^2 D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)] (\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*) \\ &\quad - \delta \frac{40}{6} e^{2A} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{Jk} \|(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \\ &\geq \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{Jk} \|(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \left(e^{-2A} - \delta \frac{40}{6} e^{2A} \right). \end{aligned}$$

Hence, one can choose $\delta > 0$ sufficiently small such that $(e^{-2A} - \delta \frac{40}{6} e^{2A})$ is strictly positive for all J and k . For example, we have $(e^{-2\delta} - \delta \frac{40}{6} e^{2\delta}) > 0$, if $\delta < 0.1$. Then, using such a δ it follows that for all $(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$,

$$|D_0(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)| \geq C(A, \mathcal{A}) \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{J} \|(\mathbf{a} - \mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha} - \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2. \quad \square$$

The proof of Proposition 2.14 is almost done. Remark that Lemma 2.20 ensures that for all $u \geq 0$ we have $\mathbb{P}(\frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \geq u) \leq \mathbb{P}(\frac{2}{C(A, \mathcal{A}, \mathbf{f}_0)} \sup_{(\mathbf{a}, \boldsymbol{\alpha}) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}} |M_0^\lambda(\mathbf{a}, \boldsymbol{\alpha}) - D_0(\mathbf{a}, \boldsymbol{\alpha})| \geq u)$. Lemma 2.19 ensures that there is a constant $C(L, s, A, \mathcal{A}, \mathbf{f}_0)$ such that for all $x > 0$,

$$\mathbb{P}\left(\frac{1}{J} \|(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda) - (\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)\|_{\mathbb{R}^{2J}}^2 \geq C(L, s, A, \mathcal{A}, \mathbf{f}_0) (F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)))\right) \leq 2e^{-x}.$$

where $V_1(k, J, x) = \gamma_{\max}(k) k^{-\frac{2s}{2s+1}} \left(1 + \sqrt{2 \frac{x}{Jk}} + 2 \frac{x}{Jk}\right)$ and $F: \mathbb{R} \rightarrow \mathbb{R}$, with $F(u) = u + \sqrt{u}$. $\square$

2.A.6. Proof of Proposition 2.16

First of all remark that, thanks to formulas (2.15) and (2.20), we have explicit expressions of $\mathbf{b}_{\Theta_0}^* = \operatorname{argmin}_{\mathbf{b} \in \Theta_0^b} \bar{D}(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b})$ and of $\hat{\mathbf{b}}^\lambda = \operatorname{argmin}_{\mathbf{b} \in \Theta_0^b} \bar{M}(\hat{\mathbf{a}}^\lambda, \hat{\boldsymbol{\alpha}}^\lambda, \mathbf{b})$. Indeed, we have,

$$\mathbf{b}_{\Theta_0}^* = (\mathbf{b}_1^* - e^{a_1^*} \bar{\mathbf{b}}^* (\overline{e^{a^*} R_{\alpha^*}})^{-1} R_{\alpha_1^*}, \dots, \mathbf{b}_J^* - e^{a_J^*} \bar{\mathbf{b}}^* (\overline{e^{a^*} R_{\alpha^*}})^{-1} R_{\alpha_J^*}),$$

where $\overline{e^{a^*} R_{\alpha^*}} = \frac{1}{J} \sum_{j=1}^J e^{a_j^*} R_{\alpha_j^*} \in \mathbb{R}^{2 \times 2}$, $\bar{\mathbf{b}}^* = \frac{1}{J} \sum_{j=1}^J \mathbf{b}_j^* \in \mathbb{R}^2$ and

$$\hat{\mathbf{b}}^\lambda = (\bar{\mathbf{Y}}_1 - e^{\hat{a}_1^\lambda} \bar{\mathbf{Y}} (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1} R_{\hat{\alpha}_1^\lambda}, \dots, \bar{\mathbf{Y}}_J - e^{\hat{a}_J^\lambda} \bar{\mathbf{Y}} (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1} R_{\hat{\alpha}_J^\lambda})$$

where $\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}} = \frac{1}{J} \sum_{j=1}^J e^{\hat{a}_j^\lambda} R_{\hat{\alpha}_j^\lambda} \in \mathbb{R}^{2 \times 2}$, $\bar{\mathbf{Y}}_j = \frac{1}{k} \sum_{\ell=1}^k [\mathbf{Y}_j]_\ell = e^{a_j^*} \bar{\mathbf{f}} R_{\alpha_j^*} + \mathbf{b}_j^* + e^{a_j^*} \bar{\boldsymbol{\zeta}}_j R_{\alpha_j^*} \in \mathbb{R}^2$ where $\bar{\mathbf{f}} = \frac{1}{k} \sum_{\ell=1}^k \mathbf{f}_\ell \in \mathbb{R}^2$, $\bar{\boldsymbol{\zeta}}_j = \frac{1}{k} \sum_{\ell=1}^k [\boldsymbol{\zeta}_j]_\ell \in \mathbb{R}^2$ and $\bar{\mathbf{Y}} = \frac{1}{J} \sum_{j=1}^J \bar{\mathbf{Y}}_j = \bar{\mathbf{f}} (\overline{e^{a^*} R_{\alpha^*}}) + \bar{\mathbf{b}}^* + \frac{1}{J} \sum_{j=1}^J e^{a_j^*} \bar{\boldsymbol{\zeta}}_j R_{\alpha_j^*} \in \mathbb{R}^2$. Thence, we have

$$\frac{1}{J} \|\mathbf{b}_{\Theta_0}^* - \hat{\mathbf{b}}^\lambda\|_{\mathbb{R}^{2J}}^2 \leq \frac{1}{J} \sum_{j=1}^J \|\bar{\mathbf{f}} (e^{a_j^*} R_{\alpha_j^*} - e^{\hat{a}_j^\lambda} (\overline{e^{a^*} R_{\alpha^*}}) (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1} R_{\hat{\alpha}_j^\lambda})\|_{\mathbb{R}^2}^2 \quad (2.35)$$

$$+ \|\bar{\mathbf{b}}^* (e^{a_j^*} (\overline{e^{a^*} R_{\alpha^*}})^{-1} R_{\alpha_j^*} - e^{\hat{a}_j^\lambda} (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1} R_{\hat{\alpha}_j^\lambda})\|_{\mathbb{R}^2}^2 \quad (2.36)$$

$$+ \|e^{a_j^*} \bar{\boldsymbol{\zeta}}_j R_{\alpha_j^*}\|_{\mathbb{R}^2}^2 + \|e^{\hat{a}_j^\lambda} (\overline{e^{a^*} R_{\alpha^*}}) (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1} R_{\hat{\alpha}_j^\lambda}\|_{\mathbb{R}^2}^2 \quad (2.37)$$

where $(\overline{e^{a^*} \zeta R_{\alpha^*}}) = \frac{1}{J} \sum_{j=1}^J e^{a_j^*} \bar{\zeta}_j R_{\alpha_j^*} \in \mathbb{R}^2$. The rest of the proof is devoted to the control of the terms (2.35), (2.36) and (2.37).

In this section, we denote by $\|\cdot\|_{op}$ the operator norm of a 2×2 matrix, *i.e.* $\|A\|_{op} = |\gamma_{\max}(A)|$ where $\gamma_{\max}(A)$ denotes the largest eigenvalue of a matrix $A \in \mathbb{R}^{2 \times 2}$. Note by the way that the eigenvalues of the matrix $\frac{1}{J} \sum_{j=1}^J e^{a_j} R_{\alpha_j} \in \mathbb{R}^{2 \times 2}$ are $\frac{1}{J} \sum_{j=1}^J e^{a_j} (\cos(\alpha_j) \pm i \sin(\alpha_j))$ for any J and $(a_1, \dots, a_J, \alpha_1, \dots, \alpha_J) \in [-A, A]^J \times [-\mathcal{A}, \mathcal{A}]^J$. It yields $\|e^{-a_j} R_{-\alpha_j}\|_{op} \leq e^A$ and $\|(\overline{e^a R_{\alpha}})^{-1}\|_{op} \leq e^A (\cos(\mathcal{A}) - \sin(\mathcal{A}))^{-1}$ which is a positive real number since by hypothesis we have $\mathcal{A} < \frac{\pi}{4}$. We are now able to derive an upper bound for (2.35),

$$\begin{aligned} (2.35) &\leq \|\bar{\mathbf{f}}\|_{\mathbb{R}^2}^2 \frac{1}{J} \sum_{j=1}^J \|e^{a_j^*} R_{\alpha_j^*} - e^{\hat{a}_j^\lambda} (\overline{e^{a^*} R_{\alpha^*}}) (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1} R_{\hat{\alpha}_j^\lambda}\|_{op}^2 \\ &\leq e^{2A} \|\bar{\mathbf{f}}\|_{\mathbb{R}^2}^2 \frac{1}{J} \sum_{j=1}^J \|e^{a_j^*} R_{\alpha_j^*} (e^{-\hat{a}_j^\lambda} R_{-\hat{\alpha}_j^\lambda}) - (\overline{e^{a^*} R_{\alpha^*}}) (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1}\|_{op}^2 \\ &\leq 2e^{2A} \|\bar{\mathbf{f}}\|_{\mathbb{R}^2}^2 \frac{1}{J} \sum_{j=1}^J \|e^{-\hat{a}_j^\lambda} R_{-\hat{\alpha}_j^\lambda}\|_{op}^2 \|e^{a_j^*} R_{\alpha_j^*} - e^{\hat{a}_j^\lambda} R_{\hat{\alpha}_j^\lambda}\|_{op}^2 \\ &\quad + 2e^{2A} \|\bar{\mathbf{f}}\|_{\mathbb{R}^2}^2 \|(\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})^{-1}\|_{op}^2 \|(\overline{e^{a^*} R_{\alpha^*}}) - (\overline{e^{\hat{a}^\lambda} R_{\hat{\alpha}^\lambda}})\|_{op}^2 \end{aligned}$$

It is now easy to show that there exists a constant $C(A, \mathcal{A}, \mathbf{f}) > 0$ independent of k and J such that

$$(2.35) \leq C(A, \mathcal{A}, \mathbf{f}) \frac{1}{J} \sum_{j=1}^J \|(a_j^* - \hat{a}_j^\lambda, \alpha_j^* - \hat{\alpha}_j^\lambda)\|_{\mathbb{R}^2}^2. \quad (2.38)$$

The term (2.36) is very similar to the term (2.35) and we have

$$(2.36) \leq C(A, \mathcal{A}, B) \frac{1}{J} \sum_{j=1}^J \|(a_j^* - \hat{a}_j^\lambda, \alpha_j^* - \hat{\alpha}_j^\lambda)\|_{\mathbb{R}^2}^2, \quad (2.39)$$

for some constant $C(A, \mathcal{A}, B) > 0$ independent of k and J . By using formula (2.38), (2.39) and Proposition 2.14 together, there exists a constant $C(L, s, A, \mathcal{A}, B, \mathbf{f})$ independent of k and J such that for all $x > 0$,

$$\mathbb{P}\left((2.35) + (2.36) \geq C(L, s, A, \mathcal{A}, B, \mathbf{f}) (F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)))\right) \leq 4e^{-x}, \quad (2.40)$$

where $V_1(k, J, x)$ is defined in statement of proposition 2.14 and $F(u) = u + \sqrt{u}$ for $u \geq 0$.

The term (2.37) can be bounded in probability as follows. First of all, remark that there exists a constant $C(A, \mathcal{A}) > 0$ such that

$$(2.37) \leq C(A, \mathcal{A}) \frac{1}{J} \sum_{j=1}^J \|\bar{\zeta}_j\|_{\mathbb{R}^2}^2.$$

Then, for all $j = 1, \dots, J$, the random variable $\bar{\zeta}_j = \frac{1}{k} \sum_{\ell=1}^k ([\zeta_j^{(1)}]_\ell, [\zeta_j^{(2)}]_\ell) \in \mathbb{R}^2$ can be written

$$(\bar{\zeta}_j^1, \bar{\zeta}_j^2)' = \frac{1}{k} \begin{pmatrix} \mathbb{1}'_k & 0'_k \\ 0'_k & \mathbb{1}'_k \end{pmatrix} ([\zeta_j^{(1)}]_1, \dots, [\zeta_j^{(1)}]_k, [\zeta_j^{(2)}]_1, \dots, [\zeta_j^{(2)}]_k)'$$

where 0_k is a column vector of k zeros. The random vector $(\bar{\zeta}_j^1, \bar{\zeta}_j^2)'$ is a two dimensional centered Gaussian vector of variance $\mathbf{V} = \frac{1}{k^2} \begin{pmatrix} \mathbb{1}'_k & 0'_k \\ 0'_k & \mathbb{1}'_k \end{pmatrix} \Sigma \begin{pmatrix} \mathbb{1}_k & 0_k \\ 0_k & \mathbb{1}_k \end{pmatrix}$. The 2×2 matrix $\mathbf{V}$ is of trace less or

equal to $\frac{2}{k}\gamma_{\max}(k)$. Hence, the random variable $\frac{1}{J}\sum_{j=1}^J\|\bar{\zeta}_j\|_{\mathbb{R}^2}^2 = \frac{1}{J}\sum_{j=1}^J(\bar{\zeta}_j^{(1)}, \bar{\zeta}_j^{(2)})(\bar{\zeta}_j^{(1)}, \bar{\zeta}_j^{(2)})'$ has the same probability distribution as $\frac{1}{J}\xi'[Id_J \otimes \mathbf{V}]\xi$ where ξ is a centered Gaussian vector of variance Id_{2J} . A standard concentration inequality of χ^2 distribution (see e.g. [LM00] Lemma 1) is $\mathbb{P}\left(\xi'[Id_J \otimes \mathbf{V}]\xi \geq J\frac{2}{k}\gamma_{\max}(k) + \frac{4}{k}\gamma_{\max}(k)\sqrt{2Jx} + x\frac{4}{k}\gamma_{\max}(k)\right) \leq e^{-x}$ for any $x > 0$. It yields that there exists a constant $C(A, \mathcal{A}) > 0$ such that for all $x > 0$

$$\mathbb{P}\left((2.37) \geq C(A, \mathcal{A})\frac{\gamma_{\max}(k)}{k}\left(1 + \sqrt{2\frac{x}{J}} + \frac{x}{J}\right)\right) \leq e^{-x}. \quad (2.41)$$

To end the proof remark that for all $u \geq 0$ we have $\mathbb{P}(\frac{1}{J}\|\mathbf{b}_{\Theta_0}^* - \hat{\mathbf{b}}^\lambda\|_{\mathbb{R}^{2J}}^2 \geq u) \leq \mathbb{P}((2.35) + (2.36) + (2.37) \geq u)$. This together with (2.40) and (2.41) yield that there exists a constant $C(L, s, A, \mathcal{A}, B, \mathbf{f}) > 0$ independent of k and J such that for all $x > 0$ we have

$$\mathbb{P}\left(\frac{1}{J}\|\mathbf{b}_{\Theta_0}^* - \hat{\mathbf{b}}^\lambda\|_{\mathbb{R}^{2J}}^2 \geq C(L, s, A, \mathcal{A}, B, \mathbf{f})(F(k^{-\frac{2s}{2s+1}}) + F(V_1(k, J, x)) + V_2(k, J, x))\right) \leq 5e^{-x},$$

where $V_2(k, J, x) = \frac{\gamma_{\max}(k)}{k}\left(1 + \sqrt{2\frac{x}{J}} + \frac{x}{J}\right)$ and the proof of Proposition 2.16 is completed. $\square$

2.B. Technical Lemma

Lemma 2.21. Assume that Assumption 2.1 holds, i.e. $f \in H_s(L)$ with $s > \frac{3}{2}$ (see (2.21)) and $\mathbf{f} = (f(\frac{\ell}{k}))_{\ell=1}^k \in \mathbb{R}^{k \times 2}$. If $\lambda(k) = k^{\frac{1}{2s+1}}$ then there exists a constant $C(L, s)$ such that for all $f \in H_s(L)$ we have

$$\frac{1}{k}\|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(L, s)k^{-\frac{2s}{2s+1}},$$

where $\mathbf{A}^\lambda$ is the projection matrix defined in (2.17).

Proof. Recall the notations introduced Section 2.4.1 and 2.5.2: $c_m(f) = (c_m(f^{(1)}), c_m(f^{(2)})) \in \mathbb{C}^2$ is the m -th Fourier coefficient of $f \in L^2([0, 1], \mathbb{R}^2)$ and $c_m(\mathbf{f}) = (c_m^{(1)}(\mathbf{f}), c_m^{(2)}(\mathbf{f})) \in \mathbb{C}^2$ is the m -th discrete Fourier coefficient of $\mathbf{f} \in \mathbb{R}^{k \times 2}$. Thus, we have by Parseval's equality

$$\frac{1}{k}\|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 = \frac{1}{k}\left\|\left(\frac{1}{k}\sum_{|m|>\lambda} c_m(\mathbf{f})e^{i2\pi m \frac{\ell}{k}}\right)_{\ell=1}^k\right\|_{\mathbb{R}^{k \times 2}}^2 = \frac{1}{k^2}\sum_{|m|>\lambda}\|c_m(\mathbf{f})\|_{\mathbb{C}^2}^2.$$

where for all $c = (c^{(1)}, c^{(2)}) \in \mathbb{C}^2$, $\|c\|_{\mathbb{C}^2}^2 = |c^{(1)}|^2 + |c^{(2)}|^2$. It yields

$$\frac{1}{k}\|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq 2\sum_{|m|>\lambda}\left\|\frac{1}{k}c_m(\mathbf{f}) - c_m(f)\right\|_{\mathbb{C}^2}^2 + 2\sum_{|m|>\lambda}\|c_m(f)\|_{\mathbb{C}^2}^2.$$

Firstly, Lemma 1.10 in [Tsy09] ensures that if $s > \frac{1}{2}$ then $\left|\frac{1}{k}c_m(\mathbf{f}^{(i)}) - c_m(f^{(i)})\right| \leq C(L, s)k^{-s+\frac{1}{2}}$ for any $m \in \mathbb{N}$ and $i = 1, 2$. Secondly, equation (1.87) in [Tsy09] ensure that if $\lambda(k) = k^{\frac{1}{2s+1}}$ then $\sum_{|m|>\lambda}|c_m(f^{(i)})|^2 \leq C(L, s)k^{-\frac{2s}{2s+1}}$. Thence, there is a constant $C(L, s)$ independent of k such that

$$\frac{1}{k}\|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(L, s)(k^{2-2s} + k^{-\frac{2s}{2s+1}}).$$

If $s \geq \frac{3}{2}$ then we have

$$\frac{1}{k}\|\mathbf{A}^\lambda \mathbf{f} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(L, s)k^{-\frac{2s}{2s+1}}. \quad \square$$

We now give a general expression of the gradient and of the Hessian of the criterion D . Recall that we have

$$D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{1}{J} \sum_{j=1}^J \left\| e^{-a_j} (e^{a_j^*} \mathbf{f} R_{\alpha_j^*} + \mathbb{1}_k \otimes (b_j^* - b_j)) R_{-\alpha_j} + \frac{1}{J} \sum_{j'=1}^J e^{-a_{j'}} (e^{a_{j'}^*} \mathbf{f} R_{\alpha_{j'}^*} + \mathbb{1}_k \otimes (b_{j'}^* - b_{j'})) R_{-\alpha_{j'}} \right\|_{\mathbb{R}^{k \times 2}}^2$$

where $(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = (a_1, \dots, a_J, \alpha_1, \dots, \alpha_J, b_1, \dots, b_J) \in \mathbb{R}^J \times [-\pi, \pi]^J \times \mathbb{R}^{2J}$. To shorten the formulas below we note $g_j = (g_j^{(1)}, g_j^{(2)}, g_j^{(3)}, g_j^{(4)}) = (a_j, \alpha_j, b_j)$, that is $g_j^{(1)} = a_j$, $g_j^{(2)} = \alpha_j$, $g_j^{(3)} = b_j^{(1)}$ and $g_j^{(4)} = b_j^{(2)}$. Let $\mathbf{f}_{g_j} = g_j^{-1} \cdot g_j^* \cdot \mathbf{f} = e^{-a_j} (e^{a_j^*} \mathbf{f} R_{\alpha_j^*} + \mathbb{1}_k \otimes (b_j^* - b_j)) R_{-\alpha_j}$, and for all $j_1 = 1, \dots, J$ and $p_1 = 1, \dots, 4$,

$$\partial_{g_{j_1}^{(p_1)}} D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = \frac{2}{Jk} \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}}, \mathbf{f}_{g_{j_1}} - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{g_{j'}} \right\rangle_{\mathbb{R}^{k \times 2}}. \quad (2.42)$$

The second order derivatives are

$$\partial_{g_{j_2}^{(p_2)}} \partial_{g_{j_1}^{(p_1)}} D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) = -\frac{2}{J^2 k} \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}}, \partial_{g_{j_2}^{(p_2)}} \mathbf{f}_{g_{j_2}} \right\rangle_{\mathbb{R}^{k \times 2}}, \quad \text{if } j_1 \neq j_2, \quad (2.43)$$

$$\begin{aligned} \partial_{g_{j_1}^{(p_2)}} \partial_{g_{j_1}^{(p_1)}} D(\mathbf{a}, \boldsymbol{\alpha}, \mathbf{b}) &= \frac{2}{Jk} \left\langle \partial_{g_{j_1}^{(p_2)}} \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}}, \left(\mathbf{f}_{g_{j_1}} - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{g_{j'}} \right) \right\rangle_{\mathbb{R}^{k \times 2}} \\ &\quad + \left(\frac{2}{Jk} - \frac{2}{J^2 k} \right) \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}}, \partial_{g_{j_1}^{(p_2)}} \mathbf{f}_{g_{j_1}} \right\rangle_{\mathbb{R}^{k \times 2}}. \end{aligned} \quad (2.44)$$

The expressions of the gradient and the Hessian of D simplify on the set $(\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * \mathcal{G}$, see Lemma 2.5. For any $g_0 = (a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$, we have $\mathbf{f}_{g_{j_1}^* \cdot g_0} - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{g_{j'}^* \cdot g_0} = e^{-a_0} (\mathbf{f} - \mathbb{1}_k \otimes b_0) R_{-\alpha_0} - \frac{1}{J} \sum_{j'=1}^J e^{-a_0} (\mathbf{f} - \mathbb{1}_k \otimes b_0) R_{-\alpha_0} = 0$. It yields that for all $(a_0, \alpha_0, b_0) \in \mathbb{R} \times [-\pi, \pi] \times \mathbb{R}^2$ we have $\nabla D((\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0)) = 0$, and

$$\partial_{g_{j_2}^{(p_2)}} \partial_{g_{j_1}^{(p_1)}} D((\mathbf{a}^*, \boldsymbol{\alpha}^*, \mathbf{b}^*) * (a_0, \alpha_0, b_0)) = \begin{cases} -\frac{2}{J^2 k} \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}^* \cdot g_0}, \partial_{g_{j_2}^{(p_2)}} \mathbf{f}_{g_{j_2}^* \cdot g_0} \right\rangle_{\mathbb{R}^{k \times 2}}, & \text{if } j_1 \neq j_2, \\ \left(\frac{2}{Jk} - \frac{2}{J^2 k} \right) \left\langle \partial_{g_{j_1}^{(p_1)}} \mathbf{f}_{g_{j_1}^* \cdot g_0}, \partial_{g_{j_2}^{(p_2)}} \mathbf{f}_{g_{j_1}^* \cdot g_0} \right\rangle_{\mathbb{R}^{k \times 2}}. \end{cases} \quad (2.45)$$

Lemma 2.22. *The smallest eigenvalue of $\nabla^2 D_0(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*)$ restricted to the subset Θ_0 is greater than $e^{-2A} \frac{2}{Jk} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2$.*

Proof. In this proof, $\mathbf{h} = (\mathbf{a}, \boldsymbol{\alpha}) \in \mathbb{R}^J \times [-\pi, \pi]^J$ and $\mathbf{f}_{h_j} = e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j}$. We have $\partial_{a_{j_1}} \mathbf{f}_{h_{j_1}}^0 = -\mathbf{f}_{h_{j_1}}^0$ and $\partial_{\alpha_{j_1}} \mathbf{f}_{h_{j_1}}^0 = \mathbf{f}_{h_{j_1}}^0 R_{-\frac{\pi}{2}}$ for all $j_1 = 1, \dots, J$. By using Formulas (2.45), the Hessian of D_0 at the point $(\mathbf{a}_{\Theta_0}^*, \boldsymbol{\alpha}_{\Theta_0}^*) = (\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*) \in \Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$ is given by,

$$\partial_{\alpha_{j_1}} \partial_{\alpha_{j_2}} D_0(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*) = \partial_{a_{j_1}} \partial_{a_{j_2}} D_0(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*) = \begin{cases} -\frac{2}{J^2 k} \|e^{\bar{\mathbf{a}}^*} \mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2, & \text{if } j_1 \neq j_2, \\ \left(\frac{2}{Jk} - \frac{2}{J^2 k} \right) \|e^{\bar{\mathbf{a}}^*} \mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \end{cases}$$

and the second order cross derivatives are 0. The Hessian of D_0 at $(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*)$ can be written,

$$\nabla^2 D_0(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*) = \frac{2}{J^2 k} \|e^{\bar{\mathbf{a}}^*} \mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \begin{pmatrix} JId_J - \mathbb{1}_{J \times J} & 0 \\ 0 & JId_J - \mathbb{1}_{J \times J} \end{pmatrix}, \quad (2.46)$$

where Id_J is the identity in $\mathbb{R}^J$ and $\mathbb{1}_J$ is the $J \times J$ matrix of ones. The eigenvalues of $JId_J - \mathbb{1}_{J \times J}$ are 0 with eigenvector $\mathbb{1}_J$ and J with eigenspace $\mathbb{1}_J^\perp$. It yields that on $\Theta_0^{\mathbf{a}, \boldsymbol{\alpha}}$ the smallest eigenvalue of $\nabla^2 D_0(\mathbf{a}^* - \bar{\mathbf{a}}^*, \boldsymbol{\alpha}^* - \bar{\boldsymbol{\alpha}}^*)$ is $e^{2\bar{\mathbf{a}}^*} \frac{2}{Jk} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2$. To finish the proof, remark that $\bar{\mathbf{a}}^* \leq A$. $\square$

Lemma 2.23. Let $\delta = \max\{A, \mathcal{A}\}$. For all $\mathbf{c} = (a_1, \dots, a_J, \alpha_1, \dots, \alpha_J) \in \Theta_0^{a, \alpha}$ and $(\mathbf{a}, \boldsymbol{\alpha}) \in \mathbb{R}^{2J}$ we have

$$\left| [\nabla^3 D_0(\mathbf{c})](\mathbf{a}, \boldsymbol{\alpha}) \right| \leq 40\delta e^{2A} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{Jk} \|(\mathbf{a}, \boldsymbol{\alpha})\|_{\mathbb{R}^{2J}}^2.$$

Proof. In this proof, $\mathbf{h} = (\mathbf{a}, \boldsymbol{\alpha}) \in \mathbb{R}^J \times [-\pi, \pi]^J$, that is $h_j^{(1)} = a_j$ is the parameter of scaling and $h_j^{(2)} = \alpha_j$ is the parameter of rotation. We have $\mathbf{f}_{h_j}^0 = e^{a_j^* - a_j} \mathbf{f}_0 R_{\alpha_j^* - \alpha_j}$. Then, from equations (2.43) and (2.44), it follows that for all $j_1, j_2, j_3 = 1, \dots, J$ and $p_1, p_2, p_3 = 1, \dots, 2$,

$$\begin{aligned} \partial_{h_{j_3}^{(p_3)}} \partial_{h_{j_2}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{a}, \boldsymbol{\alpha}) &= 0, \quad \text{if } j_1 \neq j_2 \text{ and } j_2 \neq j_3 \text{ and } j_1 \neq j_3, \\ \partial_{h_{j_2}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{a}, \boldsymbol{\alpha}) &= -\frac{2}{J^2 k} \left\langle \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \partial_{h_{j_2}^{(p_3)}} \mathbf{f}_{h_{j_2}}^0 \right\rangle_{\mathbb{R}^{k \times 2}}, \quad \text{if } j_1 \neq j_2, \\ \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{a}, \boldsymbol{\alpha}) &= \frac{2}{Jk} \left\langle \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \left(\mathbf{f}_{h_{j_1}}^0 - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{h_{j'}}^0 \right) \right\rangle_{\mathbb{R}^{k \times 2}} \\ &\quad + \left(\frac{2}{Jk} - \frac{2}{J^2 k} \right) \left(\left\langle \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \partial_{h_{j_1}^{(p_3)}} \mathbf{f}_{h_{j_1}}^0 \right\rangle_{\mathbb{R}^{k \times 2}} + \left\langle \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \right. \right. \\ &\quad \left. \left. \partial_{h_{j_1}^{(p_2)}} \mathbf{f}_{h_{j_1}}^0 \right\rangle_{\mathbb{R}^{k \times 2}} + \left\langle \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \mathbf{f}_{h_{j_1}}^0, \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0 \right\rangle_{\mathbb{R}^{k \times 2}} \right) \end{aligned}$$

By Cauchy-Schwarz inequality we have,

$$\left| \left\langle \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \partial_{h_{j_2}^{(p_3)}} \mathbf{f}_{h_{j_2}}^0 \right\rangle_{\mathbb{R}^{k \times 2}} \right| \leq \|\partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0\|_{\mathbb{R}^{k \times 2}} \|\partial_{h_{j_2}^{(p_3)}} \mathbf{f}_{h_{j_2}}^0\|_{\mathbb{R}^{k \times 2}} \leq e^{2A} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2. \quad (2.47)$$

and

$$\begin{aligned} &\left| \left\langle \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0, \left(\mathbf{f}_{h_{j_1}}^0 - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{h_{j'}}^0 \right) \right\rangle_{\mathbb{R}^{k \times 2}} \right| \\ &\leq \left\| \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} \mathbf{f}_{h_{j_1}}^0 \right\|_{\mathbb{R}^{k \times 2}} \left\| \mathbf{f}_{h_{j_1}}^0 - \frac{1}{J} \sum_{j'=1}^J \mathbf{f}_{h_{j'}}^0 \right\|_{\mathbb{R}^{k \times 2}} \leq 2e^{2A} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \quad (2.48) \end{aligned}$$

For $\kappa = (\kappa_1, \dots, \kappa_{2J}) \in \mathbb{N}^{2J}$, denote by $|\kappa| = \kappa_1 + \dots + \kappa_{2J}$ and

$$(\partial_{\mathbf{h}})^\kappa = (\partial_{a_1})^{\kappa_1} (\partial_{\alpha_1})^{\kappa_2} \dots (\partial_{a_J})^{\kappa_{2J-1}} (\partial_{\alpha_J})^{\kappa_{2J}}.$$

Then, the differential of order 3 of D_0 at $\mathbf{c} \in \Theta_0^{a, \alpha}$ applied at $(\mathbf{a}, \boldsymbol{\alpha}) \in \mathbb{R}^{2J}$ writes as

$$[\nabla^3 D_0(\mathbf{c})](\mathbf{a}, \boldsymbol{\alpha}) = \sum_{|\kappa|=3} (\partial_{\mathbf{h}})^\kappa D_0(\mathbf{c}) \mathbf{h}^\kappa$$

where $\mathbf{h}^\kappa = a_1^{\kappa_1} \alpha_1^{\kappa_2} \dots a_J^{\kappa_{2J-1}} \alpha_J^{\kappa_{2J}}$. This formula together with equations (2.47) and (2.48) give,

$$\begin{aligned}
 \left| [\nabla^3 D_0(\mathbf{c})](\mathbf{a}, \boldsymbol{\alpha}) \right| &= \left| \sum_{p_1, p_2, p_3=1}^2 \sum_{j_1=1}^J \partial_{h_{j_1}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{c}) h_{j_1}^{(p_1)} h_{j_1}^{(p_2)} h_{j_1}^{(p_3)} \right. \\
 &\quad \left. + 3 \sum_{j_1 \neq j_2=1}^J \partial_{h_{j_2}^{(p_3)}} \partial_{h_{j_1}^{(p_2)}} \partial_{h_{j_1}^{(p_1)}} D_0(\mathbf{c}) h_{j_1}^{(p_1)} h_{j_1}^{(p_2)} h_{j_2}^{(p_3)} \right| \\
 &\leq 2e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \sum_{p_1, p_2, p_3=1}^2 \left(\frac{4}{J} \sum_{j_1=1}^J |h_{j_1}^{(p_1)} h_{j_1}^{(p_2)} h_{j_1}^{(p_3)}| + \frac{6}{J^2} \sum_{j_1 \neq j_2=1}^J |h_{j_1}^{(p_1)} h_{j_1}^{(p_2)} h_{j_2}^{(p_3)}| \right) \\
 &\leq 2\delta e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \sum_{p_1, p_2=1}^2 \left(\frac{4}{J} \sum_{j_1=1}^J |h_{j_1}^{(p_1)} h_{j_1}^{(p_2)}| + \frac{6(J-1)}{J^2} \sum_{j_1=1}^J |h_{j_1}^{(p_1)} h_{j_1}^{(p_2)}| \right) \\
 &\leq 20\delta e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{J} \sum_{j=1}^J \sum_{p_1, p_2=1}^2 |h_j^{(p_1)} h_j^{(p_2)}| \\
 &= 20\delta e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{J} \sum_{j=1}^J \left(\sum_{p_1=1}^2 |h_j^{(p_1)}| \right)^2 \leq 40\delta e^{2A} \frac{1}{k} \|\mathbf{f}_0\|_{\mathbb{R}^{k \times 2}}^2 \frac{1}{J} \sum_{j=1}^J \sum_{p_1=1}^2 |h_j^{(p_1)}|^2.
 \end{aligned}$$

And the claim is now proved. $\square$

Lemma 2.24. For all $\mathbf{f} \in \mathbb{R}^{k \times 2}$ and $(a_1, \alpha_1, b_1), (a_2, \alpha_2, b_2) \in [-A, A] \times [-\mathcal{A}, \mathcal{A}] \times [-B, B]^2$, let $g_i \cdot \mathbf{f} = e^{a_i} \mathbf{f} R_{\alpha_i} + \mathbf{1}_k \otimes b_i$, $i = 1, 2$. Then, we have

$$\frac{1}{k} \|g_1 \cdot \mathbf{f} - g_2 \cdot \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq C(A, \mathbf{f}) \|(a_1, \alpha_1, b_1) - (a_2, \alpha_2, b_2)\|_{\mathbb{R}^4}^2,$$

where $C(A, \mathbf{f}) = 2 \max\{4e^{4A} \frac{1}{k} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2, 1\}$.

Proof. We have

$$\frac{1}{k} \|g_1 \cdot \mathbf{f} - g_2 \cdot \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq 2e^{2A} \frac{1}{k} \|e^{a_1 - a_2} \mathbf{f} R_{\alpha_1 - \alpha_2} - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 + 2 \frac{1}{k} \|\mathbf{1}_k \otimes (b_1 - b_2)\|_{\mathbb{R}^{k \times 2}}^2 \quad (2.49)$$

Let now $F(a, \alpha) = \frac{1}{\sqrt{k}} \|e^a \mathbf{f} R_\alpha - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}$. We have $|\partial_a F(a, \alpha)| = \frac{1}{\sqrt{k} \|e^a \mathbf{f} R_\alpha - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}} |\langle e^a \mathbf{f} R_\alpha, e^a \mathbf{f} R_\alpha - \mathbf{f} \rangle_{\mathbb{R}^{k \times 2}}| \leq \frac{e^A}{\sqrt{k}} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}$ and $|\partial_\alpha F(a, \alpha)| = \frac{1}{\sqrt{k} \|e^a \mathbf{f} R_\alpha - \mathbf{f}\|_{\mathbb{R}^{k \times 2}}} |\langle e^a \mathbf{f} R_{\alpha + \frac{\pi}{2}}, e^a \mathbf{f} R_\alpha - \mathbf{f} \rangle_{\mathbb{R}^{k \times 2}}| \leq \frac{e^A}{\sqrt{k}} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}$. The Euclidean norm of the gradient of F satisfies

$$\|\nabla F(a, \alpha)\|_{\mathbb{R}^2} = \sqrt{|\partial_a F(a, \alpha)|^2 + |\partial_\alpha F(a, \alpha)|^2} \leq \sqrt{2} e^A \frac{1}{\sqrt{k}} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}.$$

Since we have $|F(a, \alpha)| = |F(a, \alpha) - F(0, 0)| \leq \sqrt{2} e^A \frac{1}{\sqrt{k}} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}} \|(a, \alpha)\|_{\mathbb{R}^2}$, equation (2.49) yields

$$\frac{1}{k} \|g_1 \cdot \mathbf{f} - g_2 \cdot \mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2 \leq 2 \max\{4e^{4A} \frac{1}{k} \|\mathbf{f}\|_{\mathbb{R}^{k \times 2}}^2, 1\} (|a_1 - a_2|^2 + |\alpha_1 - \alpha_2|^2 + |b_1^{(1)} - b_2^{(1)}|^2 + |b_1^{(2)} - b_2^{(2)}|^2),$$

which concludes the proof. $\square$

Condition nécessaire et suffisante d'existence de la moyenne de Fréchet sur le cercle

Ce chapitre est indépendant des chapitres 1 et 2. Il est rédigé sous forme d'un article et a été soumis à la revue *ESAIM : Probability and Statistics*. Il est donc écrit en anglais et son titre original est *Necessary and sufficient condition for the existence of a Fréchet mean on the circle*. Il est autonome du reste de ce mémoire du point de vue des notations et des définitions.

Résumé

Soit $\mathbb{S}^1$ le cercle unité du plan muni de la distance $d_{\mathbb{S}^1}$ donnée par la longueur d'arc entre deux points. La moyenne de Fréchet d'une mesure de probabilité μ sur $(\mathbb{S}^1, d_{\mathbb{S}^1})$ est bien définie si le minimiseur de la fonctionnelle de Fréchet F_μ existe et est unique. Dans ce chapitre nous donnons une condition nécessaire et suffisante garantissant la bonne définition de la moyenne de Fréchet sur $(\mathbb{S}^1, d_{\mathbb{S}^1})$. Ce critère nous permet d'étendre les conditions suffisantes déjà connues et d'en donner une nouvelle sans restriction sur le support de la mesure et sans hypothèse de parité sur la densité. On étudie ensuite la convergence de la moyenne de Fréchet empirique vers la moyenne de Fréchet. Enfin, nous donnons un algorithme pour calculer la moyenne de Fréchet d'une mesure empirique.

Abstract

Let $(\mathbb{S}^1, d_{\mathbb{S}^1})$ be the unit circle in $\mathbb{R}^2$ endowed with the arclength distance. We give a sufficient and necessary condition for a general probability measure μ to admit a well defined Fréchet mean on $(\mathbb{S}^1, d_{\mathbb{S}^1})$. This criterion allows to recover already known sufficient conditions of existence. We also derive a new sufficient condition without restriction on

the support of the measure. Then, we study the convergence of the empirical Fréchet mean to the Fréchet mean. An algorithm to compute the empirical Fréchet mean is also given.

Mots clés : Circular data; Fréchet mean; Uniqueness.

3.1. Introduction

Statistics for non-Euclidean data In many fields of interest, results of an experiment are objects taking values in non-Euclidean spaces. A rather general framework to model such data is Riemannian geometry and more particularly quotient manifolds. As an illustration, in biology or geology, directional data are often used, see e.g. [MJ00] or [Fis93] and references therein. In this case, observations take their values in the circle or a sphere, that is, an Euclidean space quotiented by the action of scaling. Another well known example of non-Euclidean data in the statistical literature is Kendall's Shape Space, see e.g. [KBCL99] or [DM98] and references therein. The shape of a k -ads of the plane is invariant by the action of the group of similarity in the plane. After a convenient renormalization, the shape space is identified to a complex projective space, that is the quotient of a sphere by the action of a group of rotation.

Definition of basic statistical concepts, such as mean, must be adapted for random variables with values in non-Euclidean spaces such as manifolds. To describe the localization of a probability distribution, one needs to define a central value. In an Euclidean space, problems of existence and uniqueness may arise when one tries to defined such a central value. For example, for a real valued random variable, the two most popular notion of central value are Euclidean mean and median. If $\mathbb{R}$ is endowed with a probability measure ν satisfying $\int_{\mathbb{R}} t^2 d\nu(t) < +\infty$, the Euclidean mean $m(\nu) \in \mathbb{R}$ is defined to be the argmin of $F_{\nu}(x) = \int_{\mathbb{R}} |x - t|^2 d\nu(t)$, $x \in \mathbb{R}$. The functional F_{ν} is quadratic, its argmin is unique and is given by $m(\nu) = \int_{\mathbb{R}} t d\nu(t)$. For the Euclidean mean, existence problems appear when the distribution is not sufficiently concentrated as in the case of Cauchy distribution, i.e with density $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$. Again on $(\mathbb{R}, \nu)$, a median of ν is a real $med(\nu)$ at which the functional $x \rightarrow \int_{\mathbb{R}} |x - t| d\nu(t)$ is minimum. Equivalently $med(\nu)$ satisfies $\nu([-\infty, med(\nu)]) \geq \frac{1}{2}$ and $\nu([med(\nu), +\infty]) \geq \frac{1}{2}$. In general, $med(\nu)$ is not reduced to a point and the definition suffers from a lack of uniqueness. As an illustration, consider the Bernoulli distribution of parameter $\frac{1}{2}$, $\nu = \frac{1}{2}(\delta_0 + \delta_1)$ for which $med(\nu) = [0, 1]$. In general metric spaces, the same problems of existence and uniqueness arise and the situation is further complicated by extra phenomenons that do not happen in the Euclidean spaces. Therefore, there has been multiple attempts to give a definition of a mean in non Euclidean space, see among many others [BP03, CK99, Le01, Le04, Kar77, ÉM91, OC95, Fré48].

In this chapter, we consider the so-called Fréchet mean, see [Fré48, Kar77, Le01, BP03] and references therein. We are particularly interested in the study of its uniqueness. The Fréchet mean is defined on general metric spaces by extending the fact that Euclidean mean minimizes the sum of square the distance to the data, see equation (3.3) below. To study the well definiteness of the Fréchet mean on a manifold, two facts must be taken into account: non uniqueness of geodesics from one point to another (existence of a cut locus) and the effect of curvature, see e.g. [CK99]. Due to the cut-locus, the distance function is no longer convex and finding conditions to ensure the uniqueness of the Fréchet mean is not obvious. Two main directions have been explored in the literature: bounding

the support of the measure in [Kar77, KBCL99, Le01, BF01, Afs11], or consider special cases of absolutely continuous radial distributions, see [KBCL99, Le98, KS08]. In a sense, these two conditions control the concentration of the probability measure. The philosophy behind these works is to ensure a convexity property of the Fréchet functional given equation (3.2) below, see e.g. the introduction of [Afs11] for a review of the above cited papers.

Fréchet mean on the circle We focus here on the so-called Fréchet mean of a probability measure μ on the unit circle $\mathbb{S}^1$ of the plane,

$$\mathbb{S}^1 = \{x_1^2 + x_2^2 = 1, (x_1, x_2) \in \mathbb{R}^2\}.$$

We denote by $d_{\mathbb{S}^1}$ the arclength distance on $\mathbb{S}^1$ given for all $x = (x_1, x_2), p = (p_1, p_2) \in \mathbb{S}^1$ by

$$d_{\mathbb{S}^1}(x, p) = 2 \arcsin \left(\frac{\|x - p\|}{2} \right), \quad (3.1)$$

where $\|x - p\| = \sqrt{(x_1 - p_1)^2 + (x_2 - p_2)^2}$ is the Euclidean norm in $\mathbb{R}^2$. The advantage of dealing with a simple object such as the circle is that curvature problems disappear and we only face the cut-locus problem. In this sense, it allows us to completely understand its effect on the non-convexity of the distance function $d_{\mathbb{S}^1}$, and to give a complete answer about the problem of uniqueness. In what follows, we fully characterize probability measures that admit a well defined Fréchet mean on the circle $(\mathbb{S}^1, d_{\mathbb{S}^1})$. In particular a necessary and sufficient condition is given in Theorem 3.8, which links the existence of a Fréchet mean for a measure μ to the comparison between the distribution μ and the uniform measure λ on $\mathbb{S}^1$. The surprising fact is that λ appears as a benchmark to discriminate measures having a well defined Fréchet mean. The uniform measure λ is the 'worst' possible case as all points of the circle is a Fréchet mean, indeed the Fréchet functional (3.2) is constant and equals to $\frac{\pi^3}{3}$.

In opposition to what have been done before we do not try to ensure convexity property on the Fréchet functional. Indeed, the definition of the Fréchet mean relies on the *global* optimization problem (3.3) which is, in general, non convex. The advantage of our approach is that we do not need to restrict the support or suppose restrictive conditions of symmetry on the density. As the geometry of flat manifold is simple, we can derive explicit form on the Fréchet functional and its derivative which can be hard to compute in non-flat manifolds such as n -dimensional spheres.

3.1.1. Organization of the paper

In Section 3.2, we recall the definition of the Fréchet mean, and we review already known sufficient conditions that ensure the well definiteness of the Fréchet mean. In Section 3.3, we introduce notations that will be used throughout the paper. In Section 3.4, we give explicit expressions for the Fréchet functional and its derivative. A study of the critical points of the Fréchet functional is also done. Section 3.5 contains the main result with the necessary and sufficient condition of Theorem 3.8 for the existence of the Fréchet mean for a general measure. We also recover the conditions introduced in Section 3.2, and we propose a new criterion that ensures the well definiteness of the Fréchet mean. In Section 3.6, we study the convergence of the empirical Fréchet mean to the Fréchet mean, and describe an algorithm to compute the empirical Fréchet mean.

3.2. Fréchet Mean

A standard way to extend the definition of mean in non euclidean metric space is to use the minimization property of the Euclidean mean. This definition is usually credited to M. Fréchet in [Fré48] although some authors credit it to E. Cartan (see e.g. [KBCL99]). Let M be a metric measured space endowed with the metric d and the probability measure μ . The Fréchet functional is defined for all $p \in M$ by

$$F_\mu(p) = \frac{1}{2} \int_M d^2(x, p) d\mu(x). \quad (3.2)$$

Definition 3.1. We say that the Fréchet mean of μ in (M, d) is well defined if F_μ is finite and admits a *unique* argmin. That is, there exists a unique $p^* \in M$ satisfying $F_\mu(p^*) = \min_{p \in M} F_\mu(p)$, and we note

$$p^* = \operatorname{argmin}_{p \in M} F_\mu(p). \quad (3.3)$$

When (M, d) is a Riemannian manifold, some authors call the argmins of F_μ the Riemannian center of mass in [OC95] or intrinsic mean in [BP03]. In what follows we restrict our attention to the case of the circle $\mathbb{S}^1$ endowed with the arclength distance $d_{\mathbb{S}^1}$ defined in (3.1). The circle $(\mathbb{S}^1, d_{\mathbb{S}^1})$ is a simple one dimensional compact Riemannian manifold of finite diameter $\max\{d_{\mathbb{S}^1}(x, p), x, p \in \mathbb{S}^1\} = \pi$. The Fréchet functional F_μ is thus finite for all $p \in \mathbb{S}^1$ and is continuous. Hence F_μ attains its minimum in at least one point and the only issue at hand is uniqueness.

To avoid problems of uniqueness in the *global* optimization problem (3.3), it is proposed in [Kar77] to compute the Fréchet mean of the restriction of the measure to some subset B of M . Typically, B is a geodesic ball of small radius. It yields to a *local* version of the Fréchet mean called Karcher mean, namely $\operatorname{argmin}_{p \in B} F_\mu(p)$. This point of view was mainly used to ensure convexity property on F_μ in order to guaranty the uniqueness of the Karcher mean.

The definition of the Fréchet mean was also extended by modifying the integrand in (3.2). Some authors introduce the so called ρ -Fréchet functional given by $\int_M \rho(x, p) d\mu(x)$ where $\rho : M \times M \rightarrow \mathbb{R}$. Various assumptions can be made on ρ , e.g. to be a (Riemannian) distance, a semi-metric or only to vanish on the diagonal. For example, in [Afs11] the author studies the case where M is a complete manifold with its Riemannian metric d and $\rho(p, x) = \frac{1}{t} d^t(p, x)$ where $t \in [1, +\infty[$.

3.2.1. Previous work on the uniqueness of the Fréchet mean

There exists few general conditions to ensure that a measure on a metric space admits a well defined Fréchet mean. As in Euclidean spaces, they are related to a certain notion of concentration. We refer to [Hel01] for the definitions of notions in Riemannian geometry used in this part. To the best of our knowledge, the only case treated in the literature is when (M, d) is a complete Riemannian manifold. In this framework, the non-uniqueness of the Fréchet mean is due to two facts: existence of a cut locus and the effect of curvature, see [CK99] and references therein for more detailed discussions on this point. In the rest of this section, we review some criteria that ensure the well definiteness of the Fréchet mean of a manifold (M, d) endowed with a probability measure μ .

Manifold with negative curvature The most general result concerns the simply connected Riemannian manifold (M, d) with negative sectional curvature. They are usually called Cartan-Hadamard manifolds and are globally diffeomorphic to $\mathbb{R}^n$ with some change in the metric. In particular there is no cut locus, i.e there is a unique minimizing geodesic between two points. In this case, a general probability measure μ admits a Fréchet mean provided $F_\mu(p)$ is finite for some $p \in M$, see [BP03] Theorem 2.1.

The situation is more complex when (M, d) has a non negative curvature. In this setting, it exists two kinds of sufficient conditions that ensure the well definiteness of the Fréchet mean.

Bound on the support The first condition concerns the complete connected Riemannian manifold (M, d) with non negative scalar curvature. The condition consists in a restriction on the support of the measure that must be contained in a sufficiently small geodesic ball [Kar77]. If (M, d) is a flat Riemannian manifold, a sufficient condition for the existence of the Fréchet mean is that the support of μ is contained in a geodesically convex open normal neighborhood of M and that $F_\mu(p)$ is finite for some p . For example, on the circle, this imposes μ to be supported in a half-circle. Suppose now that (M, d) has a scalar curvature bounded from above by $C > 0$ and that $R > 0$ denotes the infimum of the injectivity radius. Then, a sufficient condition to ensure the existence of the Fréchet mean of a measure μ is that the support of μ is contained in a geodesic ball of radius less than $\frac{1}{4} \min \left\{ R, \frac{\pi}{\sqrt{C}} \right\}$. For a precise statement and definitions see [Le01] Theorem 1. In the special case of the n -dimensional unit sphere $\mathbb{S}^n = \{x_1^2 + \dots + x_{n+1}^2 = 1, x_1, \dots, x_{n+1} \in \mathbb{R}\}$ of $\mathbb{R}^{n+1}$, the Fréchet mean of a measure μ is well defined as soon as its support is strictly included in a geodesic ball of radius $\frac{\pi}{4}$. In [BF01], the authors improve the bound when the measure has finite support, i.e is a finite sum of Dirac's masses. The Fréchet mean of such a measure is well defined if the support of the measure lies in a closed hemisphere, that is a geodesic ball of radius $\frac{\pi}{2}$, and is not contained entirely in the boundary of this hemisphere. The philosophy behind this condition of boundedness of the support, is to guaranty the *convexity* of F_μ in the interior of the support of μ .

Radial distribution The second kind of conditions that ensure the well definiteness of the Fréchet mean concerns particular examples of Riemannian manifolds. We denote by (Σ_2^k, ρ) the Kendall shape space endowed with the Procrustes distance ρ , see e.g. [KBCL99] Chapter 9 for definitions. It turns out that this space is isometric to the complex projective space $\mathbb{C}P^{k-2}$ endowed with the Fubini-Study metric. In [Le98] and [KBCL99], the authors consider the case of an absolutely continuous probability measure μ with respect to the uniform law in (Σ_2^k, ρ) . If there is a $p^* \in \Sigma_2^k$ such that the density of μ is a decreasing function of $\rho(p^*, \cdot)$ (i.e μ is radially distributed probability measure around p^*) then p^* is the Fréchet mean of μ . See Theorem 9.2 and 9.3 of [KBCL99] for details and proofs. There exists a similar result for the special case of the circle $(\mathbb{S}^1, d_{\mathbb{S}^1})$ which is proved in [KS08]. Let $p^* \in \mathbb{S}^1$ and μ be an absolute continuous probability measure with density $f(d_{\mathbb{S}^1}(p^*, \cdot))$ where $f : [0, \pi] \rightarrow \mathbb{R}$ is a decreasing function. Then p^* is the Fréchet mean of μ . Note, that this decreasing radial distribution condition can be interpreted as a concentration condition around the Fréchet mean p^* .

3.2.2. Other definition of mean in $\mathbb{S}^1$

In the literature of circular data, another definition of mean is usually used, see e.g. [Fis93, MJ00, BP03]. Let $\pi_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ be the canonical projection on the i -th coordinate, $i = 1, 2$. Let $\mu = \mu_{\mathbb{R}^2}$ be a probability measure on $\mathbb{S}^1$ viewed as a singular measure on the plane with support in $\mathbb{S}^1$ and consider the marginal distribution $\mu_1 = \mu \circ \pi_1$ and $\mu_2 = \mu \circ \pi_2$. The Euclidean mean of $\mu_{\mathbb{R}^2}$ is the point of the plane defined by

$$m(\mu_{\mathbb{R}^2}) = \left(\int_{\mathbb{R}} x_1 d\mu_1(x_1), \int_{\mathbb{R}} x_2 d\mu_2(x_2) \right)'.$$

Remark that the Euclidean mean is the well defined Fréchet mean of $\mu_{\mathbb{R}^2}$ in $(\mathbb{R}^2, \|\cdot\|)$ and that it belongs to the unit ball $\{x \in \mathbb{R}^2, \|x\| \leq 1\}$. Moreover, provided $m(\mu_{\mathbb{R}^2})$ is not the origin of the plane, *i.e.* $m(\mu_{\mathbb{R}^2}) \neq 0$, it can be projected back on $\mathbb{S}^1$ in a unique way, namely $p^* = \frac{m(\mu_{\mathbb{R}^2})}{\|m(\mu_{\mathbb{R}^2})\|}$. This point can also be interpreted as a Fréchet mean on $\mathbb{S}^1$ endowed with the chordal distance $d_{\mathbb{R}^2}(x, y) = \|x - y\|$. That is,

$$p^* = \operatorname{argmin}_{p \in \mathbb{S}^1} \int_{\mathbb{S}^1} d_{\mathbb{R}^2}^2(p, x) d\mu(x).$$

In [BP03] the Fréchet mean of $(\mathbb{S}^1, d_{\mathbb{R}^2})$ is called the extrinsic mean whereas the Fréchet mean of $(\mathbb{S}^1, d_{\mathbb{S}^1})$ is called the intrinsic mean. The two notions do not coincide in general, see Figure 3.1(a) for an example, but conditions can be given to ensure that they are equal, see Theorem 3.3 in [BP03]. The advantage of the extrinsic mean is that it is easy to compute since the projection on $\mathbb{S}^1$ is simply a renormalization. Moreover conditions of existence are well understood. Indeed, the uniqueness of the definition is guaranteed as soon as $m(\mu_{\mathbb{R}^2})$, the Euclidean mean of μ viewed as a measure in the plane, is not the origin.

Note also that, in general, the well-definiteness of the intrinsic mean and the extrinsic mean are not related. There is situation with a well defined intrinsic mean and no unique extrinsic mean and, conversely, where the extrinsic is unique and with no well defined intrinsic mean. The main drawback of the extrinsic mean is that it can be very unstable when $m(\mu_{\mathbb{R}^2})$ is near the origin. Consider the case where $\mu_{\mathbb{R}^2} = \omega \delta_{(0,1)} + (1 - \omega) \delta_{(0,-1)}$ with $\omega \in [0, 1]$. For all $0 < \varepsilon < \frac{1}{2}$, if $\omega = \frac{1}{2} + \varepsilon$ then $\frac{m(\mu_{\mathbb{R}^2})}{\|m(\mu_{\mathbb{R}^2})\|} = (0, 1)$ and if $\omega = \frac{1}{2} - \varepsilon$ then $\frac{m(\mu_{\mathbb{R}^2})}{\|m(\mu_{\mathbb{R}^2})\|} = (0, -1)$. Note that, in these case the intrinsic mean of $\mu_{\mathbb{R}^2}$ is not well defined, as there is two global argmin, see commentary after the proof of Proposition 3.10 and Figure 3.3.

3.3. Notations

In what follows, $\mathbb{1}_A$ denotes the indicator function of the set $A \subset \mathbb{R}$ and the notation $\int_a^b f(t) d\mu_p(t)$ stands for the Lebesgue integral $\int_{[a,b]} f(t) d\mu_p(t)$ if $a \leq b$ and $\int_{[b,a]} f(t) d\mu_p(t)$ if $b > a$.

3.3.1. The distance function

The one dimensional sphere $\mathbb{S}^1$ can be identified with the torus $\mathbb{R}/(2\pi\mathbb{Z})$, that is the real line $\mathbb{R}$ quotiented by the equivalent relation $\sim$ defined for all $\theta_1, \theta_2 \in \mathbb{R}$ by $\theta_1 \sim \theta_2$ if and only if $\theta_1 - \theta_2 =$

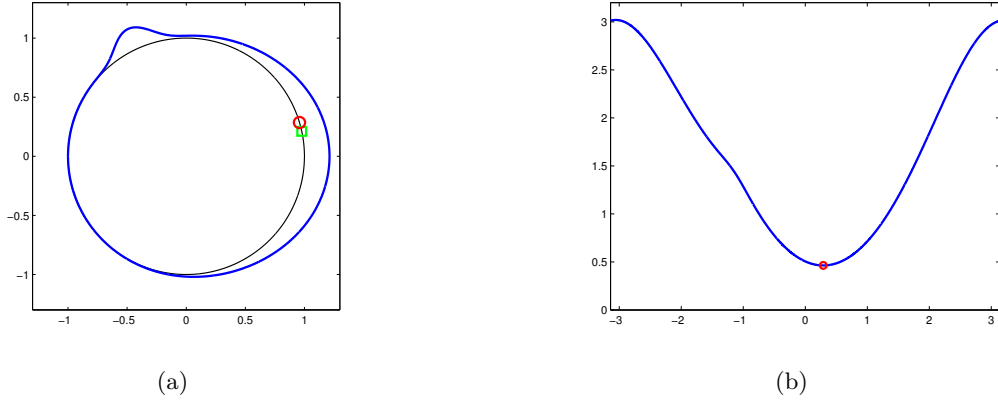


Figure 3.1: Let μ be a mixture of wrapped Gaussian on $\mathbb{S}^1$. (a) This is an example of distribution with distinct means. The (rescaled) density appears in blue. The red circle is the intrinsic mean, i.e the Fréchet mean of $(\mathbb{S}^1, d_{\mathbb{S}^1})$, and the green square is the extrinsic mean, i.e the Fréchet mean of $(\mathbb{S}^1, d_{\mathbb{R}^2})$. (b) plot of the Fréchet functional of F_μ , the global minimum (i.e. the intrinsic mean) is the red point.

$2\pi k$, $k \in \mathbb{Z}$. The equivalence class of $\theta \in \mathbb{R}$ is denoted by $[\theta]$. By choosing an arbitrary $p_0 \in \mathbb{S}^1$, we can identify any $p \in \mathbb{S}^1$ with $\theta_p^{p_0}$, the angle between p_0 and p . The arclength distance defined in (3.1) reads now, for $\theta_1, \theta_2 \in \mathbb{R}$,

$$d_{\mathbb{S}^1}(\theta_1, \theta_2) = \min\{|\theta_1 - \theta_2 + 2\pi k|, k \in \mathbb{Z}\}.$$

The circle $\mathbb{S}^1$ is locally isometric to the real line $\mathbb{R}$ and we use the same notation for the spherical distance between points in $\mathbb{S}^1 \subset \mathbb{R}^2$ and points in $\mathbb{R}/(2\pi\mathbb{Z})$.

The cut locus of a point $p \in \mathbb{S}^1$ is denoted by $\tilde{p}$ which is equals to the opposite point of p , that is $\tilde{p} = -p$. In $\mathbb{R}/(2\pi\mathbb{Z})$, the cut locus of $[\theta]$ is $[\theta + \pi]$. We refer to [Hel01] for details about cut loci.

3.3.2. Normal coordinates

To make explicit computation on $\mathbb{S}^1$, we use charts (also called coordinate systems), that is smooth one to one maps between $\mathbb{R}$ and $\mathbb{S}^1$. This is the terminology used in Riemannian geometry as the circle $\mathbb{S}^1$ is a flat sub-manifold of dimension 1. Even if the geometry of $\mathbb{S}^1$ is rather simple, this terminology allows us to connect the concepts used in this paper with more general situations. For each $p \in \mathbb{S}^1$, there is a canonical chart called the exponential map $e_p : \mathbb{R} \rightarrow \mathbb{S}^1$. It is defined for all $\theta \in \mathbb{R}$ by $e_p(\theta) = R_\theta p$ where $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ is the rotation matrix of angle θ that fixes 0 in $\mathbb{R}^2$. Equivalently, we could define $e_p : \mathbb{R} \rightarrow \mathbb{R}/(2\pi\mathbb{Z})$ by $e_p(\theta) = [\theta]$ and the exponential map corresponds in this case to the quotient map $\mathbb{R} \rightarrow \mathbb{R}/(2\pi\mathbb{Z})$.

For any $p \in \mathbb{S}^1$ the exponential map e_p , is onto but not one to one. In this paper, we choose to restrict the domain of definition of e_p to $[-\pi, \pi[$, for all $p \in \mathbb{S}^1$. Thus, there is now a unique $\theta_{p_2}^{p_1} \in [-\pi, \pi[$ satisfying $e_{p_1}(\theta_{p_2}^{p_1}) = p_2$ and we have for all $p \in \mathbb{S}^1$

$$e_p : [-\pi, \pi[\rightarrow \mathbb{S}^1 \quad \text{and} \quad e_p^{-1} : \mathbb{S}^1 \rightarrow [-\pi, \pi[.$$

Such parametrizations are called normal coordinate systems centered at p and $\theta_{p_2}^{p_1}$ is nothing else but the coordinate of p_2 read in a normal coordinate system centered in p_1 . In these coordinate systems,

the cut locus of a point p is $\theta_p - \pi$. To simplify notations, we will omit the exponent p_1 if no confusion is possible and we will write $\theta_{p_2}^{p_1} = \theta_{p_2}$.

In the case of the circle $\mathbb{S}^1$, a changing of coordinate system is particularly simple. It corresponds to translation in $\mathbb{R}/(2\pi\mathbb{Z})$. For any $p_1, p_2 \in \mathbb{S}^1 \simeq \mathbb{R}/(2\pi\mathbb{Z})$ we have

$$e_{p_2}^{-1} \circ e_{p_1}(\theta) = \theta - \theta_{p_1}^{p_2},$$

where $e_{p_2}^{-1} \circ e_{p_1} : \mathbb{R} \longrightarrow \mathbb{R}$.

3.3.3. Probability measure on the circle

Unless specified, in what follows, μ denotes a general probability measure on $(\mathbb{S}^1, \mathcal{B}(\mathbb{S}^1))$ where $\mathcal{B}(\mathbb{S}^1)$ is the Borel set of $\mathbb{S}^1 \subset \mathbb{R}^2$. Given a $p \in \mathbb{S}^1$, μ_p is the image measure of μ through $e_p^{-1} : \mathbb{S}^1 \longrightarrow [-\pi, \pi[$. It defines a measure on $\mathbb{R}$ given by

$$\mu_p(A) = \mu \circ e_p(A \cap [-\pi, \pi[), \quad \text{for all } A \in \mathcal{B}(\mathbb{R}), \quad (3.4)$$

where $\mathcal{B}(\mathbb{R})$ is the Borel set in $\mathbb{R}$. In words, μ_p is then nothing else but the measure μ read in the normal coordinate system centered at p . Note that we have $\mu_p([-\pi, \pi]) = 1$. For a measure μ_p on $\mathbb{R}$, let

$$m(\mu_p) = \int_{\mathbb{R}} t d\mu_p(t) \quad \text{and} \quad \text{Var}(\mu_p) = \int_{\mathbb{R}} t^2 d\mu_p(t) - (m(\mu_p))^2$$

be the usual Euclidean mean/expectation and variance of μ . The Fréchet functional on $(\mathbb{R}, d_{\mathbb{S}^1})$ endowed with the probability measure μ_p is

$$F_{\mu_p}(\theta) = \frac{1}{2} \int_{\mathbb{R}} d_{\mathbb{S}^1}^2(t, \theta) d\mu_p(t),$$

for all $\theta \in \mathbb{R}$. It attains its unique minimum at $m(\mu_p)$ and its value at this minimum is $\text{Var}(\mu_p)$. Note that we have for all $p_0, p \in \mathbb{S}^1$

$$F_{\mu}(p) = F_{\mu} \circ e_{p_0}(\theta_p^{p_0}) = F_{\mu_{p_0}}(\theta_p^{p_0}).$$

Finally, following [ÉM91], a point $p \in \mathbb{S}^1$ satisfying $m(\mu_p) = 0$ is called an exponential barycenter.

3.4. The Fréchet functional on the Circle

Using the notations introduced in Section 3.3, we give an expression of the Fréchet functional F_{μ} in the normal coordinate system centered at $p_0 \in \mathbb{S}^1$. For all $\theta_p^{p_0} \in [-\pi, \pi[$,

$$F_{\mu}(p) = \frac{1}{2} \begin{cases} \int_{-\pi}^{\theta_p^{p_0}-\pi} (\theta + 2\pi - \theta_p^{p_0})^2 d\mu_{p_0}(\theta) + \int_{\theta_p^{p_0}-\pi}^{\pi} (\theta - \theta_p^{p_0})^2 d\mu_{p_0}(\theta), & \text{if } 0 \leq \theta_p^{p_0} < \pi, \\ \int_{-\pi}^{\theta_p^{p_0}+\pi} (\theta - \theta_p^{p_0})^2 d\mu_{p_0}(\theta) + \int_{\theta_p^{p_0}+\pi}^{\pi} (\theta - 2\pi - \theta_p^{p_0})^2 d\mu_{p_0}(\theta), & \text{if } -\pi \leq \theta_p^{p_0} < 0. \end{cases} \quad (3.5)$$

Then, the Fréchet functional is 2π -Lipschitz since by the triangle inequality we have $|d_{\mathbb{S}^1}^2(p_1, x) - d_{\mathbb{S}^1}^2(x, p_2)| \leq 2\pi d_{\mathbb{S}^1}(p_1, p_2)$ for any $p_1, p_2, x \in \mathbb{S}^1$ which yields $|F_{\mu}(p_1) - F_{\mu}(p_2)| \leq 2\pi d_{\mathbb{S}^1}(p_1, p_2)$. Then, F_{μ} is continuous everywhere on $\mathbb{S}^1$. We can prove something better, it is continuously differentiable everywhere except at the cut locus of points of strictly positive measure (the atoms of the measure), see Section 3.4.1 below.

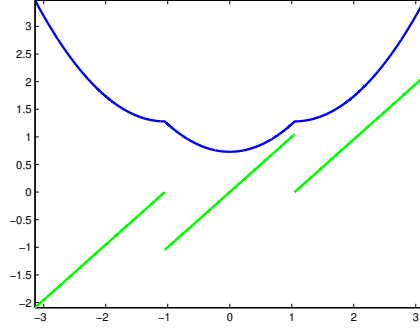


Figure 3.2: Let $\mu = \frac{1}{6}\delta_{p_1} + \frac{2}{3}\delta_{p^*} + \frac{1}{6}\delta_{p_2}$ with $p_1 = R_{\frac{2\pi}{3}}p^*$ and $p_2 = R_{-\frac{2\pi}{3}}p^*$. The blue curve is the graph of the function $F_{\mu_{p^*}}$ and the green curve is the graph of $\frac{d}{d\theta}F_{\mu_{p^*}}$.

3.4.1. The derivative of the Fréchet functional

A function $f : [-\pi, \pi[\rightarrow \mathbb{R}$ is said left continuous on $[-\pi, \pi[$ if it is left continuous everywhere on $]-\pi, \pi[$ and with $\lim_{\varepsilon \rightarrow 0^-} f(\pi + \varepsilon) = f(-\pi)$. Similarly, f is said to be continuous on $[-\pi, \pi[$ if it is left and right continuous on $[-\pi, \pi[$. We provide an explicit expression of the derivative of F_μ ,

Proposition 3.2. *Let μ be a probability measure on $(\mathbb{S}^1, d_{\mathbb{S}^1})$ and fix an arbitrary $p_0 \in \mathbb{S}^1$. Then, $F_\mu : \mathbb{S}^1 \rightarrow \mathbb{R}$ is differentiable in following sense :*

1. *Let $p \in \mathbb{S}^1$ be a point with a cut locus of μ -measure 0, i.e $\mu(\{-p\}) = 0$. Then F_μ is continuously differentiable at p and we have*

$$\frac{d}{d\theta}F_{\mu_{p_0}}(\theta_p^{p_0}) = \begin{cases} \theta_p^{p_0} - 2\pi\mu_{p_0}([-\pi, -\pi + \theta_p^{p_0}[) - m(\mu_{p_0}), & \text{if } 0 \leq \theta_p^{p_0} < \pi, \\ \theta_p^{p_0} + 2\pi\mu_{p_0}([\pi + \theta_p^{p_0}, \pi[) - m(\mu_{p_0}), & \text{if } -\pi \leq \theta_p^{p_0} < 0. \end{cases} \quad (3.6)$$

2. *The function $\frac{d}{d\theta}F_{\mu_{p_0}}$ is left continuous on $[-\pi, \pi[$. Then we extend the definition of the derivative of F_μ by setting for all $\theta \in [-\pi, \pi[$*

$$\frac{d}{d\theta}F_{\mu_{p_0}}(\theta) := \lim_{\varepsilon \rightarrow 0^-} \frac{d}{d\theta}F_{\mu_{p_0}}(\theta + \varepsilon). \quad (3.7)$$

3. *Let $p \in \mathbb{S}^1$ be a point with a cut locus of positive measure, i.e $\mu(\{-p\}) > 0$. Then, p is a cusp point of F_μ in the sense that $\lim_{\varepsilon \rightarrow 0^-} \frac{d}{d\theta}F_{\mu_{p_0}}(\theta_p^{p_0} + \varepsilon) - \lim_{\varepsilon \rightarrow 0^+} \frac{d}{d\theta}F_{\mu_{p_0}}(\theta_p^{p_0} + \varepsilon) = -\mu(\{-p\})$.*

Note that the left-continuity comes from our convention on the exponential map which is defined on $[-\pi, \pi[$. If a measure μ is such that $\mu(\{p\}) = 0$ for all $p \in \mathbb{S}^1$ then F_μ is of class $\mathcal{C}^1$ on $[-\pi, \pi[$, i.e. its derivative is continuous on $[-\pi, \pi[$. Differentiability issues appear when the measure μ has atoms, i.e. points p such that $\mu(\{p\}) > 0$. See Figure 3.2 where $F_{\mu_{p^*}}$ has three cusp points at $-\pi$, $-\frac{\pi}{3}$ and $\frac{\pi}{3}$ corresponding to cut loci of the three Dirac masses.

Proof. First of all, fix an arbitrary $p_0 \in \mathbb{S}^1$. For convenience we omit in this proof the superscript p_0 by writing $\theta_p = \theta_p^{p_0}$ for all $p \in \mathbb{S}^1$. Then, in the coordinate system centered at p_0 we have for all

$$\theta \in [-\pi, \pi[$$

$$\begin{aligned} F_{\mu_{p_0}}(\theta) &= \frac{1}{2} \int_{\mathbb{R}} t^2 d\mu_{p_0}(t) - \theta m(\mu_{p_0}) \\ &\quad + \frac{1}{2} \theta^2 + 2\pi (g_{\mu_{p_0}}^+(\theta) \mathbb{1}_{[0, \pi[}(\theta) + g_{\mu_{p_0}}^-(\theta) \mathbb{1}_{[-\pi, 0[}(\theta)) \end{aligned}$$

where $g_{\mu_{p_0}}^+(\theta) = \int_{-\pi}^{-\pi+\theta} (\pi + t - \theta) d\mu_{p_0}(t)$ and $g_{\mu_{p_0}}^-(\theta) = \int_{\theta+\pi}^{\pi} (\pi - t + \theta) d\mu_{p_0}(t)$. Hence to prove the claim in Proposition 3.2, we just have to compute the derivative of $g_{\mu_{p_0}}^+$. The case of $g_{\mu_{p_0}}^-$ can be deduced in the same way. For all $\theta \in]0, \pi[$ and $\varepsilon \in \mathbb{R}$ such that $\theta + \varepsilon \in]0, \pi[$ we have,

$$\frac{1}{\varepsilon} (g_{\mu_{p_0}}^+(\theta_p + \varepsilon) - g_{\mu_{p_0}}^+(\theta_p)) = \frac{1}{\varepsilon} \int_{-\pi+\theta_p}^{-\pi+\theta_p+\varepsilon} (\pi + t - \theta_p) d\mu_{p_0}(t) - \int_{-\pi}^{-\pi+\theta_p+\varepsilon} d\mu_{p_0}(t). \quad (3.8)$$

Then, let $p \in \mathbb{S}^1$ with $\theta_p > 0$ and with a cut locus of μ -measure 0. This is equivalent to the fact that $\mu_{p_0}(\{\theta_p - \pi\}) = 0$. The first term in the right hand side of (3.8) tends to zero as $\varepsilon \rightarrow 0$ since the integrand is, in absolute value, less than $|\varepsilon|$ and if $\varepsilon > 0$ (resp. $\varepsilon < 0$), then $\mu_{p_0}([-\pi + \theta_p, -\pi + \theta_p + \varepsilon[)$ (resp. $\mu_{p_0}([-\pi + \theta_p + \varepsilon, -\pi + \theta_p])$) tends to zero as $\varepsilon \rightarrow 0$ since $\mu_{p_0}(\{\theta_p - \pi\}) = 0$. Similarly, the second term tends to $\mu_{p_0}([-\pi, -\pi + \theta_p])$ when $\varepsilon \rightarrow 0$. The same arguments can be applied for $g_{\mu_{p_0}}^-$ and we have, $\frac{d}{d\theta} g_{\mu_{p_0}}^+(\theta_p) = -\mu_{p_0}([-\pi, -\pi + \theta_p])$, if $\theta_p \in]0, \pi[$ and $\frac{d}{d\theta} g_{\mu_{p_0}}^-(\theta_p) = \mu_{p_0}([\pi + \theta_p, \pi])$, if $\theta_p \in]-\pi, 0[$. If $\theta_p = 0$, (i.e $p_0 = p$) and $\mu_{p_0}(\{-\pi\}) = 0$ we can extend by continuity and let $\frac{d}{d\theta} g_{\mu_{p_0}}^+(\theta_p) = \lim_{\varepsilon \rightarrow 0-} \frac{d}{d\theta} g_{\mu_{p_0}}^+(0 + \varepsilon) = 0$. Again, at $\theta_p = -\pi$ with $\mu_{p_0}(\{0\}) = 0$ we can extend by continuity and let $\frac{d}{d\theta} g_{\mu_{p_0}}^+(-\pi) = \lim_{\varepsilon \rightarrow 0-} \frac{d}{d\theta} g_{\mu_{p_0}}^+(\pi + \varepsilon) = \mu_{p_0}^d([0, \pi[$.

Suppose now that $p \in \mathbb{S}^1$ is such that $\mu_{p_0}(\theta_p - \pi) > 0$. By the same arguments as above, we have if $\theta_p > 0$, $\lim_{\varepsilon \rightarrow 0-} \frac{1}{\varepsilon} (g_{\mu_{p_0}}^+(\theta_p + \varepsilon) - g_{\mu_{p_0}}^+(\theta_p)) = -\mu_{p_0}([-\pi, -\pi + \theta_p])$, and if $\theta_p < 0$, $\lim_{\varepsilon \rightarrow 0-} \frac{1}{\varepsilon} (g_{\mu_{p_0}}^-(\theta_p + \varepsilon) - g_{\mu_{p_0}}^-(\theta_p)) = \mu_{p_0}([\pi + \theta_p, \pi])$, and the formula can be extended to 0 and $-\pi$. Then, the function $\frac{d}{d\theta} F_{\mu_{p_0}}$ is left continuous everywhere on $[-\pi, \pi[$, continuous at each point with a cut locus of μ -measure 0 and with a (negative) jump of size $-\lim_{\varepsilon \rightarrow 0+} \mu_{p_0}([-\pi, -\pi + \theta_p + \varepsilon]) = -\mu_{p_0}(\{\pi + \theta_p\})$ at a point with a cut locus of positive measure. $\square$

3.4.2. Local minimum of the Fréchet functional

In this section, we precise the link between critical points of F_{μ} (i.e points at which the derivative of F_{μ} , in the sens of Proposition 3.2, is 0), exponential barycenters of μ (i.e points p satisfying $m(\mu_p) = 0$) and local argmins of F_{μ} . Note that part of the results presented here are already known for general Riemannian manifold in [BP03] (in particular see Theorem 2.1), [Pen06] Theorem 2 or [ÉM91]. The first result shows us that the critical points are exactly the exponential barycenter.

Corollary 3.3. *Let μ be a measure of probability on $\mathbb{S}^1$. We have*

$$\frac{d}{d\theta} F_{\mu_{p_0}}(\theta_p^{p_0}) = -m(\mu_p),$$

where the derivative is in the sens of equation (3.7).

Proof. We simply give an expression of $m(\mu_p) \in \mathbb{R}$ in normal coordinate centered at a $p_0 \in \mathbb{S}^1$,

$$\begin{aligned} m(\mu_p) &= \int_{\mathbb{R}} t d\mu_p(t) = \begin{cases} \int_{-\pi}^{\theta_p^{p_0}-\pi} (t - \theta_p^{p_0} + 2\pi) d\mu_{p_0}(t) + \int_{\theta_p^{p_0}-\pi}^{\pi} (t - \theta_p^{p_0}) d\mu_{p_0}(t) & \text{if } 0 < \theta_p^{p_0} < \pi, \\ \int_{-\pi}^{\theta_p^{p_0}+\pi} (t - \theta_p^{p_0}) d\mu_{p_0}(t) + \int_{\theta_p^{p_0}+\pi}^{\pi} (t - \theta_p^{p_0} - 2\pi) d\mu_{p_0}(t) & \text{if } -\pi \leq \theta_p^{p_0} < 0 \end{cases} \\ &= -\frac{d}{d\theta} F_{\mu_{p_0}}(\theta_p^{p_0}). \end{aligned}$$

Where the last inequality is given by Proposition 3.2. $\square$

When $\mu(\{p\}) = 0$ for all $p \in \mathbb{S}^1$, to be a critical point of F_μ is equivalent to be an extremum of F_μ . Unfortunately, when the measure μ has atoms, to be a critical point of F_μ is not a sufficient condition for p to be an extremum of F_μ . Consider example of Figure 3.2 where the Fréchet mean is well defined and is located at p^* , the mass point of weight $\frac{2}{3}$. Then, $m(\mu_p) = 0$ for $p = p^*$ (that is $\theta_p^{p^*} = 0$) and $p = \tilde{p}_2 = -p_2$ (that is $\theta_p^{p^*} = -\frac{\pi}{3}$). Note that for $p = \tilde{p}_1 = -p_1$, we have $m(\mu_p) < 0$. This example is particular since the critical points are at the cut loci of the atoms of μ . Nevertheless, a local argmin of F_μ cannot be a cusp point as we have the following result :

Corollary 3.4. *Let μ be a probability measure on $\mathbb{S}^1$. The cut locus of a (local or global) minimum of F_μ is of μ -measure 0.*

Proof. Choose an arbitrary $p_0 \in \mathbb{S}^1$ and let $p_m \in \mathbb{S}^1$ be a minimum of F_μ satisfying $\mu(\{\tilde{p}_m\}) = \delta > 0$. Recall Statement 3 in Proposition 3.2, where it is shown that $\frac{d}{d\theta} F_{\mu_{p_0}}$ has a negative jump of size $-\delta$ at $\theta_{p_m}^{p_0}$. Then, the possible sign of

$$\left(\lim_{\varepsilon \rightarrow 0^-} \frac{d}{d\theta} F_{\mu_{p_0}}(\theta_{p_m}^{p_0} + \varepsilon), \lim_{\varepsilon \rightarrow 0^+} \frac{d}{d\theta} F_{\mu_{p_0}}(\theta_{p_m}^{p_0} + \varepsilon) \right)$$

are $(+, +)$, $(+, -)$ and $(-, -)$. This means that $\theta_{p_m}^{p_0}$ cannot be a minimum of $F_{\mu_{p_0}}$ since it would correspond to the case $(-, +)$. $\square$

Remark 3.5. Note that assumptions of Corollary 1 in [Pen06] and Theorem 1 in [BP03] contain a condition of the form $\mu\{\tilde{p}^*\} = 0$ to ensure the (classical) differentiability of the Fréchet functional at its minimum. In the case of the circle, Corollary 3.4 shows us that the Fréchet functional is *automatically* differentiable at its minimum.

As an illustration, consider again Figure 3.2. The point $\theta_{p_m}^{p^*} = -\frac{\pi}{6}$ corresponds to the case $(-, -)$, the point $\theta_{p_m}^{p^*} = \frac{\pi}{6}$ to the case $(+, +)$ and the point $\theta_{p_m}^{p^*} = -\pi$ to the case $(+, -)$. It can be shown that when μ is a purely atomic measure we have a better result: the “regular” critical points (i.e critical point that are not cusp point) are local minima of F_μ . More precisely,

Corollary 3.6. *Let $\mu = \sum_{i=1}^n a_i \delta_{x_i}$ with $x_1, \dots, x_n \in \mathbb{S}^1$ and $a_1 + \dots + a_n = 1$ with $a_1, \dots, a_n > 0$. Then, the critical points of F_μ that are not in the cut locus of the x_i ’s are the local minima of F_μ .*

Proof. If p_m is a local argmin of F_μ then, by Proposition 3.4, it is not in the cut locus of the x_i ’s. Moreover, F_μ is continuously differentiable at p_m and then p_m is a critical point of F_μ . Consider

now a critical point p_c of F_μ that is not in the cut locus of the x_i 's which are denoted by $\tilde{x}_i$'s. As $\mu = \sum_{i=1}^n a_i \delta_{x_i}$ we have in normal coordinate centered at some $p_0 \in \mathbb{S}^1$

$$F_{\mu_{p_0}}(\theta) = \frac{1}{2} \begin{cases} \sum_{i=1}^n a_i (\theta - \theta_{x_i}^{p_0} - 2\pi \mathbb{1}_{[-\pi, \theta - \pi](\theta_{x_i}^{p_0})})^2, & \text{if } 0 \leq \theta < \pi, \\ \sum_{i=1}^n a_i (\theta - \theta_{x_i}^{p_0} + 2\pi \mathbb{1}_{[\pi + \theta, \pi](\theta_{x_i}^{p_0})})^2, & \text{if } -\pi \leq \theta < 0 \end{cases}$$

Then $\frac{d^2}{d\theta^2} F_{\mu_{p_0}}(\theta) = 1 > 0$ for all $\theta \neq \theta_{x_i}^{p_0}$ and $\theta_{p_c}^{p_0}$ is now clearly a minimum of $F_{\mu_{p_0}}$ by the second derivative test. $\square$

We end this section with a result that allows us to compute efficiently the critical point of F_μ . The proof is omitted as it is an immediate consequence of Corollary 3.3.

Corollary 3.7. *Let μ be a measure of probability on $\mathbb{S}^1$, then the following propositions are equivalent*

1. $p \in \mathbb{S}^1$ is a critical point of F_μ
2. $p \in \mathbb{S}^1$ satisfy $m(\mu_p) = 0$
3. for any $p_0 \in \mathbb{S}^1$,

$$\begin{cases} \frac{1}{2\pi}(\theta_p^{p_0} - m(\mu_{p_0})) = \mu_{p_0}([-\pi, -\pi + \theta_p^{p_0}[), & \text{if } 0 \leq \theta_p^{p_0} < \pi, \\ \frac{1}{2\pi}(-\theta_p^{p_0} - m(\mu_{p_0})) = \mu_{p_0}([\pi + \theta_p^{p_0}, \pi]), & \text{if } -\pi \leq \theta_p^{p_0} < 0. \end{cases} \quad (3.9)$$

The Corollaries 3.6 and 3.7 will be used to compute the empirical Fréchet mean as follow: Corollaries 3.6 suggests that the set of local argmins is included in the set of the exponential barycenter and equations (3.9) allows us to compute theses points. In particular, the measure μ is a purely atomic measure and it has a simple cumulative distribution function and the solutions of (3.9) are easy to compute. Finally, it suffices to compare the values of F_μ at points p satisfying $m(\mu_p) = 0$. See Section 3.6.2 for details.

3.5. A necessary and sufficient condition for the existence of the Fréchet mean

3.5.1. Main result

In what follows, p^* denotes a critical point of F_μ , i.e a point at which the derivative of F_μ , in the sens of Proposition 3.2, is 0. Let us consider for all $p \in \mathbb{S}^1$, the functional

$$G_\mu(p) = F_\mu(p) - F_\mu(p^*),$$

which vanishes at $p = p^*$. We give here our main result,

Theorem 3.8. *Let μ be a general probability measure and $p^* \in \mathbb{S}^1$ be a critical point of F_μ . Then, the following propositions are equivalent,*

1. p^* is a well defined Fréchet mean of $(\mathbb{S}^1, \mu)$.
2. For all $p \neq p^*$ $G(p) > 0$

3. For all $0 < \theta < \pi$

$$\int_0^\theta \lambda([- \pi, -\pi + t]) - \mu_{p^*}([- \pi, -\pi + t]) dt > 0,$$

and for all $-\pi \leq \theta < 0$,

$$\int_\theta^0 \lambda([\pi + t, \pi]) - \mu_{p^*}([\pi + t, \pi]) dt > 0,$$

where λ is the uniform measure on $[-\pi, \pi[$ and μ_{p^*} is defined in (3.4).

Theorem 3.8 gives a necessary and sufficient condition for the existence of the Fréchet mean of a general measure μ on the circle $\mathbb{S}^1$. This condition is given in terms of comparison between the μ -measure and the uniform measure λ of balls centered at the cut locus of a global minimum. Note that the uniform measure λ has a density with respect to the Lebesgue measure equals to the constant function $\frac{1}{2\pi}$.

As μ is a probability measure, the functions $t \mapsto \lambda([- \pi, -\pi + t]) - \mu_{p^*}([- \pi, -\pi + t])$ and $t \mapsto \lambda([\pi - t, \pi]) - \mu_{p^*}([\pi - t, \pi])$ do not need to be always positive for $t \in [-\pi, 0[$ and $t \in [0, \pi[$ respectively. A situation where it is positive is studied in Proposition 3.11 of the next paragraph. The point is that the μ -measure of a (small) neighborhood of the cut locus of p^* cannot be larger than the uniform measure of this neighborhood.

3.5.2. Proof of Theorem 3.8

First of all, there exists at least one global argmin $p^* \in \mathbb{S}^1$ of F_μ since F_μ is a continuous function defined on the compact set $\mathbb{S}^1$. Then, Proposition 3.2 and Corollary 3.4 ensure that p^* is a critical point of F_μ . Note also that Statement 1. and 2. are equivalent as they are simple reformulations of the definition of a well defined Fréchet mean. To prove Theorem 3.8 we only need to show that Statement 2. is equivalent to Statement 3. The proof relies on the computation of the derivative of G_μ expressed in a well chosen coordinate system.

Corollary 3.3 ensures that p^* is an exponential barycenter, i.e satisfy $m(\mu_{p^*}) = 0$. Hence in the normal coordinate system centered at such a p^* the functional $G_{\mu_{p^*}} = F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0)$ has a particularly simple expression, that is

$$\begin{aligned} G_{\mu_{p^*}}(\theta) &= \frac{1}{2}\theta^2 + 2\pi \mathbb{1}_{[0, \pi[}(\theta) \int_{-\pi}^{\theta - \pi} (\pi + t - \theta) d\mu_{p^*}(t) \\ &\quad + 2\pi \mathbb{1}_{[-\pi, 0[}(\theta) \int_{\theta + \pi}^{\pi} (\pi - t + \theta) d\mu_{p^*}(t). \end{aligned}$$

Using Proposition 3.2 we have the following result,

Lemma 3.9. *Let μ be a probability measure on $\mathbb{S}^1$ and $p^* \in \mathbb{S}^1$ be an argmin of F_μ . Then for any $\theta \in [-\pi, \pi[$*

$$G_{\mu_{p^*}}(\theta) = 2\pi \begin{cases} \int_0^\theta \frac{t}{2\pi} - \mu_{p^*}([- \pi, -\pi + t]) dt, & \text{if } 0 \leq \theta < \pi, \\ \int_\theta^0 \frac{-t}{2\pi} - \mu_{p^*}([\pi + t, \pi]) dt, & \text{if } -\pi \leq \theta < 0 \end{cases}$$

Proof. The probability measure μ can be decomposed as follow,

$$\mu = a\mu^d + (1-a)\mu^\delta, \quad 0 \leq a \leq 1, \quad (3.10)$$

where μ^d is a probability measure such that $\mu_d(\{p\}) = 0$ for all $p \in \mathbb{S}^1$ and $\mu^\delta = \sum_{j=1}^{+\infty} \omega_j \delta_{p_j}$ where $\sum_{j=1}^{+\infty} \omega_j = 1$ and $p_1, \dots, p_n, \dots \in \mathbb{S}^1$. Hence, we consider the two cases separately : first, when the measure is non atomic, and then, when it is purely atomic. The general case follows immediately in view of equation (3.10).

First, assume that μ is an atomless measure of $\mathbb{S}^1$. Proposition 3.2 ensures that F_μ is continuously differentiable everywhere and the real function $F_{\mu_{p^*}}$ is of class $\mathcal{C}^1$ on $[-\pi, \pi[$. Formula (3.6) and the fundamental theorem of calculus gives for all $\theta \in [-\pi, \pi[$

$$g_{\mu_{p^*}}(\theta) = \int_0^\theta t dt - 2\pi \begin{cases} \int_0^\theta \mu_{p^*}([-\pi, -\pi + t]) dt, & \text{if } 0 \leq \theta < \pi, \\ \int_\theta^0 \mu_{p^*}([\pi + t, \pi]) dt, & \text{if } -\pi \leq \theta < 0. \end{cases} \quad (3.11)$$

Consider now the case where μ is a purely atomic measure. First, we treat the case where the number of mass of Dirac in the sum is finite, i.e $\mu = \sum_{j=1}^n \omega_j \delta_{p_j}$, $n \in \mathbb{N}$. Recall that $F_{\mu_{p^*}}$ is a Lipschitz function on $[-\pi, \pi[$. Proposition 3.2 ensures that the derivative is piecewise continuous and formula (3.6) holds for all $\theta \in [-\pi, \pi[\setminus \{\theta_{p_j}\}_{j=1}^n$, i.e points that have a cut locus of μ -measure 0. Hence for all $\theta \in [-\pi, \pi[$, equation (3.11) holds too.

To treat the case where $\mu = \sum_{j=1}^{+\infty} \omega_j \delta_{x_j}$ we proceed by approximation. Let $\phi(n) = \{j \in \mathbb{N} \mid \omega_j \geq \frac{1}{2^n}\}$ and remark that $\text{Card}(\phi(n)) < +\infty$ for all $n \in \mathbb{N}$ since $\sum_{j=1}^{+\infty} \omega_j = 1$. Then if $\nu_{p^*}^n = \frac{1}{c(n)} \sum_{j \in \phi(n)} \omega_j \delta_{x_j}$, where $c(n) = \sum_{j \in \phi(n)} \omega_j$ is a normalizing constant, we have for all $\theta \in [-\pi, \pi[$,

$$g_{\nu_{p^*}^n}(\theta) = \int_0^\theta t dt - 2\pi \begin{cases} \int_0^\theta \nu_{p^*}^n([-\pi, -\pi + t]) dt, & \text{if } 0 \leq \theta < \pi, \\ \int_\theta^0 \nu_{p^*}^n([\pi + t, \pi]) dt, & \text{if } -\pi \leq \theta < 0. \end{cases}$$

The sequence $(\nu_{p^*}^n)_{n \geq 1}$ converges to μ in total variation. By the dominated convergence Theorem for all $\theta_p \in [-\pi, \pi[$, $g_{\nu_{p^*}^n}(\theta)$ converge as $n \rightarrow \infty$ to (3.11). $\square$

The proof of Theorem 3.8 is almost done. The uniform measure on $[-\pi, \pi[$ has a density of $\frac{1}{2\pi}$, that is $\lambda([-\pi, -\pi + t]) = \frac{t}{2\pi}$ for all $0 \leq t < \pi$. Then Lemma 3.9 ensures that Statement 3. is equivalent to $G_{\mu_p}(\theta) > 0$ for $\theta \in [-\pi, 0[\cup]0, \pi[$ and thus Statement 2. $\square$

3.5.3. Sufficient conditions of existence

In practice, it is convenient to derive simple sufficient conditions on probability measures to ensure the well definiteness of the Fréchet mean. The necessary and sufficient condition of Theorem 3.8 allows us to recover and extend the already known cases described Section 3.2.1. We also give a new criterion that does not restrict the support of the measure.

Bound on the support. A closed ball in $(\mathbb{S}^1, d_{\mathbb{S}^1})$ is the set $B(p_0, r) = \{p \in \mathbb{S}^1, d_{\mathbb{S}^1}(p, p_0) \leq r\}$. The next result was proved in [BF01] for measure with finite support in $\mathbb{S}^n$, $n \in \mathbb{N}$. In $\mathbb{S}^1$, it is easy to extend the result to general measure.

Proposition 3.10. *Let μ be a probability measure on the circle with support included in a closed ball $B(\hat{p}, \frac{\pi}{2})$ centered at some $\hat{p} \in \mathbb{S}^1$ and of length $\frac{\pi}{2}$. If $\mu_{\hat{p}}(\{-\frac{\pi}{2}, \frac{\pi}{2}\}) < 1$ then, the Fréchet mean p^* of μ is well defined and belongs to $B(\hat{p}, \frac{\pi}{2})$.*

Proof. If p^* is a minimum of F_μ then p^* belongs to $B(\hat{p}, \frac{\pi}{2})$. To see this, we shortly present a reflection argument given in [BF01]. For any $p \in \mathbb{S}^1$ lying outside $B(\hat{p}, \frac{\pi}{2})$ consider $\bar{p}$, the symmetric of p with respect to the border of $B(\hat{p}, \frac{\pi}{2})$. In the normal coordinate system centered at $\hat{p}$, the point p satisfies $|\theta_p^{\hat{p}}| > \frac{\pi}{2}$ and $\theta_{\bar{p}}^{\hat{p}} = \pi - \theta_p^{\hat{p}}$ if $\theta_p^{\hat{p}}$ is positive or $-\pi - \theta_p^{\hat{p}}$ if $\theta_p^{\hat{p}}$ is negative. Assume now that $\frac{\pi}{2} < \theta_p^{\hat{p}} \leq \pi$, the other case being similar. Since $\mu_{\hat{p}}(\{-\frac{\pi}{2}, \frac{\pi}{2}\}) < 1$, we have

$$\begin{aligned} F_\mu(p) &= \frac{1}{2} \int_{-\frac{\pi}{2}}^{\theta_p^{\hat{p}}} (\theta_p^{\hat{p}} - 2\pi - t)^2 d\mu_{\hat{p}}(t) + \frac{1}{2} \int_{\theta_p^{\hat{p}}}^{\frac{\pi}{2}} (\theta_p^{\hat{p}} - t)^2 d\mu_{\hat{p}}(t) \\ &> \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\theta_p^{\hat{p}} - t)^2 d\mu_{\hat{p}}(t) = F_\mu(\bar{p}). \end{aligned}$$

Again with the condition $\mu_{\hat{p}}(\{-\frac{\pi}{2}, \frac{\pi}{2}\}) < 1$, we have $F_\mu(p) = F_\mu(\bar{p})$ if and only if $\theta_p^{\hat{p}} = \pm \frac{\pi}{2}$. Hence, we have proved that $|\theta_{p^*}^{\hat{p}}| < \frac{\pi}{2}$, i.e the argmin lies inside the open ball $\dot{B}(\hat{p}, \frac{\pi}{2})$.

Now remark that $F_{\mu_{\hat{p}}}$ is quadratic on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and then achieves its unique minimum at $m(\mu_{\hat{p}})$. To conclude the proof, we take $p^* = e_{\hat{p}}(m(\mu_{\hat{p}}))$. $\square$

It is possible to use the criterion of Theorem 3.8 to prove the preceding Proposition. Indeed, one can show that $G_{\mu_{p^*}}(\theta) > \frac{1}{2}(\theta + \pi - 2\theta_{\hat{p}})^2$ for $\theta \in [-\pi, \theta_{\hat{p}} - \frac{\pi}{2}[$, $G_{\mu_{p^*}} = \frac{\theta^2}{2}$ for $\theta \in [\theta_{\hat{p}} - \frac{\pi}{2}, \theta_{\hat{p}} + \frac{\pi}{2}[$ and $G_{\mu_{p^*}} > \frac{1}{2}(\theta - \pi - 2\theta_{\hat{p}})^2$ for all $\theta \in [\theta_{\hat{p}} + \frac{\pi}{2}, \pi[$. The case of equality corresponds to the distribution $\mu_{p^*} = (1 - \varepsilon)\delta_{\theta_{\hat{p}}^* - \frac{\pi}{2}} + \varepsilon\delta_{\theta_{\hat{p}}^* + \frac{\pi}{2}}$ with $\varepsilon = \frac{1}{\pi}\theta_{\hat{p}}^* + \frac{1}{2}$ and in this case, there are two global argmins at 0 and $2(\theta_{\hat{p}} \pm \frac{\pi}{2})$, see Figure 3.3.

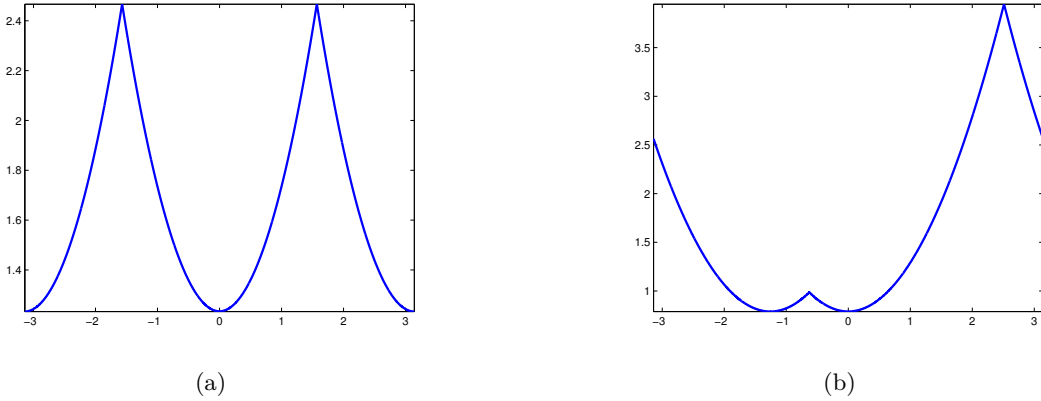


Figure 3.3: Let $\mu_\theta = (1 - \varepsilon)\delta_{\theta - \frac{\pi}{2}} + \varepsilon\delta_{\theta + \frac{\pi}{2}}$ with $\varepsilon = \frac{\theta}{\pi} + \frac{1}{2}$. (a) Plot of F_{μ_θ} with $\theta = 0$. (b) Plot of F_{μ_θ} with $\theta = \frac{3\pi}{10}$

Condition on the density. In this part we consider absolute continuous probability measure. First, we present a result for a subclass of radially distributed measures due to [KS08],

Proposition 3.11. *Let μ be a probability measure with density $f : \mathbb{S}^1 \rightarrow \mathbb{R}$ which can be written as $f(p) = \rho(d(p, p^*))$ for some $p^* \in \mathbb{S}^1$ where $\rho : [0, \pi] \rightarrow \mathbb{R}$ is a non constant decreasing function. Then p^* is the Fréchet mean of μ .*

Proof. Consider the function $f_{p^*} = f \circ e_{p^*} : [-\pi, \pi[\rightarrow \mathbb{R}$. By hypothesis, f_{p^*} is an even function decreasing on $[0, \pi[$ and increasing on $[-\pi, 0[$. The point p^* is now clearly a critical point of F_μ as $m(\mu_{p^*}) = 0$, see Corollary 3.3. Let for all $\theta \in [-\pi, \pi[$

$$\begin{aligned} k_{\mu_{p^*}}(\theta) &= \mathbb{1}_{[0, \pi[}(\theta) \mu_{p^*}([-\pi, -\pi + \theta]) + \mathbb{1}_{[-\pi, 0[}(\theta) \mu_{p^*}([\pi + \theta, \pi]) \\ &= \mathbb{1}_{[0, \pi[}(\theta) \int_0^\theta f_{p^*}(t - \pi) dt + \mathbb{1}_{[-\pi, 0[}(\theta) \int_0^\theta -f_{p^*}(\pi + t) dt \end{aligned}$$

This function is even and convex as $t \mapsto f_{p^*}(t - \pi)$ and $t \mapsto -f_{p^*}(\pi + t)$ are increasing on $[0, \pi[$ and $[-\pi, 0[$ respectively. Now, the hypothesis on f implies that $k_{\mu_{p^*}}(-\pi) = \lim_{\theta \rightarrow \pi} k_{\mu_{p^*}}(\theta) = \frac{1}{2}$. Moreover, there is a $0 < t_0 < \pi$ such that $f(\theta) < \frac{1}{2\pi}$ for all $t_0 < \theta < \pi$ and $-\pi < \theta < -t_0$. This yields that $\frac{|t|}{2\pi} - k_{\mu_{p^*}}(t)$ is strictly positive for all $t \in]-\pi, 0[\cup]0, \pi[$ and vanishes at $t = 0$ and $t = -\pi$. Then, Statement 3 of Theorem 3.8 is satisfied and the claim is proved. $\square$

A lot of classical probability distributions used in circular data analysis follow the hypothesis of Lemma 3.11: von Mises distribution, wrapped normal distribution, geodesic normal distribution [Pen06], etc...

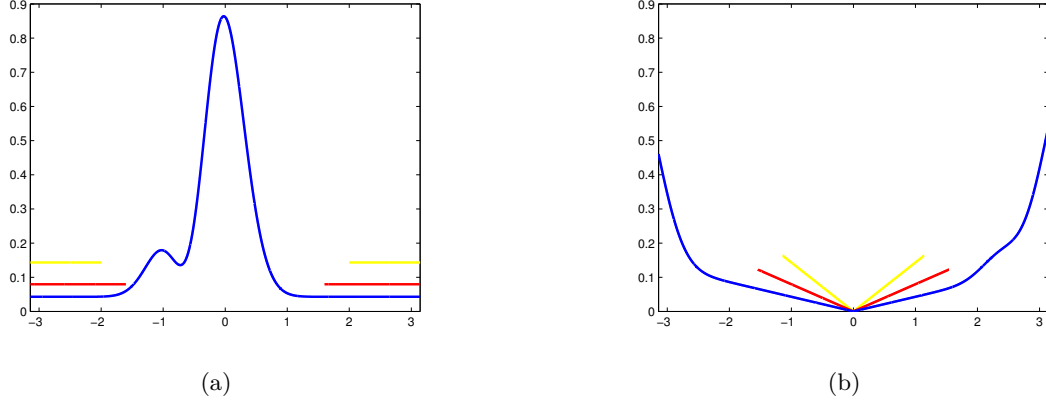


Figure 3.4: An example of distribution satisfying $P(\alpha, \varphi)$. (a) Plot of the density f_p in blue with the bounds $P(p, 0.1, 2)$ in yellow and $P(p, 0.5, 1.6)$ in red. (b) Plot of the function $\theta \mapsto \mathbb{1}_{[0, \pi[}(\theta) \int_0^\theta f_p(t - \pi) dt + \mathbb{1}_{[-\pi, 0[}(\theta) \int_0^\theta -f_p(\pi + t) dt$ in blue with the bounds corresponding to $P(p, 0.1, 2)$ in yellow and $P(p, 0.5, 1.6)$ in red.

In the rest of this section, we present a concentration criterion that do not impose bounds on the support nor the symmetry of the density. Let us introduce the following definition:

Definition 3.12. Let $f : \mathbb{S}^1 \rightarrow \mathbb{R}^+$ be a probability density, $p \in \mathbb{S}^1$, $\alpha \in]0, 1]$ and $\varphi \in]0, \pi[$. We say that f satisfies the property $P(p, \alpha, \varphi)$ if for all $|\theta| \geq \varphi$

$$\frac{1}{2\pi} - f_p(\theta) \geq \frac{\alpha}{2\pi}, \quad (3.12)$$

where $f_p = f \circ e_p : [-\pi, \pi[\rightarrow \mathbb{R}^+$. Moreover, we say that $f \in P(\alpha, \varphi)$ if there is a $p \in \mathbb{S}^1$ such that f satisfies $P(p, \alpha, \varphi)$.

The parameters α and φ control the concentration of μ around p . The idea is to control the mass lying in the complementary of the ball $B(p, \varphi)$. Note that is equivalent to the fact that $f_p(\theta) \leq \frac{1-\alpha}{2\pi}$ for $|\theta| \geq \varphi$, see Figure 3.4. We have the following properties :

Lemma 3.13. *Let $f : \mathbb{S}^1 \rightarrow \mathbb{R}$ be a probability density on the circle. Then*

1. $P(p, \alpha_1, \varphi_1) \implies P(p, \alpha_2, \varphi_2)$ if $\alpha_1 \geq \alpha_2$ and $\varphi_1 \leq \varphi_2$.
2. Let $p_1, p_2 \in \mathbb{S}^1$ and $\varphi < \frac{\pi}{2}$. If $d_{\mathbb{S}^1}(p_1, p_2) < \pi - \varphi$ then $P(p_1, \alpha, \varphi) \implies P(p_2, \alpha, \varphi + d_{\mathbb{S}^1}(p_1, p_2))$.
3. If f satisfies $P(p, \alpha, \varphi)$ then $|m(\mu_p)| \leq \varphi + \frac{1-\alpha}{4\pi}(\pi - \varphi)^2$.

Proof. The first proposition is obvious in view of the Definition 3.12.

To prove the second claim suppose that $0 \leq \theta_{p_2}^{p_1} \leq \pi - \varphi$ (the other case is similar) and write

$$f_{p_1}(\theta) = f_{p_2}(\theta - \theta_{p_2}^{p_1})\mathbb{1}_{[\theta_{p_2}^{p_1}-\pi, \pi[}(\theta) + f_{p_2}(\theta - \theta_{p_2}^{p_1} + 2\pi)\mathbb{1}_{[-\pi, \theta_{p_2}^{p_1}-\pi[}(\theta).$$

In particular, it implies that $\theta_{p_2}^{p_1} - \pi \leq -\varphi$ and since $P(p_1, \alpha, \varphi)$ holds, we have $f_{p_2}(t) \leq \frac{1-\alpha}{2\pi}$, if $t \geq \varphi - \theta_{p_2}^{p_1}$ or $t \leq -\varphi - \theta_{p_2}^{p_1}$. This is equivalent to the fact that $P(p_2, \alpha, \min\{|\varphi - \theta_{p_2}^{p_1}|, |\varphi - \theta_{p_2}^{p_1}|\}) = P(p_2, \alpha, \varphi + \theta_{p_2}^{p_1})$ holds. The case $\varphi - \pi \leq \theta_{p_2}^{p_1} \leq 0$ is similar and we have $P(p_2, \alpha, \varphi - \theta_{p_2}^{p_1})$. Finally recall that $|\theta_{p_2}^{p_1}| = d_{\mathbb{S}^1}(p_1, p_2)$ and the property is proved.

For the last proposition we show that the upper bound is attained for $\mu_p = (1 - \frac{\pi-\varphi}{2\pi})\delta_\varphi + (1 - \alpha)\mathbb{1}_{[\varphi, \pi[}d\lambda$, i.e. 'push the mass as far as possible'. First, it is necessary to consider only the case where μ_p has its support on $[0, \pi[$. Indeed, $\mu_p = \omega\mu_p^- + (1 - \omega)\mu_p^+$ where $\mu_p^-([-\pi, 0]) = \mu_p^+([0, \pi]) = 1$ and $0 \leq \omega \leq 1$. It yields that $m(\mu_p) = \int_{\mathbb{R}} td(\omega\mu_p^- + (1 - \omega)\mu_p^+) = \int_{\mathbb{R}^+} td(-\omega\mu_p^- + (1 - \omega)\mu_p^+) \leq \int_{\mathbb{R}^+} td\mu_p^+$. Then, if the density f_p of μ_p has its support in $[0, \pi[$ we have

$$\begin{aligned} m(\mu_p) &\leq \varphi \left(1 - \int_{\varphi}^{\pi} f_p(t) dt \right) + \int_{\varphi}^{\pi} t f_p(t) dt \\ &= \varphi + \int_{\varphi}^{\pi} (t - \varphi) f_p(t) dt \leq \varphi + \frac{1-\alpha}{2\pi} \int_{\varphi}^{\pi} (t - \varphi) dt, \end{aligned}$$

which gives the result. $\square$

If the density f is sufficiently concentrated around a critical point p^* of F_μ then, this point is the Fréchet mean of μ . More precisely we have the following,

Proposition 3.14. *Let μ be a probability measure with density $f : \mathbb{S}^1 \rightarrow \mathbb{R}^+$. Consider a critical point p^* of F_μ , i.e. a point satisfying $m(\mu_{p^*}) = \int_{-\pi}^{\pi} t f_{p^*}(t) dt = 0$. If f satisfies $P(p^*, \alpha, \varphi)$ with $\alpha \in]0, 1]$ and $0 < \varphi < \varphi_\alpha = \pi \frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}$ then, μ admits a well defined Fréchet mean at p^* .*

Before the proof we make two remarks about Proposition 3.14. For all $\alpha \in]0, 1]$ we have $\varphi_0 = 0 \leq \varphi_\alpha < \frac{\pi}{2} = \varphi_1$. Note that if $\alpha = 1$ then μ has its support included in the ball $B(p^*, \frac{\pi}{2})$. When $\alpha < 1$ the support of μ can be the entire circle $\mathbb{S}^1$. Now if α is small, the density f_{p^*} is allowed to

approach 'from below' the uniform density on $[-\pi, \pi[\setminus [-\varphi_\alpha, \varphi_\alpha]$ and f_{p^*} can be greater than $\frac{1}{2\pi}$ only in the 'small' interval $[-\varphi_\alpha, \varphi_\alpha]$.

This result can also be used to generate absolutely continuous probability distribution on the circle with a given Fréchet mean. Proceed as follows: fix a $p^* \in \mathbb{S}^1$ and choose a function $f_{p^*} : [-\pi, \pi[\rightarrow \mathbb{R}^+$ satisfying $\int_{-\pi}^{\pi} f_{p^*}(t) dt = 1$, $\int_{-\pi}^{\pi} t f_{p^*}(t) dt = 0$ and such that equation (3.12) holds for $\alpha \in]0, 1]$ and $\varphi_\alpha = \pi \frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}$. Then the probability measure μ on the circle with density $f = f_{p^*} \circ e_{p^*}^{-1}$ has a Fréchet mean at p^* .

Proof. As the measure μ admits a density, F_μ is twice differentiable and p^* is a critical point of F_μ since $m(\mu_{p^*}) = 0$, see Corollary 3.3. Moreover Lemma 3.9 ensures that the second derivative of $F_{\mu_{p^*}}$ is equal to $\frac{d^2}{d\theta^2} G_{\mu_{p^*}}(\theta) = 1 - 2\pi f(-\pi + \theta)$, if $0 \leq \theta < \pi$, and $\frac{d^2}{d\theta^2} G_{\mu_{p^*}}(\theta) = 1 - 2\pi f(\pi + \theta)$, if $-\pi \leq \theta < 0$. Thus by hypothesis, the function $F_{\mu_{p^*}}$ is convex on $[-\pi + \varphi, \pi - \varphi]$ and has a unique minimum at 0. Let us show that it is the only argmin of $F_{\mu_{p^*}}$. Let $\theta \in [\pi - \varphi, \pi[$, we have

$$G_{\mu_{p^*}}(\theta) = G_{\mu_{p^*}}(\pi - \varphi) + \int_{\pi - \varphi}^{\theta} t - 2\pi \mu_{p^*}([-\pi, -\pi + t]) dt$$

By hypothesis f satisfies $P(p^*, \alpha, \varphi)$ which implies that $G_{\mu_{p^*}}(\pi - \varphi) \geq \frac{\alpha}{2}(\pi - \varphi)^2$. The second term is bounded from below by $\int_{\pi - \varphi}^{\theta} t - 2\pi \nu([-\pi, -\pi + t]) dt$ where $\nu = \frac{1}{2}(\delta_{\pi - \varphi} + \delta_{\varphi - \pi})$, that is

$$\begin{aligned} G_{\mu_{p^*}}(\theta) &\geq \frac{\alpha}{2}(\pi - \varphi)^2 + \int_{\pi - \varphi}^{\theta} (t - \pi) dt \\ &\geq \frac{1}{2}((\alpha - 1)(\pi - \varphi)^2 + 2\pi(\pi - \varphi) - \pi^2) \end{aligned} \quad (3.13)$$

The right hand side of the preceding inequality is strictly positive if $\varphi < \varphi_\alpha = \pi \frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}$. Similarly, for $\theta \in [-\pi, -\pi + \varphi[$ $G_{\mu_{p^*}}(\theta) > 0$. The result now follows by Theorem 3.8. $\square$

To use the preceding Proposition, we need to know a critical point p^* of F_μ , that is we need to localize the Fréchet mean. This condition is not very realistic in practice and to relax it, one needs more concentration than in the preceding Proposition. Indeed, if there is a $p \in \mathbb{S}^1$ with f_p sufficiently concentrated, then there is a critical point in the neighborhood of p . Moreover, this critical point is the Fréchet mean of μ .

Theorem 3.15. *Let $\delta \in]0, \frac{1}{2}[$ and α_δ be the square of the root of $(5 - 6\delta + \delta^2)X^3 + (1 - \delta^2)X^2 - (2\delta + 1)X - 1$ that lies in $]0, 1]$. If μ is a probability measure with density $f \in P(\alpha, \varphi)$ (see Definition 3.12) with $\alpha_\delta \leq \alpha \leq 1$ and $\varphi \leq \delta\varphi_\alpha = \delta\pi \frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}$ then μ admits a well defined Fréchet mean.*

This result gives a functional class of densities that admit a well defined Fréchet mean. The parameter δ controls the concentration of f via the inequality $\alpha_\delta < \alpha$ and $\varphi \leq \delta\varphi_\alpha$. There is a tradeoff between α and the possible value of φ . The smaller α is (i.e the less f is concentrated) the smaller φ must be (i.e we need to control the value of the density on a bigger interval). As a typical example, take $\delta = \frac{1}{3}$. In this case $\alpha_\delta = \alpha_{\frac{1}{3}} < 0.69$ and $\delta\varphi_{\alpha_\delta} = \frac{1}{3}\varphi_{\alpha_{\frac{1}{3}}} \leq 0.47$. If $f \in P(\alpha, \varphi)$ with $0.69 < \alpha \leq 1$ and $\varphi \leq 0.48$ then there is a well defined Fréchet mean. In Tabular 3.1 we give some numerical values. Note that the column corresponding to $\delta = 0$ is given as a reference only as the set $P(\alpha_\delta, \delta\varphi_\alpha)$ is empty for this values of δ .

$\delta =$	0	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{2}$
$\alpha_\delta \leq$	0.39	0.46	0.54	0.69	1
$\delta\varphi_{\alpha_\delta} \geq$	0	0.12	0.26	0.47	$\frac{\pi}{4}$

 Table 3.4: Some values of α_δ and $\delta\varphi_{\alpha_\delta}$ depending on $\delta \in]0, \frac{1}{2}[$.

Proof. If we show that under the hypothesis of the Theorem, there is a critical point p^* of F_μ satisfying $d_{\mathbb{S}^1}(p, p^*) \leq (1 - \delta)\varphi_\alpha$ where p is a point such that f satisfies $P(p, \alpha, \varphi)$, then, by Lemma 3.13, f will satisfy $P(p^*, \alpha, \delta\varphi_\alpha + (1 - \delta)\varphi_\alpha) = P(p^*, \alpha, \varphi_\alpha)$ and Proposition 3.14 will ensure that p^* is the Fréchet mean of μ .

Hence the rest of the proof is devoted to show that there is a $p^* \in \mathbb{S}^1$ such that $\frac{d}{d\theta} F_{\mu_p}(\theta_{p^*}^p) = 0$ with $d_{\mathbb{S}^1}(p, p^*) \leq (1 - \delta)\varphi_\alpha$. Suppose that $m(\mu_p) \geq 0$ (the case $m(\mu_p) < 0$ is similar). We have

$$\frac{d}{d\theta} F_{\mu_p}(0) = -m(\mu_p) \leq 0.$$

Then, remark that

$$\frac{d}{d\theta} F_{\mu_p}((1 - \delta)\varphi_\alpha) = (1 - \delta)\varphi_\alpha - 2\pi\mu_p([-\pi, -\pi + (1 - \delta)\varphi_\alpha]) - m(\mu_p).$$

We have $-2\pi\mu_p([-\pi, -\pi + (1 - \delta)\varphi_\alpha]) \geq (\alpha - 1)(1 - \delta)\varphi_\alpha$ since f satisfies the $P(p, \alpha, \delta\varphi_\alpha)$ and that $[-\pi + (1 - \delta)\varphi_\alpha] \geq \delta\varphi_\alpha$. Moreover, $-m(\mu) \geq -|m(\mu_p)|$ which is controlled Lemma 3.13 Statement 3. It gives,

$$\begin{aligned} \frac{d}{d\theta} F_{\mu_p}((1 - \delta)\varphi_\alpha) &\geq (1 - \delta)\alpha\varphi_\alpha - \delta\varphi_\alpha - \frac{1 - \alpha}{4\pi} (\pi - \delta\varphi_\alpha)^2 \\ &= \pi \frac{(5 - 6\delta + \delta^2)\alpha\sqrt{\alpha} + (1 - \delta^2)\alpha - (2\delta + 1)\sqrt{\alpha} - 1}{1 + \sqrt{\alpha}} \end{aligned}$$

This quantity is positive as soon as $1 \geq \alpha > \alpha_\delta$, where $\sqrt{\alpha_\delta}$ is the root of the polynomial $X \mapsto (5 - 6\delta + \delta^2)X^3 + (1 - \delta^2)X^2 - (2\delta + 1)X - 1$ that lies in $]0, 1]$. Numerical experiment shows that the function $\delta \mapsto \alpha_\delta$ takes its value in $]0.39, 1[$ for $\delta \in]0, \frac{1}{2}[$. Nevertheless it is easy to see that $\alpha_{\frac{1}{2}} = 1$. Since the derivative of F_μ is continuous, the intermediate value Theorem ensures that there is a critical point such that $|\theta_{p^*}^p| \leq (1 - \delta)\varphi_\alpha$. This is what we need to complete the proof. $\square$

3.6. Fréchet mean of an empirical measure

3.6.1. Consistency of the empirical Fréchet mean

Let $X_1, \dots, X_n$ be independent and identically distributed random variables with value in $(\mathbb{S}^1, d_{\mathbb{S}^1})$ and of probability distribution μ . The empirical measure is defined as usual by $\mu^n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. Following [BP03], we call empirical Fréchet mean set the set of argmins of

$$p \mapsto F_{\mu^n}(p) = \frac{1}{2n} \sum_{i=1}^n d_{\mathbb{S}^1}^2(p, X_i), \quad p \in \mathbb{S}^1.$$

If the argmin of F_{μ^n} is unique it is called the empirical Fréchet mean, and will be denoted by p_n^* . In [Zie77] a strong law of large number is given for the empirical Fréchet mean in a semi metric space which is the case of $(\mathbb{S}^1, d_{\mathbb{S}^1})$. If μ admits a well defined Fréchet mean, any measurable choice in the empirical Fréchet mean set of μ^n is a consistent estimator of p^* . In particular if p_n^* exists for each $n \in \mathbb{N}$, the empirical Fréchet mean is a consistent estimator of the Fréchet mean. Nevertheless, the empirical Fréchet mean is well defined almost surely for a wide class of probability measures. The following fact is from [BP03] Remark 2.6:

Lemma 3.16. *Let μ be a non atomic probability measure on the circle, i.e satisfying $\mu(\{p\}) = 0$ for all $p \in \mathbb{S}^1$. Then for all $n \in \mathbb{N}$ the empirical Fréchet mean exists almost surely.*

Lemma 3.16 shows that the empirical Fréchet mean p_n^* of a probability measure can be computed even if this measure does not possess a well defined Fréchet mean. If the Fréchet mean p^* of μ is well defined, we study the rate of convergence of the empirical Fréchet mean p_n^* to p^* . To this end, we derive a concentration inequality that shows the consistency of the empirical Fréchet mean.

Proposition 3.17. *Let μ be a measure with density $f : \mathbb{S}^1 \rightarrow \mathbb{R}$ that admits a well defined Fréchet mean p^* . Then there exists a strictly increasing function $\rho : [0, \pi[\rightarrow \mathbb{R}^+$ such that for all $p \in \mathbb{S}^1$*

$$F_\mu(p) \geq \rho(d_{\mathbb{S}^1}(p, p^*)).$$

If p_n^ denotes the empirical Fréchet mean, we have for all $x > 0$*

$$\mathbb{P}\left(\rho(d_{\mathbb{S}^1}(p_n^*, p^*)) \geq C(s)\sqrt{\frac{x}{n}}\right) \leq 2e^{-x}.$$

where $s = \max\{|x - y|, x, y \in \text{supp}(\mu)\}$ and $C(s) = (4\pi^2 + 4\pi^2 s + 2s) \leq 4\pi(2\pi^2 + \pi + 1)$.

The function ρ in the statement of the preceding Proposition 3.17 determines the rate of convergence of p_n^* to p^* . Indeed, the rate of convergence of p_n^* to p^* will depend on how fast $\rho^{-1}(t)$ is going to 0 when $t \rightarrow 0$. For example, when the support of μ is strictly included in an hemisphere, then ρ is a polynomial of order 2. Below, we will see that under the assumptions of Theorem 3.15 we keep similar asymptotic rates of convergence but without this bound on the support.

Proof. The first claim about the lower bound ρ is a direct consequence of the following Lemma:

Lemma 3.18. *Let $f : [-\pi, \pi[\rightarrow \mathbb{R}^+$ be a continuous function on $[-\pi, \pi[$ which satisfies $\lim_{\theta \rightarrow \pi^-} f(\theta) = \lim_{\theta \rightarrow -\pi^+} f(\theta) = f(-\pi)$. Suppose that f vanishes at a unique point $\theta_0 \in [-\pi, \pi[$. Then, there exists a strictly increasing function $\rho : [0, \pi[\rightarrow \mathbb{R}^*$ such that for all $\theta \in [-\pi, \pi[$,*

$$f(\theta) \geq \rho(d_{\mathbb{S}^1}(\theta_0, \theta)).$$

Proof. As the function f is periodic we can assume, without loss of generality, that $\theta_0 = 0$. Define the function $g : [-\pi, \pi[\rightarrow \mathbb{R}^+$ by

$$g(\theta) = \min \left\{ f(-\theta), f(\theta), \min_{|t| > \theta} f(t) \right\}.$$

Remark that the considered minima are attained as f is a continuous function on $\mathbb{R}/(2\pi\mathbb{Z})$ which is compact. We have $g \leq f$. The function g is even, is increasing on $[0, \pi[$ and vanishes only in 0. We now have to construct an even, positive, strictly increasing function that bounds g from below. Consider for all $\theta \in [-\pi, \pi[$,

$$G(\theta) = \frac{1}{|\theta|} \int_0^\theta g(t) dt.$$

We have $G(\theta) \leq g(\theta)$ for all $\theta \in [0, \pi[$ since g is an increasing function on $[0, \pi[$. Moreover G is even and we have $G \leq g$ on $[-\pi, \pi[$. The function G is strictly increasing since its derivative and $g - G$ have the same sign. Now $g(\theta) - G(\theta) = 0$ if and only if $g(t) = g(0)$ for all $t \in]0, \theta]$ which is impossible by the construction of g . We conclude the proof of Lemma 3.18 by setting $G(\theta) = \rho(|\theta|)$. $\square$

We now focus on the concentration inequality of Proposition 3.17. The proof is divided in two steps. First we show the uniform convergence in probability of $F_{\mu_p^n}$ to F_{μ_p} . Then, we deduce the convergence of their argmins by using the lower bound given by ρ . Recall the notations, $p^* = \operatorname{argmin}_{p \in \mathbb{S}^1} F_\mu(p)$ and $p_n^* = \operatorname{argmin}_{p \in \mathbb{S}^1} F_{\mu_p^n}(p)$. We fix an arbitrary $p \in \mathbb{S}^1$ and in normal coordinate centered at p we have for all $\theta \in [-\pi, \pi[$

$$F_{\mu_p}(\theta) = \int_{\mathbb{R}} d_{\mathbb{S}^1}^2(\theta, t) d\mu_p(t) \quad \text{and} \quad F_{\mu_p^n}(\theta) = \int_{\mathbb{R}} d_{\mathbb{S}^1}^2(\theta, t) d\mu_p^n(t).$$

Let $\theta_{p^*} = \theta_{p^*}^p = \operatorname{argmin}_{\theta \in \mathbb{R}} F_{\mu_p}(\theta)$ and $\theta_{p_n^*} = \theta_{p_n^*}^p = \operatorname{argmin}_{\theta \in \mathbb{R}} F_{\mu_{p_n^*}}(\theta)$.

For all $\theta \in [-\pi, \pi[$, let $H(\theta) = \mu_p([-\pi, \theta])$ be the cumulative distribution function (c.d.f) of μ_p and $H_n(\theta) = \mu_p^n([-\pi, \theta])$ be and the empirical c.d.f. Recall Proposition 3.2 where we have shown that $\frac{d}{d\theta} F_{\mu_p}(\theta) = \theta - 2\pi H(-\pi + \theta) - m(\mu_p)$ and $\frac{d}{d\theta} F_{\mu_p^n}(\theta) = \theta - 2\pi H_n(-\pi + \theta) - m(\mu_p^n)$ if $\theta \in [0, \pi[$ and $\frac{d}{d\theta} F_{\mu_p}(\theta) = \theta + 2\pi - 2\pi H(\pi + \theta) - m(\mu_p)$ and $\frac{d}{d\theta} F_{\mu_p^n}(\theta) = \theta + 2\pi - 2\pi H_n(\pi + \theta) - m(\mu_p^n)$ if $\theta \in [-\pi, 0[$. Then,

$$\begin{aligned} 2 \sup_{\theta \in \mathbb{R}} |F_{\mu_p^n}(\theta) - F_{\mu_p}(\theta)| &\leq 2\pi \sup_{\theta \in \mathbb{R}} \left| \frac{d}{d\theta} F_{\mu_p^n}(\theta) - \frac{d}{d\theta} F_{\mu_p}(\theta) \right| + 2 |F_{\mu_p^n}(0) - F_{\mu_p}(0)| \\ &\leq 4\pi^2 \sup_{\theta \in [-\pi, \pi[} |H(\theta) - H_n(\theta)| + 4\pi^2 |m(\mu_p) - m(\mu_p^n)| \end{aligned} \quad (3.14)$$

$$+ 2 |m_2(\mu_p) - m_2(\mu_p^n)| \quad (3.15)$$

where $m_2(\nu) = \int_{\mathbb{R}} t^2 d\nu(t)$, for a measure ν on $\mathbb{R}$. The first term of the preceding upper bound can be controlled in probability using the Dvoretzky-Kiefer-Wolfowitz inequality (see e.g [Mas90]), we have for all $x > 0$,

$$\mathbb{P} \left(4\pi^2 \sup_{\theta \in [-\pi, \pi[} |H(\theta) - H_n(\theta)| \geq 4\pi^2 \sqrt{\frac{x}{n}} \right) \leq 2e^{-x}.$$

For the second and third term which involve the first and second moment of μ_p and μ_p^n , we use an Hoeffding type inequality which gives for all $x > 0$, $\mathbb{P} \left(4\pi^2 |m(\mu_p) - m(\mu_p^n)| + 2 |m_2(\mu_p) - m_2(\mu_p^n)| \geq s(4\pi^2 + 2s) \sqrt{\frac{x}{n}} \right) \leq 2e^{-x}$, where $s = \max\{|x - y|, x, y \in \operatorname{support} \mu\}$ is the diameter of the support of μ . Combining the two concentration inequalities and (3.15), it yields for all $x > 0$,

$$\mathbb{P} \left(2 \sup_{\theta \in \mathbb{R}} |F_{\mu_p}(\theta) - F_{\mu_p^n}(\theta)| \geq (4\pi^2 + 4\pi^2 s + 2s) \sqrt{\frac{x}{n}} \right) \leq 2e^{-x}. \quad (3.16)$$

Now that the uniform convergence in probability is shown, we use a classical inequality in M-estimation, $F_{\mu_p}(\theta_{p_n^*}) - F_{\mu_p}(\theta_{p^*}) \leq 2 \sup_{\theta \in \mathbb{R}} |F_{\mu_p^n}(\theta) - F_{\mu_p}(\theta)|$. By Lemma 3.18, there exists an increasing function $\rho : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $F(\theta_{p_n^*}) - F(\theta_{p^*}) \geq \rho(d_{\mathbb{S}^1}(\theta_{p_n^*}, \theta_{p^*}))$. Plugging this in equation (3.16) we have,

$$\mathbb{P}\left(\rho(d_{\mathbb{S}^1}(\theta_{p_n^*}, \theta_{p^*})) \geq (4\pi^2 + 4\pi^2 s + 2s)\sqrt{\frac{x}{n}}\right) \leq 2e^{-x},$$

and the proof of Proposition 3.17 is completed. $\square$

Function ρ that appears in the statement of Proposition 3.17 can be explicitly computed if the density $f : \mathbb{S}^1 \rightarrow \mathbb{R}^+$ satisfies property $P(p, \alpha, \varphi)$ for some $p \in \mathbb{S}^1$, see Definition 3.12. The parameter $\alpha \in]0, 1]$ can be interpreted as a measure of the convexity of F_{μ_p} on the interval $[-\varphi, \varphi]$. For example, if $\alpha = 1$ and $\varphi = \varphi_\alpha = \frac{\pi}{2}$, then μ has its support contained in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and F_{μ_p} is quadratic on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ with a second derivative equals to 1.

Proposition 3.19. *Let μ be a probability measure with density $f : \mathbb{S}^1 \rightarrow \mathbb{R}^+$ satisfying the hypothesis of Theorem 3.15. Then, for all $x > 0$ we have*

$$\mathbb{P}\left(d_{\mathbb{S}^1}(p_n^*, p^*) \geq \sqrt{B(\alpha, \varphi)} \left(\frac{x}{n}\right)^{\frac{1}{4}}\right) \leq 2e^{-x},$$

where $B(\alpha, \varphi) = C \max\left\{\frac{\pi^2}{\gamma(\alpha, \varphi)}, \frac{2}{\alpha}\right\}$ with $\gamma(\alpha, \varphi) = \frac{1}{2}((\alpha - 1)(\pi - \varphi)^2 + 2\pi(\pi - \varphi) - \pi^2)$ and $C = 4\pi(2\pi^2 + \pi + 1)$.

Proof. This result follows from Proposition 3.17 and we only have to find a strictly increasing function $\rho : [0, \pi] \rightarrow \mathbb{R}^+$ satisfying for all $\theta \in [-\pi, \pi[$, $F_{\mu_p}(\theta) - F_{\mu_p}(p^*) \geq \rho(d_{\mathbb{S}^1}(p, p^*))$. By Theorem 3.15, the Fréchet mean p^* of μ exists and Lemma 3.18 ensures that there is a strictly increasing function ρ which satisfies,

$$F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) \geq \rho(|\theta|).$$

for all $\theta \in [-\pi, \pi[$. As μ_{p^*} admits a density f_{p^*} , the Fréchet functional $F_{\mu_{p^*}}$ is twice differentiable. Moreover f satisfies $P(p^*, \alpha, \varphi_\alpha)$, see proof of Theorem 3.15. For all $\theta \in [-\pi + \varphi_\alpha, \varphi_\alpha - \pi]$, a second order Taylor expansion of $F_{\mu_{p^*}}$ at 0 ensures that for some $\tilde{\theta} \in [-\pi + \varphi_\alpha, \pi - \varphi_\alpha]$,

$$F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) = \frac{1}{2}\theta^2 \frac{d^2}{d\theta^2} F_{\mu_{p^*}}(\tilde{\theta}) \geq \frac{\alpha}{2}\theta^2.$$

The last inequality is a direct consequence of property $P(p^*, \alpha, \varphi_\alpha)$ as we have $\frac{d^2}{d\theta^2} F_{\mu_p}(\theta) = 1 - 2\pi f(-\pi + \theta)$, if $0 \leq \theta < \pi$ and $\frac{d^2}{d\theta^2} F_{\mu_p}(\theta) = 1 - 2\pi f(\pi + \theta)$, if $-\pi \leq \theta < 0$. For all $\theta \in [-\pi, -\varphi[\cup]\varphi, \pi[$ we have by inequality (3.13),

$$F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) \geq \frac{1}{2}((\alpha - 1)(\pi - \varphi)^2 + 2\pi(\pi - \varphi) - \pi^2) = \gamma(\alpha, \varphi) > 0.$$

Then, let $\rho(t) = t^2 \min\left\{\frac{\gamma(\alpha, \varphi)}{\pi^2}, \frac{\alpha}{2}\right\}$ and the proof is completed. $\square$

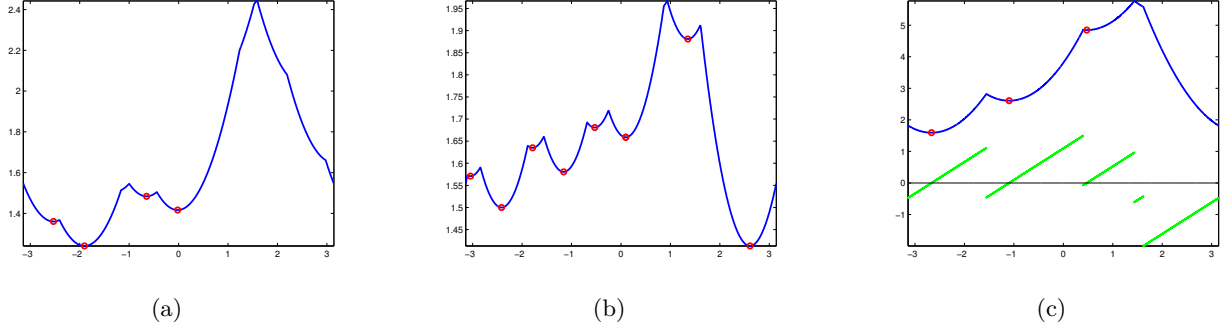


Figure 3.5: (a) and (b) Plots of F_{μ^n} where $n = 10$ and $\mu^n = \sum_{i=1}^n \delta_{X_i}$ with the X_i 's i.i.d of uniform law λ . (c) Same plot but with $n = 5$ and the plot of the derivative in green. The critical points are given by the intersection between the green curve and the x -axis in black. In (a), (b) and (c) the red dots are the local minima computed with the described algorithm.

3.6.2. Computation of the Empirical Fréchet mean

Computation of the Fréchet mean of a general probability measure may not be an easy task as it is a *global* optimization problem. In practice the Fréchet functional is not a convex function and a gradient descent algorithm will only give a *local* minimum which depends on the initialization point chosen.

In the following we will use the results of Section 3.4.2 to derive an algorithm to compute the empirical Fréchet mean. Recall that the regular critical points (i.e no cusp point) of F_{μ} are the local minima of F_{μ} . Moreover, Corollary 3.7 gives us a simple mean to compute them by solving at most n affine relations, see equation (3.9). In a coordinate system centered at some $p \in \mathbb{S}^1$, it amounts to compute the cumulative distribution function of μ^n which is, here, piecewise constant with jumps of size $\frac{1}{n}$. Indeed, we have,

$$\mu_p^n([- \pi, t]) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[- \pi, t]}(\theta_{X_i}^p) = \frac{1}{n} \text{Card}\{\theta_{X_i}^p < t\}.$$

Note, that in practice, there are less than n solutions, see e.g. Figure 3.5. The following algorithm takes as input the values $\{X_i\}_{i=1}^n$ and returns the Fréchet mean of $\mu^n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. See also Figure 3.5(c) for an illustration.

Initialization: Choose an arbitrarily point $p \in \mathbb{S}^1$, compute the coordinates $\{\theta_{X_i}^p\}_{i=1}^n$ and reorder them in increasing order.

We denote $\tau_0^- = -\pi \leq \tau_1^- \leq \tau_2^- \leq \dots \leq \tau_{n_1}^- < 0 = \tau_{n_1+1}^-$ the n_1 negative sorted terms and $\tau_0^+ = \pi > \tau_1^+ \geq \dots \geq \tau_{n_2}^+ \geq 0 = \tau_{n_2+1}^+$ the $n_2 = n - n_1$ positive sorted terms.

Compute the mean $m(\mu_p^n) = \frac{1}{n}(\tau_1^- + \dots + \tau_{n_1}^- + \tau_1^+ + \dots + \tau_{n_2}^+)$ and initialize $\theta_{p^*}^p$ to 0, says.

Step 1: # We compare all the local minima in $[0, \pi[$.

For i from 0 to n_1 do

$\theta_{p^*,new}^p$ is the candidate to be a critical point between τ_i^- and τ_{i+1}^-

Let $\theta_{p^*,new}^p = 2\pi \frac{i}{n} + m(\mu_p^n)$

```
# Verify if  $\theta_{p^*,new}^p$  is a critical point and test its value
  if  $\tau_i^- + \pi \leq \theta_{p^*,new}^p \leq \tau_{i+1}^- + \pi$  and  $F_{\mu_p^n}(\theta_{p^*,new}^p) \leq F_{\mu_p^n}(\theta_{p^*}^p)$  then  $\theta_{p^*}^p := \theta_{p^*,new}^p$ 
end for
```

Step 2: # The Step 2 is the same as Step 1 but for local minima in $[-\pi, 0[$

```
For  $i = 0$  to  $n_2$  do
  Let  $\theta_{p^*,new}^p = -2\pi \frac{i}{n} + m(\mu_p^n)$ 
  if  $\tau_{i+1}^+ - \pi \leq \theta_{p^*,new}^p \leq \tau_i^+ - \pi$  and  $F_{\mu_p^n}(\theta_{p^*,new}^p) \leq F_{\mu_p^n}(\theta_{p^*}^p)$  then  $\theta_{p^*}^p := \theta_{p^*,new}^p$ 
end for
```

Output # The value of $\theta_{p^*}^p$ is now the best argmin

```
Return  $p^* = e_p(\theta_{p^*}^p)$ .
```

This algorithm can be extended to more general measures than the empirical one. The approach will be the same: find the critical points of the Fréchet functional with formula of Corollary 3.7. Unfortunately, there may be some computational issues as general cumulative distribution function will be not piecewise constant anymore.

3.7. Conclusion

It is not straightforward to extend criterion such as the one given in Theorem 3.8 to more general spaces, e.g. for the n dimensional sphere $\mathbb{S}^n$. Recall that the circle $\mathbb{S}^1$ is a flat space in the sense that it is locally isometric to the Euclidean space $\mathbb{R}$. Then, the only phenomenon that induces uniqueness issues of the Fréchet mean is the presence of a cut locus. The criterion presented in this note relies on an explicit formula for the gradient of the Fréchet mean. Curvature has an extra effect on the metric and makes difficult to derive exact computation on the Fréchet functional and its gradient. Moreover, it is not clear if the role played by the uniform measure as a benchmark in the well definiteness of the Fréchet mean in $\mathbb{S}^1$ can be extended to n -spheres or non flat manifolds.

The extension of the criterion to general flat manifold is undoubtedly a more tractable program. To conclude we shortly depict the situation on the flat torus of dimension n .

3.7.1. Flat Torus

A flat torus $\mathbb{T}^n$ of dimension n is the Cartesian product $\mathbb{S}^1 \times \dots \times \mathbb{S}^1$ of n sphere of dimension 1. It is endowed with the flat metric given for all $p = (p_1, \dots, p_n)$ and $x = (x_1, \dots, x_n)$ in $\mathbb{T}^n$ by

$$d_{\mathbb{T}^n}(x, p) = \sqrt{\sum_{i=1}^n d_{\mathbb{S}^1}^2(x_i, p_i)}$$

The exponential map at p is $e_p^n : [-\pi, \pi]^n \rightarrow \mathbb{T}^n$ is given by $e_p^n(p) = (e_{p_1}(p_1), \dots, e_{p_n}(p_n))$. The cut locus of a point $p \in \mathbb{T}$ is now the set $\{\{-p^1\} \times \mathbb{S}^1 \times \dots \times \mathbb{S}^1\} \cup \{\mathbb{S}^1 \times \{-p^2\} \times \mathbb{S}^1 \times \dots \times \mathbb{S}^1\} \cup \dots \cup \{\mathbb{S}^1 \times \dots \times \mathbb{S}^1 \times \{-p^n\}\}$. We denote μ_i the i -th marginal measure. Then the gradient of the Fréchet

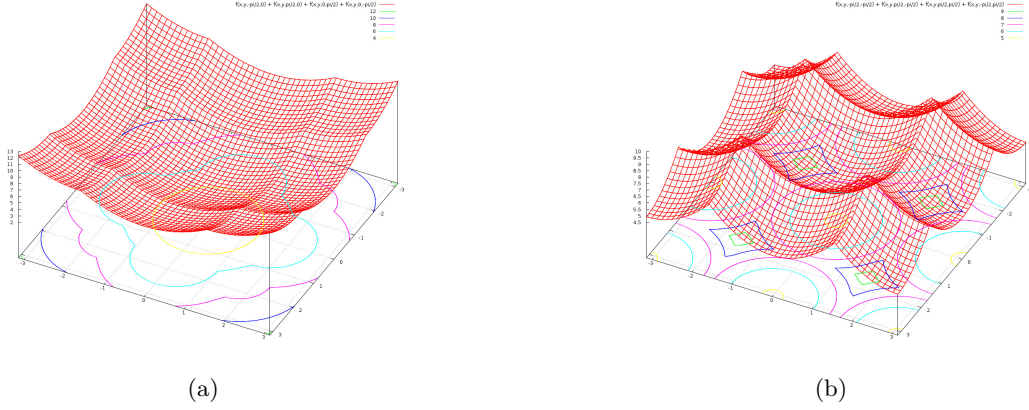


Figure 3.6: Plots of the Fréchet functional F_μ and its level sets for the flat torus $\mathbb{T}^2$. (a) $\mu = \frac{1}{2}(\delta_{(-\frac{\pi}{2},0)} + \delta_{(\frac{\pi}{2},0)})$ with a well defined Fréchet mean at 0. (b) $\mu = \frac{1}{4}(\delta_{(\frac{\pi}{2},\frac{\pi}{2})} + \delta_{(\frac{\pi}{2},-\frac{\pi}{2})} + \delta_{(-\frac{\pi}{2},-\frac{\pi}{2})} + \delta_{(-\frac{\pi}{2},\frac{\pi}{2})})$ with 4 global minima at $(0,0)$, $(-\pi,0)$, $(-\pi,-\pi)$ and $(0,-\pi)$.

functional is given by,

$$\nabla F_{\mu_p}(\theta_1, \dots, \theta_n) = \left[\begin{cases} \theta_i - m(\mu_i) - 2\pi\mu_i([- \pi, -\pi + \theta_i]), & \text{if } -\pi \leq \theta_i < 0 \\ \theta_i - m(\mu_i) + 2\pi\mu_i([\pi + \theta_i, \pi]), & \text{if } 0 \leq \theta_i < \pi \end{cases} \right]_{i=1}^n.$$

To derive a criterion for μ to admit a well defined Fréchet mean, one could use the following formula for all $\theta = (\theta_1, \dots, \theta_n) \in [-\pi, \pi]^n$,

$$G_{\mu_{p^*}}(\theta) = F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) = \int_0^\tau \left\langle \nabla F_{\mu_{p^*}} \left(\frac{\theta}{\|\theta\|} t \right), \frac{\theta}{\|\theta\|} \right\rangle dt.$$

where $\tau = d_{\mathbb{T}^n}(\theta, 0)$ and $\|\cdot\|$ is the standard Euclidean norm in $\mathbb{R}^n$. Hence, the idea is to integrate the gradient vector field ∇F_{μ_p} along the line $\theta\mathbb{R}$ for all $\theta = (\theta_1, \dots, \theta_n) \in [-\pi, \pi]^n$ and μ admits a well defined Fréchet mean p^* if and only if $G_{\mu_{p^*}}$ vanishes only at $\theta = 0$. It seems possible to derive a similar criterion as in the case of the circle which links the existence of the Fréchet mean and the comparison of the marginal law and the uniform measure λ on $[-\pi, \pi[$. We give two examples of Fréchet functional on the torus $\mathbb{T}^2$ of dimension two at Figure 3.6.

Conclusion

Dans ce travail nous avons abordé essentiellement deux axes de recherche autour des statistiques de données observées avec déformations. On donne ici quelques pistes de réflexion laissées ouvertes par ce travail.

Le premier axe concerne la modélisation statistique avec la mise en place de modèles de déformations ainsi que l'étude des objets que l'on peut estimer de manière consistante. Pour poursuivre la réflexion du chapitre 1, une piste possible est l'extension des résultats obtenus ici avec des modèles utilisant des déformations paramétriques (*i.e.* de dimension finie) à des modèles de déformations de dimension infinie. En effet, les modèles réalistes de déformations d'images ou de courbes temporelles utilisent des ensembles de déformations bien plus riches que les groupes de dimensions finie comme le groupe des similarités. L'exemple typique est l'action des difféomorphismes sur les courbes ou images. Dans cette optique, il faut sans doute aborder d'autres techniques que la M-estimation utilisée ici. En effet, les conditions d'identifiabilité décrites dans les chapitres 1 et 2 ne passent pas en dimension infinie où l'action des déformations ne peut être raisonnablement supposée libre. On peut difficilement faire l'hypothèse qu'une image à reconstruire ne possède pas de zones de couleurs uniformes, ces zones étant précisément celles qui caractérisent les groupes d'isotropie de l'action des difféomorphismes.

Une deuxième piste de réflexion, dans la continuité du chapitre 2 peut être la recherche de conditions suffisantes moins restrictives que l'isotropie pour garantir une asymptotique en J (le nombre de courbes observées) et avec k (le nombre de points dans l'échantillonnage) fixé. En effet, les simulations du chapitre 1 et 2 montrent que l'augmentation du nombre de courbes J améliore nettement l'estimation de la forme moyenne. Pour cela, il faudrait étudier précisément les cas de bruit pour lesquelles la consistance n'est pas vérifiée et comprendre comment, sans estimer correctement les paramètres de déformations, on peut estimer la forme moyenne.

Le deuxième axe de recherche de ce travail est orienté vers les problèmes plus géométriques avec l'étude des structures engendrées par les actions de groupes. Ceci a permis une reformulation des problèmes statistiques issus des modèles de déformations et aussi d'étudier plus avant les propriétés de la moyenne de Fréchet. Une continuation possible du chapitre 3 est bien sûr l'extension du critère à des espaces plus généraux. Une autre piste, indépendante, peut-être l'étude de statistiques robustes dans les espaces non-euclidiens. Comment garantir la définition de valeurs centrales stables même en observant des données très dispersées ou des valeurs aberrantes, ce qui est courant en analyse d'images. Une définition de la médiane géométrique a été proposée mais dont il reste à étudier les propriétés.

Bibliographie

- [Afs11] B. Afsari. Riemannian l^p center of mass : existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2) : 655–673, 2011.
- [AKT10] S. Allasonnière, E. Kuhn, and A. Trouvé. Construction of Bayesian deformable models via a stochastic approximation algorithm : a convergence study. *Bernoulli*, 16(3) : 641–678, 2010.
- [Arn95] M. Arnaudon. Barycentres convexes et approximations des martingales continues dans les variétés. In *Séminaire de Probabilités, XXIX*, volume 1613 of *Lecture Notes in Math.*, pages 70–85. Springer, Berlin, 1995.
- [AST07] Allasonnière, Y. S., Amit, and A. Trouvé. Toward a coherent statistical framework for dense deformable template estimation. *Journal of the Statistical Royal Society (B)*, 69 : 3–29, 2007.
- [BF01] S. R. Buss and J. P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *ACM Trans. Graph.*, 20 : 95–126, April 2001.
- [BG10a] J. Bigot and S. Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.*, 38(4) : 2422–2464, 2010.
- [BG10b] J. Bigot and S. Gadat. Smoothing under diffeomorphic constraints with homeomorphic splines. *SIAM J. Numer. Anal.*, 48(1) : 224–243, 2010.
- [BGL09] J. Bigot, S. Gadat, and J.-M. Loubes. Statistical M-estimation and consistency in large deformable models for image warping. *J. Math. Imaging Vision*, 34(3) : 270–290, 2009.
- [BGV09] J. Bigot, F. Gamboa, and M. Vimond. Estimation of translation, rotation and scaling between noisy images using the fourier mellin transform. *SIAM Journal on Imaging Sciences*, 2(2) : 614–645, 2009.
- [Big06a] J. Bigot. Landmark-based registration of curves via the continuous wavelet transform. *J. Comput. Graph. Statist.*, 15(3) : 542–564, 2006.
- [Big06b] J. Bigot. Landmark-based registration of curves via the continuous wavelet transform. *Journal of Computational and Graphical Statistics*, 15(3) : 542–564, 2006.
- [BLV10] J. Bigot, J.-M. Loubes, and M. Vimond. Semiparametric estimation of shifts on compact lie groups for image registration. *Probability Theory and Related Fields*, pages 1–49, 2010. 10.1007/s00440-010-0327-2.

- [BM98] L. Birgé and P. Massart. Minimum contrast estimators on sieves : exponential bounds and rates of convergence. *Bernoulli*, 4(3) : 329–375, 1998.
- [BP03] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds (i). *Annals of statistics*, 31(1) : 1–29, 2003.
- [BP05] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds (ii). *Annals of statistics*, 33 : 1225–1259, 2005.
- [Car46] É. Cartan. *Leçons sur la géométrie des espaces de Riemann*. Les Grands Classiques Gauthier-Villars. Éditions Jacques Gabay, Sceaux, 1946. Reprint of the second edition.
- [Cha06] I. Chavel. *Riemannian geometry*, volume 98 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2006. A modern introduction.
- [CK99] J. M. Corcuera and W. S. Kendall. Riemannian barycentres and geodesic convexity. *Math. Proc. Cambridge Philos. Soc.*, 127(2) : 253–269, 1999.
- [CL09] I. Castillo and J.-M. Loubes. Estimation of the distribution of random shifts deformation. *Math. Methods Statist.*, 18(1) : 21–42, 2009.
- [CMM91] V. Cervera, F. Mascaró, and P. W. Michor. The action of the diffeomorphism group on the space of immersions. *Differential Geom. Appl.*, 1(4) : 391–401, 1991.
- [DF04] D. S. Dummit and R. M. Foote. *Abstract algebra*. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2004.
- [DM98] I. L. Dryden and K. V. Mardia. *Statistical shape analysis*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1998.
- [ÉM91] M. Émery and G. Mokobodzki. Sur le barycentre d’une probabilité dans une variété. In *Séminaire de Probabilités, XXV*, volume 1485 of *Lecture Notes in Math.*, pages 220–233. Springer, Berlin, 1991.
- [Fis93] N. I. Fisher. *Statistical analysis of circular data*. Cambridge University Press, Cambridge, 1993.
- [Fré48] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré, Sect. B, Prob. et Stat.*, 10 : 235–310, 1948.
- [FVJ09] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi. The geometric median on riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1, Supplement 1) : S143–S152, 2009. Mathematics in Brain Imaging.
- [GK92] T. Gasser and A. Kneip. Statistical tools to analyze data representing a sample of curves. *Annals of Statistics*, 20(3) : 1266–1305, 1992.
- [GL95] R. D. Gill and B. Y. Levit. Applications of the Van Trees inequality : a Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2) : 59–79, 1995.
- [GLM07] F. Gamboa, J.-M. Loubes, and E. Maza. Semi-parametric estimation of shifts. *Electron. J. Stat.*, 1 : 616–640, 2007.

-
- [GM01] C. A. Glasbey and K. V. Mardia. A penalized likelihood approach to image warping. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(3) : 465–514, 2001.
 - [GM07] U. Grenander and M. Miller. *Pattern Theory : From Representation to Inference*. Oxford Univ. Press, Oxford, 2007.
 - [Goo91] C. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B*, 53(2) : 285–339, 1991.
 - [Gra06] M. Gray. *Toeplitz and circulant matrices : a review*. Now Publishers Inc, 2006.
 - [Gre63] U. Grenander. *Probabilities on algebraic structures*. John Wiley & Sons Inc., New York, 1963.
 - [Gre93] U. Grenander. *General pattern theory - A mathematical study of regular structures*. Clarendon Press, Oxford, 1993.
 - [Hel01] S. Helgason. *Differential geometry, Lie groups, and symmetric spaces*, volume 34 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. Corrected reprint of the 1978 original.
 - [Hen91] H. Hendriks. A Cramér-Rao type lower bound for estimators with values in a manifold. *J. Multivariate Anal.*, 38(2) : 245–261, 1991.
 - [HHM10a] S. Huckemann, T. Hotz, and A. Munk. Intrinsic manova for riemannian manifolds with an application to kendalls spaces of planar shapes. *IEEE Trans. Patt. Anal. Mach. Intell. Special Section on Shape Analysis and its Applications in Image Understanding*, 32(4) : 593–603, 2010.
 - [HHM10b] S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis : Geodesic principal component analysis for riemannian manifolds modulo lie group actions. discussion paper with rejoinder. *Statistica Sinica*, 20 : 1–100, 2010.
 - [HJ90] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990.
 - [HKCF92] T. J. Hastie, E. Kishon, M. Clark, and J. Fan. A model for signature verification. Technical report, AT&T Bell Laboratories Technical Report, 1992.
 - [Huc10] S. Huckemann. Inference on 3d procrustes means : Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scand. J. Statist.*, To appear, 2010.
 - [Huc11] S. Huckemann. Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.*, 39(2) : 1098–1124, 2011.
 - [JDJG04] S. Joshi, B. Davis, B. Jomier, and G. Guido. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23 : 151–160, 2004.
 - [Kar77] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5) : 509–541, 1977.
 - [KBCL99] D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and shape theory*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1999.

- [Ken84] D. G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.*, 16(2) : 81–121, 1984.
- [Ken90] W. S. Kendall. Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proc. London Math. Soc.* (3), 61(2) : 371–406, 1990.
- [Ken92] W. S. Kendall. The propeller : A counterexample to a conjectured criterion for the existence of certain convex functions. *J. Lond. Math. Soc., II. Ser.*, 46(2) : 364–374, 1992.
- [KG88] A. Kneip and T. Gasser. Convergence and consistency results for self-modelling regression. *Annals of Statistics*, 16 : 82–112, 1988.
- [KL00] A. Kume and H. Le. Estimating Fréchet means in Bookstein’s shape space. *Adv. in Appl. Probab.*, 32(3) : 663–674, 2000.
- [KM97] J. T. Kent and K. V. Mardia. Consistency of Procrustes estimators. *J. Roy. Statist. Soc. Ser. B*, 59(1) : 281–290, 1997.
- [KS08] D. Kaziska and A. Srivastava. The karcher mean of a class of symmetric distributions on the circle. *Statistics & Probability Letters*, 78(11) : 1314–1316, 2008.
- [KSMJ04] E. Klassen, A. Srivastava, W. Mio, and S. H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26 : 372–383, March 2004.
- [Küh02] W. Kühnel. *Differential geometry*, volume 16 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2002. Curves-surfaces-manifolds, Translated from the 1999 German original by Bruce Hunt.
- [Le98] H. Le. On the consistency of procrustean mean shapes. *Adv. in Appl. Probab.*, 30(1) : 53–63, 1998.
- [Le01] H. Le. Locating Fréchet means with application to shape spaces. *Adv. in Appl. Probab.*, 33(2) : 324–338, 2001.
- [Le04] H. Le. Estimation of Riemannian barycentres. *LMS J. Comput. Math.*, 7 : 193–200 (electronic), 2004.
- [Lee00] J. M. Lee. *Introduction to topological manifolds*, volume 202 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.
- [Lee11] J. M. Lee. *Introduction to topological manifolds*, volume 202 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2011.
- [Le193] S. Lele. Euclidean distance matrix analysis (EDMA) : estimation of mean form and mean form difference. *Math. Geol.*, 25(5) : 573–602, 1993.
- [LK00] H. Le and A. Kume. The fréchet mean shape and the shape of the means. *Advances in Applied Probability*, 32 : 101–113, 2000.
- [LM00] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Stat.*, 28(5) : 1302–1338, 2000.

- [LM04] X. Liu and H.G. Muller. Functional convex averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association*, 99(467) : 687–699, 2004.
- [Lor99] G. Lorette. Handwriting recognition or reading ? what is the situation at the dawn of the 3rd millenium ? *International Journal on Document Analysis and Recognition*, 2(1) : 2–12, 1999.
- [Mas90] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.*, 18(3) : 1269–1283, 1990.
- [Mas07] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.
- [Mic] P. W. Michor. Isometric actions of lie groups and invariants. Lecture Notes, 1997.
- [MJ00] K. V. Mardia and P. E. Jupp. *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2000.
- [MM06] P. W. Michor and D. Mumford. Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc. (JEMS)*, 8(1) : 1–48, 2006.
- [MR93] M. K. Murray and J. W. Rice. *Differential geometry and statistics*, volume 48 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993.
- [MSDD06] G. Moise, M. Secrieru, A. Dinescu, and S. Diaconescu. System and methods of acquisition, analysis and authentication of the handwritten signature. Patent WO/2006/085783, 2006.
- [MY01] M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching : A general framework. *International Journal of Computer Vision*, 41 : 61–84, 2001.
- [Nas54] J. Nash. C^1 isometric imbeddings. *The Annals of Mathematics*, 60(3) : 383–396, 1954.
- [OC95] J. M. Oller and J. M. Corcuera. Intrinsic analysis of statistical estimation. *Ann. Statist.*, 23(5) : 1562–1581, 1995.
- [Pal61] R. S. Palais. On the existence of slices for actions of non-compact Lie groups. *Ann. of Math. (2)*, 73 : 295–323, 1961.
- [Pen06] X. Pennec. Intrinsic statistics on Riemannian manifolds : basic tools for geometric measurements. *J. Math. Imaging Vision*, 25(1) : 127–154, 2006.
- [Pen09] X. Pennec. Statistical computing on manifolds : From riemannian geometry to computational anatomy. In Frank Nielsen, editor, *Emerging Trends in Visual Computing*, pages 347–386. Springer-Verlag, Berlin, Heidelberg, 2009.
- [Pic94] J. Picard. Barycentres et martingales sur une variété. *Ann. Inst. H. Poincaré Probab. Statist.*, 30(4) : 647–702, 1994.
- [PS00] R. Plamondon and S.N. Srihari. Online and off-line handwriting recognition : a comprehensive survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1) : 63–84, jan 2000.
- [RL01] J.O. Ramsay and X. Li. Curve registration. *Journal of the Royal Statistical Society (B)*, 63 : 243–259, 2001.

- [Røn01] B. Rønn. Nonparametric maximum likelihood estimation for shifted curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2) : 243–259, 2001.
- [TIR11] T. Trigano, U. Isserles, and Y. Ritov. Semiparametric curve alignment and shift density estimation for biological data. *Signal Processing, IEEE Transactions on*, 59(5) : 1970–1984, 2011.
- [Tsy09] A. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. New York, NY : Springer. xii, 214 p., 2009.
- [TY05a] A. Trouvé and L. Younes. Local geometry of deformable templates. *SIAM J. Math. Anal.*, 37(1) : 17–59 (electronic), 2005.
- [TY05b] A. Trouvé and L. Younes. Metamorphoses through lie group action. *Foundations of Computational Mathematics*, 5(2) : 173–198, 2005.
- [TY11] A. Trouvé and L. Younes. Shape spaces. In *Handbook of Mathematical Methods in Imaging*. Springer, 2011.
- [vdV98] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [Vim10] M. Vimond. Efficient estimation for a subclass of shape invariant models. *Ann. Statist.*, 38(3) : 1885–1912, 2010.
- [WG97] K. Wang and T. Gasser. Alignment of curves by dynamic time warping. *Annals of Statistics*, 25(3) : 1251–1276, 1997.
- [You98] L. Younes. Computable elastic distances between shapes. *SIAM J. Appl. Math.*, 58(2) : 565–586 (electronic), 1998.
- [You10] L. Younes. *Shapes and diffeomorphisms*. Berlin : Springer, 2010.
- [Zie77] H. Ziezold. On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974)*, Vol. A, pages 591–602. Reidel, Dordrecht, 1977.

Auteur : Benjamin CHARLIER

Directeurs de thèse : Dominique BAKRY et Jérémie BIGOT

Date et lieu de soutenance : le 8 décembre 2011 à l'Université Toulouse III Paul Sabatier

Étude des propriétés statistiques des moyennes de Fréchet dans des modèles de déformations pour l'analyse de courbes et d'images de grande dimension

Résumé

Cette thèse porte sur l'analyse statistique de données sur lesquelles agissent des déformations. Dans un premier temps, nous présentons une nouvelle classe de modèles statistiques semi-paramétriques dits de déformations. Ces modèles peuvent s'appliquer à l'étude de courbes temporelles ou d'images de grande dimension. Les données sont supposées être générées par une courbe/image moyenne qui est bruitée et sur laquelle agit un opérateur de déformation. Nous étudions l'estimation des paramètres d'intérêt de ces modèles dans le cas général, puis dans le cas particulier des courbes du plan sur lesquelles agissent les rotations, translations et homothéties.

Dans un second temps, nous considérons les structures non-euclidiennes induites par les actions de groupes de déformations. Un des enjeux des statistiques dans de tels espaces est de généraliser la notion de moyenne euclidienne. C'est ainsi que nous étudions les propriétés qui garantissent l'existence de la moyenne de Fréchet dans le cas particulier du cercle unité muni de la distance de la longueur d'arc.

Mots-clés : Modèles de déformations ; moyenne de Fréchet ; statistique semi-paramétrique ; espace quotient.

Statistical properties of Fréchet means in deformable models for high-dimensional curves and images analysis

Abstract

We are concerned with the statistical analysis of data observed with extra nuisance deformations. To this end, we first introduce a new class of semi-parametric deformable models. These models can be used to study the variability of time dependent curves or high dimensional images. We suppose that the curves or images at hand are generated by a noisy ideal mean pattern on which act some deformations operators. We then study the estimation of the parameters of interest of such models in the general case and in the particular case of planar curves observed with some rotation, translation and scaling.

In a second part, we study the notion of mean in non-Euclidean spaces. More precisely, we study the conditions of existence of the Fréchet mean in the unit circle of the plane endowed with the arclength distance.

Keywords : Deformable models; Fréchet mean; semi-parametric statistics; quotient spaces.
