



**HAL**  
open science

# Template estimation in computational anatomy: Fréchet means in top and quotient spaces are not consistent

Loïc Devilliers, Stéphanie Allasonnière, Alain Trouvé, Xavier Pennec

## ► To cite this version:

Loïc Devilliers, Stéphanie Allasonnière, Alain Trouvé, Xavier Pennec. Template estimation in computational anatomy: Fréchet means in top and quotient spaces are not consistent. *SIAM Journal on Imaging Sciences*, 2017, 10 (3), pp.1139-1169. 10.1137/16M1083931 . hal-01352707v4

**HAL Id: hal-01352707**

**<https://hal.archives-ouvertes.fr/hal-01352707v4>**

Submitted on 25 Apr 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Template estimation in computational anatomy: Fréchet means in top and quotient spaces are not consistent

Loïc Devilliers\*, Stéphanie Allasonnière†, Alain Trouvé‡  
and Xavier Pennec§

April 25, 2017

## Abstract

In this article, 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.

Keyword : Template, Fréchet mean, group action, quotient space, inconsistency, consistency bias, empirical Fréchet mean, Hilbert space, manifold

---

\*Université Côte d'Azur, Inria, France, [loic.devilliers@inria.fr](mailto:loic.devilliers@inria.fr)

†CMAP, Ecole polytechnique, CNRS, Université Paris-Saclay, 91128, Palaiseau, France

‡CMLA, ENS Cachan, CNRS, Université Paris-Saclay, 94235 Cachan, France

§Université Côte d'Azur, Inria, France

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Definitions, notations and generative model</b>	<b>5</b>
<b>3</b>	<b>Inconsistency for finite group when the template is a regular point</b>	<b>8</b>
3.1	Presence of inconsistency . . . . .	9
3.2	Upper bound of the consistency bias . . . . .	12
3.3	Study of the consistency bias in a simple example . . . . .	13
<b>4</b>	<b>Inconsistency for any group when the template is not a fixed point</b>	<b>14</b>
4.1	Presence of an inconsistency . . . . .	14
4.2	Analysis of the condition in theorem 4.1 . . . . .	15
4.3	Lower bound of the consistency bias . . . . .	18
4.4	Upper bound of the consistency bias . . . . .	20
4.5	Empirical Fréchet mean . . . . .	22
4.6	Examples . . . . .	22
4.6.1	Action of translation on $L^2(\mathbb{R}/\mathbb{Z})$ . . . . .	23
4.6.2	Action of discrete translation on $\mathbb{R}^{\mathbb{Z}/\mathbb{N}\mathbb{Z}}$ . . . . .	23
4.6.3	Action of rotations on $\mathbb{R}^n$ . . . . .	23
<b>5</b>	<b>Fréchet means top and quotient spaces are not consistent when the template is a fixed point</b>	<b>25</b>
5.1	Result . . . . .	25
5.2	Proofs of these theorems . . . . .	26
5.2.1	Proof of theorem 5.1 . . . . .	26
5.2.2	Proof of theorem 5.2 . . . . .	28
<b>6</b>	<b>Conclusion and discussion</b>	<b>28</b>
<b>A</b>	<b>Proof of theorems for finite groups' setting</b>	<b>29</b>
A.1	Proof of theorem 3.2: differentiation of the variance in the quotient space . . . . .	29
A.2	Proof of theorem 3.1: the gradient is not zero at the template . . . . .	32
A.3	Proof of theorem 3.3: upper bound of the consistency bias . . . . .	32
A.4	Proof of proposition 3.2: inconsistency in $\mathbb{R}^2$ for the action of translation . . . . .	34
<b>B</b>	<b>Proof of lemma 5.1: differentiation of the variance in the top space</b>	<b>35</b>

# 1 Introduction

In Kendall's shape space theory [Ken89], in computational anatomy [GM98], 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 [DPC<sup>+</sup>14] or an aircraft [LAJ<sup>+</sup>12], an electrical signal of the human body, a MR image etc. To analyse 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 [GM98] 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 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 top space (or ambient space) by the group. In this particular setting, the template estimation is most of the time based on the minimisation of the empirical variance in the quotient space (for instance [KSW11, JDJG04, SBG08] among many others). The points that minimise the empirical variance are called the empirical Fréchet mean. The Fréchet means introduced in [Fré48] is comprised of the elements minimising the variance. This generalises the notion of expected value in non linear 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 [Kar77] and [Ken90]).

Several group actions are used in practice: some signals can be shifted in time compared to other signals (action of translations [HCG<sup>+</sup>13]), landmarks can be transformed rigidly [Ken89], shapes can be deformed by diffeomorphisms [DPC<sup>+</sup>14], 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 5, allows 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 Allasonnière, Amit and Trouvé [AAT07]: they took a step function as a template and they 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, minimising 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 ([AKT10, BG14] or [ZSF13]). But there is a need to have a better understanding of the failure of the template estimation with the Fréchet mean. One can study the inconsistency of the template estimation. Bigot and Charlier [BC11] 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 et al. [MP15, MHP16] 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 article, 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 top space while the article of Miolane et al. is restricted to finite dimensional manifolds. In a preliminary unpublished version of this work [ADP15], we proved the inconsistency when the transformations come from a finite group acting by translation. The current article extends these results by generalizing to any isometric action of finite and non-finite groups.

This article is organised as follows. Section 2 details the mathematical terms that we use and the generative model. In sections 3 and 4, we exhibit 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 is dedicated to finite groups. Then we generalise our result in section 4 to non-finite groups. To complete this study, we extend in section 5 the result when the template is a fixed point under the group action and when the top 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.

## 2 Definitions, notations and generative model

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

$$\begin{aligned} G \times M &\rightarrow M \\ (g, m) &\mapsto g \cdot m \end{aligned}$$

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 article we use these two ways. We project an element  $m$  of the top 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 top 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:

$$d_Q([a], [b]) = \inf_{g \in G} d_M(g \cdot a, b). \quad (1)$$

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

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

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 article, 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 article, except in section 5, we

suppose that the the group acts isometrically on an 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 article for  $M$  a vector space. Let us take a template  $t_0 \in M$  to which we add a unbiased noise  $\epsilon$ :  $X = t_0 + \epsilon$ . Finally we transform  $X$  with a random shift  $S$  of  $G$ . We assume that this variable  $S$  is independent of  $X$  and the only observed variable is:

$$Y = S \cdot X = S \cdot (t_0 + \epsilon), \text{ with } \mathbb{E}(\epsilon) = 0, \quad (2)$$

while  $S$ ,  $X$  and  $\epsilon$  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  $Y' = S \cdot t_0 + \epsilon'$ . But when the noise is isotropic and the action is isometric, one can show that the two models have the same law, since  $S \cdot \epsilon$  and  $\epsilon$  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 (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  $S$  by  $Sg^{-1}$  and the noise  $\epsilon$  by  $g\epsilon$ , 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  $Y = S \cdot X$  and the noisy template  $X$  are different random variables in the top space, their projections on the quotient space lead to the same random orbit  $[Y] = [X]$ . That is why we consider the generative model (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:

$$Q \ni [m] \mapsto \mathbb{E}(d_Q([m], [X])^2) \quad (3)$$

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

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

$$Q \ni [m] \mapsto \frac{1}{n} \sum_{i=1}^n d_Q([m], [Y_i])^2 = \frac{1}{n} \sum_{i=1}^n \inf_{g_i \in G} \|m - g_i \cdot Y_i\|^2. \quad (4)$$

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 minimise the empirical quotient variance (4), the max-max algorithm<sup>1</sup> alternatively minimises the function  $J(m, (g_i)_i) = \frac{1}{n} \sum_{i=1}^n \|m - g_i \cdot Y_i\|^2$  over a point  $m$  of the orbit  $[m]$  and over the hidden transformation  $(g_i)_{1 \leq i \leq n} \in G^n$ .

With these notations we can reformulate our questions as:

1. Is the orbit of the template  $[t_0]$  a minimiser of the quotient variance defined in (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 article shows that the answer to the first question is usually "no" in the framework of an Hilbert space  $M$  on which a group  $G$  acts linearly and isometrically. The only exception is theorem 5.1 where the top 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 top space. This is the subgroup which leaves this point unchanged:

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

We start in section 3 with the simple example where the group is finite and the isotropy group of the template is reduced to the identity element ( $\text{Iso}(t_0) = \{e_G\}$ , in this case  $t_0$  is called a regular point). We turn in section 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 5 that the template  $t_0$  is a fixed point which means that  $\text{Iso}(t_0) = G$ .

In sections 3 and 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 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 top 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 top 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:  $\sigma^2 = \mathbb{E}(\|t_0 - X\|^2) = E(t_0)$ .

---

<sup>1</sup>The term max-max algorithm is used for instance in [AAT07], and we prefer to keep the same name, even if it is a minimisation.



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 top 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 and 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 top 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. [CWS16] 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 top space is:

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

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  $\mathbb{E}(e(t)e(s)) = \sigma^2 \mathbf{1}_{s=t}$ , 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.

### 3 Inconsistency for finite group when the template is a regular point

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  $\| \cdot \|$ , associated to the dot product  $\langle \cdot, \cdot \rangle$ .

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 appendix A.1).

**Example 3.1.** *The action of translation on coordinates: 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  $m : \mathbb{T} \rightarrow \mathbb{R}$ , where  $m(\tau)$  is the grey value at pixel  $\tau$ . 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:*

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

*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:

$$F : m \in M \mapsto \mathbb{E}(d_Q([X], [m])^2). \quad (5)$$

In this framework,  $F$  is non-negative and continuous. Thanks to Cauchy-Schwarz inequality we have:

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

Thus for some  $R > 0$  we have: for all  $m \in M$  if  $\|m\| > R$  then  $F(m) \geq F(0) + 1$ . The closed ball  $B(0, R)$  is a compact set (because  $M$  is a finite vector space) then  $F$  restricted to this ball reached its minimum  $m_*$ . Then for all  $m \in M$ , if  $m \in B(0, R)$ ,  $F(m_*) \leq F(m)$ , if  $\|m\| > R$  then  $F(m) \geq F(0) + 1 > F(0) \geq F(m_*)$ . Therefore  $[m_*]$  is a Fréchet mean of  $[X]$  in the quotient  $Q = M/G$ . Note that this ensure the existence but not the 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.

### 3.1 Presence of 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.*

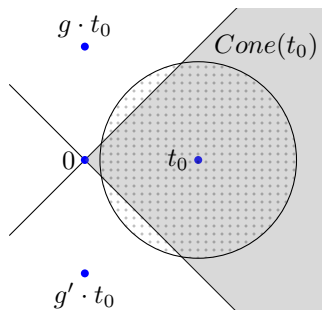


Figure 1: Planar representation of a part of the orbit of the template  $t_0$ . The lines are the hyperplanes whose points are equally distant of two distinct elements of the orbit of  $t_0$ ,  $Cone(t_0)$  represented in points is the set of points closer from  $t_0$  than any other points in the orbit 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.

We define  $Cone(t_0)$  as the set of points closer from  $t_0$  than any other points of the orbit  $[t_0]$ , see fig. 1 or item 6 in appendix A.1 for a formal definition. In other words,  $Cone(t_0)$  is defined as the set of points already registered with  $t_0$ . Suppose that:

$$\mathbb{P}(X \notin Cone(t_0)) > 0, \quad (6)$$

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

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  $g \in G$  minimising  $\|m_0 - g \cdot x\|$  (in other words,  $g(x, m_0) \cdot x \in Cone(m_0)$ ). This allows us to compute the gradient of  $F$  at  $m_0$ :*

$$\nabla F(m_0) = 2(m_0 - \mathbb{E}(g(X, m_0) \cdot X)). \quad (7)$$

This Theorem is proved in appendix A.1. 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 fig. 2: on fig. 2a there is a mass under the two hyperplanes outside

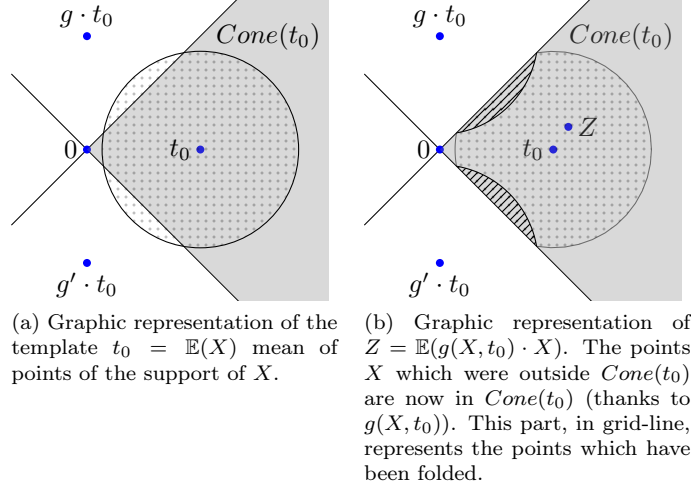


Figure 2:  $Z$  is the mean of points in  $Cone(t_0)$  where  $Cone(t_0)$  is the set of points closer from  $t_0$  than  $g \cdot t_0$  for  $g \in G \setminus e_G$ . Therefore it seems that  $Z$  is higher than  $t_0$ , therefore  $\nabla F(t_0) = 2(t_0 - Z) \neq 0$ .

$Cone(t_0)$ , so this mass is nearer from  $gt_0$  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 grid-line on fig. 2. 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 appendix A.2.

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 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 do 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 fig. 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 3.2, 3.3, 4.3 and 4.4 investigate how far this hypothesis is true.

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 it happens that condition (6) is necessary and sufficient. More generally the following proposition provides a partial converse:

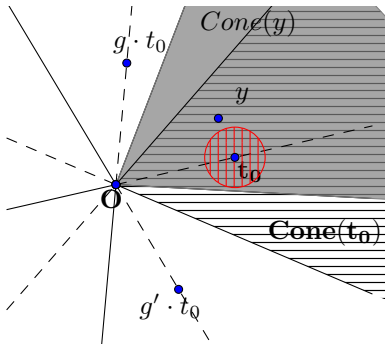


Figure 3:  $y \mapsto Cone(y)$  is continuous. When the support of the  $X$  is bounded and included in the interior of  $Cone(t_0)$  the hatched cone. For  $y$  sufficiently close to the template  $t_0$ , the support of the  $X$  (the ball in red) is still included in  $Cone(y)$  (in grey), then  $F(y) = (\mathbb{E}(\|X - y\|^2))$ . Therefore in this case,  $[t_0]$  is at least a Karcher mean of  $[X]$ .

**Proposition 3.1.** *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).*

*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 top space. Now by continuity of the distance map (see fig. 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 top 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.  $\square$

## 3.2 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:  $X \sim \mathcal{N}(t_0, s^2 Id_{\mathbb{R}^n})$ , and  $m_\star \in \operatorname{argmin} F$ , then we have the upper bound of the consistency bias:*

$$d_Q([t_0], [m_\star]) \leq s\sqrt{8 \log(|G|)}. \quad (8)$$

The proof is postponed in appendix A.3. When  $X \sim \mathcal{N}(t_0, s^2 Id_n)$  the variability of  $X$  is  $\sigma^2 = \mathbb{E}(\|X - t_0\|^2) = ns^2$  and we can write the upper bound of the bias:  $d_Q([t_0], [m_\star]) \leq \frac{\sigma}{\sqrt{n}} \sqrt{8 \log |G|}$ . This Theorem shows that the consistency bias is low when the variability of  $X$  is small, which tends to confirm our hypothesis in section 3.1. It is important to notice that this upper bound explodes when the cardinal of the group tends to infinity.

### 3.3 Study of the consistency bias in a simple example

In this Subsection, we take a particular case of example 3.1: the action of 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 proved in appendix A.4:

**Proposition 3.2.** *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:  $X \sim \mathcal{N}(t_0, s^2 Id_2)$ , 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  $d = \operatorname{dist}(t_0, L)$  given by:*

$$\tilde{\rho}(d, s) = s \frac{2}{\pi} \int_{\frac{d}{s}}^{+\infty} r^2 \exp\left(-\frac{r^2}{2}\right) g\left(\frac{d}{rs}\right) dr, \quad (9)$$

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

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

$$\tilde{\rho}(d, s) \underset{s \rightarrow \infty}{\sim} s \frac{2}{\pi} \int_0^{+\infty} r^2 \exp\left(-\frac{r^2}{2}\right) dr. \quad (10)$$

- (b) If  $d > 0$ , then  $\tilde{\rho}(d, s) = o(s^k)$  when  $s \rightarrow 0$ , for all  $k \in \mathbb{N}$ .  
(c)  $s \mapsto \tilde{\rho}(0, s)$  is linear with respect to  $s$  (for  $d = 0$  the template is a fixed point).

**Remark 3.1.** Here, contrarily to the case of the action of rotation in [MHP16], it is not the ratio  $\|\mathbb{E}(X)\|$  over the noise which matters to estimate the consistency bias. Rather the ratio  $\text{dist}(\mathbb{E}(X), L)$  over the noise. However in both cases we measure the distance between the signal and the singularities which was  $\{0\}$  in [MHP16] for the action of rotations,  $L$  in this case.

## 4 Inconsistency for any group when the template is not a fixed point

In section 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 an Hilbert space with its dot product noted by  $\langle \cdot, \cdot \rangle$  and its associated norm  $\|\cdot\|$ . 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 4.1.** *The action of continuous translation:* We take  $G = (\mathbb{R}/\mathbb{Z})^D$  acting on  $M = \mathbb{L}^2((\mathbb{R}/\mathbb{Z})^D, \mathbb{R})$  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  $D$ .*

### 4.1 Presence of an inconsistency

We state here a generalization of theorem 3.1:

**Theorem 4.1.** *Let  $G$  be a group acting isometrically on  $M$  an Hilbert space, and  $X$  a random variable in  $M$ ,  $\mathbb{E}(\|X\|^2) < +\infty$  and  $\mathbb{E}(X) = t_0 \neq 0$ . If:*

$$\mathbb{P}(d_Q([t_0], [X]) < \|t_0 - X\|) > 0, \tag{11}$$

*or equivalently:*

$$\mathbb{P}\left(\sup_{g \in G} \langle g \cdot X, t_0 \rangle > \langle X, t_0 \rangle\right) > 0. \tag{12}$$

*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  $gt_0$  for some  $g$  than  $t_0$ . Thus the condition (12) 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 top 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 (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  $a \in \mathbb{R}^+$   $f(a) = F(at_0) = \mathbb{E}(\inf_{g \in G} \|g \cdot X - a\|^2)$ . Thanks to the isometric action we can expand  $f(a)$  by:

$$f(a) = a^2 \|t_0\|^2 - 2a \mathbb{E} \left( \sup_{g \in G} \langle g \cdot X, t_0 \rangle \right) + \mathbb{E}(\|X\|^2), \quad (13)$$

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

$$a_\star = \frac{\mathbb{E} \left( \sup_{g \in G} \langle g \cdot X, t_0 \rangle \right)}{\|t_0\|^2}. \quad (14)$$

For all  $x \in M$ , we have  $\sup_{g \in G} \langle g \cdot x, t_0 \rangle \geq \langle x, t_0 \rangle$  and thanks to condition (12) we get:

$$\mathbb{E}(\sup_{g \in G} \langle g \cdot X, t_0 \rangle) > \mathbb{E}(\langle X, t_0 \rangle) = \langle \mathbb{E}(X), t_0 \rangle = \|t_0\|^2, \quad (15)$$

which implies  $a_\star > 1$ . Then  $F(a_\star t_0) < F(t_0)$ .  $\square$

Note that  $\|t_0\|^2(a_\star - 1) = \mathbb{E}(\sup_{g \in G} \langle g \cdot X, t_0 \rangle) - \mathbb{E}(\langle X, t_0 \rangle)$  (which is positive) is exactly  $-\langle \nabla F(t_0), t_0 \rangle / 2$  in the case of finite group, see Equation (44). Here we find the same expression without having to differentiate the variance  $F$ , which may be not possible in the current setting.

## 4.2 Analysis of the condition in theorem 4.1

We now look for general cases when we are sure that Equation (12) holds which implies the presence of inconsistency. We saw in section 3 that when the group was 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 4.1 cannot be applied are not generic cases.



First we can notice that it is not possible to have the condition (12) if  $t_0$  is a fixed point under the action of  $G$ . Indeed in this case  $\langle g \cdot X, t_0 \rangle = \langle X, g^{-1}t_0 \rangle = \langle X, t_0 \rangle$ . So from now, we suppose that  $t_0$  is not a fixed point. Now let us see some settings when we have the condition (11) and thus condition (12).

**Proposition 4.1.** *Let  $G$  be a group acting isometrically on an 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.  $[t_0] \setminus \{t_0\}$  is a dense set in  $[t_0]$ .
2. There exists  $\eta > 0$  such that the support of  $X$  contains a ball  $B(t_0, \eta)$ .

Then condition (12) holds, and the estimator is inconsistent according to theorem 4.1.

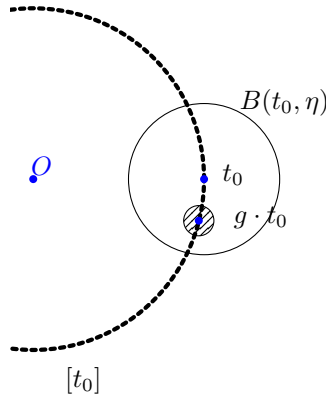


Figure 4: The smallest disk is included in the support of  $X$  and the points in that disk is closer from  $g \cdot t_0$  than from  $t_0$ . According to theorem 4.1 there is an inconsistency.

*Proof.* By density, one takes  $g \cdot t_0 \in B(t_0, \eta) \setminus \{t_0\}$  for some  $g \in G$ , now if we take  $r < \min(\|g \cdot t_0 - t_0\|/2, \eta - \|g \cdot t_0 - t_0\|)$  then  $B(g \cdot t_0, r) \subset B(t_0, \eta)$ . Therefore by the assumption we made on the support one has  $\mathbb{P}(X \in B(g \cdot t_0, r)) > 0$ . For  $y \in B(g \cdot t_0, r)$  we have that  $\|gt_0 - y\| < \|t_0 - y\|$  (see fig. 4). Then we have:  $\mathbb{P}(d_Q([X], [t_0]) < \|X - t_0\|) \geq \mathbb{P}(X \in B(g \cdot t_0, r)) > 0$ . Then we verify condition (12), and we can apply theorem 4.1.  $\square$

Proposition 4.1 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 4.1 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:

**Proposition 4.2.** *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.  $\varphi : (-a, a) \rightarrow [t_0]$  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 (12) is fulfilled, which leads to an inconsistency thanks to Theorem 4.1.*

*Proof.* Thanks to the isometric action:  $\langle t_0, v \rangle = 0$ . We choose  $y \notin v^\perp$  in the support of  $X$  and make a Taylor expansion of the following square distance (see also Figure 5) 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:  $\exists x_* \in (-a, a)$  s.t.  $\|x_*\| < a$ ,  $x \langle y, v \rangle > 0$  and  $\|\varphi(x_*) - y\| < \|t_0 - y\|$ . For some  $g \in G$ ,  $\varphi(x_*) = g \cdot t_0$ . By continuity of the norm we have:

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

Then  $\mathbb{P}(\|g \cdot t_0 - X\| < \|t_0 - X\|) \geq \mathbb{P}(X \in B(y, r)) > 0$ . Theorem 4.1 applies.  $\square$

Proposition 4.2 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:

**Proposition 4.3.** *Let  $G$  be a group acting isometrically on an Hilbert space  $M$ ,  $X$  a random variable in  $M$ , with  $\mathbb{E}(\|X\|^2) < +\infty$ . Assume  $X = t_0 + \sigma\epsilon$ , where  $t_0 \neq 0$  and  $\mathbb{E}(\epsilon) = 0$ , and  $\mathbb{E}(\|\epsilon\|) = 1$ . 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:*

$$\mathbb{P}(X \notin T_{t_0}[t_0]^\perp) > 0, \tag{16}$$

*which is equivalent to:*

$$\mathbb{P}(\epsilon \notin T_{t_0}[t_0]^\perp) > 0, \tag{17}$$

*then there is an inconsistency.*

*Proof.* First  $t_0 \perp T_{t_0}[t_0]$  (because the action is isometric),  $T_{t_0}[t_0]^\perp = t_0 + T_{t_0}[t_0]^\perp$ , then the event  $\{X \in T_{t_0}[t_0]^\perp\}$  is equal to  $\{\epsilon \in T_{t_0}[t_0]^\perp\}$ . This proves that equations (16) and (17) are equivalent. Thanks to assumption (16), we can choose  $y$  in the support of  $X$  such that  $y \notin T_{t_0}[t_0]^\perp$ . Let us take  $v \in T_{t_0}[t_0]$  such that  $\langle y, v \rangle \neq 0$  and choose  $\varphi$  a  $\mathcal{C}^1$  curve in  $[t_0]$ , such that  $\varphi(0) = t_0$  and  $\varphi'(0) = v$ . Applying proposition 4.2 we get the inconsistency.  $\square$

Note that Condition (16) is very weak, because  $T_{t_0}[t_0]$  is a strict linear subspace of  $M$ .

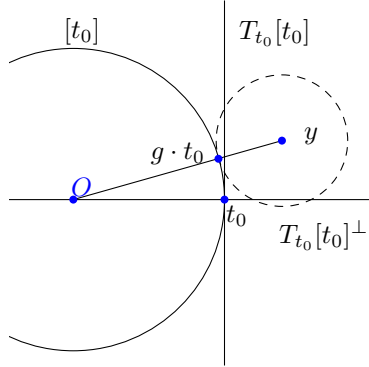


Figure 5:  $y \notin T_{t_0}[t_0]^\perp$  therefore  $y$  is closer from  $g \cdot t_0$  for some  $g \in G$  than  $t_0$  itself. In conclusion, if  $y$  is in the support of  $X$ , there is an inconsistency.

### 4.3 Lower bound of the consistency bias

Under the assumption of Theorem 4.1, we have an element  $a_\star t_0$  such that  $F(a_\star t_0) < F(t_0)$  where  $F$  is the variance of  $[X]$ . From this element, we deduce lower bounds of the consistency bias:

**Theorem 4.2.** *Let  $\delta$  be the unique positive solution of the following equation:*

$$\delta^2 + 2\delta (\|t_0\| + \mathbb{E}\|X\|) - \|t_0\|^2(a_\star - 1)^2 = 0. \quad (18)$$

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

$$\delta^2 + 2\delta\|t_0\| \left(1 + \sqrt{1 + \sigma^2/\|t_0\|^2}\right) - \|t_0\|^2(a_\star - 1)^2 = 0, \quad (19)$$

where  $\sigma^2 = \mathbb{E}(\|X - t_0\|^2)$  is the variability of  $X$ . Then  $\delta$  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  $a_\star t_0$ , where  $a_\star$  was defined in Equation (14): thanks to the expansion of the function  $f$  we did in (13) we get :

$$F(t_0) - F(a_\star t_0) = \|t_0\|^2(a_\star - 1)^2 > 0, \quad (20)$$

Moreover we can show (exactly like equation (43)) that for all  $x \in M$ :

$$\begin{aligned} |F(t_0) - F(x)| &\leq \mathbb{E} \left( \left| \inf_{g \in G} \|g \cdot X - t_0\|^2 - \inf_{g \in G} \|g \cdot X - x\|^2 \right| \right) \\ &\leq \|x - t_0\| (2\|t_0\| + \|x - t_0\| + \mathbb{E}(\|2X\|)). \end{aligned} \quad (21)$$

With Equations (20) and (21), for all  $x \in B(t_0, \delta)$  we have  $F(x) > F(a_\star t_0)$ . No point in that ball mapped in the quotient space is a Fréchet mean of  $[X]$ . So

$\delta$  is a lower bound of the consistency bias. Now by using the fact that  $\mathbb{E}(\|X\|) \leq \sqrt{\|t_0\|^2 + \sigma^2}$ , we get:  $2|F(t_0) - F(x)| \leq 2\|x - t_0\| \times \|t_0\| \left(1 + \sqrt{1 + \sigma^2/\|t_0\|^2}\right) + \|x - t_0\|^2$ . This proves that  $\delta_*$  is also a lower bound of the consistency bias.  $\square$

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

**Proposition 4.4.** *Under the hypotheses of Theorem 4.2, we write  $X = t_0 + \sigma\epsilon$ , with  $\mathbb{E}(\epsilon) = 0$ , and  $\mathbb{E}(\|\epsilon\|^2) = 1$  and note  $\nu = \mathbb{E}(\sup_{g \in G} \langle g\epsilon, t_0/\|t_0\| \rangle) \in (0, 1]$ , we have that:*

$$\delta_* \underset{\sigma \rightarrow +\infty}{\sim} \sigma(\sqrt{1 + \nu^2} - 1),$$

In particular, the consistency bias explodes when the variability of  $X$  tends to infinity.

*Proof.* First, let us prove that  $\nu \in (0, 1]$  under the condition (12). We have  $\nu \geq \mathbb{E}(\langle \epsilon, t_0/\|t_0\| \rangle) = 0$ . By a *reductio ad absurdum*: if  $\nu = 0$ , then  $\sup_{g \in G} \langle g\epsilon, t_0 \rangle = \langle \epsilon, t_0 \rangle$  almost surely. We have then almost surely:  $\langle X, t_0 \rangle \leq \sup_{g \in G} \langle gX, t_0 \rangle \leq \|t_0\|^2 + \sup_{g \in G} \sigma \langle g\epsilon, t_0 \rangle = \|t_0\|^2 + \sigma \langle \epsilon, t_0 \rangle \leq \langle X, t_0 \rangle$ , which is in contradiction with (12). Besides  $\nu \leq \mathbb{E}(\|\epsilon\|) \leq \sqrt{\mathbb{E}\|\epsilon\|^2} = 1$

Second, we exhibit equivalent of the terms in equation (19) when  $\sigma \rightarrow +\infty$ :

$$2\|t_0\| \left(1 + \sqrt{1 + \sigma^2/\|t_0\|^2}\right) \sim 2\sigma. \quad (22)$$

Now by definition of  $a_*$  in Equation (14) and the decomposition of  $X = t_0 + \sigma\epsilon$  we get:

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

$$\|t_0\|(a_* - 1) \leq \frac{1}{\|t_0\|} \mathbb{E} \left( \sup_{g \in G} \langle g \cdot \sigma\epsilon, t_0 \rangle \right) = \sigma\nu \quad (23)$$

$$\|t_0\|(a_* - 1) \geq \frac{1}{\|t_0\|} \mathbb{E} \left( \sup_{g \in G} \langle g \cdot \sigma\epsilon, t_0 \rangle \right) - 2\|t_0\| = \sigma\nu - 2\|t_0\|, \quad (24)$$

The lower bound and the upper bound of  $\|t_0\|(a_* - 1)$  found in (23) and (24) are both equivalent to  $\sigma\nu$ , when  $\sigma \rightarrow +\infty$ . Then the constant term of the quadratic Equation (19) has an equivalent:

$$-\|t_0\|^2(a_* - 1)^2 \sim -\sigma^2\nu^2. \quad (25)$$

Finallye if we solve the quadratic Equation (19), we write  $\delta_*$  as a function of the coefficients of the quadratic equation (19). We use the equivalent of each of these terms thanks to equation (22) and (25), this proves proposition 4.4.  $\square$

**Remark 4.1.** Thanks to inequality (24), if  $\frac{\|t_0\|}{\sigma} < \frac{\nu}{2}$ , then  $\|t_0\|^2(1 - a_\star)^2 \geq (\sigma\nu - 2\|t_0\|)^2$ , then we write  $\delta_\star$  as a function of the coefficients of Equation (19), we obtain a lower bound of the inconsistency bias as a function of  $\|t_0\|$ ,  $\sigma$  and  $\nu$  for  $\sigma > 2\|t_0\|/\nu$ :

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

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

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

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  $\text{Fix}(M)$ . Then for  $x \in M$ , we have:  $\text{dist}(x, \text{Fix}(M)) = \|x - p(x)\|$ . Which yields:

$$\langle g\epsilon, t_0 \rangle = \langle g\epsilon, t_0 - p(t_0) \rangle + \langle \epsilon, p(t_0) \rangle. \quad (26)$$

The right hand side of Equation (26) does not depend on  $g$  as  $p(t_0) \in \text{Fix}(M)$ . Then:

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

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

$$\nu \leq \frac{1}{\|t_0\|} \text{dist}(t_0, \text{Fix}(M)) \mathbb{E}(\|\epsilon\|) = \text{dist}(t_0/\|t_0\|, \text{Fix}(M)) \mathbb{E}(\|\epsilon\|). \quad (27)$$

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.

#### 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 4.5.** Let  $X$  be a random variable in  $M$ , such that  $X = t_0 + \sigma\epsilon$  where  $\sigma > 0$ ,  $\mathbb{E}(\epsilon) = 0$  and  $\mathbb{E}(\|\epsilon\|^2) = 1$ . 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  $[X]$ :

$$d_Q([m_\star], [t_0]) \leq \sigma\nu(m_\star - m_0) + \sqrt{\sigma^2\nu(m_\star - m_0)^2 + 2\text{dist}(t_0, \text{Fix}(M))\sigma\nu(m_\star - m_0)}, \quad (28)$$

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

Note that we made no hypothesis on the template in this proposition. We deduce from Equation (28) that  $d_Q([m_\star], [t_0]) \leq \sigma + \sqrt{\sigma^2 + 2\sigma \text{dist}(t_0, \text{Fix}(M))}$  is a  $O(\sigma)$  when  $\sigma \rightarrow \infty$ , but a  $O(\sqrt{\sigma})$  when  $\sigma \rightarrow 0$ , in particular the consistency bias can be neglected when  $\sigma$  is small.

*Proof.* First we have:

$$F(m_\star) \leq F(t_0) = \mathbb{E}(\inf_g \|t_0 - g(t_0 + \sigma\epsilon)\|^2) \leq \mathbb{E}(\|\sigma\epsilon\|^2) = \sigma^2. \quad (29)$$

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

$$\begin{aligned} F(m) &= \mathbb{E}(\inf_g (\|m - gt_0\|^2 + \sigma^2\|\epsilon\|^2 - 2\langle g\sigma\epsilon, m - gt_0 \rangle)) \\ &\geq d_Q([m], [t_0])^2 + \sigma^2 - 2\mathbb{E}(\sup_g \langle \sigma\epsilon, gm \rangle). \end{aligned} \quad (30)$$

With Inequalities (29) and (30) one gets:

$$d_Q([m_\star], [t_0])^2 \leq 2\mathbb{E}(\sup_g \langle \sigma\epsilon, gm_\star \rangle) = 2\sigma\nu(m_\star)\|m_\star\|,$$

note that at this point, if  $m_\star = 0$  then  $\mathbb{E}(\sup_g \langle \sigma\epsilon, gm_\star \rangle) = 0$  and  $\nu(m_\star) = 0$  although Equation (4.4) is still true even if  $m_\star = 0$ . Moreover with the triangular inequality applied at  $[m_\star]$ ,  $[0]$  and  $[t_0]$ , one gets:  $\|m_\star\| \leq \|t_0\| + d_Q([m_\star], [t_0])$  and then:

$$d_Q([m_\star], [t_0])^2 \leq 2\sigma\nu(m_\star)(d_Q([m_\star], [t_0]) + \|t_0\|). \quad (31)$$

We can solve inequality (31) and we get:

$$d_Q([m_\star], [t_0]) \leq \sigma\nu(m_\star) + \sqrt{\sigma^2\nu(m_\star)^2 + 2\|t_0\|\sigma\nu(m_\star)}, \quad (32)$$

We note by  $F_X$  instead of  $F$  the variance in the quotient space of  $[X]$ , and we want to apply inequality (32) to  $X - m_0$ . As  $m_0$  is a fixed point:

$$F_X(m) = \mathbb{E} \left( \inf_{g \in G} \|X - m_0 - g \cdot (m - m_0)\|^2 \right) = F_{X-m_0}(m - m_0)$$

Then  $m_\star$  minimises  $F_X$  if and only if  $m_\star - m_0$  minimises  $F_{X-m_0}$ . We apply Equation (32) 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:

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

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

## 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 [BB08], 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 [Zie77] 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 (18). 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 4.6.** *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  $\delta_n$  is a lower bound of the quotient distance between the orbit of the template and  $[m_{n\star}]$ , where  $\delta_n$  is the unique positive solution of:*

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

$a_{n\star}$  is defined like  $a_\star$  in section 4.1 by:

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

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

The proof is a direct application of theorem 4.2, but applied to the empirical law of  $X$  given by the realization of  $X_1, \dots, X_n$ .

## 4.6 Examples

In this Subsection, we discuss, in some examples, the application of theorem 4.1 and see the behaviour of the constant  $\nu$ . This constant intervened in lower bound of the consistency bias.

#### 4.6.1 Action of 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  $\mathbb{R}f'_0$ . 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 4.3. 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:  $\text{Fix}(M) = \mathbb{R}\mathbb{1}$  and:

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

This distance to the fixed points is used in the upper bound of the constant  $\nu$  in Equation (27). Note that if  $f_0$  is not differentiable, then  $[f_0]$  is not necessarily a manifold, and (4.3) does not apply. However proposition 4.1 does: if  $f_0$  is not a constant function, then  $[f_0] \setminus \{f_0\}$  is dense in  $[f_0]$ . Therefore as soon as the support of  $X$  contains a ball around  $f_0$ , there is an inconsistency.

#### 4.6.2 Action of discrete translation on $\mathbb{R}^{\mathbb{Z}/N\mathbb{Z}}$

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

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

Therefore if we have a sample of size  $I$  of  $\epsilon$  iid, then:

$$\nu = \frac{1}{\|t_0\|} \lim_{I \rightarrow +\infty} \frac{1}{I} \sum_{i=1}^I \max_{\tau_i \in \mathbb{Z}/N\mathbb{Z}} \langle \epsilon_i, \tau_i \cdot t_0 \rangle,$$

By an exhaustive research, we can find the  $\tau_i$ 's which maximise the dot product, then with this sample and  $t_0$  we can approximate  $\nu$ . We have done this approximation for several signals  $t_0$  on fig. 6. According the previous results, the bigger  $\nu$  is, the more important the lower bound of the consistency bias is. We remark that the  $\nu$  estimated is small,  $\nu \ll 1$  for different signals.

#### 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:

---

<sup>2</sup>Indeed  $\varphi : \begin{matrix} ]-\frac{1}{2}, \frac{1}{2}[ & \rightarrow & O \\ t & \mapsto & f_0(\cdot - t) \end{matrix}$  is a local parametrisation of  $O$ :  $f_0 = \varphi(0)$ , and we check that:  $\lim_{x \rightarrow 0} \|\varphi(x) - \varphi(0) - x f'_0\|_{L^2} = 0$  with Taylor-Lagrange inequality at the order 2. As a conclusion  $\varphi$  is differentiable at 0, and it is an immersion (since  $f'_0 \neq 0$ ), and  $D_0\varphi : x \mapsto x f'_0$ , then  $O$  is a manifold of dimension 1 and the tangent space of  $O$  at  $f_0$  is:  $T_{f_0}O = D_0\varphi(\mathbb{R}) = \mathbb{R}f'_0$ .



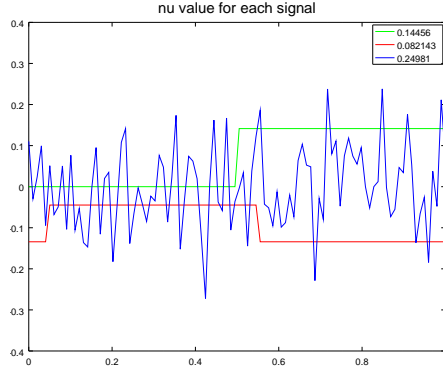


Figure 6: Different signals and their  $\nu$  approximated with a sample of size  $10^3$  in  $\mathbb{R}^{\mathbb{Z}/100\mathbb{Z}}$ .  $\epsilon$  is here a Gaussian noise in  $\mathbb{R}^{\mathbb{Z}/100\mathbb{Z}}$ , such that  $\mathbb{E}(\epsilon) = 0$  and  $\mathbb{E}(\|\epsilon\|^2) = 1$ . For instance the blue signal is a signal defined randomly, and when we approximate the  $\nu$  which corresponds to that  $t_0$  we find  $\simeq 0.25$ .

$X = t_0 + \sqrt{n}s\epsilon$  with  $\mathbb{E}(\epsilon) = 0$  and  $\mathbb{E}(\|\epsilon\|^2) = 1$ . According to proposition 4.4 we have by noting  $\delta_\star$  the lower bound of the consistency bias when  $s \rightarrow \infty$ :

$$\frac{\delta_\star}{s} \rightarrow \sqrt{n}(-1 + \sqrt{1 + \nu^2}).$$

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

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

We compare this result with the exact computation of the consistency bias (noted here  $CB$ ) made by Miolane et al. [MHP16], which writes with our current notations:

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

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

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

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. [MP15].

## 5 Fréchet means top and quotient spaces are not consistent when the template is a fixed point

In this Section, we do not assume that the top 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  $Y = S \cdot X$  is the observed variable where  $S$  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  $G$ .

### 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 5.1.** *Assume that  $M$  is a complete finite dimensional Riemannian manifold and that  $d_M$  is the geodesic distance on  $M$ . 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 5.2.** *Assume that  $M$  is a Hilbert space and that  $d_M$  is given by the Hilbert norm on  $M$ . 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 5.1.** *Theorem 5.2 covers the interesting case of the Fisher Rao metric on functions:*

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

Then considering for  $G$  the group of smooth diffeomorphisms  $\gamma$  on  $[0, 1]$  such that  $\gamma(0) = 0$  and  $\gamma(1) = 1$ , we have a right group action  $G \times \mathcal{F} \rightarrow \mathcal{F}$  given by  $\gamma \cdot f = f \circ \gamma$ . The Fisher Rao metric is built as a pull back metric of the  $L^2([0, 1], \mathbb{R})$  space through the map  $Q : \mathcal{F} \rightarrow L^2$  given by:  $Q(f) = \dot{f} / \sqrt{|\dot{f}|}$ . This square root trick is often used, see for instance [KSW11]. 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:  $\gamma \cdot q = q \circ \gamma \sqrt{\dot{\gamma}}$ , for which one can check easily by a change of variable that:

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

So up to the mapping  $Q$ , the Fisher Rao metric on curve corresponds to the situation  $M$  where theorem 5.2 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 5.1.** *Under the assumptions of theorem 5.2, we write  $X = t_0 + \sigma\epsilon$  where  $t_0$  is a fixed point,  $\sigma > 0$ ,  $\mathbb{E}(\epsilon) = 0$  and  $\mathbb{E}(\|\epsilon\|^2) = 1$ , if there is a Fréchet mean of  $[X]$ , then the consistency bias is linear with respect to  $\sigma$  and it is equal to:*

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

*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:

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

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

## 5.2 Proofs of these theorems

### 5.2.1 Proof of theorem 5.1

We start with the following simple result, which aims to differentiate the variance of  $X$ . This classical result (see [Pen06] for instance) is proved in appendix B in order to be the more self-contained as possible:

**Lemma 5.1.** *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  $\log_m : M \setminus C(m) \rightarrow T_m M$  is defined for any  $x \in M \setminus C(m)$  as the unique  $u \in T_m M$  such that  $\exp_m(u) = x$  and  $\|u\|_m = d_M(x, m)$ .

We are now ready to prove theorem 5.1.

*Proof.* (of theorem 5.1) 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 since  $x \mapsto g_0 x$  is a symmetry (the distance is equivariant under the action of  $G$ ) have that  $m_1 = g_0 m_0 \notin C(g_0 t_0) = C(t_0)$  ( $t_0$  is a fixed point under the action of  $G$ ). Let  $v_0 = \log_{t_0}(m_0)$  and  $v_1 = \log_{t_0}(m_1)$ . We have  $v_0 \neq v_1$  and since  $C(t_0)$  is closed and the  $\log_{t_0}$  is continuous application on  $M \setminus C(t_0)$  we have:

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\mathbb{P}(X \in B(m_0, \epsilon))} \mathbb{E}(\mathbb{1}_{X \in B(m_0, \epsilon)} \log_{t_0}(X)) = v_0.$$

(we use here the fact that since  $m_0$  is in the support of the law of  $X$ ,  $\mathbb{P}(X \in B(m_0, \epsilon)) > 0$  for any  $\epsilon > 0$  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:

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\mathbb{P}(X \in B(m_0, \epsilon))} \mathbb{E}(\mathbb{1}_{X \in B(m_0, \epsilon)} \log_{t_0}(g_0 X)) = v_1.$$

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

$$\mathbb{E}(\log_{t_0}(X) \mathbb{1}_{X \in B(m_0, \epsilon)}) \neq \mathbb{E}(\log_{t_0}(g_0 X) \mathbb{1}_{X \in B(m_0, \epsilon)}). \quad (33)$$

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 (33). In order to do that we introduce simple functions as the function  $x \mapsto \mathbb{1}_{x \in B(m_0, \epsilon)}$  which intervenes in Equation (33). Let  $s : M \rightarrow G$  be a simple function (i.e. a measurable function with finite number of values in  $G$ ). Then  $x \mapsto h(x) = s(x)x$  is a measurable function<sup>3</sup>. Now, let  $E_s(x) = \mathbb{E}(d(x, s(X)X)^2)$  be the variance of the variable  $s(X)X$ . Note that (and this is the main point):

$$\forall g \in G \quad d_M(t_0, x) = d_M(gt_0, gx) = d_M(t_0, gx) = d_Q([t_0], [x]),$$

---

<sup>3</sup>Indeed if:  $s = \sum_{i=1}^n g_i \mathbb{1}_{A_i}$  where  $(A_i)_{1 \leq i \leq n}$  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.

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:

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

Now, we want to apply lemma 5.1 to the random variables  $s(X)X$  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  $s(X)X \notin C(t_0)$  we get  $\mathbb{P}(s(X)X \in C(t_0)) = 0$  and the lemma 5.1 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:

$$\mathbb{E}(\log_{t_0}(X)) = \mathbb{E}(\log_{t_0}(s(X)X)) = 0. \quad (34)$$

Now we apply Equation (34) to a particular simple function defined by  $s(x) = g_0 \mathbb{1}_{x \in B(m_0, \epsilon)} + e_G \mathbb{1}_{x \notin B(m_0, \epsilon)}$ . We split the two expected values in (34) into two parts:

$$\mathbb{E}(\log_{t_0}(X) \mathbb{1}_{X \in B(m_0, \epsilon)}) + \mathbb{E}(\log_{t_0}(X) \mathbb{1}_{X \notin B(m_0, \epsilon)}) = 0, \quad (35)$$

$$\mathbb{E}(\log_{t_0}(g_0 X) \mathbb{1}_{X \in B(m_0, \epsilon)}) + \mathbb{E}(\log_{t_0}(X) \mathbb{1}_{X \notin B(m_0, \epsilon)}) = 0. \quad (36)$$

By substrating (35) from (36), one gets:

$$\mathbb{E}(\log_{t_0}(X) \mathbb{1}_{X \in B(m_0, \epsilon)}) = \mathbb{E}(\log_{t_0}(g_0 X) \mathbb{1}_{X \in B(m_0, \epsilon)}),$$

which is a contradiction with (33). Which concludes.  $\square$

### 5.2.2 Proof of theorem 5.2

*Proof.* The extension to theorem 5.2 is quite straightforward. In this setting many things are now explicit since  $d(x, y) = \|x - y\|$ ,  $\nabla_x d(x, y)^2 = 2(x - y)$ ,  $\log_x(y) = y - x$  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.  $\square$

## 6 Conclusion and discussion

In this article, 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. The behaviour of the consistency bias is summarized in table 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

Table 1: Behaviour of the consistency bias with respect to  $\sigma^2$  the variability of  $X = t_0 + \sigma\epsilon$ . The constants  $K_i$ 's depend on the kind of noise, on the template  $t_0$  and on the group action.

Consistency bias : $CB$	$G$ is any group	Supplementary properties for $G$ a finite group
Upper bound of $CB$	$CB \leq \sigma + 2\sqrt{\sigma^2 + K_1\sigma}$ (proposition 4.5)	$CB \leq K_2\sigma$ (theorem 3.3)
Lower bound of $CB$ for $\sigma \rightarrow \infty$ when the template is not a fixed point	$CB \geq L \underset{\sigma \rightarrow \infty}{\sim} K_3\sigma$ (proposition 4.4)	
Behavior of $CB$ for $\sigma \rightarrow 0$ when the template is not a fixed point	$CB \leq U \underset{\sigma \rightarrow 0}{\sim} K_4\sqrt{\sigma}$	$CB \underset{0}{=} o(\sigma^k)$ , $\forall k \in \mathbb{N}$ in the section 3.3, can we extend this result for finite group?
$CB$ when the template is a fixed point	$CB = \sigma \sup_{\ v\ =1} \mathbb{E}(\sup_{g \in G} \langle v, g\epsilon \rangle)$ (proposition 5.1)	

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 [AKT10] or develop new methods.

## A Proof of theorems for finite groups' setting

### A.1 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 strict 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 :

$$d_Q([m], [x]) = \min_{g \in G} \|m - g \cdot x\|, \quad (37)$$

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  $g$  maximises  $\langle m, g \cdot x \rangle$ .
5. If  $m$  is regular and  $x \notin A_m$  then:

$$\forall g \in G \setminus \{g(x, m)\}, \|m - g(x, m) \cdot x\| < \|m - g \cdot x\|,$$

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  $y \in B(x, \alpha)$ :

$$\forall g \in G \setminus \{g(x, m)\} \|\mu - g(x, m) \cdot y\| < \|\mu - g \cdot y\|. \quad (38)$$

Therefore for such  $y$  and  $\mu$  we have:

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

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

$$\begin{aligned} Cone(m) &= \{x \in \mathbb{R}^n / \forall g \in G \|x - m\| \leq \|x - g \cdot m\|\} \\ &= \{x \in \mathbb{R}^n / \forall g \in G \langle m, x \rangle \geq \langle gm, x \rangle\}. \end{aligned} \quad (39)$$

This is the intersection of  $|G| - 1$  half-spaces: each half space is delimited by  $H(m, gm)$  for  $g \neq e_G$  (see fig. 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 minimises 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:  $[x] \cap Cone(m) \neq \emptyset$ ,  $card([x] \cap Cone(m)) \geq 2$  if and only if  $x \in A_m$ . The borders of the cone is  $Cone(m) \setminus \text{Int}(Cone(m)) = Cone(m) \cap A_m$  (we denote by  $\text{Int}(A)$  the interior of a part  $A$ ). Therefore  $Q = \mathbb{R}^n / G$  can be seen like  $Cone(m)$  whose border 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  $F$ :

$$\forall m, x \in M \quad f(x, m) = \min_{g \in G} \|m - g \cdot x\|^2. \quad (40)$$

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 A.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:*

$$\frac{\partial f}{\partial m}(x, m_0) = 2(m_0 - g(x, m_0) \cdot x) \quad (41)$$

*Proof.* If  $m_0$  is regular and  $x \notin A_{m_0}$  then we know from the item 5 of the appendix A.1 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 A.1.  $\square$

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 A.2.** *For every  $m_0 \in M$  we have the existence of an integrable function  $\Phi : M \rightarrow \mathbb{R}^+$  such that:*

$$\forall m \in B(m_0, 1), \forall x \in M \quad |f(x, m_0) - f(x, m)| \leq \|m - m_0\| \Phi(x). \quad (42)$$

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

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

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

$$|f(x, m_0) - f(x, m)| \leq \|m_0 - m\| (2\|m_0\| + \|m - m_0\| + \|2x\|) \quad (43)$$

The function  $\Phi$  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 (43), which concludes.  $\square$



## A.2 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  $t_0$ :

$$\langle \nabla F(t_0), t_0 \rangle = 2\mathbb{E}(\langle X, t_0 \rangle - \langle g(X, t_0) \cdot X, t_0 \rangle) = 0. \quad (44)$$

The item 4 of  $(x, m) \mapsto g(x, m)$  seen at appendix A.1 leads to:

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

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

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

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 (6).

## A.3 Proof of theorem 3.3: upper bound of the consistency bias

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

**Lemma A.3.** *We write  $X = t_0 + \epsilon$  where  $\mathbb{E}(\epsilon) = 0$  and we make the assumption that the noise  $\epsilon$  is a subgaussian random variable. This means that it exists  $c > 0$  such that:*

$$\forall m \in M = \mathbb{R}^n, \mathbb{E}(\exp(\langle \epsilon, m \rangle)) \leq c \exp\left(\frac{s^2 \|m\|^2}{2}\right). \quad (45)$$

If for  $m \in M$  we have:

$$\tilde{\rho} := d_Q([m], [t_0]) \geq s\sqrt{2\log(c|G|)}, \quad (46)$$

then we have:

$$\tilde{\rho}^2 - \tilde{\rho}s\sqrt{8\log(c|G|)} \leq F(m) - \mathbb{E}(\|\epsilon\|^2). \quad (47)$$

*Proof.* (of lemma A.3) First we expand the right member of the inequality (47):

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

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

$$\mathbb{E}(\|\epsilon\|^2) - F(m) = \mathbb{E}\left[\max_{g \in G} (-2\langle X - t_0, t_0 - gm \rangle - \|t_0 - gm\|^2)\right] = \mathbb{E}(\max_{g \in G} \eta_g), \quad (48)$$

with  $\eta_g = -\|t_0 - gm\|^2 + 2\langle \epsilon, gm - t_0 \rangle$ . Our goal is to find a lower bound of  $F(m) - \mathbb{E}(\|\epsilon\|^2)$ , 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 (45):

$$\begin{aligned}
\exp(x\mathbb{E}(\max_{g \in G} \eta_g)) &\leq \mathbb{E}(\exp(\max_{g \in G} x\eta_g)) \leq \mathbb{E}\left(\sum_{g \in G} \exp(x\eta_g)\right) \\
&\leq \sum_g \exp(-x\|t_0 - gm\|^2) \mathbb{E}(\exp(\langle \epsilon, 2x(gm - t_0) \rangle)) \\
&\leq c \sum_g \exp(-x\|t_0 - gm\|^2) \exp(2s^2x^2\|gm - t_0\|^2) \\
&\leq c \sum_g \exp(\|gm - t_0\|^2(-x + 2x^2s^2)) \tag{49}
\end{aligned}$$

Now if  $(-x + 2t^2x^2) < 0$ , we can take an upper bound of the sum sign in (49) 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  $(-x + 2x^2s) < 0 \iff 0 < x < \frac{1}{2s^2}$ . Then we have:

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

Then by taking the log:

$$\mathbb{E}(\max_{g \in G} \eta_g) \leq \frac{\log c|G|}{x} + (2xs^2 - 1)\tilde{\rho}^2. \tag{50}$$

Now we find the  $x$  which optimizes inequality (50). By differentiation, the right member of inequality (50) is minimal for  $x_\star = \sqrt{\log c|G|/2}/(s\tilde{\rho})$  which is a valid choice because  $x_\star \in (0, \frac{1}{2s^2})$  by using the assumption (46). With the equations (48) and (50) and  $x_\star$  we get the result.  $\square$

*Proof.* (of theorem 3.3) We take  $m_\star \in \operatorname{argmin} F$ ,  $\tilde{\rho} = d_Q([m_\star], [t_0])$ , and  $\epsilon = X - t_0$ . We have:  $F(m_\star) \leq F(t_0) \leq \mathbb{E}(\|\epsilon\|^2)$  then  $F(m_\star) - \mathbb{E}(\|\epsilon\|^2) \leq 0$ . If  $\tilde{\rho} > s\sqrt{2\log(|G|)}$  then we can apply lemma A.3 with  $c = 1$ . Thus:

$$\tilde{\rho}^2 - \tilde{\rho}s\sqrt{8\log(|G|)} \leq 2F(m_\star) - \mathbb{E}(\|\epsilon\|^2) \leq 0,$$

which yields to  $\tilde{\rho} \leq s\sqrt{8\log(|G|)}$ . If  $\tilde{\rho} \leq s\sqrt{2\log(|G|)}$ , we have nothing to prove.  $\square$

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.

#### A.4 Proof of proposition 3.2: inconsistency in $\mathbb{R}^2$ for the action of translation

*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 minimises  $\|\tau \cdot x - m\|$  see (37) 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 completely write what take  $\tau(x, m)$  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  $-m$ ).

This allows use to write the variance at the point  $m \in HP_A$ :

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

Then we define the random variable  $Z$  by:  $Z = X \mathbf{1}_{X \in HP_A \cup L} + 1 \cdot X \mathbf{1}_{X \in HP_B}$ , 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  $m_*$  is a global minimiser of  $F$ , then  $m_* = \mathbb{E}(Z)$  or  $m_* = 1 \cdot \mathbb{E}(Z)$ . 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  $X = X \mathbf{1}_{X \in HP_A} + X \mathbf{1}_{X \in HP_B \cup L}$ , we have:

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

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:  $(u, v)^T = \mathbb{E}(X) + (r \cos \theta, r \sin \theta)^T$  where  $r > 0$  et  $\theta \in [0, 2\pi]$ . We also define:  $d = \text{dist}(\mathbb{E}(X), L)$ ,  $\mathbb{E}(X)$  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  $(r, \theta)$  the points which are in  $HP_B$ :

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

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

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E}(X \mathbf{1}_{X \in HP_A}) + \mathbb{E}(1 \cdot X \mathbf{1}_{X \in HP_B}) \\ \mathbb{E}(Z) &= \int_0^d \int_0^{2\pi} \begin{pmatrix} \alpha + r \cos \theta \\ \beta + r \sin \theta \end{pmatrix} \frac{\exp\left(-\frac{r^2}{2s^2}\right)}{2\pi s^2} r d\theta dr \\ &\quad + \int_d^{+\infty} \int_{\arccos(\frac{d}{r}) - \frac{\pi}{4}}^{2\pi - \frac{\pi}{4} - \arccos(\frac{d}{r})} \begin{pmatrix} \alpha + r \cos \theta \\ \beta + r \sin \theta \end{pmatrix} \frac{\exp\left(-\frac{r^2}{2s^2}\right)}{2\pi s^2} r dr d\theta \\ &\quad + \int_d^{+\infty} \int_{-\frac{\pi}{4} - \arccos(\frac{d}{r})}^{-\frac{\pi}{4} + \arccos(\frac{d}{r})} \begin{pmatrix} \beta + r \sin \theta \\ \alpha + r \cos \theta \end{pmatrix} \frac{\exp\left(-\frac{r^2}{2s^2}\right)}{2\pi s^2} r dr d\theta \\ &= \mathbb{E}(X) + \int_d^{+\infty} \frac{r^2 \exp\left(-\frac{r^2}{2s^2}\right)}{\pi s^2} \sqrt{2} g\left(\frac{d}{r}\right) dr \times (-1, 1)^T, \end{aligned}$$

We compute  $\tilde{\rho} = d_Q([\mathbb{E}(X)], [\mathbb{E}(Z)])$  where  $d_Q$  is the distance in the quotient space defined in (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 eq. (9), note that items 2a to 2c are the direct consequence of eq. (9) and basic analysis.  $\square$

## B Proof of lemma 5.1: differentiation of the variance in the top space

*Proof.* 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  $m \in M \setminus C(z)$  (i.e.  $z \notin C(m)$ ) with derivative  $-2 \log_m(z)$ . Now since:

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

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

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

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

## References

- [AAT07] Stéphanie Allasonnière, 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)*, 69(1):3–29, 2007.
- [ADP15] Stéphanie Allasonnière, 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.
- [AKT10] Stéphanie Allasonnière, Estelle Kuhn, and Alain Trouvé. Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli*, 16(3):641–678, 2010.
- [BB08] Abhishek Bhattacharya and Rabi Bhattacharya. Statistics on riemannian manifolds: asymptotic distribution and curvature. *Proceedings of the American Mathematical Society*, 136(8):2959–2967, 2008.
- [BC11] Jérémie Bigot and Benjamin Charlier. On the consistency of fréchet means in deformable models for curve and image analysis. *Electronic Journal of Statistics*, 5:1054–1089, 2011.
- [BG14] Dominique Bontemps and Sébastien Gadat. Bayesian methods for the shape invariant model. *Electronic Journal of Statistics*, 8(1):1522–1568, 2014.
- [CWS16] 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*, 28(2):338–359, 2016.
- [DPC<sup>+</sup>14] Stanley Durrleman, Marcel Prastawa, Nicolas Charon, Julie R Kornberg, Sarang Joshi, Guido Gerig, and Alain Trouvé. Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage*, 101:35–49, 2014.
- [Fré48] 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.

- [GM98] Ulf Grenander and Michael I. Miller. Computational anatomy: An emerging discipline. *Q. Appl. Math.*, LVI(4):617–694, December 1998.
- [HCG<sup>+</sup>13] Sebastian Hitziger, Maureen Clerc, Alexandre Gramfort, Sandrine Sillion, Christian Bénar, and Théodore Papadopoulos. Jitter-adaptive dictionary learning-application to multi-trial neuroelectric signals. *arXiv preprint arXiv:1301.3611*, 2013.
- [JDJG04] Sarang Joshi, Brad Davis, Mathieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23:S151–S160, 2004.
- [Kar77] Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.
- [Ken89] David G Kendall. A survey of the statistical theory of shape. *Statistical Science*, pages 87–99, 1989.
- [Ken90] Wilfrid S Kendall. Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 3(2):371–406, 1990.
- [KSW11] Sebastian A. 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.
- [LAJ<sup>+</sup>12] Sidonie Lefebvre, Stéphanie Allasonnière, Jérémie Jakubowicz, Thomas Lasne, and Eric Moulines. Aircraft classification with a low resolution infrared sensor. *Machine Vision and Applications*, 24(1):175–186, 2012.
- [MHP16] Nina Miolane, Susan Holmes, and Xavier Pennec. Template shape estimation: correcting an asymptotic bias. *arXiv preprint arXiv:1610.01502*, 2016.
- [MP15] 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.
- [Pen06] Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

- [SBG08] Mert Sabuncu, Serdar K. Balci, and Polina Golland. Discovering modes of an image population through mixture modeling. *Proceeding of the MICCAI conference*, LNCS(5242):381–389, 2008.
- [Zie77] 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.
- [ZSF13] Miaomiao Zhang, Nikhil Singh, and P.Thomas Fletcher. Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In JamesC. Gee, Sarang Joshi, KilianM. Pohl, WilliamM. 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.