Consistency of statistics in infinite dimensional quotient spaces

Loïc Devilliers

ÉCOLE DOCTORALE STIC
INRIA SOPHIA-ANTIPOLIS

THÈSE DE DOCTORAT

Présentée en vue de l'obtention du grade de docteur
Mention automatique et traitement du signal et des images de

L'Université Côte d'Azur

par
Loïc Devilliers

Consistance des statistiques dans les espaces quotients de dimension infinie

Dirigée par Xavier Pennec
et codirigée par Stéphanie Allassonnière

Défendue le 20 novembre 2017
Devant le jury constitué de :

Stéphanie Allassonnière	Professeur des Universités	Université Paris Descartes	Co-directrice de thèse
Marc Arnaudon	Professeur des Universités	Université de Bordeaux	Rapporteur
Charles Bouveyron	Professeur des Universités	Université Côte d'Azur	Examinateur
Stephan Huckemann	Professeur des Universités	Université de Göttingen	Rapporteur
Xavier Pennec	Directeur de Recherche	Université Côte d'Azur, Inria	Directeur de thèse
Stefan Sommer	Professeur Associé	Université de Copenhague	Rapporteur
Alain Trouvé	Professeur des Universités	ENS Paris-Saclay	Examinateur

ÉCOLE DOCTORALE STIC

INRIA SOPHIA-ANTIPOLIS

PHD THESIS

To obtain the title of PhD of science,
specialty automation, signal and image processing of

THE UNIVERSITÉ CÔTE D'AZUR

by

Loïc DEVILLIERS

Consistency of statistics in infinite dimensional quotient spaces

Thesis Advisor: Xavier PENNEC
Thesis Co-advisor: Stéphanie ALLASSONNIÈRE

Defended on November 20, 2017
Jury:

Stéphanie Allassonnière	Professor	Université Paris Descartes	Co-advisor
Marc Arnaudon	Professor	Université de Bordeaux	Reviewer
Charles Bouveyron	Professor	Université Côte d'Azur	Examiner
Stephan Huckemann	Professor	University of Göttingen	Reviewer
Xavier Pennec	Senior Researcher	Université Côte d'Azur, Inria	Advisor
Stefan Sommer	Associated Professor	University of Copenhagen	Reviewer
Alain Trouvé	Professor	ENS Paris-Saclay	Examiner

Politique : la division éloigne les gens.

Mathématiques : le quotient les rapproche.

Remerciements

Je souhaiterais vivement remercier mes deux directeurs de thèse. Merci à Stéphanie pour avoir pensé à moi pour ce beau sujet de thèse ainsi que pour tes constants encouragements et conseils durant ces années. Je remercie également Xavier pour les intéressantes discussions que l'on a pu avoir et pour toutes les questions soulevées. Merci de m'avoir guidé tout en me laissant une grande liberté. Je n'oublie pas non plus les bons moments passés ensemble en conférences ainsi que dans la mine.

Je tiens également à remercier Alain d'avoir participé aux discussions scientifiques. Ta vision a également contribué à cette thèse. Meri de l'intérêt que tu as porté à mon travail.

Je souhaiterais remercier Marc pour le temps passé à lire mon manuscrit. Merci d'avoir écrit ce gentil rapport et pour les suggestions que vous avez faites. Thank you Stephan for your report. The suggestion you made fit perfectly with my thesis, and I am glad to see that you offer me some tracks for future work. I would like to thank Stefan for your nice report. It was a pleasure to read it. Thank you for having accepted to read my manuscript during your paternity leave.

J'aimerais également remercier Charles, qui a accepté de présider mon jury de thèse.

Pendant la durée de cette thèse j'ai eu la chance d'avoir une mission d'enseignement dans la préparation à l'agrégation de l'ENS Paris Saclay. Un grand merci à toute l'équipe enseignante si dévouée : Alain, Claudine, Frédéric, Alain D, Arthur, Lionel, Pierre, Anaïs, Sandrine et Tuong-Huy.

Je n'oublie pas non plus tous les gens du CMAP qui m'ont accueilli dans leur laboratoire.

Merci également à l'équipe d'Asclepios avec laquelle j'ai toujours passé de bons moments que ce soit à Sophia ou Auron. Merci à Nicholas de si bien diriger cette équipe. Je souhaite également remercier Isabelle de m'avoir si souvent (trop souvent ?) aidé pour les missions et pour l'organisation de cette soutenance. Merci aux doctorants, ingénieurs et anciens doctorants de leur compagnie. Je souhaite en particulier remercier tous ceux qui m'ont prodigué leurs excellents conseils lors des répétitions. Merci aussi à Nina pour les discussions scientifiques que l'on a eues.

Merci à tous mes amis et à ma famille pour votre constant soutien durant cette thèse.

Enfin, je voudrais remercier ma femme Maëlle, qui tous les jours me supporte. Faire ma thèse à ses côtés était la plus belle chose qui puisse m'arriver.

Merci à tous !

Abstracts

Consistance des statistiques dans les espaces quotient de dimension infinie

En anatomie computationnel, on suppose que les formes d'organes sont issues de déformation d'un template commun. Les données peuvent être des images ou des surfaces d'organes, les déformations peuvent être des difféomorphismes. Pour estimer le template, on utilise souvent un algorithme, appelé «max-max», qui minimise parmi tous les template candidats, la somme des carrées des distances après recalage entre les données et le template candidat. Le recalage étant une étape dans l'algorithme qui trouve la meilleur déformation pour passer d'une forme à une autre.

Le but de cette thèse est d'étudier cet algorithme max-max d'un point de vue mathématique. En particulier, on prouve que cet algorithme est inconsistant à cause du bruit. Cela veut dire que même avec un nombre infini de données et avec un algorithme de minimisation parfait, on estime le template original avec une erreur. Pour prouver cette inconsistance, différentes hypothèses sont requises dans différent résultats de cette thèse. Nous devons donc expliquer ces hypothèses, et surtout produire des résultats avec les hypothèses les plus faibles possibles, pour s'approcher du cadre utilisé dans les applications.

Pour prouver l'inconsistance, on formalise mathématiquement l'estimation de template. On suppose que les déformations sont des éléments aléatoires d'un groupe qui agit sur l'espace des observations. De plus, l'algorithme étudié est interprété comme le calcul de la moyenne de Fréchet dans l'espace des observations quotienté par le groupe des déformations. Dans cette thèse, on prouve que l'inconsistance est dû à la contraction de la distance quotient par rapport à la distance dans l'espace des observations. Dans cette thèse, les observations appartiennent à des espaces comme les espaces de Hilbert ou les variétés Riemanniennes, l'inconsistance est obtenue pour presque tous les bruits.

Un autre but de cette thèse est de quantifier l'inconsistance. On estime l'erreur entre le template originel et le template estimé. Cela met en évidence les paramètres qui gouvernent l'inconsistance. On obtient un équivalent de biais de consistance en fonction du niveau de bruit. Ainsi, l'inconsistance est inévitable quand le niveau de bruit est suffisamment grand.

Mots clés: Moyenne de Fréchet, action de groupe, espace quotient, consistance, espace de Hilbert, variétés, recalage

Consistency of statistics in infinite dimensional quotient spaces

In computational anatomy, organ shapes are assumed to be deformation of a common template. The data can be organ images but also organ surfaces, and the deformations are often assumed to be diffeomorphisms. In order to estimate the template, one often uses the max-max algorithm which minimizes, among all the prospective templates, the sum of the squared distance after registration between the data and a prospective template. Registration is here the step of the algorithm which finds the best deformation between two shapes.

The goal of this thesis is to study this template estimation method from a mathematically point of view. We prove in particular that this algorithm is inconsistent due to the noise. This means that even with an infinite number of data, and with a perfect minimization algorithm, one estimates the original template with an error. In order to prove inconsistency, various hypotheses are required in different results in this thesis. We are committed to explain these hypotheses, and we aim at providing results with the weakest hypotheses possible, in order to approach the frameworks used for applications.

In order to prove inconsistency, we formalize the template estimation into a mathematical framework. Deformations are assumed to be random elements of a group which acts on the space of observations. Besides, the studied algorithm is interpreted as the computation of the Fréchet mean in the space of observations quotiented by the group of deformations. In this thesis, we prove that the inconsistency comes from the contraction of the distance in the quotient space with respect to the distance in the space of observations. As a result, we consider that observations belong to general spaces such as Hilbert spaces and Riemannian manifolds, in these spaces, the inconsistency appears for general noise.

Another goal of this thesis is to quantify this inconsistency. We estimate the error between the original template and the estimated template. This highlights the parameters which govern the inconsistency. We obtained a Taylor expansion of the consistency bias with respect to the noise level. As a consequence, the inconsistency is unavoidable when the noise level is high.

Keywords: Fréchet mean, group action, quotient space, consistency, Hilbert space, manifolds, registration, max-max

Notation

Les statistiques sont l'art d'utiliser toutes les lettres.

The following notations are commonly used in this thesis:

- "$\cdot$": group action of G on M, for $x \in M$, $g \in G$, $g \cdot x \in M$.
- $\langle \ , \ \rangle$: dot product on M when M is a Hilbert space.
- $\| \ \|$: Hilbert norm on M when M is a Hilbert space.
- $[m]$: orbit of m, $[m] = \{g \cdot m, g \in G\}$.
- $B(x, r)$: open ball of center x and radius r.
- CB: consistency bias.
- $C(m)$: cut locus of a point m belonging to a complete Riemannian manifold.
- $\text{Cone}(t_0)$: the Voronoï Cone associated to the template t_0 defined as the set of points of M closer from t_0 than the other points of $[t_0]$.
- d_M: distance in the ambient space, often given by the euclidean norm in Hilbert space, or the Riemannian distance in complete Riemannian manifold.
- d_Q (when d_M is invariant under the group action): quotient (pseudo-)distance between two orbits, $d_Q([a], [b]) = \inf\limits_{g \in G} d_M(a, g \cdot b)$.
- E: variance of Y (or X) in the ambient space M, $E(m) = \mathbb{E}(d_M^2(m, Y))$.
- e_G: the identity element of G.
- $\mathbb{E}$: expectation of a random variable in $\mathbb{R}$ or in a Hilbert space.
- ε: a noise in M with $\mathbb{E}(\varepsilon) = 0$ (sometimes $\mathbb{E}(\|\varepsilon\|^2) = 1$).
- Exp: the Riemannian exponential map in complete Riemannian manifold.
- F: variance in the quotient space of Y (or X), $F(m) = \mathbb{E}(d_Q^2([m], [Y]))$.
- FM: set of all the Fréchet means of a random variable in a metric space $(\mathcal{X}, d_{\mathcal{X}})$ $FM(Z) = \text{argmin}_{x \in \mathcal{X}} \mathbb{E}(d_{\mathcal{X}}^2(x, Z))$.
- F_n: empirical variance in the quotient space of Y, $F_n(m) = \frac{1}{n} \sum\limits_{i=1}^{n} d_Q^2([m], [Y_i])$.
- $\text{Fix}(M)$: set of fixed point under the group action, $\text{Fix}(m) = \{m \in M, \ \forall g \in G \ g \cdot m = m\}$.

- η: a small positive number.
- $\theta(v)$: registration score of the unit vector v of the noise: $\theta(v) = \mathbb{E}\left(\sup_{g\in G}\langle v, g\cdot\varepsilon\rangle\right)$.
- G: a group acting on M.
- $g(a, b)$: an element of G which registers a to b.
- H: a subgroup of G
- $\mathrm{Iso}(m)$: the isotropy group of $m \in M$, $\mathrm{Iso}(m) = \{g \in G,\ g\cdot m = m\}$.
- Int: $\mathrm{Int}(A)$ is the interior of a set A.
- J: an auxiliary map used to minimize F_n:

$$J(m, g_1, \dots, g_n) = \frac{1}{n}\sum_{i=1}^{n}\|m - g_i\cdot Y_i\|^2.$$

- Log the logarithm map in complete Riemannian manifold.
- $L^2(\mathbb{R}/\mathbb{Z})$ or $L^2([0,1])$: set of measurable functions on $[0,1]$ or $\mathbb{R}/\mathbb{Z}$ which are squared integrable.
- $\lambda(v)$: registration score of the unit vector v, $\lambda(v) = \mathbb{E}\left(\sup_{g\in G}\langle v, g\cdot Y\rangle\right)$.
- M: ambient space, most of the time a Hilbert space, sometimes M is a complete Riemannian manifold.
- $m_\star$: an element which minimizes F.
- $\mathbb{P}$: probability measure.
- π: projection in the quotient space $\pi(x) = [x]$.
- Q: quotient space of M by G, $Q = M/G = \{[m],\ m \in M\}$.
- Reg: regularization term on the group G.
- s: section of the quotient, $s : Q \to M$ with $\pi \circ s = Id$.
- $\mathcal{S}$: the image of the section $\mathcal{S} = s(Q)$.
- S: the unit sphere in M.
- σ: the noise level.
- t_0: the template.
- τ: time-shift.

- v: an unit vector in M.
- w: parameter in the Gaussian noise, $\sigma = w\sqrt{n}$ in an Euclidean space of dimension n.
- Φ: a random variable in G which deforms data.
- $X = t_0 + \sigma\varepsilon$: noisy template.
- $Y = \Phi \cdot t_0 + \sigma\varepsilon$ or $Y = \Phi \cdot (t_0 + \sigma\varepsilon)$ the observable variable.

Contents

Chapter 1

Introduction

Contents

1.1 Computational anatomy.

The visionary work of D'Arcy Thompson [Thompson 1942], in the beginning of the 20th century, consisted in the study of the form of animals. The main goal was to create a classification between two species. This raised the question of how to do one study of the difference of form between two animal drawings. His answer was to introduce a grid superimposed onto the image of the first animal, and deform this grid to match to the second one. The more the grid needs to be deformed the more different the two species are.

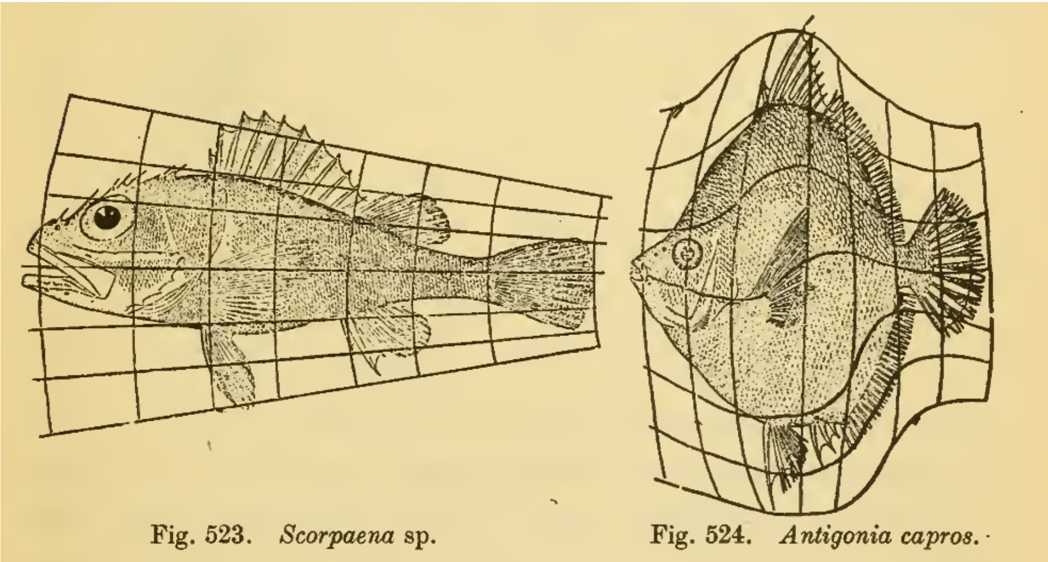


Figure 1.1: Comparison between two fishes.

Since this pioneering work, which was not completely written mathematically, computational anatomy has arisen [Grenander 1998]. This recent discipline aims to build statistics on data with mathematical framework which can be numerically implemented.

There are several reasons for the development of this newly created discipline:

- The increasing use of medical imaging, with high resolution images.
- These images are digitized and are accessible across the world.
- The increasing of computational power which allows us to work with a large number of high resolutions images.
- The mathematical tools and theory have been developed in order to tackle medical issues, for instance shape space theory [Kendall 1989].

One use of computational anatomy would be to predict disease. For instance, one could imagine that we use this deformation model in order to compare a new patient to a known patient. However, as this known patient is singular, he may not be representative among the population, this could bias the study. Moreover, a doctor never compares a new patient to an old one. A doctor would rather want to compare a new patient to all the previous patients. A doctor had synthesized the common features of the previous patients who share the same illness. Computing the mean patient (in this thesis we use the word *template*) would allow us to mime this synthesis. This leads us to study statistics in computational anatomy, in particular, the template estimation.

1.2 Statistics in computational anatomy

Once one has data like organs images, one wants to perform statistics on this data. There are several levels we can think of:

- Given two elements (for instance two medical images), one can estimate the amount of deformations to match the first image (the source image) to the second one (the target image). This step is called registration and formalize the idea of D'Arcy Thompson [Thompson 1942]. One active field of research is to build admissible group of deformations in order to register elements [Trouvé 1995, Miller 2006].

- If we are able to perform the registration step, then we can estimate the mean of images. Due to the deformation, data do not belong to an Euclidean space, one can generalize the notion of mean with the Fréchet mean introduced by Fréchet [Fréchet 1948]. Then we obtain an element which is representative of our population images [Guimond 2000, Joshi 2004]. Many names are given for this element: *template*, *prototype*, *virtual patient* etc; in the following we will use the word *template*. This can be done for instance with the LDDMM framework [Beg 2005] or with the Demons algorithm [Thirion 1998, Vercauteren 2009, Lombaert 2013].

  Note that in practice, we obtain both the template and the deformations between the template and each images [Durrleman 2014, Durrleman 2013, Allassonnière 2015]. The set of the template and the deformations is called an *atlas*.

- The template estimation can be a useful step, but it does not explain the variability of data. Then *Principal Geodesic Analysis* [Fletcher 2003, Fletcher 2004, Sommer 2010, Sommer 2014] (PGA) or *Principal Composant Analysis* [Huckemann 2010, Seiler 2012] (PCA), barycentric subspaces [Pennec 2016, Rohé 2016] etc. can be developed in order to estimate which directions best explain the data.

In this thesis we focus on the estimation of the template (the second level in the previous list). One goal of this thesis is to study the consistency of the minimization algorithm used to estimate the template. As a result, we prove that this method is not consistent. This means that even with the whole distribution (and not only a sample), one does not find the original template, as soon as noise is added. Besides we also focus on finding the error between the original template and the estimated template.

1.3 Geometrical framework of template estimation

1.3.1 A brief overview of manifolds and Riemannian manifolds

The goal of this section is not to recall some formal definitions but to just make a brief recap on some notions used in this thesis. We refer to [do Carmo Valero 1992] for precise definitions and proofs. In this thesis, we prove several properties of the template estimation, where data belong to Hilbert spaces and Riemannian manifold. Besides, in this thesis, deformations leads us to consider a kind of spaces called quotient space which are not manifolds in general. However we will see in chapter 5 that someway the quotient spaces behave as manifolds. Making an analogy between quotient space and manifold will be a useful tool in to understand the property of the template estimation in chapter 5.

We say that a set M is a finite dimensional manifold of dimension n if M locally looks like an open set of $\mathbb{R}^n$:

Definition 1.1. *Let M be a topological space. We say that M is a differentiable manifold of dimension n if there exists a family of injective maps $\varphi_a : U_a \subset R^n \to M$, where for $a \in I$, U_a is an open set of $\mathbb{R}^n$ such that:*

$$M = \bigcup_{a\in I} \varphi(U_a)$$

and for any $a \in I$, $b \in I$, with $W = \varphi_a(U_a) \cap \varphi_b(U_b) \neq \varnothing$, the sets $\varphi^{-1}(W) and \varphi^{-1}(W)$ are open sets of $\mathbb{R}^n$ and the map $\varphi_b^{-1} \circ \varphi_a$ is differentiable.

With this definition, it is possible to define for every point of a manifold M has a tangent space.

In particular, at each point of a manifold m, the space can be approximated by its tangent plan noted $T_m M$.

Moreover, we say that a manifold M is a Riemannian manifold if on each tangent plan m, we have a dot product $\langle \quad , \quad \rangle_m$, which depends continuously of the point m (we note $\| \quad \|_m$ the associated norm).

This allows us to differentiate curve included in the manifold. As a consequence, we are able to define the length of a curve of the manifold. Then, when M is a path-connected, (this means that given two points a and b, it exists a continuous curve connecting a to b), we can define the geodesic distance between a and b as the infimum of the length of all the curves connecting a to b. One can prove that this provide actually a distance called the geodesic distance (or the Riemannian

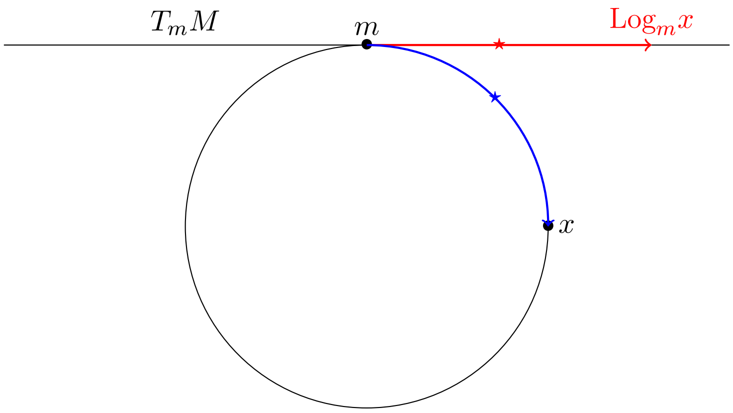


Figure 1.2: Example of the unit sphere in $\mathbb{R}^2$, a Riemannian manifold.

distance) on M noted d_M. From now on, we assume that M is a connected path space.

Besides, by considering that the geodesics are critical point of the energy, one can prove that the geodesics are exactly the solution of a differential equation of order 2. Due to the theory of differential equation, there is always one local solution given one initial point $m \in M$, and one initial vector speed $v \in T_mM$. As a result, there exists one geodesic curve $\gamma :]-\eta, \eta[\to M$ such that $\gamma(0) = m$ and $\dot{\gamma}(0) = v$.

Let m be a point in M, let v be a vector belonging to T_mM. If a geodesic curve γ which satisfies $\gamma(0) = m$ and $\dot{\gamma}(0) = v$ is defined, on an interval I such that $1 \in I$, then we define the exponential map as $\text{Exp}_m(v) = \gamma(1) \in M$. Besides, it is possible to prove that Exp_m is a local diffeomorphism at 0 the null vector in T_mM. We call Log_m the inverse diffeomorphism. Besides we have: $\|\text{Log}_m(a)\|_m = d_M(m, a)$ (see figure 1.2).

Next, we wonder if the geodesic curve can be extended globally to a map defined on $\mathbb{R}$. The Hopf-Rinow theorem give the answer:

Theorem 1.1 (Hopf-Rinow Theorem). *Let M be a connected finite dimensional Riemannian manifold. Then the following statements are equivalent:*

1. *(M, d_M) is a complete metric space where d_M is the geodesic distance.*
2. *M is geodesically complete this means that for every $m \in M$, the exponential map is defined on the entire tangent space T_mM.*
3. *The closed and bounded subsets of M are compact;*

From now on, we assume that M is a complete connected Riemannian manifold. Let γ be such a geodesic curve, then it is possible that γ stops to be a minimal geodesic at the time t_0, in this case, we call $\gamma(t_0)$ a cut point of m. And we call cut locus of m, noted $C(m)$ the set of all cut point of m.

Finally we give a useful result, which allows to differentiate the square distance in manifold:

Proposition 1.1. *Let M be a complete Riemannian manifolld, let $x_0, y \in M$. We assume that $y \notin C(x_0)$. Then $x \mapsto d_M(x,y)^2$ is differentiable at x_0. Besides, $\nabla_1 d_M^2(x_0, y) = -2Log_{x_0}(y)$, where ∇_1 is the gradient of $x \mapsto d_M(x,y)^2$.*

This extends the well known result of the differentiation of the square Euclidean norm in Hilbert space $x \mapsto \|x-y\|^2$ is differentiable at x_0, and $\nabla_1 \|x_0 - y\|^2 = -2(y-x_0)$. This leads us to say that the Log map in complete Riemannian manifolds is like the subtraction in Hilbert spaces.

1.3.2 Deformations and quotient space

Since we assume that there are some deformations of our images. We need to provide a mathematical framework of this idea. In this thesis, we assume that the deformations come from a group action defined below:

Definition 1.2. *We say that a gropup G acts on our ambient space M if there is map $G \times M \to M$, $(g,m) \mapsto g \cdot m$ s.t. $e_G \cdot x = x$ and $g \cdot (g' \cdot x) = (gg') \cdot x$ for every g, $g \in G$, $x \in M$, where e_G is the identity element of G. For a point $m \in M$ we call orbit of m, the set $[m] = \{g \cdot m,\ g \in G\}$. The orbits of M forms a partition of M. Besides, we call quotient of M by G, noted $Q = M/G$ the set of all orbits.*

Due to the deformations induced by the group action, statistics cannot be performed in the ambient M, however its can be performed in the quotient space.

We want to insist on this point: the quotient space is not a manifold in general. The differential structure of the quotient space is an orbifold (see [Thurston 1979] for a definition of orbifold). However, we do not use the formal definition of an orbifold in this thesis. One interesting idea about the differential structure of quotient spaces or orbifolds is that the local dimension of the structure is not constant. This may make the analysis more difficult since it prevents us to apply any statistical theorems involving linear spaces or manifolds.

1.3.3 What are the data?

- 1D/2D/3D images: $f : \mathbb{R}^D \to \mathbb{R}$ at the point/pixel/voxel p, $f(p)$ is the intensity of the images. For instance with $D = 1$, we obtain some signals with D points, the study of this signal is important in ECG [Hitziger 2013, Bigot 2013]

- Landmarks: set of points of interest, for instance, an organ can be encoded numerically, by selecting some characteristic points of the surface of the organ [Bookstein 1986, Joshi 2000]. This selection can been done manually or automatically. One drawback of landmarks is that it is hard to compare two shapes, when these two shapes are encoded with a different number of points. This means that all the data do not belong to the same space, which make difficult the comparison between two shapes.

- Currents [Vaillant 2005] and varifolds [Charon 2013] circumvent this obstacle: these two concepts embed surfaces into a Hilbert space. Then it is possible to do statistical analysis in these frameworks [Durrleman 2014, Durrleman 2009].

1.3.4 What are the actions?

- Diffeomorphisms on images: if $f : \mathbb{R}^D \to \mathbb{R}$ is an image (in dimension D), and φ a diffeomorphism, $\varphi \cdot f = f \circ \varphi$ deforms the image.

- Diffeomorphisms on landmarks: $\varphi \cdot (x_1, \ldots, x_n) = (\varphi(x_1), \ldots, \varphi(x_n))$. Note that there is the particular case of rigid transformations when φ is a combination of rotations and translations.

- Diffeomorphisms on currents/varifolds.

As we will see it in this thesis, there are many actions which can be studied. Each property of the action can be used in order to prove the inconsistency. One important class of action, in this thesis, is the isometric actions in Hilbert spaces (the action is linear, and the Hilbert norm is conserved), for instance:

Example 1.1 (horizontal translation). *The action of continuous translation of functions defined on a torus: we take $G = (\mathbb{R}/\mathbb{Z})^D$ acting on $M = L^2((\mathbb{R}/\mathbb{Z})^D, \mathbb{R})$ with:*

$$\forall \tau \in G \quad \forall f \in M \quad (\tau \cdot f) : t \mapsto f(t + \tau)$$

Then $f \mapsto \tau \cdot f$ is linear, besides $\|\tau \cdot f\| = \|f\|$ for $\| \quad \|$ the norm in M given by $\|f\|^2 = \int_{(\mathbb{R}/Z)^D} f^2(x)dx$.

Remark 1.1. *In this thesis, translation is used in two different context, one is the translation of the variable in a function: $f \mapsto f(\quad + \tau)$. This translation is linear with respect with f. We call this translation, «horizontal translation» (because we translate the variable). On the contrary there is the translation of a point x by a vector v: $x \mapsto x + v$. This map is no longer linear but only affine. We call this translation «vertical translation.*

Remark 1.2. *In all this thesis, we use the expression isometric action for the action on a Hilbert space with a linear action which lets the norm invariant. We are aware that this is not the standard definition which would be "an action is isometric if $d_M(m, m') = d_M(g \cdot m, g \cdot m')$ for every m, $m' \in M$, for every $g \in G$". We prefer using the word isometric for the **linear** isometries in **linear** spaces. And we say that the distance d_M is invariant if $d_M(m, m') = d_M(g \cdot m, g \cdot m')$ for every m, $m' \in M$, for every $g \in G$.*

1.3.5 Distance in quotient space

We recall the definition of pseudo distance and distance on any set $\mathcal{X}$.

Definition 1.3. *Let $\mathcal{X}$ be a set, and $d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ be a map. We say that $d_{\mathcal{X}}$ is a distance in $\mathcal{X}$ (respectively a pseudo-distance) if the following properties are satisfied for every $x, y, z \in \mathcal{X}$:*

- $x = y \iff d_{\mathcal{X}}(x, y) = 0$ *(respectively $x = y \implies d_{\mathcal{X}}(x, y) = 0$)*
- $d_{\mathcal{X}}(x, y) = d_{\mathcal{X}}(y, x)$ *symmetry*
- $d_{\mathcal{X}}(x, y) \leqslant d_{\mathcal{X}}(x, z) + d_{\mathcal{X}}(z, y)$ *triangular inequality*

One assumes that the distance d_M in the ambient space M is invariant with respect to in the group action, this means that:

$$\forall g \in G \quad \forall m, n \in M \quad d_M(g \cdot m, g \cdot n) = d_M(m, n)$$

In this case, we can define d_Q in the quotient space:

$$d_Q([a], [b]) = \inf_{g \in G} d_M(a, g \cdot b)$$

Proposition 1.2. *The map d_Q is a pseudo-distance in Q.*

Proof. First d_Q is a well defined map, indeed, due to the invariant distance, one can verify that

$$\forall h, j \in G \quad \inf_{g \in G} d_M(a, g \cdot b) = \inf_{g \in G} d_M(h \cdot a, g \cdot (j \cdot b)).$$

Namely, that the definition of d_Q does not depend on the chosen element in the orbit of $[a]$ or $[b]$.

Secondly, let $[a], [b], [c]$ three points of Q:

- $d_Q([a], [a]) \leqslant d_M(a, e_g \cdot a) = d_M(a, a) = 0$ (because d_M is a distance). Therefore $d_Q([a], [a]) = 0$.
- $d_M(a, g \cdot b) = d_M(g^{-1} \cdot a, b) = d_M(b, g^{-1} \cdot a)$ (because d_M is symmetric). Therefore $d_Q([a], [b]) = d_Q([b], [a])$ by taking the infimum in the previous inequality and by using the fact that $g \mapsto g^{-1}$ is a bijective map in G.
- $d_M(a, g \cdot b) \leqslant d_M(a, h \cdot c) + d_M(h \cdot c, g \cdot b)$ for every $g, h \in G$ by using the triangular inequality in M. Besides we have:

  $$d_Q([a], [b]) \leqslant d_M(a, g \cdot b) \leqslant d_M(a, h \cdot c) + d_M(c, (h^{-1}g) \cdot b).$$

  By taking the infimum over g, we get:

  $$d_Q([a], [b]) \leqslant d_M(a, h \cdot c) + d_Q([c], [b]).$$

  Finally, by taking the infimum over h we get:

  $$d_Q([a], [b]) \leqslant d_Q([a], [c]) + d_Q([c], [b]).$$

□

Proposition 1.3. *When the orbits are closed set in M for the topology defined by the distance d_M, d_Q is a distance in Q.*

Proof. All we have to verify is if $d_Q([a],[b]) = 0$ implies $[a] = [b]$. Let us assume that $d_Q([a],[b]) = 0$, then for every $n \in \mathbb{N}$ it exists $g_n \in G$ such that $d_M(a, g_n \cdot b) \to 0$. This means in particular that $g_n \cdot b$ is a convergent sequence. Moreover this sequence converges to a. As a consequence a is the limit of elements which are all in the orbit of b. Therefore a is in the closeness of $[b]$. Then, as $[b]$ is a closed set (by assumption) of M, we conclude that $a \in [b]$ and then $[a] = [b]$. □

The orbits are automatically closed if the group G is compact and acts continuously on the ambient space M. In this thesis, when the distance is invariant under a group action, we have a pseudo-distance in the quotient and not necessarily a distance, however this is enough for the analysis of statistics we make. Besides, we may call d_Q quotient distance even if it is only a pseudo-distance.

Remark 1.3. *For every a, $b \in M$, $d_Q([a],[b]) \leqslant d_M(a,b)$. Moreover, when a and b are in generic position we have:*[1]

$$d_Q([a],[b]) < d_M(a,b).$$

In this thesis, we use several concepts which are similar, therefore we make a brief recap to avoid any confusion:

- We say that the action is *isometric* if M is an Hilbert space and if $x \mapsto g \cdot x$ is a linear map which leaves the norm invariant: $\|g \cdot x\| = \|x\|$ for all $g \in G$ and $x \in M$.

- For M a metric space, we say that the distance d_M is invariant under the group action G if:

$$\forall x,\ y \in M \quad \forall g \in G \quad d_M(g \cdot x, g \cdot y) = d_M(x,y).$$

- Let $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ be two sets with discrepancy measures $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$, we only assume that $d_{\mathcal{X}} : \mathcal{X} \to \mathcal{X} \to \mathbb{R}^+$ same for $d_{\mathcal{Y}}$. Let $\Psi : \mathcal{X} \to \mathcal{Y}$ be a map, we say that Ψ is congruent if:

$$\forall (x,x') \in \mathcal{X} \quad d_{\mathcal{X}}(x,x') = d_{\mathcal{Y}}(\Psi(x), \Psi(x')).$$

[1] This childish remark contains all this thesis, since it is the cause of the presence of the inconsistency.

1.4 Statistical estimation method

Since we are in a quotient space, the estimation of the mean by the empirical sum does not apply. However one can use another approach known as the Fréchet mean which can be define in quotient spaces. Indeed, as soon as we have a measure of discrepancy between orbits, one can define and minimize a variance (if the discrepancy is a distance) or a pre-variance (when one of the property of the quotient distance is not fulfilled).

1.4.1 Fréchet mean in Hilbert space

Let M be a Hilbert space. We consider Z a random variable in M, such that $\mathbb{E}(\|Z\|^2) < +\infty$. We consider the variance as the map $E : M \to \mathbb{R}^+$ defined as $E(m) = \mathbb{E}(\|Z - m\|^2)$. We call Fréchet mean of Z the set of global minimizers of E. The empirical Fréchet means are defined as the global minimizer of the empirical variance $E_n : M \to \mathbb{R}^+$ with $E_n(m) = \frac{1}{n}\sum_{i=1}^{n} \|Z_i - m\|^2$ for $Z_1, \ldots, Z_n$ n observations of Z. The following proposition states that the Fréchet mean of Z and the expected value of Z are the same element:

Proposition 1.4. *Let M be a Hilbert space, and d_M the distance given by the euclidean norm. Starting from Z a random variable such that $\mathbb{E}(\|Z\|^2) < +\infty$, then $m \mapsto E(m) = \mathbb{E}(\|Z - m\|^2)$ is well defined and m minimizes E if and only if $m = \mathbb{E}(Z)$.*

Since the square norm is differentiable, it is easy to verify that $\mathbb{E}(Z)$ is the only critical point of the map E, but this is not enough to show the property of being a global minimum.

Proof. First it is easy to expand $E(m)$:

$$E(m) = \|m\|^2 - 2\langle m, \mathbb{E}(Z)\rangle + \mathbb{E}(\|Z\|^2),$$

in particular we can compute the variance at the point $\mathbb{E}(Z)$:

$$E(\mathbb{E}(Z)) = -\|\mathbb{E}(Z)\|^2 + \mathbb{E}(\|Z\|^2).$$

For all $m \in M$, one gets:

$$E(m) \geqslant \|m\|^2 - 2\|m\| \times \|\mathbb{E}(Z)\| + \mathbb{E}(\|Z\|^2) \quad (1.1)$$

$$\geqslant -\|\mathbb{E}(Z)\|^2 + \mathbb{E}(\|Z\|^2) = E(\mathbb{E}(Z)). \quad (1.2)$$

The inequality (1.1) is the use of the Cauchy-Schwarz inequality. Besides, and there is an equality in the Cauchy-Schwarz inequality if and only if m and $\mathbb{E}(Z)$ are positively dependent (this means that $m = \lambda\mathbb{E}(Z)$ or $\mathbb{E}(Z) = \lambda m$ for $\lambda \geqslant 0$). Moreover $x \mapsto x^2 - 2x\|\mathbb{E}(Z)\|$ reaches this minimum at an unique point $x = \|\mathbb{E}(Z)\|$. This implies inequality (1.2).

This shows that $\mathbb{E}(Z)$ minimizes the map E, and that any other minimizer should be positively dependant to $\mathbb{E}(Z)$ and having the same norm that $\mathbb{E}(Z)$, this proves the uniqueness. □

We get the same result with empirical Fréchet mean in a Hilbert space, the empirical Fréchet mean of Z being $\frac{1}{n}\sum_{i=1}^{n} Z_i$.

As we have said, we want to perform statistics of data which are not in linear spaces, but only in metric spaces. It requires to generalize the notion of Fréchet mean in metric spaces.

1.4.2 Fréchet mean in metric spaces

Definition 1.4. *Let (M, d_M) be a metric space. If Z is a random variable in M, such that for every $m \in M$ $\mathbb{E}(d_M(Z,m)^2) < +\infty$, then we can define the variance of Z at any point $m \in M$ by:*

$$E(m) = \mathbb{E}(d_M^2(Z,m)).$$

We say that m is a Fréchet mean of Z if m minimizes globally E. Likewise, if we have $Z_1, \dots, Z_n$ a sample of Z (independent and identically distributed), we define the empirical variance of Z at the point m by:

$$E_n(m) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(d_M^2(Z_i,m)),$$

and we say that m_n is an empirical Fréchet mean of Z if m_n minimizes globally E_n. If an element minimizes locally E (respectively E_n), we call it Karcher mean (respectively empirical Karcher mean).

Proposition 1.5. *Let (M, d_M) be a metric space. If Z is a random variable in M, such that there exists $m_0 \in M$ such that $\mathbb{E}(d_M(Z,m_0)^2) < +\infty$ then for every $m \in M$, $\mathbb{E}(d_M(Z,m)^2) < +\infty$. Therefore we can define the variance E, besides the variance is a continuous map.*

Proof. First, by the triangular inequality we have for all $m \in M$:

$$\begin{aligned} d_M^2(Z,m) &\leqslant (d_M(Z,m_0) + d_M(m_0,m))^2 \\ &\leqslant d_M(Z,m_0)^2 + 2d_M(m_0,m)d_M(Z,m_0) + d_M(m_0,m)^2 \\ &\leqslant d_M(Z,m_0)^2 + 2d_M(m_0,m)(d_M(Z,m_0)^2 + 1) + d_M(m_0,m)^2. \end{aligned}$$

In the left hand side, every term has a finite expected value.

Secondly for m and m' we have, by the triangular inequality:

$$\begin{aligned} |E(m) - E(m')| &\leqslant \mathbb{E}(|d_M(Z,m)^2 - d_M(Z,m')^2| \\ &\leqslant \mathbb{E}(|d_M(Z,m) - d_M(Z,m')| \times (d_M(Z,m) + d_M(Z,m'))) \\ &\leqslant d_M(m,m') \times \mathbb{E}(d_M(Z,m) + d_M(Z,m')), \end{aligned}$$

proving the continuity of the map E.

□

In all this thesis, we always consider random variable Z such that there exists $m_0 \in M$ satisfying $\mathbb{E}(d_M(Z, m_0)^2) < +\infty$.

In order to find the global/local minima of the variance, one can show that, when M is a complete Riemannian manifold and if the cut locus is a null set for the probability measure of Z, then the variance is a differentiable map, and one can compute its gradient (see [Pennec 2006] for instance, it will be also proven in lemma 3.9). As global/local minimum of the variance, the Fréchet/Karcher means are critical point of the variance. This leads to the following definition taken from [Émery 1991] (see also figure 1.3):

Definition 1.5. *Let M be a complete Riemannian manifold. Let Z be a random variable in M, such that $\mathbb{E}(d_M(Z, m_0)^2) < +\infty$ for some $m_0 \in M$. We say that m is an exponentiel barycenter if $\mathbb{P}(Z \in C(m)) = 0$ and if the differential of the variance at the point m is 0, namely if:*

$$\nabla E(m) = -2\mathbb{E}(Log_m(Z)) = 0$$

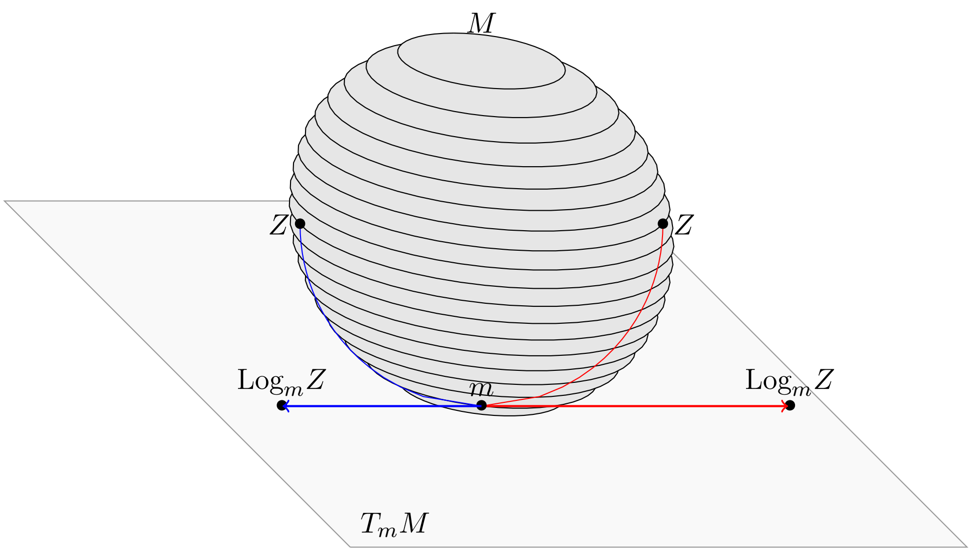


Figure 1.3: If m is the Fréchet mean, then m is an exponentiel barycenter: $\mathbb{E}(\mathrm{Log}_m(Z)) = 0$. The length of the blue curve is equal to the length of the blue segment (same with red).

In this thesis, we will deal with the Fréchet mean in the quotient space. This means that we will consider the quotient distance d_Q previously defined in order to define the Fréchet mean in quotient space.

One issue is that in metric spaces (including manifolds), the existence or the uniqueness is not guarantied. Therefore existence or uniqueness of the Fréchet mean is commonly studied with interesting results [Karcher 1977, Kendall 1990, Afsari 2011, Arnaudon 2005, Arnaudon 2014, Charlier 2013] among many others. These results are mainly proved for data living in manifolds. Besides, one the existence is ensured, methods are developed in order to estimate this Fréchet mean [Le 2004].

Likewise the issue of convergence of Fréchet empirical mean to Fréchet mean is studied [Ziezold 1977] (for data living in metric spaces). Besides, it is also possible to get an evaluation of the speed of convergence provided in [Bhattacharya 2008, Bhattacharya 2003, Bhattacharya 2005] with central limit theorem (for data living in manifolds). It is also possible to study the speed of convergence for data living in quotient space [Huckemann 2011].

In this thesis, we deal with quotient spaces of finite or infinite dimension, which do not fall into the previous frameworks of manifold preventing us from most of these results (except [Ziezold 1977], for instance).

1.4.3 Geometrical and statistical sufficient conditions for inconsistency

In this thesis, we prove, in several different contexts, the inconsistency of the template estimation with the Fréchet mean in quotient space. In most of situations, Theorems can be summarized in two steps:

1. First, due to the geometry, we can define a geometrical set of points in M.
2. Secondly, we require that the support our random variable Y, X or ε is not included in this particular set.

This structure of proof is also common, for instance in the study of stochastic algoritms [Delyon 2000]. Let us review some results of this thesis when this sketch of proof applies:

- Theorems 3.1 and 3.4, we define the cone of the template as the set of points closer from the template than any other points in the orbit of the template. As soon as the support of X is not included in this cone, there is inconsistency.
- Theorems 3.6 and 3.7 (when the template is a fixed point under the action of G), the particular set is the fixed points. We require that the support of the noisy template X is not included in this set, then there is inconsistency.
- Theorem 4.1, we require that the support of the noise ε is not included in the fixed point under the action of G, then there is inconsistency as soon as the noise is large enough.
- Theorem 5.2, we require that the observable Y takes value, with a non zero probability, in the set of points which have a isotropy group reduced to $\{e_G\}$.

To this end, we recall now the support a measure (or a random variable):

Definition 1.6. *Let $\mathbb{P}$ be a (probability) measure on (M, d_M) a metric space, we define the support of $\mathbb{P}$ as:*

$$supp(\mathbb{P}) = \{x \in M, \quad \forall r > 0 \quad \mathbb{P}(B(x,r)) > 0\},$$

where $B(x,r)$ is the open ball for the distance d_M centred at x with a radius r. Let be X a random variable with a law $\mathbb{P}$ we define the support of X as the support of $\mathbb{P}$.

1.4.4 Noise and probability in Hilbert Spaces

Generally, there is no difficulty to generate noise. For instance in $\mathbb{R}$, the Gaussian noise is commonly used, and other noise can be considered. Gaussian noise can easily be generalized in finite dimensional space through the choice a covariance matrix. Moreover it is also possible to generalize in finite dimensional manifolds [Pennec 2006].

Let us talk about the noise in Hilbert spaces, since we will use often in this thesis a noise in the Hilbert space.

Let us begin by a simple example: $L^2([0,1],\mathbb{R})$ the set of square integrable real functions defined on $[0,1]$. If we want to simulate a noise in this space, one could think that it is enough to chose for all $t \in [0,1]$ $f(t)$ randomly. However there is no reason *a priori* for the resulting function f to be measurable. Therefore f does not belong to $L^2([0,1],\mathbb{R})$.

One way to make it works, would be for instance to take the Fourier basis of L^2 which is a one to one application form L^2 to l^2 and define the noise on the coefficient. Let $(\gamma_n)_{n\in\mathbb{Z}}$ a random sequence, where $\gamma_n \in \mathbb{C}$. We choose the law of γ_n in order to have $\sum |\gamma_n|^2 < +\infty$ almost surely. We consider the random function:

$$X : t \mapsto \sum_{n\in\mathbb{Z}} \gamma_n e^{2\pi i n x}$$

This function has a sense as soon as $\sum_n |\gamma_n|^2 < +\infty$. Moreover f will be a real function as soon as $\gamma_n = \overline{\gamma_{-n}}$ for all $n \in \mathbb{Z}$. The expected value will be (by linearity), the function:

$$x \mapsto \sum_{n\in\mathbb{Z}} \mathbb{E}(\gamma_n) e^{2\pi i n x}.$$

Likewise one can compute the variance of its random variable by $\sum_n \mathbb{E}\left(|\gamma_n|^2\right) - \mathbb{E}(|\gamma_n|)^2)$

Note that we do not need to simulate an infinity number of random variables. Indeed, one can first to choose the number of frequencies N randomly (a Poisson distribution for instance), then random numbers: $p_1, \dots, p_N$ (the frequencies), simulate $\gamma_1, \dots, \gamma_N$ (the amplitude of each frequencies) and the random function would be

$$X \mapsto \sum_{i=1}^{N} \gamma_n e^{2\pi i p_n x}.$$

This kind of noise can be immediately generalized in any separable Hilbert space, since it suffices to choose $(e_n)_{n\in\mathbb{N}}$, an orthonormal Hilbert basis, and to choose γ_n randomly as previously, then the random point would be $\sum_n \gamma_n e_n$

1.4.5 Backward/forward model, backward/forward estimation

In this thesis, but also in the literature, there are two ways of considering deformed objects:

- On the one hand, one can assume that the noise is added after deformation, namely that the observable variable is $\Phi \cdot t_0 + \varepsilon$, where t_0 is the template, Φ the unknown and random deformation and ε the additive noise. This is the model introduced by [Grenander 1998] and often considered in computational anatomy. Under this assumption the noise ε is a noise in measurement. In chapter 4, we study the consistency of the template estimation with the Fréchet mean in quotient spaces when observations are created by this model. This model is call forward model in opposition to the backward model defined below.

- On the other hand, one can assume that the noise added before deformation, namely that the observable variable is $\Phi \cdot (t_0 + \varepsilon)$. In this case the noise ε is rather a variability in the shape. Once this variability has been added, the deformation operates on the sum. This is the backward model. In chapter 3, we study the consistency of the template estimation with Fréchet mean in quotient spaces when observations are sampled by this model .

In fact, the model which is the most realistic is probably a mixture of both models. This leads to assume that the observations are $Y = \Phi \cdot (t_0 + \varepsilon) + \varepsilon'$, we will prove the inconsistency of the template estimation with the Fréchet mean when observations are generated by this model in chapter 5.

Once we have the observations, we have to choose between two methods of estimation:

- On the one hand, the forward estimation method: one defines and minimizes the variance

$$F(m) = \mathbb{E}\left(\inf_{g \in G} d_M(g \cdot m, Y)^2\right).$$

  In other words, ones tries to fit the template to the data. This estimation method will be studied, for non isometric action in chapter 4.

- On the other hand, the backward estimation method: we define and minimize the variance

$$F(m) = \mathbb{E}\left(\inf_{g \in G} d_M(m, g \cdot Y)^2\right).$$

  In other words, ones tries to fit the data to the template. This estimation method will be studied, for non isometric action in chapter 5.

When the distance d_M is invariant under the group action the two methods of estimation are equivalent, because d_Q is symmetric (since it is a pseudo-distance as we have see it in proposition 1.2). Therefore in this case, we do not have a choice to make. When d_M is no longer invariant, the forward estimation method may seem more realistic, for example if you have computed a template, it may seem more realistic to deform the template to the data, than the opposite if you have assumed a forward generative model. But the backward estimation method is also easier to implement.

All these questions are about points of view, which can be discussed, for instance, in [Durrleman 2009]. One could think that the forward model is the true instead of the backward (or *vice versa*). But at the end of the day, every statistical models are wrong [Box 1976]. Therefore, to the best of our knowledge, this kind of choice are rather philosophical than mathematical: there is no obvious choice because there is no mathematical statements proving than one is better than the other.

1.5 Template estimation in this thesis and in related works

The consistency was first studied in the particular case of Procustes means: Procustes mean is the mean of data when rotations, translations (and sometimes scaling) have been removed [Lele 1993, Kent 1997, Le 1998, Huckemann 2011]. Note that Procustes means are related to Fréchet means as noticed in [Le 1998]. However, there is no contradiction between this work and, for instance, the article of [Kent 1997]. Indeed, [Kent 1997] showed the consistency when the scaling parameters were taken into account. In most of our work, we deal with isometric actions which exclude this scaling effect.

Bigot and Charlier [Bigot 2011] studied the question of the template estimation with a finite sample in the case of translated signals or images by providing a lower bound of the consistency bias. This lower bound was unfortunately not so informative as it is converging to zero asymptotically when the dimension of the space tends to infinity.

Miolane [Miolane 2017] provided a general explanation of why the template is badly estimated for a isometric group action thanks to a geometric interpretation. She showed that the external curvature of the orbit of the template is responsible for the inconsistency. This result was quantified with Gaussian noise for general manifolds.

In this thesis, we study the template estimation by computing the Fréchet mean in quotient space of the observable variable. Our thesis is that this estimation is not consistent, this means that the template is generally not a Fréchet mean of the observable variable mapped in the quotient space. We call consistency bias the distance between the template and the Fréchet means in the quotient space of the observable variable. Contrarily to [Lele 1993, Kent 1997, Le 1998, Huckemann 2011, Bigot 2011, Huckemann 2012], we study the template estimation with an abstract action. Therefore, we use a general framework described, for instance, in [Huckemann 2010].

1.6 Heuristic of inconsistency

In this thesis, we want to prove that the template estimation is not consistent with the Fréchet mean in the quotient space. Before giving any proofs, we can give an heuristic: Let us take M a Hilbert space, G a group acting, such that d_M is invariant under the group action, we note $\pi : M \to Q = M/G$ the canonical projection in the quotient space: $\pi(m) = [m]$. For $Y = \Phi \cdot (t_0 + \sigma\varepsilon)$ a random deformation of the template added to the noise. We define $X = t_0 + \sigma\varepsilon$, therefore $Y = \Phi \cdot X$. Thus, there is consistency if and only if:

$$\pi(t_0) \in \mathrm{FM}(\pi(Y)).$$

As $t_0 = \mathbb{E}(X)$ (because we add a centred noise), and $\pi(Y) = \pi(X)$, there is consistency if and only if $\pi(\mathbb{E}(X)) \in \mathrm{FM}(\pi(X))$. As we have seen it, the Fréchet mean of X is reduced to $\mathbb{E}(X)$ because X lives in M a Hilbert space. Therefore we can state that:

$$\pi(\mathrm{FM}(X)) \in \mathrm{FM}(\pi(X)) \text{ if and only if there is consistency.}$$

Then, the question of consistency is reduced to know if the projection into the quotient space commutes with the Fréchet mean. There are, at least, two trivial cases where this is true:

- If $\sigma = 0$, then $X = t_0$ almost surely, and $\pi(X) = \pi(t_0)$ is a constant variable, then $\mathrm{FM}(\pi(X)) = \pi(\mathbb{E}(X))$, and there is consistency.
- If Q is a linear space, the notion of Fréchet mean is just the expectation, therefore there is consistency if and only if $\pi(\mathbb{E}(X)) = \mathbb{E}(\pi(X))$. However, it is a well know fact that in general $f(\mathbb{E}(X)) \neq \mathbb{E}(f(X))$ (then in this case why we would have $\pi(\mathrm{FM}(X)) \in \mathrm{FM}(\pi(X))$? . One remarkable exception is for f a affine map. As a result, if Q is a linear space and if π the canonical projection is an affine, there is consistency.

Therefore when $\sigma \neq 0$ and when the quotient space is not a linear space, there is no reason *a priori* for $\pi(\mathrm{FM}(X)) \in \mathrm{FM}(\pi(X))$. And this thesis aims to provide proofs of this heuristic.

1.7 Manuscript overview

- Chapter 3 establishes theorems which proves inconsistency in the case of isometric action in a Hilbert space. This works has been partially presented in the *Mathematical Foundations of Computational Anatomy* workshop [Allassonnière 2015], then published in the *SIAM imaging science* journal [Devilliers 2017c].
- Chapter 4 establishes an asymptotic behaviour of the consistency bias in the case of isometric action in a Hilbert space. Moreover the used method paves

the way to the proof of inconsistency when the action is no longer isometric. We prove in this chapter that the inconsistency appears as soon as the noise is large enough for non isometric actions in the foreward estimation. It has been partially published in the *Information Processing in Medical Imaging* conference [Devilliers 2017a] and extended in the *Entropy* journal [Devilliers 2017b].

- Chapter 5 is an opening to generalization. We find an implicit formula for the Fréchet mean in quotient space used to provide new proofs of inconsistency. In this chapter, we restrict ourselves in the case of a backward estimation. In particular, we also prove inconsistency in complete Riemannian manifold. As a result, we prove that the Fréchet mean in the quotient space is noisier than the original template.

Chapter 2

Introduction (français)

Contents

2.1 Anatomie computationnelle

Au début du 20ème siècle, D'Arcy Thompson effectua un travail visionnaire en étudiant les formes des animaux. Une des idées principales était de créer une classification entre deux espèces. Pour ce faire, il introduisit une grille sur l'image du premier animal et déforma cette grille pour faire coïncider l'image du premier animal sur le second. Plus la grille doit être déformée, plus les espèces sont différents.

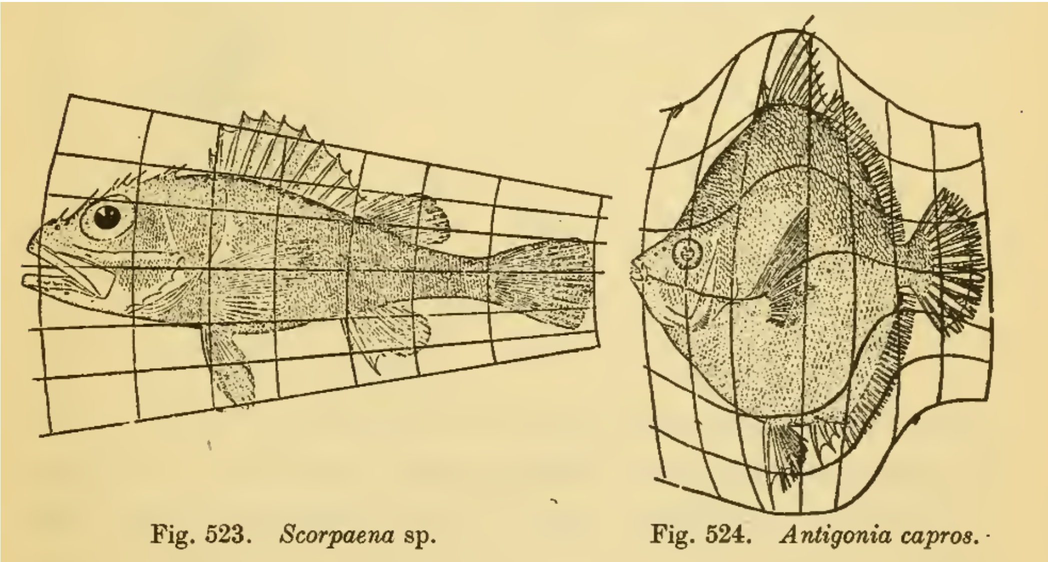


Figure 2.1: Comparison entre deux poissons.

Ce travail n'avait pas été écrit complètement mathématiquement, mais a inspiré une nouvelle discipline : l'anatomie computationnelle [Grenander 1998] Cette discipline récente cherche à construire des statistiques sur des données avec un cadre mathématique qui peuvent être implantées sur un ordinateur.

Il y a plusieurs raisons qui ont conduit au développement de cette nouvelle discipline:

- L'augmentation de l'utilisation des images médicales en haute résolution.
- Le fait que ces images sont numérisées et puissent être transmises dans le monde entier.
- L'augmentation de la capacité de calcul des ordinateurs.
- Le développement d'outils et de théories mathématiques, par exemple la théorie des espaces de formes [Kendall 1989].

L'anatomie computationelle pourrait être utilisée pour prédire l'apparition de maladies. Par exemple, on peut imaginer utiliser ce modèle de déformation pour comparer un nouveau patient avec un patient connu. Cependant, ce patient connu est singulier, il peut ne pas être représentatif de la population. Ceci pourrait biaiser l'étude. De plus, un docteur ne compare jamais un nouveau patient avec un ancien. Il compare plutôt ce patient avec tous les patients précédents qui partagent la même maladie. Calculer le patient moyen (dans cette thèse on utilisera le mot *template*) nous permettrait de faire mimer le processus mental qu'effectue le médecin. Ceci nous conduit à étudier les statistiques en anatomie computationnelle, et en particulier l'estimation de ce template.

2.2 Statistique en anatomie computationnelle

Lorsqu'on a des données comme des images d'organes, on veut effectuer des statistiques sur ces données. On peut séparer cette analyse en plusieurs niveaux :

- Étant donné deux éléments (par exemple deux images médicales), on peut estimer la déformation qui permet de faire coller la première image à la seconde. Cette étape, appelée «recalage», formalise l'idée de D'Arcy Thompson. L'un des champs de recherche est de construire des groupes de déformations pour recaler ces éléments [Trouvé 1995, Miller 2006].

- Si on est capable d'effectuer la première étape, alors on peut estimer la moyenne des images. À cause des déformations, les données n'appartiennent plus à un espace euclidien, cependant on peut généraliser la notion de moyenne avec la moyenne de Fréchet [Fréchet 1948]. Ainsi, on obtient un élément qui est représentatif de notre population [Guimond 2000, Joshi 2004]. Cet élément peut être appelé de différentes façons : template, prototype, patient virtuel etc. Dans toute cette thèse, on utilisera le mot template. Cette estimation peut être effectuée avec la méthode LDDMM [Beg 2005] ou avec l'algorithme des Démons [Thirion 1998, Vercauteren 2009, Lombaert 2013].

- Dans la pratique, on obtient le template, et les déformations entre le template et chacune des images [Durrleman 2014, Durrleman 2013, Allassonnière 2015]. L'ensemble formé par le template et ces déformations est appelé *atlas*.

- L'estimation de template peut être une étape utile, mais elle n'explique pas la variabilité des données. Ainsi, *Principal Composant Analysis* [Huckemann 2010, Seiler 2012] (PCA) ou *Principal Geodesic Analysis* [Fletcher 2003, Fletcher 2004, Sommer 2010, Sommer 2014] (PGA), les espaces barycentriques [Pennec 2016, Rohé 2016] etc. peuvent être développés pour estimer les directions qui expliquent le mieux les données.

Dans cette thèse, nous nous concentrons sur l'estimation de template (le second niveau dans la liste ci-dessus). Un des buts de cette thèse est d'étudier la consistance de l'algorithme de minimisation utilisé pour estimer le template. On prouve que cette méthode n'est pas consistante. Cela veut dire que, même avec la distribution entière (et non un échantillon), on ne trouve pas le template originel dès que du bruit a été ajouté. De plus, on cherche aussi à estimer l'erreur entre le template originel et le template estimé.

2.3 Cadre géométrique pour l'estimation de template

2.3.1 Distance dans le quotient

Soit (M, d_M) un espace métrique, on suppose que G est un groupe qui agit sur M. On suppose que la distance d_M est invariante par l'action de groupe, cela veut dire que :

$$\forall g \in G \quad \forall m, n \in M \quad d_M(g \cdot m, g \cdot n) = d_M(m, n)$$

Dans ce cas, on peut définir une pseudo distance dans le quotient $Q = M/G$:

$$d_Q([a], [b]) = \inf_{g \in G} d_M(a, g \cdot b)$$

Remarque 2.1. *Pour tout a, $b \in M$, $d_Q([a], [b]) \leqslant d_M(a, b)$. De plus, si a et b sont en position générique, on a:*[1]

$$d_Q([a], [b]) < d_M(a, b).$$

Lorsqu'on est dans un quotient, l'estimation de la moyenne par la moyenne empirique ne s'applique pas. Cependant on peut utiliser la moyenne de Fréchet.

2.3.2 Moyenne de Fréchet dans les espaces de Hilbert

Soit M un espace de Hilbert et soit Z une variable aléatoire dans M telle que $\mathbb{E}(\|Z\|^2) < +\infty$. On définit la variance comme l'application $E : M \to \mathbb{R}^+$ définie par $E(m) = \mathbb{E}(\|Z - m\|^2$. On appelle moyenne de Fréchet l'ensemble des minimiseurs globaux de E. La moyenne de Fréchet empirique est définie comme les minimiseurs globaux de la variance empirique $E_n : M \to \mathbb{R}^+$ avec $E_n(m) = \frac{1}{n}\sum_{i=1}^{n} \|Z_i - m\|^2$ pour $Z_1, \ldots, Z_n$ n observations de Z. La proposition suivante établit que la moyenne de Fréchet de Z est l'espérance de Z :

Proposition 2.1. *Soit M un espace de Hilbert, et d_M la distance donnée par la norme euclidienne. Soit Z une variable aléatoire telle que $\mathbb{E}(\|Z\|^2) < +\infty$, alors $m \mapsto E(m) = \mathbb{E}(\|Z - m\|^2)$ est bien définie et m minimise E si et seulement si $m = \mathbb{E}(Z)$.*

On obtient le même résultat avec les moyennes de Fréchet dans les espaces de Hilbert, la moyenne de Fréchet empirique de Z étant $\frac{1}{n}\sum_{i=1}^{n} Z_i$.

Comme dit précédemment, nous voulons effectuer des statistiques sur des données qui ne vivent pas dans un espace vectoriel mais plutôt dans un espace métrique. On généralise donc la notion de moyenne de Fréchet dans les espaces métriques.

[1] Cette remarque peut sembler innocente, mais en fait c'est ce fait là qui conduit à l'inconsistance.

2.3.3 Moyenne de Fréchet dans les espaces métriques

Definition 2.1. *Soit (M, d_M) un espace métrique. Soit Z une variable aléatoire dans M, telle que pour tout $m \in M$ $\mathbb{E}(d_M(Z,m)^2) < +\infty$, alors on peut définir la variance de Z en tout point $m \in M$ par :*

$$E(m) = \mathbb{E}(d_M^2(Z,m))$$

On dit que m est une moyenne de Fréchet de Z si m minimise E. De même, pour un échantillon $Z_1, \dots, Z_n$ un échantillon de Z, on définit la variance empirique de Z au point m par :

$$E_n(m) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(d_M^2(Z_i,m))$$

On dit que m_n est une moyenne de Fréchet empirique de Z si m_n minimise E_n. Un élément minimisant localement E (respectivement E_n) est appelé moyenne de Karcher (respectivement moyenne de Karcher locale).

Proposition 2.2. *Soit (M, d_M) un espaces métrique. Si Z est une variable aléatoire dans M telle qu'il existe $m_N \in M$ telle que $\mathbb{E}(d_M(Z,m_0)^2) < +\infty$, alors pour tout m, la variance au point m - $E(m) = \mathbb{E}(d_M(Z,m)^2)$ est bien définie, de plus E est une application continue.*

Pour trouver les minima locaux/globaux de la variance, on peut montrer que lorsque M est une variété Riemannienne complète et si le lieu de coupure est un ensemble de mesure nulle, alors la variance est différentiable et on peut calculer son gradient (voir [Pennec 2006] par exemple, ce résultat sera aussi prouvé au lemme 3.9). Comme ce sont minima globaux/locaux les moyennes de Fréchet/Karcher sont les points critiques de la variance. Cela conduit, dans [Émery 1991], à la définition suivante :

Definition 2.2. *Soit M une variété Riemannienne complète. Soit Z une variable aléatoire dans M telle que $\mathbb{E}(d_M(Z,m_0)^2) < +\infty$ pour un certain $m_0 \in M$. On dit que m est barycentre exponentiel si $\mathbb{P}(Z \in C(m)) = 0$ et si :*

$$\nabla E(m) = -2\mathbb{E}(Log_m(Z)) = 0$$

Dans cette thèse, on va étudier la moyenne de Fréchet dans les espaces quotients. Cela signifie que l'on va utiliser la distance d_Q définie précédemment pour définir la moyenne de Fréchet dans les espaces quotients.

Dans les espaces métriques (y compris les variétés), l'existence ou l'unicité n'est pas garantie. Par conséquent, l'existence ou l'unicité de la moyenne de Fréchet est souvient étudiée avec des résultats intéressants par exemple [Karcher 1977, Kendall 1990, Afsari 2011, Arnaudon 2005, Arnaudon 2014, Charlier 2013]. Ces résultats sont souvent prouvés dans le cas des variétés. De plus, une fois que l'existence est assurée, on peut développer des méthodes pour estimer la moyenne de Fréchet [Le 2004].

De même, on peut étudier la convergence de la moyenne de Fréchet empirique [Ziezold 1977] (dans le cas des espaces métriques). On peut aussi évaluer la vitesse de convergence [Bhattacharya 2008, Bhattacharya 2003, Bhattacharya 2005] avec des théorèmes central limite dans le cas des variétés ou dans le cas des espaces quotients [Huckemann 2011].

Dans toute cette thèse, on s'intéresse aux espaces quotients de dimension finie ou infinie, ainsi on ne peut pas utiliser la plupart de ces résultats (excepté un résultat comme celui de [Ziezold 1977] par exemple).

2.4 Estimation de template dans cette thèse et dans d'autres travaux

La consistance a d'abord été étudiée dans le cas particulier des moyennes de Procuste : La moyenne de Procuste est la moyenne des données quand les rotations, translations (et parfois les homothéties) ont été enlevées [Lele 1993, Kent 1997, Le 1998, Huckemann 2011]. Le [Le 1998] a remarqué que les moyennes de Procuste sont reliées aux moyennes de Fréchet. Cependant il n'y a pas de contradiction entre ce travail et par exemple l'article [Kent 1997]. En effet, [Kent 1997] montre la consistance quand le paramètre d'homothétie a été pris en compte. Dans ce travail, on étudiera souvent les actions isométriques, ce qui exclut les homothéties.

Bigot et Charlier [Bigot 2011] ont étudié la question de l'estimation de template avec un échantillon de taille fini dans les cas des signaux/images translatés en fournissant une minoration du biais de consistance. Cette minoration n'était malheureusement pas assez informative car elle convergeait vers zéro quand la dimension de l'espace tend vers plus l'infini.

Miolane [Miolane 2017] a expliqué de façon générale pourquoi le template était aussi mal estimé lorsque la distance est invariante sous l'action de groupe grâce à une interprétation géométrique. Elle a montré que la courbure externe de l'orbite du template cause l'inconsistance. Ce résulat a été quantifié avec un bruit gaussien dans des variétés.

Dans cette thèse, on étudie l'estimation de template par le calcul de la moyenne de Fréchet dans l'espace quotient de la variable aléatoire observable. Notre thèse est que cette estimation n'est pas consistante. Cela veut dire que le template n'est généralement pas une moyenne de Fréchet de la variable aléatoire observable projetée dans l'espace quotient. On appelle biais de consistance la distance entre le template et les moyennes de Fréchet. Contrairement à [Lele 1993, Kent 1997, Le 1998, Huckemann 2011, Bigot 2011, Huckemann 2012], on étudie l'estimation de template avec une action de groupe abstraite. Par conséquent on utilise un cadre général, décrit par exemple dans [Huckemann 2010].

2.5 Organisation du manuscrit

- Dans le chapitre 3, on établit des théorèmes qui prouvent le biais dans le cas d'actions isométriques dans les espaces de Hilbert. Ce travail a été partiellement présenté dans le workshop *Mathematical Foundations of Computational Anatomy* [Allassonnière 2015], puis publié dans la revue *SIAM imaging science* [Devilliers 2017c].

- Dans le chapitre 4, on établit un comportement asymptotique du biais de consistance dans le cas d'une action isométrique dans un espace de Hilbert. De plus, la méthode utilisée permet de prouver l'inconsistance lorsque l'action est non isométrique, en effet on prouve dans ce chapitre l'inconsistance dès lors que le niveau de bruit et suffisamment grand. Ce travail a été partiellement publié dans la conférence *Information Processing in Medical Imaging* [Devilliers 2017a] et étendu dans la revue *Entropy* [Devilliers 2017b].

- Dans le chapitre 5, on tente de généraliser l'étude déjà faite. On trouve une formule implicite de la moyenne de Fréchet dans les espaces quotients qu'on utilise pour faire de nouvelles preuves de l'inconsistance. Dans ce chapitre, on se restreint à l'estimation «*backward*». On prouve notamment que la moyenne de Fréchet dans l'espace quotient est plus bruitée que le template originel.

Chapter 3

Inconsistency in Hilbert space for isometric action

Contents

A short version of this chapter has been presented in a workshop [Allassonnière 2015], then has been extended in the *SIAM imaging sciences* journal [Devilliers 2017c]. Note that, comparing to this journal paper, we add section 3.4.7 proving that the variance in the quotient space is not differentiable everywhere.

Abstract: In this chapter, we study the consistency of the template estimation with the Fréchet mean in quotient spaces. The Fréchet mean in quotient spaces is often used when the observations are deformed or transformed by a group action. We show that in most cases this estimator is actually inconsistent. We exhibit a sufficient condition for this inconsistency, which amounts to the folding of the distribution of the noisy template when it is projected to the quotient space. This condition appears to be fulfilled as soon as the support of the noise is large enough. To quantify this inconsistency we provide lower and upper bounds of the bias as a function of the variability (the noise level). This shows that the consistency bias cannot be neglected when the variability increases.

3.1 Introduction

In Kendall's shape space theory [Kendall 1989], in computational anatomy [Grenander 1998], in statistics on signals, or in image analysis, one often aims at estimating a template. A template stands for a prototype of the data. The data can be the shape of an organ studied in a population [Durrleman 2014] or an aircraft [Lefebvre 2012], an electrical signal of the human body, a MR image, etc. To analyze the observations, one assumes that these data follow a statistical model. One often models observations as random deformations of the template with additional noise. This deformable template model proposed in [Grenander 1998] is commonly used in computational anatomy. The concept of deformation introduces the notion of group action: the deformations we consider are elements of a group which acts on the space of observations, called here the ambient space (also called top space). Since the deformations are unknown, one usually considers equivalent classes of observations under the group action. In other words, one considers the quotient space of the ambient space (or top space) by the group. In this particular setting, the template estimation is most of the time based on the minimization of the empirical variance in the quotient space (for instance, [Kurtek 2011b, Joshi 2004, Sabuncu 2008] among many others). The points that minimize the empirical variance are called the empirical Fréchet mean. The Fréchet means introduced in [Fréchet 1948] is comprised of the elements minimizing the variance. This generalizes the notion of expected value in nonlinear spaces. Note that the existence or uniqueness of Fréchet mean is not ensured. But sufficient conditions may be given in order to reach existence and uniqueness (for instance, [Karcher 1977, Kendall 1990]).

Several group actions are used in practice: some signals can be shifted in time compared to other signals (action of horizontal translations [Hitziger 2013]), landmarks can be transformed rigidly [Kendall 1989], shapes can be deformed by diffeomorphisms [Durrleman 2014], etc. In this paper we restrict to transformation which leads the norm unchanged. Rotations, for instance, leave the norm unchanged, but it may seem restrictive. In fact, the square root trick detailed in section 3.5, allows one to build norms which are unchanged, for instance, by reparametrization of curves with a diffeomorphism, where our work can be applied.

We raise several issues concerning the estimation of the template.

1. Is the Fréchet mean in the quotient space equal to the original template projected in the quotient space? In other words, is the template estimation with the Fréchet mean in quotient space consistent?

2. If there is an inconsistency, how large is the consistency bias? Indeed, we may expect the consistency bias to be negligible in many practicable cases.

3. If one gets only a finite sample, one can only estimate the empirical Fréchet mean. How far is the empirical Fréchet mean from the original template?

These issues originated from an example exhibited by Allassonnière, Amit, and Trouvé [Allassonnière 2007]: they took a step function as a template and added some noise and shifted in time this function. By repeating this process they created a data sample from this template. With this data sample, they tried to estimate the template with the empirical Fréchet mean in the quotient space. In this example, minimizing the empirical variance did not succeed in estimating well the template when the noise added to the template increases, even with a large sample size.

One solution to ensure convergence to the template is to replace this estimation method with a Bayesian paradigm ([Allassonnière 2010, Bontemps 2014] or [Zhang 2013]). But there is a need to have a better understanding of the failure of the template estimation with the Fréchet mean. One can studied the inconsistency of the template estimation. Bigot and Charlier [Bigot 2011] first studied the question of the template estimation with a finite sample in the case of translated signals or images by providing a lower bound of the consistency bias. This lower bound was unfortunately not so informative as it is converging to zero asymptotically when the dimension of the space tends to infinity. Miolane and co-authors [Miolane 2015, Miolane 2017] later provided a more general explanation of why the template is badly estimated for a general group action thanks to a geometric interpretation. They showed that the external curvature of the orbits is responsible for the inconsistency. This result was further quantified with Gaussian noise. In this chapter, we provide sufficient conditions on the noise for which inconsistency appears, and we quantify the consistency bias in the general (non necessarily Gaussian) case. Moreover, we mostly consider a vector space (possibly infinite dimensional) as the ambient space while the article of Miolane and co-authors is restricted to finite dimensional manifolds.

This chapter is organized as follows. Section 3.2 details the mathematical terms that we use and the generative model. In sections 3.3 and 3.4, we exhibit a sufficient condition that lead to an inconsistency when the template is not a fixed point under the group action. This sufficient condition can be roughly understand as follows: with a non zero probability, the projection of the random variable on the orbit of the template is different from the template itself. This condition is actually quite general. In particular, this condition it is always fulfilled with the Gaussian noise or with any noise whose support is the whole space. Moreover we quantify the consistency bias with lower and upper bounds. We restrict our study to Hilbert spaces and isometric actions. This means that the space is linear, the group acts linearly and leaves the norm (or the dot product) unchanged. Section 3.3 is dedicated to finite groups. Then we generalise our result in section 3.4 to non-finite groups. To complete this study, we extend in section 3.5 the result when the template is a fixed point under the group action and when the ambient space is a manifold. As a result we show that the inconsistency exists for almost all noises. Although the bias can be neglected when the noise level is sufficiently small, its linear asymptotic behaviour with respect to the noise level show that it becomes unavoidable for large noises.

3.2 Definitions, notation and generative model

We denote by M the ambient space, which is the image/shape space, and G the group acting on M. The action is a map:

$$\begin{array}{ccc} G \times M & \to & M \\ (g, m) & \mapsto & g \cdot m \end{array},$$

satisfying the following properties: for all $g,\ g' \in G,\ m \in M$ $(gg') \cdot m = g \cdot (g' \cdot m)$ and $e_G \cdot m = m$ where e_G is the neutral element of G. For $m \in M$ we note by $[m]$ the orbit of m (or the class of m). This is the set of points reachable from m under the group action: $[m] = \{g \cdot m,\ g \in G\}$. Note that if we take two orbits $[m]$ and $[n]$ there are two possibilities:

1. The orbits are equal: $[m] = [n]$ i.e. $\exists g \in G$ s.t. $n = g \cdot m$.

2. The orbits have an empty intersection: $[m] \cap [n] = \emptyset$.

We call quotient of M by the group G the set all orbits. This quotient is noted by:

$$Q = M/G = \{[m],\ m \in M\}.$$

The orbit of an element $m \in M$ can be seen as the subset of M of all elements $g \cdot m$ for $g \in G$ or as a point in the quotient space. In this chapter we use these two ways. We project an element m of the ambient space M into the quotient by taking $[m]$.

Now we are interested in adding a structure on the quotient from an existing structure in the ambient space: take M a metric space, with d_M its distance. Suppose that d_M is invariant under the group action which means that:

$$\forall g \in G,\ \forall a,\ b \in M\ d_M(a, b) = d_M(g \cdot a, g \cdot b).$$

Then we obtain a pseudo-distance on Q defined by:

$$d_Q([a],[b]) = \inf_{g \in G} d_M(g \cdot a, b). \tag{3.1}$$

We remind that a distance on M is a map $d_M : M \times M \mapsto \mathbb{R}^+$ such that for all m, n, $p \in M$ the following hold:

1. $d_M(m,n) = d_M(n,m)$ (symmetry).
2. $d_M(m,n) \leqslant d_M(m,p) + d_M(p,n)$ (triangular inequality).
3. $d_M(m,m) = 0$.
4. $d_M(m,n) = 0 \Longleftrightarrow m = n$.

A pseudo-distance satisfies only the first three conditions. If we suppose that all the orbits are closed sets of M, then one can show that d_Q is a distance. In this chapter, we assume that d_Q is always a distance, even if a pseudo-distance would be sufficient. $d_Q([a],[b])$ can be interpreted as the distance between the shapes a and b, once one has removed the parametrisation by the group G. In other words, a and b have been registered. In this chapter, except in section 3.5, we suppose that the the group acts isometrically on a Hilbert space, this means that the map $x \mapsto g \cdot x$ is linear, and that the norm associated to the dot product is conserved: $\|g \cdot x\| = \|x\|$. Then $d_M(a,b) = \|a - b\|$ is a particular case of invariant distance.

We now introduce *the generative model* used in this chapter for M a vector space. Let us take a template $t_0 \in M$ to which we add a unbiased noise ε: $X = t_0 + \varepsilon$. Finally we transform X with a random shift Φ of G. We assume that this variable Φ is independent of X and the only observed variable is:

$$Y = \Phi \cdot X = \Phi \cdot (t_0 + \varepsilon), \text{ with } \mathbb{E}(\varepsilon) = 0, \tag{3.2}$$

while Φ, X and ε are hidden variables.

Note that it is not the generative model defined by Grenander and often used in computational anatomy. Where the observed variable is rather $Y' = \Phi \cdot t_0 + \varepsilon'$. But when the noise is isotropic and the action is isometric, one can show that the two models have the same law, since $\Phi \cdot \varepsilon$ and ε have the same probability distribution. As a consequence, the inconsistency of the template estimation with the Fréchet mean in quotient space with one model implies the inconsistency with the other model. Because the former model (3.2) leads to simpler computation we consider only this model.

We can now set the inverse problem: given the observation Y, how to estimate the template t_0 in M? This is an ill-posed problem. Indeed for some element group $g \in G$, the template t_0 can be replaced by the translated $g \cdot t_0$, the shift Φ by Φg^{-1} and the noise ε by $g \cdot \varepsilon$, which leads to the same observation Y. So instead of estimating the template t_0, we estimate its orbit $[t_0]$. By projecting the observation Y in the quotient space we obtain $[Y]$. Although the observation $Y = \Phi \cdot X$ and the noisy template X are different random variables in the ambient space, their projections on the quotient space lead to the same random orbit $[Y] = [X]$. That is why we consider the generative model (3.2): the projection in the quotient space remove the transformation of the group G. From now on, we use the random orbit $[X]$ in lieu of the random orbit of the observation $[Y]$.

The variance of the random orbit $[X]$ (sometimes called the Fréchet functional or the energy function) at the quotient point $[m] \in Q$ is the expected value of the square distance between $[m]$ and the random orbit $[X]$, namely:

$$Q \ni [m] \mapsto \mathbb{E}(d_Q([m],[X])^2) \tag{3.3}$$

An orbit $[m] \in Q$ which minimizes this map is called a Fréchet mean of $[X]$.

If we have an *i.i.d* sample of observations $Y_1, \ldots, Y_n$ we can write the *empirical quotient variance*:

$$Q \ni [m] \mapsto \frac{1}{n}\sum_{i=1}^{n} d_Q([m],[Y_i])^2 = \frac{1}{n}\sum_{i=1}^{n} \inf_{g_i \in G} \|m - g_i \cdot Y_i\|^2. \tag{3.4}$$

Thanks to the equality of the quotient variables $[X]$ and $[Y]$, an element which minimises this map is an *empirical Fréchet mean* of $[X]$.

In order to minimize the empirical quotient variance (3.4), the max-max algorithm[1] alternatively minimizes the function $J(m,(g_i)_i) = \frac{1}{n}\sum\limits_{i=1}^{n}\|m - g_i \cdot Y_i\|^2$ over a point m of the orbit $[m]$ and over the hidden transformation $(g_i)_{1\leqslant i \leqslant n} \in G^n$.

With these notations we can reformulate our questions as follows:

1. Is the orbit of the template $[t_0]$ a minimiser of the quotient variance defined in (3.3)? If not, the Fréchet mean in quotient space is an inconsistent estimator of $[t_0]$.

2. In this last case, can we quantify the quotient distance between $[t_0]$ and a Fréchet mean of $[X]$?

3. Can we quantify the distance between $[t_0]$ and an empirical Fréchet mean of a n-sample?

This chapter shows that the answer to the first question is usually "no" in the framework of a Hilbert space M on which a group G acts linearly and isometrically. The only exception is theorem 3.6 where the ambient space M is a manifold. In order to prove inconsistency, an important notion in this framework is the isotropy group of a point m in the ambient space. This is the subgroup which leaves this point unchanged:

$$\mathrm{Iso}(m) = \{g \in G,\ g \cdot m = m\}.$$

We start in section 3.3 with the simple example where the group is finite and the isotropy group of the template is reduced to the identity element ($\mathrm{Iso}(t_0) = \{e_G\}$, in

[1] The term max-max algorithm is used for instance in [Allassonnière 2007], and we prefer to keep the same name, even if it is a minimization.

this case t_0 is called a regular point). We turn in section 3.4 to the case of a general group and an isotropy group of the template which does not cover the whole group ($\text{Iso}(t_0) \neq G$) i.e t_0 is not a fixed point under the group action. To complete the analysis, we assume in section 3.5 that the template t_0 is a fixed point which means that $\text{Iso}(t_0) = G$.

In sections 3.3 and 3.4 we show lower and upper bounds of the consistency bias which we define as the quotient distance between the template orbit and the Fréchet mean in quotient space. These results give an answer to the second question. In section 3.4, we show a lower bound for the case of the empirical Fréchet mean which answers to the third question.

As we deal with different notions whose name or definition may seem similar, we use the following vocabulary:

1. The variance of the noisy template X in the ambient space is the function $E : m \in M \mapsto \mathbb{E}(\|m - X\|^2)$. The unique element which minimises this function is the Fréchet mean of X in the ambient space. With our assumptions it is the template t_0 itself.

2. We call variability (or noise level) of the template the value of the variance at this minimum: $\sigma^2 = \mathbb{E}(\|t_0 - X\|^2) = E(t_0)$.

3. The variance of the random orbit $[X]$ in the quotient space is the function $F : m \mapsto \mathbb{E}(d_Q([m], [X])^2)$. Notice that we define this function from the ambient space and not from the quotient space. With this definition, an orbit $[m_\star]$ is a Fréchet mean of $[X]$ if the point $m_\star$ is a global minimiser of F.

In sections 3.3 and 3.4, we exhibit a sufficient condition for the inconsistency, which is: the noisy template X takes value with a non zero probability in the set of points which are strictly closer to $g \cdot t_0$ for some $g \in G$ than the template t_0 itself. This is linked to the folding of the distribution of the noisy template when it is projected to the quotient space. The points for which the distance to the template orbit in the quotient space is equal to the distance to the template in the ambient space are projected without being folded. If the support of the distribution of the noisy template contains folded points (we only assume that the probability measure of X, noted $\mathbb{P}$, is a regular measure), then there is inconsistency. The support of the noisy template X is defined by the set of points x such that $\mathbb{P}(X \in B(x, r)) > 0$ for all $r > 0$. For different geometries of the orbit of the template, we show that this condition is fulfilled as soon as the support of the noise is large enough.

The recent article of Cleveland et al. [Cleveland 2016] may seem contradictory with our current work. Indeed the consistency of the template estimation with the Fréchet mean in quotient space is proved under hypotheses which seem to satisfy our framework: the norm is unchanged under their group action (isometric action) and a noise is present in their generative model. However we believe that the noise they consider might actually not be measurable. Indeed, their ambient space is:

$$L^2([0,1]) = \left\{ f : [0,1] \to \mathbb{R} \text{ such that } f \text{ is measurable and } \int_0^1 f^2(t)dt < +\infty \right\}.$$

The noise e is supposed to be in $L^2([0,1])$ such that for all t, $s \in [0,1]$, $\mathbb{E}(e(t)) = 0$ and $\mathbb{E}(e(t)e(s)) = \sigma^2 \mathbb{1}_{s=t}$, for $\sigma > 0$. This means that $e(t)$ and $e(s)$ are chosen without correlation as soon as $s \neq t$. In this case, it is not clear for us that the resulting function e is measurable, and thus that its Lebesgue integration makes sense. Thus, the existence of such a random process should be established before we can fairly compare the results of both works. Furthermore, no discrete version of their theorem is given.

3.3 Study of consistency for finite group

In this section, we consider a finite group G acting isometrically and effectively on $M = \mathbb{R}^n$ a finite dimensional space equipped with the euclidean norm $\| \ \|$, associated to the dot product $\langle \ , \ \rangle$.

We say that the action is effective if $x \mapsto g \cdot x$ is the identity map if and only if $g = e_G$. Note that if the action is not effective, we can define a new effective action by simply quotienting G by the subgroup of the element $g \in G$ such that $x \mapsto g \cdot x$ is the identity map.

The template is assumed to be a regular point which means that the isotropy group of the template is reduced to the neutral element of G. Note that the measure of singular points (the points which are not regular) is a null set for the Lebesgue measure (see item 1 in section 3.3.2).

Example 3.1. *The action of horizontal translation: this action is a simplified setting for image registration, where images can be obtained by the translation of one scan to another due to different poses. More precisely, we take the vector space* $M = \mathbb{R}^{\mathbb{T}}$ *where* $G = \mathbb{T} = (\mathbb{Z}/N\mathbb{Z})^D$ *is the finite torus in* D*-dimension. An element of* $\mathbb{R}^{\mathbb{T}}$ *is seen as a function* $m : \mathbb{T} \to \mathbb{R}$*, where* $m(\tau)$ *is the grey value at pixel* τ*. When* $D = 1$*,* m *can be seen like a discretised signal with* N *points, when* $D = 2$*, we can see* m *like an image with* $N \times N$ *pixels etc. We then define the group action of* $\mathbb{T}$ *on* $\mathbb{R}^{\mathbb{T}}$ *by:*

$$\tau \in \mathbb{T},\ m \in \mathbb{R}^{\mathbb{T}} \quad \tau \cdot m : x \mapsto m(x + \tau).$$

This group acts isometrically and effectively on $M = \mathbb{R}^{\mathbb{T}}$*.*

In this setting, if $\mathbb{E}(\|X\|^2) < +\infty$ then the variance of $[X]$ is well defined:

$$F : m \in M \mapsto \mathbb{E}(d_Q([X],[m])^2). \tag{3.5}$$

In this framework, F is non-negative and continuous. Then we can prove the existence of the Fréchet mean in the quotient space:

Proposition 3.1. *Let* G *be a group acting isometrically on* M *an Euclidean space, then* F *has a minimizer.*

Proof. Thanks to Cauchy-Schwarz inequality we have:

$$\lim_{\|m\| \to \infty} F(m) \geqslant \lim_{\|m\| \to \infty} \|m\|^2 - 2\|m\|\mathbb{E}(\|X\|) + \mathbb{E}(\|X\|^2) = +\infty.$$

Thus for some $R > 0$ we have: for all $m \in M$ if $\|m\| > R$ then $F(m) \geqslant F(0) + 1$. The closed ball $\overline{B(0,R)}$ is a compact set (because M is a finite vector space) then F restricted to this ball reaches its minimum at some point $m_\star$ (since F is continuous). Then for all $m \in M$, we have:

$$\begin{aligned} &\text{If } \|m\| \leqslant R \text{ then} && F(m) \geqslant F(m_\star) \\ &\text{If } \|m\| > R \text{ then} && F(m) \geqslant F(0) + 1 > F(0) \geqslant F(m_\star) \end{aligned}$$

Therefore $[m_\star]$ is a Fréchet mean of $[X]$ in the quotient $Q = M/G$. □

Note that this ensure the existence of the Fréchet mean in quotient spaces but not its uniqueness.

In this section, we show that as soon as the support of the distribution of X is big enough, the orbit of the template is not a Fréchet mean of $[X]$. We provide an upper bound of the consistency bias depending on the variability of X and an example of computation of this consistency bias.

3.3.1 Presence of inconsistency

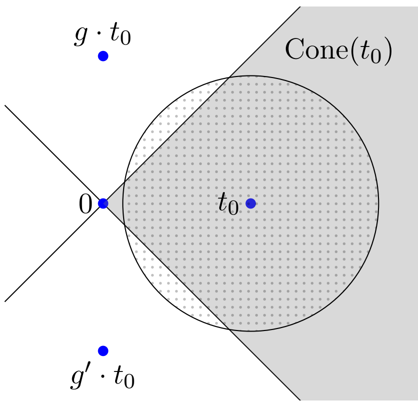


Figure 3.1: Planar representation of a part of the orbit of the template t_0. The lines are the hyperplanes whose points are equally distant of two distinct elements of the orbit of t_0, Cone(t_0) represented in points is the set of points closer from t_0 than any other points in the orbit of t_0. Theorem 3.1 states that if the support (the dotted disk) of the random variable X is not included in this cone, then there is an inconsistency.

The following theorem gives a sufficient condition on the random variable X for an inconsistency:

Theorem 3.1. *Let G be a finite group acting on $M = \mathbb{R}^n$ isometrically and effectively. Assume that the random variable X is absolutely continuous with respect to the Lebesgue's measure, with $\mathbb{E}(\|X\|^2) < +\infty$. We assume that $t_0 = \mathbb{E}(X)$ is a regular point.*

We define Cone(t_0) as the set of points closer from t_0 than any other points of the orbit $[t_0]$, see figure 3.1 or item 6 in section 3.3.2 for a formal definition.

In other words, Cone(t_0) is defined as the set of points already registered with t_0. Suppose that:

$$\mathbb{P}\left(X \notin Cone(t_0)\right) > 0, \tag{3.6}$$

then $[t_0]$ is not a Fréchet mean of $[X]$.

Because the action is isometric, this set is really a cone (this point is proved in section 3.3.2), this justifies the name. Note that is the Voronoï cell associated to the template. The Voronoï cells have been defined and used by Voronoï [Voronoi 1908] and Dirichlet [Dirichlet 1850].

The proof of theorem 3.1 is based on two steps: first, differentiating the variance F of $[X]$. Second, showing that the gradient at the template is not zero, therefore the template can not be a minimum of F. Theorem 3.2 makes the first step.

Theorem 3.2. *The variance F of $[X]$ is differentiable at any regular points. For m_0 a regular point, we define $g(x, m_0)$ as the almost unique $g \in G$ minimizing $\|m_0 - g \cdot x\|$ (in other words, $g(x, m_0) \cdot x \in Cone(m_0)$). This allows us to compute the gradient of F at m_0:*

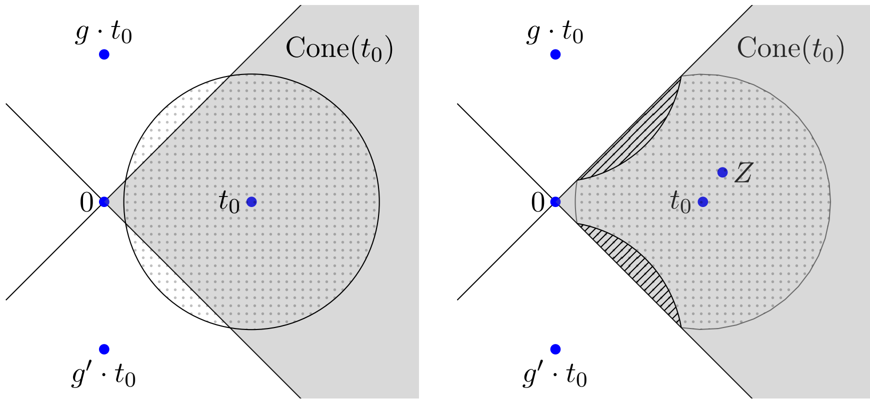
$$\nabla F(m_0) = 2(m_0 - \mathbb{E}(g(X, m_0) \cdot X)). \tag{3.7}$$

Remark 3.1. *It may seem that it is not obvious that $g(X, m_0) \cdot X$ is measurable. This is a natural requirement before taking the expected value. However, $x \mapsto m_0 - g(X, m_0) \cdot X$ is nothing else that the gradient of a measurable and differentiable function which is $m_0 \mapsto d_Q^2([X], [m_0])$ (if we admit theorem 3.2). Therefore its measurability is obvious. In chapter 5, the group will be no longer finite, and the measurability will become an issue, which will be discussed.*

This Theorem is proved in section 3.3.2. Then we show that the gradient of F at t_0 is not zero. To ensure that F is differentiable at t_0 we suppose in the assumptions of theorem 3.1 that $t_0 = \mathbb{E}(X)$ is a regular point. Thanks to theorem 3.2 we have:

$$\nabla F(t_0) = 2(t_0 - \mathbb{E}(g(X, t_0) \cdot X)).$$

Therefore $\nabla F(t_0)/2$ is the difference between two terms, which are represented on figure 3.2: on figure 3.2a there is a mass under the two hyperplanes outside Cone(t_0), so this mass is nearer from $g \cdot t_0$ for some $g \in G$ than from t_0. In the following expression $Z = \mathbb{E}(g(X, t_0) \cdot X)$, for $X \notin$ Cone(t_0), $g(X, t_0)X \in$ Cone(t_0) such points are represented in red (grid-line) on figure 3.2, (in this case, we say that X is folded). This suggests that the point $Z = \mathbb{E}(g(X, t_0) \cdot X)$ which is the mean of points in Cone(t_0) is further away from 0 than t_0. Then $\nabla F(t_0)/2 = t_0 - Z$ should be not zero, and $t_0 = \mathbb{E}(X)$ is not a critical point of the variance of $[X]$. As a conclusion $[t_0]$ is not a Fréchet mean of $[X]$. This is turned into a rigorous proof in section 3.3.3.



(a) Graphic representation of the template $t_0 = \mathbb{E}(X)$ mean of points of the support of X.

(b) Graphic representation of $Z = \mathbb{E}(g(X,t_0) \cdot X)$. The points X which were outside Cone(t_0) are now in Cone(t_0) (thanks to $g(X,t_0)$). This part, in grid-line, represents the points which have been folded.

Figure 3.2: Z is the mean of points in Cone(t_0) where Cone(t_0) is the set of points closer from t_0 than $g \cdot t_0$ for $g \in G \backslash e_G$. Therefore it seems that Z is higher that t_0, therefore $\nabla F(t_0) = 2(t_0 - Z) \neq 0$.

Note also that theorem 3.2 gives a criteria of the Fréchet means of $[X]$; if $[m_\star]$ is a Fréchet mean of $[X]$ there are two cases: $m_\star$ is not a regular point or

$$m_\star = \mathbb{E}(g(X, m_\star) \cdot X). \tag{3.8}$$

In the proof of theorem 3.1, we took M an Euclidean space and we work with the Lebesgue's measure in order to have $\mathbb{P}(X \in H) = 0$ for every hyperplane H. Therefore the proof of theorem 3.1 can be extended immediately to any Hilbert space M, if we make now the assumption that $\mathbb{P}(X \in H) = 0$ for every hyperplane H, as long as we keep a finite group acting isometrically and effectively on M.

Figure 3.2 illustrates the condition of theorem 3.1: if there is no mass beyond the hyperplanes, then the two terms in $\nabla F(t_0)$ are equal (because almost surely $g(X,t_0) \cdot X = X$). Therefore in this case we have $\nabla F(t_0) = 0$. This does not prove necessarily that there is no inconsistency, just that the template t_0 is a critical point of F.

Moreover this figure can give us an intuition on what the consistency bias (the distance between $[t_0]$ and the set of all Fréchet mean in the quotient space) depends: for t_0 a fixed regular point, when the variability of X (defined by $\mathbb{E}(\|X - t_0\|^2)$) increases the mass beyond the hyperplanes on figure 3.2 also increases, the distance between $\mathbb{E}(g(X,t_0) \cdot X)$ and t_0 (i.e. the norm of $\nabla F(t_0)$) augments. Therefore q the Fréchet mean should be further from t_0, (because at this point one should have $\nabla F(q) = 0$ or q is a singular point). Therefore the consistency bias appears to increase with the variability of X. By establishing a lower and upper bound of the consistency bias and by computing the consistency bias in a very simple case, sections 3.3.5, 3.3.6, 3.4.3 and 3.4.4 investigate how far this hypothesis is true.

3.3.2 Proof of theorem 3.2: differentiation of the variance in the quotient space

In order to show theorem 3.2 we proceed in three steps. First we see some following properties and definitions which will be used. Most of these properties are the consequences of the fact that the group G is finite. Then we show that the integrand of F is differentiable. Finally we show that we can permute gradient and integral signs.

1. The set of singular points in $\mathbb{R}^n$, is a null set (for the Lebesgue's measure), since it is equal to:
$$\bigcup_{g \neq e_G} \ker(x \mapsto g \cdot x - x),$$
a finite union of proper linear subspaces of $\mathbb{R}^n$ thanks to the linearity and effectively of the action and to the finite group.

2. If m is regular, then for g, g' two different elements of G, we pose:
$$H(g \cdot m, g' \cdot m) = \{x \in \mathbb{R}^n,\ \|x - g \cdot m\| = \|x - g' \cdot m\|\}.$$
Moreover $H(g \cdot m, g' \cdot m) = (g \cdot m - g' \cdot m)^{\perp}$ is an hyperplane.

3. For m a regular point we define the set of points which are equally distant from two different points of the orbit of m:
$$A_m = \bigcup_{g \neq g'} H(g \cdot m, g' \cdot m).$$
Then A_m is a null set. For m regular and $x \notin A_m$ the minimum in the definition of the quotient distance :
$$d_Q([m], [x]) = \min_{g \in G} \|m - g \cdot x\|, \tag{3.9}$$
is reached at a unique $g \in G$, we call $g(x, m)$ this unique element.

4. By expansion of the squared norm: g minimises $\|m - g \cdot x\|$ if and only if g maximizes $\langle m, g \cdot x \rangle$.

5. If m is regular and $x \notin A_m$ then:
$$\forall g \in G \backslash \{g(x, m)\},\ \|m - g(x, m) \cdot x\| < \|m - g \cdot x\|,$$

by continuity of the norm and by the fact that G is a finite group, we can find $\alpha > 0$, such that for $\mu \in B(m, \alpha)$ and $y \in B(x, \alpha)$:

$$\forall g \in G \backslash \{g(x, m)\} \; \|\mu - g(x, m) \cdot y\| < \|\mu - g \cdot y\|.$$

Therefore for such y and μ we have:

$$g(x, m) = g(y, \mu).$$

6. For m a regular point, we define Cone(m) the convex cone of $\mathbb{R}^n$:

$$\begin{aligned} \text{Cone}(m) &= \{x \in \mathbb{R}^n \,/\, \forall g \in G \; \|x - m\| \leqslant \|x - g \cdot m\|\} \\ &= \{x \in \mathbb{R}^n \,/\, \forall g \in G \; \langle m, x\rangle \geqslant \langle gm, x\rangle\}. \end{aligned} \tag{3.10}$$

   This is the intersection of $|G| - 1$ half-spaces (some of them could be equal): each half space is delimited by $H(m, gm)$ for $g \neq e_G$ (see figure 3.1). Cone(m) is the set of points whose projection on $[m]$ is m, (where the projection of one point p on $[m]$ is one point $g \cdot m$ which minimizes the set $\{\|p - g \cdot m\|, \, g \in G\}$).

7. Taking a regular point m allows us to see the quotient. For every point $x \in \mathbb{R}^n$ we have: $[x] \bigcap \text{Cone}(m) \neq \varnothing$, $card([x] \bigcap \text{Cone}(m)) \geqslant 2$ if and only if $x \in A_m$. The borders of the cone are $\text{Cone}(m) \backslash \text{Int}(\text{Cone}(m)) = \text{Cone}(m) \cap A_m$ (we denote by Int(A) the interior of a part A). Therefore $Q = \mathbb{R}^n / G$ can be seen like Cone(m) whose borders have been glued together.

The proof of theorem 3.2 is the consequence of the following lemmas. The first lemma studies the differentiability of the integrand, and the second allows us to permute gradient and integral sign. Let us denote by f the integrand of F:

$$\forall \, m, \, x \in M \quad f(x, m) = \min_{g \in G} \|m - g \cdot x\|^2. \tag{3.11}$$

Thus we have: $F(m) = \mathbb{E}(f(X, m))$. The min of differentiable functions is not necessarily differentiable, however we prove the following result:

Lemma 3.1. *Let m_0 be a regular point, if $x \notin A_{m_0}$ then $m \mapsto f(x, m)$ is differentiable at m_0, besides we have:*

$$\frac{\partial f}{\partial m}(x, m_0) = 2(m_0 - g(x, m_0) \cdot x) \tag{3.12}$$

Proof. If m_0 is regular and $x \notin A_{m_0}$ then we know from the item 5 of the section 3.3.2 that $g(x, m_0)$ is locally constant. Therefore around m_0, we have:

$$f(x, m) = \|m - g(x, m_0) \cdot x\|^2,$$

which can differentiate with respect to m at m_0. This proves the lemma 3.1. □

Now we want to prove that we can permute the integral and the gradient sign. The following lemma provides us a sufficient condition to permute integral and differentiation signs thanks to the dominated convergence theorem:

Lemma 3.2. *For every $m_0 \in M = \mathbb{R}^n$, we have the existence of an integrable function $\Psi : M \to \mathbb{R}^+$ (for $\mathbb{P}$the law of X) such that:*

$$\forall m \in B(m_0, 1),\ \forall x \in M \quad |f(x, m_0) - f(x, m)| \leqslant \|m - m_0\|\Psi(x), \tag{3.13}$$

where $\Psi(x) = 2\|m_0\| + 1 + 2\|x\|$ +

Proof. For all $g \in G$, $m \in M$ we have:

$$\begin{aligned}
\|g \cdot x - m_0\|^2 - \|g \cdot x - m\|^2 &= \langle m - m_0, 2g \cdot x - (m_0 + m)\rangle \\
&\leqslant \|m - m_0\| \times (\|m_0 + m\| + \|2x\|) \\
\min_{g \in G} \|g \cdot x - m_0\|^2 &\leqslant \|m - m_0\| (\|m_0 + m\| + \|2x\|) + \|g \cdot x - m\|^2 \\
\min_{g \in G} \|g \cdot x - m_0\|^2 &\leqslant \|m - m_0\| (\|m_0 + m\| + \|2x\|) + \min_{g \in G} \|g \cdot x - m\|^2 \\
\min_{g \in G} \|g \cdot x - m_0\|^2 - \min_{g \in G} \|g \cdot x - m\|^2 &\leqslant \|m - m_0\| (2\|m_0\| + \|m - m_0\| + \|2x\|)
\end{aligned}$$

By symmetry we get also the same control of $f(x, m) - f(x, m_0)$, then:

$$|f(x, m_0) - f(x, m)| \leqslant \|m_0 - m\| (2\|m_0\| + \|m - m_0\| + \|2x\|) \tag{3.14}$$

The function Ψ should depend on x or m_0, but not on m. That is why we take only $m \in B(m_0, 1)$, then we replace $\|m - m_0\|$ by 1 in (3.14), which concludes. □

3.3.3 Proof of theorem 3.1: the gradient is not zero at the template

To prove it, we suppose that $\nabla F(t_0) = 0$, and we take the dot product with t_0:

$$\langle \nabla F(t_0), t_0\rangle = 2\mathbb{E}(\langle X, t_0\rangle - \langle g(X, t_0) \cdot X, t_0\rangle) = 0. \tag{3.15}$$

The item 4 of $(x, m) \mapsto g(x, m)$ seen at section 3.3.2 leads to:

$$\langle X, t_0\rangle - \langle g(X, t_0) \cdot X, t_0\rangle \leqslant 0 \text{ almost surely.}$$

So the expected value of a non-positive random variable is null. Then

$$\begin{aligned}
\langle X, t_0\rangle - \langle g(X, t_0) \cdot X, t_0\rangle &= 0 \text{ almost surely} \\
\langle X, t_0\rangle &= \langle g(X, t_0) \cdot X, t_0\rangle \text{ almost surely.}
\end{aligned}$$

Then $g = e_G$ maximizes the dot product almost surely. Therefore (as we know that $g(X, t_0)$ is unique almost surely, since t_0 is regular):

$$g(X, t_0) = e_G \text{ almost surely,}$$

which is a contradiction with Equation (3.6).

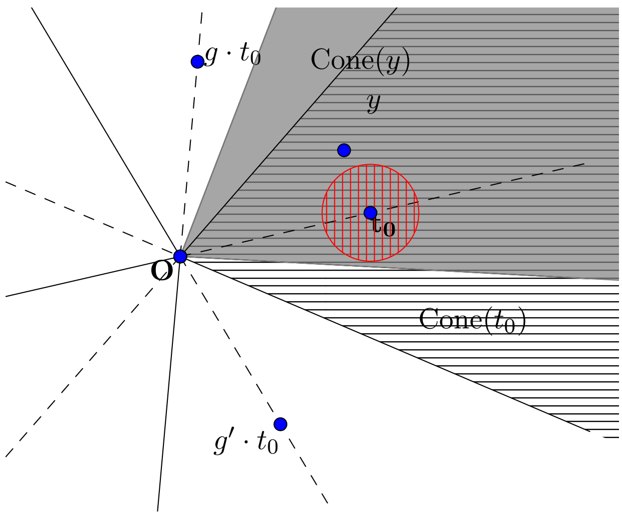


Figure 3.3: $y \mapsto \text{Cone}(y)$ is continuous. When the support of the X is bounded and included in the interior of Cone(t_0) the hatched cone. For y sufficiently close to the template t_0, the support of the X (the ball in red) is still included in Cone(y) (in grey), then $F(y) = (\mathbb{E}(\|X - y\|^2)$. Therefore in this case, $[t_0]$ is at least a Karcher mean of $[X]$.

3.3.4 Study of consistency when the support of X is included in the cone of the template

We can also wonder if the converse of theorem 3.1 is true: if the support is included in Cone(t_0), is there consistency? We do not have a general answer to that. In the simple example section 3.3.6 it happens that condition (3.6) is necessary and sufficient. More generally the following proposition provides a partial converse:

Proposition 3.2. *If the support of X is a compact set included in the interior of Cone(t_0), then the orbit of the template $[t_0]$ is at least a Karcher mean of $[X]$ (a Karcher mean is a local minimum of the variance).*

Sketch of the proof: If the support of X is a compact set included in the interior of Cone(t_0) then we know that X-almost surely: $d_Q([X], [t_0]) = \|X - t_0\|$. Thus the variance at t_0 in the quotient space is equal to the variance at t_0 in the ambient space. Now if we assume a sort of continuity of the cone (see figure 3.3) for y in a small neighbourhood of t_0, the support of X is still included in the interior of Cone(y). We still have $d_Q([X], [y]) = \|X - y\|$ X-almost surely. In other words, locally around t_0, the variance in the quotient space is equal to the variance in the ambient space. Moreover we know that $t_0 = \mathbb{E}(X)$ is the only global minimiser of the variance of X: $m \mapsto E(\|m - X\|^2) = E(m)$. Therefore t_0 is a local minimum of F the variance in the quotient space (since the two variances are locally equal). Therefore $[t_0]$ is at least a Karcher mean of $[X]$ in this case.

However, our notion of continuity of the cone is too vague, therefore we give a more rigorous proof based on the same idea:

Proof. We know that we have $\text{Support}(X) \subset \text{Cone}(t_0)$ and we want to show that:

$$\exists \eta > 0 \text{s.t.} \forall y \in B(t_0, r) \quad \text{Support}(X) \subset \text{Cone}(y) \tag{3.16}$$

Let us assume that equation (3.16) does not hold. In this case we have:

$$\forall n \in \mathbb{N} \quad \exists y_n \in B(t_0, \frac{1}{n}) \quad \exists z_n \in \text{Support}(X) \text{ and } z_n \notin \text{Cone}(y_n)$$

We have that $(y_n)_n$ converges to t_0. As the support of ε is assumed to be compact, by extraction, without loss of generality we can assume that z_n converges to z. First $z \in \text{Support}(X)$ (because $\text{Support}(X)$ is a closed set. Now as $z_n \notin \text{Cone}(y_n)$, there exists $g_n \in G$ such that $\langle z_n, y_n - g_n \cdot y_n \rangle < 0$. As G is a finite group, by extraction, without loss of generality, we can assume that $(g_n)_n$ is a constant sequence. Then we have $\langle z_n, y_n - gy_n \rangle < 0$. By taking the limit, we have that $\langle z, t_0 - gt_0 \rangle \leqslant 0$. As a consequence z is not in the interior of the cone of the template. This is absurd, since $z \in \text{Support}(X) \subset \text{Int}(\text{Cone}(t_0))$. Therefore we have proved that equation (3.16) holds. As a consequence,

$$\forall y \in B(t_0, \eta), F(y) = \mathbb{E}(\|X - y\|^2) \geqslant \mathbb{E}(\|X - t_0\|^2) = F(t_0).$$

Lastly for η sufficiently small we have $B(t_0, \eta) \subset \text{Cone}(t_0)$. This proves that $[t_0]$ is a minimum of the variance restricted to $B([t_0], \eta)$ (the open ball in Q of center $[t_0]$ and radius η). This proves that $[t_0]$ is a Karcher mean of $[X]$.

□

Can the template be a Fréchet mean instead of being only a Karcher mean? The following simple example show that the template can be Karcher mean without being a Fréchet mean.

Example 3.2. *Let us take* $M = \mathbb{R}^2$ *the euclidean plane.* $G = \{I_2, -I_2\}$ *acts isometrically on* M. *On this example, the quotient distance is* $d_Q([a],[b]) = \min(\|a - b\|, \|a + b\|)$. *Let us take* A *and* B *two points, let* X *a random variable such that* $\mathbb{P}(X = A) = \mathbb{P}(X = B) = \frac{1}{2}$. *Then* $X = t_0 + \varepsilon$, *where* $t_0 = \frac{A+B}{2}$, *and* ε *is a noise such that* $\mathbb{E}(\varepsilon) = 0$.

Now, all we have to do is to chose wisely A *and* B. *We propose* $A = (1, 2.7)$ *and* $B = (1, 0.7)$ *then* $t_0 = (1, 1)$, A, B *are inside the interior of Cone*(t_0), *(here Cone*$(t_0) = \{x \in \mathbb{R}^2, \text{ s.t. } \langle x, t_0 \rangle \geqslant 0\}$*). The variance in the quotient space is defined by:*

$$F(m) = \frac{1}{2} \left(\min(\|m - A\|^2, \|m + A\|^2) + \min(\|m - B\|^2, \|m + B\|^2) \right)$$

On this example it easy to verify that $2 = F(\frac{A-B}{2}) < F(t_0) = 2.89$. *And the Fréchet mean of* $[X]$ *is exactly* $[m_\star] = [\frac{A-B}{2}]$. *Therefore it is possible that the template estimation is inconsistent even if the random variable* X *is included into the cone of the template.*

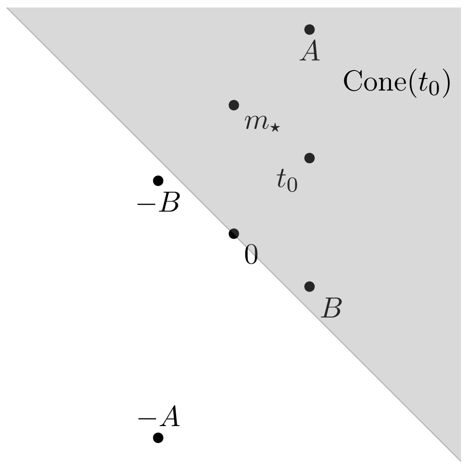


Figure 3.4: Representation of the template t_0, A, B, in gray the cone of t_0.

3.3.5 Upper bound of the consistency bias

In this Subsection we show an explicit upper bound of the consistency bias.

Theorem 3.3. *When G is a finite group acting isometrically on $M = \mathbb{R}^n$, we denote $|G|$ the cardinal of the group G. If X is Gaussian vector: $X \sim \mathcal{N}(t_0, w^2 Id_{\mathbb{R}^n})$, and $m_\star \in$ argmin F, then we have the upper bound of the consistency bias:*

$$d_Q([t_0],[m_\star]) \leqslant w\sqrt{8\ln(|G|)}. \tag{3.17}$$

When $X \sim \mathcal{N}(t_0, w^2 Id_n)$ the variability of X is $\sigma^2 = \mathbb{E}(||X - t_0||^2) = nw^2$ and we can write the upper bound of the bias: $d_Q([t_0],[m_\star]) \leqslant \frac{\sigma}{\sqrt{n}}\sqrt{8\ln|G|}$. This Theorem shows that the consistency bias is low when the variability of X is small, which tends to confirm our hypothesis in section 3.3.1. It is important to notice that this upper bound explodes very slowly when the cardinal of the group tends to infinity.

In order to show this Theorem, we use the following lemma:

Lemma 3.3. *We write $X = t_0 + \varepsilon$ where $\mathbb{E}(\varepsilon) = 0$ and we make the assumption that the noise ε is a subgaussian random variable. This means that it exists $c > 0$ and $w > 0$ such that:*

$$\forall m \in M = \mathbb{R}^n,\ \mathbb{E}(\exp(\langle \varepsilon, m\rangle)) \leqslant c\exp\left(\frac{w^2\|m\|^2}{2}\right). \tag{3.18}$$

For $m \in M$ we note $\tilde{\rho} = d_Q([m],[t_0]) = \inf_{g\in G} \|g\cdot m - t_0\|$. If we have:

$$\tilde{\rho} \geqslant w\sqrt{2\ln(c|G|)}, \tag{3.19}$$

then we have:

$$\tilde{\rho}^2 - \tilde{\rho}w\sqrt{8\ln(c|G|)} \leqslant F(m) - \mathbb{E}(\|\varepsilon\|^2). \tag{3.20}$$

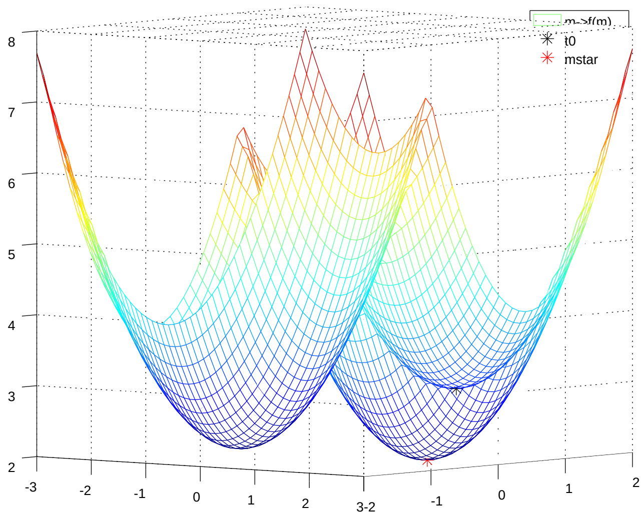


Figure 3.5: Variation of the variance $m \mapsto F(m)$.

Proof of lemma 3.3. First we expand the right member of the inequality (3.20):

$$\mathbb{E}(\|\varepsilon\|^2) - F(m) = \mathbb{E}\left(\max_{g\in G}(\|X-t_0\|^2 - \|X - g\cdot m\|^2)\right)$$

We use the formula $\|A\|^2 - \|A+B\|^2 = -2\langle A, B\rangle - \|B\|^2$ with $A = X - t_0$ and $B = t_0 - g\cdot m$:

$$\mathbb{E}(\|\varepsilon\|^2) - F(m) = \mathbb{E}\left[\max_{g\in G}\left(-2\langle X - t_0, t_0 - g\cdot m\rangle - \|t_0 - gm\|^2\right)\right] = \mathbb{E}(\max_{g\in G}\eta_g), \tag{3.21}$$

with $\eta_g = -\|t_0 - g\cdot m\|^2 + 2\langle \varepsilon, gm - t_0\rangle$. Our goal is to find a lower bound of $F(m) - \mathbb{E}(\|\varepsilon\|^2)$, that is why we search an upper bound of $\mathbb{E}(\max_{g\in G}\eta_g)$ with the Jensen's inequality. We take $x > 0$ and we get by using the assumption (3.18):

$$\begin{aligned}
\exp(x\mathbb{E}(\max_{g\in G}\eta_g)) &\leqslant \mathbb{E}(\exp(\max_{g\in G} x\eta_g)) \leqslant \mathbb{E}\left(\sum_{g\in G}\exp(x\eta_g)\right) \\
&\leqslant \sum_{g\in G}\exp(-x\|t_0 - gm\|^2)\mathbb{E}(\exp(\langle\varepsilon, 2x(gm - t_0)\rangle)) \\
&\leqslant c\sum_{g\in G}\exp(-x\|t_0 - gm\|^2)\exp(2w^2x^2\|gm - t_0\|^2) \\
&\leqslant c\sum_{g\in G}\exp(\|gm - t_0\|^2(-x + 2x^2w^2)) \qquad (3.22)
\end{aligned}$$

Now if $(-x + 2w^2x^2) < 0$, we can take an upper bound of the sum sign in (3.22) by taking the smallest value in the sum sign, which is reached when g minimizes $\|g \cdot m - t_0\|$ multiplied by the number of elements summed. Moreover:

$$(-x + 2x^2w^2) < 0 \Longleftrightarrow 0 < x < \frac{1}{2w^2}.$$

Then we have:

$$\exp(x\mathbb{E}(\max_{g\in G} \eta_g)) \leqslant c|G| \exp(\tilde{\rho}^2(-x + 2x^2w^2)) \text{ as soon as } 0 < x < \frac{1}{2w^2}.$$

Then by taking the logarithm map:

$$\mathbb{E}(\max_{g\in G}\eta_g) \leqslant \frac{\ln c|G|}{x} + (2xw^2 - 1)\tilde{\rho}^2. \tag{3.23}$$

Now we find the x which optimizes inequality (3.23). By differentiation, the right member of inequality (3.23) is minimal for $x_\star = \sqrt{\ln c|G|/2}/(w\tilde{\rho})$ which is a valid choice because $x_\star \in (0, \frac{1}{2w^2})$ by using the assumption (3.19). With the equations (3.21) and (3.23) and $x_\star$ we get the result. □

Proof of theorem 3.3. We take $m_\star \in \operatorname{argmin} F$, $\tilde{\rho} = d_Q([m_\star], [t_0])$, and $\varepsilon = X - t_0$. We have: $F(m_\star) \leqslant F(t_0) \leqslant \mathbb{E}(\|\varepsilon\|^2)$ then $F(m_\star) - \mathbb{E}(\|\varepsilon\|^2) \leqslant 0$. If $\tilde{\rho} > w\sqrt{2\ln(|G|)}$ then we can apply lemma 3.3 with $c = 1$. Thus:

$$\tilde{\rho}^2 - \tilde{\rho}w\sqrt{8\ln(|G|)} \leqslant 2F(m_\star) - \mathbb{E}(\|\varepsilon\|^2) \leqslant 0,$$

which yields to $\tilde{\rho} \leqslant w\sqrt{8\ln(|G|)}$. If $\tilde{\rho} \leqslant w\sqrt{2\ln(|G|)}$, we have nothing to prove. □

Note that the proof of this upper bound does not use the fact that the action is isometric, therefore this upper bound is true for every finite group action. More precisely, when the action is not isometric, if $m_\star$ minimizes $m \mapsto F(m) = \mathbb{E}(\inf_{g\in G} \|X - g \cdot m\|^2)$ then $\inf_{g\in G}\|t_0 - g \cdot m_\star\| \leqslant w\sqrt{8\ln|G|}$.

3.3.6 Study of the consistency bias in a simple example

In this Subsection, we take a particular case of example 3.1: the action of horizontal translation with $\mathbb{T} = \mathbb{Z}/2\mathbb{Z}$. We identify $\mathbb{R}^{\mathbb{T}}$ with $\mathbb{R}^2$ and we note by $(u, v)^T$ an element of $\mathbb{R}^{\mathbb{T}}$. In this setting, one can completely describe the action of $\mathbb{T}$ on $\mathbb{R}^{\mathbb{T}}$: $0 \cdot (u, v)^T = (u, v)^T$ and $1 \cdot (u, v)^T = (v, u)^T$. The set of singularities is the line $L = \{(u, u)^T,\ u \in \mathbb{R}\}$. We note $HP_A = \{(u, v)^T,\ v > u\}$ the half-plane above L and HP_B the half-plane below L.

This simple example will allow us to provide necessary and sufficient condition for an inconsistency at regular and singular points. Moreover we can compute exactly the consistency bias, and exhibit which parameters govern the bias. We can then find an equivalent of the consistency bias when the noise tends to zero or infinity. More precisely, we have the following theorem:

Proposition 3.3. *Let X be a random variable such that $\mathbb{E}(\|X\|^2) < +\infty$ and $t_0 = \mathbb{E}(X)$.*

1. *If $t_0 \in L$, there is no inconsistency if and only if the support of X is included in the line $L = \{(u,u),\ u \in \mathbb{R}\}$. If $t_0 \in HP_A$ (respectively in HP_B), there is no inconsistency if and only if the support of X is included in $HP_A \cup L$ (respectively in $HP_B \cup L$).*

2. *If X is Gaussian: $X \sim \mathcal{N}(t_0, w^2 Id_2)$, then the Fréchet mean of $[X]$ exists and is unique. This Fréchet mean $[m_\star]$ is on the line passing through $\mathbb{E}(X)$ and perpendicular to L and the consistency bias $\tilde{\rho} = d_Q([t_0],[m_\star])$ is the function of s and $d = dist(t_0, L)$ given by:*

$$\tilde{\rho}(d,s) = w\frac{2}{\pi}\int_{\frac{d}{w}}^{+\infty} r^2 \exp\left(-\frac{r^2}{2}\right)\kappa\left(\frac{d}{rw}\right)dr, \tag{3.24}$$

where κ is a non-negative function on $[0,1]$ defined by $\kappa(x) = \sin(\arccos(x)) - x\arccos(x)$.

   (a) *If $d > 0$ then $w \mapsto \tilde{\rho}(d,w)$ has an asymptotic linear expansion:*

$$\tilde{\rho}(d,w) \underset{w\to\infty}{\sim} w\frac{2}{\pi}\int_0^{+\infty} r^2 \exp\left(-\frac{r^2}{2}\right)dr. \tag{3.25}$$

   (b) *If $d > 0$, then $\tilde{\rho}(d,w) = o(w^k)$ as $w \to 0$, for all $k \in \mathbb{N}$.*

   (c) *$w \mapsto \tilde{\rho}(0,w)$ is linear with respect to w (for $d = 0$ the template is a fixed point).*

Remark 3.2. *Here, contrarily to the case of the action of rotation in [Miolane 2017], it is not the ratio $\|\mathbb{E}(X)\|$ over the noise which matters to estimate the consistency bias. Rather the ratio $dist(\mathbb{E}(X), L)$ over the noise. However in both cases we measure the distance between the signal and the singularities which was $\{0\}$ in [Miolane 2017] for the action of rotations, L in this case.*

Proof. We suppose that $\mathbb{E}(X) \in HP_A \cup L$. In this setting we call $\tau(x,m)$ one of element of the group $G = \mathbb{T}$ which minimizes $\|\tau \cdot x - m\|$ see (3.9) instead of $g(x,m)$. The variance in the quotient space at the point m is:

$$F(m) = \mathbb{E}\left(\min_{\tau\in\mathbb{Z}/2\mathbb{Z}} \|\tau\cdot X - m\|^2\right) = \mathbb{E}(\|\tau(X,m)\cdot X - m\|^2).$$

As we want to minimize F and $F(1\cdot m) = F(m)$, we can suppose that $m \in HP_A \cup L$. We can find the value of $\tau(x,m)$ for $x \in M$:

- If $x \in HP_A \cup L$ we can set $\tau(x,m) = 0$ (because in this case x, m are on the same half plane delimited by L the perpendicular bisector of m and $-m$).

- If $x \in HP_B$ then we can set $\tau(x,m) = 1$ (because in this case x, m are not on the same half plane delimited by L the perpendicular bisector of m and $-m$).

The map $[x] \mapsto g(x,m) \cdot x$ will be called an congruent section in section 5.4.1. This allows use to write the variance at the point $m \in HP_A$:

$$F(m) = \left(\mathbb{E}\left(\|X - m\|^2 \mathbb{1}_{\{X \in HP_A \cup L\}}\right) + \mathbb{E}\left(\|1 \cdot X - m\|^2 \mathbb{1}_{\{X \in HP_B\}}\right)\right)$$

Then we define the random variable Z by: $Z = X\mathbb{1}_{X \in HP_A \cup L} + 1 \cdot X\mathbb{1}_{X \in HP_B}$, such that for $m \in HP_A$ we have: $F(m) = \mathbb{E}(\|Z - m\|^2)$ and $F(m) = F(1 \cdot m)$. Thus if $m_\star$ is a global minimiser of F, then $m_\star = \mathbb{E}(Z)$ or $m_\star = 1 \cdot \mathbb{E}(Z)$. So the Fréchet mean of $[X]$ is $[\mathbb{E}(Z)]$. Here instead of using theorem 3.1, we can work explicitly: Indeed there is no inconsistency if and only if $\mathbb{E}(Z) = \mathbb{E}(X)$, ($\mathbb{E}(Z) = 1 \cdot \mathbb{E}(X)$ would be another possibility, but by assumption $\mathbb{E}(Z)$, $\mathbb{E}(X) \in HP_A$), by writing $X = X\mathbb{1}_{X \in HP_A} + X\mathbb{1}_{X \in HP_B \cup L}$, we have:

$$\begin{aligned}
\mathbb{E}(Z) = \mathbb{E}(X) &\iff \mathbb{E}(1 \cdot X\mathbb{1}_{X \in HP_B \cup L}) = \mathbb{E}(X\mathbb{1}_{X \in HP_B \cup L}) \\
&\iff 1 \cdot \mathbb{E}(X\mathbb{1}_{X \in HP_B \cup L}) = \mathbb{E}(X\mathbb{1}_{X \in HP_B \cup L}) \\
&\iff \mathbb{E}(X\mathbb{1}_{X \in HP_B \cup L}) \in L \\
&\iff \mathbb{P}(X \in HP_B) = 0,
\end{aligned}$$

Therefore there is an inconsistency if and only if $\mathbb{P}(X \in HP_B) > 0$ (we remind that we made the assumption that $\mathbb{E}(X) \in HP_A \cup L$). If $\mathbb{E}(X)$ is regular (i.e. $\mathbb{E}(X) \notin L$), then there is an inconsistency if and only if X takes values in HP_B, (this is exactly the condition of theorem 3.1, but in this particular case, this is a necessarily and sufficient condition). This proves point 1.

Now we make the assumption that X follows a Gaussian noise in order compute $\mathbb{E}(Z)$ (note that we could take another noise, as long as we are able to compute $\mathbb{E}(Z)$). For that we convert to polar coordinates: we write $(u,v)^T$, a vector of $\mathbb{R}^{\mathbb{Z}/2\mathbb{Z}}$, under the form:

$$(u,v)^T = \mathbb{E}(X) + (r\cos\gamma, r\sin\gamma)^T,$$

where $r > 0$ and $\gamma \in [0, 2\pi]$. We also define: $d = \operatorname{dist}(\mathbb{E}(X), L)$, $\mathbb{E}(X)$ is a regular point if and only if $d > 0$. We still suppose that $\mathbb{E}(X) = (\alpha, \beta)^T \in HP_A \cup L$. First we parametrise in function of (r, γ) the points which are in HP_B:

$$\begin{aligned}
v < u &\iff \beta + r\sin\gamma < \alpha + r\cos\gamma \iff \frac{\beta - \alpha}{r} < \sqrt{2}\cos(\gamma + \frac{\pi}{4}) \\
&\iff \frac{d}{r} < \cos(\gamma + \frac{\pi}{4}) \\
&\iff \gamma \in \left[-\frac{\pi}{4} - \arccos(d/r), -\frac{\pi}{4} + \arccos(d/r)\right] \text{ and } d < r
\end{aligned}$$

Then we compute $\mathbb{E}(Z)$:

$$\mathbb{E}(Z) = \mathbb{E}(X\mathbb{1}_{X\in HP_A}) + \mathbb{E}(1\cdot X\mathbb{1}_{X\in HP_B})$$

$$\mathbb{E}(Z) = \int_0^d\int_0^{2\pi}\begin{pmatrix}\alpha + r\cos\gamma\\ \beta + r\sin\gamma\end{pmatrix}\frac{\exp\left(-\frac{r^2}{2w^2}\right)}{2\pi w^2}rd\gamma dr$$

$$+\int_d^{+\infty}\int_{\arccos(\frac{d}{r})-\frac{\pi}{4}}^{2\pi-\frac{\pi}{4}-\arccos(\frac{d}{r})}\begin{pmatrix}\alpha + r\cos\gamma\\ \beta + r\sin\gamma\end{pmatrix}\frac{\exp\left(-\frac{r^2}{2w^2}\right)}{2\pi w^2}rdrd\gamma$$

$$+\int_d^{+\infty}\int_{-\frac{\pi}{4}-\arccos\left(\frac{d}{r}\right)}^{-\frac{\pi}{4}+\arccos\left(\frac{d}{r}\right)}\begin{pmatrix}\beta + r\sin\gamma\\ \alpha + r\cos\gamma\end{pmatrix}\frac{\exp\left(-\frac{r^2}{2w^2}\right)}{2\pi w^2}rdrd\gamma$$

$$= E(X) + \int_d^{+\infty}\frac{r^2\exp(-\frac{r^2}{2w^2})}{\pi w^2}\sqrt{2}\kappa\left(\frac{d}{r}\right)dr\times(-1,1)^T,$$

We compute $\tilde{\rho} = d_Q([\mathbb{E}(X)],[\mathbb{E}(Z)])$ where d_Q is the distance in the quotient space defined in (3.1). As we know that $\mathbb{E}(X),\mathbb{E}(Z)$ are in the same half-plane delimited by L, we have: $\tilde{\rho} = d_Q([\mathbb{E}(Z)],[\mathbb{E}(X)]) = \|\mathbb{E}(Z)-\mathbb{E}(X)\|$. This proves equation (3.24), note that items 2a to 2c are the direct consequence of equation (3.24) and basic analysis. □

3.4 Inconsistency for finite and infinite group when the template is not a fixed point

In the previous section, we prove the inconsistency when the group was finite. Being a finite group was a restriction to applications, therefore we now extend to non finite group. However, we still assume isometric action. In section 3.3 we exhibited sufficient condition to have an inconsistency, restricted to the case of finite group acting on an Euclidean space. We now generalize this analysis to Hilbert spaces of any dimension included infinite. Let M be such a Hilbert space with its dot product noted by $\langle\ ,\ \rangle$ and its associated norm $\|\ \|$. In this section, we do not anymore suppose that the group G is finite. In the following, we prove that there is an inconsistency in a large number of situations, and we quantify the consistency bias with lower and upper bounds.

Example 3.3. *The action of continuous horizontal translation: We take* $G = (\mathbb{R}/\mathbb{Z})^D$ *acting on* $M = L^2((\mathbb{R}/\mathbb{Z})^D,\mathbb{R})$ *with:*

$$\forall\tau\in G\quad\forall f\in M\quad(\tau\cdot f): t\mapsto f(t+\tau)$$

This isometric action is the continuous version of the example 3.1: the elements of M *are now continuous images in dimension* D*.*

3.4.1 Presence of an inconsistency

We state here a generalization of theorem 3.1:

Theorem 3.4. *Let G be a group acting isometrically on M a Hilbert space, and X a random variable in M, $\mathbb{E}(\|X\|^2) < +\infty$ and $\mathbb{E}(X) = t_0 \neq 0$. If:*

$$\mathbb{P}\left(d_Q([t_0],[X]) < \|t_0 - X\|\right) > 0, \tag{3.26}$$

or equivalently:

$$\mathbb{P}\left(\sup_{g\in G} \langle g\cdot X, t_0\rangle > \langle X, t_0\rangle\right) > 0. \tag{3.27}$$

Then $[t_0]$ is not a Fréchet mean of $[X]$ in $Q = M/G$.

The condition of this Theorem is the same condition of theorem 3.1: the support of the law of X contains points closer from $g \cdot t_0$ for some g than t_0. Thus the condition (3.27) is equivalent to $\mathbb{E}(d_Q([X],[t_0])^2) < \mathbb{E}(\|X - t_0\|^2)$. In other words, the variance in the quotient space at t_0 is strictly smaller than the variance in the ambient space at t_0.

Proof. First the two conditions are equivalent by definition of the quotient distance and by expansion of the square norm of $\|t_0 - X\|$ and of $\|t_0 - gX\|$ for $g \in G$.

As above, we define the variance of $[X]$ by:

$$F(m) = \mathbb{E}\left(\inf_{g\in G}\|g\cdot X - m\|^2\right).$$

In order to prove this Theorem, we find a point m such that $F(m) < F(t_0)$, which directly implies that $[t_0]$ is not be a Fréchet mean of $[X]$.

In the proof of theorem 3.1, we showed that under condition (3.6) we had $\langle\nabla F(t_0), t_0\rangle < 0$. This leads us to study F restricted to $\mathbb{R}^+ t_0$, we define for $\lambda \in \mathbb{R}^+$:

$$f(\lambda) = F(\lambda t_0) = \mathbb{E}\left(\inf_{g\in G}\|g\cdot X - \lambda t_0\|^2\right).$$

Thanks to the isometric action we can expand $f(\lambda)$ by:

$$f(\lambda) = \lambda^2\|t_0\|^2 - 2\lambda\mathbb{E}\left(\sup_{g\in G}\langle g\cdot X, t_0\rangle\right) + \mathbb{E}(\|X\|^2), \tag{3.28}$$

and explicit the unique element of $\mathbb{R}^+$ which minimizes f:

$$\lambda(t_0) = \frac{\mathbb{E}\left(\sup_{g\in G}\langle g\cdot X, t_0\rangle\right)}{\|t_0\|^2}. \tag{3.29}$$

For all $x \in M$, we have $\sup_{g\in G}\langle g\cdot x, t_0\rangle \geqslant \langle x, t_0\rangle$ and thanks to condition (3.27) we get:

$$\begin{aligned}\mathbb{E}\left(\sup_{g\in G}\ \langle g\cdot X,t_0\rangle\right) =&\mathbb{E}\left(\sup_{g\in G}\ \langle g\cdot X,t_0\rangle \mathbb{1}_{X\in \mathrm{Cone}(t_0)}\right)\\ &+\mathbb{E}\left(\sup_{g\in G}\ \langle g\cdot X,t_0\rangle \mathbb{1}_{X\notin \mathrm{Cone}(t_0)}\right)\\ =&\mathbb{E}\left(\langle X,t_0\rangle \mathbb{1}_{X\in \mathrm{Cone}(t_0)}\right)\\ &+\mathbb{E}\left(\sup_{g\in G}\ \langle g\cdot X,t_0\rangle \mathbb{1}_{X\notin \mathrm{Cone}(t_0)}\right)\\ >&\mathbb{E}(\langle X,t_0\rangle) = \langle \mathbb{E}(X),t_0\rangle = \|t_0\|^2, \end{aligned} \tag{3.30}$$

which implies $\lambda(t_0) > 1$. Then $F(\lambda(t_0)t_0) < F(t_0)$. □

Note that $\|t_0\|^2(\lambda(t_0)-1) = \mathbb{E}\left(\sup_{g\in G}\ \langle g\cdot X,t_0\rangle\right) - \mathbb{E}(\langle X,t_0\rangle)$ (which is positive) is exactly $-\langle \nabla F(t_0),t_0\rangle/2$ in the case of finite group, see Equation (3.15). Here we find the same expression without having to differentiate the variance F, which may be not possible in the current setting.

3.4.2 Analysis of the condition in theorem 3.4

We now look for general cases when we are sure that Equation (3.27) holds which implies the presence of inconsistency. We saw in section 3.3 that when the group is finite, it is possible to have no inconsistency only if the support of the law is included in a cone delimited by some hyperplanes. The hyperplanes were defined as the set of points equally distant of the template t_0 and $g\cdot t_0$ for $g \in G$. Therefore if the cardinal of the group becomes more and more important, one could think that in order to have no inconsistency the space where X should takes value becomes smaller and smaller. At the limit it leaves only at most an hyperplane. In the following, we formalise this idea to make it rigorous. We show that the cases where theorem 3.4 cannot be applied are not generic cases.

First we can notice that it is not possible to have the condition (3.27) if t_0 is a fixed point under the action of G. Indeed in this case $\langle g\cdot X,t_0\rangle = \left\langle X,g^{-1}t_0\right\rangle = \langle X,t_0\rangle)$. So from now, we suppose that t_0 is not a fixed point. Now let us see some settings when we have the condition (3.26) and thus condition (3.27). First, let us recall the definition of the Voronoï cell of a point t_0:

$$\mathrm{Cone}(t_0) = \{x\in M,\quad \forall g\in G, \|x-t_0\| \leqslant \|x-g\cdot t_0\|\}$$

The structure of this paragraph will be always the same, first we give, in some lemmas, a property of this cone, then we deduce a case where we ensure inconsistency.

Lemma 3.4. *Let G be a group acting isometrically on a Hilbert space M. Let t_0 be a point in M. Assume that t_0 is a limit point of $[t_0]$, this means that t_0 can be a limit of a sequence belonging to $[t_0]\backslash\{t_0\}$. In this case we have:*

$$t_0 \notin Int(Cone(t_0)),$$

where $Int(A)$ is the interior of A.

Proof. Let η be a positive number, we have to prove that $B(t_0, \eta)$, the open ball centred at t_0 with a radius η, is not included in the cone of t_0. By density, one takes $g \cdot t_0 \in B(t_0, \eta)\backslash\{t_0\}$ for some $g \in G$, now if we take r such that

$$r < \min(\|g \cdot t_0 - t_0\|/2, \eta - \|g \cdot t_0 - t_0\|),$$

then $B(g \cdot t_0, r) \subset B(t_0, \eta)$. Besides, for every $x \in B(g \cdot t_0, r)$ we have $\|x - g \cdot t_0\| < \|x - t_0\|$. Then $x \notin \text{Cone}(t_0)$. □

Proposition 3.4. *Let G be a group acting isometrically on a Hilbert space M, and X a random variable in M, with $\mathbb{E}(\|X\|^2) < +\infty$ and $\mathbb{E}(X) = t_0 \neq 0$. If:*

1. *The template t_0 is a limit point in $[t_0]$.*
2. *There exists $\eta > 0$ such that the support of X contains a ball $B(t_0, \eta)$.*

Then condition (3.27) holds, and the estimator is inconsistent according to theorem 3.4.

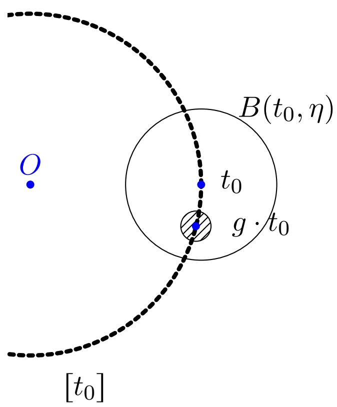


Figure 3.6: The smallest disk is included in the support of X and the points in that disk is closer from $g \cdot t_0$ than from t_0. According to theorem 3.4 there is an inconsistency.

Proof. Thanks to lemma 3.4, the template t_0 is not in the interior of $\text{Cone}(t_0)$. Therefore $B(t_0, \eta)$ is not included in $\text{Cone}(t_0)$. Then we verify condition (3.27), and we can apply theorem 3.4. □

Proposition 3.4 proves that there is a large number of cases where we can ensure the presence of an inconsistency. For instance when M is a finite dimensional vector space and the random variable X has a continuous positive density (for the Lebesgue's measure) at t_0, condition 2 of proposition 3.4 is fulfilled. Unfortunately this proposition do not cover the case where there is no mass at the expected value $t_0 = \mathbb{E}(X)$. This situation could appear if X has two modes for instance. The following proposition deals with this situation:

Lemma 3.5. *Let G be a group acting isometrically on M. If we assume that the orbit of a point t_0 contains a differential curve:*

$$\exists \varphi \text{ s.t. } \varphi : (-a, a) \to [t_0] \text{ is } \mathcal{C}^1 \text{ with } \varphi(0) = t_0, \varphi'(0) = v \neq 0.$$

In this case we have (we note $v^\perp$, the set of points orthogonal to v).

$$Cone(t_0) \subset v^\perp.$$

Proof. In order to prove this lemma, we take a point not in $v^\perp$ and we prove that this point is not in $\text{Cone}(t_0)$: Let $y \notin v^\perp$. We have also, thanks to the isometric action: $\langle t_0, v\rangle = 0$. We make a Taylor expansion of the following square distance (see also figure 3.7) at 0:

$$\|\varphi(x) - y\|^2 = \|t_0 + xv + o(x) - y\|^2 = \|t_0 - y\|^2 - 2x\langle y, v\rangle + o(x).$$

Then: $\exists x_\star \in (-a, a)$ s.t. $\|x_\star\| < a$, $x\langle y, v\rangle > 0$ and $\|\varphi(x_\star) - y\| < \|t_0 - y\|$. For some $g \in G$, $\varphi(x_\star) = g \cdot t_0$. By continuity of the norm we have:

$$\exists r > 0 \text{ s.t. } \forall z \in B(y, r) \quad \|g \cdot t_0 - z\| < \|t_0 - z\|.$$

Proposition 3.5. *Let G be a group acting isometrically on M. Let X be a random variable in M, such that $\mathbb{E}(\|X\|^2) < +\infty$ and $\mathbb{E}(X) = t_0 \neq 0$. If:*

1. *$\exists \varphi$ s.t. $\varphi : (-a, a) \to [t_0]$ is $\mathcal{C}^1$ with $\varphi(0) = t_0, \varphi'(0) = v \neq 0$.*
2. *The support of X is not included in the hyperplane $v^\perp$: $\mathbb{P}(X \notin v^\perp) > 0$.*

Then condition (3.27) is fulfilled, which leads to an inconsistency thanks to theorem 3.4.

Proof. Thanks to lemma 3.5, we have: $\mathbb{P}(X \notin \text{Cone}(t_0)) \geqslant \mathbb{P}(X \notin v^\perp) > 0$. Therefore theorem 3.4 applies. □

Proposition 3.5 was a sufficient condition on inconsistency in the case of an orbit which contains a curve. This brings us to extend this result for orbits which are manifolds:

Lemma 3.6. *Let G be a group acting isometrically on a Hilbert space M. If the orbit of a point t_0 is a manifold, then we have:*

$$Cone(t_0) \subset T_{t_0}[t_0]^\perp$$

where $T_{t_0}[t_0]$ the linear tangent space of $[t_0]$ at t_0.

Proof. Once again, let us prove that a point which is not in $T_{t_0}[t_0]^\perp$ is not in Cone(t_0): Let $y \notin T_{t_0}[t_0]^\perp$. Let us take $v \in T_{t_0}[t_0]$ such that $\langle y, v\rangle \neq 0$ and choose φ a $\mathcal{C}^1$ curve in $[t_0]$, such that $\varphi(0) = t_0$ and $\varphi'(0) = v$. Applying lemma 3.5, we get that $y \notin \text{Cone}(t_0)$. □.

Proposition 3.6. *Let G be a group acting isometrically on a Hilbert space M, X a random variable in M, with $\mathbb{E}(\|X\|^2) < +\infty$. Assume $X = t_0 + \sigma\varepsilon$, where $t_0 \neq 0$ and $\mathbb{E}(\varepsilon) = 0$, and $\mathbb{E}(\|\varepsilon\|) = 1$. We suppose that $[t_0]$ is a sub-manifold of M and write $T_{t_0}[t_0]$ the linear tangent space of $[t_0]$ at t_0. If:*

$$\mathbb{P}(X \notin T_{t_0}[t_0]^\perp) > 0, \tag{3.31}$$

which is equivalent to:

$$\mathbb{P}(\varepsilon \notin T_{t_0}[t_0]^\perp) > 0, \tag{3.32}$$

then there is an inconsistency.

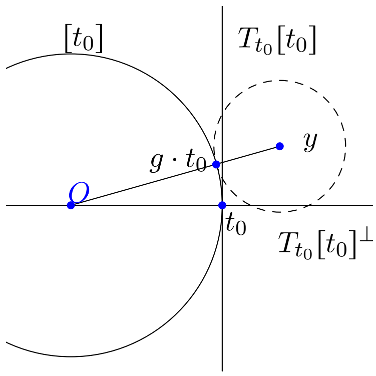


Figure 3.7: $y \notin T_{t_0}[t_0]^\perp$ therefore y is closer from $g \cdot t_0$ for some $g \in G$ than t_0 itself. In conclusion, if y is in the support of X, there is an inconsistency.

Proof. First equations (3.31) and (3.32) are equivalent, because $X = t_0 + \varepsilon$ and $t_0 \in T_{t_0}[t_0]^\perp$. Secondly, thanks to lemma 3.6 and equation (3.31) we get that $\mathbb{P}(X \notin \text{Cone}(t_0)) > 0$. Therefore, we have proved the inconsistency by applying theorem 3.4. □

Note that Condition (3.31) is very weak, because $T_{t_0}[t_0]$ is a proper linear subspace of M.

3.4.3 Lower bound of the consistency bias

Under the assumption of theorem 3.4, we have an element $\lambda(t_0)t_0$ such that $F(\lambda(t_0)t_0) < F(t_0)$ where F is the variance of $[X]$. From this element, we deduce lower bounds of the consistency bias:

Theorem 3.5. *Let δ be the unique positive solution of the following equation:*

$$\delta^2 + 2\delta\left(\|t_0\| + \mathbb{E}\|X\|\right) - \|t_0\|^2(\lambda(t_0)-1)^2 = 0. \tag{3.33}$$

Let $\delta_\star$ be the unique positive solution of the following equation:

$$\delta^2 + 2\delta\|t_0\|\left(1+\sqrt{1+\sigma^2/\|t_0\|^2}\right) - \|t_0\|^2(\lambda(t_0)-1)^2 = 0, \tag{3.34}$$

where $\sigma^2 = \mathbb{E}(\|X - t_0\|^2)$ is the variability of X. Then δ and $\delta_\star$ are two lower bounds of the consistency bias.

Proof. In order to prove this Theorem, we exhibit a ball around t_0 such that the points on this ball have a variance bigger than the variance at the point $\lambda(t_0)t_0$, where $\lambda(t_0)$ was defined in Equation (3.29): thanks to the expansion of the function f we did in (3.28) we get :

$$F(t_0) - F(\lambda(t_0)t_0) = \|t_0\|^2(\lambda(t_0)-1)^2 > 0, \tag{3.35}$$

Moreover we can show (exactly like equation (3.14)) that for all $x \in M$:

$$\begin{aligned} |F(t_0) - F(x)| &\leqslant \mathbb{E}\left(\left|\inf_{g\in G}\|g\cdot X - t_0\|^2 - \inf_{g\in G}\|g\cdot X - x\|^2\right|\right) \\ &\leqslant \|x - t_0\|\left(2\|t_0\| + \|x-t_0\| + \mathbb{E}(\|2X\|)\right). \end{aligned} \tag{3.36}$$

With Equations (3.35) and (3.36), for all $x \in B(t_0,\delta)$ we have $F(x) > F(\lambda(t_0)t_0)$. No point in that ball mapped in the quotient space is a Fréchet mean of $[X]$. So δ is a lower bound of the consistency bias. Now by using the fact that $\mathbb{E}(\|X\|) \leqslant \sqrt{\|t_0\|^2+\sigma^2}$, we get:

$$2|F(t_0) - F(x)| \leqslant 2\|x-t_0\| \times \|t_0\|\left(1+\sqrt{1+\sigma^2/\|t_0\|^2}\right) + \|x-t_0\|^2$$

This proves that $\delta_\star$ is also a lower bound of the consistency bias. □

$\delta_\star$ is smaller than δ, but the variability of X intervenes in $\delta_\star$. Therefore we propose to study the asymptotic behaviour of $\delta_\star$ when the variability tends to infinity. We have the following proposition:

Proposition 3.7. *Under the hypotheses of theorem 3.5, we write $X = t_0 + \sigma\varepsilon$, with $\mathbb{E}(\varepsilon) = 0$, and $\mathbb{E}(\|\varepsilon\|^2) = 1$ and note $\theta(t_0) = \mathbb{E}(\sup_{g\in G}\langle g\cdot\varepsilon, t_0/\|t_0\|\rangle) \in (0,1]$, we have that:*

$$\delta_\star \underset{\sigma\to+\infty}{\sim} \sigma\left(\sqrt{1+\theta(t_0)^2} - 1\right),$$

In particular, the consistency bias explodes when the variability of X tends to infinity. First let us proove the following lema which states that $\theta(t_0) \in (0,1]$.

Lemma 3.7. *Thanks to condition (3.27), we have $\theta(t_0) \in (0,1]$.*

Proof of lemma 3.7. We have $\theta(t_0) \geqslant \mathbb{E}(\langle \varepsilon, t_0/\|t_0\| \rangle = 0$. By a *reductio ad absurdum*: if $\theta(t_0) = 0$, then $\sup_{g\in G} \langle g \cdot \varepsilon, t_0 \rangle = \langle \varepsilon, t_0 \rangle$ almost surely. We have then almost surely:

$$\langle X, t_0 \rangle \leqslant \sup_{g\in G} \langle gX, t_0 \rangle \leqslant \|t_0\|^2 + \sup_{g\in G} \sigma \langle g \cdot \varepsilon, t_0 \rangle = \|t_0\|^2 + \sigma \langle \varepsilon, t_0 \rangle \leqslant \langle X, t_0 \rangle,$$

which is in contradiction with (3.27). Besides $\theta(t_0) \leqslant \mathbb{E}(\|\varepsilon\|) \leqslant \sqrt{\mathbb{E}\|\varepsilon\|^2} = 1$. □

Proof of proposition 3.7. We exhibit equivalent of the terms in equation (3.34) when $\sigma \to +\infty$:

$$2\|t_0\| \left(1 + \sqrt{1 + \sigma^2/\|t_0\|^2}\right) \sim 2\sigma. \tag{3.37}$$

Now by definition of $\lambda(t_0)$ in Equation (3.29) and the decomposition of $X = t_0 + \sigma\varepsilon$ we get:

$$\|t_0\|(\lambda(t_0) - 1) = \frac{1}{\|t_0\|}\mathbb{E}\left(\sup_{g\in G} (\langle g \cdot t_0, t_0 \rangle + \langle g \cdot \sigma\varepsilon, t_0 \rangle)\right) - \|t_0\|$$

$$\|t_0\|(\lambda(t_0) - 1) \leqslant \frac{1}{\|t_0\|}\mathbb{E}\left(\sup_{g\in G} \langle g \cdot \sigma\varepsilon, t_0 \rangle\right) = \sigma\theta(t_0) \tag{3.38}$$

$$\|t_0\|(\lambda(t_0) - 1) \geqslant \frac{1}{\|t_0\|}\mathbb{E}\left(\sup_{g\in G} \langle g \cdot \sigma\varepsilon, t_0 \rangle\right) - 2\|t_0\| = \sigma\theta(t_0) - 2\|t_0\|, \tag{3.39}$$

The lower bound and the upper bound of $\|t_0\|(\lambda(t_0) - 1)$ found in (3.38) and (3.39) are both equivalent to $\sigma\theta(t_0)$, when $\sigma \to +\infty$. Then the constant term of the quadratic Equation (3.34) has an equivalent:

$$-\|t_0\|^2(\lambda(t_0) - 1)^2 \sim -\sigma^2\theta(t_0)^2. \tag{3.40}$$

Finallye if we solve the quadratic Equation (3.34), we write $\delta_\star$ as a function of the coefficients of the quadratic equation (3.34). We use the equivalent of each of these terms thanks to equation (3.37) and (3.40), this proves proposition 3.7. □

Remark 3.3. *Thanks to inequality (3.39), if* $\frac{\|t_0\|}{\sigma} < \frac{\theta(t_0)}{2}$*, then* $\|t_0\|^2(1 - \lambda(t_0))^2 \geqslant (\sigma\theta(t_0) - 2\|t_0\|)^2$*, then we write* $\delta_\star$ *as a function of the coefficients of Equation (3.34), we obtain a lower bound of the inconsistency bias as a function of* $\|t_0\|$*,* σ *and* $\theta(t_0)$ *for* $\sigma > 2\|t_0\|/\theta(t_0)$*:*

$$\frac{\delta_\star}{\|t_0\|} \geqslant -(1 + \sqrt{1 + \sigma^2/\|t_0\|^2}) + \sqrt{(1 + \sqrt{1 + \sigma^2/\|t_0\|^2})^2 + (\sigma\theta(t_0)/\|t_0\| - 2)^2}.$$

Although the constant $\theta(t_0)$ intervenes in this lower bound, it is not an explicit term. We now explicit its behaviour depending on t_0. We remind that:

$$\theta(t_0) = \frac{1}{\|t_0\|}\mathbb{E}\left(\sup_{g\in G} \langle g \cdot \varepsilon, t_0 \rangle\right).$$

To this end, we first note that the set of fixed points under the action of G is a closed linear space, (because we can write it as an intersection of the kernel of the continuous and linear functions: $x \mapsto g \cdot x - x$ for all $g \in G$). We denote by p the orthogonal projection on the set of fixed points $\mathrm{Fix}(M)$. Then for $x \in M$, we have: $\mathrm{dist}(x, \mathrm{Fix}(M)) = \|x - p(x)\|$. Which yields:

$$\langle g \cdot \varepsilon, t_0 \rangle = \langle g \cdot \varepsilon, t_0 - p(t_0) \rangle + \langle \varepsilon, p(t_0) \rangle . \tag{3.41}$$

The right hand side of Equation (3.41) does not depend on g as $p(t_0) \in \mathrm{Fix}(M)$. Then:

$$\|t_0\|\theta(t_0) = \mathbb{E}\left(\sup_{g\in G} \langle g\varepsilon, t_0 - p(t_0) \rangle\right) + \langle \mathbb{E}(\varepsilon), p(t_0) \rangle .$$

Applying the Cauchy-Schwarz inequality and using $\mathbb{E}(\varepsilon) = 0$, we can conclude that:

$$\theta(t_0) \leqslant \frac{1}{\|t_0\|}\mathrm{dist}(t_0, \mathrm{Fix}(M))\mathbb{E}(\|\varepsilon\|) = \mathrm{dist}(t_0/\|t_0\|, \mathrm{Fix}(M))\mathbb{E}(\|\varepsilon\|). \tag{3.42}$$

This leads to the following comment: our lower bound of the consistency bias is smaller when our normalized template $t_0/\|t_0\|$ is closer to the set of fixed points.

3.4.4 Upper bound of the consistency bias

In this section, we find an upper bound of the consistency bias. More precisely we have the following Theorem:

Proposition 3.8. *Let X be a random variable in M, such that $X = t_0 + \sigma\varepsilon$ where $\sigma > 0$, $\mathbb{E}(\varepsilon) = 0$ and $\mathbb{E}(||\varepsilon||^2) = 1$. We suppose that $[m_\star]$ is a Fréchet mean of $[X]$. Then we have the following upper bound of the quotient distance between the orbit of the template t_0 and the Fréchet mean of $[X]$:*

$$d_Q([m_\star], [t_0]) \leqslant \sigma\theta(m_\star) + \sqrt{\sigma^2\theta(m_\star)^2 + 2\|t_0\|\sigma\theta(m_\star)}, \tag{3.43}$$

It is also possible to improve this inequality:

$$d_Q([m_\star], [t_0]) \leqslant \sigma\theta(m_\star - m_0) + \sqrt{\sigma^2\theta(m_\star - m_0)^2 + 2 dist(t_0, Fix(M))\sigma\theta(m_\star - m_0)}, \tag{3.44}$$

where we have noted $\theta(m) = \mathbb{E}(\sup_{g\in G} \langle g \cdot \varepsilon, m/\|m\| \rangle) \in [0, 1]$ if $m \neq 0$ and $\theta(0) = 0$, and m_0 the orthogonal projection of t_0 on $Fix(M)$.

Note that we made no hypothesis on the template in this proposition. We deduce from Equation (3.44) that $d_Q([m_\star], [t_0]) \leqslant \sigma + \sqrt{\sigma^2 + 2\sigma \mathrm{dist}(t_0, \mathrm{Fix}(M))}$ is a $O(\sigma)$ when $\sigma \to \infty$, but a $O(\sqrt{\sigma})$ when $\sigma \to 0$, in particular the consistency bias can be neglected when σ is small.

In order to prove equation (3.44), we give a useful lemma:

Lemma 3.8. *Let m_0 be a fixed point under the action of G, then the translation map*

$$T : [x] \mapsto [x - m_0], \tag{3.45}$$

is well defined on the quotient. Moreover, this map is a congruent map of Q:

$$\forall (a,b) \in Q^2 \quad d_Q(T(a),T(b)) = d_Q(a,b).$$

Proof of lemma 3.8. First, we need to prove that this map is well defined: let us assume that x and y are in the same orbit, we need to prove that $[x-m_0] = [y-m_0]$. We know that $y = g\cdot x$ for some $g \in G$, therefore:

$$y - m_0 = g\cdot x - gm_0 = g(x-m_0),$$

indeed m_0 is a fixed point, and G acts linearly on M. Therefore we have $[y-m_0] = [x-m_0]$.

Secondly, we need to prove that for every x and $y \in M$, we have $d_Q([x],[y]) = d_Q([x-m_0],[y-m_0])$. This is the consequence of:

$$\forall g \in G \quad (x-m_0) - g\cdot(y-m_0) = x - m_0 + m_0 - g\cdot y = x - g\cdot y,$$

once again, this equation is true because g acts linearly, and because m_0 is a fixed point. By taking the infimum over $g \in G$, we prove that the map is congruent. □

Proof of proposition 3.8. First we have:

$$F(m_\star) \leqslant F(t_0) = \mathbb{E}(\inf_{g\in G} ||t_0 - g(t_0+\sigma\varepsilon)||^2) \leqslant \mathbb{E}(||\sigma\varepsilon||^2) = \sigma^2. \tag{3.46}$$

Secondly, we have for all $m \in M$, (in particular for $m_\star$):

$$\begin{aligned} F(m) &= \mathbb{E}(\inf_{g\in G}(\|m - g\cdot t_0\|^2 + \sigma^2\|\varepsilon\|^2 - 2\langle g\cdot\sigma\varepsilon, m - g\cdot t_0\rangle)) \\ &\geqslant d_Q([m],[t_0])^2 + \sigma^2 - 2\mathbb{E}(\sup_{g\in G}\ \langle \sigma\varepsilon, g\cdot m\rangle). \end{aligned} \tag{3.47}$$

With Inequalities (3.46) and (3.47) one gets:

$$d_Q([m_\star],[t_0])^2 \leqslant 2\mathbb{E}(\sup_{g\in G}\ \langle\sigma\varepsilon, g\cdot m_\star\rangle) = 2\sigma\theta(m_\star)||m_\star||, \tag{3.48}$$

note that at this point, if $m_\star = 0$ then $\mathbb{E}(\sup_{g\in G}\ \langle\sigma\varepsilon, g\cdot m_\star\rangle) = 0$ and $\theta(m_\star) = 0$ although equation (3.48) is still true even if $m_\star = 0$. Moreover with the triangular inequality applied at $[m_\star]$, $[0]$ and $[t_0]$, one gets: $\|m_\star\| \leqslant \|t_0\| + d_Q([m_\star],[t_0])$ and then:

$$d_Q([m_\star],[t_0])^2 \leqslant 2\sigma\theta(m_\star)(d_Q([m_\star],[t_0]) + \|t_0\|). \tag{3.49}$$

We can solve inequality (3.49) and we get:

$$d_Q([m_\star],[t_0]) \leqslant \sigma\theta(m_\star) + \sqrt{\sigma^2\theta(m_\star)^2 + 2\|t_0\|\sigma\theta(m_\star)}, \tag{3.50}$$

We note by F_X instead of F the variance in the quotient space of $[X]$, and we want to apply inequality (3.43) to $X - m_0$. As m_0 is a fixed point, we have thanks to lemma 3.8:

$$F_X(m) = \mathbb{E}\left(d_Q([X],[m])^2\right) = \mathbb{E}(\left(d_Q([X-m_0],[m-m0]^2\right)) = F_{X-m_0}(m-m_0)$$

Then:

$$m_\star \in \text{argmin } F_X \iff m_\star - m_0 \in \text{argmin } F_{X-m_0}.$$

We apply Equation (3.43) to $X - m_0$, with $\mathbb{E}(X - m_0) = t_0 - m_0$ and $[m_\star - m_0]$ a Fréchet mean of $[X - m_0]$. We get:

$$d_Q([m_\star - m_0], [t_0 - m_0]) \leqslant \sigma\theta(m_\star - m_0) + \sqrt{\sigma^2\theta(m_\star - m_0)^2 + 2\|t_0 - m_0\|\sigma\theta(m_\star - m_0)}.$$

Moreover $d_Q([m_\star], [t_0]) = d_Q([m_\star - m_0], [t_0 - m_0])$ (see lemma 3.8), which concludes the proof. □

3.4.5 Empirical Fréchet mean

In practice, we never compute the Fréchet mean in quotient space, only the empirical Fréchet mean in quotient space when the size of a sample is supposed to be large enough. If the empirical Fréchet in the quotient space means converges to the Fréchet mean in the quotient space then we can not use these empirical Fréchet mean in order to estimate the template. In [Bhattacharya 2008], it has been proved that the empirical Fréchet mean converges to the Fréchet mean with a $\frac{1}{\sqrt{n}}$ convergence speed, however the law of the random variable is supposed to be included in a ball whose radius depends on the geometry on the manifold. Here we are not in a manifold, indeed the quotient space contains singularities, moreover we do not suppose that the law is necessarily bounded. However in [Ziezold 1977] the empirical Fréchet means is proved to converge to the Fréchet means but no convergence rate is provided.

We propose now to prove that the quotient distance between the template and the empirical Fréchet mean in quotient space have an lower bound which is the asymptotic of the one lower bound of the consistency bias found in (3.33). Take $X, X_1, \dots, X_n$ independent and identically distributed (with $t_0 = \mathbb{E}(X)$ not a fixed point). We define the empirical variance of $[X]$ by:

$$m \in M \mapsto F_n(m) = \frac{1}{n}\sum_{i=1}^n d_Q([m], [X_i])^2 = \frac{1}{n}\sum_{i=1}^n \inf_{g \in G} \|m - g \cdot X_i\|^2,$$

and we say that $[m_{n\star}]$ is a empirical Fréchet mean of $[X]$ if $m_{n\star}$ is a global minimiser of F_n.

Proposition 3.9. *Let $X, X_1, \dots, X_n$ independent and identically distributed random variables, with $t_0 = \mathbb{E}(X)$. Let be $[m_{n\star}]$ be an empirical Fréchet mean of $[X]$. Then δ_n is a lower bound of the quotient distance between the orbit of the template and $[m_{n\star}]$, where δ_n is the unique positive solution of:*

$$\delta^2 + 2\left(||t_0|| + \frac{1}{n}\sum_{i=1}^n \|X_i\|\right)\delta - \|t_0\|^2(\lambda_n(t_0) - 1)^2 = 0.$$

$\lambda_n(t_0)$ is defined like $\lambda(t_0)$ in section 3.4.1 by:

$$\lambda_n(t_0) = \frac{\frac{1}{n}\sum_{i=1}^n \sup_{g \in G} \langle g \cdot X_i, t_0\rangle}{\|t_0\|^2}.$$

We have that $\delta_n \to \delta$ by the law of large numbers.

The proof is a direct application of theorem 3.5, but applied to the empirical law of X given by the realization of $X_1, \dots, X_n$.

3.4.6 Examples

In this Subsection, we discuss, in some examples, the application of theorem 3.4 and see the behaviour of the constant $\theta(t_0)$. This constant intervened in lower bound of the consistency bias.

3.4.6.1 Action of horizontal translation on $L^2(\mathbb{R}/\mathbb{Z})$

We take an orbit $O = [f_0]$, where $f_0 \in \mathcal{C}^2(\mathbb{R}/\mathbb{Z})$, non constant. We show easily that O is a manifold of dimension 1 and the tangent space at f_0 is[2] $\mathbb{R}f_0'$. Therefore a sufficient condition on X such that $\mathbb{E}(X) = f_0$ to have an inconsistency is: $\mathbb{P}(X \notin f_0'^{\perp}) > 0$ according to proposition 3.6. Now if we denote by $\mathbb{1}$ the constant function on $\mathbb{R}/\mathbb{Z}$ equal to 1. We have in this setting: that the set of fixed points under the action of G is the set of constant functions: $\mathrm{Fix}(M) = \mathbb{R}\mathbb{1}$ and:

$$\mathrm{dist}(f_0, \mathrm{Fix}(M)) = \|f_0 - \langle f_0, \mathbb{1}\rangle \mathbb{1}\| = \sqrt{\int_0^1 \left(f_0(t) - \int_0^1 f_0(s)ds\right)^2 dt}.$$

This distance to the fixed points is used in the upper bound of the constant $\theta(t_0)$ in Equation (3.42). Note that if f_0 is not differentiable, then $[f_0]$ is not necessarily a manifold, and (3.6) does not apply. However proposition 3.4 does: if f_0 is not a constant function, then $[f_0]\backslash\{f_0\}$ is dense in $[f_0]$. Therefore as soon as the support of X contains a ball around f_0, there is an inconsistency.

3.4.6.2 Action of discrete horizontal translation on $\mathbb{R}^{\mathbb{Z}/\mathbb{N}\mathbb{Z}}$

We come back on example 3.1, with $D = 1$ (discretised signals). For some signal t_0, $\theta(t_0)$ previously defined is:

$$\theta(t_0) = \frac{1}{\|t_0\|}\mathbb{E}\left(\max_{\tau \in \mathbb{Z}/\mathbb{N}\mathbb{Z}} \langle \varepsilon, \tau \cdot t_0\rangle\right).$$

Therefore if we have a sample of size I of ε *iid*, then:

$$\theta(t_0) = \frac{1}{\|t_0\|}\lim_{I \to +\infty} \frac{1}{I}\sum_{i=1}^{I} \max_{\tau_i \in \mathbb{Z}/N\mathbb{Z}} \langle \varepsilon_i, \tau_i \cdot t_0\rangle,$$

[2] Indeed $\varphi : \begin{array}{ccc}]-\frac{1}{2},\frac{1}{2}[& \to & O \\ t & \mapsto & f_0(.-t) \end{array}$ is a local parametrisation of O: $f_0 = \varphi(0)$, and we check that: $\lim_{x\to 0} \|\varphi(x)-\varphi(0)-xf_0'\|_{L^2} = 0$ with Taylor-Lagrange inequality at the order 2. As a conclusion φ is differentiable at 0, and it is an immersion (since $f_0' \neq 0$), and $D_0\varphi : x \mapsto xf_0'$, then O is a manifold of dimension 1 and the tangent space of O at f_0 is: $T_{f_0}O = D_0\varphi(\mathbb{R}) = \mathbb{R}f_0'$.

By an exhaustive research, we can find the τ_i's which maximise the dot product, then with this sample and t_0 we can approximate $\theta(t_0)$. We have done this approximation for several signals t_0 on figure 3.8. According the previous results, the bigger $\theta(t_0)$ is, the more important the lower bound of the consistency bias is. We remark that the $\theta(t_0)$ estimated is small, $\theta(t_0) \ll 1$ for different signals.

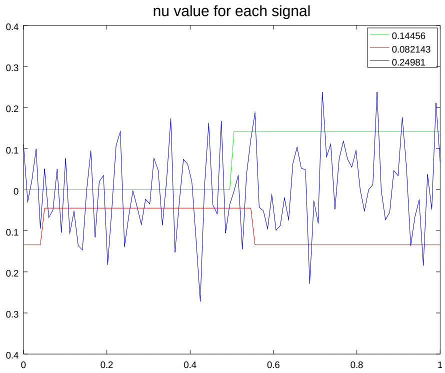


Figure 3.8: Different signals and their $\theta(t_0)$ approximated with a sample of size 10^3 in $\mathbb{R}^{\mathbb{Z}/100\mathbb{Z}}$. ε is here a Gaussian noise in $\mathbb{R}^{\mathbb{Z}/100\mathbb{Z}}$, such that $\mathbb{E}(\varepsilon) = 0$ and $\mathbb{E}(\|\varepsilon\|^2) = 1$. For instance the blue signal is a signal defined randomly, and when we approximate the $\theta(t_0)$ which corresponds to that t_0 we find $\simeq 0.25$.

3.4.6.3 Action of rotations on $\mathbb{R}^n$

Now we consider the action of rotations on $\mathbb{R}^n$ with a Gaussian noise. Take $X \sim \mathcal{N}(t_0, s^2 Id_n)$ then the variability of X is ns^2, then X has a decomposition: $X = t_0 + \sqrt{n}s\varepsilon$ with $\mathbb{E}(\varepsilon) = 0$ and $\mathbb{E}(\|\varepsilon\|^2) = 1$. According to proposition 3.7 we have by noting $\delta_\star$ the lower bound of the consistency bias when $s \to \infty$:

$$\frac{\delta_\star}{s} \to \sqrt{n}(-1 + \sqrt{1 + \theta(t_0)^2}).$$

Now $\theta(t_0) = \mathbb{E}(\sup_{g \in G} \langle g \cdot \varepsilon, t_0 \rangle)/\|t_0\| = \mathbb{E}(\|\varepsilon\|) \to 1$ when n tends to infinity (expected value of the Chi distribution) we have that for n large enough:

$$\lim_{s \to \infty} \frac{\delta_\star}{s} \simeq \sqrt{n}(\sqrt{2} - 1).$$

We compare this result with the exact computation of the consistency bias (noted here CB) made by Miolane et al. [Miolane 2017], which writes with our current notations:

$$\lim_{s \to \infty} \frac{\text{CB}}{s} = \sqrt{2}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}.$$

Using a standard Taylor expansion on the Gamma function, we have that for n large enough:

$$\lim_{s \to \infty} \frac{\text{CB}}{s} \simeq \sqrt{n}.$$

As a conclusion, when the dimension of the space is large enough our lower bound and the exact computation of the bias have the same asymptotic behaviour. It differs only by the constant $\sqrt{2}-1 \simeq 0.4$ in our lower bound, 1 in the work of Miolane et al. [Miolane 2015].

3.4.7 Differentiability of the variance in the quotient space

In section 3.3.1, we study the consistency of the template estimation with the Fréchet mean in quotient space, by assuming that the group is finite. In order to prove the inconsistency we establish that the variance is differentiable at some points. Then, one natural question we ask is the following: is the variance always differentiable?

In this section, we show that the variance in the quotient space is not differentiable at 0 when the ambient space is an Hilbert spaces and for isometric action: $F(m) = \mathbb{E}(\inf\limits_{g\in G}\|Y - g\cdot m\|^2)$ is not differentiable at 0.

This question matters, since one may want to compute the Fréchet mean in quotient space with a gradient descent, but this requires, at least, that the variance is differentiable. For instance what if the Fréchet mean of $[Y]$ is $[m_\star]$ and that the variance F is not differentiable at $m_\star$?

First, we would to insist on this point: it is a well known fact that, for all $y \in M$, $m \mapsto \|m-y\|^2$ is differentiable with a gradient equal to $m-y$. Then

$$m \mapsto d_Q([m],[y]) = \inf_{g\in G}\|m - g\cdot y\|^2$$

is defined as the infimum of differentiable functions. There is no guarantee that the resulting function is differentiable. $x \mapsto \|x\| = -\min(x,-x)$ is a toy example of the infimum of differentiable functions which is not differentiable.

One could think that we could deal with sub-differentiability, however the variance is probably not convex nor concave.

If the group is compact then, we have an element which reaches this infimum, $d_Q([m],[y]) = \|m - g_\star y\|^2$ for some $g_\star \in G$. In this case, one could think that $m \mapsto d_Q([m],[y])$ is differentiable with a gradient equal to $2(m - g_\star y)$. However this is more complicated than that. Indeed $g_\star$ depends on m and y. We should call this element $g(m,y)$ rather than $g_\star$. One gets:

$$d_Q([m],[y])^2 = \|m - g(m,y)\cdot y\|^2.$$

Now, $g(m,y)\cdot y$ depends on m, and it is more difficult to show the differentiability or to compute the gradient, since we have the existence of the element $g(m,y)$ but we have not an explicit formula of this element. If one is not able to prove the differentiability of the squared distance, then it will be harder to prove the differentiablity of the variance (defined as the expectation of the squared distance). In the proof of theorem 3.1, we proved that for finite group $g(m,y)$ was locally constant, that is why we could be able to differentiate the squared quotient distance and the variance.

Example 3.4. *Example of the action of rotation Taking M an Euclidean space, and G the group of rotations of M. Then the quotient distance is given by: $d_Q([a],[b]) = |\|a\| - \|b\||$. Therefore, the variance in the quotient space is equal to:*

$$F(m) = \mathbb{E}((\|Y\| - \|m\|)^2),$$

by expansion one gets:

$$F(m) = \mathbb{E}(\|Y\|^2) - 2\|m\|\mathbb{E}(\|Y\|) + \|m\|^2,$$

now $m \mapsto \|m\|^2$ is differentiable everywhere, however $m \mapsto \|m\|$ is not differentiable at 0. We can conclude that F is not differentiable at 0.

This example can be generalized to any isometric group action:

Proposition 3.10. *Let G acting isometrically on M a Hilbert space. We do not assume that G is a finite group. Let us take Y a random variable in M such that $\mathbb{E}(\|Y\|^2) < +\infty$. Let us assume that it exists a point $p \in M$ such that*

$$\mathbb{P}(d_Q([p],[Y]) < \|p - Y\|) > 0,$$

then the variance $m \mapsto F(m) = \mathbb{E}(d_Q^2([m],[Y]))$ is not differentiable at 0.

Note that it is easy to fulfill proposition 3.10, for instance with $p = t_0$, as we have seen it in theorem 3.1.

Remark 3.4. *We can prove at the same time that the squared distance is also not differentiable: $m \mapsto d_Q^2([m],[y])$ is not differentiable at 0, if y is not a fixed point. Indeed, let us take Y a random variable equal to y, then $F(m) = \mathbb{E}(d^2([m],[Y])) = d^2([m],[y])$ is not differentiable at 0 according to proposition 3.10.*

Proof of proposition 3.10. We first expand the variance:

$$F(m) = \|m\|^2 + \mathbb{E}\left(\|Y\|^2) - 2\mathbb{E}(\sup_{g\in G} \langle m, g\cdot Y\rangle\right),$$

then if this function is differentiable so is the function $f : m \mapsto \mathbb{E}(\sup_{g\in G} \langle m, g\cdot Y\rangle)$, now by linearity of the action we remark that we have $f(\lambda m) = \lambda f(m)$ for all $m \in M$ and $\lambda \geqslant 0$, then if f was differentiable, f would be a linear function. Let us assume that f is linear. In this case we have for $m, m' \in M$:

$$\begin{aligned} f(m+m') &= \mathbb{E}(\overbrace{\sup_{g\in G} \left\langle m+m', g\cdot Y\right\rangle}^{A}) \\ &\leqslant \underbrace{\mathbb{E}(\sup_{g\in G} \langle m, g\cdot Y\rangle + \sup_{g\in G}\left\langle m', g\cdot Y\right\rangle)}_{B} \\ &\leqslant f(m) + f(m') \\ &\leqslant f(m+m'), \end{aligned}$$

where the inequality comes from the fact that the supremum of the sum of two terms is smaller than the sum of the two supremum. Then we have two random variables A, B such that $A \leqslant B$ and $\mathbb{E}(A) = \mathbb{E}(B)$, when we can conclude that $A = B$ almost surely. It comes that:

$$\sup_{g\in G} \left\langle m + m', g \cdot Y \right\rangle) = \sup_{g\in G} \left\langle m, g \cdot Y \right\rangle) + \sup_{g\in G} \left\langle m', g \cdot Y \right\rangle) \quad Y - \text{a.s.}.$$

Now let us choose $m' = -m$ we have:

$$\sup_{g\in G} \left\langle m, g \cdot Y \right\rangle + \sup_{g\in G} -\left\langle m, g \cdot Y \right\rangle = 0 \quad Y - \text{almost surely.}$$

If we reformulate we have:

$$\sup_{g\in G} \left\langle m, g \cdot Y \right\rangle = \inf_{g\in G} \left\langle m, g \cdot Y \right\rangle \quad Y - \text{almost surely.}$$

Then for any $m \in M$ we have Y-almost surely:

$$\forall g \in G \left\langle m, Y \right\rangle = \left\langle g \cdot m, Y \right\rangle.$$

Then we have Y-almost surely $\|m - Y\| = d_Q([m], [Y])$. Taking $m = p$ shows the contradiction with our hypothesis. □

In order to prove the remark remark 3.4, we can take Y a random variable which is constantly equal to y a non fixed point, as y is not a fixed point it exists $g \in G$ such that $g \cdot y \neq y$. We can take $m_0 = g \cdot y$, this m_0 fulfills proposition 3.10. □

Remark 3.5. 0 *is one fixed point among all the other, thanks to lemma 3.8, we can also conclude that the variance is not differentiable at any other fixed points.*

Proof. Let m_0 be a fixed point. We can consider two random variables X and $X - m_0$, if we note F_X the variance of X, thanks to lemma 3.8, we have:

$$F_X(m) = \mathbb{E}(d_Q^2([m], [X])) = \mathbb{E}(d_Q^2([m - m_0], [X - m_0])) = F_{X-m_0}(m - m_0)$$

Now we know that F_{X-m_0} is not differentiable at 0 (proposition 3.10), therefore $m \mapsto F_{X-m_0}(m - m_0)$ is not differentiable at m_0. We conclude that F_X is not differentiable at m_0. □

One remaining question is: can we study the differentiability of the square distance for non fixed points?

3.5 Fréchet means ambient and quotient spaces are not consistent when the template is a fixed point

In this section, we do not assume that the ambient space M is a vector space, but rather a manifold. We need then to rewrite the generative model likewise: let $t_0 \in M$, and X any random variable of M such as t_0 is a Fréchet mean of X. Then $Y = \Phi \cdot X$ is the observed variable where Φ is a random variable whose value are in G. In this section we make the assumption that the template t_0 is a fixed point under the action of G.

3.5.1 Result

Let X be a random variable on M and define the variance of X as:

$$E(m) = \mathbb{E}(d_M(m, X)^2).$$

We say that t_0 is a Fréchet mean of X if t_0 is a global minimiser of the variance E. We prove the following result:

Theorem 3.6. *Assume that M is a complete finite dimensional Riemannian manifold and that d_M is the geodesic distance on M. We suppose that d_M is invariant under the group action. Let X be a random variable on M, with $\mathbb{E}(d(x, X)^2) < +\infty$ for some $x \in M$. We assume that t_0 is a fixed point and a Fréchet mean of X and that $\mathbb{P}(X \in C(t_0)) = 0$ where $C(t_0)$ is the cut locus of t_0. Suppose that there exists a point in the support of X which is not a fixed point nor in the cut locus of t_0. Then $[t_0]$ is not a Fréchet mean of $[X]$.*

The previous result is finite dimensional and does not cover interesting infinite dimensional setting concerning curves for instance. However, a simple extension to the previous result can be stated when M is a Hilbert vector space since then the space is flat and some technical problems like the presence of cut locus point do not occur.

Theorem 3.7. *Assume that M is a Hilbert space and that d_M is given by the Hilbert norm on M. We suppose that d_M is invariant under the group action. Let X be a random variable on M, with $\mathbb{E}(\|X\|^2) < +\infty$. We assume that $t_0 = \mathbb{E}(X)$. Suppose that there exists a point in the support of the law of X that is not a fixed point for the action of G. Then $[t_0]$ is not a Fréchet mean of $[X]$.*

Note that the reciprocal is true: if all the points in the support of the law of X are fixed points, then almost surely, for all $m \in M$ and for all $g \in G$ we have:

$$d_M(X, m) = d_M(g \cdot X, m) = d_Q([X], [m]).$$

Up to the projection on the quotient, we have that the variance of X is equal to the variance of $[X]$ in M/G, therefore $[t_0]$ is a Fréchet mean of $[X]$ if and only if t_0 is a Fréchet mean of X. There is no inconsistency in that case.

Example 3.5. *Theorem 3.7 covers the interesting case of the Fisher Rao metric on functions:*

$$\mathcal{F} = \{f : [0, 1] \to \mathbb{R} \quad | \quad f \text{ is absolutely continuous}\}.$$

Then considering for G the group of smooth diffeomorphisms φ on $[0, 1]$ such that $\varphi(0) = 0$ and $\varphi(1) = 1$, we have a right group action $G \times \mathcal{F} \to \mathcal{F}$ given by $\varphi \cdot f = f \circ \varphi$. The Fisher Rao metric is built as a pull back metric of the $L^2([0, 1], \mathbb{R})$ space through the map $Q : \mathcal{F} \to L^2$ given by: $Q(f) = \dot{f}/\sqrt{|\dot{f}|}$. This square root trick is often used, see for instance [Kurtek 2011b]. Note that in this case, Q is a bijective mapping with *inverse given by* $q \mapsto f$ *with* $f(t) = \int_0^t q(s)|q(s)|ds$. *We can define a group action on* $M = L^2$ *as:* $\varphi \cdot q = q \circ \varphi\sqrt{\dot{\varphi}}$, *for which one can check easily by a change of variable that:*

$$\|\varphi \cdot q - \varphi \cdot q'\|^2 = \|q \circ \varphi\sqrt{\dot{\varphi}} - q' \circ \varphi\sqrt{\dot{\varphi}}\|^2 = \|q - q'\|^2.$$

So up to the mapping Q, *the Fisher Rao metric on curve corresponds to the situation* M *where theorem 3.7 applies. Note that in this case the set of fixed points under the action of* G *corresponds in the space* $\mathcal{F}$ *to constant functions.*

We can also provide an computation of the consistency bias in this setting:

Proposition 3.11. *Under the assumptions of theorem 3.7, we write* $X = t_0 + \sigma\varepsilon$ *where* t_0 *is a fixed point,* $\sigma > 0$, $\mathbb{E}(\varepsilon) = 0$ *and* $\mathbb{E}(\|\varepsilon\|^2) = 1$. *Furthermore, we assume that the group* G *acts isometrically on* M. *If there is a Fréchet mean of* $[X]$, *then the consistency bias is linear with respect to* σ *and it is equal to:*

$$\sigma \sup_{\|v\|=1} \mathbb{E}(\sup_{g \in G} \langle v, g \cdot \varepsilon \rangle).$$

Proof. For $\lambda > 0$ and $\|v\| = 1$, we compute the variance F in the quotient space of $[X]$ at the point $t_0 + \lambda v$. Since t_0 is a fixed point we get:

$$F(t_0+\lambda v) = \mathbb{E}(\inf_{g \in G}\|t_0+\lambda v - gX\|^2) = \mathbb{E}(\|X\|^2) - \|t_0\|^2 - 2\lambda\mathbb{E}(\sup_{g \in G} \langle v, g(X - t_0)\rangle) + \lambda^2.$$

Then we minimise F with respect to λ, and after we minimise with respect to v (with $\|v\| = 1$). Which concludes. ▫

3.5.2 Proofs of these theorems

3.5.2.1 Proof of theorem 3.6

We start with the following simple result, which aims to differentiate the variance of X. This classical result (see [Pennec 2006] for instance) is proved again here to be the more self-contained as possible:

Lemma 3.9. *Let* X *a random variable on* M *such that* $\mathbb{E}(d(x, X)^2) < +\infty$ *for some* $x \in M$. *Then the variance* $m \mapsto E(m) = \mathbb{E}(d_M(m, X)^2)$ *is a continuous function which is differentiable at any point* $m \in M$ *such that* $\mathbb{P}(X \in C(m)) = 0$ *where* $C(m)$ *is the cut locus of* m. *Moreover at such point one has:*

$$\nabla E(m) = -2\mathbb{E}(Log_m(X)),$$

where $Log_m : M \backslash C(m) \to T_m M$ *is defined for any* $x \in M \backslash C(m)$ *as the unique* $u \in T_m M$ *such that* $Exp_m(u) = x$ *and* $\|u\|_m = d_M(x, m)$.

Proof of lemma 3.9. By triangle inequality it is easy to show that E is finite and continuous everywhere. Moreover, it is a well known fact that $x \mapsto d_M(x, z)^2$ is differentiable at any $m \in M\backslash C(z)$ (i.e. $z \notin C(m)$) with derivative $-2\text{Log}_m(z)$. Now since:

$$\begin{aligned}|d_M(x,z)^2 - d_M(y,z)^2| &= |d_M(x,z) - d_M(y,z)\|d_M(x,z) + d_M(y,z)| \\ &\leqslant d_M(x,y)(2d_M(x,z) + d_M(y,x)),\end{aligned}$$

we get in a local chart $\phi : U \to V \subset \mathbb{R}^n$ at $t = \phi(m)$ we have locally around t that:

$$h \mapsto d_M(\phi^{-1}(t), \phi^{-1}(t+h)),$$

is smooth and $|d_M(\phi^{-1}(t), \phi^{-1}(t+h))| \leqslant C|h|$ for a $C > 0$. Hence for sufficiently small h, $|d_M(\phi^{-1}(t), z)^2 - d_M(\phi^{-1}(t+h), z)^2| \leqslant C|h|(2d_M(m,z)+1)$. We get the result from dominated convergence Lebesgue theorem with $\mathbb{E}(d_M(m,X)) \leqslant \mathbb{E}(d_M(m,X)^2 + 1) < +\infty$. □

We are now ready to prove theorem 3.6.

Proof. (of theorem 3.6) Let m_0 be a point in the support of M which is not a fixed point and not in the cut locus of t_0. Then there exists $g_0 \in G$ such that $m_1 = g_0 m_0 \neq m_0$. Note that as the distance is equivariant under the action of G, we have that $m_1 = g_0 \cdot m_0 \notin C(g_0 \cdot t_0) = C(t_0)$ (t_0 is a fixed point under the action of G). Let $v_0 = \text{Log}_{t_0}(m_0)$ and $v_1 = \text{Log}_{t_0}(m_1)$. We have $v_0 \neq v_1$ and since $C(t_0)$ is closed and the Log_{t_0} is a continuous application on $M\backslash C(t_0)$ we have:

$$\lim_{\eta \to 0} \frac{1}{\mathbb{P}(X \in B(m_0,\eta))} \mathbb{E}(\mathbb{1}_{X \in B(m_0,\eta)} \text{Log}_{t_0}(X)) = v_0.$$

(we use here the fact that since m_0 is in the support of the law of X, $\mathbb{P}(X \in B(m_0,\eta)) > 0$ for any $\eta > 0$ so that the denominator does not vanish and the fact that since M is a complete manifold, it is a locally compact space (the closed balls are compacts) and Log_{t_0} is locally bounded). Similarly:

$$\lim_{\eta \to 0} \frac{1}{\mathbb{P}(X \in B(m_0,\eta))} \mathbb{E}(\mathbb{1}_{X \in B(m_0,\eta)} \text{Log}_{t_0}(g_0 \cdot X)) = v_1.$$

Thus for sufficiently small $\eta > 0$ we have (since $v_0 \neq v_1$):

$$\mathbb{E}(\text{Log}_{t_0}(X)\mathbb{1}_{X \in B(m_0,\eta)}) \neq \mathbb{E}(\text{Log}_{t_0}(g_0 \cdot X)\mathbb{1}_{X \in B(m_0,\eta)}). \tag{3.51}$$

By using using a *reductio ad absurdum*, we suppose that $[t_0]$ is a Fréchet mean of $[X]$ and we want to find a contradiction with (3.51). In order to do that we introduce simple functions as the function $x \mapsto \mathbb{1}_{x \in B(m_0,\eta)}$ which intervenes in Equation (3.51). Let $s : M \to G$ be a simple function (i.e. a measurable function with finite number of values in G). Then $x \mapsto h(x) = s(x) \cdot x$ is a measurable function[3]. Now, let

[3]Indeed if: $s = \sum_{i=1}^{n} g_i \mathbb{1}_{A_i}$ where $(A_i)_{1 \leqslant i \leqslant n}$ is a partition of M (such that the sum is always defined). Then for any Borel set $B \subset M$ we have: $h^{-1}(B) = \bigcup_{i=1}^{n} g_i^{-1}(B) \cap A_i$ is a measurable set since $x \mapsto g_i x$ is a measurable function.

$E_s(x) = \mathbb{E}(d(x, s(X) \cdot X)^2)$ be the variance of the variable $s(X) \cdot X$. Note that (and this is the main point):

$$\forall g \in G \qquad d_M(t_0, x) = d_M(g \cdot t_0, g \cdot x) = d_M(t_0, g \cdot x) = d_Q([t_0], [x]),$$

we have: $E_s(t_0) = E(t_0)$. Assume now that $[t_0]$ a Fréchet mean for $[X]$ on the quotient space and let us show that E_s has a global minimum at t_0. Indeed for any m, we have:

$$E_s(m) = \mathbb{E}(d_M(m, s(X) \cdot X)^2) \geqslant \mathbb{E}(d_Q([m], [X])^2) \geqslant \mathbb{E}(d_Q([t_0], [X])^2) = E_s(t_0).$$

Now, we want to apply lemma 3.9 to the random variables $s(X) \cdot X$ and X at the point t_0. Since we assume that $X \notin C(t_0)$ almost surely and $X \notin C(t_0)$ implies $s(X) \cdot X \notin C(t_0)$ we get $\mathbb{P}(s(X) \cdot X \in C(t_0)) = 0$ and the lemma 3.9 applies. As t_0 is a minimum, we already know that the differential of E_s (respectively E) at t_0 should be zero. We get:

$$\mathbb{E}(\mathrm{Log}_{t_0}(X)) = \mathbb{E}(\mathrm{Log}_{t_0}(s(X) \cdot X)) = 0. \tag{3.52}$$

Now we apply Equation (3.52) to a particular simple function defined by $s(x) = g_0 \mathbb{1}_{x \in B(m_0,\eta)} + e_G \mathbb{1}_{x \notin B(m_0,\eta)}$. We split the two expected values in (3.52) into two parts:

$$\mathbb{E}(\mathrm{Log}_{t_0}(X) \mathbb{1}_{X \in B(m_0,\eta)}) + \mathbb{E}(\mathrm{Log}_{t_0}(X) \mathbb{1}_{X \notin B(m_0,\eta)}) = 0, \tag{3.53}$$

$$\mathbb{E}(\mathrm{Log}_{t_0}(g_0 \cdot X) \mathbb{1}_{X \in B(m_0,\eta)}) + \mathbb{E}(\mathrm{Log}_{t_0}(X) \mathbb{1}_{X \notin B(m_0,\eta)}) = 0. \tag{3.54}$$

By substrating (3.53) from (3.54), one gets:

$$\mathbb{E}(\mathrm{Log}_{t_0}(X) \mathbb{1}_{X \in B(m_0,\eta)}) = \mathbb{E}(\mathrm{Log}_{t_0}(g_0 \cdot X) \mathbb{1}_{X \in B(m_0,\eta)}),$$

which is a contradiction with (3.51). Which concludes. □

3.5.2.2 Proof of theorem 3.7

Proof. The extension to theorem 3.7 is quite straightforward. In this setting many things are now explicit since $d(x, y) = \|x - y\|$, $\nabla_x d(x, y)^2 = 2(x - y)$, $\mathrm{Log}_x(y) = y - x$ and the cut locus is always empty. It is then sufficient to go along the previous proof and to change the quantity accordingly. Note that the local compactness of the space is not true in infinite dimension. However this was only used to prove that the log was locally bounded but this last result is trivial in this setting. □

3.6 Conclusion and discussion

In this chapter, we exhibit conditions which imply that the template estimation with the Fréchet mean in quotient space is inconsistent. These conditions are rather generic. As a result, without any more information, *a priori* there is inconsistency.

Table 3.1: Behaviour of the consistency bias with respect to σ^2 the variability of $X = t_0 + \sigma\varepsilon$. The constants K_i's depend on the kind of noise, on the template t_0 and on the group action.

Consistency bias noted CB	G is any group	Supplementary properties for G a finite group
Upper bound of CB	$\text{CB} \leqslant \sigma + 2\sqrt{\sigma^2 + K_1\sigma}$ (proposition 3.8)	$\text{CB} \leqslant K_2\sigma$ (theorem 3.3)
Lower bound of CB for $\sigma \to \infty$ when the template is not a fixed point	$\text{CB} \geqslant L \underset{\sigma\to\infty}{\sim} K_3\sigma$ (proposition 3.7)	
Behavior of CB for $\sigma \to 0$ when the template is not a fixed point	$\text{CB} \leqslant U \underset{\sigma\to 0}{\sim} K_4\sqrt{\sigma}$	$\text{CB} \underset{0}{=} o(\sigma^k)$, $\forall k \in \mathbb{N}$ in the section 3.3.6, can we extend this result for finite group?
CB when the template is a fixed point	$\text{CB} = \sigma \sup\limits_{\Vert v\Vert=1} \mathbb{E}(\sup_{g\in G} \langle v, g\cdot\varepsilon\rangle)$ (proposition 3.11)	

The behaviour of the consistency bias is summarized in table 3.1. Surely future works could improve these lower and upper bounds.

In a more general case: when we take an infinite-dimensional vector space quotiented by a non isometric group action, is there always an inconsistency? An important example of such action is the action of diffeomorphisms. Can we estimate the consistency bias? In this setting, one estimates the template (or an atlas), but does not exactly compute the Fréchet mean in quotient space, because a regularization term is added. In this setting, can we ensure that the consistency bias will be small enough to estimate the original template? Otherwise, one has to reconsider the template estimation with stochastic algorithms as in [Allassonnière 2010] or develop new methods [Kühnel 2017].

Chapter 4

Inconsistency when the noise level tends to infinity

Contents

A short version of this chapter has been published in a conference [Devilliers 2017a], then has been extend in the *Entropy* journal [Devilliers 2017b]. Compared to this journal paper, section 4.2.5 has been added.

Abstract: We tackle the problem of template estimation when data have been randomly deformed under a group action in the presence of noise. In order to estimate the template, one often minimizes the variance when the influence of the transformations have been removed (computation of the Fréchet mean in the quotient space). The consistency bias is defined as the distance (possibly zero) between the orbit of the template and the orbit of one element which minimizes the variance. In a first part, we restrict ourselves to isometric group action, in this case the Hilbert distance is invariant under the group action. We establish an asymptotic behavior of the consistency bias which is linear with respect to the noise level. As a result the inconsistency is unavoidable as soon as the noise is large enough. In practice, the template estimation with a finite sample is often done with an algorithm called "max-max". In a second part, also in the case of isometric group action, we show the convergence of this algorithm to an empirical Karcher mean. Our numerical experiments show that the bias observed in practice can not be attributed to the small sample size or to a convergence problem but is indeed due to the previously studied inconsistency. In a third part, we also present some insights of the case of a non invariant distance with respect to the group action. We will see that the inconsistency still holds as soon as the noise level is large enough. Moreover we prove the inconsistency even when a regularization term is added.

4.1 Introduction

4.1.1 General Introduction

The template estimation is a well known issue in different fields such as statistics on signals [Kurtek 2011b], shape theory, computational anatomy [Guimond 2000, Joshi 2004, Cootes 2004] etc. In these fields, the template (which can be viewed as the prototype of our data) can be (according to different vocabulary) shifted, transformed, wrapped or deformed due to different groups acting on data. Moreover, due to a limited precision in the measurement, the presence of noise is almost always unavoidable. These mixed effects on data lead us to study the consistency of algorithms which claim to compute the template. A popular algorithm consists in the minimization of the variance, in other words, the computation of the Fréchet mean in quotient space. This method has been already proved to be inconsistent [Bigot 2011, Miolane 2017, Devilliers 2017c]. In [Bigot 2011] the authors proves the inconsistency with a lower bound of the expectation of the error between the original template and the estimated template with a finite sample, they deduce that this expectation does not go to zero as the size of the sample goes to infinity. This work was done in a functional space, where functions only observed at a finite number of points of the functions were observed. In this case one can model these observable values on a grid. When the resolution of the grid goes to zero, one can show the consistency [Panaretos 2016] by using the Fréchet mean with the Wasserstein distance on the space of measures rather than in the space of functions. But in (medical) images the number of pixels or voxels is finite.

In [Miolane 2017], the authors demonstrated the inconsistency in a finite dimensional manifold with Gaussian noise, when the noise level tends to zero.

In chapter 3, we focused our study on the inconsistency with Hilbert Space (including infinite dimensional case) as ambient space, for isometric action. Although we gave some bounds of the consistency bias when the noise level tends to infinity, we did not gave an asymptotic behaviour of the consistency bias.

4.1.2 Settings and Notation

In this paper, we suppose that observations belong to a Hilbert space $(M, \langle \cdot, \cdot \rangle)$, we denote by $\| \cdot \|$ the norm associated to the dot product $\langle \cdot, \cdot \rangle$. We also consider a group of transformation G which acts on M the space of observations. This means that[1] $g' \cdot (g \cdot x) = (g'g) \cdot x$ and $e_G \cdot x = x$ for all $x \in M$, g, $g' \in G$, where e_G is the identity element of G.

The generative model is the following: we transform an unknown template $t_0 \in M$ with Φ a random and unknown element of the group G and we add some noise. Let σ be a positive noise level and ε a standardized noise: $\mathbb{E}(\varepsilon) = 0$, $\mathbb{E}(\|\varepsilon\|^2) = 1$. Moreover we suppose that ε and Φ are independent random variables. Finally, the only observable random variable is:

$$Y = \Phi \cdot t_0 + \sigma\varepsilon. \tag{4.1}$$

This generative model is commonly used in Computational anatomy in diverse frameworks, for instance with currents [Durrleman 2014], but also in functional data analysis [Kurtek 2011b].

For instance: if we assume that the noise is independent and identically distributed on each pixel or voxel with a standard deviation w on each pixel/voxel, then $\sigma = \sqrt{N}w$, where N is the number of pixels/voxels. But the noise which we consider can be more general: we do not require the fact that the noise is independent over each region of the space M.

Note that the inconsistency of template estimation can be also studied with an alternative generative model, called backward model where $Y = \Phi \cdot (t_0 + \sigma\varepsilon)$ chapter 3. Some authors also use the term *perturbation model* see [Huckemann 2011, Rohlf 2003, Goodall 1991].

Quotient space: the random transformation of the template by the group leads us to project the observation Y into the quotient space. The quotient space is defined as the set containing all the orbit $[x] = \{g \cdot x, \ g \in G\}$ for $x \in M$. The set which is constituted of all orbits is called the quotient space M by the group G and

[1] Note that in this chapter, $g \cdot x$ is the result of the action of g on x, and $\cdot$ should not to be confused with the multiplication of real numbers noted $\times$.

is noted by:

$$Q = M/G = \{[x],\ x \in M\}.$$

As we want to do statistics on this space, we aim to equip the quotient with a metric. One often requires that d_M the distance in the ambient space is invariant under the group action G (see figure 4.1), this means that

$$\forall m, n \in M,\ \forall g \in G \quad d_M(g \cdot m, g \cdot n) = d_M(m, n).$$

If d_M is invariant and if the orbits are closed sets[2], then

$$d_Q([x],[y]) = \inf_{g \in G} d_M(x, g \cdot y),$$

is well defined, and d_Q is a distance in the quotient space. The quotient distance $d_Q([x],[y])$ is the distance between x and y' where y' is the registration of y with respect to x. We say in this case that y' is in optimal position with respect to x.

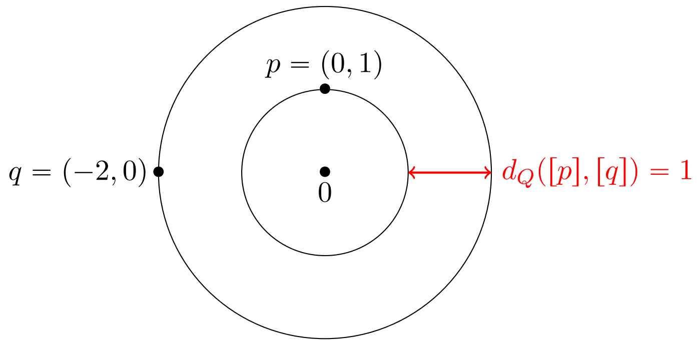


Figure 4.1: Due to the invariant action, the orbits are parallel. Here the orbits are circles centred at 0. This is the case when the group G is the group of rotations.

One particular distance in the ambient space M, which we use in all this chapter, is the distance given by the norm of the Hilbert space: $d_M(a, b) = \|a - b\|$. Moreover we say that G acts isometrically on M, if $x \mapsto g \cdot x$ is a linear map which leaves the norm unchanged. In this case d_M the distance given by the norm of the Hilbert space is invariant under the group action. The quotient (pseudo)-distance is, in this case, $d_Q([a],[b]) = \inf_{g \in G} \|a - g \cdot b\|$.

Remark 4.1. *When G acts isometrically on M a Hilbert space, by expansion of the squared norm we have:*

$$d_Q([a],[b])^2 = \|a\|^2 - 2\sup_{g \in G} \langle a, g \cdot b \rangle + \|b\|^2$$

[2]If the orbits are not closed sets, it is possible to have $d_Q([a],[b]) = 0$ even if $[a] \neq [b]$, in this case we call d_Q a pseudo-distance. Nevertheless, this has no consequence in this chapter if d_Q is only a pseudo-distance.

Thus, even if the quotient space is not a linear space, we have a "polarization identity" in the quotient space:

$$\sup_{g\in G} \langle a, g\cdot b\rangle = \frac{1}{2}\left(\|a\|^2 + \|b\|^2 - d_Q^2([a],[b])\right)$$
$$= \frac{1}{2}\left(d_Q^2([a],[0]) + d_Q^2([b],[0]) - d_Q^2([a],[b])\right) \quad (4.2)$$

When the distance given by the norm is invariant under the group action, we define the variance of the random orbit $[Y]$ as the expectation of the (pseudo)-distance between the random orbit $[Y]$ and the orbit of a point x in M:

$$F(x) = \mathbb{E}(d_Q^2([x],[Y])) = \mathbb{E}(\inf_{g\in G}\|g\cdot x - Y\|^2) = \mathbb{E}(\inf_{g\in G}\|x - g\cdot Y\|^2).$$

Note that $F(x)$ is well defined for all $x \in M$ because $\mathbb{E}(\|Y\|^2)$ is finite. Moreover, since $F(g\cdot x) = F(x)$, for all $x\in M$ and $g \in G$, the variance F is well defined in the quotient space: $[x] \mapsto F(x)$ does have a sense.

Besides, in presence of a sample of the observable variable Y noted $Y_1, \dots, Y_n$, one can define the empirical variance of a point x in M:

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n}\left(\inf_{g\in G}\|g\cdot x - Y_i\|^2\right) = \frac{1}{n}\sum_{k=1}^{n}\left(\inf_{g\in G}\|x - g\cdot Y_i\|^2\right).$$

Definition 4.1. *The template estimation is performed by minimizing F_n:*

$$\hat{t_{0n}} = argmin_{x\in M}\ F_n(x).$$

In order to study this estimation method, one can look the limit of this estimator when the number of data n tends to $+\infty$, in this case, the estimation becomes:

$$\hat{t_{0\infty}} = argmin_{x\in M}\ F(x).$$

If $m_\star \in M$ minimizes F, then $[m_\star]$ is called a Fréchet mean of $[Y]$.

Definition 4.2. *We say that the estimation is consistent if t_0 minimizes F. Moreover the consistency bias, noted CB, is the (pseudo)-distance between the orbit of the template $[t_0]$ and $[m_\star]$:*

$$CB = d_Q([t_0],[m_\star]).$$

If such a $m_\star$ does not exists, then the consistency bias is infinite.

Note that, if the action is not isometric and is not either invariant, *a priori* d_Q is no longer a (pseudo)-distance in the quotient space (this point is discussed in section 4.3). However one can still define F and wonder if the minimization of F is a consistent estimator of t_0. In this case we call F a pre-variance.

4.1.3 Questions and Contributions

This setting leads us to wonder about few things listed below:

Questions:

- Is t_0 a minimum of the variance or the pre-variance?
- What is the behavior of the consistency bias with respect to the noise level?
- How to perform such a minimization of the variance? Indeed, in practice we have only a sample and not the whole distribution.

Contribution: In the case of an isometric action, we provide a Taylor expansion of the consistency bias when the noise level σ tends to infinity. As we do not have the whole distribution, we minimize the empirical variance given a sample. An element which minimizes this empirical variance is called an empirical Fréchet mean. We already know that the empirical Fréchet mean converges to the Fréchet mean when the sample size tends to infinity [Ziezold 1977]. Therefore our problem is reduced to finding an empirical Fréchet mean with a finite but sufficiently large sample. One algorithm called the "max-max" algorithm [Allassonnière 2007] aims to compute such an empirical Fréchet mean. We establish some properties of the convergence of this algorithm. In particular, when the group is finite, the algorithm converges in a finite number of steps to an empirical Karcher mean (a local minimum of the empirical variance given a sample). This helps us to illustrate the inconsistency in this very simple framework.

We would like to insist on this point: the noise is created in the ambient space with our generative model and the computation of the Fréchet mean is done in the quotient space, this interaction induces an inconsistency. On the opposite, if one models the noise directly in the quotient space and compute the Fréchet mean in the quotient space, we have no reason to suspect any inconsistency.

Moreover it is also possible to define and use isometric actions on curves [Hitziger 2013, Kurtek 2011b] or on surfaces [Kurtek 2011a] where our work can be directly applied. The previous works related to the inconsistency of the template estimation [Bigot 2011, Miolane 2017] and chapter 3 focused to isometric action, which is a restriction to real applications. That is why, we provide, in section 4.3, some insights of the non invariant case: the inconsistency also appears as soon as the noise level is large enough.

This chapter is organized as follows: Section 4.2 is dedicated for isometric action. More precisely, in section 4.2.1, we study the presence of the inconsistency and we establish the asymptotic behavior when the noise parameter σ tends to ∞. In section 4.2.3 we detail the max-max algorithm and its properties. In section 4.2.4 we illustrate the inconsistency with synthetic data. In section 4.2.5, we see some examples of the registration score surface. This surface, seen in section 4.2.1 is a deformation of the unit sphere: for each unit vector, the amount of deformation of this vector is given by the quality of registration of data with respect to this vector.

Finally in section 4.3, we prove the inconsistency for more general group action, when the noise level is large enough. We do it in two settings, firstly, the group contains a subgroup acting isometrically on M, secondly the group acts linearly on the space M.

4.2 Inconsistency of the template estimation with an isometric action when the noise level tends to infinity.

4.2.1 Inconsistency and quantification of the consistency bias

We start with theorem 4.1 which gives us an asymptotic behavior of the consistency bias when the noise level σ tends to infinity. One key notion in theorem 4.1 is the concept of fixed point under the action G: a point $x \in M$ is a fixed point if for all $g \in G$, $g \cdot x = x$. We require that the support of the noise ε is not included in the set of fixed points. But this condition is almost always fulfilled. For instance in $\mathbb{R}^n$ the set of fixed points under a linear group action is a null set for the Lebesgue measure (unless the action is trivial: $g \cdot x = x$ for all $g \in G$ but this situation is irrelevant).

Theorem 4.1. *Let us suppose that the support of the noise ε is not included in the set of fixed points under the group action. Let Y be the observable variable defined in Equation (4.1). If the Fréchet mean of $[Y]$ exists, then we have the following lower and upper bounds of the consistency bias noted CB:*

$$\sigma K - 2\|t_0\| \leqslant CB \leqslant \sigma K + 2\|t_0\|, \tag{4.3}$$

where $K = \sup_{\|v\|=1} \mathbb{E}\left(\sup_{g\in G} \langle v, g\cdot\varepsilon\rangle\right) \in (0,1]$, K is a constant which depends only of the standardized noise and of the group action. In particular, K does not depends of the template. The consistency bias has the following asymptotic behavior when the noise level σ tends to infinity:

$$CB = \sigma K + o(\sigma) \text{ as } \sigma \to +\infty. \tag{4.4}$$

In the following we note by S the unit sphere of M. For $v \in S$, we call $\theta(v) = \mathbb{E}\left(\sup_{g\in G} \langle v, g\cdot\varepsilon\rangle\right)$, so that $K = \sup_{v\in S} \theta(v)$. The sketch of the proof is the following:

- $K > 0$ because the support of ε is not included in the set of fixed points under the action of G.
- $K \leqslant 1$ is the consequence of the Cauchy-Schwarz inequality.
- The proof of Inequalities (4.3) is based on the triangular inequalities:

$$\|m_\star\| - \|t_0\| \leqslant \text{CB} = \inf_{g\in G} \|t_0 - g\cdot m_\star\| \leqslant \|t_0\| + \|m_\star\|,$$

  where $m_\star$ minimizes F: having a piece of information about the norm of $m_\star$ is enough to deduce a piece of information about the consistency bias.

- The asymptotic Taylor expansion of the consistency bias (4.4) is the direct consequence of inequalities (4.3).

Proof of theorem 4.1. We note S the unit sphere in M. In order to prove that $K > 0$, we take x in the support of ε such that x is not a fixed point under the action of G. There exists $g_0 \in G$ such that $g_0 \cdot x \neq x$. We note $v_0 = \frac{g_0 \cdot x}{\|x\|} \in S$, we have $\langle v_0, g_0 \cdot x\rangle = \|x\| > \langle v_0, x\rangle$ and by continuity of the dot product there exists $r > 0$ such that: $\forall y \in B(x, r) \quad \langle v_0, g_0 \cdot y\rangle > \langle v_0, y\rangle$ as x is in the support of ε we have $\mathbb{P}(\varepsilon \in B(x, r)) > 0$, it follows:

$$\mathbb{P}\left(\sup_{g\in G} \langle v_0, g \cdot \varepsilon\rangle > \langle v_0, \varepsilon\rangle\right) > 0. \tag{4.5}$$

Thanks to Inequality (4.5) and the fact that $\sup_{g\in G} \langle v_0, g \cdot \varepsilon\rangle \geqslant \langle v_0, \varepsilon\rangle$ we have:

$$\theta(v_0) = \mathbb{E}\left(\sup_{g\in G} \langle v_0, g \cdot \varepsilon\rangle\right) > \mathbb{E}(\langle v_0, \varepsilon\rangle) = \langle v_0, \mathbb{E}(\varepsilon)\rangle = \langle v_0, 0\rangle = 0.$$

Then we get $K \geqslant \theta(v_0) > 0$. Moreover, if we use the Cauchy-Schwarz inequality:

$$K \leqslant \sup_{v\in S} \mathbb{E}(\|v\| \times \|\varepsilon\|) \leqslant \mathbb{E}(\|\varepsilon\|^2)^{\frac{1}{2}} = 1.$$

In order to prove Inequalities (4.3), we use the "polar" coordinates of a point in M, every point in M can be represented by (r, v) where $r \geqslant 0$ is the radius, and v belong to S the unit sphere in M, v represents the "angle". We compute $F(m)$ as a function of (r, v). In a first step, we minimize this expression as a function of r, in a second step we minimize this expression as a function of v. This makes appear the constant K (see figure 4.2).

As we said, let us take $r \geqslant 0$ and $v \in S$, we expand the variance at the point rv:

$$F(rv) = \mathbb{E}\left(\inf_{g\in G}\|rv - g \cdot Y\|^2\right) = r^2 - 2r\mathbb{E}\left(\sup_{g\in G} \langle v, g \cdot Y\rangle\right) + \mathbb{E}(\|Y\|^2). \tag{4.6}$$

Indeed $\|g \cdot Y\| = \|Y\|$ thanks to the isometric action. We note $x^+ = \max(x, 0)$ the positive part of x. Moreover we define the two following functions:

$$\lambda(v) = \mathbb{E}(\sup_{g\in G} \langle v, g \cdot Y\rangle) = \mathbb{E}(\sup_{g\in G} \langle g \cdot Y, v\rangle) \text{ and } \tilde{\lambda}(v) = \lambda(v)^+ \text{ for } v \in S,$$

since that $f : x \in \mathbb{R}^+ \mapsto x^2 - 2bx + c$ reaches its minimum at the point $r = b^+$ and $f(b^+) = c - (b^+)^2$, the $r_\star \geqslant 0$ which minimizes (4.6) is $\tilde{\lambda}(v)$ and the minimum value of the variance restricted to the half line $\mathbb{R}^+v$ is:

$$F(\tilde{\lambda}(v)v) = \mathbb{E}(\|Y\|^2) - \tilde{\lambda}(v)^2.$$

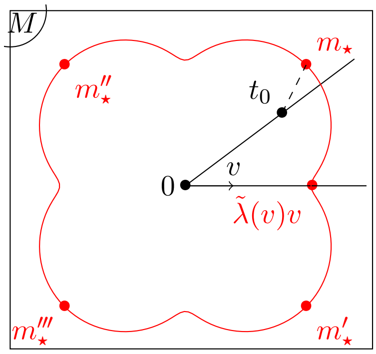


Figure 4.2: We minimize the variance on each half-line $\mathbb{R}^+ v$ where $\|v\| = 1$. The element which minimizes the variance on such a half-line is $\tilde{\lambda}(v)v$, where $\tilde{\lambda}(v) \geqslant 0$. We get a surface in M by $S \in v \mapsto \tilde{\lambda}(v)v$ (which is a curve in this figure since we draw it in dimension 2). The proof of theorem 4.1 states that if $[m_\star]$ is a Fréchet mean then $m_\star$ is an extreme point of this surface. On this picture there are four extreme points which are in the same orbit: we took here the simple example of the group of rotations of 0, 90, 180 and 270 degrees.

To find $[m_\star]$ the Fréchet mean of $[Y]$, we need to maximize $\tilde{\lambda}(v)^2$ with respect to $v \in S$:

$$m_\star = \lambda(v_\star)v_\star \text{ with } v_\star \in \underset{v \in S}{\operatorname{argmax}} \lambda(v).$$

Note that we remove the positive part and the square because argmax λ = argmax $(\lambda^+)^2$ indeed λ takes a non negative value. In order to prove it let us remark that:

$$\lambda(v) \geqslant \mathbb{E}(\langle v, \Phi \cdot t_0 + \varepsilon \rangle) = \langle v, \mathbb{E}(\Phi \cdot t_0) \rangle + 0,$$

then there are two cases: if $\mathbb{E}(\Phi \cdot t_0) = 0$ then for any $v \in S$ we have $\lambda(v) \geqslant 0$, if $w = \mathbb{E}(\Phi \cdot t_0) \neq 0$ then we take $v = \frac{w}{\|w\|} \in S$, and we get $\lambda(v) \geqslant \left\langle \frac{w}{\|w\|}, w \right\rangle = \|w\| \geqslant 0$.

As we said in the sketch of the proof we are interested in getting information about the norm of $\|m_\star\|$:

$$\|m_\star\| = \lambda(v_\star) = \sup_{v \in S} \lambda.$$

Let $v \in S$, we have: $-\|t_0\| \leqslant \langle v, g\Phi \cdot t_0 \rangle \leqslant \|t_0\|$ because the action is isometric. Now we decompose $Y = \Phi \cdot t_0 + \sigma\varepsilon$ and we get:

$$\lambda(v) = \mathbb{E}\left(\sup_{g \in G} \langle v, g \cdot Y \rangle\right) = \mathbb{E}\left(\sup_{g \in G} (\langle v, g \cdot \sigma\varepsilon \rangle + \langle v, g\Phi \cdot t_0 \rangle)\right) \quad (4.7)$$

$$\lambda(v) \leqslant \mathbb{E}\left(\sup_{g \in G} (\langle v, g \cdot \sigma\varepsilon \rangle + \|t_0\|)\right) = \sigma\mathbb{E}\left(\sup_{g \in G} \langle v, g \cdot \varepsilon \rangle\right) + \|t_0\| \quad (4.8)$$

$$\lambda(v) \geqslant \mathbb{E}\left(\sup_{g \in G} (\langle v, g \cdot \sigma\varepsilon \rangle) - \|t_0\|\right) = \sigma\mathbb{E}\left(\sup_{g \in G} \langle v, g \cdot \varepsilon \rangle\right) - \|t_0\|. \quad (4.9)$$

By taking the largest value in these inequalities with respect to $v \in S$, we get by definition of K:

$$-\|t_0\| + \sigma K \leqslant \|m_\star\| = \sup_{v \in S} \lambda(v) \leqslant \|t_0\| + \sigma K. \tag{4.10}$$

Moreover we recall the triangular inequalities:

$$\|m_\star\| - \|t_0\| \leqslant \mathrm{CB} = \inf_{g \in G} \|t_0 - g \cdot m_\star\| \leqslant \|t_0\| + \|m_\star\|, \tag{4.11}$$

Thanks to (4.10) and to (4.11), Inequalities (4.3) are proved. □

4.2.2 Remarks about theorem 4.1 and its proof

We can ensure the presence of inconsistency as soon as the signal to noise ratio satisfies $\frac{\|t_0\|}{\sigma} < \frac{K}{2}$. Moreover, if the signal to noise ratio verifies $\frac{\|t_0\|}{\sigma} < \frac{K}{3}$ then the consistency bias is not smaller than $\|t_0\|$ i.e.: $\mathrm{CB} \geqslant \|t_0\|$. In other words, the Fréchet mean in quotient space is too far from the template: the template estimation with the Fréchet mean in quotient space is useless in this case. In chapter 3 we also gave lower and upper bounds as a function of σ but these bounds were less informative than bounds given by theorem 4.1. These bounds did not give the asymptotic behaviour of the consistency bias. Moreover, in chapter 3 the lower bound goes to zero when the template becomes closed to fixed points. This may suggest that the consistency bias was small for this kind of template. We prove here that it is not the case.

Note that theorem 4.1 is not a contradiction with [Kurtek 2011b] where the authors proved the consistency of the template estimation with the Fréchet mean in quotient space for all $\sigma > 0$. Indeed their noise was included in the set of constant functions which are the fixed points under their group action.

The constant K appearing in the asymptotic behaviour of the consistency bias (4.4) is a constant of interest. We can give several (but similar) interpretations of K:

- It follows from Equation (4.3) that K is the consistency bias with a null template $t_0 = 0$ and a standardized noise ($\sigma = 1$).

- From the proof of theorem 4.1 we know that $0 < K \leqslant \mathbb{E}(\|\varepsilon\|) \leqslant 1$. On the one hand, if G is the group of rotations then $K = \mathbb{E}(\|\varepsilon\|)$, because for all v s.t. $\|v\| = 1$, $\sup_{g \in G} \langle v, g\varepsilon \rangle = \|\varepsilon\|$, by aligning v and ε. On the other hand if G acts trivially (which means that $g \cdot x = x$ for all $g \in G$, $x \in M$) then $K = 0$. The general case for K is between two extreme cases: the group where the orbits are minimal (one point) and the group for which the orbits are maximal (the whole sphere). We can state that the more the group action has the ability to align the elements, the larger the constant K is and the larger the consistency bias is.

- The squared quotient distance between two points is:

$$d_Q([a],[b])^2 = \|a\|^2 - 2\sup_{g\in G} \langle a, g\cdot b\rangle + \|b\|^2,$$

thus the larger $\sup_{g\in G} \langle a, g\cdot b\rangle$, the smaller $d_Q([a],[b])$. We get:

$$K = 1 - \frac{1}{2}\inf_{\|v\|=1} \mathbb{E}(d_Q^2([v],[\varepsilon]))$$

Therefore, the constant K encodes the level of contraction of the quotient distance (or folding). The larger K is, the more contracted the quotient space is.

- In chapter 3, we say that if the random variable was not included in the cone of the template, there was inconsistency. However in theorem 4.1 there is apparently no this notion of cone. In fact, this notion of cone appears, indeed we have:

$$K > 0 \iff \exists v \in M \text{ s.t. } \|v\| = 1 \text{Support}(\varepsilon) \not\subset \text{Cone}(v).$$

Besides, $K > 0$ is necessary to see inconsistency in theorem 4.1.

One disadvantage of theorem 4.1 is that it ensures the presence of inconsistency for σ large enough but it says nothing when σ is small, in this case one can refer to [Miolane 2017] or chapter 3. We can give a refinement of theorem 4.1.

Corollary 4.1. *Under the hypothesis of theorem 4.1, we have:*

$$\sigma K - 2\,dist(t_0, Fix(M)) \leqslant CB \leqslant \sigma K + 2\,dist(t_0, Fix(M)),$$

where $Fix(M)$ is the set of fixed points under the group action.

Proof. Indeed, in chapter 3 we have seen lemma 3.8 which states that we can translate the random variable by a fixed point, besides this translation is a congruent map. Therefore, it suffices to apply theorem 4.1 to $Y - m_0 = \Phi(t_0 - m_0) + \sigma\varepsilon$ where m_0 is the orthogonal projection of t_0 on $\text{Fix}(M)$. □

We deduce immediately from corollary 4.1 a result which was already proved in chapter 3 (proposition 3.11):

Corollary 4.2. *Under the hypothesis of theorem 4.1, if t_0 is a fixed point, we have an expression of the consistency bias:*

$$CB = \sigma K.$$

In the proof of theorem 4.1, we have seen that the minimum of the variance restricted to the half-line $\mathbb{R}^+ v$ for $v \in S$, was

$$\mathbb{E}(\|Y\|^2) - \left(\left(\mathbb{E}\left(\inf_{g\in G}\langle v, g\cdot Y\rangle\right)\right)^+\right)^2.$$

therefore $\tilde{\lambda}(v) = \left(\mathbb{E}\left(\inf_{g\in G}\langle v, g\cdot Y\rangle\right)\right)^+$ is a registration score: $\tilde{\lambda}(v)$ tells you how much it is a good idea to search the Fréchet mean of $[Y]$ in the direction pointed by v: the more $\tilde{\lambda}(v)$ is large, the more v is a good choice. On the contrary when this value is equal to zero, it is useless to search the Fréchet mean in this direction.

Likewise, for $v \in S$, $\theta(v) = \mathbb{E}(\sup_{g\in G}\langle g\cdot v, \varepsilon\rangle)$ is a registration score with respect to the noise, the larger $\theta(v)$, the more the unit vector v looks like to the noise ε after registration.

If $[m_\star]$ is a Fréchet mean of $[Y]$ we have seen that its norm verifies:

$$\|m_\star\| = \sup_{\|v\|=1}\mathbb{E}(\sup_{g\in G}\langle v, g\cdot Y\rangle).$$

Then if there are two different Fréchet means of $[Y]$ noted $[m_\star]$ and $[n_\star]$, we can deduce that $\|m_\star\| = \|n_\star\|$. Even if there is no uniqueness of the Fréchet mean in the quotient space, we can state that the representants of the different Fréchet means have all the same norm.

We can also wonder if the converse of theorem 4.1 is true: if ε is a non biased noise always included in the set of fixed points, is $[t_0]$ a Fréchet mean of $[\Phi\cdot t_0 + \sigma\varepsilon]$?

Proposition 4.1. *Let G be a group acting isometrically on M an Hilbert space. We consider a template t_0 and ε a standardized noise. We define the observable variable Y by $Y = \Phi\cdot t_0 + \sigma\varepsilon$, where $\sigma > 0$. If ε belongs almost surely in the set of fixed points, then t_0 is a Fréchet mean of $[Y]$.*

Proof. A simple computation show that t_0 is a minimum of the variance:

$$\begin{aligned}
F(m) &= \mathbb{E}\left(\inf_{g\in G}\|m - g\cdot(\Phi\cdot t_0 + \sigma\varepsilon)\|^2\right)\\
&= \|m\|^2 + \mathbb{E}(\|\Phi\cdot t_0 + \sigma\varepsilon\|^2) - 2\mathbb{E}(\sup_{g\in G}\langle m, g\Phi\cdot t_0\rangle + \langle m, g\cdot\sigma\varepsilon\rangle)\\
&= \|m\|^2 + \mathbb{E}(\|\Phi\cdot t_0 + \sigma\varepsilon\|^2) - 2\mathbb{E}\left(\sup_{g\in G}\langle m, g\cdot t_0\rangle\right) - 2\langle m, \mathbb{E}(\sigma\varepsilon)\rangle\\
&= \|m\|^2 + \mathbb{E}(\|\Phi\cdot t_0 + \sigma\varepsilon\|^2) - 2\mathbb{E}\left(\sup_{g\in G}\langle m, g\cdot t_0\rangle\right) \qquad (4.12)
\end{aligned}$$

We see that the element m which minimizes (4.12) does not depend of σ, in particular we can assume $\sigma = 0$, and wonder which elements minimizes

$$F(m) = \mathbb{E}(\inf_{g\in G}\|m - g\Phi\cdot t_0\|^2) = \inf_{g\in G}\|m - g\cdot t_0\|^2$$

it becomes clear that only the points in the closure of the orbit of t_0 can minimize this variance. □

Then when ε is included in the set of fixed points, the estimation is always consistent for all σ. This is an alternative proof of the Theorem of consistency done by Kurtek et al. [Kurtek 2011b].

In the proof of theorem 4.1, we have seen that the direction of the Fréchet mean of $[Y]$ is given by the supremum of this quantity (4.7):

$$\mathbb{E}\left(\sup_{g\in G} \langle v, g\cdot\sigma\varepsilon\rangle + \langle v, g\Phi\cdot t_0\rangle\right).$$

This Equation is a good illustration of the difficulty to compute the Fréchet mean in quotient space. Indeed, we have on one side the contribution of the noise $\langle v, g\cdot\sigma\varepsilon\rangle$ and on the other side the contribution of the template $\langle v, g\Phi\cdot t_0\rangle$, and we take the supremum of the sum of these two contributions over $g \in G$. Unfortunately the supremum of the sum of two terms is not equal to the sum of the supremum of each of these terms. Hence, it is difficult to separate these two contributions. However, we can intuit that when the noise is large, $\langle v, g\cdot\sigma\varepsilon\rangle$ prevails over $\langle v, g\Phi\cdot t_0\rangle$, and the use of the Cauchy-Schwarz inequality in Equations (4.8) and (4.9) proves it rigorously. We can conclude that, when the noise is large, the direction of the Fréchet mean in the quotient space depends more on the noise than on the template.

4.2.3 Template estimation with the Max-Max Algorithm

4.2.3.1 Max-Max Algorithm Converges to a Local Minima of the Empirical Variance

Section 4.2.1 can be understood as follows: if we want to estimate the template by minimizing the Fréchet mean in the quotient space, then there is a bias. This supposes that we are able to compute such a Fréchet mean. In practice, we cannot minimize the exact variance in quotient space, because we have only a finite sample and not the whole distribution. In this section we study the estimation of the empirical Fréchet mean with the max-max algorithm. We assume that the group is finite. In this case, the registration can always be found by an exhaustive search. Hence, the numeric experiments which we conduct in section 4.2.4 lead to an empirical Karcher mean in a finite number of steps. In a compact group acting continuously, the registration also exists but is not necessarily computable without approximation.

If we have a sample: $Y_1, \ldots, Y_I$ of independent and identically distributed copies of Y, then we define the empirical variance in the quotient space:

$$\begin{aligned} M \ni x \mapsto F_I(x) &= \frac{1}{I}\sum_{i=1}^{I} d_Q^2([x],[Y_i]) \qquad (4.13)\\ &= \frac{1}{I}\sum_{i=1}^{I} \min_{g_i\in G}\|x - g_i\cdot Y_i\|^2 \\ &= \frac{1}{I}\sum_{i=1}^{I} \min_{g_i\in G}\|g_i\cdot x - Y_i\|^2. \end{aligned}$$

The empirical variance is an approximation of the variance. Indeed thanks to the law of large number we have $\lim_{I\to\infty} F_I(x) = F(x)$ for all $x \in M$. One element which minimizes globally (respectively locally) F_I is called an empirical Fréchet mean (respectively an empirical Karcher mean). For $x \in M$ and $\underline{g} \in G^I$: $\underline{g} = (g_1, \ldots, g_I)$ where $g_i \in G$ for all $i = 1..I$ we define J an auxiliary function by:

$$J(x, \underline{g}) = \frac{1}{I}\sum_{i=1}^{I} \|x - g_i \cdot Y_i\|^2 = \frac{1}{I}\sum_{i=1}^{I} \|g_i^{-1} \cdot x - Y_i\|^2.$$

The max-max algorithm (algorithm 1) iteratively minimizes the function J in the variable $x \in M$ and in the variable $\underline{g} \in G^I$ (see also figure 4.3). This algorithm is nothing else than a gradient descent, it has also known as Procrustes Analysis [Gower 1975, Goodall 1991].

Algorithm 1 Max-Max algorithm

Require: A starting point $m_0 \in M$, a sample $Y_1, \ldots, Y_I$.
  $n = 0$.
  while Convergence is not reached **do**
    Minimizing $\underline{g} \in G^I \mapsto J(m_n, \underline{g})$: we get g_i^n by registering Y_i with respect to m_n.
    Minimizing $x \in M \mapsto J(x, \underline{g}^n)$: we get $m_{n+1} = \frac{1}{I}\sum_{i=1}^{I} g_i^n Y_i$.
    $n = n + 1$.
  end while
  $\hat{m} = m_n$

First, we note that this algorithm is sensitive to the the starting point. However we remark that $m_1 = \frac{1}{I}\sum_{i=1}^{I} g_i \cdot Y_i$ for some $g_i \in G$, thus without loss of generality, we can start from $m_1 = \frac{1}{I}\sum_{i=1}^{I} g_i \cdot Y_i$ for some $g_i \in G$. The empirical variance does not increase at each step of the algorithm since:

$$F_I(m_n) = J(m_n, \underline{g}^n) \geqslant J(m_{n+1}, \underline{g}^n) \geqslant J(m_{n+1}, \underline{g}^{n+1}) = F_I(m_{n+1})$$

Proposition 4.2. *As the group is finite, the convergence is reached in a finite number of steps.*

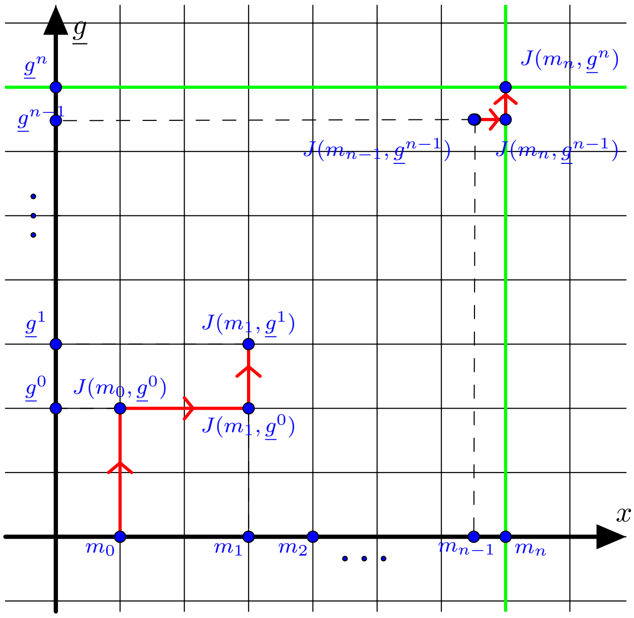


Figure 4.3: Iterative minimization of the function J on the two axes, the horizontal axis represents the variable in the space M, the vertical axis represents the set of all the possible registrations G^I. Once the convergence is reached, the point (m_n, g_n) is the minimum of the function J on the two axis in green. Is this point the minimum of J on its whole domain? There are two pitfalls: firstly this point could be a saddle point, it can be avoided with proposition 4.3, secondly this point could be a local (but not global) minimum, this is discussed in section 4.2.4.3

Proof of proposition 4.2. The sequence $(F_I(m_n))_{n\in\mathbb{N}}$ is non-increasing. Moreover the sequence $(m_n)_{n\in\mathbb{N}}$ takes value in a finite set which is: $\{\frac{1}{I}\sum_{i=1}^I g_i \cdot Y_i,\ g_i \in G\}$. Therefore, the sequence $(F_I(m_n))_{n\in\mathbb{N}}$ is stationary. Let $n \in \mathbb{N}$ such that $F_I(m_n) = F_I(m_{n+1})$. Hence the empirical variance did not decrease between step n and step $n+1$ and we have:

$$F_I(m_n) = J(m_n, \underline{g}_n) = J(m_{n+1}, \underline{g}_n) = J(m_{n+1}, \underline{g}_{n+1}) = F_I(m_{n+1}),$$

as m_{n+1} is the unique element which minimizes $m \mapsto J(m, \underline{g}_n)$ we conclude that $m_{n+1} = m_n$. □

This proposition gives us a shutoff parameter in the max-max algorithm: we stop the algorithm as soon as $m_n = m_{n+1}$. Let us call $\hat{m}$ the final result of the max-max algorithm. It may seem logical that $\hat{m}$ is at least a local minimum of the empirical variance. However this intuition may be wrong: let us give a toy counterexample, suppose that we observe $Y_1, \dots, Y_I$, due to the transformation of the group it is possible that $\sum_{i=1}^n Y_i = 0$. We can start from $m_1 = 0$ in the max-max algorithm, as Y_i and 0 are already registered, the max-max algorithm does not transform Y_i. At step two, we still have $m_2 = 0$, by induction the max-max algorithm stays at 0 even if 0 is not a Fréchet or Karcher mean of $[Y]$. Because 0 is equally distant from all the points in the orbit of Y_i, 0 is called a focal point of $[Y_i]$. The notion of focal point is important for the consistency of the Fréchet mean in manifold [Bhattacharya 2008]. Fortunately, the situation where $\hat{m}$ is not a Karcher mean is almost always avoided due to the following statement:

Proposition 4.3. *Let $\hat{m}$ be the result of the max-max algorithm. If the registration of Y_i with respect to $\hat{m}$ is unique, in other words, if $\hat{m}$ is not a focal point of Y_i for all $i \in 1..I$ then $\hat{m}$ is a local minimum of F_I: $[\hat{m}]$ is an empirical Karcher mean of $[Y]$.*

Note that, if we call z the registration of y with respect to m, then the registration is unique if and only if $\langle m, z - g \cdot z \rangle \neq 0$ for all $g \in G \backslash \{e_G\}$. Once the max-max algorithm has reached convergence, it suffices to test this condition for $\hat{m}$ obtained by the max-max algorithm and Y_i for all i. This condition is in fact generic and is always obtained in practice.

Proof of proposition 4.3. We call g_i the unique element in G which register Y_i with respect to $\hat{m}$, for all $h \in G \backslash \{g_i\}$, $\|\hat{m} - g_i \cdot Y_i\| < \|\hat{m} - h_i \cdot Y_i\|$. By continuity of the norm we have for a close enough to m: $\|a - g_i \cdot Y_i\| < \|a - h_i \cdot Y_i\|$ for all $h_i \neq g_i$ (note that this argument requires a finite group). The registrations of Y_i with respect to m and to a are the same:

$$F_I(a) = \frac{1}{I}\sum_{i=1}^{I} \|a - g_i \cdot Y_i\|^2 = J(a, \underline{g}) \geqslant J(\hat{m}, \underline{g}) = F_I(\hat{m}),$$

because $m \mapsto J(m, \underline{g})$ has one unique local minimum $\hat{m}$. □

This condition of the unique registration may be seem, odd, in fact this is a natural condition. Indeed, we can state a result which states that if $[m_{n\star}]$ is a empirical Fréchet mean, then the registration of the data is unique:

Proposition 4.4. *Let G a group acting isometrically on M a Hilbert space, let $Y_1, \dots, Y_n$ being a sample of a random variable Y, let $[m_{n\star}]$ being a empirical Fréchet mean of Y. Then $m_{n\star} = \frac{1}{n}\sum_{i=1}^{n} g(Y_i, m_{n\star}) \cdot Y_i$, where $g(Y_i, m_\star) \in \underset{g \in G}{argmin} \|g \cdot Y_i - m_{n\star}\|$. Furthermore, $g(Y_i, m_\star)$ is unique up to an element of $Iso(Y_i)$, this means that:*

$$g(Y_i, m_{n\star}), \tilde{g}(Y_i, m_{n\star}) \in \underset{g \in G}{argmin} \|gY_i - m_{n\star}\| \Longrightarrow g(Y_i, m_{n\star}) \cdot Y_i = \tilde{g}(Y_i, m_{n\star}) \cdot Y_i$$

Proof. We want to find:

$$\underset{g_1, \cdots g_n}{\text{argmin}}\ \underset{m \in M}{\text{argmin}} \frac{1}{n}\sum_{i=1}^{n} \|m - g_i \cdot Y_i\|^2.$$

Then, if $[m_{n\star}]$, we have $m_{n\star} = \frac{1}{n}\sum_{i=1}^{n} \|m - g_i \cdot Y_i\|^2$. Then we need to minimize:

$$\underset{g_1, \cdots g_n}{\text{argmin}} \frac{1}{n}\sum_{i=1}^{n} \left\| \left(\frac{1}{n}\sum_{i=1}^{n} g_i \cdot Y \right) - g_i \cdot Y_i \right\|^2.$$

Therefore $m_{n\star} = \frac{1}{n}\sum_{i=1}^{n} g(Y_i, m_{n\star}) \cdot Y_i$. For every j, let us consider:

$$g(Y_j, m_{n\star}) \cdot Y_j, \quad \tilde{g}(Y_j, m_{n\star}) \cdot Y \in \underset{g\in G}{\operatorname{argmin}} \|g \cdot Y_j - m_{n\star}\|.$$

Then we have:

$$m_{n\star} = \frac{1}{n}\sum_{i=1}^{n} g(Y_i, m_{n\star}) \cdot Y_i = \frac{1}{n}\left(\sum_{i=1, i\neq j}^{n} (g(Y_i, m_{n\star}) \cdot Y_i) + \tilde{g}(Y_j, m_{n\star}) \cdot Y_j\right).$$

By simplifying the sum, we get $g(Y_j, m_{n\star}) \cdot Y_j = \tilde{g}(Y_j, m_{n\star}) \cdot Y_j$. □

In chapter 5, proposition 5.2, we generalize proposition 4.4, at the Fréchet mean of $[Y]$ instead of the empirical Fréchet mean of $Y_1, \dots Y_n$.

Remark 4.2. *We remark the max-max algorithm is in fact a gradient descent. The gradient descent is a general method to find the minimum of a differentiable function. Here we are interested in the minimum of the variance F: let $m_0 \in M$ and we define by induction the gradient descent of the variance $m_{n+1} = m_n - \rho \nabla F(m_n)$, where $\rho > 0$ and F the variance in the quotient space. In chapter 3, the gradient of the variance in quotient space for finite group and for a regular point m was computed (m is regular as soon as $g \cdot m = m$ implies $g = e$), this leads to:*

$$m_{n+1} = m_n - 2\rho \left[m_n - \mathbb{E}(g(Y, m_n) \cdot Y)\right],$$

where $g(Y, m_n)$ is the almost-surely unique element of the group which registers Y with respect to m_n. Now if we have a set of data $Y_1, \dots, Y_n$ we can approximated the expectation which leads to the following approximated gradient descent:

$$m_{n+1} = m_n(1 - 2\rho) + \rho \frac{2}{I}\sum_{i=1}^{I} g(Y_i, m_n) \cdot Y_i,$$

now by taking $\rho = \frac{1}{2}$ we get $m_{n+1} = \frac{1}{I}\sum_{i=1}^{I} g(Y_i, m_n) \cdot Y_i$. So the approximated gradient descent with $\rho = \frac{1}{2}$ is exactly the max-max algorithm. But the max-max algorithm for finite group, is proved to be converging in a finite number of steps which is not the case for gradient descent in general.

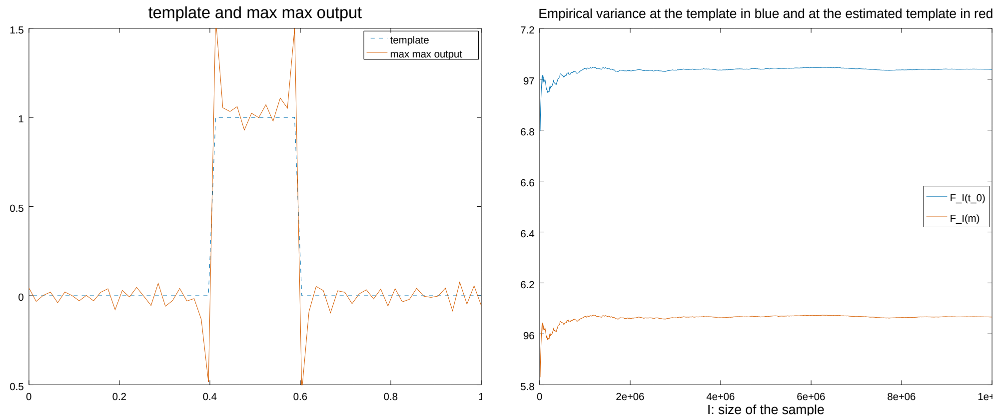
4.2.4 Simulation on synthetic data

In this section, we consider data in an Euclidean space $\mathbb{R}^N$ equipped with its canonical dot product $\langle \cdot, \cdot \rangle$, and $G = \mathbb{Z}/\mathbb{N}\mathbb{Z}$ acts on $\mathbb{R}^N$ by horizontal translation:

$$\begin{array}{ccc} \mathbb{Z}/N\mathbb{Z} \times \mathbb{R}^N & \to & \mathbb{R}^N \\ (\bar{k}, (x_1, \dots, x_N)) & \mapsto & (x_{1+k}, x_{2+k}, \dots x_{N+k}) \end{array},$$

where indexes are taken modulo N. This space models the discretization of functions defined on $[0, 1]$ with N points. This action is found in [Allassonnière 2007] and used for neuroelectric signals in [Hitziger 2013]. The registration between two vectors can be made by an exhaustive research but it is faster with the fast Fourier transform [Cooley 1965].

4.2.4.1 Max-max algorithm with a step function as template



(a) Example of a template (a step function) and the estimated template $\hat{m}$ with a sample size 10^5 in $\mathbb{R}^{64}$, ε is Gaussian noise and $\sigma = 10$. At the discontinuity points of the template, we observe a Gibbs-like phenomena.

(b) Variation of $F_I(t_0)$ (in blue) and of $F_I(\hat{m})$ (in red) as a function of I the size of the sample. Since convergence is already reached, $F(\hat{m})$, which is the limit of red curve, is below $F(t_0)$: $F(t_0)$ is the limit of the blue curve. Due to the inconsistency, $\hat{m}$ is an example of point such that $F(\hat{m}) < F(t_0)$.

Figure 4.4: Template t_0 and template estimation $\hat{m}$ on figure 4.4a. Empirical variance at the template and the template estimation with the max-max algorithm as a function of the size of the sample on figure 4.4b.

We display an example of a template and the template estimation with the max-max algorithm on figure 4.4a. This experiment was already conducted in [Allassonnière 2007], but no explanation of the appearance of the bias was provided. We know from section 4.2.3 that the max-max output is an empirical Karcher mean, and that this result can be obtained in a finite number of steps. Taking $\sigma = 10$ may seem extremely high, however the standard deviation of the noise at each point is not 10 but $\frac{\sigma}{\sqrt{N}} = 1.25$ which is reasonable.

The sample size is 10^5, the algorithm stopped after 247 steps, and $\hat{m}$ the estimated template (in red on the figure 4.4a) is not a focal points of the orbits $[Y_i]$, then proposition 4.3 applies. We call empirical bias (noted EB) the quotient distance between the true template and the point $\hat{m}$ given by the max-max result. On this experiment we have $\frac{EB}{\sigma} \simeq 0.11$. Of course, one could think that we estimate the template with an empirical bias due to a too small sample size which induces fluctuation. To reply to this objection, we keep in memory $\hat{m}$ obtained with the max-max algorithm. If there was no inconsistency then we would have $F(t_0) \leqslant F(\hat{m})$. We do not know the value of the variance F at these points, but thanks to the law of large number, we know that:

$$F(t_0) = \lim_{I \to \infty} F_I(t_0) \text{ and } F(\hat{m}) = \lim_{I \to \infty} F_I(\hat{m}),$$

Given a sample, we compute $F_I(t_0)$ and $F_I(\hat{m})$ thanks to the definition of the empirical variance F_I (4.13). We display the result on figure 4.4b, this tends to confirm that $F(t_0) > F(\hat{m})$. In other word, the variance at the template is larger that the variance at the point given by the max-max algorithm.

4.2.4.2 Max-max algorithm with a continuous template

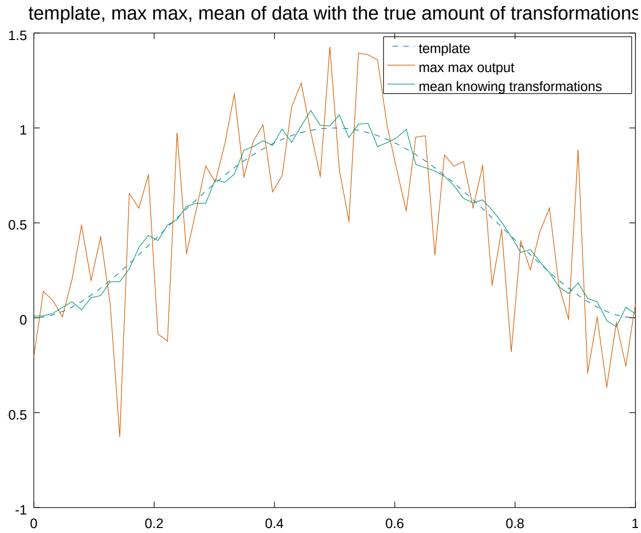


Figure 4.5: Example of an other template (here a discretization of a continuous function) and the template estimation with a sample size 10^3 in $\mathbb{R}^{64}$, ε is Gaussian noise and $\sigma = 10$. Even with a continuous function the inconsistency appears.

Figure 4.4a shows that the main source of the inconsistency was the discontinuity of the template. One may think that a continuous template would lead to a better behaviour. But it is not the case as presented on figure 4.5. Even with a large number of observations created from a continuous template we do not observe a convergence to the template. In the example of figure 4.5, the empirical bias satisfies $\frac{EB}{\sigma} = 0.23$. In green we also display the mean of data knowing transformations. This means that if the data are on the form $Y_i = \Phi_i \cdot t_0 + \sigma\varepsilon_i$, and if we know Φ_i then we can estimate the template by

$$\frac{1}{n}\sum_{i=1}^{n}\Phi_i^{-1}\cdot Y_i = t_0 + \frac{\sigma}{n}\sum_{i=1}^{n}\Phi_i^{-1}\varepsilon_i,$$

this produces a much better result, since that in this case we have $\frac{EB}{\sigma} = 0.04$. But in practice, we do not know Φ_i. A sample of size 10^3 may seem a too small. Thus we can do the same thing with a sample of size 10^6 on figure 4.6. We do not observe a much better estimation.

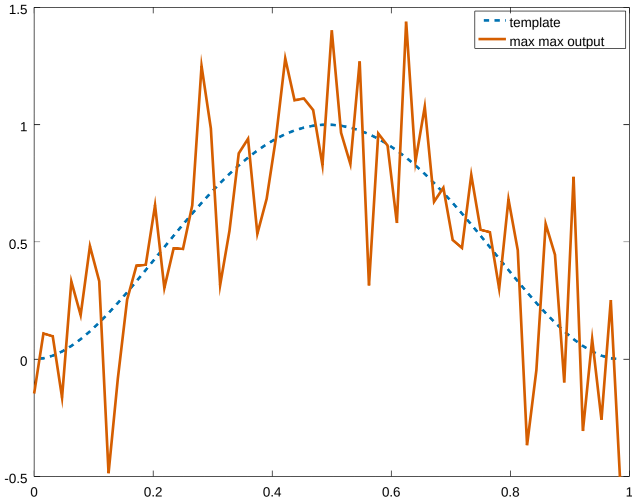


Figure 4.6: Example of the continuous template (here a discretization of a continuous function) and the template estimation with a sample size 10^6 in $\mathbb{R}^{64}$, ε is Gaussian noise and $\sigma = 10$.

4.2.4.3 Does the max-max algorithm give us a global minimum or only a local minimum of the variance?

Proposition 4.3 tells us that the output of the max-max algorithm is a Karcher mean of the variance, but we do not know whether it is Fréchet mean of the variance. In other words, is the output a global minimum of the variance? In fact, F_I has a lot of local minima which are not global. To illustrate this, we may use the max-max algorithm with different starting points and we observe different outputs (which are all local minima thanks to proposition 4.3) with different empirical variances on table 4.1.

Points	t_0	$\hat{m}_1$	$\hat{m}_2$	$\hat{m}_3$	$\hat{m}_4$	$\hat{m}_5$
Empirical variance at these points	97.068	96.073	96.074	96.074	96.074	96.074

Table 4.1: Empirical variances at the template and at 5 different outputs of the max-max algorithm coming from the same sample of size 10^5 $(\hat{m}_i)_{1\leqslant i\leqslant 5}$, but with different starting points. We use $\sigma = 10$ and the action of horizontal translation in $\mathbb{R}^{64}$. Conclusion: on these five points, only $\hat{m}_1$ is an eventual global minima.

We observe that the variances at these points are very close. We also display these local minima on figure 4.7

Moreover we can compute all the quotient distances given two points in the set $\{t_0, \hat{m}_i\}$ (table 4.2). We observe that all the Karcher means are concentrated compared to the distance to the real template. One can intuit that the whole set of Karcher means (including the Fréchet means) are concentrated in this neighborhood.

Of course all these results are sensitive to the size of the sample. In the following

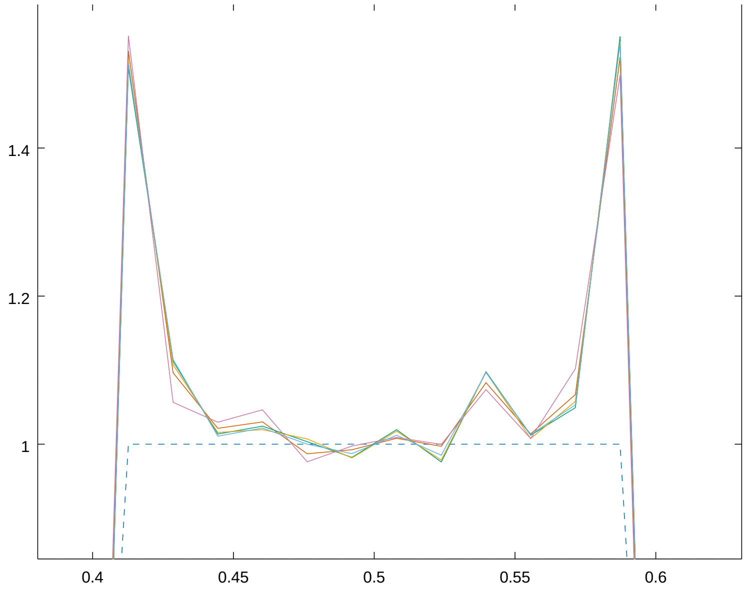


Figure 4.7: Several local minima computed from the same sample of size 10^5, we took the same step function as template. We zoomed in order to see better the small differences between the different outputs of the max-max algorithm given five different starting points.

we launch 100 times the max-max algorithm with 100 different starting points. We did it for a sample of size 10^5 and with a sample of size 10^4. In the figure 4.8 we plot an histogram of the distance between the template and the outputs of the max-max algorithm. We observe that this distance depends less on the starting point with a large sample. On this example, we can intuit that with an infinite number of observations the distance between the template and the Fréchet mean is unique: whatever the starting point we would have the same bias.

4.2.5 Example of registration score surface

In the proof of theorem 4.1, we have seen that the direction of the Fréchet mean was given by maximizing the $\tilde{\lambda}$ function on the sphere: $S \ni v \mapsto \left(\mathbb{E}\left(\sup_{g\in G} \langle v, g\cdot Y\rangle\right)\right)^+$. Through this process, we obtain a hyper-surface in the Hilbert space M which is a deformation of the unit sphere. The current section aims to study such hyper-surfaces on an example.

In order to obtain such a hyper-surface, we approximate the expectation in the computation of $\tilde{\lambda}$ by the empirical expectation: we have an random hyper-surface depending on the sample of Y we simulate. Then we can draw $S \ni v \mapsto \tilde{\lambda}(v)v$ in dimension 2 or 3. In dimension 2 for the action of rotation of angle $\frac{2\pi}{n}$ we observe

Quotient distance	t_0	$\hat{m}_1$	$\hat{m}_2$	$\hat{m}_3$	$\hat{m}_4$	$\hat{m}_5$
t_0	0	1.10584	1.10632	1.10589	1.10745	1.10584
$\hat{m}_1$	1.10584	0	0.09735	0.09807	0.13743	0.08668
$\hat{m}_2$	1.10632	0.09735	0	0.03114	0.20498	0.04093
$\hat{m}_3$	1.10589	0.09807	0.03114	0	0.20268	0.04140
$\hat{m}_4$	1.10745	0.13743	0.20498	0.20268	0	0.20398
$\hat{m}_5$	1.10584	0.08668	0.04093	0.04140	0.20398	0

Table 4.2: Quotient distance between two pairs of points among the template and the five Karcher means, we remark that the different Karcher means are closed to each other compared to the distance between the template, in other words the set of Karcher means seems localized in a small neighborhood far from the template

roughly a rose with n petals. Moreover this curve depends on the noise level.

Example 4.1. *For the action of horizontal translation on* $\mathbb{R}^3$. *We can change this noise level and observe the resulting movie on* *http://loic.devilliers.free.fr/rosace3animation.gif.*

For $\sigma = 0$ we observe that the hyper-surface is three pieces of three spheres. This can be shown and generalized in any dimension (including infinite). Indeed for $\sigma = 0$ we have:

$$\tilde{\lambda}(v) = \mathbb{E}\left(\sup_{g \in G} \langle g\Phi \cdot t_0, v\rangle\right)^+ .$$

Since $\tilde{\lambda}$ is invariant under the group action: $\tilde{\lambda}(v) = \tilde{\lambda}(g \cdot v)$, without loss of generality we can assume that v belongs to the cone of the template. In this case we have $\tilde{\lambda}v = (\langle v, t_0\rangle)^+$. Then the point on the hyper-surface is $\langle v, t_0\rangle v$ or 0 for v belonging to $Cone(t_0)$, and $\langle v, t_0\rangle \geqslant 0$. Moreover:

$$\{0\} \bigcup \{\langle v, t_0\rangle v,\ v \in S\} \subset \left\{ p \in \mathbb{R}^3 \left\| p - \frac{\|t_0\|}{2} \right\|^2 = \frac{\|t_0\|^2}{4} \right\}.$$

In other words, these points are on the sphere of center $\frac{t_0}{2}$ and radius $\frac{\|t_0\|}{2}$. In particular the segment $[0, t_0]$ is a diameter of this circle. We can conclude that, when $\sigma = 0$ this hyper-surface is a union of part of the spheres. There are as many spheres as points in the orbits of the template.

When $\sigma > 0$, note that it is possible to have $\tilde{\lambda}(v) = 0$. For instance assume that the template is $t_0 \in (\mathbb{R}^+)^3$, if the noise is sufficiently small, then Y belongs almost surely in $(\mathbb{R}^+)^3$. Let $v \in S$ be an unit vector in $(\mathbb{R}^-)^3$ then $\sup_{g \in G} \langle v, g \cdot Y\rangle < 0$ almost surely: the angle between v and $g \cdot Y$ is obtuse for every $g \in G$. Then $\tilde{\lambda}v = 0$. This makes the hyper-surface singular to zero, since some vectors of the sphere collapse to zero.

When σ is not equal to zero, it is harder to interpret the resulting surface. However, we believe that there is a link between the geometrical property of the

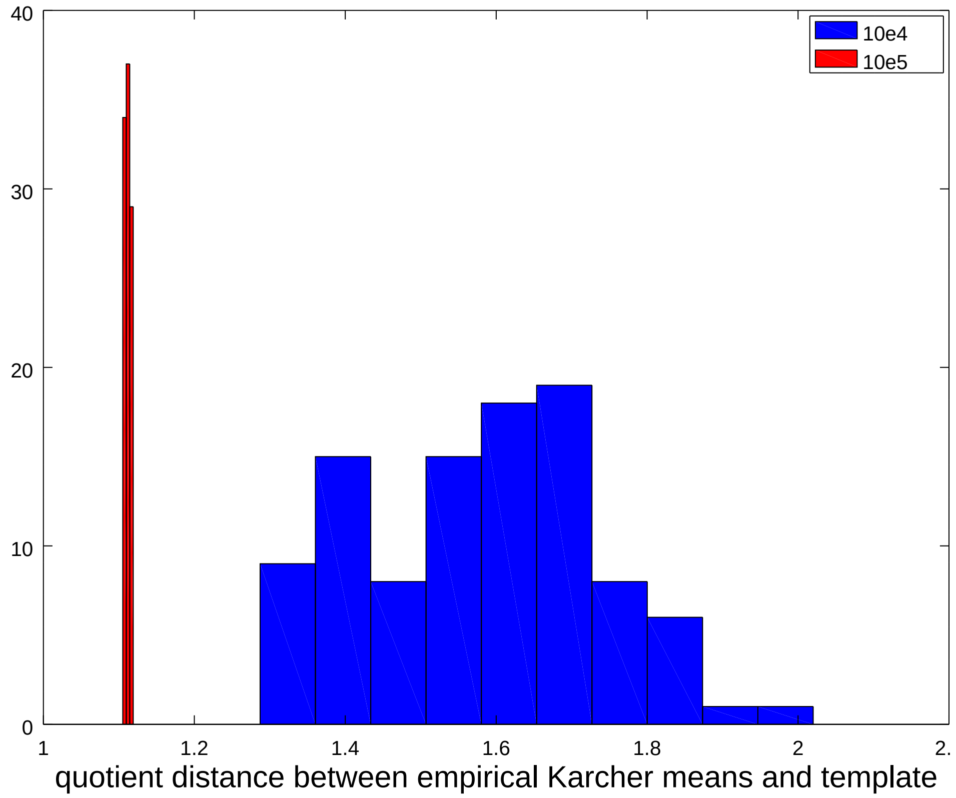


Figure 4.8: Histogram of the quotient distance between the template and the output of the max-max algorithm, in red we have the distribution of the distance of 100 max-max algorithm with the same sample of size 10^5, in blue the distribution with a sample of size 10^4. The larger the size of the sample, the more concentrated the distance is.

surface and the statistical property. We advocate, that the discovery of such a links matters for a better understanding of the consistency bias.

4.3 Inconsistency in the case of non invariant distance under the group action

4.3.1 Notation and hypothesis

In this section, data still come from a Hilbert space M. But we take a group of deformation G which acts in a non invariant way on M. Starting from a template t_0 we consider a random deformation in the group G namely a random variable Φ which takes value in G and ε an standardized noise in M independent of Φ. We

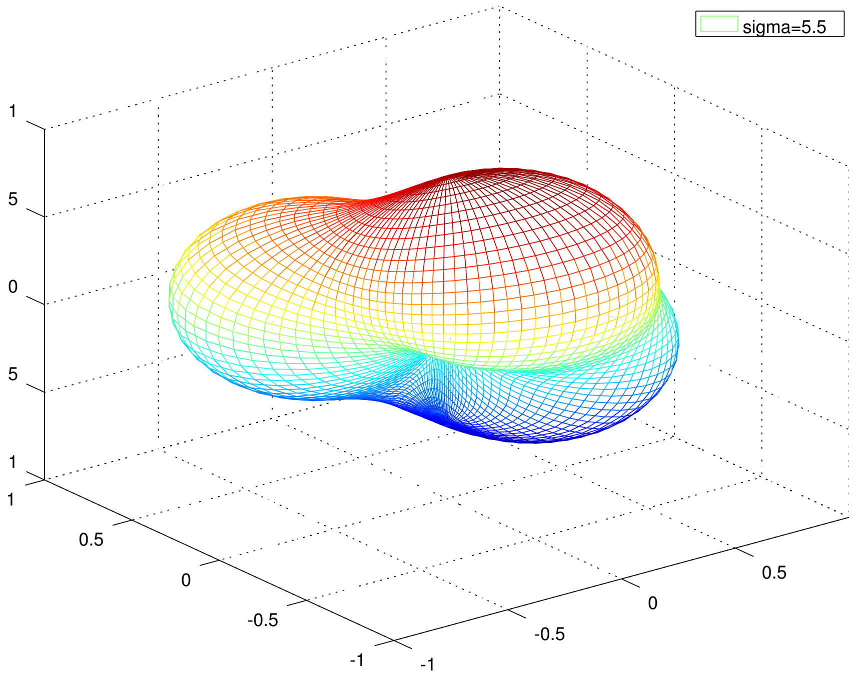


Figure 4.9: Rose in the case of horizontal translation in $\mathbb{R}^3$ with a noise level $\sigma = 5.5$, we use a parametrization of a discretized of the sphere with 2500 points.

suppose that our observable random variable is:

$$Y = \Phi \cdot t_0 + \sigma\varepsilon \text{ with } \sigma > 0,\ \mathbb{E}(\varepsilon) = 0,\ \mathbb{E}(\|\varepsilon\|^2) = 1,$$

where σ is the noise level. We suppose that $\mathbb{E}(\|Y\|^2) < +\infty$, and we define the pre-variance of Y in M/G as the map defined by:

$$F(m) = \mathbb{E}\left(\inf_{g \in G} \|g \cdot m - Y\|^2\right).$$

In this part we still study the inconsistency of the template estimation by minimizing F.

We present two frameworks where we can ensure the presence of inconsistency: in section 4.3.3 we suppose that the group G contains a non trivial group H which acts isometrically on M. But some groups do not satisfy this hypothesis, that is why, in section 4.3.4 we do not suppose that G contains a subgroup acting isometrically but we require that G acts linearly on M. In both sections we prove inconsistency as soon as the variance σ^2 is large enough.

These hypothesis are not unacceptable as for example, deformations that are considered in computational anatomy may include rotations which form a subgroup

H of the diffeomorphic deformations which acts isometrically. Concerning the second case, an important example is:

Example 4.2. *Let G be a subgroup of the group of C^∞ diffeomorphisms on $\mathbb{R}^n$ G acts linearly on $L^2(\mathbb{R}^n)$ with the map:*

$$\forall \varphi \in G \quad \forall f \in L^2(\mathbb{R}^n) \qquad \varphi \cdot f = f \circ \varphi^{-1}.$$

Note that this action is not isometric: indeed, $f \circ \varphi^{-1}$ has generally a different L^2-norm than f, because a Jacobian determinant appears in the computation of the integral.

4.3.2 Where did we need an isometric action previously?

Let M be a Hilbert space, and G a group acting on M. Can we define a distance in the quotient space $Q = M/G$ defined as the set which contains all the orbits? When the action is invariant, the orbits are parallel in the sense where $d_M(m,n) = d_M(g \cdot m, g \cdot n)$ for all $m, n \in M$ and for all $g \in G$. This implies that:

$$d_Q([m],[n]) = \inf_{g \in G} \|m - g \cdot n\|,$$

is a distance on Q. But it is not necessarily the case when the action is no longer invariant. Let us take the following example:

Example 4.3. *We call $C^\infty_{diff}(\mathbb{R}^2)$ the set of the C^∞ diffeomorphisms of $M = \mathbb{R}^2$. We equip $\mathbb{R}^2$ with its canonical Euclidean structure. We take $p = (-1,-1)$, $q = (1,1)$ and $r = (2,0)$ and we define a group G by:*

$$G = \left\{ f \in C^\infty_{diff}(\mathbb{R}^2) \mid f(q) = (q),\ f(p) = (p),\ \forall x \in \mathbb{R}\ f(x,0) \in \mathbb{R}r \right\},$$

G acts on $\mathbb{R}^2$ by $f \cdot (x,y) = f(x,y)$. In this case d_Q (define by $d_Q(a,b) = \inf_{g \in G} d_M(a, g \cdot b)$) is not a distance.

Indeed, first let us notice that q and p are fixed points under this group action and the orbit of r is the horizontal line $\{(x,0), x \in \mathbb{R}\}$. On this example:

$$\inf_{g \in G} \|q - g \cdot r\| = \|q - (1,0)\| = 1 \qquad \text{however} \qquad \inf_{g \in G} \|r - g \cdot q\| = \|r - q\| = \sqrt{2},$$

then the function d_Q is not symmetric. One could think define a distance by:

$$\tilde{d}_Q([m],[n]) = \inf_{h,g \in G} \|h \cdot m - g \cdot n\|.$$

Unfortunately, in this case we have:

$$\tilde{d}_Q([p],[q]) = \|p - q\| = 2\sqrt{2} \text{ and } \tilde{d}_Q([p],[r]) = 1 = \tilde{d}_Q([r],[q]),$$

then we do not have $\tilde{d}_Q([p],[q]) \leqslant \tilde{d}_Q([p],[r]) + \tilde{d}_Q([r],[q])$. In other words we do not have the triangular inequality (see figure 4.10).

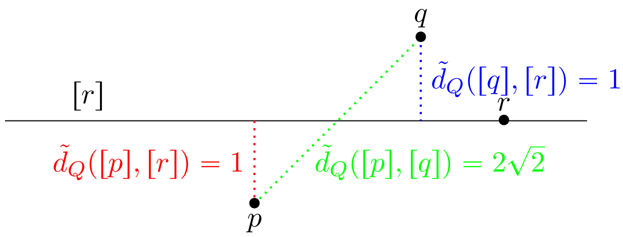


Figure 4.10: Example of three orbits, when $\tilde{d}_Q$ does not satisfy the inequality triangular.

Therefore when the action is no longer invariant, *a priori* one cannot define a distance in the quotient anymore. If Y is a random variable in M, $F(m) = \mathbb{E}(\inf_{g\in G} \|g\cdot m - Y\|^2)$ cannot be interpreted as the variance of $[Y]$.

However $\inf_{g\in G} \|g \cdot a - b\|$ is positive and is equal to zero if $a = b$, then $\inf_{g\in G} \|g \cdot a - b\|$ is a pre-distance in M. Then $\inf_{g\in G} \|g\cdot m - Y\|$ measures the discrepancy between the random point Y and the current point m. Even if the discrepancy measure is not symmetric or does not satisfy the triangular inequality, we can still define $F(x) = \mathbb{E}(\inf_{g\in G} \|g\cdot x - Y\|^2)$ and call it the pre-variance of the projection of Y into M/G, if $\mathbb{E}(\|Y\|^2) < +\infty$. The elements which minimize this function are the elements which orbit are the closest of the random point Y. Hence, we wonder if the template can be estimated by minimizing this pre-variance. Note that, once again $F(x) = F(g\cdot x)$ for all $x \in M$ and $g \in G$. Then the pre-variance is well defined in the quotient space by $[x] \mapsto F(x)$.

It is not surprising to use a discrepancy measure which is not a distance, for instance the Kullback-Leibler divergence [Kullback 1951] is not symmetric although it is commonly used.

In the proof of inconsistency of theorem 4.1, we used that the action was isometric in order to simplify the expansion of the variance in Equation (4.6):

$$
\begin{aligned}
F(m) &= \mathbb{E}\left(\inf_{g\in G} \|m - g\cdot Y\|^2\right) \\
&= \mathbb{E}\left(\inf_{g\in G} \left[\|m\|^2 - \langle m, g\cdot Y\rangle + \|g\cdot Y\|^2\right]\right),
\end{aligned}
$$

with $\|g\cdot Y\|^2 = \|Y\|^2$ there was only one term which depends on g: $\langle g\cdot m, Y\rangle$ and the two other terms could be pulled out of the infimum. When the action is no longer isometric we cannot do this trick anymore. To remedy this situation, in this chapter we require that the orbit of the template is a bounded set.

4.3.3 Non invariant group action, with a subgroup acting isometrically

In this Subsection, G acts on M a Hilbert space. We assume that there exists a subgroup $H \subset G$ such that H acts isometrically on M. As H is included in

G, we deduce a useful link between the variance of Y projected in M/H and the pre-variance of Y projected in M/G:

$$F(m) = \mathbb{E}\left(\inf_{g\in G}\|g\cdot m - Y\|^2\right) \leqslant \mathbb{E}\left(\inf_{h\in H}\|h\cdot m - Y\|^2\right) = F_H(m).$$

The orbit of a point m under the group action G is $[m] = \{g\cdot m, g \in G\}$, whereas the orbit of the point m under the group action H is $[m]_H = \{h\cdot m, h \in H\}$. Moreover, we call F_H the variance of $[Y]_H$ in the quotient space M/H, and F the pre-variance of $[Y]$ in the quotient space M/G.

4.3.3.1 Inconsistency when the template is a fixed point

We begin by assuming that the template t_0 is a fixed point under the action of G:

Proposition 4.5. *Suppose that t_0 is a fixed point under the group action G. Let ε be a standardized noise which support is not included in the fixed points under the group action of H, and $Y = \Phi\cdot t_0 + \sigma\varepsilon = t_0 + \sigma\varepsilon$. Then t_0 is not a minimum of the pre-variance F*

Proof. We have:

1. Thanks to corollary 4.2 of section 4.2.1 we know that $[t_0]_H = [\mathbb{E}(Y)]_H$ is not the Fréchet mean of $[Y]_H$ the projection of Y into M/H: we can find $m \in M$ such that:
$$F_H(m) < F_H(t_0). \tag{4.14}$$
Note that in order to apply corollary 4.2, we do not need that Φ is included in H, because t_0 is a fixed point.

2. Because we take the infimum over more elements we have:
$$F(m) \leqslant F_H(m). \tag{4.15}$$

3. As t_0 is a fixed point under the action of G and under the action of H:
$$F_H(t_0) = F(t_0) = \mathbb{E}(\|t_0 - Y\|^2). \tag{4.16}$$

With Equations (4.14), (4.15) and (4.16), we conclude that t_0 does not minimize F.
□

4.3.3.2 Inconsistency in the general case for the template

The following proposition 4.6 tells us that when σ is large enough then there is an inconsistency.

Proposition 4.6. *We suppose that the template is not a fixed point and that its orbit under the group G is bounded. We consider $A \geqslant \sup_{g\in G} \frac{\|g\cdot t_0\|}{\|t_0\|}$ and $a \leqslant \inf_{g\in G} \frac{\|g\cdot t_0\|}{\|t_0\|}$, note that $a \leqslant 1 \leqslant A$ and we have:*

$$\forall g \in G \quad a\|t_0\| \leqslant \|g \cdot t_0\| \leqslant A\|t_0\|.$$

We note:

$$\theta(t_0) = \frac{1}{\|t_0\|}\mathbb{E}(\sup_{g\in G} \langle g \cdot t_0, \varepsilon\rangle) \text{ and } \theta_H = \frac{1}{\|t_0\|}\mathbb{E}\left(\sup_{h\in H} \langle h \cdot t_0, \varepsilon\rangle\right).$$

We suppose that $\theta_H > 0$. If σ is bigger than a critical noise level noted σ_c defined as:

$$\sigma_c = \frac{\|t_0\|}{\theta_H}\left[\left(\frac{\theta(t_0)}{\theta_H} + A\right) + \sqrt{\left(\frac{\theta(t_0)}{\theta_H} + A\right)^2 + A^2 - a^2}\right]. \tag{4.17}$$

Then we have inconsistency.

Note that in section 4.2.1 we have proved inconsistency in the isometric case as soon as $\sigma > \frac{2\|t_0\|}{K}$, where $K \geqslant \theta_H$, then we find in this theorem an analogical sufficient condition on σ where $\left[\left(\frac{\theta(t_0)}{\theta_H} + A\right) + \sqrt{\left(\frac{\theta(t_0)}{\theta_H} + A\right)^2 + A^2 - a^2}\right]$ is a corrective term due to the non invariant action.

We have shown in chapter 3 that if the orbit of the template $[t_0]_H$ is a manifold, then $\theta_H > 0$ as soon as the support of ε is not included in $T_{t_0}[t_0]^\perp$ (the normal space of the orbit of the template t_0 at the point t_0). If $[t_0]$ is not a manifold, we have also seen in chapter 3 that $\theta_H > 0$ as soon as t_0 is a limit point of $[t_0]_H$ and the support of ε contains a ball $B(0, \eta)$ for $\eta > 0$. Hence, $\theta_H > 0$ is a rather generic condition. Condition (4.17) can be reformulated as follows: as soon as the signal to noise ratio $\frac{\|t_0\|}{\sigma}$ is sufficiently small:

$$\frac{\|t_0\|}{\sigma} < \frac{\theta_H}{\left(\frac{\theta(t_0)}{\theta_H} + A\right) + \sqrt{\left(\frac{\theta(t_0)}{\theta_H} + A\right)^2 + A^2 - a^2}},$$

then there is inconsistency.

We remark the presence of the constants $\theta(t_0)$ and θ_H in proposition 4.6. This kind of constants were already here in the isometric case under the form $\theta(\frac{t_0}{\|t_0\|}) = \frac{1}{\|t_0\|}\mathbb{E}(\sup_{g\in G} \langle t_0, g \cdot \varepsilon\rangle)$, due to the polarization identity (4.2), we can state that it measures how much the template looks like to the noise after registration, but only in the isometric case. However we can intuit that this constant plays a analogical role in the non isometric case.

Example 4.4. *Let G acting on M, we suppose that G contains $H = O(M)$ the orthogonal group of M. Assume that G can modify the norm of the template by* *multiplying its norm by at most 2. Then we can set up $A = 2$ and $a = 0$. By aligning ε and $\|t_0\|$ we have $\theta_H = \mathbb{E}(\|\varepsilon\|) > 0$, and $\theta(t_0) = A\mathbb{E}(\|\varepsilon\|)$ then when the signal to noise ratio $\frac{\|t_0\|}{\sigma}$ is smaller that $\frac{\mathbb{E}(\|\varepsilon\|)}{4+\sqrt{20}}$ then there is inconsistency. By Cauchy-Schwarz inequality we have $\mathbb{E}(\|\varepsilon\|) \leqslant \mathbb{E}(\|\varepsilon\|^2) = 1$, thus the signal to noise ratio has to be rather small in order to fulfill this condition.*

4.3.3.3 Proof of proposition 4.6

We define the following values:

$$\lambda_H = \frac{1}{\|t_0\|^2}\mathbb{E}\left(\sup_{h\in H} \langle h\cdot t_0, Y\rangle\right) \text{ and } \lambda(t_0) = \frac{1}{\|t_0\|^2}\mathbb{E}\left(\sup_{g\in G} \langle g\cdot t_0, Y\rangle\right).$$

Note that λ_H and $\lambda(t_0)$ are registration scores which definitions are the same than the registration score used in the proof of theorem 4.1 in section 4.2 (only the normalization by $\|t_0\|$ is different). The proof of proposition 4.6 is based on the following Lemma:

Lemma 4.1. *If:*

$$\lambda_H \geqslant 0, \tag{4.18}$$

$$a^2 - 2\lambda(t_0) + \lambda_H^2 > 0, \tag{4.19}$$

then t_0 is not a minimizer of the pre-variance of $[Y]$ in M/G.

How condition (4.19) can be understood? In order to answer to that question, let us imagine that $G = H$ acts isometrically, then a can be set up to 1, and $\lambda(t_0) = \lambda_H$ the condition (4.19) becomes $\lambda_H^2 - 2\lambda_H + 1 = (\lambda_H - 1)^2 > 0$ and the condition (3.26) of theorem 3.4 aimed to ensure that $\lambda_H > 1$. Now let us return to the non invariant case: if H is strictly included in G such that a is closed enough to 1 and $\lambda(t_0)$ closed enough to λ_H, then on can think that condition (4.19) still holds. But the *closed enough* seems hard to be quantified.

Proof of lemma 4.1. The proof is based on the following points:

1. $F(\lambda_H t_0) \leqslant F_H(\lambda_H t_0)$,
2. $F_H(\lambda_H t_0) < F(t_0)$.

With items 1 and 2 we get that $F(\lambda_H t_0) < F(t_0)$. Item 1 is just based on the fact that in the map F, we take the infimum on a larger set than on F_H. We now prove item 2, in order to do that we expand the two quantities, firstly:

$$F_H(\lambda_H t_0) = \mathbb{E}\left(\inf_{h\in H} \|h\cdot\lambda_H t_0\|^2 + \|Y\|^2 - 2\langle h\cdot\lambda_H t_0, Y\rangle\right) \tag{4.20}$$

$$= \lambda_H^2\|t_0\|^2 + \mathbb{E}(\|Y\|^2) - 2\lambda_H\mathbb{E}\left(\sup_{h\in H} \langle h\cdot t_0, Y\rangle\right) \tag{4.21}$$

$$= \mathbb{E}(\|Y\|^2) - \lambda_H^2\|t_0\|^2,$$

We use the fact that H acts isometrically between Equations (4.20) and (4.21) and the fact that $\lambda_H \geqslant 0$ because $\inf_{a\in A} -\lambda a = -\lambda \sup_{a\in A} a$ is true for any A subset of $\mathbb{R}$ if $\lambda \geqslant 0$. Secondly:

$$\begin{aligned} F(t_0) &= \mathbb{E}\left(\inf_{g\in G} \|g\cdot t_0\|^2 + \|Y\|^2 - 2\langle g\cdot t_0, Y\rangle\right) \\ &\geqslant a^2\|t_0\|^2 + \mathbb{E}(\|Y\|^2) - 2\mathbb{E}\left(\sup_{g\in G} \langle g\cdot t_0, Y\rangle\right) \\ &\geqslant a^2\|t_0\|^2 + \mathbb{E}(\|Y\|^2) - 2\lambda(t_0)\|t_0\|^2 \end{aligned}$$

Then:

$$F(t_0) - F_H(\lambda_H t_0) \geqslant \|t_0\|^2\left[a^2 - 2\lambda(t_0) + \lambda_H^2\right] > 0,$$

thanks to hypothesis (4.19). □

Proof of proposition 4.6. In order to prove proposition 4.6, all we have to do is proving $\lambda_H \geqslant 0$ and proving that Condition (4.19) is fulfilled when $\sigma > \sigma_c$. Firstly, thanks to Cauchy-Schwarz inequality, we have:

$$\begin{aligned} \lambda_H &= \frac{1}{\|t_0\|^2}\mathbb{E}\left(\sup_{h\in H} \langle h\cdot t_0, \Phi\cdot t_0 + \sigma\varepsilon\rangle\right) \\ &\geqslant \frac{1}{\|t_0\|^2}\left[-A\|t_0\|^2 + \mathbb{E}(\sup_{h\in H} \langle h\cdot t_0, \sigma\varepsilon\rangle)\right] \geqslant -A + \sigma\frac{\theta_H}{\|t_0\|} \end{aligned}$$

Note that as $\sigma > \sigma_c \geqslant A\frac{\|t_0\|}{\theta_H}$ we get $\lambda_H \geqslant 0$, this proves (4.18). We also have:

$$\begin{aligned} \lambda(t_0) &= \frac{1}{\|t_0\|^2}\mathbb{E}\left(\sup_{g\in G} \langle g\cdot t_0, \Phi\cdot t_0 + \sigma\varepsilon\rangle\right) \\ &\leqslant \frac{1}{\|t_0\|^2}\left[A^2\|t_0\|^2 + \sigma\mathbb{E}\left(\sup_{g\in G} \langle g\cdot t_0, \varepsilon\rangle\right)\right] \leqslant A^2 + \sigma\frac{\theta(t_0)}{\|t_0\|}, \end{aligned}$$

Then we can find a lower bound of $a^2 - 2\lambda(t_0) + \lambda_H^2$:

$$\begin{aligned} a^2 - 2\lambda(t_0) + \lambda_H^2 &\geqslant a - 2\left(A^2 + \sigma\frac{\theta(t_0)}{\|t_0\|}\right) + \left(\frac{\sigma\theta_H}{\|t_0\|} - A\right)^2 \\ &\geqslant a^2 - A^2 - 2\frac{\sigma\theta_H}{\|t_0\|}\left(\frac{\theta(t_0)}{\theta_H} + A\right) + \left(\frac{\sigma\theta_H}{\|t_0\|}\right)^2 := P(\sigma) \end{aligned}$$

For $\sigma > \sigma_c$ where σ_c is the biggest solution of the quadratic Equation $P(\sigma) = 0$, we get $a^2 - 2\lambda(t_0) + \lambda_H^2 > 0$ and the template estimation is inconsistent thanks to lemma 4.1. The critical σ_c is exactly the one given by proposition 4.6. □

4.3.4 Linear action

The result of the previous part has a drawback, it requires that the group of deformations contains a non trivial subgroup which acts isometrically. Now, we remove this hypothesis, but we require that the group acts linearly on data. This means that for all $g \in G$, $m \mapsto g\cdot m$ is a linear map.

4.3.4.1 Inconsistency

In this Subsection we suppose that the group G acts linearly on M. Once again, we can give a criteria on the noise level which leads to inconsistency:

Proposition 4.7. *We suppose that the orbit of the template is bounded, therefore we can consider the following two constants:*

$$\exists a \geqslant 0, A > 0 \text{ such that } \forall g \in G \quad a\|t_0\| \leqslant \|g \cdot t_0\| \leqslant A\|t_0\|.$$

We suppose that $A < \sqrt{2}$. In other words, the deformation of the template can multiply the norm of the template by less than $\sqrt{2}$. We also suppose that:

$$\theta(t_0) = \frac{1}{\|t_0\|}\mathbb{E}\left(\sup_{g \in G} \langle g \cdot t_0, \varepsilon\rangle\right) > 0. \tag{4.22}$$

There is inconsistency as soon as

$$\sigma \geqslant \sigma_c = \frac{\|t_0\|}{\theta(t_0)}\left[A^2 + \frac{1+\sqrt{1-a^2(2-A^2)}}{2-A^2}\right].$$

Example 4.5. *For instance if $A \leqslant 1.2$, then there is inconsistency if $\sigma \geqslant 7\frac{\|t_0\|}{\theta(t_0)}$.*

Once again we find a condition which is similar to the isometric case, but due to the non invariant action we have here a corrective term which depends on A and a. In chapter 3, we have seen lemma 3.7 which states that $\theta(t_0) > 0$. However G does not act isometrically, therefore we can no longer apply lemma 3.7 in order to fulfill Condition (4.22). However it is easy to fulfill this Condition thanks to the following Proposition:

Proposition 4.8. *If t_0 is not a fixed point, and if the support of ε contains a ball $B(0, \eta)$ for $\eta > 0$ then*

$$\theta(t_0) = \frac{1}{\|t_0\|}\mathbb{E}\left(\sup_{g \in G} \langle g \cdot t_0, \varepsilon\rangle\right) > 0.$$

Remark 4.3. *It is possible to remove the condition $A < \sqrt{2}$ in proposition 4.7. Indeed Let be $h \in G$ such that:*

$$\frac{\sup_{g \in G} \|g \cdot t_0\|}{\|h \cdot t_0\|} < \sqrt{2}.$$

The template t_0 can be replaced by $h \cdot t_0$ since $\Phi \cdot t_0 + \sigma\varepsilon$ is equal to $\Phi h^{-1} \cdot ht_0$ and applying proposition 4.7 to the new template $h \cdot t_0$. We get that $h \cdot t_0$ does not minimize the variance F with $A \leqslant \sqrt{2}$ (because the new template is $h \cdot t_0$). Since $h \cdot t_0$ does not minimize F, the original template t_0 does not minimize the pre-variance F neither, since $F(t_0) = F(h \cdot t_0)$.

This changes the critical σ_c since we apply proposition 4.7 to $h \cdot t_0$ instead of t_0 itself.

4.3.4.2 Proofs of proposition 4.7 and proposition 4.8

As in section 4.3.3 we first prove a Lemma:

Lemma 4.2. *We define:*

$$\lambda(t_0) = \frac{1}{\|t_0\|^2}\mathbb{E}\left(\sup_{g\in G} \langle g\cdot t_0, Y\rangle\right).$$

Suppose that $\lambda(t_0) \geqslant 0$ *and that:*

$$a^2 - 2\lambda(t_0) + \lambda(t_0)^2(2 - A^2) > 0. \tag{4.23}$$

Then t_0 *is not a minimum of* F.

Proof of lemma 4.2. Since

$$\forall g \in G \quad a\|t_0\| \leqslant \|g\cdot t_0\| \leqslant A\|t_0\|, \tag{4.24}$$

then by linearity of the action we get:

$$\forall g \in G,\ \mu\in\mathbb{R} \quad a\|\mu t_0\| \leqslant \|g\cdot \mu t_0\| \leqslant A\|\mu t_0\|. \tag{4.25}$$

We remind that:

$$F(m) = \mathbb{E}\left(\inf_{g\in G}\|g\cdot m\|^2 - 2\langle g\cdot m, Y\rangle + \|Y\|^2\right).$$

By using Equations (4.24) and (4.25) we get:

$$F(t_0) \geqslant a^2\|t_0\|^2 - 2\lambda(t_0)\|t_0\|^2 + \mathbb{E}(\|Y\|^2),$$

We get:

$$\begin{aligned} F(\lambda(t_0)t_0) &\leqslant \mathbb{E}\left(A^2\|\lambda(t_0)t_0\|^2 + \|Y\|^2 + \inf_{g\in G}\left(-2\lambda(t_0)\langle g\cdot t_0, Y\rangle\right)\right) \\ &\leqslant A^2\lambda(t_0)^2\|t_0\|^2 + \mathbb{E}(\|Y\|^2) - 2\lambda(t_0)^2\|t_0\|^2. \end{aligned} \tag{4.26}$$

Note that we use the fact that the action is linear in Equation (4.26). We obtain that t_0 is not the minimum of the F:

$$F(t_0) - F(\lambda(t_0)t_0) \geqslant \|t_0\|^2\left[a^2 - 2\lambda(t_0) + \lambda(t_0)^2(2 - A^2)\right] > 0.$$

□

Proof of proposition 4.7. By solving the following quadratic inequality we remark that:

$$a^2 - 2\lambda(t_0) + (2 - A^2)\lambda(t_0)^2 > 0 \text{ if } \lambda(t_0) > \frac{1 + \sqrt{1 - a^2(2 - A^2)}}{2 - A^2},$$

Besides, as in section 4.3.3.2 we can take a lower bound of $\lambda(t_0)$ by decomposing $Y = \Phi \cdot t_0 + \sigma\varepsilon$ and applying Cauchy-Schwarz inequality $\langle \Phi \cdot t_0, g \cdot t_0 \rangle \geqslant -A^2 \|t_0\|^2$, we get:

$$\lambda(t_0) \geqslant -A^2 + \frac{\sigma}{\|t_0\|}\theta(t_0). \tag{4.27}$$

Thanks to Condition (4.27) and the fact that $\sigma > \sigma_c$ we get:

$$\lambda(t_0) \geqslant -A^2 + \frac{\sigma}{\|t_0\|}\theta(t_0) > \frac{1+\sqrt{1-a^2(2-A^2)}}{2-A^2}$$

Then $\lambda(t_0) \geqslant 0$ and Condition (4.23) is fulfilled. Thus, there is inconsistency, according to lemma 4.2. □

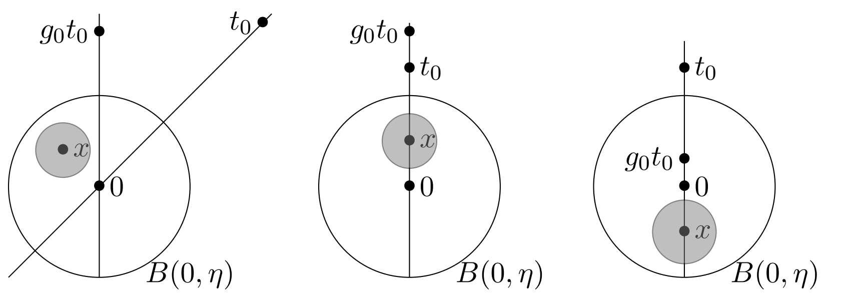
Proof of proposition 4.8. First we notice that:

$$\|t_0\|\theta(t_0) = \mathbb{E}\left(\sup_{g\in G} \langle g \cdot t_0, \varepsilon\rangle\right) \geqslant \mathbb{E}(\langle t_0, \varepsilon\rangle) = \langle t_0, \mathbb{E}(\varepsilon)\rangle = 0. \tag{4.28}$$

In order to have $\theta(t_0) > 0$, first we show that there exists $x \in B(0,\eta)$ and $g_0 \in G$ such that

$$\sup_{g\in G} \langle g \cdot t_0, x\rangle \geqslant \langle g_0 \cdot t_0, x\rangle > \langle t_0, x\rangle .$$

Let $g_0 \in G$ such that $g_0 \cdot t_0 \neq t_0$. There are three cases to be distinguished (see figure 4.11:



(a) Case 1: t_0 and $g \cdot t_0$ are linearly independent.

(b) Case 2: $g \cdot t_0$ is proportional to t_0 with a factor > 1.

(c) Case 3: $g \cdot t_0$ is proportional to t_0 with a factor < 1.

Figure 4.11: Representation of the three cases, on each we can find an x in the support of the noise such as $\langle x, g_0 \cdot t_0\rangle > \langle x, t_0\rangle$ and by continuity of the dot product $\langle \varepsilon, g_0 \cdot t_0\rangle > \langle \varepsilon, t_0\rangle$ with is an event with a non zero probability, (for instance the ball in gray). This is enough in order to show that $\theta(t_0) > 0$.

1. The vectors $g_0 \cdot t_0$ and t_0 are linearly independent. In this case $t_0^\perp \not\subset (g_0 \cdot t_0)^\perp$, then we can find $x \in t_0^\perp$ and $x \notin (g \cdot t_0)^\perp$. Then $\langle t_0, x\rangle = 0$ and $\langle g \cdot t_0, x\rangle \neq 0$,

without loss of generality we can assume that $\langle g \cdot t_0, x\rangle > 0$ (replacing x by $-x$ if necessary). We also can assume that $x \in B(0, \eta)$ (replacing x by $\frac{x\eta}{2\|x\|}$ if necessary). Then we have $x \in B(0, \eta)$ and:

$$\langle g_0 \cdot t_0, x\rangle > 0 = \langle t_0, x\rangle .$$

2. If $g_0 \cdot t_0 = wt_0$ with $w > 1$, we take $x = \frac{\eta}{2\|t_0\|} t_0 \in B(0, \eta)$ and we have:

$$\langle g \cdot t_0, x\rangle = w\frac{\eta}{2}\|t_0\| > \frac{\eta}{2}\|t_0\| = \langle t_0, x\rangle .$$

3. If $g_0 \cdot t_0 = wt_0$ with $w < 1$ we take $x = -\frac{\eta}{2\|t_0\|} t_0 \in B(0, \eta)$ and we have:

$$\langle g_0 \cdot t_0, x\rangle = -w\frac{\eta}{2}\|t_0\| > -\frac{\eta}{2}\|t_0\| = \langle t_0, x\rangle .$$

In all these cases, we can find x such that $\langle g_0 \cdot t_0, x\rangle > \langle t_0, x\rangle$ By continuity there exists $r > 0$ such that for all y on this ball we have $\langle g \cdot t_0, y\rangle > \langle t_0, y\rangle$. Then the event $\{\sup_{g\in G} \langle g \cdot t_0, \varepsilon\rangle > \langle t_0, \varepsilon\rangle\}$ has non zero probability, since x is in the support of ε we have $\mathbb{P}(\varepsilon \in B(x, r)) > 0$. Thus Inequality in (4.28) will be strict. This proves that $\theta(t_0) > 0$. □

4.3.5 Example of a template estimation which is consistent

In order to underline the importance of the hypotheses, we give an example where the method is consistent:

Example 4.6 (affine action). *Let M be a Hilbert space and V a closed sub-linear space of M (see figure 4.12). Then $G = V$ acts on M by:*

$$(v, m) \in G \times M \mapsto m + v.$$

This action is not isometric, indeed $m \mapsto m + v$ is not linear (except if $v = 0$). However the distance is invariant under this group action ($\|v \cdot m - v \cdot n\| = \|m - v\|$), let us consider $V^\perp$ the orthogonal space of V. The variance in the quotient space is:

$$F(m) = \mathbb{E}\left(\inf_{v\in V} \|m + v - Y\|^2\right) = \mathbb{E}(\|p(m) - p(Y)\|^2) = \mathbb{E}(\|p(m) - p(t_0) + \varepsilon\|^2),$$

where $p : M \to V^\perp$ the orthogonal projection on $V^\perp$. Then it is clear that t_0 minimizes F.

The map $[x] \mapsto p(x)$ will be called an congruent section in section 5.4.1. Hence, is there a contradiction with proposition 4.6 or proposition 4.7 which prove inconsistency as soon as the noise level is large enough? In proposition 4.6, we require that there is a subgroup acting isometrically, in this example the only element which acts linearly is the identity element $m \mapsto m + 0$, then $H = \{0\}$ is the only possibility, however the support of the noise should not be included in the set of fixed point under the group action of H. Here, all points are fixed under H, hence it is not possible to fulfill this condition. Example 4.6 is not a contradiction with proposition 4.6, it is also not a contradiction with proposition 4.7 since it does not act linearly on data.

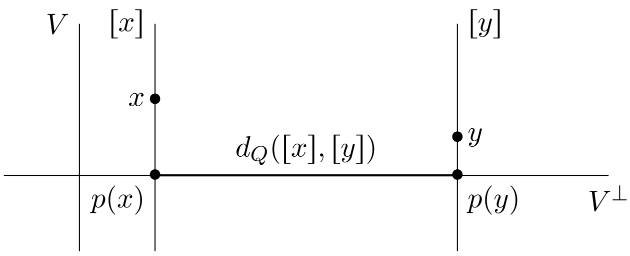


Figure 4.12: In the case of vertical translation by vectors of V, the orbits are affine subspaces parallel to V. The distance between two orbits $[x]$ and $[y]$ is given by the distance between the orthogonal projection of x and y in $V^\perp$. This is an example where the template estimation is consistent.

4.3.6 Inconsistency with non invariant action and regularization

In practice people add a regularization term in the function they minimize in LDDMM [Beg 2005, Durrleman 2014], or in Demons [Lombaert 2013] etc. Because, if one considers two points, one does not want necessarily to fit one with the other. Indeed, even if one deformation matches exactly these two points, it may be an unrealistic deformation. So far, we did not study the use of such a term in the inconsistency.

4.3.6.1 Case of deformations close to the identity element of G

If we suppose that the deformations Φ of the template is closed to identity, it is useless to take the infimum over G because G contains big deformations. Perhaps one of these big deformations can reache the infimum in F, but this element is not the one which deforme the template in the generative model. Then such big deformations should not be taken into account. That is why, if we suppose that G can be equipped with a distance d_G, then we can assume that there exists $r > 0$ such that the deformation Φ belongs almost surely to

$$\mathcal{B} = B(e,r) = \{g \in G, \quad d_G(e,g) < r\}.$$

Instead of defining $F(m)$ as $\mathbb{E}(\inf_{g\in G} \|g \cdot m - Y\|^2)$, one can define $F(m) = \mathbb{E}(\inf_{g\in\mathcal{B}} \|g \cdot m - Y\|^2)$, and the previous proofs will still be true, when replacing for instance $\lambda(t_0)$ by $\lambda(t_0) = \frac{1}{\|t_0\|^2}\mathbb{E}(\sup_{g\in\mathcal{B}} \langle g \cdot t_0, Y\rangle)$ etc. Likewise we need to replace the hypothesis "the support of ε is not included in the set of fixed points" by "the support of ε in not included is the set of fixed points under the action restricted to $\mathcal{B}$.

Note that restraining ourselves to $\mathcal{B}$ is equivalent to add a following regularization on the function F:

$$F(m) = \mathbb{E}\left(\inf_{g\in G} \|g \cdot m - Y\|^2 + Reg(g)\right) \text{ with } Reg(g) = \begin{cases} 0 & \text{if} \quad g \in \mathcal{B} \\ +\infty & \text{if} \quad g \notin \mathcal{B} \end{cases}.$$

Moreover considering only the elements in $\mathcal{B}$ will automatically satisfy the condition $A < \sqrt{2}$ in proposition 4.7 as long as the group G acts continuously on the template, if r is small enough.

4.3.6.2 Inconsistency in the case of a group acting linearly with a bounded regularization

In this section we suppose that the group G acts linearly. We also suppose that $A < \sqrt{2}$. The regularization term is a bounded map $Reg : G \to [0, \Omega]$. With this framework, we still able to prove that there is inconsistency as soon as the noise level is large enough:

Proposition 4.9. *Let G be a group acting linearly on M. We suppose that the orbit of the template t_0 is bounded with $A = \sup_{g\in G} \frac{\|g\cdot t_0\|}{\|t_0\|} < \sqrt{2}$, the generative model is still $Y = \Phi \cdot t_0 + \sigma\varepsilon$. We define the pre-variance as:*

$$F(m) = \mathbb{E}\left(\inf_{g\in G}\left(\|Y - g\cdot m\|^2 + Reg(g)\right)\right).$$

Then as soon as the noise level is large enough, i.e.:

$$\sigma > \sigma_c = \frac{\|t_0\|}{\theta(t_0)}\left[A^2 + \frac{1 + \sqrt{1 - (a^2 + \frac{\Omega}{\|t_0\|^2})(2 - A^2)}}{2 - A^2}\right].$$

Then t_0 is not a minimizser of F.

The proof is exactly the same as the Proof of proposition 4.7, we take 0 as a lower bound of the the regularization term in the lower bound of $F(t_0)$, and we take Ω as a upper bound of the regularization term in the upper bound of $F(\lambda(t_0)t_0)$. We solve a similar quadratic equation in order to find the critical σ.

4.4 Conclusion and discussion

We provided an asymptotic behavior of the consistency bias when the noise level σ tends to infinity in the case of isometric action. As a consequence, the inconsistency can not be neglected when σ is large. When the action is no longer isometric, inconsistency has been also shown when the noise level is large.

However we have not answered this question: can the inconsistency be neglected? When the noise level is small enough, then the consistency bias is small [Miolane 2017] or chapter 3, hence it can be neglected. Note that the quotient space is not a manifold, this prevents us to use *a priori* the Central Limit theorem for manifold proved in [Bhattacharya 2008]. But if the Central Limit theorem could be applied to quotient space, the fluctuations induces an error which would be approximately equal to $\frac{\sigma}{\sqrt{I}}$ and if $K \ll \frac{1}{\sqrt{I}}$, then the inconsistency could be neglected because it is small compared to fluctuation. One way to avoid the inconsistency is to use another framework, for a instance a Bayesian paradigm [Cheng 2016].

In the numerical experiments we presented, we have seen that the estimated template is more crispy that the true template. The intuition is that the estimated template in computational anatomy with a group of diffeomorphisms is also more detailed. But the true template is almost always unknown. It is then possible that one think that the computation of the template succeeded to capture small details of the template while it is just an artifact due to the inconsistency. Moreover in order to tackle this question, one needs to have a good modelisation of the noise, for instance in [Kurtek 2011b], the observations are curves, what is a relevant noise in the space of curves?

In this chapter, we have considered actions which do not let the distance invariant. Although we have only shown the inconsistency as soon as the noise level is large enough, the inequality used where not optimal at all, surely future works could improve this work and prove that inconsistency appears for small noise level. Moreover a quantification of the inconsistency should be established.

Chapter 5

Study of consistency with a backward estimation

Contents

5.1 Introduction

Let us make a brief overview of what we did so far. In chapter 3, we established inconsistency for isometric action in Hilbert space. In chapter 4, we provided an asymptotic behaviour of the consistency bias when the noise level goes to infinity also for isometric action in Hilbert space. Besides, we extended the study of the consistency for non isometric action in Hilbert space. As a result, we saw that the inconsistency also appears when the noise level was large enough. Therefore, our previous results have two flaws. The first one is that those results say nothing in the case of small noise level for non isometric action. The second one is that our results are restricted to ambient spaces which are assumed to be Hilbert spaces. But ambient spaces are not always Hilbert spaces. This chapter gathers some results which complete our previous study, to solve the two flaws mentioned below. In this chapter, we consider only backward estimation.

In section 5.2, we concentrate on compact and continuous group action. We find an implicit expression of an element which minimizes the variance/pre-variance in the quotient space which is used this implicit expression in order to prove inconsistency in the Hilbert space in the case of a invariant distance under the group action. Moreover this work can be generalized to Riemannian manifolds contrary to most of our previous works.

Section 5.3 is a conjecture to generalize the presence of inconsistency when data belong to ambient space which can be metric spaces (including infinite dimensional manifold) restricted to compact group action.

In section 5.4, we discuss the existence of a section of the quotient which satisfies geometrical properties. The first property is that the section is a congruent map between the quotient and the ambient space. When such a section exists, the quotient space can be embedded into the ambient space. The section is said to be congruent, when the distance is preserved though the section. Moreover we can find an explicit equation of the Fréchet mean in the quotient space. However, we advocate that the congruent section does not exists often. We give some examples and we provide an algorithm which is able to dismiss the existence of a congruent section. The second property is that the section is a measurable map. The existence of a measurable map which registers elements with respect to a point is a cornerstone in order to complete the proofs of Theorems seen in section 5.2.

In section 5.5 we prove that the template estimation can be inconsistent even for small σ for non isometric action which are a perturbation of an isometric action.

5.2 Implicit equation of an element which minimizes the variance/pre-variance and proofs of inconsistency

In this section, we consider a compact group. We also assume that the group acts continuously on our ambient space. This implies that the registration between two points has always a solution (non necessarily unique). We exhibit an implicit equation of an element which minimizes the variance/pre-variance in the quotient space. We use this expression to show inconsistency for isometric action in Hilbert space, but also in more general spaces as complete Riemannian manifolds.

One difficulty in all this thesis is that we do not have an explicit equation of the Fréchet mean in the quotient space. Therefore, proving the consistency or the inconsistency cannot be reduced to show that the template is a solution of this possible explicit equation. In order to overcome this difficulty, we have worked with inequalities in chapter 4 and we have studied the variance restricted to an half-line in chapter 3 (theorem 3.4). But these two strategies were based on avoiding the difficulty. On the contrary in theorem 3.1, we have found the gradient of the variance in the quotient space. Solving this gradient equal to zero gives an implicit equation of the Fréchet mean in the quotient space. In this section, we aim to do the same: we want to have an implicit (or even better an explicit) equation of the elements which minimize the variance in the quotient space.

In the following, we will use the theoretical advantages of the backward estimation method: minimizing $F(m) = \mathbb{E}\left(\inf_{g\in G}\|m - g\cdot Y\|^2)\right)$. This estimation differs from the forward estimation method namely minimizing $F(m) = \mathbb{E}\left(\inf_{g\in G}\|g\cdot m - Y\|^2\right)$, used in section 4.3. However, these two estimations are the same in the case of a invariant distance under the group action. Note that both estimation methods are used in practice, see [Glaunes 2006, Joshi 2004, Du 2014, Glasbey 2001] for instance.

5.2.1 Implicit equation of an element which minimizes the variance/pre-variance in quotient spaces

Let M be our ambient space; we just assume that M is a metric space, with d_M its metric. Let G be a group acting on M and Y be a variable in M. We define the pre-variance of Y in the quotient space as:

$$F(m) = \mathbb{E}\left(\inf_{g\in G} d_M(m, g\cdot Y)^2\right).$$

If d_M is invariant under the group action, then we say that F is the variance in the quotient space For the moment, we make no assumption of the random variable Y, except that $\mathbb{E}(d_M(m,Y)^2) < +\infty$ for at least one m in M. Therefore F is well defined, thanks to the triangular inequality.

We assume the existence of a minimum of the pre-variance (otherwise the template estimation is necessarily inconsistent). We denote by $m_\star$ one minimizer of the pre-variance F. We assume that the group G is compact and that the action is continuous (namely $g \mapsto g \cdot x$ is a continuous map for all $x \in M$). Therefore the registration of Y with respect to $m_\star$ does exist. Then it exists $g(Y, m_\star) \in G$ such that:

$$F(m_\star) = \mathbb{E}(d_M^2(m_\star, g(Y, m_\star) \cdot Y)).$$

We define by E the variance of $g(Y, m_\star) \cdot Y$:

$$E(m) = \mathbb{E}(d_M^2(m, g(Y, m_\star) \cdot Y)).$$

For Z a random variable in M, we define $\mathrm{FM}(Z)$ as the set of all Fréchet means of Z:

$$\mathrm{FM}(Z) = \underset{m \in M}{\operatorname{argmin}}\ \mathbb{E}(d_M^2(m, Z)).$$

Proposition 5.1. *Let (M, d_M) be a metric space, G a compact group acting continuously on M. We assume that the variance/pre-variance, defined by*

$$F(m) = \mathbb{E}\left(\inf_{g \in G} d_M(m, g \cdot Y)^2\right) \text{ for } m \in M,$$

reaches its minimum a point $m_\star$. We assume that there exits $y \mapsto g(y, m_\star) \cdot y$ a deterministic measurable function such that:

$$\forall y \in M \quad g(y, m_\star) \in \underset{g \in G}{\operatorname{argmin}}\ d_M(m_\star, g \cdot y).$$

Then, the set $FM(g(Y, m_\star) \cdot Y)$ is non empty, besides we have:

$$m_\star \in FM(g(Y, m_\star) \cdot Y) \tag{5.1}$$

Remark 5.1. *Note that we do not need the distance to be invariant under the group action. However it requires a backward estimation. It requires also a compact group contrarily to previous results as theorem 3.4 for instance.*
When the ambient space M is a Hilbert space and Z a squarred integrable variable, $FM(Z)$ exists, and is unique and equals $\mathbb{E}(Z)$ (proposition 1.4), in this case we can conclude that:

$$m_\star = \mathbb{E}(g(Y, m_\star) \cdot Y). \tag{5.2}$$

Note that equation (5.2) was already obtained in the case of the finite and isometric group action (see equation (3.8)). However, this was done by differentiating the variance, which is not possible everywhere as we have already see it (section 3.4.7).
In the case of a finite sample $Y_1, \dots, Y_n$, we define the empirical pre-variance by:

$$F_n(m) = \frac{1}{n} \sum_{i=1}^{n} \inf_{g \in G} \|m - g \cdot Y_i\|^2$$

if $m_\star$ is a minimizer of the empirical pre-variance F_n, we can also deduce that:

$$m_\star = \frac{1}{n} \sum_{i=1}^{n} g(Y, m_\star) \cdot Y$$

Proof of proposition 5.1. We have two things to prove: $\mathrm{FM}(g(Y,m_\star)\cdot Y)$ is not empty as a first step and $m_\star \in \mathrm{FM}(g(Y,m_\star)\cdot Y)$ as a second step.

- First, let us prove that the set $\mathrm{FM}(g(Y,m_\star)\cdot Y)$ is not empty. If $\mathrm{FM}(g(Y,m_\star)\cdot Y)$ was empty, then in particular, $m_\star$ would not minimize E the variance of $g(Y,m_\star)\cdot Y$. Therefore, we can find $z \in M$ such that $E(z) < E(m_\star)$ in this case we have:

$$\begin{aligned} F(m_\star) &= E(m_\star) \\ &> E(z) \\ &> \mathbb{E}\left(\inf_{g\in G} d_M^2(z, g\cdot Y)\right) \\ &> F(z), \end{aligned}$$

  which is a contradiction, since $m_\star$ minimizes F.

- Secondly, let us take $z \in \mathrm{FM}(g(Y,m_\star)\cdot Y)$, we get:

$$\begin{aligned} F(m_\star) &= \mathbb{E}\left(\inf_{g\in G} d_M^2(g\cdot m, Y)\right) \\ &= E(m_\star) \\ &\geqslant E(z) = \mathbb{E}(d_M^2(z, g(Y,m_\star)\cdot Y)) \\ &\geqslant \mathbb{E}\left(\inf_{g\in G} d_M^2(z, g\cdot Y)\right) \\ &\geqslant F(z) \\ &\geqslant F(m_\star), \end{aligned}$$

  Then we conclude that $E(m_\star) = E(z) = \min E$, then:

$$m_\star \in \mathrm{FM}(g(Y,m_\star)\cdot Y).$$

□

It may seem odd, that we get an implicit equation of $m_\star$ which depends on the choice we made on the element $g(Y,m_\star)$ of G which registers Y with respect to $m_\star$. Indeed, there is *a priori* no uniqueness of the element which register Y to $m_\star$. What if there are such two elements $g(Y,m_\star)$ and $\tilde{g}(Y,m_\star)$? The following proposition prove that $g(Y,m_\star)\cdot Y = \tilde{g}(Y,m_\star)\cdot Y$ almost surely. In other words, if there is no uniqueness in the registration of Y to m, then the choice of the element $g(Y,m_\star)$, has no consequence:

Proposition 5.2. *Let G be a compact group acting continuously on M a Hilbert space, let Y be a random variable, and $m_\star$ an element which minimizes the pre-variance (defined in section 5.2.1). We assume that there are two random variables $g(Y, m_\star)$ and $\tilde{g}(Y, m_\star)$ which minimize $d_M(m_\star, g \cdot Y)$. In this case we have almost surely:*

$$g(Y, m_\star) \cdot Y = \tilde{g}(Y, m_\star) \cdot Y$$

Proof. Applying equation (5.2) to $g(Y, m_\star)$ and $\tilde{g}(Y, m_\star)$ leads to:

$$m_\star = \mathbb{E}(g(Y, m_\star) \cdot Y) = \mathbb{E}(\tilde{g}(Y, m_\star) \cdot Y).$$

Now we mix g and $\tilde{g}$: let us choose A a measurable set, and define the following random variable:

$$Z = g(Y, m_\star) \cdot Y \mathbb{1}_{Y \in A} + \tilde{g}(Y, m_\star) \cdot Y \mathbb{1}_{Y \notin A}.$$

For every Y, Z is an element which reach the minimum in $d_M(m_\star, y)$ for $y \in [Y]$. Once again, we can apply equation (5.2) to Z, we get $m_\star = \mathbb{E}(Z)$, we can deduce that:

$$m_\star = \mathbb{E}\left(g(Y, m_\star) \cdot Y \mathbb{1}_{Y \in A} + \tilde{g}(Y, m_\star) \cdot Y \mathbb{1}_{Y \notin A}\right) = \mathbb{E}\left(\tilde{g}(Y, m_\star) \cdot Y\right).$$

By splitting the second expectation into two parts and simplifying, we have:

$$\mathbb{E}(g(Y, m_\star) \cdot Y \mathbb{1}_{Y \in A}) = \mathbb{E}(\tilde{g}(Y, m_\star) \cdot Y \mathbb{1}_{Y \in A}),$$

and we do it for any measurable set A, which yields

$$g(Y, m_\star) \cdot Y = \tilde{g}(Y, m_\star) \cdot Y, \ Y\text{-almost surely.}$$

□

We can generalize the uniqueness of the registration variable $g(Y, m_\star) \cdot Y$ in the case of complete Riemannian manifolds:

Proposition 5.3. *Let G be a compact group acting continuously on M a complete Riemannian manifold, let Y be a random variable and $m_\star$ an element which minimizes the pre-variance. We assume that*

$$g(Y, m_\star) \in \underset{g \in G}{\operatorname{argmin}}\, d_M(m_\star, g \cdot Y) \text{ and } \tilde{g}(Y, m_\star) \in \underset{g \in G}{\operatorname{argmin}}\, d_M(m_\star, g \cdot Y),$$

for $g(Y, m_\star)$ and $\tilde{g}(Y, m_\star)$ two random variables in G. We assume that

$$\mathbb{P}(g(Y, m_\star) \cdot Y \in C(m_\star)) = \mathbb{P}(\tilde{g}(Y, m_\star) \cdot Y \in C(m_\star)) = 0.$$

Where $C(m_\star)$ is the cut locus of $m_\star$. In this case we have:

$$g(Y, m_\star) \cdot Y = \tilde{g}(Y, m_\star) \cdot Y \quad \text{almost surely.}$$

Proof. Applying equation (5.1) to $g(Y, m_\star)$ and $\tilde{g}(Y, m_\star)$ leads to:

$$m_\star = \mathrm{FM}(g(Y, m_\star) \cdot Y) = \mathrm{FM}(\tilde{g}(Y, m_\star) \cdot Y).$$

Now we can mix g and $\tilde{g}$: let us choose A a measurable set, and define:

$$Z = g(Y, m_\star) \cdot Y \mathbb{1}_{Y \in A} + \tilde{g}(Y, m_\star) \cdot Y \mathbb{1}_{Y \notin A},$$

The random variable Z reaches the infimum which is in the definition of $F(m_\star)$. Then from equation (5.1), we get $m_\star = \mathrm{FM}(Z)$. By differentiation of the variance at $m_\star$ (because of the probability to fall in the cut locus is zero), we have:

$$\mathbb{E}(\mathrm{Log}_{m_\star} g(Y, m_\star) \cdot Y \mathbb{1}_{Y \in A}) = \mathbb{E}(\mathrm{Log}_{m_\star} \tilde{g}(Y, m_\star) \cdot Y \mathbb{1}_{Y \in A}),$$

and we do it for any measurable sets which yields

$$\mathrm{Log}_{m_\star} g(Y, m_\star) \cdot Y = \mathrm{Log}_{m_\star} \tilde{g}(Y, m_\star) \cdot Y, \quad Y\text{-almost surely.}$$

This proves that $g(Y, m_\star) \cdot Y = g(Y, m_\star) \cdot Y$, Y-almost surely. □

5.2.2 Interpretation of equation (5.2) in Hilbert spaces

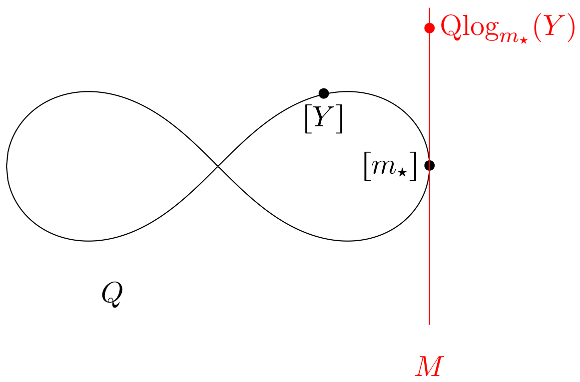


Figure 5.1: Representation of the quotient as if it was a manifold, with M the tangent plane. If $[m_\star]$ is a Fréchet mean of $[Y]$ then $\mathbb{E}(\mathrm{Qlog}_{m_\star}(Y)) = 0$, where Qlog is a quotient logarithm function.

In the previous subsection, we have found an implicit equation of the elements which minimize the pre-variance in the quotient spaces. Can we explain this equation? First, we suppose that the ambient space is an Hilbert space. We can rewrite equation (5.2) as follows:

$$\mathbb{E}(g(Y, m_\star) \cdot Y - m_\star) = 0.$$

Then, if we define a *quotient logarithm function* as:

$$\mathrm{Qlog}_{m_\star}(y) = g(y, m_\star) \cdot y - m_\star.$$

We have the equation $\mathbb{E}(\mathrm{Qlog}_{m_\star}(Y)) = 0$ in M. Then even if Q is not a manifold and M is not the tangent plane of Q at the point $m_\star$, equation (5.2) can be reinterpreted as the equation of a exponential barycenter: $\mathbb{E}(\mathrm{Qlog}_{m_\star}(Y)) = 0$ when $m_\star$ is one element which minimizes the pre-variance (see figure 5.1). This equation looks like the one find for the Fréchet mean in a manifold, when Fréchet mean are, in particular, exponential barycenter (see definition 1.5).

5.2.3 Interpretation of equation (5.1) in complete Riemannian manifolds

Now, we want to find also an explanation of equation (5.1) when the ambient space is a complete Riemannian manifold. Contrarily to Hilbert space, there is no easy expression of $\mathrm{FM}(Z)$ for Z a random variable in M when M is a manifold. This may complicate our analysis of equation (5.1). However we can adapt our previous interpretation. We call Qlog of Y the element $g(Y, m_\star) \cdot Y$, then we have $m_\star = \mathrm{FM}(\mathrm{Qlog}_{m_\star}(Y)$. Therefore $m_\star$ is a critical point of the variance of $\mathrm{Qlog}_{m_\star}(Y)$. By differentiating the variance of $\mathrm{Qlog}_{m_\star}(Y)$ (defined as $m \mapsto \mathbb{E}(d^2(m, \mathrm{Qlog}_{m_\star}(Y)))$ (see lemma 3.9) we have that:

$$\mathbb{E}\left(\mathrm{Log}_{m\star}(\mathrm{Qlog}_{m_\star}(Y))\right) = 0.$$

In other words, $m_\star$ is a *double exponential barycenter*, since it is the expectation of the log of the log which is equal to zero (see figure 5.2).

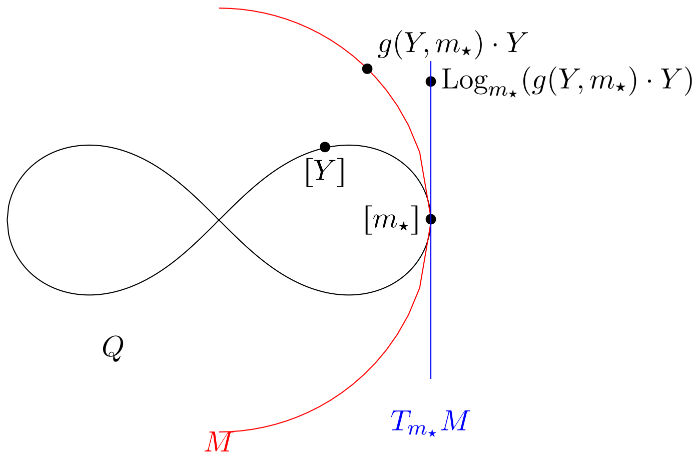


Figure 5.2: Representation of the quotient, M acts as an intermediary between the quotient space and the tangent plan at $m_\star$. If $m_\star$ is a minimizer of the pre-variance in the quotient space, then $\mathbb{E}(\mathrm{Log}_{m_\star}(\mathrm{Qlog}_{m_\star})Y) = 0$. Where $\mathrm{Qlog}_{m_\star}(Y)$ is the registration of Y to $m_\star$ and $\mathrm{Log}_{m_\star}$ is the Riemannian logarithm map.

Note that this interpretation has been done without differentiating the pre-variance in the quotient space. Because, we have already seen that it may not be possible. Moreover $g(Y, m_\star)$ may be not unique, this is an analogy with the cut locus issue in manifolds. In manifold when points are in the cut locus, the differentiability fails because, for instance, of the non uniqueness of the geodesics. However, the non uniqueness in this case, is not really an issue, indeed if $g(Y, m_\star) \cdot Y$ and $\tilde{g}(Y, m_\star) \cdot Y$ are two choices in order to register Y with respect to $m_\star$, then $\mathbb{E}(g(Y, m_\star) \cdot Y) = \mathbb{E}(\tilde{g}(Y, m_\star) \cdot Y) = m_\star$ for any measurable choice of $g(Y, m_\star)$. Therefore, we have avoided the difficulty of differentiating the pre-variance in the quotient space. Without differentiating the pre-variance, we have found that if $m_\star$ minimizes the pre-variance, then $m_\star$ satisfies a certain implicit equation, which is an analogy with the implicit equation $\nabla F(m_\star) = 0$ found in section 3.3 for finite group.

Remark 5.2. *Let f be the function defined by:*

$$f(m) \in FM(g(Y, m) \cdot Y) \tag{5.3}$$

then starting from a point m_0, we define the sequence $m_{n+1} = f(m_n)$. If this sequence is well defined and converge to a point $\hat{m}$, and if f is continuous at $\hat{m}$, this point will satisfied:

$$\hat{m} \in FM(g(Y, \hat{m}) \cdot Y),$$

in other words the limit of this sequence is a good candidate to be a Fréchet mean of $[Y]$. This sequence is nothing else than the max-max algorithm. Indeed, we have seen in chapter 4, that the max-max algorithm is the repetition of two steps, the first one registers data to a current point, the second one takes the mean of the registered data in order to update the current point. This is exactly what the sequence $(m_n)_n$ defined by equation (5.3) does. The only difference is that the max-max algorithm seen in chapter 4 was restricted to a finite sample. Equation (5.3) deals with the whole distribution.

5.2.4 Inconsistency in Hilbert space thanks to equation (5.2)

In this subsubsection we give an alternative proof of theorem 3.4 based on equation (5.2).

The first advantage of this current proof is that the point which variance is strictly smaller than the variance at the template is not proportional to the template. The second advantage is that this proof works with forward generative model contrarily to theorem 3.4 which was stated and proved for backward generative model only.

Theorem 5.1. *Let M be a Hilbert space, and G a compact group acting isometrically and continuously. We assume a forward model: $Y = \Phi \cdot t_0 + \sigma\varepsilon$, where ε is a standardized noise ($\mathbb{E}(\varepsilon) = 0$ and $\mathbb{E}(\|\varepsilon\|^2) = 1$) and $\sigma > 0$. We assume that there exits $y \mapsto g(y, t_0) \cdot y$ a deterministic measurable function such that:*

$$\forall y \in M \quad g(y, t_0) \in \underset{g \in G}{argmin}\ d_M(t_0, g \cdot y).$$

We assume that $\mathbb{P}(Y \notin Cone(\Phi \cdot t_0)) > 0$. Then $[t_0]$ is not a Fréchet mean of $[Y]$ in the quotient space.

Proof. Thanks to equation (5.2), in order to show inconsistency, all we have to do is proving that the following equation is not true:

$$t_0 = \mathbb{E}(g(Y, t_0) \cdot Y). \tag{5.4}$$

By a *reductio ad absurdum*, if we assume that this equation holds. By taking the dot product of (5.4) with t_0, one get:

$$\|t_0\|^2 = \mathbb{E}(\sup_{g \in G} \langle Y, g \cdot t_0 \rangle)$$

Indeed, because the action is isometric we have that maximizing the dot product is the same thing than minimizing the distance: $\langle g(Y, t_0) \cdot Y, t_0 \rangle = \max_{g \in G} \langle g \cdot Y, t_0 \rangle$. By replacing g by Φ, and thanks to $\mathbb{P}(Y \notin \text{Cone}(\Phi \cdot t_0)) > 0$, we get:

$$\|t_0\|^2 > \mathbb{E}(\langle Y, \Phi \cdot t_0 \rangle = \mathbb{E}(\langle \Phi \cdot t_0 + \sigma\varepsilon, \Phi \cdot t_0 \rangle) = \|t_0\|^2 + 0.$$

Thus we have a contradiction, $\mathbb{E}(g(Y, t_0) \cdot Y)$ is different of t_0, then t_0 does not satisfy equation (5.2). We can conclude that t_0 does not minimize the variance in the quotient space.

This proves the inconsistency of the template estimation by computing the Fréchet mean in quotient space when the observable variable verifies a forward model. □

Moreover, we know that the variance F at the point $\mathbb{E}(g(Y, t_0) \cdot Y)$ is strictly smaller than the variance at the template. In the proof of theorem 3.4 (for Hilbert spaces) we also found $\lambda(t_0)t_0$, a point which variance was smaller than the variance at the template. By reading the proof of theorem 3.4, one could argue that: "If the only points which variance is smaller than the variance at the template are proportional to the template, inconsistency is not really an issue. Because the estimated template would be the real template up to a scaling.". This results exhibits an example of point with a smaller variance than at the template, which is not necessairly proportional to t_0.

5.2.5 Inconsistency: variation of the isotropy group

We use equation (5.2) in order to prove inconsistency, by studying the isotropy group of the template, of the observations Y and of $m_\star$. Indeed, we want to prove that the template t_0 is not equal to $m_\star$ the Fréchet mean of $[Y]$. One way to prove that two quantities are not equal, is to prove that these two quantities do not share the same properties. Here we prove that t_0 and $m_\star$ do not share the same isotropy group. We remind that the isotropy group of a point m is the group of element $g \in G$ such that $g \cdot m$:

$$\text{Iso}(m) = \{g \in G \quad g \cdot m = m\}.$$

Let us take a simple example. Let $M = C^0([0, 1], \mathbb{R})$ be the set of continuous real fonctions defined on $[0, 1]$. Then the group G defined by

$$G = \{\varphi : [0, 1] \to [0, 1] \text{ such as } \varphi \text{is an increasing homeomorphism}\},$$

acts on M by:

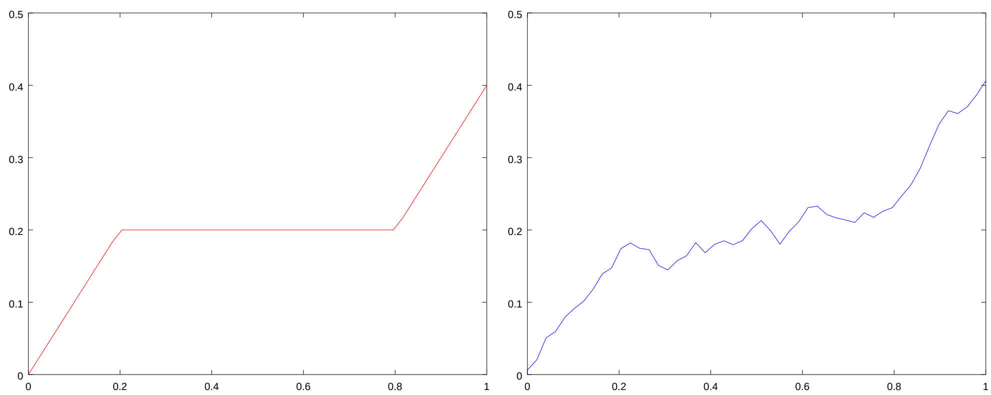
$$(\varphi, f) \mapsto f \circ \varphi.$$

Starting from a template t_0, there can be bijective maps which leave this template invariant: any bijection which lets every isolevel invariant, in this case $\mathrm{Iso}(t_0) \neq 0$. An isolevel associated to a fonction f defined on Ω and a real y is defined as:

$$\mathrm{isolevel}(f, y) = \{x \in \Omega, \text{ such as } f(x) = y\}$$

Now by adding some noise, it is likely that the observations Y with $\mathrm{Iso}(Y) = \{e_G\}$ almost surely. Then it is no longer possible to find a bijective map in G which lets every isolevel invariant except for the identity map. We give this explanation for 1D functions, but we could also explain it in 2D ou 3D. For an image, we can find many diffeomorphisms which leaves the isolevel invariant, when such isolevel are smooth enough: it is sufficient to follow the tangent at each point of the isolevel.

If we prove that $m_\star$ have the same property $\mathrm{Iso}(m_\star) = \{e_G\}$ then it will become obvious that $t_0 \neq m_\star$, in other words t_0 does not minimize the variance.



(a) The original template has a large isotropy group: any bijective map which is equal to identity on $(0, 0.2) \cup (0.8, 1)$.

(b) Due to the noise, there is no more bijective map which leaves invariant this noisy template.

Figure 5.3: Example of signals, which are 1D images.

Theorem 5.2. *Let G be a compact group acting continuously on M. We assume that M is a Hilbert space or a complete Riemannian manifold. We suppose that the distance d_M (the Hilbert norm or the Riemannian distance) is invariant under the group action G.*
Let Y be a random variable in M, we assume that it exists $m_0 \in M$ such that $\mathbb{E}(d_M^2(m_0, Y)) < +\infty$. Suppose that

$$\mathbb{P}(Iso(Y) = \{e_G\}) > 0 \tag{5.5}$$

Let us note $m_\star$ a minimizer of the variance $F(m) = \mathbb{E}\left(\inf_{g \in G} d_M(m, g \cdot Y)^2\right)$. We assume that it exists a deterministic and measurable function $y \mapsto g(y, m_\star)$ such that:

$$g(y, m_\star) \in \underset{g \in G}{argmin}\; d_M(m_\star, g \cdot y).$$

In the case of a complete Riemannian manifold M, we make the extra assumption that

$$\mathbb{P}(g(Y, m_\star) \cdot Y \in C(m_\star)) = 0,$$

where $C(m_\star)$ is the cut locus of $m_\star$.

Then $m_\star$ satisfies $Iso(m_\star) = \{e_G\}$.

Corollary 5.1. *Under the same hypotheses as theorem 5.2, we take a template t_0 such that $Iso(t_0) \neq \{e_G\}$. Then we create an observable variable Y from this template, (for instance in a Hilbert space $Y = \Phi(t_0 + \sigma\varepsilon) + \sigma'\varepsilon'$: a mixture of backward/forward model). Then the template t_0 is not a minimizer of the variance in the quotient space (since $Iso(t_0) \neq \{e_G\} = Iso(m_\star)$ for any $m_\star$ minimizer of the variance.*

We can understand corollary 5.1 as follows: $m_\star$, obtained by minimization of the variance, looses all symmetry due to the noise. In a sense we can state that $m_\star$ looks like Y and do not look like t_0; $m_\star$ is noisy.

Note that Huckemann [Huckemann 2012] has already proved that the Fréchet mean never lie on a singular orbit, in finite dimensional Riemannian manifolds for Lie group proper action (stability theorem). However, the proof we give here, works also in infinite dimensional Hilbert spaces. It is our understanding that the proof of Huckemann relies, for instance, on the fact that the set of singular points in the quotient space is a null set. Such results are not that obvious to generalize in infinite dimensional spaces. However, we should point out that Huckemann proves the measurability of the congruent section. This last result may be generalizable in our current setting.

Proof of theorem 5.2. The detail of the proof is based on these two steps:

- If $\text{Iso}(m) \neq \{e_G\}$ then it exist $h \in G \backslash \{e_G\}$ such that $h \cdot m = m$. We have:

$$F(m_\star) = \mathbb{E}(d_M(m_\star, g(Y, m_\star) \cdot Y))^2).$$

As $m_\star = h \cdot m_\star$, we also have $m_\star = h^{-1} \cdot m_\star$. Thanks to the invariant distance:

$$d_M(m_\star, g(Y, m_\star) \cdot Y) = d_M(h^{-1} \cdot m_\star, g(Y, m_\star) \cdot Y) = d_M(m_\star, hg(Y, m_\star) \cdot Y).$$

Therefore, $hg(Y, m_\star)$ also reaches the infimum in the computation of $F(m_\star)$, which proves that we have two different ways for reaching this minimum.

- Thanks to propositions 5.2 and 5.3, we get a relation between the two registration variables which reach the infimum:

$$\begin{aligned}
g(Y,m_\star)\cdot Y &= hg(Y,m_\star)\cdot Y && Y-\text{almost surely}\\
g(Y,m_\star)^{-1}hg(Y,m_\star)\cdot Y &= Y && Y-\text{almost surely}\\
g(Y,m_\star)^{-1}hg(Y,m_\star) &\in \mathrm{Iso}(Y) && Y-\text{almost surely}\\
g(Y,m_\star)^{-1}hg(Y,m_\star) &= e_G && \text{with a non zero probability (5.5)}\\
h &= e_G && \text{with a non zero probability}
\end{aligned}$$

  This is a contradiction, because we supposed $h \neq e_G$.

Conclusion: $\mathrm{Iso}(m_\star) = \{e_G\}$. Thus $t_0 \neq m_\star$, the estimation of the template is inconsistent. □

Remark 5.3. *We also need to prove that $g(Y,m_\star)\cdot Y$ is a measurable variable, if not $m \mapsto \mathbb{E}(d_M^2(m, g(Y,m_\star)\cdot Y))$ could have no sense. At this moment, we do not have a proof that $g(Y,m_\star)\cdot Y$ is a measurable variable. This technical point will be discussed in section 5.4.3.*

Remark 5.4. *When M is a Hilbert space, and when $G = V$ is a closed linear subspace of M acting by vertical translation, we have seen that the template estimation is consistent. However, d_M is invariant under the group action. Therefore, is there a contradiction with corollary 5.1? First G is not a compact set, but as previously said, this is not a problem, indeed the compacity was used in order to have an element in the group which reaches the infimum in $\|m - g\cdot n\|$. In the affine action, we have the existence of such $g \in G = V$: this is $p(m) - p(n)$, where p is the orthogonal projection into V.*

For this action, we have Iso$(m) = \{e_G\}$ for every $m \in M$. Therefore on this example, it is not possible to choose a template t_0 with Iso$(t_0) \neq \{e_G\}$, this example is not a counter example of corollary 5.1.

5.2.6 Towards an extension to other spaces

We previously assumed that the ambient space was a Hilbert or a complete Riemannian manifold only to have this result: if Z and Z' are two random variables in M such that the Fréchet mean of $Z\mathbb{1}_A + Z'\mathbb{1}_{A^c}$ does not depend of the measurable set A, then $Z = Z'$ almost surely. Here we noted A^c the complementary of the measurable set A. In other words, if all possible mixture of Z and Z' have the same Fréchet mean, then $Z = Z'$.

This leads us to the following definition.

Definition 5.1. *Let (M, d_M) a metric space, we say that M is a space with good mixtures if we have the following property:*

$$\forall Z, Z' \in L^2(\Omega, M) \quad \left[\forall A \text{ measurable } FM(Z\mathbb{1}_A + Z'\mathbb{1}_{A^c}) = FM(Z)\right] \Longrightarrow Z = Z' \text{ a.s.},$$

where $L^2(\Omega, M)$ is the set of all random variable which takes value in M and such that $\mathbb{E}(d_M^2(m, X)) < +\infty$ for one $m \in M$ (and thus for all $m \in M$ by triangular inequality).

Theorem 5.2 can be immediately generalized in *any spaces with good mixture*. Can we exhibit other *spaces with good mixtures* different from Hilbert spaces, for instance metric spaces?

5.3 Conjecture of inconsistency for metric space with non invariant distance under the group action.

This section is an attempt to generalize the proof of inconsistency in spaces which are not Hilbert spaces anymore but just metric spaces. We do not suppose that the distance in the ambient space is invariant under the group action.

Although we do not give a proof, we provide an intuition that it will not be consistent either. We state two conjectures: the first one is dedicated to metric space when no regularization is added. On the contrary, the second one is a generalization when a regularization term is added. Working in a metric space with a regularization term is the real framework of applications.

5.3.1 Conjecture 1: in metric space without regularization

Let (M, d_M) be a metric space, G a compact group acting continuously. We do not suppose that the distance d_M is invariant under the group action. In particular, $\inf_{g \in G} d_M(x, g \cdot y)$ will not define a distance in the quotient space.

In this section, we consider the backward model: $Y = \Phi \cdot X$ where Φ is a random variable in G, X is a random variable defined as a noisy version of the template: the template is t_0, the unique Fréchet mean of X:

$$t_0 = \operatorname*{argmin}_{m \in M} \mathbb{E}(d_M^2(m, X)).$$

We also assume that Φ and X are independent random variables.

We are interested in the template estimation given the observation Y by the minimization of the pre-variance:

$$F(m) = \mathbb{E}\left(\inf_{g \in G} d_M^2(m, g \cdot Y)\right).$$

Because $X = \Phi \cdot Y$ we have also:

$$F(m) = \mathbb{E}\left(\inf_{g \in G} d_M^2(m, g \cdot X)\right).$$

We use the term pre-variance, because it is not a variance (because we do not have a quotient distance). As the group is compact and the action is continuous $\inf_{g \in G} d_M(m, g \cdot X)$ is reached for some $g \in G$; we note $g(X, t_0)$ one element which reaches the infimum. Therefore, $F(t_0) = \mathbb{E}(d_M^2(t_0, g(X,t_0)\cdot X))$. We define the variance of the random variable $g(X,t_0)\cdot X$:

$$m \mapsto E(m) = \mathbb{E}(d_M^2(m, g(X,t_0)\cdot X)).$$

We define the Voronoï cell associated to the point x defined as the set of elements closer to x than to all the point in the orbit of x:

$$\text{VC}(x) = \{t \in M \text{ s.t. } \forall g \in G \quad d_M(x,t) \leqslant d_M(g\cdot x, t)\}.$$

Conjecture 5.1. *Let (M, d_M) be a metric space, let G be a compact group acting continuously on M. We assume that a point t_0 is the unique Fréchet mean of X. Let us suppose that the template t_0 does not always belong to $VC(X)$:*

$$\mathbb{P}(t_0 \notin VC(X)) > 0.$$

In this case, we think that t_0 is not a minimizer of the pre-variance F.

Unfortunately, this conjecture is false, indeed, there is a counter example: the action of vertical translation $(v,m) \mapsto v+m$ is a counter example for $m \in M$ (M a Hilbert space), and $v \in V$ (V is a closed linear sub-space. It may be possible to fix this conjecture by adding an hypothesis which excludes this counter-example.

This hypothesis is very similar to the one in theorems 3.1 and 3.4 (the theorems of inconsistency in chapter 3). Indeed in these theorems, the condition was $\mathbb{P}(X \notin \text{Cone}(t_0)) > 0$, where $\text{Cone}(t_0) = \text{VC}(t_0)$ was the Voronoï cell associated to t_0, (this cell was a cone, due to the isometric action). In theorems 3.1 and 3.4, the action was isometric, and this condition could be also written $\mathbb{P}(t_0 \notin \text{VC}(X)) > 0$. Therefore in conjecture 5.1, we have the same condition, however, as the space is no longer a linear space, the voronoï cell associated to X is no longer a cone.

Let us detail why we think that this conjecture may be true:

Proposition 5.4. *There are two cases:*

- *If $g(X,t_0)\cdot X$ does not have a Fréchet mean in M, then t_0 does not minimize F: the template estimation is inconsistent.*
- *If z is one Fréchet mean of $g(X,t_0)\cdot X$, then $F(z) \leqslant F(t_0)$. Therefore z is a good candidate to show that t_0 does not minimize F.*

Proof of proposition 5.4.
- Let us prove the first point: if $g(X,t_0)\cdot X$ does not have a Fréchet mean in M, this means in particularly that t_0 does not minimize E. In this case, it exists $z' \in M$ such that $E(z') < E(t_0)$. Then once again:

$$\begin{aligned} F(t_0) &= E(t_0) \\ &> E(z') \\ &> \mathbb{E}(d_M^2(z', g(X,t_0)\cdot X)) \\ &> \mathbb{E}\left(\inf_{g\in G} d_M^2(z', g\cdot X)\right) \\ &> F(z'), \end{aligned}$$

- For the second point, if z is one Fréchet mean of $g(X,t_0)\cdot X$, then:

$$\begin{aligned} F(t_0) &= \mathbb{E}\left(\inf_{g\in G} d_M^2(t_0, g\cdot X)\right) \\ &= E(t_0) \\ &\geqslant E(z) \\ &\geqslant \mathbb{E}(d_M^2(z, g(X,t_0)\cdot X)) \\ &\geqslant \mathbb{E}\left(\inf_{g\in G} d_M^2(z, g\cdot X)\right) \\ &\geqslant F(z), \end{aligned}$$

□

Is it possible that $t_0 \in \mathrm{FM}(g(X,t_0)\cdot X)$? In this case t_0 would minimize two functions:

$$m \mapsto \mathbb{E}(d_M^2(m,X)) \text{ and } m \mapsto \mathbb{E}^2(d_M^2(m, g(X,t_0)\cdot X)).$$

In the case where $g(X,t_0) = e_G$ X-a.s. It is obvious that t_0 minimizes these two functions. This would mean that t_0 belongs almost surely in the Voronoï Cone of X. Therefore, as we have assumed that it is not the case, we can conclude that $g(X,t_0) \neq e_G$ with a non zero probability. Therefore, the only thing remaining to prove that t_0 does not minimize the pre-variance, is proving that $z \neq t_0$, for every $z \in \mathrm{FM}(g(X,t_0)\cdot X)$. Maybe, this can be done by adding extra assumptions.

The difficulty, in order to finish the proof of this conjecture, is that we are not able to compute the minimum of the variance E.

Note that we require a compact group acting continuously so that the registration problem with respect to the template has at least a solution. Therefore an immediate extension for non compact group is when the registration problem has at least a solution.

Remark 5.5. *As in section 5.2.5, we also need to prove that $g(X,t_0)\cdot X$ is a measurable variable, if not $m \mapsto \mathbb{E}(d_M^2(m, g(X,t_0)\cdot X))$ could have no sense. At this point, we do not have a proof that $g(X,t_0)\cdot X$ is a measurable variable. This technical point will be discussed in section 5.4.3.*

5.3.2 Conjecture 2: metric space with regularization

Let (M, d_M) be a metric space, G a compact group acting continuously. We take $Y = \Phi\cdot X$ where X is a noisy template: the template is t_0 the unique Fréchet mean of X:

$$t_0 = \operatorname*{argmin}_{m\in M} \mathbb{E}(d_M^2(m,X)).$$

We are interested in the template estimation given the observation Y with the minimization of the regularized pre-variance defined by:

$$F(m) = \mathbb{E}\left(\inf_{g\in G} d_M^2(m, g\cdot Y) + \text{Reg}(g)\right),$$

where Reg is a regularization over the group. Here we suppose that

$$\begin{array}{ccc} G & \to & \mathbb{R}^+ \\ g & \mapsto & \text{Reg}(g) \end{array}$$

is a continuous map, and that $\text{Reg}(e_G) = 0$. In most non linear registration algorithm, for instance, in LDDMM framework [Beg 2005], one makes a trades off between an inexact matching thanks to the term $d_M^2(m, g\cdot Y)$, and a realistic selection of the chosen deformation via the regularization term $\text{Reg}(g)$. In LDDMM, the regularization is the squared norm of the vector which defined the flow equation satisfying by the chosen diffeomorphim. For a point $x \in M$ and an element $h \in G$, we define the regularized Voronoï cell of x as the element closer from x than the other element of the orbit of x. However, here the "closer" needs to be understood as the distance d_M regularized by Reg, the regularization term:

$$\text{VC}_{\text{reg}}(x,h) = \{t \in M,\ \text{s.t.}\forall g\in G \quad d_M(t,x) + \text{Reg}(h^{-1}) \leqslant d_M(t, g\cdot x) + \text{Reg}(gh^{-1})\}$$

The definition is similar to the definition of the Voronoï cell, but here, the regularization acts as a deformation of the metric d_M. As $\text{Reg}(e_G) = 0$, and $\text{Reg}(g) \geqslant 0$, it is easy to see that: $\text{VC}(x) \subset \text{VC}_{\text{reg}}(x, e_G)$, in other words, the regularized Voronoï cell is bigger than the original Voronoï cell.

Conjecture 5.2. *Let X be a random variable, we assume that t_0 is the unique Fréchet mean of X. $Y = \Phi \cdot X$, where Φ is a random variable in G, X and Φ are assumed to be independant. We suppose that t_0 does not belong almost surely to the random regularized Voronoï cell $VC_r(X,\Phi)$. Then we think that t_0 is not a minimizer of F.*

Note that *a priori* for x and h, $\text{VC}_{\text{reg}}(x,h)$ is a strictly subset of M. Therefore, it is possible to fulfill the condition t_0 does not belong almost surely in $\text{VC}_{\text{reg}}(X,\Phi)$.

Once again, let us explain why we think that this conjecture may be true: As the group G is compact and acts continuously, we have the existence of an element which reaches the infimum in the regularized pre-variance:

$$\begin{aligned} F(t_0) &= \mathbb{E}(\inf_{g\in G} d_M^2(t_0, g\cdot Y) + \text{Reg}(g)) \\ &= \mathbb{E}(d_M^2(t_0, g(Y,t_0)\cdot Y) + \text{Reg}(g(Y,t_0)) \\ &= \mathbb{E}(d_M^2(t_0, g(Y,t_0)\cdot Y) + \text{Reg}(g(Y,t_0)) \\ &\geqslant \mathbb{E}(d_M^2(z, g(Y,t_0)\cdot Y)) + \mathbb{E}(\text{Reg}(g(Y,t_0))) \\ &\geqslant \mathbb{E}(\inf_{g\in G} d_M^2(z, g\cdot Y) + \text{Reg}(g)) \\ &\geqslant F(z), \end{aligned}$$

where z is one minimizer of the function $m \mapsto \mathbb{E}(d_M^2(m, g(Y, t_0) \cdot Y)$. The point z is then a serious candidate which variance is smaller than the variance at the template. Note that the $g(Y, t_0)$ is the element which minimizes $g \mapsto d_M^2(t_0, g \cdot Y) + \mathrm{Reg}(g)$, contrarily to conjecture 1, where $g(Y, t_0)$ minimized $g \mapsto d_M^2(t_0, g \cdot Y)$.

Is it possible to have $t_0 \in \mathrm{argmin}_{m \in M} \mathbb{E}(d_M^2(m, g(Y, t_0) \cdot Y)$? If true, t_0 minimizes two functions:

$$m \mapsto \mathbb{E}(d_M^2(m, \Phi^{-1} Y)) \text{ and } m \mapsto \mathbb{E}^2(d_M^2(m, g(Y, t_0) \cdot Y)).$$

In the case where $g(Y, t_0) = \Phi^{-1}$ Y-a.s. it is obvious that it is the case, since that the two functions are equal. By definition of $g(Y, t_0)$, if $g(Y, t_0) = \Phi^{-1}$ Y-a.s., this means that:

$$\forall g \in G \quad d_M^2(t_0, g \cdot Y) + \mathrm{Reg}(g) \geqslant d_M^2(t_0, X) + \mathrm{Reg}(\Phi^{-1}).$$

Besides $Y = \Phi \cdot X$, this leades to:

$$\forall g \in G \quad d_M^2(t_0, g \cdot X) + \mathrm{Reg}(g\Phi^{-1}) \geqslant d_M^2(t_0, X) + \mathrm{Reg}(\Phi^{-1}).$$

This means that the template t_0 belongs almost surely to $\mathrm{VC}_{\mathrm{reg}}(X, \Phi)$, which is exactly the hypothesis we excluded. The only thing we need to prove is that t_0 is not a Fréchet mean of $g(Y, t_0) \cdot Y$. Because in this case, for $z \in \mathrm{FM}(g(Y, t_0) \cdot Y)$, we would have $F(z) < F(t_0)$ proving the inconsistency.

The difficulty in order to prove that t_0 does not minimize $\mathrm{FM}(g(Y, t_0) \cdot Y)$ is that we do not know $g(Y, t_0)$ therefore it hard to know if t_0 minimize $\mathrm{FM}(g(Y, t_0) \cdot Y)$.

5.4 Congruent and measurable sections

In this section, we discuss the existence of a section of the quotient space satisfying a certain property (being congruent or being measurable). First, a section of the quotient space is a map which associate at each orbit an element of this orbit. Note that a section always exists: it suffices, for each orbit, to chose one element in the orbit (with the axiom of choice).

In sections 5.4.1 and 5.4.2, we discuss the existence of congruent section, this means that this section leaves the distance invariant. We see that if such a section exists, then computing the Fréchet mean in the quotient space, and the consistency bias is straightforward. For instance, for the group of rotations acting on $\mathbb{R}^n$ a congruent section exists, and this way, we are able to have an explicit formula of the consistency bias. Unfortunately, we show that a congruent section does not always exists.

In section 5.4.3, we discuss the existence of a measurable section such that the image of this section is included in a Voronoï cell. If a measurable section exists, then we solve the missing technical details of sections 5.2 and 5.3: we needed to prove that $g(Y, t_0) \cdot Y$ and $g(Y, m_\star) \cdot Y$ are measurable variables, Y is the observable variable, t_0 the template, and $m_\star$ the element which minimizes the pre-variance, and $g(a, b)$ is one element which registers a with respect to b.

5.4.1 Congruent section and Computation of Fréchet Mean in Quotient Space

Let M be an Hilbert space. Given points $m \in M$ and $y \in M$, there is *a priori* no closed-form expression in order to compute the quotient distance $\inf_{g \in G} \|g \cdot m - y\|$. Therefore computing and minimizing the variance in the quotient does not seem straightforward. There is one case where it may be possible: the existence of a congruent section. We say that $s : Q \to M$ is a section if $\pi \circ s = Id$, where $\pi : M \to Q$ is the canonical projection into the quotient space. Moreover we say that the section s is congruent if:

$$\forall o, o' \in Q \quad \|s(o) - s(o')\| = d_Q(o, o').$$

Then the image of the quotient by the section $\mathcal{S} = s(Q)$ is a part of M which has an interesting property:

$$\forall p, q \in \mathcal{S}, \|p - q\| = d_Q([p], [q]).$$

In other words, the section gives us a part of M containing a point of each orbit such that all points in $\mathcal{S}$ are already registered. Moreover, if s is a section, $s' : [m] \mapsto g \cdot s([m])$ is also a section, without loss of generality we can assume that $t_0 = s([t_0])$.

In this case, the variance is equal to:

$$F(m) = \mathbb{E}(\|s([m]) - s([Y])\|^2),$$

where we recognize the variance of the random variable $s([Y])$. As we know that the element which minimizes the variance in a linear space is given by the expected value, we have that:

$$F(m) \geqslant F(\mathbb{E}(s([Y]))).$$

Moreover this inequality is strict if and only if m and $\mathbb{E}(s([Y]))$ are not in the same orbit.

Therefore, we have a method to determine if the estimation is consistent or not: computing $\mathbb{E}(s([Y]))$ and verifying if t_0 and $\mathbb{E}(s([Y]))$ are in the same orbit, and the consistency bias is given by $d_Q([t_0], [\mathbb{E}(s([Y]))])$. Moreover if we take $m \in \mathcal{S}$, we have $F(m) = \mathbb{E}(\|m - s([Y])\|^2)$ and it is now straightforward that the restriction of F to $\mathcal{S}$, noted $F|_{\mathcal{S}}$, is differentiable[1] on $\mathcal{S}$, and that $\nabla F|_{\mathcal{S}}(m) = m - \mathbb{E}(s([Y])$. In particular $\|\nabla F|_{\mathcal{S}}(t_0)\| = \|t_0 - \mathbb{E}(s([Y]))\|$ gives us the value of the bias.

Note that in this thesis, we have already seen two examples of action with a congruent section see section 3.3.6 and example 4.6. Each time, we were able to compute the bias, as in the following example of rotations (studied in [Miolane 2017]):

[1] We say that $F|_{\mathcal{S}}$ is differentiable on $\mathcal{S}$, even if $\mathcal{S}$ is not open, because $m \mapsto \mathbb{E}(\|m - s([Y])\|^2)$ is defined and differentiable on M, and is equal to $F|_{\mathcal{S}}$.

Example 5.1. *The action of rotations: $G = SO(n)$ acts isometrically on $M = \mathbb{R}^n$. We notice that the quotient distance is $d_Q([x],[y]) = |\|x\| - \|y\||$. We can check that $s([x]) = \|x\|v$ is a section for an unitary vector v . Therefore the computation of the bias is given by $d_Q([t_0], [\mathbb{E}(s([Y])]) = |\mathbb{E}(\|Y\|) - \|t_0\|)|$.*

Unfortunately, the congruent section generally does not exist. Let us give an example:

Example 5.2. *Taking $N \in \mathbb{N}$ with $N \geqslant 3$, we consider the action of $G = \mathbb{Z}/N\mathbb{Z}$ on $M = \mathbb{R}^{\mathbb{Z}/N\mathbb{Z}} = \mathbb{R}^G$ by horizontal translation: for $\tau \in \mathbb{Z}/N\mathbb{Z}$, and $(x_1, x_2, \ldots, x_N) = \mathbb{R}^G$:*

$$\tau \cdot (x_1, x_2, \ldots, x_N) = (x_{1+\tau}, x_{2+\tau}, \ldots, x_{N+\tau}),$$

Let us take three points p_1, p_2 and p_3 in M defined by:

$$p_1 = (0, 5, 0, \ldots, 0), \quad p_2 = (0, 3, 2, 0, \ldots, 0) \text{ and } p_3 = (2, 3, 0, \ldots, 0).$$

By hand we can check that there is no $x \in [p_1]$, $y \in [p_2]$ and $z \in [p_3]$ such that $\|x-y\| = d_Q([p_1],[p_2])$, $\|x-z\| = d_Q([p_1],[p_3])$, and $\|y-z\| = d_Q([p_2],[p_3])$. Thus, a congruent section in $Q = M/G$ does not exists.

We can generalize this simple example by taking a non finite group:

Example 5.3. *Let us take $M = L^2(\mathbb{R}/\mathbb{Z})$ the set of 1-periodic functions such that $\int_0^1 f^2(t)dt < +\infty$ (this example was already introduced in 3.3). $G = \mathbb{R}/\mathbb{Z}$ acts on $L^2(\mathbb{R}/\mathbb{Z})$ by horizontal translation:*

$$\tau \in \mathbb{R}/\mathbb{Z},\ f \in L^2(\mathbb{R}/\mathbb{Z}) \mapsto f_\tau \text{ with } f(x) = f(x+\tau).$$

Then a congruent section in $Q = M/G$ does not exists.

Proof. Let us take $f_1 = \mathbb{1}_{[\frac{1}{4},\frac{3}{4}]}$, $f_2 = f_1 + 2\mathbb{1}_{[\frac{1}{4},\frac{1}{4}+\eta]}$ and $f_3 = f_1 + 2\mathbb{1}_{[\frac{1}{4}+\eta,\frac{1}{4}+2\eta]}$ for some $\eta \in (0, \frac{1}{4})$ (see figure 5.4). Let us suppose that a congruent section s exists. Without loss of generality we can assume that $s([f_1]) = f_1$, we should have

$$\|f_1 - s([f_2])\| = \|s([f_1]) - s([f_2])\| = d_Q([f_1],[f_2]).$$

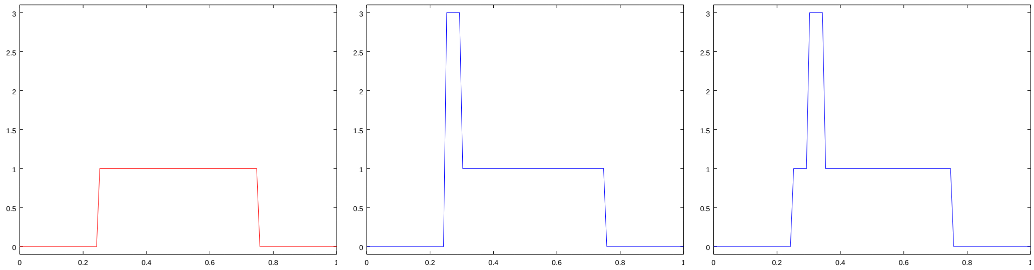
In other words, $s([f_2])$ should be registered with respect to f_1. For $\tau \in \mathbb{R}/\mathbb{Z}$ we can verify that $\|f_1 - \tau \cdot f_2\| \geqslant \|f_1 - f_2\|$ and that this inequality is strict as soon as $\tau \neq 0$. Then f_2 is the only element of $[f_2]$ registered with f_1 then $s([f_2]) = f_2$. Likewise for $s([f_3]) = f_3$, then we should have:

$$d_Q([f_2],[f_3]) = \|f_2 - f_3\|.$$

However it is easy to verify that

$$d_Q^2([f_2],[f_3]) \leqslant \|\eta \cdot f_2 - f_3\|^2 = 2\eta < 8\eta = \|f_2 - f_3\|^2 = d_Q([f_2],[f_3])$$

This is a contradiction. Therefore, a congruent section does not exist. □



(a) f_1 (b) f_2 (c) f_3

Figure 5.4: Representation of the three functions f_1, f_2 and f_3 with $\eta = 0.05$. the functions f_2 and f_3 are registered with respect to f_1. However f_2 and f_3 are not registered with each other, since it is more profitable to shift f_2 in order to align the highest parts of f_2 and f_3.

This can be done for every $\eta > 0$, therefore, we can also conclude that a local congruent section around f_1 does not exist either: because $\lim_{\eta \to 0} f_2 = \lim_{\eta \to 0} f_3 = f_1$.

Ziezold [Ziezold 1977] had already noticed that being registered (he uses the expression "optimal position" instead of registered) was not always a transitive relation in other examples.

The existence of a congruent section indicates us that the quotient space is not so complicated. Indeed when a congruent section exists, the quotient space is embedded in the ambient space with respect to the distances in the quotient space and in the ambient space. The computation of the Fréchet mean would be easier, it suffices to project data on $\mathcal{S}$ and to take the mean. When such a congruent section does not exist, computing the Fréchet mean in quotient space is not so easy. However, we can established proofs of inconsistency which are less tight.

Remark 5.6 (Link between congruent section and equation (5.2)). *When M is a Hilbert space, equation (5.2) is an implicit equation which gives the expression of $m_\star$, an element which minimizes the pre-variance in the quotient space:*

$$m_\star = \mathbb{E}(g(Y, m_\star) \cdot Y).$$

If we knew $m_\star$, we could compute $g(Y, m_\star)$. Therefore we could compute $\mathbb{E}(g(Y, m_\star))$ and therefore we could compute $m_\star$. The serpents eats it own tail. However, if there is a congruent section s, things will become much easier. Indeed let us assume that a congruent section s exists, if $m_\star$ minimizes F. Without loss of generality we can assume that $s([m_\star]) = m_\star$, then $g(Y, m_\star) \cdot Y = s([Y])$ does not depend on $m_\star$, and equation (5.2) becomes $m_\star = \mathbb{E}(s([Y]))$: finding $m_\star$ is just the computation of an expected value.

5.4.2 Congruent section in Euclidean space

We have seen that the existence of a congruent section is useful to compute an element which minimize the variance. In section 5.4.1, we were able to prove by an example that for horizontal translation action in $\mathbb{R}^d$ a discrete space or $L^2([0,1])$ a continuous space, we do not have the existence of a congruent section. But, for a given action, what if we are not able to find such an example to reject the existence of a congruent section by ourselves? In this section we propose an algorithm which is able to reject the existence of such a congruent section.

Let G be a group acting on M (a Hilbert space). First, we can notice that, if $d_Q([a],[b]) = \inf_{g\in G} \|a - g \cdot b\|$ is not a distance in a quotient space, then a section can not exist. Indeed if we have d_M a distance in the ambient space and s a congruent section, every property of the distance d_M would be verified by d_Q through the congruent section. Therefore in the following, we restrict ourselves to isometric action.

Now, we propose a simple algorithm which can reject the existence of a congruent section in Euclidean space. Let $M = \mathbb{R}^d$ be an Euclidean space. Taking random points in the ambient space $(y_1, \dots, y_n)$. If there is a congruent section, in particular we have:

$$\forall i \in [\![1,n]\!] \quad \exists x_i \in [y_i] \quad \text{s.t.} \quad \forall j \in [\![1,n]\!] \quad \|x_i - x_j\|^2 = d_Q([y_i],[y_j])^2. \tag{5.6}$$

By expanding the square norms, we get:

$$\langle x_i - x_n, x_j - x_n\rangle = \frac{1}{2}\left(\|x_i - x_n\|^2 + \|x_j - x_n\|^2 - \|x_i - x_j\|^2\right). \tag{5.7}$$

Even if we do not know the family of vectors $(x_i - x_n)_{1\leqslant i\leqslant n-1}$, thanks to the chosen vectors $(y_i)_i$, and thanks to equations (5.6) and (5.7), we can compute $A \in M_{n-1}(\mathbb{R})$, the Gram matrix associated to the family of vectors $(x_i - x_n)_{1\leqslant i\leqslant n-1}$ (if these vectors (x_i) exist):

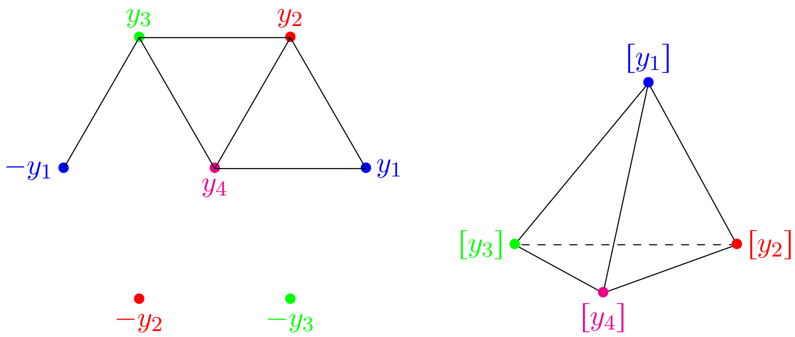
$$\begin{aligned}\forall (i,j) \in [\![1,n]\!]^2 \quad A_{ij} &= \langle x_i - x_n, x_j - x_n\rangle \\ &= \frac{1}{2}\left(d_Q([y_i],[y_n])^2 + d_Q([y_j],[y_n])^2 - d_Q([y_i],[y_j])^2\right).\end{aligned}$$

If we denote by B the matrix such that the i-th column of B are the coordinate of $x_i - x_n$, we have $A = (B)^T B$, moreover the rank of B is equal to the rank of A. Then we should have $\text{rank}(A) = \text{rank}(B) \leqslant \dim M$ (since the vectors $x_i - x_n$ belongs to M).

Example 5.4. *Let us chose $G = \{I_2, -I_2\}$ acting isometrically on $M = \mathbb{R}^2$, an orbit of a point m is $[m] = \{m, -m\}$, the quotient distance is:*

$$d_Q([m],[n]) = \min(\|m - n\|, \|m + n\|)$$

We chose $y_1 = (1,0)$, $y_2 = (\frac{1}{2}, \frac{\sqrt{3}}{2})$, $y_3 = (-\frac{1}{2}, \frac{\sqrt{3}}{2})$, $y_4 = (0,0)$ (see figure 5.5). Then we have $d_Q([y_i],[y_j]) = 1$ for every $(i,j) \in [\![1,4]\!]^2$ (except if $i = j$; $d_Q([y_i],[y_i]) = 0$).



(a) The four orbits in $\mathbb{R}^2$, the segments are the lines which realize the quotient distance.

(b) Representation of the four orbits in the quotient space, the quotient space can not be embedded into $\mathbb{R}^2$ in a congruent way.

Figure 5.5: Representation of the four points in $\mathbb{R}^2$ and representation of the four orbits

Therefore, if a congruent section exists, it should be possible to find 4 points in $\mathbb{R}^2$ *such that any two different points among these 4 points would have a distance equal to 1. Of course, this is not possible, in order to prove it, we compute the matrix* A *as explained above, we find:*

$$A = \frac{1}{2}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix},$$

then $\mathrm{rank}(A) = 3 > \dim M$*. In order to build a regular tetrahedron, we need to be at least in* $\mathbb{R}^3$*.*

Let us return to a general isometric action on a general Euclidean space. By choosing the $y_i's$ randomly, if we get an example where $\mathrm{rank}(A) > \dim M$, then we reject the existence the x_i such that $\|x_i - x_j\|^2 = d_Q([y_i],[y_j])^2$ for every $(i,j) \in [\![1,n]\!]^2$. Therefore, we reject the existence of a congruent section for the chosen group action.

Note that the rank of A is very sensitive to error in the computation of the squared quotient distance. That is why it is preferable to chose some vectors y_i with coordinates which are integer.

We implement this algorithm, with the action of horizontal translation in $\mathbb{R}^{64}$. For instance, if we take 100 random points we find that the rank of the matrix A is 99. In other words, in order to embed the quotient space $\mathbb{R}^{64}/(\mathbb{Z}/64\mathbb{Z})$ with respect to the quotient and Euclidean distances, we need at least an Euclidean space of dimension 99. But if we take more points for instance 500 we find that we need a at least an Euclidean space of dimension 499. The more we take points the larger the dimension must be. In other words it is probably not possible to embed the quotient space in an Euclidean space with respect to the quotient distance. In particular, the original ambient space of dimension 64 is not sufficient: there is no congruent section.

Remark 5.7. *Here, we can notice that the Nash embedding theorem [Nash 1956] states that any Riemannian manifold can be included in $\mathbb{R}^d$ with $d \in \mathbb{N}$ sufficiently large. Furthermore, this embedding is cconsistent with the Riemannian metric. Then, if the embedded Riemannian manifold is totally geodesic in $\mathbb{R}^d$. Then the quotient distance is given by the Euclidean norm in $\mathbb{R}^d$. Here the situation is more complex, indeed the quotient space Q is not a manifold, and even if we remove some points in Q to obtain a Riemannian manifold $Q^\star$, we do not want to embed $Q^\star$ into $\mathbb{R}^d$ with d sufficiently large, we want to embed $Q^\star$ into M with an application s such that $s([x]) \in [x]$.*

5.4.3 Measurable section

We now assume that (M, d_M) is a metric set, and G a compact group acting continuously on M. We have seen in sections 5.2 and 5.3 that we need to prove the existence of the measurable variables $g(Y, t_0) \cdot Y$ and $g(Y, m_\star) \cdot Y$, where $g(a, b)$ is one element which registers a to b (see remarks 5.3 and 5.5). In fact, in these previous sections, we have admitted the existence of such measurable variables. This technical detail matters because we need it in order to make the proof of theorem 5.2 rigorous. In this section we make the link between this problem and the existence of an measurable section.

Definition 5.2. *Let μ be a point in M (μ could be t_0 or $m_\star$ depending on the context). We say that s is a μ-measurable-section if:*

- *s is a section namely, s is a map; $s : M/G \to M$ with $\pi \circ s = Id$ (where π is the canonical projection into the quotient space).*
- *s is a measurable map.*
- *$\mathcal{S} = s(M/G)$ is included in the Voronoï cell of μ:*

$$s : \begin{array}{ccc} M/G & \to & VC(\mu) \\ [x] & \mapsto & y \text{ where } y \in [x] \end{array}.$$

Once again, it is easy to show that a section $s : M/G \to \mathrm{VC}(\mu)$ exists: for every orbit, it suffices to choose one element among the element of the orbit which are the closest to μ. As the group is compact and acts continuously, the existence of this element is ensured. Therefore, by the axiom of choice, s exits. However, nothing in this argument ensure the measurability of the section. On the contrary, using the axiom of choice is the best thing to do in order to build non measurable functions or non measurable sets.

When such a μ-measurable-section exists, then $g(Y, \mu) \cdot Y = s(Y)$ is a measurable variable because Y and s are measurable. Therefore in order to solve the technical problem seen in sections 5.2 and 5.3, all we have to do is reduced to prove the existence of μ-measurable-section.

If we have the existence of a congruent section, the result is straightforward: let s be a congruent section, without loss of generally we can assume that $s([\mu]) = \mu$, then for every point $x \in \mathcal{S} = s(Q)$ we have $d_Q([x], [\mu]) = d_M(x, \mu)$, then $x \in \mathrm{VC}(\mu)$ therefore $\mathcal{S} \subset \mathrm{VC}(\mu)$, then $s : Q \to \mathrm{VC}(\mu)$ is measurable since continuous since congruent. Unfortunately, we have already seen that a congruent section does not always exists.

Fortunately, being measurable is much weaker assumption that being congruent. Then, it should be possible to show that a μ-measurable-section exists even if there is no congruent section.

5.5 Inconsistency for non isometric action by perturbation of an isometric group action

In this Section, we exhibit a sufficient condition in order to prove inconsistency even for non isometric action. However, knowing if this condition is satisfied is difficult, except if the action is isometric. Therefore, we propose to prove that this condition is only verified for action which are small perturbation of an isometric action.

5.5.1 Notation, hypothesis and theorem of inconsistency

Let M be a Hilbert space, G acting on M, Y a random variable such that $\mathbb{E}(\|Y\|^2) < +\infty$. The observable variable Y is built with a template t_0 added to unbiased noise ε and a random transformation Φ in the group G. The random deformation Φ and the noise ε are independent variables. This leads to two different generative models:

$$Y = \Phi \cdot t_0 + \varepsilon,$$

which is called forward model. Or

$$Y = \Phi \cdot (t_0 + \varepsilon),$$

which is called a backward model.

In this Section we exhibit results for this two models, however in the case of a forward model we will need an extra hypothesis: the action is linear which means that $x \mapsto g \cdot x$ is linear for all $g \in G$.

In this Section, we suppose that there exists some non negative constants A, a, B, b such that:

$$\forall x \in M \quad \forall g \in G \qquad a\|x\|^2 + b \leqslant \|g \cdot x\|^2 \leqslant A\|x\|^2 + B, \tag{5.8}$$

This can be seen as a relaxation of the isometric action (previously studied) where we had:

$$\forall x \in M \quad \forall g \in G \qquad \|x\|^2 = \|g \cdot x\|^2.$$

Note that it is always possible to take $a = b = 0$, but we will see that it is possible to have a more precise result if a and b are greater than 0.

Example 5.5. *Let us give three examples where (5.8) is verified:*

- *For an isometric action we have $A = a = 1$ and $B = b = 0$.*
- *More generally, if one has a compact group of linear operators for the euclidean norm $\| \quad \|$, we can take:*

$$A = \sup_{g \in G} \|\!|g|\!\|^2 \quad a = \frac{1}{A} \quad B = b = 0,$$

where $\|\!|g|\!\|$ is the subordinate norm of the linear application $x \mapsto g \cdot x$.

- *In Example (4.2), this condition is fulfilled as soon as the diffeomorphims in the group G has a Jacobian determinant uniformly bounded since:*

$$\|f \circ \phi^{-1}\|_2^2 = \int_{\mathbb{R}^n} f(x)^2 |Jac_x| dx\phi \leqslant A\|f\|_2^2.$$

We can define the function $m \mapsto F(m) = \mathbb{E}(\inf_{g \in G} \|m - g \cdot Y\|^2)$ and provide frameworks in which t_0 does not always minimize F. As we do not suppose an isometric action, F is a called pre-variance instead of being called variance. We define

$$\lambda(t_0) = \frac{\mathbb{E}\left(\sup_{g \in G} \langle t_0, g \cdot Y \rangle\right)}{\|t_0\|^2}.$$

Proposition 5.5. *Let G a group acting on M a Hilbert space. We assume that Y satisfies a backward or forward generative model from a template t_0 (with the extra assumption that G acts linearly in the case of a forward model). We suppose that:*

$$[\|t_0\|(\lambda(t_0) - 1)]^2 > (A - a)\mathbb{E}(\|Y\|^2) + B - b. \tag{5.9}$$

We also assume that $\lambda(t_0) > 1$. Then t_0 does not minimize F.

Condition (5.9) can be understood as follows: the non isometric action leads to some constants A, a, B, b such that A and a are closed enough and so are B and b. In the case of an isometric action $B = b = 0$ and $A = a = 1$ and the right member is equal to 0, besides we have seen that $\lambda(t_0) > 1$ as soon as the quotient distance is a contraction with respect to the ambient distance (see chapter 3). Therefore, condition (5.9) is a generalization of the sufficient condition for inconsistency in the case of isometric action.

Proof. We define for $\lambda \geqslant 1$, $f(\lambda) = F(\lambda t_0)$. Then we have:

$$f(\lambda) = \lambda^2 \|t_0\|^2 + \mathbb{E}\left(\inf_{g \in G} \|g \cdot Y\|^2 - 2\lambda \langle t_0, g \cdot Y \rangle\right).$$

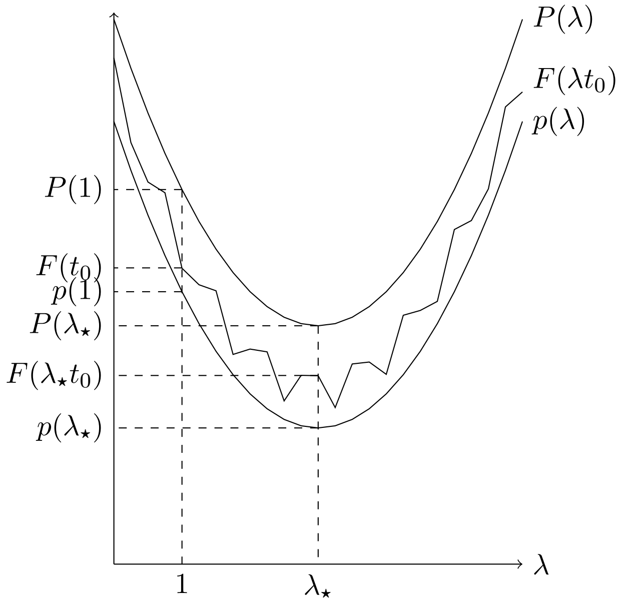


Figure 5.6: Sketch of the proof: thanks to condition (5.8) we find two parabola P and p such that $p(\lambda) \leqslant F(\lambda t_0) \leqslant P(\lambda)$ for all $\lambda \geqslant 0$. P and p reach their minimum at $\lambda(t_0) > 1$. Condition (5.9) ensures that $P(\lambda(t_0)) < p(1)$. Then $F(\lambda(t_0)t_0) < F(t_0)$ as a conclusion t_0 is not the minimum of F. In chapter 3, the action was isometric and $P(\lambda) = p(\lambda) = F(\lambda t_0)$ for all $\lambda > 0$ leading to a simpler proof.

By using condition (5.8), we get:

$$f(\lambda) \leqslant \lambda^2 \|t_0\|^2 - 2\lambda \mathbb{E}(\sup_{g \in G} \langle t_0, g \cdot Y \rangle + A\mathbb{E}(\|Y\|^2) + B = P(\lambda).$$

Similarly,

$$f(\lambda) \geqslant \lambda^2 \|t_0\|^2 - 2\lambda \mathbb{E}(\sup_{g \in G} \langle t_0, g \cdot Y \rangle + a\mathbb{E}(\|Y\|^2) + b = p(\lambda).$$

This allows us to determine the unique $\lambda \in \mathbb{R}^+$ which minimizes P and p:

$$\lambda(t_0) = \text{argmin } P = \text{argmin } p = \frac{\mathbb{E}(\sup_{g \in G} \langle t_0, g \cdot Y \rangle)}{\|t_0\|^2}.$$

Now, we know that:

$$\begin{cases} F(t_0) & = f(1) \quad \geqslant \quad p(1) \\ F(\lambda(t_0)t_0) & = f(\lambda(t_0)) \quad \leqslant \quad P(\lambda(t_0)) \end{cases}.$$

thus if $p(1) > P(\lambda(t_0))$, we have $F(t_0) > F(\lambda(t_0)t_0)$. So the only thing we have to prove is $p(1) > P(\lambda(t_0))$:

$$P(\lambda(t_0)) = \lambda(t_0)^2 \|t_0\|^2 - 2\lambda(t_0)\|t_0\|^2 \lambda(t_0) + A\mathbb{E}(\|Y\|^2) + B, \tag{5.10}$$

$$p(1) = \|t_0\|^2 - 2\|t_0\|^2 \lambda(t_0) + a\mathbb{E}(\|Y\|^2) + b. \tag{5.11}$$

Now we can compute $p(1) - P(\lambda(t_0))$, thanks to equations (5.10) and (5.11):

$$p(1) - P(\lambda(t_0)) = [\|t_0\|(\lambda(t_0) - 1)]^2 - \left[(A - a)\mathbb{E}(\|Y\|^2) + B - b\right] > 0,$$

by using Condition (5.9). This concludes that t_0 does not minimize F. □

5.5.2 Is it possible to fulfill condition (5.9)?

As we have seen, we need condition (5.9) in order to prove that $p(1) > P(\lambda(t_0))$ which suffices to prove that t_0 do not minimizes F. But is it possible to have this condition? First, we can notice that when the action is isometric we can always take $A = a = 1$ and $B = b = 0$ then as soon as $t_0 \neq 0$ and $\lambda(t_0) \neq 1$ (and we have seen that $\lambda(t_0) > 1$) we have the condition (5.9), since the left member of condition (5.9) is positive whereas its right member is equal to zero.

So in fact, in the case of isometric action we just prove again what we proved in chapter 3. However let us suppose that we have a isometric which satisfies (5.9) then the left hand side is strictly greater than the right hand side. Now if it was possible to deform our group action on a continuous way, such that the left and right member are continuous with respect to the group action, then starting from an isometric group action we can deform a little this group action to have a non isometric group action which still satisfies (5.9).

However, we need to specify what is meant by deforming our group action in a continuous way.

Definition 5.3. *We suppose that for all $\tau \in [0, 1]$ we have G_τ a group acting on M. Moreover we suppose that G_0 acts isometrically on M. We say that the group action is continuous with respect to τ if for all $\tau_0 \in [0, 1]$ we have:*

$$\lim_{\tau \to \tau_0} \sup_{g_\tau \in G_{\tau_0}} \inf_{g_{\tau_0} \in G_{\tau_0}} \sup_{x \in M \setminus \{0\}} \frac{\|g_\tau \cdot x - g_{\tau_0} \cdot x\|}{\|x\|} = 0. \tag{5.12}$$

This means that for τ sufficiently close to τ_0, $g_\tau \cdot x$ will behave like a $g_{\tau_0} \cdot x$ for some $g_{\tau_0} \in G_{\tau_0}$. Note that *a priori* we do not know how to find this g_{τ_0}.

In particular, if G_τ acts linearly for all τ, condition (5.12) is equivalent to:

$$\forall \tau_0 \in [0, 1] \quad \lim_{\tau \to \tau_0} \sup_{g_\tau \in G_\tau} \inf_{g_{\tau_0} \in G_{\tau_0}} |||g_\tau - g_{\tau_0}||| = 0, \tag{5.13}$$

where $|||\ \ |||$ is the subordinate norm associated to the euclidean norm $\|\ \ \|$. Let us give an example:

Example 5.6. *We take $M = \mathbb{R}^n$, let be G a compact group of the linear group $GL_n(\mathbb{R})$ which linearly. We assume additionally that G do not acts isometrically. In other words G is not included in $O_n(\mathbb{R})$ the orthogonal group. A classical result [Bourbaki 2012] on maximal compact subgroup states that it exists S a symmetric definite positive matrix such that:*

$$S^{-1} G S \subset O_n(\mathbb{R}).$$

Now, the space of symmetric and definite positive matrices is path connected. Therefore we can find $\tau \in [0,1] \mapsto S_\tau$ such that $S_0 = S$ and $S_1 = I_n$. We define then $G_\tau = S_\tau^{-1} G S_\tau$. In particular, $G_0 = S^{-1}GS \subset O_n(\mathbb{R})$ and $G_1 = G$. For every $\tau \in [0,1]$ G_τ acts linearly on M. It is possible that, for some τ, G_τ is included in $O_n(\mathbb{R})$ but we can remove these τ in order that G_0 is the only group which acts isometrically: We define $\tau_\star \in [0,1($ as:

$$\tau_\star = \sup \{t \in [0,1] \text{ such that } G_t \subset O_n(\mathbb{R})\}.$$

Then by the fact that $O_n(\mathbb{R})$ is a closed set we have that $G_{\tau_\star} \subset O_n(\mathbb{R})$. Then we can consider G_τ only for $[\tau_\star, 1]$ and re-scale the τ parameter so that $\tau_\star = 0$ then G_τ acts linearly but not isometrically as soon as $\tau > 0$.

Now let us prove that equation (5.13) is satisfied:

Proof. Let $g_\tau = S_\tau^{-1} g S_\tau \in G_\tau$ with $g \in G$. We have ${S_\tau}^{-1} g S_\tau \in G_{\tau_0}$ then we have:

$$\begin{aligned}
\inf_{g_{\tau_0} \in G_{\tau_0}} |||g_\tau - g_{\tau_0}||| &\leqslant |||g_\tau - {S_{\tau_0}}^{-1} g S_{\tau_0}||| \\
&\leqslant |||{S_\tau}^{-1} g (S_\tau - S_{\tau_0})||| + |||({S_\tau}^{-1} - {S_{\tau_0}}^{-1}) g S_{\tau_0}||| \\
\sup_{g_\tau \in G_\tau} \inf_{g_{\tau_0} \in G_{\tau_0}} |||g_\tau - g_{\tau_0}||| &\leqslant |||{S_\tau}^{-1}||| N |||S_\tau - S_{\tau_0}||| + |||{S_\tau}^{-1} - {S_{\tau_0}}^{-1}||| N |||S_{\tau_0}||| \underset{\tau \to \tau_0}{\to} 0.
\end{aligned}$$

because it exists N such that for all $g \in G$ $|||g||| \leqslant N$ (the group is compact). □

Now, if one have a continuous path of group action $(G_\tau, \cdot)_{\tau \in [0,1]}$ such that G_0 acts isometrically on M then (5.9) is a inequality between two terms which are continuous maps with respect of τ. When (5.9) is fulfilled for $\tau = 0$ (in chapter 3 we have seen that this condition is fulfilled as soon as the quotient distance to the template is contracted compared to the ambient distance), then this condition (5.9) will be still fulfilled for $\tau \in [0, c]$ for some $c > 0$ by continuity. Besides, if the condition $\lambda(t_0) > 1$ is true for $\tau = 0$ (the isometric action), $\lambda(t_0) > 1$ for a small interval $[0, c']$.

Conclusion: if is possible to connect a non isometric action to an isometric one with a continuous path and if the template estimation is inconsistent for the isometric group, then it will also be inconsistent for a few non isometric groups among the path connecting the two groups. The inconsistency appears even for small σ and for non isometric groups. However, we do not know how to estimate the $c \in (0,1]$ such that the action of G_τ leads to inconsistency for $\tau < c$.

5.6 Conclusion

In section 5.2, we have found an implicit equation of an element which minimizes the variance/pre-variance. This implicit equation was used in order to prove that the element which minimize the variance in Hilbert space have less symmetry than the original template in theorem 5.2. In section 5.3, we have provided a conjecture of inconsistency. This conjecture is based on the Voronoï cell, this concept is similar to the notion of the cone of the template seen in chapter 3. It is important to notice that, we do neither suppose an invariant distance under the group action nor that the ambient space is a Hilbert space in this conjecture,. Besides, in this conjecture, we could add a regularization term in the template estimation. Finally, in section 5.5, we do not assume a isometric action, and we give a condition which leaves to inconsistency. Unfortunately, this condition is not easy to verify. However, we can state that when the action is sufficiently closed to an isometric one, then inconsistency holds.

Chapter 6

Conclusion and Perspectives

Contents

6.1 Synthesis of contributions

6.1.1 Hypotheses leading to inconsistency

In this thesis, we have considered the estimation of the template in computational anatomy with the Fréchet mean in quotient space in the case where data in the ambient space were deformed and noisy. We have mathematically proved that the presence of the noise and the presence of the deformation on data leads to an inconsistent estimator of the template. In chapter 3 we proved the inconsistency mostly for isometric action in Hilbert space. Chapter 4 generalized the inconsistency in more general action in Hilbert space, but we restrict ourselves on the case where the noise level was high. We extended our results in chapter 5 for nonlinear spaces, where we proved in theorem 5.2 that the Fréchet mean in quotient space can be more noisy than the original template. Table 6.1, summaries of the main assumptions of the different results leading to inconsistency.

We have highlighted the origin of the inconsistency: it is the contraction of the quotient distance with respect to the distance in the ambient space (see theorems 3.1, 3.4 and 4.1 for instance). We have also provided a quantification of the consistency bias (theorem 4.1 for instance). As a result, the consistency bias is asymptotically linear with respect to the noise level σ, when $\sigma \to +\infty$ in the case of an isometric action in Hilbert spaces.

Besides, we have done a little bit more than providing quantification of the consistency bias. We have proved that the estimated element is less symmetric than the original template (theorem 5.2).

Results	Property of the action	Generative Model	Estimation	Note
Theorem 3.1	isometric	B	B/F	in HS, finite group, template is a regular point
Theorem 3.4	isometric	B	B/F	in HS template is a non fixed point
Theorem 3.6	invariant distance	B	B/F	in CRM, template is a fixed point
Theorem 4.1	isometric action	F	B/F	in HS, Taylor expansion of the consistency bias
Proposition 4.6	subgroup acting isometrically	F	F	in HS, when the noise level is large enough
Proposition 4.7	linear action	F	F	in HS, when the noise level is large enough
Theorem 5.2	invariant distance	general model	B	in HS or CRM
Proposition 5.5	inconsistency when the action is closed enough to an isometric one.	both model	B	in HS

Table 6.1: Recapitulation of the main hypotheses in the different results on this thesis which leads to inconsistency. B stands for backward, F stands for forward, HS for Hilbert space, CRM for complete Riemannian manifold.

In this thesis, we have divided the proofs of inconsistency in several results with all different hypothesis on the action or on the template. On figure 6.1 we draw a Venn diagram of some properties of the action with a few examples that we have seen in this thesis. One important property was the fact that the distance was invariant under the group action, a particular case was the isometric action in Hilbert spaces. The most advanced results was for isometric action in Hilbert spaces. Even if we were able to prove inconsistency for non isometric action, but for a large enough noise level (propositions 4.6 and 4.7).

6.1.2 The role played by the constants

One important constant in the consistency bias is the noise level σ. Indeed, in theorem 4.1 (for isometric action in Hilbert space), we have seen that the consistency bias was asymptotically linear with respect to σ. Besides for non isometric action, we were able to show inconsistency as soon as the noise level was above a certain threshold (propositions 4.6 and 4.7).

But σ is not the only constant which appears in the study of the consistency: there were also the constants K, $\lambda(t_0)$, $\theta(m_\star)$ and $\theta(t_0)$. Let us make a brief overview

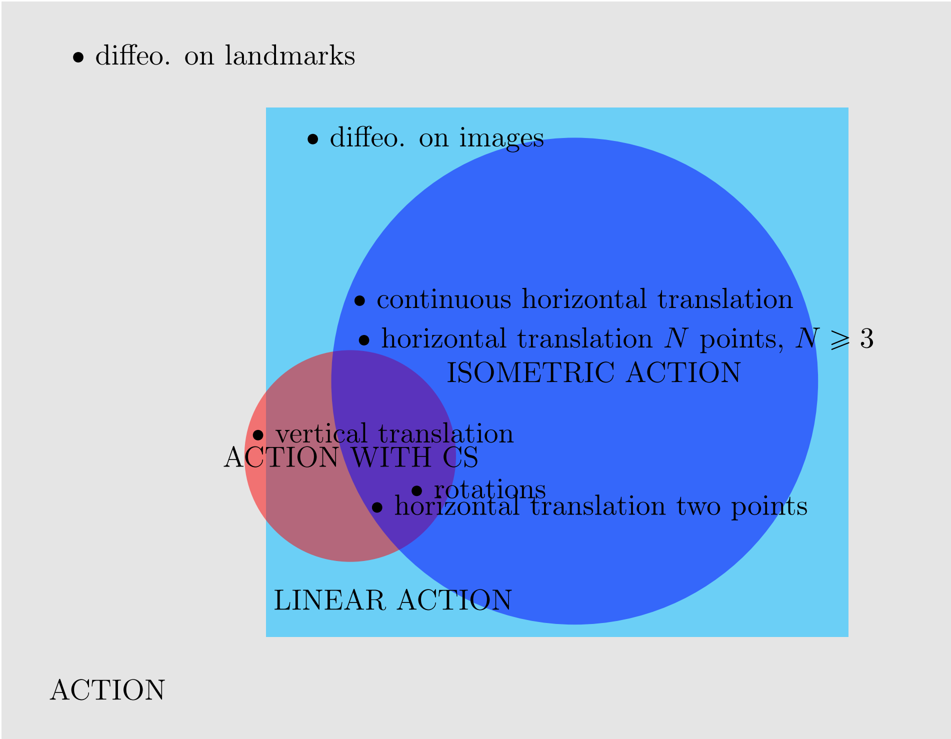


Figure 6.1: Different kind of actions according to their property, the actions in bullet are examples of each cases. The definition and the existence of such congruent section (CS) was discussed in section 5.4.1, isometric actions were studied in chapters 3 and 4, linear actions in section 4.3.4. The examples of the diffeomorphism on landmarks or images were described in section 1.3.3, the vertical translation in example 4.6, the horizontal translation on two points in section 3.3.6, the horizontal translation on more than two points in examples 3.1 and 5.2, the continuous horizontal translation in examples 3.3 and 5.3, rotations in section 3.4.6.3 and example 5.1

of the significance of these constants and how they are related to each other:

- In theorem 3.5, we have found a lower bound of the consistency bias which depended on $\lambda(t_0)$. We recall, that $\lambda(m) = \frac{1}{\|m\|^2}\mathbb{E}(\sup_{g\in G} \langle m, g\cdot Y\rangle)$. This quantity indicates how much the observable variable Y looks like the template t_0 after registration. Later, in proposition 3.7, we take a Taylor expansion of the lower bound found in theorem 3.5 when σ tends to infinity. This Taylor expansion was linear with respect to σ. And the linearity constants was depending on $\theta(t_0)$ defined as $\theta(t_0) = \frac{1}{\|t_0\|}\mathbb{E}(\sup_{g\in G} \langle \varepsilon, g\cdot t_0\rangle)$. This quantity indicates how much the template t_0 looks like the standardized noise ε after registration. In proposition 3.8, we found an upper bound of the consistency bias, this upper bound depends on $\theta(m_\star) = \frac{1}{\|m_\star\|}\mathbb{E}(\sup_{g\in G} \langle \varepsilon, g\cdot m_\star\rangle)$, where $m_\star$ is a minimizer of the variance.

- Therefore, in chapter 3, we have seen that we could have lower and upper bounds of the consistency bias. However the Taylor expansion of these two bounds were different (since they depends on two different constants $\theta(t_0)$ and $\theta(m_\star)$), this prevent us to find a Taylor expansion of the consistency bias. This flaw is fixed in chapter 4, as we found in the proof of theorem 4.1 new lower and upper bounds. These bounds are also asymptotically linear with respect to σ, but with the same constant $K = \sup_{v \in S} \theta(v)$, where S is the unit sphere, and $\theta(v) = \mathbb{E}(\sup_{g \in G} \langle v, g \cdot \varepsilon \rangle)$. Therefore K is bigger than $\theta(t_0)$ and $\theta(m_\star)$. Therefore we were able to establish a Taylor expansion of the consistency bias for isometric action with respect to σ when σ tends to infinity: $\mathrm{CB} = K\sigma + o(\sigma)$. This result is complementary with the one of [Miolane 2017] which provides a Taylor expansion of CB when $\sigma \to 0$.

- In propositions 4.6 and 4.7, we found that the threshold for the noise level, depends on $\theta(t_0)$, besides the proofs make $\lambda(t_0)$ intervenes.

- Finally, in proposition 5.5, we have also give a sufficient condition for inconsistency where $\lambda(t_0)$ also appeared.

6.2 Questions to be investigated

In order to conclude this thesis, we give some questions which need to be solved. We can classify the questions to be investigated into two parts: on the one hand, there are statistically issues directly related to the template estimation with the Fréchet mean in quotient space. On the other hand, there are also geometrical questions, which can be solved independently from the template estimation. But whose solutions would bring new answers to the template estimation issue.

6.2.1 Statistical questions on the template estimation

- What is the behaviour of the consistency bias for $\sigma \to 0$ in Hilbert space for isometric action or in infinite dimensional manifolds?

- How can we prove that the inconsistency appears for all σ for non isometric action? Solving conjectures 5.1 and 5.2, for instance?

- What is the behaviour of the consistency bias with respect to σ when a regularization term is added? Solving this issue could tell which regularization minimizes the inconsistency.

- In all this thesis, the consistency bias was studied under the condition that the Fréchet mean in quotient space exists. If not, the consistency is, by convention, infinite. Can we given some results ensuring such existence? Indeed, the theorems on the existence or the uniqueness of the Fréchet means [Kendall 1989, Karcher 1977] are all, (to the best of our knowledge), restricted to random variable with a support included in a ball which is small enough. Therefore these interesting results applies only in the case of small noise level.

6.2.2 Geometrical questions

- Given a metric space and a group action on this metric space, is it possible to obtain a criteria which indicates the existence (or the non existence) of a congruent section? Indeed, we have seen in chapter 5, than when such a congruent section exists, computing the bias is just a computation of an expectation.

- Is it possible to establish that there exists a measurable section of the quotient space which takes value in the Voronoï cell of a certain point? This would make the proof of theorem 5.2 perfectly rigorous.

- Can we find *spaces with good mixtures* (defined in definition 5.1) different from Hilbert space? This question may matter since we have seen that theorem 5.2 can be directly extended to this kind of space.

Chapter 7

Conclusion et Perspectives (français)

Contents

7.1 Synthèse des contributions

Dans cette thèse, nous avons étudié l'estimation de template en anatomie computationnelle avec la moyenne de Fréchet dans les quotients dans le cas où les données ont été déformées et bruitées. Mathématiquement, nous avons prouvé que la présence de ce bruit et des déformations conduit à l'inconsistance de l'estimateur du template.

Dans le chapitre 3, nous avons montré l'inconsistance pour les actions isométriques dans les espaces de Hilbert. Dans le chapitre 4, nous avons généralisé l'inconsistance pour des actions plus générales mais seulement quand le niveau de bruit était suffisamment grand.

Dans le chapitre 5 nous avons étendu cette étude dans des espaces non linéaires. Dans le théorème 5.2, nous avons prouvé que la moyenne de Fréchet dans l'espace quotient était plus bruitée que le template original. Dans le tableau 7.1, nous avons résumé les différents résultats qui conduisent à l'inconsistance.

Nous avons de plus mis en valeur l'origine du bruit : c'est la contraction de la distance quotient par rapport à la distance dans l'espace ambiant. Par exemple dans les théorèmes 3.1, 3.4, et 4.1. Nous avons aussi fourni une quantification du biais de consistance (théorème 4.1 par exemple) : Le biais de consistance est asymptotiquement linéaire par rapport au niveau de bruit σ quand $\sigma \to +\infty$ dans le cas d'une action isométrique dans un espace de Hilbert.

De plus, nous avons fait un peu plus que fournir une quantification du biais de consistance. Nous avons prouvé que l'élément estimé est moins symétrique que le template original (théorème 5.2).

Résultat	Type d'action	Modèle Generatif	Estimation	Note
Théorème 3.1	isométrique	B	B/F	dans les EH, groupe fini, template est un point régulier
Théorème 3.4	isométrique	B	B/F	in EH, template n'est pas un point fixe
Théorème 3.6	distance invariante	B	B/F	in VRC, template est un point fixe
Théorème 4.1	action isométrique	F	B/F	in EH, équivalent du biais quand $\sigma \to +\infty$
Proposition 4.6	sous-groupe agissant isométriquement	F	F	dans EH quand le niveau de bruit est grand
Proposition 4.7	action linéaire	F	F	dans EH quand le niveau de bruit est grand
Théorème 5.2	distance invariante	modèle général	B	dans EH ou VRC
Proposition 5.5	inconsistance quand l'action est suffisamment isométrique.	B/F	B	dans EH

Table 7.1: Récapitulatif des hypothèses principales dans les différents résultats de cette thèse qui conduisent à l'inconsistance. Avec les abréviations suivantes, B : *backward*, F : *forward*, EH : espace de Hilbert, VRC : variété riemannienne complète.

Dans cette thèse, on a divisé les preuves d'inconsistances en plusieurs résultats avec différentes hypothèses sur l'action ou sur le template. Sur la figure 7.1, on a dessiné un diagramme de Venn de certaines propriétés de l'action avec quelques exemples vus dans cette thèse.

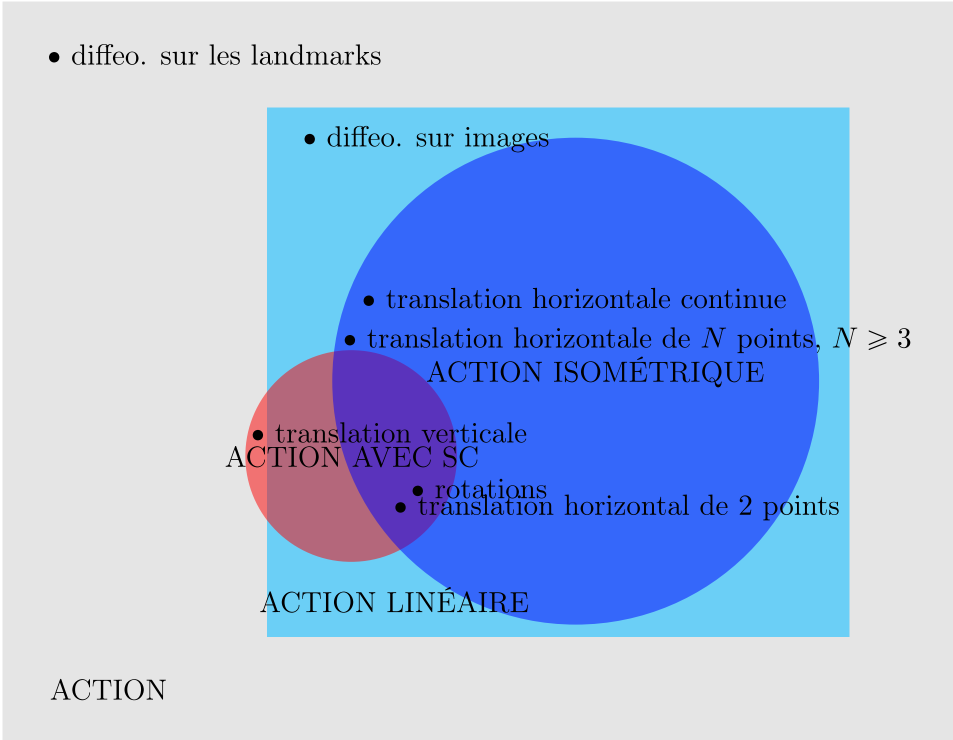


Figure 7.1: Différents types d'actions selon leurs propriétés. Les actions représentées par des points sont des exemples de chaque case. La définition et l'existence des sections congruentes (SC) a été discuté dans la sous-section 5.4.1, les actions isométriques ont été étudiées aux chapitres 3 et 4, les actions linéaires en sous-section 4.3.4.

7.2 Questions ouvertes

Pour conclure cette thèse, on peut donner quelques questions ouvertes concernant les problèmes statistiques et géométriques.

7.2.1 Questions statistiques reliées à l'estimation de template

- Quel est le comportement du biais de consistance lorsque $\sigma \to 0$ dans des espaces de Hilbert pour des actions isométriques ou dans des variétés de dimension infinie ?
- Comment prouver l'inconsistance pour tout $\sigma > 0$ et pour toute action isométrique. En résolvant les conjectures 5.1 et 5.2, par exemple ?
- Quel est le comportement du biais quand un terme de régularisation est ajouté ? Ainsi on pourrait dire quel type de régularisation il est préférable de considérer.

- Dans toute cette thèse, le biais de consistance a été étudié à la condition que la moyenne de Fréchet dans l'espace quotient existait. Sinon le biais de consistance est par convention infinie. Peut-on assurer cette existence ? En effet les théorèmes d'existence et d'unicité de la moyenne de Fréchet [Kendall 1989, Karcher 1977] sont tous restreints (nous semble-t'il) à des variables aléatoires dont le support est contenu dans une boule suffisamment petite. Donc, ces résultats ne s'appliquent que dans le cas des niveaux de faible bruit.

7.2.2 Questions géométriques

- Pour un certain espace métrique et une certaine action de groupe, est-ce possible de trouver un critère indiquant l'existence (ou la non-existence) d'une section congruente ? En effet, au chapitre 5, nous avons vu que lorsqu'une section congruente existe, le calcul du biais est réduit au calcul d'une espérance.

- Est-il possible de montrer qu'il existe une section mesurable de l'espace quotient dont l'image est incluse dans la cellule de Voronoï d'un certain point ? Cela simplifierait le théorème 5.2.

- Peut-on trouver des *espaces à bons mélanges* (définition 5.1) différents des espaces de Hilbert ? En effet, nous avons vu que le théorème 5.2 peut être directement étendu à ce type d'espaces.

Chapter 8

Résumé détaillé (français)

Contents

8.1 Cadre de la thèse

Le but de cette thèse est de démontrer l'inconsistence de l'estimation de template en calculant la moyenne de Fréchet dans l'espace quotient. Pour cela on se place dans un espace de départ noté M. On suppose souvent que M est un espace de Hilbert. On considère G un groupe agissant sur M. On définit alors l'orbite d'un élément $m \in M$ par :

$$[m] = \{g \cdot m, \quad g \in G\}$$

De plus on définit le quotient de M par G noté M/G comme l'ensemble des orbites.

Si on suppose que la distance d_M de M est invariante sous l'action de G, c'est-à-dire que :

$$\forall g \in G \quad \forall m \in M \quad \forall n \in M \quad d_M(g \cdot m, g \cdot n) = d_M(m, n)$$

Alors on peut munir d_Q d'une (pseudo)-distance définie par $d_Q([m], [n]) = \inf_{g \in G} d_M(a, g \cdot b)$.

Soit t_0 un élément de M, que l'on nomme template, partant de ce template on crée des données en déformant ce template par l'action de groupe et en ajoutant du bruit. Suivant que l'on ajoute le bruit avant ou après la déformation on obtient deux modèles génératifs :

$$Y = \Phi \cdot t_0 + \sigma\varepsilon$$

ou

$$Y = \Phi \cdot (t_0 + \sigma\varepsilon)$$

Où on a noté t_0 le template, Φ une variable aléatoire dans G, ε une variable aléatoire centrée réduite ($\mathbb{E}(\varepsilon) = 0$ et $\mathbb{E}(\|\varepsilon\|^2) = 1$, on suppose également que Φ et ε sont deux variables aléatoires indépendantes.

Le premier modèle est appelé «*forward*» et le second «*backward*». Une fois que l'on a généré Y, on s'intéresse au problème inverse : on cherche à estimer le template, ou plus précisément on cherche à estimer son orbite. Une méthode souvent utilisée est le calcul de la moyenne de Fréchet dans l'espace quotient, c'est-à-dire que l'on minimise :

$$m \mapsto F(m) = \mathbb{E}(d_Q([m], [Y])^2) = \mathbb{E}(\inf_{g \in G} \|m - g \cdot Y\|^2)$$

On obtient donc un estimateur du template. Dans cette thèse, on étudie donc les propriétés statistiques de cet estimateur.

Dans cette thèse, on va montrer que généralement cet estimateur est inconsistant. De plus on cherche à estimer le biais de consistance, c'est-à-dire la distance entre le template et la moyenne de Fréchet dans le quotient.

8.2 Simulation sur des données synthétiques

Avant d'étudier de manière théorique l'estimation de template, on peut essayer de faire des expériences numériques pour se donner une intuition de la (in)consistance de l'estimation du template. Pour ce faire, on prend par exemple des signaux vus comme des fonctions discrétisées de $[0, 1]$ dans $\mathbb{R}$ avec N points. On définit une action de groupe par translation :

$$\tau \in \mathbb{Z}/N\mathbb{Z}, (x_1, \dots, x_n) \in \mathbb{R}^n \mapsto (x_{1+\tau}, \dots, x_{n+\tau})$$

Partant d'un template t_0 fixé, on crée un échantillon $Y_1, \dots Y_I$ en agissant sur ce template et en y ajoutant du bruit; $Y_k = \Phi_k \cdot t_0 + \varepsilon_k$. Ainsi avec cet échantillon, on peut définir la variance empirique :

$$m \mapsto F_I(m) = \frac{1}{I}\sum_{i=1}^{I} d_Q([m], [Y_i])$$

Puis en minimisant cette variance (via un algorithme de minimisation alternée appelé «*max-max*»), on peut comparer la moyenne de Fréchet empirique au vrai template :

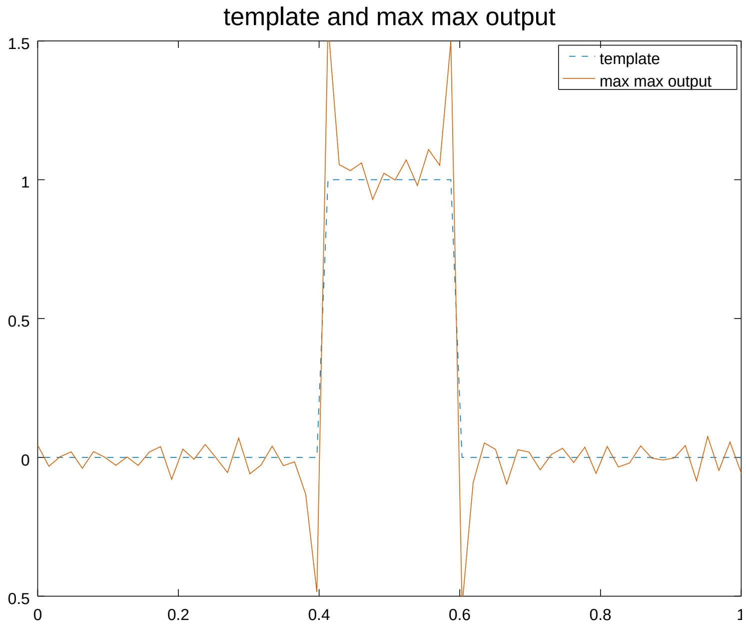


Figure 8.1: Exemple d'un template (une fonction escalier) et du template estimé avec un échantillon de taille 10^5 dans $\mathbb{R}^{64}$, ε est un bruit gaussien et $\sigma = 10$.

8.3 Preuve d'inconsistance pour des actions isométriques

On appelle action isométrique sur un espace de Hilbert M, toute action d'un groupe G tel que pour tout $g \in G$, $m \mapsto g \cdot m$ soit linéaire et que $\|g \cdot m\| = \|m\|$ pour tout $m \in M$.

Pour étudier la consistance on peut commencer par se restreindre aux groupes finis. On a le résultat suivant :

Théorème 8.1. *Soit G un groupe fini agissant sur $M = \mathbb{R}^n$ effectivement et isométriquement. Supposons que la variable aléatoire X est absolument continue par rapport à la mesure de Lebesgue. On suppose aussi que $\mathbb{E}(\|X\|^2) < +\infty$ et que $t_0 = \mathbb{E}(X)$ est un point régulier (son groupe d'isotropie est réduit à $\{0\}$).*

On définit $Cone(t_0)$ comme l'ensemble des points plus proches de t_0 que des autres points gt_0 pour $g \in G$ (voir figure 8.2). $Cone(t_0)$ est en fait définit comme l'ensemble des points déjà recalés avec t_0. Si:

$$\mathbb{P}\left(X \notin Cone(t_0)\right) > 0, \tag{8.1}$$

alors $[t_0]$ n'est pas une moyenne de Fréchet de $[X]$.

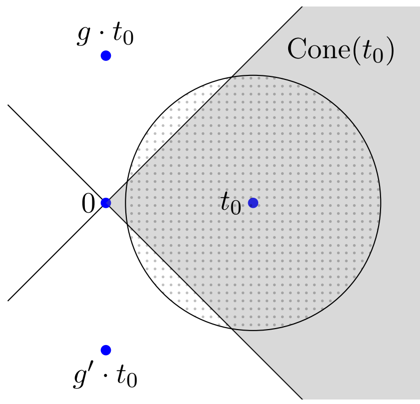


Figure 8.2: Cone(t_0) en gris.

Théorème 8.2. *Soit G un groupe agissant isométriquement sur un espace de Hilbert M. Soit X une variable aléatoire dans M telle que $\mathbb{E}(\|X\|^2) < +\infty$. On suppose que $t_0 = \mathbb{E}(X) \neq 0$. Si:*

$$\mathbb{P}\left(d_Q([t_0],[X]) < \|t_0 - X\|\right) > 0 \tag{8.2}$$

ou de manière équivalente

$$\mathbb{P}\left(\sup_{g \in G} \langle g \cdot X, t_0 \rangle > \langle X, t_0 \rangle\right) > 0 \tag{8.3}$$

Alors $[t_0]$ n'est pas une moyenne de Fréchet de $[X]$ dans $Q = M/G$.

De plus, on peut prouver que la condition 8.2 est souvent vérifiée. Par exemple, lorsque l'orbite du template est une sous variété de M, il suffit que le support de X ne soit pas inclus dans l'orthogonal de l'espace tangent de $[t_0]$ au point t_0.

8.4 Quantification du biais pour les actions isométriques

Dans cette thèse, on prouve également le théorème suivant qui fournit un équivalent du biais lorsque le niveau de bruit diverge vers $+\infty$:

Théorème 8.3. *Soit M un espace de Hilbert, et G agissant isométriquement sur M. Soit $Y = \Phi \cdot t_0 + \sigma\varepsilon$, t_0 est le template, Φ une variable aléatoire dans G, ε une variable aléatoire centrée réduite ($\mathbb{E}(\varepsilon) = 0$ et $\mathbb{E}(\|\varepsilon\|^2) = 1$. Si le support du bruit ε n'est pas inclus dans l'ensemble des points fixes, alors on a l'encadrement du biais de consistance suivant :*

$$\sigma K - 2\|t_0\| \leqslant CB \leqslant \sigma K + 2\|t_0\|, \tag{8.4}$$

où $K = \sup_{\|v\|=1} \mathbb{E}\left(\sup_{g\in G} \langle v, g\cdot\varepsilon\rangle\right) \in (0,1]$. La constante K dépend seulement du bruit standardisé et de l'action de groupe mais pas du template. On obtient alors l'équivalent du biais suivant quand σ tend vers plus l'infini :

$$CB = \sigma K + o(\sigma) \text{ as } \sigma \to +\infty. \tag{8.5}$$

8.5 Actions non isométriques

Lorsque les actions sont non isométriques, les preuves des théorèmes vues précédemment ne s'appliquent pas. Cependant, si on suppose que l'orbite du template est bornée, on peut montrer l'inconsistance dès que le niveau de bruit est suffisamment grand. On prouve, en particulier, la proposition suivante :

Proposition 8.1. *Soit G un groupe agissant sur M un espace de Hilbert. On suppose que le template t_0 n'est pas un point fixe, et que son orbite sous l'action du groupe G est bornée. On suppose que G contient un sous-groupe H agissant isométriquement. On considère $A \geqslant \sup_{g\in G} \frac{\|g\cdot t_0\|}{\|t_0\|}$ et $a \leqslant \inf_{g\in G} \frac{\|g\cdot t_0\|}{\|t_0\|}$. Notons que $a \leqslant 1 \leqslant A$ et on a :*

$$\forall g \in G \quad a\|t_0\| \leqslant \|g\cdot t_0\| \leqslant A\|t_0\|.$$

On pose :

$$\theta(t_0) = \frac{1}{\|t_0\|}\mathbb{E}(\sup_{g\in G} \langle g\cdot t_0, \varepsilon\rangle) \text{ et } \theta_H = \frac{1}{\|t_0\|}\mathbb{E}\left(\sup_{h\in H} \langle h\cdot t_0, \varepsilon\rangle\right).$$

On suppose que $\theta_H > 0$. Si σ est plus grand qu'un certain niveau de bruit critique noté σ_c défini par :

$$\sigma_c = \frac{\|t_0\|}{\theta_H}\left[\left(\frac{\theta(t_0)}{\theta_H} + A\right) + \sqrt{\left(\frac{\theta(t_0)}{\theta_H} + A\right)^2 + A^2 - a^2}\right]. \tag{8.6}$$

Alors l'estimation est inconsistante.

8.6 Variation du groupe d'isotropie

Un autre résultat de cette thèse, est la preuve que sous certaines conditions le groupe d'isotropie de la moyenne de Fréchet dans le quotient est plus petit que le template. Ce théorème est prouvé dans les espaces de Hilbert ainsi que les variétés riemanniennes complètes lorsque la distance est invariante sous l'action de groupe, ce résultat est similaire à celui de [Huckemann 2012].

Cela prouve non seulement une nouvelle fois l'inconsistance mais aussi que le template et la moyenne de Fréchet dans le quotient n'ont pas les mêmes propriétés géométriques.

Bibliography

[Afsari 2011] Bijan Afsari. *Riemannian L^p center of mass: existence, uniqueness, and convexity.* Proceedings of the American Mathematical Society, vol. 139, no. 2, pages 655–673, 2011. (Cited on pages 26 and 37.)

[Allassonnière 2007] Stéphanie Allassonnière, Yali Amit and Alain Trouvé. *Towards a coherent statistical framework for dense deformable template estimation.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 69, no. 1, pages 3–29, 2007. (Cited on pages 42, 45, 88, 99 and 100.)

[Allassonnière 2010] Stéphanie Allassonnière, Estelle Kuhn and Alain Trouvé. *Construction of Bayesian deformable models via a stochastic approximation algorithm: a convergence study.* Bernoulli, vol. 16, no. 3, pages 641–678, 2010. (Cited on pages 42 and 81.)

[Allassonnière 2015] Stéphanie Allassonnière, Loïc Devilliers and Xavier Pennec. *Estimating the Template in the Total Space with the Fréchet Mean on Quotient Spaces may have a Bias: a Case Study on Vector Spaces Quotiented by the Group of Translations.* In Mathematical Foundations of Computational Anatomy (MFCA'15), 2015. (Authors given by alphabetic order, LD is in fact the prinicpal author). (Cited on pages 31, 39 and 41.)

[Allassonnière 2015] S. Allassonnière, S. Durrleman and E. Kuhn. *Bayesian Mixed Effect Atlas Estimation with a Diffeomorphic Deformation Model.* SIAM Journal on Imaging Sciences, vol. 8, no. 3, pages 1367–1395, 2015. (Cited on pages 17 and 35.)

[Arnaudon 2005] Marc Arnaudon and Xue-Mei Li. *Barycenters of measures transported by stochastic flows.* Annals of probability, pages 1509–1543, 2005. (Cited on pages 26 and 37.)

[Arnaudon 2014] Marc Arnaudon and Laurent Miclo. *Means in complete manifolds: uniqueness and approximation.* ESAIM: Probability and Statistics, vol. 18, pages 185–206, 2014. (Cited on pages 26 and 37.)

[Beg 2005] M Faisal Beg, Michael Miller, Alain Trouvé and Laurent Younes. *Computing large deformation metric mappings via geodesic flows of diffeomorphisms.* International journal of computer vision, vol. 61, no. 2, pages 139–157, 2005. (Cited on pages 17, 35, 117 and 136.)

[Bhattacharya 2003] Rabi Bhattacharya and Vic Patrangenaru. *Large sample theory of intrinsic and extrinsic sample means on manifolds.* The Annals of Statistics, vol. 31, no. 1, pages 1–29, 2003. (Cited on pages 27 and 38.)

[Bhattacharya 2005] Rabi Bhattacharya and Vic Patrangenaru. *Large sample theory of intrinsic and extrinsic sample means on manifolds: II.* Annals of statistics, pages 1225–1259, 2005. (Cited on pages 27 and 38.)

[Bhattacharya 2008] Abhishek Bhattacharya and Rabi Bhattacharya. *Statistics on Riemannian manifolds: asymptotic distribution and curvature.* Proceedings of the American Mathematical Society, vol. 136, no. 8, pages 2959–2967, 2008. (Cited on pages 27, 38, 71, 98 and 118.)

[Bigot 2011] Jérémie Bigot and Benjamin Charlier. *On the consistency of Fréchet means in deformable models for curve and image analysis.* Electronic Journal of Statistics, vol. 5, pages 1054–1089, 2011. (Cited on pages 30, 38, 42, 84 and 88.)

[Bigot 2013] Jérémie Bigot. *Fréchet means of curves for signal averaging and application to ECG data analysis.* The Annals of Applied Statistics, vol. 7, no. 4, pages 2384–2401, 2013. (Cited on page 20.)

[Bontemps 2014] Dominique Bontemps and Sébastien Gadat. *Bayesian methods for the Shape Invariant Model.* Electronic Journal of Statistics, vol. 8, no. 1, pages 1522–1568, 2014. (Cited on page 42.)

[Bookstein 1986] Fred L Bookstein. *Size and shape spaces for landmark data in two dimensions.* Statistical science, pages 181–222, 1986. (Cited on page 20.)

[Bourbaki 2012] Nicolas Bourbaki. Algebre: Chapitre 8. Springer Science & Business Media, 2012. (Cited on page 147.)

[Box 1976] George EP Box. *Science and statistics.* Journal of the American Statistical Association, vol. 71, no. 356, pages 791–799, 1976. (Cited on page 30.)

[Charlier 2013] Benjamin Charlier. *Necessary and sufficient condition for the existence of a Fréchet mean on the circle.* ESAIM: Probability and Statistics, vol. 17, pages 635–649, 2013. (Cited on pages 26 and 37.)

[Charon 2013] Nicolas Charon and Alain Trouvé. *The varifold representation of nonoriented shapes for diffeomorphic registration.* SIAM Journal on Imaging Sciences, vol. 6, no. 4, pages 2547–2580, 2013. (Cited on page 21.)

[Cheng 2016] Wen Cheng, Ian L Dryden and Xianzheng Huang. *Bayesian registration of functions and curves.* Bayesian Analysis, vol. 11, no. 2, pages 447–475, 2016. (Cited on page 118.)

[Cleveland 2016] Jason Cleveland, Wei Wu and Anuj Srivastava. *Norm-preserving constraint in the Fisher–Rao registration and its application in signal estimation.* Journal of Nonparametric Statistics, vol. 28, no. 2, pages 338–359, 2016. (Cited on page 46.)

[Cooley 1965] James Cooley and John Tukey. *An algorithm for the machine calculation of complex Fourier series.* Mathematics of computation, vol. 19, no. 90, pages 297–301, 1965. (Cited on page 99.)

[Cootes 2004] Timothy F Cootes, Stephen Marsland, Carole J Twining, Kate Smith and Christopher J Taylor. *Groupwise diffeomorphic non-rigid registration for automatic model building.* In European conference on computer vision, pages 316–327. Springer, 2004. (Cited on page 84.)

[Delyon 2000] Bernard Delyon. *Stochastic approximation with decreasing gain: Convergence and asymptotic theory.* Unpublished lecture notes, Université de Rennes, 2000. (Cited on page 27.)

[Devilliers 2017a] Loïc Devilliers, Xavier Pennec and Stéphanie Allassonnière. Inconsistency of template estimation with the fréchet mean in quotient space, pages 16–27. Springer International Publishing, Cham, 2017. (Cited on pages 32, 39 and 84.)

[Devilliers 2017b] Loïc Devilliers, Stéphanie Allassonnière, Alain Trouvé and Xavier Pennec. *Inconsistency of Template Estimation by Minimizing of the Variance/Pre-Variance in the Quotient Space.* Entropy, vol. 19, no. 6, 2017. (Cited on pages 32, 39 and 84.)

[Devilliers 2017c] Loïc Devilliers, Stéphanie Allassonnière, Alain Trouvé and Xavier Pennec. *Template Estimation in Computational Anatomy: Fréchet Means Top and Quotient Spaces Are Not Consistent.* SIAM Journal on Imaging Sciences, vol. 10, no. 3, pages 1139–1169, 2017. (Cited on pages 31, 39, 41 and 84.)

[Dirichlet 1850] G Lejeune Dirichlet. *Über die Reduction der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen.* Journal für die reine und angewandte Mathematik, vol. 40, pages 209–227, 1850. (Cited on page 49.)

[do Carmo Valero 1992] Manfredo Perdigao do Carmo Valero. Riemannian geometry. Birkhäuser, 1992. (Cited on page 18.)

[Du 2014] Jia Du, A. Pasha Hosseinbor, Moo K. Chung, Barbara B. Bendlin, Gaurav Suryawanshi, Andrew L. Alexander and Anqi Qiu. *Diffeomorphic metric mapping and probabilistic atlas generation of hybrid diffusion imaging based on BFOR signal basis.* Medical Image Analysis, vol. 18, no. 7, pages 1002 – 1014, 2014. (Cited on page 122.)

[Durrleman 2009] Stanley Durrleman, Xavier Pennec, Alain Trouvé and Nicholas Ayache. *Statistical models of sets of curves and surfaces based on currents.* Medical image analysis, vol. 13, no. 5, pages 793–808, 2009. (Cited on pages 21 and 30.)

[Durrleman 2013] Stanley Durrleman, Stéphanie Allassonnière and Sarang Joshi. *Sparse adaptive parameterization of variability in image ensembles*. International Journal of Computer Vision, vol. 101, no. 1, pages 161–183, 2013. (Cited on pages 17 and 35.)

[Durrleman 2014] Stanley Durrleman, Marcel Prastawa, Nicolas Charon, Julie Korenberg, Sarang Joshi, Guido Gerig and Alain Trouvé. *Morphometry of anatomical shape complexes with dense deformations and sparse parameters*. NeuroImage, vol. 101, pages 35–49, 2014. (Cited on pages 17, 21, 35, 41, 42, 85 and 117.)

[Émery 1991] Michel Émery and Gabriel Mokobodzki. *Sur le barycentre d'une probabilité dans une variété*. In Séminaire de probabilités XXV, pages 220–233. Springer, 1991. (Cited on pages 26 and 37.)

[Fletcher 2003] Thomas Fletcher, Conglin Lu and Sarang Joshi. *Statistics of shape via principal geodesic analysis on Lie groups*. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 1, pages I–I. IEEE, 2003. (Cited on pages 17 and 35.)

[Fletcher 2004] Thomas Fletcher, Conglin Lu, Stephen M Pizer and Sarang Joshi. *Principal geodesic analysis for the study of nonlinear statistics of shape*. IEEE transactions on medical imaging, vol. 23, no. 8, pages 995–1005, 2004. (Cited on pages 17 and 35.)

[Fréchet 1948] Maurice Fréchet. *Les elements aléatoires de nature quelconque dans un espace distancié*. In Annales de l'institut Henri Poincaré, volume 10, pages 215–310, 1948. (Cited on pages 17, 35 and 41.)

[Glasbey 2001] CA Glasbey and KV Mardia. *A penalized likelihood approach to image warping*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 63, no. 3, pages 465–492, 2001. (Cited on page 122.)

[Glaunes 2006] Joan Alexis Glaunes and Sarang Joshi. *Template estimation form unlabeled point set data and surfaces for computational anatomy*. In 1st MICCAI Workshop on Mathematical Foundations of Computational Anatomy: Geometrical, Statistical and Registration Methods for Modeling Biological Shape Variability, 2006. (Cited on page 122.)

[Goodall 1991] Colin Goodall. *Procrustes methods in the statistical analysis of shape*. Journal of the Royal Statistical Society. Series B (Methodological), pages 285–339, 1991. (Cited on pages 85 and 96.)

[Gower 1975] John C Gower. *Generalized procrustes analysis*. Psychometrika, vol. 40, no. 1, pages 33–51, 1975. (Cited on page 96.)

[Grenander 1998] Ulf Grenander and Michael Miller. *Computational Anatomy: An Emerging Discipline*. Q. Appl. Math., vol. LVI, no. 4, pages 617–694, December 1998. (Cited on pages 16, 29, 34 and 41.)

[Guimond 2000] Alexandre Guimond, Jean Meunier and Jean-Philippe Thirion. *Average brain models: A convergence study*. Computer vision and image understanding, vol. 77, no. 2, pages 192–210, 2000. (Cited on pages 17, 35 and 84.)

[Hitziger 2013] Sebastian Hitziger, Maureen Clerc, Alexandre Gramfort, Sandrine Saillet, Christian Bénar and Théodore Papadopoulo. *Jitter-adaptive dictionary learning-application to multi-trial neuroelectric signals*. arXiv preprint arXiv:1301.3611, 2013. (Cited on pages 20, 42, 88 and 99.)

[Huckemann 2010] Stephan Huckemann, Thomas Hotz and Axel Munk. *Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric lie group actions*. Statistica Sinica, pages 1–58, 2010. (Cited on pages 17, 30, 35 and 38.)

[Huckemann 2011] Stephan Huckemann. *Inference on 3d procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models*. Scandinavian Journal of Statistics, vol. 38, no. 3, pages 424–446, 2011. (Cited on pages 27, 30, 38 and 85.)

[Huckemann 2012] Stephan Huckemann. *On the meaning of mean shape: manifold stability, locus and the two sample test*. Annals of the Institute of Statistical Mathematics, vol. 64, no. 6, pages 1227–1259, 2012. (Cited on pages 30, 38, 131 and 166.)

[Joshi 2000] Sarang Joshi and Michael Miller. *Landmark matching via large deformation diffeomorphisms*. IEEE Transactions on Image Processing, vol. 9, no. 8, pages 1357–1370, 2000. (Cited on page 20.)

[Joshi 2004] Sarang Joshi, Brad Davis, Mathieu Jomier and Guido Gerig. *Unbiased diffeomorphic atlas construction for computational anatomy*. Neuroimage, vol. 23, pages S151–S160, 2004. (Cited on pages 17, 35, 41, 84 and 122.)

[Karcher 1977] Hermann Karcher. *Riemannian center of mass and mollifier smoothing*. Communications on pure and applied mathematics, vol. 30, no. 5, pages 509–541, 1977. (Cited on pages 26, 37, 41, 154 and 160.)

[Kendall 1989] David Kendall. *A survey of the statistical theory of shape*. Statistical Science, pages 87–99, 1989. (Cited on pages 16, 34, 41, 42, 154 and 160.)

[Kendall 1990] Wilfrid Kendall. *Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence*. Proceedings of the London Mathematical Society, vol. 3, no. 2, pages 371–406, 1990. (Cited on pages 26, 37 and 41.)

[Kent 1997] John Kent and Kanti Mardia. *Consistency of Procrustes estimators.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 59, no. 1, pages 281–290, 1997. (Cited on pages 30 and 38.)

[Kullback 1951] Solomon Kullback and Richard Leibler. *On information and sufficiency.* The annals of mathematical statistics, vol. 22, no. 1, pages 79–86, 1951. (Cited on page 108.)

[Kurtek 2011a] Sebastian Kurtek, Eric Klassen, Zhaohua Ding, Malcolm Avison and Anuj Srivastava. *Parameterization-invariant shape statistics and probabilistic classification of anatomical surfaces.* In Biennial International Conference on Information Processing in Medical Imaging, pages 147–158. Springer, 2011. (Cited on page 88.)

[Kurtek 2011b] Sebastian Kurtek, Anuj Srivastava and Wei Wu. *Signal Estimation Under Random Time-Warpings and Nonlinear Signal Alignment.* In J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira and K.Q. Weinberger, editors, Advances in Neural Information Processing Systems 24, pages 675–683. Curran Associates, Inc., 2011. (Cited on pages 41, 77, 84, 85, 88, 92, 94 and 119.)

[Kühnel 2017] Line Kühnel, Stefan Sommer, Akshay Pai and Lars Lau Raket. *Most likely separation of intensity and warping effects in image registration.* SIAM Journal on Imaging Sciences, vol. 10, no. 2, pages 578–601, 2017. (Cited on page 81.)

[Le 1998] Huiling Le. *On the consistency of procrustean mean shapes.* Advances in Applied Probability, vol. 30, no. 1, pages 53–63, 1998. (Cited on pages 30 and 38.)

[Le 2004] Huiling Le. *Estimation of Riemannian barycentres.* LMS J. Comput. Math, vol. 7, no. 193200, page 20, 2004. (Cited on pages 27 and 37.)

[Lefebvre 2012] Sidonie Lefebvre, Stéphanie Allassonnière, Jérémie Jakubowicz, Thomas Lasne and Eric Moulines. *Aircraft classification with a low resolution infrared sensor.* Machine Vision and Applications, vol. 24, no. 1, pages 175–186, 2012. (Cited on page 41.)

[Lele 1993] Subhash Lele. *Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference.* Mathematical Geology, vol. 25, no. 5, pages 573–602, 1993. (Cited on pages 30 and 38.)

[Lombaert 2013] Herve Lombaert, Leo Grady, Xavier Pennec, Nicholas Ayache and Farida Cheriet. *Spectral Log-Demons: Diffeomorphic Image Registration with Very Large Deformations.* International Journal of Computer Vision (IJCV), pages 1–18, 2013. (Cited on pages 17, 35 and 117.)

[Miller 2006] Michael Miller, Alain Trouvé and Laurent Younes. *Geodesic shooting for computational anatomy.* Journal of mathematical imaging and vision, vol. 24, no. 2, pages 209–228, 2006. (Cited on pages 17 and 35.)

[Miolane 2015] Nina Miolane and Xavier Pennec. *Biased estimators on Quotient spaces.* In Geometric Science of Information. Second International Conference, GSI 2015, Palaiseau, France, October 28-30, 2015, Proceedings, volume 9389. Springer, 2015. (Cited on pages 42 and 74.)

[Miolane 2017] Nina Miolane, Susan Holmes and Xavier Pennec. *Template Shape Estimation: Correcting an Asymptotic Bias.* SIAM Journal on Imaging Sciences, vol. 10, no. 2, pages 808–844, 2017. (Cited on pages 30, 38, 42, 59, 73, 84, 85, 88, 93, 118, 138 and 154.)

[Nash 1956] John Nash. *The imbedding problem for Riemannian manifolds.* Annals of mathematics, pages 20–63, 1956. (Cited on page 143.)

[Panaretos 2016] Victor Panaretos and Yoav Zemel. *Amplitude and phase variation of point processes.* The Annals of Statistics, vol. 44, no. 2, pages 771–812, 2016. (Cited on page 85.)

[Pennec 2006] Xavier Pennec. *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements.* Journal of Mathematical Imaging and Vision, vol. 25, no. 1, pages 127–154, 2006. (Cited on pages 26, 28, 37 and 78.)

[Pennec 2016] Xavier Pennec. *Barycentric Subspace Analysis on Manifolds.* arXiv preprint arXiv:1607.02833, 2016. (Cited on pages 17 and 35.)

[Rohé 2016] Marc-Michel Rohé, Maxime Sermesant and Xavier Pennec. *Barycentric Subspace Analysis: a new Symmetric Group-wise Paradigm for Cardiac Motion Tracking.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 300–307. Springer International Publishing, 2016. (Cited on pages 17 and 35.)

[Rohlf 2003] James Rohlf. *Bias and error in estimates of mean shape in geometric morphometrics.* Journal of Human Evolution, vol. 44, no. 6, pages 665–683, 2003. (Cited on page 85.)

[Sabuncu 2008] Mert Sabuncu, Serdar K. Balci and Polina Golland. *Discovering Modes of an Image Population through Mixture Modeling.* Proceeding of the MICCAI conference, vol. LNCS, no. 5242, pages 381–389, 2008. (Cited on page 41.)

[Seiler 2012] Christof Seiler, Xavier Pennec and Mauricio Reyes. *Capturing the multiscale anatomical shape variability with polyaffine transformation trees.* Medical image analysis, vol. 16, no. 7, pages 1371–1384, 2012. (Cited on pages 17 and 35.)

[Sommer 2010] Stefan Sommer, François Lauze, Søren Hauberg and Mads Nielsen. *Manifold valued statistics, exact principal geodesic analysis and the effect of linear approximations.* Computer Vision–ECCV 2010, pages 43–56, 2010. (Cited on pages 17 and 35.)

[Sommer 2014] Stefan Sommer, François Lauze and Mads Nielsen. *Optimization over geodesics for exact principal geodesic analysis.* Advances in Computational Mathematics, vol. 40, no. 2, pages 283–313, 2014. (Cited on pages 17 and 35.)

[Thirion 1998] J-P Thirion. *Image matching as a diffusion process: an analogy with Maxwell's demons.* Medical image analysis, vol. 2, no. 3, pages 243–260, 1998. (Cited on pages 17 and 35.)

[Thompson 1942] Darcy Wentworth Thompson. *On growth and form.* On growth and form., 1942. (Cited on pages 16 and 17.)

[Thurston 1979] William Thurston and John Willard Milnor. The geometry and topology of three-manifolds. Princeton University Princeton, NJ, 1979. (Cited on page 20.)

[Trouvé 1995] Alain Trouvé. *An approach of pattern recognition through infinite dimensional group action.* 1995. (Cited on pages 17 and 35.)

[Vaillant 2005] Marc Vaillant and Joan Glaunès. *Surface matching via currents.* In Information processing in medical imaging, pages 1–5. Springer Berlin/Heidelberg, 2005. (Cited on page 21.)

[Vercauteren 2009] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Diffeomorphic demons: Efficient non-parametric image registration.* NeuroImage, vol. 45, no. 1, pages S61–S72, 2009. (Cited on pages 17 and 35.)

[Voronoi 1908] Georges Voronoi. *Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs.* Journal für die reine und angewandte Mathematik, vol. 134, pages 198–287, 1908. (Cited on page 49.)

[Zhang 2013] Miaomiao Zhang, Nikhil Singh and P.Thomas Fletcher. *Bayesian Estimation of Regularization and Atlas Building in Diffeomorphic Image Registration.* In JamesC. Gee, Sarang Joshi, Kilian. Pohl, William. Wells and Lilla Zöllei, editors, Information Processing in Medical Imaging, volume 7917 of *Lecture Notes in Computer Science*, pages 37–48. Springer Berlin Heidelberg, 2013. (Cited on page 42.)

[Ziezold 1977] Herbert 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 1974 European Meeting of Statisticians, pages 591–602. Springer, 1977. (Cited on pages 27, 38, 71, 88 and 140.)