HAL archives-ouvertes.fr

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

Stéphanie Allassonnière, Loïc Devilliers, Xavier Pennec

▶ To cite this version:

Stéphanie Allassonnière, Loïc Devilliers, 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. Mathematical Foundations of Computational Anatomy (MFCA'15), Oct 2015, Munich, Germany. pp.131-142. hal-01203816

HAL Id: hal-01203816

https://hal.inria.fr/hal-01203816

Submitted on 25 Sep 2015

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.

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

Stéphanie Allassonnière[1], Loïc Devilliers[1 2], and Xavier Pennec[3]

[1] École Polytechnique, Palaiseau, France.
[2] École Normale Supérieure de Cachan, France. loic.devilliers@ens-cachan.fr
[3] Asclepios project-team, INRIA Sophia Antipolis, France.

Abstract. When we have a deformation group acting on a vector space of observations, these data are not anymore elements of our space but rather orbits for the group action we consider. If the data are generated from an unknown template with noise, to estimate this template, one may want to minimise the variance in the quotient set. In this article we study statistics on a particular quotient space. We prove that the expected value of a random variable in our vector space mapped in the quotient space is different from the Fréchet mean in the quotient space when the observations are noisy.

Introduction

In the theory of shape introduced by Kendall [6], in Computational anatomy [4] or in image analysis, one often aims at estimating a template (which stands for the mean of the data) of shapes (for instance an average shape of an organ from a population of subject scans). To understand the observations, one assumes that these data follow a statistical model. A very popular one is that the observations are random deformations of the template with additional noise. This is the model proposed in [4] which is the foundation of Computational Anatomy. This introduces the notion of group action where the deformations we considered are elements of a group which acts on the set of objects, namely the images. In this particular setting, the template estimation is most of the time based on the minimization of the empirical variance in the quotient space (called the empirical Fréchet mean) (see for instance [7,5,9] among many others).

More precisely here, we consider M a finite dimensional vector space with a euclidean norm, G a finite group acting on M, such that the action is isometric with the respect of the euclidean norm on M. Thus the quotient M/G is equipped with a quotient distant, noted here ρ, moreover we call $[m]$ the orbit of $m \in M$.

We consider the following generative model: Y is a random variable on M with a density h for the Lebesgue measure. Y is not a random variable constant: it is the sum of a template and a white noise. And we aim to estimate this template. If Y was an observed variable, this question will be meaningless it would suffice to compute: $\int_M yh(y)dy$ which corresponds at the case where we have an infinite number of observations. Instead, here the random variable Y which lives in the total space M is not an observable variable, only $[Y]$ is observable.

In the following, the deformations of the group G will be restricted to translations and the noise will follow a general distribution. This kind of action is a simplified setting for image registration, for instance medical images can be obtained by translation of one scan to another due to different poses. More precisely, we work in the vector space $M = \mathbb{R}^{\mathbb{T}}$ where $\mathbb{T} = (\mathbb{Z}/N\mathbb{Z})^D$ is a discrete torus in D-dimension, an element of $\mathbb{R}^{\mathbb{T}}$ is seen as a function $y : \mathbb{T} \to \mathbb{R}$, $y(\tau)$ is the value at the pixel τ. When $D = 1$, y can be seen like a discretised signal with N pixels, when $D = 2$, we can see y like a picture with $N \times N$ pixels etc. We then define the group action of $G = \mathbb{T}$ on $\mathbb{R}^{\mathbb{T}}$ by:

$$\tau \in \mathbb{T},\ y \in \mathbb{R}^{\mathbb{T}} \quad \tau \cdot y : \sigma \mapsto y(\sigma + \tau). \tag{1}$$

We note $|| \quad ||$ the canonical Euclidean norm over $\mathbb{R}^{\mathbb{T}}$. We define a distance in the quotient space by:

$$\rho([y],[z]) = \inf_{\tau,\sigma \in \mathbb{T}} ||\tau \cdot y - \sigma \cdot z|| = \inf_{\tau \in \mathbb{T}} ||\tau \cdot y - z||. \tag{2}$$

Now the fact that Y has a density for the Lebesgue measure, implies that $[Y]$ has a density in M/G for the image measure noted ν. This density is given by: $\tilde{h}([y]) = \frac{1}{|G|}\sum_{g\in G} h(g \cdot y)$, therefore we can write the variance of $[Y]$ at the point $[\mu] \in M/G$ by:

$$\begin{aligned}
F([\mu]) &= \mathbb{E}(\rho([\mu],[Y])^2) && (3)\\
&= \int_{M/G} \rho([\mu],z)^2 \tilde{h}(z)\nu(dz) && (4)\\
&= \frac{1}{|G|}\int_{M/G} \rho([\mu],[y])^2 \sum_{g\in G} h(g\cdot y)\nu(d[y])) && (5)\\
&= \int_M \rho([\mu],[y])^2 h(y)\lambda(dy) && (6)\\
&= \int_{\mathbb{R}^{\mathbb{T}}} \inf_{\tau\in\mathbb{T}} ||\tau y - \mu||^2 h(y)\lambda(dy) = J(\mu). && (7)
\end{aligned}$$

J is non-negative, continuous, $\lim_{||\mu||\to+\infty} J(\mu) = +\infty$, therefore J reaches its minimum. The points in $\mathbb{R}^{\mathbb{T}}/\mathbb{T}$ which minimises F are the Fréchet means of $[Y]$. In this article, the central question is: is the template - which generates the random variable Y in the total space - mapped in the quotient space, a Fréchet mean of $[Y]$ or not?

About this kind of questions, previous works have been done before: for instance Allassonnière, Yali and Trouvé in [1] show an example of translated step function. They compared the iterative algorithm which numerically estimates the empirical Frechet mean in the quotient space to the Expectation-Maximization [3] algorithm which approximates the maximum likelihood estimator. In this example, even with a large number of observations, estimating the empirical Fréchet mean did not succeed to estimate well the template (the step function from which the synthetic samples were generated) when the noise on the observation was large enough.

To understand the example found in [1], different algorithms and theorems have been proposed (for instance in [2,7] or [10]), to improve or ensure the convergence of the empirical Fréchet mean in a more general case than presented in this article. A first contribution to provide a clue to know if even with an infinite number of observations, we could estimate the template has been given by Miolane and Pennec in [8]. They show that the presence of noise may imply that the template mapped in the quotient space is not a Fréchet mean in the quotient space. Then estimating the template in the total space with the Fréchet mean in the quotient space produces a bias. Considering the action of rotations on an euclidean space, they highlight the influence of dimension of the considered vector space and the influence of the ratio signal over noise on the bias. Although they showed a general result with a finite dimensional manifold and an isometric Lie group action, they made the assumption of a Gaussian noise. Here we do not make this assumption to show the presence of bias. For instance, here even with a bounded support of the density, under some condition we may have a bias. Moreover the method proposed here is different from [8], which can provide another explanation to the presence of bias in this context.

This paper is organised as follows. In Section 1, we show that the expected value of Y mapped in the quotient space is not a Fréchet mean of $[Y]$ as soon as the density of Y satisfies a certain condition. In Section 2, we compute the bias in a special case of torus with a Gaussian noise. This trivial example aims to give us an intuition of which parameters the bias depends on.

1 Existence of a bias for any discrete torus

In this section, we show that under some conditions of the density, the expected value of the random variable Y is not a minimum of J (defined in Equation (7)). To show that, we first study the differentiability of the integrand of J. Then we justify that the gradient of the variance J is the integral of the gradient's integrand. Finally we show that the gradient of the variance J at the expected value of the random variable Y is not zero. It will imply that the expected value of Y mapped in the quotient space in not a Fréchet mean of $[Y]$.

1.1 Study of the Differentiate of the Integrand

In this sub-part, we search to see when the integrand is differentiable, and to compute its gradient. In order to do that we defined:

$$\forall\, \mu,\ y \in \mathbb{R}^{\mathbb{T}} \quad f(\mu, y) = \inf_{\tau \in \mathbb{T}} ||\tau \cdot y - \mu||^2 h(y) dy. \tag{8}$$

Then we have: $J(\mu) = \int_{\mathbb{R}^{\mathbb{T}}} f(\mu, y) dy$, we will see that the differentiability of $\mu \mapsto f(\mu, y)$ at the point μ_0 depends on y and μ_0, more precisely the question on differentiability is related to the isotropy group of μ_0 and to the distances between y and $\tau \cdot \mu_0$ for $\tau \in \mathbb{T}$. Indeed one difficulty appears here: the inf of several differentiable functions is not necessary differentiable.

Remark 1. Let $f_1, \dots, f_r : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions at a point x_0, $f = \inf_{1 \le i \le r} f_i$ is differentiable at x_0 if: $\forall i,\ j \in [\![1, r]\!]^2,\ i \neq j \implies f_i(x_0) \neq f_j(x_0)$. Indeed in this case, we take $k = \operatorname{argmin}\{f_i(x_0), i \in [\![1, r]\!]\}$, we have locally around x_0: $f = f_k$. Then f is differentiable at x_0, and $\nabla f(x_0) = \nabla f_k(x_0)$ (where $\nabla f(x)$ is the gradient of f at the point x).

In Equation (8), let τ, τ' be two distinct elements of $\mathbb{T}$ then: $||\tau y - \mu|| = ||\tau' y - \mu||$ is equivalent to $||y - (-\tau)\mu|| = ||y - (-\tau')\mu||$,[4] there are two cases:

- If $(-\tau)\mu = (-\tau')\mu$ then $\forall y \in \mathbb{R}^{\mathbb{T}}$ $||\tau y - \mu|| = ||\tau' y - \mu||$.
- If for all $\tau \neq \tau'$ we have $\tau\mu \neq \tau'\mu$, i.e. the isotropy group is reduced to $\{0\}$, (the isotropy group is defined by: $\mathrm{Iso}(\mu) = \{\tau \in \mathbb{T},\ \tau\mu = \mu\}$). We call such a μ a regular point, otherwise we say that μ is a singular point. We note

$$A_\mu = \bigcup_{\tau, \tau' \in \mathbb{T},\ \tau \neq \tau'} \{x \in \mathbb{R}^{\mathbb{T}},\ ||x - \tau \cdot \mu|| = ||x - \tau' \cdot \mu||\}. \tag{9}$$

  Foj μ regual, A_μ is the set of points equally distant from two points of the orbit of μ, A_μ is a finite union of hyperplanes, therefore the Lebesgue's measure of A_μ is null. In this case for every regular point μ and for all most every y (y does not belong to A_μ), the infimum in Equation (8) is reached at a unique $\tau \in \mathbb{T}$. When the infimum in Equation (8) is reached at a unique $\tau \in T$, we note this τ by:

$$\tau(y, \mu) = \operatorname{argmin}\{||\tau \cdot y - \mu||,\ \tau \in \mathbb{T}\}. \tag{10}$$

We note $\mathrm{Sing} = \{\mu \in \mathbb{R}^{\mathbb{T}},\ \text{such that } \mathrm{Iso}(\mu) \neq \{0\}\}$ the set of singular points. Notice that: $\mathrm{Sing} = \bigcup_{\tau \neq 0} \ker(x \mapsto \tau \cdot x - x)$ is a finite union of strict linear subspaces of $\mathbb{R}^{\mathbb{T}}$, then Sing is a null set for the Lebesgue's measure. For $\mu \notin \mathrm{Sing}$ we have then for almost all y:

$$f(\mu, y) = \inf_{\tau \in \mathbb{T}} ||\mu - \tau y||^2 h(y) = ||\mu - \tau(y, \mu) y||^2 h(y). \tag{11}$$

[4] Because $||x|| = ||\tau x||$, and $\tau(x + y) = \tau x + \tau y$.

We can now apply the remark 1 to differentiate the integrand f defined in (8). But first we need to see how τ variates. Let μ be a regular point and $y \notin A_\mu$ therefore: $||\mu-\tau(y,\mu)\cdot y|| < \inf_{\alpha\neq\tau(y,\mu)} ||\mu-\alpha y||$. For continuity reason we have the existence of $\alpha_{\mu,y} > 0$, $\beta_{\mu,y} > 0$ such that for $\nu, z \in \mathbb{R}^{\mathbb{T}}$ verifying $||\mu-\nu|| < \alpha_{\mu,y}$, $||y-z|| < \beta_{\mu,y}$ we still have:

$$||\nu-\tau(y,\mu)\cdot z|| < \inf_{\alpha\neq\tau(y,\mu)} ||\nu-\alpha\cdot z||. \tag{12}$$

And then we have:

$$\forall \nu, z \in \mathbb{R}^{\mathbb{T}},\ ||\mu-\nu|| < \alpha_{\mu,y},\ ||y-z|| < \beta_{\mu,y} \Longrightarrow \tau(z,\nu) = \tau(y,\mu). \tag{13}$$

Finally, we can differentiate $\mu \mapsto f(\mu, y)$ with respect to μ in $\mu_0 \notin \text{Sing}$ and $y \notin A_{\mu_0}$, Equation (13) allow us to differentiate $\mu \mapsto \tau(y,\mu)$ (which is locally constant) which yields:

$$\frac{\partial f}{\partial \mu}(\mu_0, y) = 2(\mu_0 - \tau(y,\mu_0)y)h(y). \tag{14}$$

Now that we have seen the differentiability of the integrand, we justify in the next part that we can permute the differentiation and the integral sign.

1.2 Justification of the Differentiation of the Integral

In order to differentiate the variance in the quotient space (noted J), we propose to do the following things:

- Showing that $\mu \mapsto f(\mu, y)$ is weakly differentiable for almost all y, and computing its weak gradient.
- Deducing that J is weakly differentiable and finding its weak gradient ∇J.
- Showing that ∇J is continuous at some point, therefore by integration J is differentiable at these points, and ∇J is its strong gradient.

Remark 2. We can not apply here the theorem of differentiation under the integral sign, because $\mu \mapsto f(\mu, y)$ is differentiable at μ_0 for almost all y, but "the allmost y" is $\mathbb{R}^{\mathbb{T}} \setminus A_{\mu_0}$ depends of μ_0.

Weak differentiation of *f(,y)* for almost all *y*. First we define $\mathcal{C}_c^\infty(\mathbb{R}^{\mathbb{T}}, \mathbb{R})$ as the set of functions of infinite class whose support is a compact set. We want here to show that for almost all y ($y \notin \text{Sing}$):

$$\forall \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^{\mathbb{T}}, \mathbb{R}) \quad \int f(\mu, y)\nabla\varphi(\mu)d\mu = -\int \varphi(\mu)\frac{\partial f}{\partial \mu}(\mu, y)d\mu. \tag{15}$$

Let $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^{\mathbb{T}}, \mathbb{R})$ and by linearity it is sufficient to show:

$$\int f(\mu, y)\frac{\partial \varphi(\mu)}{\partial \mu_1}d\mu = -\int \frac{\partial f}{\partial \mu_1}(\mu, y)\varphi(\mu)d\mu.$$

Which is equivalent to:[5]

$$\int\left(\int f(\mu,y)\frac{\partial\varphi(\mu)}{\partial\mu_1}d\mu_1\right)d\mu_2\dots d\mu_{|\mathbb{T}|}=-\int\left(\int\frac{\partial f}{\partial\mu_1}(\mu,y)\varphi(\mu)d\mu_1\right)d\mu_2\dots d\mu_{|\mathbb{T}|}.$$

Hence, it is sufficient to show that for almost all $(\mu_2,\dots,\mu_{|T|})$ we have:

$$\int f(\mu,y)\frac{\partial\varphi(\mu)}{\partial\mu_1}d\mu_1=-\int\frac{\partial f}{\partial\mu_1}(\mu,y)\varphi(\mu)d\mu_1. \tag{16}$$

There are two cases: $L=\{(x,\mu_2,\dots,\mu_{|\mathbb{T}|}),\ x\in\mathbb{R}\}$ is included in A_y, or not.
If L is not included in A_y then $L\cap A_y$ is finite, so in a connected component of $L\setminus A_y$, $f_L : x\mapsto f(x,\mu_2,\dots,\mu_{\mathbb{T}},y)$ is derivable with a strong derivative:[6]

$$f'_L(x)=\langle(\mu-\tau(y,\mu)y)h(y)\,|e\rangle\ \text{ with } \mu=(x,\mu_2,\dots,\mu_{|\mathbb{T}|}),\ e=(1,0,\dots,0).$$

Therefore by cutting the integral by pieces, where each piece is a connected component on $L\setminus A_y$ and by integrating by part on each piece, we get[7] in this case Equation (16).
We note $B=\{(\mu_2,\dots\mu_{|\mathbb{T}|})\in\mathbb{R}^{|\mathbb{T}|-1},\ \forall x\in\mathbb{R}\,(x,\mu_2,\dots\mu_{|\mathbb{T}|})\in A_y\}$ and we define for all $(\tau,\alpha)\in\mathbb{T}^2$ with $\tau\neq\alpha$:

$$\begin{aligned}\Psi_{\tau,\alpha}\ &:\ \begin{pmatrix}\mathbb{R}^{|\mathbb{T}|-1} & \to & \mathrm{Aff}\\ (\mu_2,\dots\mu_{|\mathbb{T}|}) & \mapsto & x\mapsto\Psi_{\tau,\alpha}(\mu_2,\dots,\mu_{|\mathbb{T}|})(x)\end{pmatrix}\\ \Psi_{\tau,\alpha}(\mu_2,\dots,\mu_{|\mathbb{T}|})(x)&=||(x,\mu_2,\dots\mu_{|\mathbb{T}|})-\tau y||^2-||(x,\mu_2,\dots\mu_{|\mathbb{T}|})-\alpha y||^2\\ &=2\left\langle(x,\mu_2,\dots,\mu_{|\mathbb{T}|})\,|\alpha y-\tau y\right\rangle.\end{aligned}$$

where Aff is set of real affine maps, $\Psi_{\tau,\alpha}$ is well defined, affine, non zero (because $y\notin$ Sing), so $\Psi^{-1}_{\tau,\alpha}(\{0\})$ is a strict affine subspace of $\mathbb{R}^{|\mathbb{T}|-1}$ therefore: $B=\bigcup\limits_{\tau\neq\alpha}\Psi^{-1}_{\tau,\alpha}(\{0\})$ is a null set.

To conclude, we have for almost all $(\mu_2,\dots\mu_{|\mathbb{T}|})$ the equation (16) which proves (15).

Remark 3. We did not show here that f belongs to a Sobolev space, because generally a Sobolev space is defined as the set of L^2 (or L^p) functions whose weak derivative exist and are in L^2 (or L^p), here $f(\ ,y)\notin L^p$ for every $p>1$, because $f(\mu,y)\to+\infty$ when $||\mu||\to+\infty$. Instead we have shown that the derivative of the distribution associated to $\mu\mapsto f(\mu,y)$ is a distribution associated to another function (namely $\mu\mapsto\frac{\partial f}{\partial\mu}(\mu,y)$). The only thing we need in order to speak about a distribution associated to a function is that the function is integrable over each compact set of $\mathbb{R}^{\mathbb{T}}$ which is the case here.

[5] For writing $\mu\in\mathbb{R}^{\mathbb{T}}$ $\mu=(\mu_1,\dots,\mu_{|\mathbb{T}|})$ we suppose that we have chosen (once for all) an arbitrary order between the $|\mathbb{T}|$ real variables.

[6] By using the result in (14) by permuting the role of μ and y to ensure that the inf in (8) is unique.

[7] In fact, this is a particular case of the theorem of derivation of a distribution represented by a function with jumps, the derivative of a jump at the position a is obtained by a Dirac distribution function in a, here there is no Dirac because the function f is continuous.

Weakly differentiability of the variance in the quotient space We now prove that $J(\mu) = \int f(\mu, y)dy$ is weakly differentiable: Let $\varphi \in \mathcal{C}_c(\mathbb{R}^{\mathbb{T}}, \mathbb{R})$ then (by permutation of integrals thanks to Fubini's theorem):

$$\begin{aligned}\int J(\mu)\varphi'(\mu)d\mu &= \int \left(\left[\int f(\mu,y)dy\right]\varphi'(\mu)\right)d\mu \\ &= \int \left(\int f(\mu,y)\varphi'(\mu)d\mu\right)dy \\ &= -\int\int \varphi(\mu)\frac{\partial f}{\partial \mu}(\mu,y)d\mu dy \\ &= -\int\left(\int \frac{\partial f}{\partial \mu}(\mu,y)dy\right)\varphi(\mu)d\mu.\end{aligned}$$

Thus J is weakly differentiable, and its weakly derivative is:

$$\nabla J(\mu) = \int \frac{\partial f}{\partial \mu}(\mu,y)dy = 2\left(\mu - \int \tau(y,\mu)yh(y)dy\right). \tag{17}$$

Continuity of the weak gradient at the regular points. We show the continuity of the weak gradient at the regular points by simply applying the continuity under integral sign:
Let $\mu_0 \in \mathbb{R}^{\mathbb{T}} \setminus \mathrm{Sing}$, then for $y \notin A_{\mu_0}$, $\mu \mapsto f(\mu, y)$ is continuous at μ_0 by Equation (13), moreover $||\tau(y,\mu)\cdot y||h(y) \leq ||y||h(y)$ with $y \mapsto ||y||h(y)$ an integrable function independent of μ. Therefore by the continuity under integral sign:

$$\nabla J(\mu) = 2\left(\mu - \int \tau(y,\mu)\cdot yh(y)dy\right), \tag{18}$$

is continuous at μ_0. This implies that the variance in the quotient space, (noted J) is differentiable over $\mathbb{R} \setminus \mathrm{Sing}$ and its strong gradient over $\mathbb{R} \setminus \mathrm{Sing}$ is:

$$\nabla J(\mu) = 2\left(\mu - \int_{\mathbb{R}^{\mathbb{T}}} \tau(y,\mu)\cdot yh(y)dy\right). \tag{19}$$

1.3 The expected value of Y mapped in the quotient space is not necessarily a Fréchet mean of $[Y]$

We suppose that $\mathbb{E}(Y)$ (noted y_0) the expected value of the random variable Y is a regular point (to ensure that J is differentiable at y_0) and verifies $\nabla J(y_0) = 0$, and we want to find a contradiction, we know from Equation (19) that:

$$\frac{1}{2}\nabla J(y_0) = \int_{\mathbb{R}^{\mathbb{T}}} yh(y)dy - \int_{\mathbb{R}^{\mathbb{T}}} \tau(y,y_0)\cdot yh(y)dy. \tag{20}$$

Therefore:

$$\frac{1}{2}\left\langle \nabla J(y_0)\,|y_0\right\rangle = \int_{\mathbb{R}^{\mathbb{T}}} \left(\left\langle y\,|y_0\right\rangle - \left\langle \tau(y,y_0)\cdot y\,|y_0\right\rangle\right)h(y)dy = 0. \tag{21}$$

We shall remember that $\tau(y, y_0)$ minimises $\{||\tau \cdot y - y_0||,\ \tau \in \mathbb{T}\}$ for almost all y. Then it minimises for almost all y:

$$\{||\tau \cdot y - y_0||^2 = ||y||^2 + ||y_0||^2 - 2\,\langle \tau y\,|y_0\,\rangle\,,\ \tau \in \mathbb{T}\},$$

and then almost surely $\tau(y, y_0)$ maximises:

$$\{\langle \tau \cdot y\,|y_0\,\rangle\,, \tau \in \mathbb{T}\},$$

This leads to:

$$\langle y\,|y_0\,\rangle - \langle \tau(y, y_0) \cdot y\,|y_0\,\rangle \leq 0 \text{ almost surely.}$$

So the integral of a non-positive function is null, so if we note $\mathrm{Supp}(h)$ the support of h we have then:

$$\forall y \in \mathrm{Supp}(h),\ \langle y\,|y_0\,\rangle = \langle \tau(y, y_0) \cdot y\,|y_0\,\rangle \text{ almost surely.} \tag{22}$$

Then $\tau = 0$ maximises the dot product almost surely. Therefore (as we know that $\tau(y, y_0)$ is unique almost surely, since y_0 is regular):

$$\forall y \in \mathrm{Supp}(h),\ \tau(y, y_0) = 0 \text{ almost surely.} \tag{23}$$

Let us suppose that the support of h contains a neighbourhood of $y \in \mathbb{R}^{\mathbb{T}}$ such that $\tau(y, y_0)$ is unique and $\tau(y, y_0) = \alpha \neq 0$, therefore: $||\alpha y - y_0|| < ||\tau y - y_0|| \quad \forall \tau \in \mathbb{T} \setminus \{\alpha\}$, and like in Equation (12), we have the existence of $r > 0$ such that:

$$\forall z \in B(y, r) \quad ||\alpha \cdot z - y_0|| < \inf_{\tau \in \mathbb{T}, \tau \neq \alpha} ||\tau z - y_0||. \tag{24}$$

Then for $z \in B(y, r)$ $\tau(z, y_0)$ is unique and $\tau(z, y_0) = \alpha \neq 0$, which is a contradiction with Equation (23). We have therefore proved the following theorem:

Theorem 1. *Let Y be a random variable of density h, whose expected value has a isotropy group reduced to $\{0\}$. If $\mathrm{Supp}(h)$ contains a neighbourhood of a point y such that $\tau(y, \mathbb{E}(Y)) \neq 0$ (which means that y is strictly closer to $\tau \cdot \mathbb{E}(Y)$ with some $\tau \neq 0$ than $\mathbb{E}(Y)$ itself), then we can say that $[\mathbb{E}(Y)]$ is not a Fréchet mean of $[Y]$ in the quotient of the space quotiented by the action of translations.*

2 Example in a very simple torus

In the previous part, we have shown that the expected value of Y can not be estimated by the Fréchet mean estimator. But we did not say how far this expected value of Y was from the set of all the Fréchet means in the quotient space. In this section, we take a very simple example: we take only two pixels: in other words here we work with $\mathbb{R}^{\mathbb{T}}$ where $\mathbb{T} = \mathbb{Z}/2\mathbb{Z}$, we can identify $\mathbb{R}^{\mathbb{T}}$ with $\mathbb{R}^2$ and work with the canonical basis of $\mathbb{R}^2$, we note by (u, v) the coordinates of an element of $\mathbb{R}^{\mathbb{T}}$. $0 \cdot (u, v) = (u, v)$ and $1 \cdot (u, v) = (v, u)$,. We note $L = \{(u, u),\ u \in \mathbb{R}\}$, and $\mathrm{HP} = \{(u, v),\ v > u\}$ the half-plane above the line L. Here we suppose that Y follows a Gaussian law of variance σ^2 and expected value $\mathbb{E}(Y) \in \mathrm{HP}$.

2.1 Graphical Interpretation of the Presence of the Bias

In this subpart we explain why there is a biasin this situation. Let $\mu \in \mathrm{HP}$ [8] (as it is the case for $\mathbb{E}(Y)$). We remind that $\tau(y, \mu)$ is an element of $\mathbb{T}$ which minimises $||\tau \cdot y - \mu||$ see (10)):

- If $y \in \mathrm{HP}$ then $\tau(y, \mu) \cdot y = y$, because μ, y are in the same half-plane delimited by L, and L is the perpendicular bisector of y and $1 \cdot y$.
- If $y \notin \overline{\mathrm{HP}}$ then $\tau(y, \mu) \cdot y = 1 \cdot y \in \mathrm{HP}$.

For $\mu \in \mathrm{HP}$, we define $Z = \tau(Y, \mu) \cdot Y$ (Z do not depend of $\mu \in \mathrm{HP}$ see above) we have $J(\mu) = \mathbb{E}(||Z - \mu||^2)$.

Lemma 1. *The global minimums of J are exactly: $\mathbb{E}(Z)$ and $1 \cdot \mathbb{E}(Z)$.*

Proof. Let μ_0 be a global minimum of J, we know that $J(\mu) = J(1 \cdot \mu)$. Without loss of generality we can assume that $\mu_0 \in \overline{\mathrm{HP}}$. Now as $\mathbb{E}(Z)$ is the expected value of Z we know that $\mathbb{E}(Z)$ is the only point where the variance of Z:

$$\begin{pmatrix} \mathbb{R}^{\mathbb{T}} \rightarrow & \mathbb{R}^+ \\ \mu \mapsto & \mathbb{E}(||Z - \mu||^2) \end{pmatrix} \tag{25}$$

is minimal, moreover we know that Z takes value in $\overline{\mathrm{HP}}$, then for convexity reason $\mathbb{E}(Z) \in \overline{\mathrm{HP}}$. Then by restriction to $\overline{\mathrm{HP}}$, $\mathbb{E}(Z)$ is still the unique minimum of:

$$\begin{pmatrix} \overline{\mathrm{HP}} \rightarrow & \mathbb{R}^+ \\ \mu \mapsto & \mathbb{E}(||Z - \mu||^2) \end{pmatrix}. \tag{26}$$

As a conclusion we have $\mu_0 = \mathbb{E}(Z)$. □

When we represent graphically these two random variables Y, Z, we can see that Y, Z have different means. This case shows graphically the bias. On the Fig. 1(a) and Fig. 1(b), we see the noise's influence: more the noise is important, more the mass under the line L is big and more the mean of Z is far from the expected value of Y. On the Fig. 1 we understand the condition of the density in the theorem 1, if the density's support is too small, then there are no mass under the line L, therefore $Y = Z$ and in this case there is no bias.

2.2 Localize the Fréchet mean

Thanks to the lemma 1, we can compute a Fréchet mean by computing $\mathbb{E}(Z)$, which is the sum of the area of the grey part (for the density h) and of the area of the black part (for the density h) in the Fig. 1(b), we have:

$$\mathbb{E}(Z) = \int_{v>u} (u, v)h(u, v)dudv + \int_{v<u} (v, u)h(u, v)dudv, \tag{27}$$

[8] For symmetry reason, because $J(\mu) = J(\tau\mu)$.

where (u, v) are the coordinates of a point in $\mathbb{R}^{\mathbb{T}} \simeq \mathbb{R}^2$. To compute (27) we convert to polar coordinates: $(u, v) = \mathbb{E}(Y) + (r\cos\theta, r\sin\theta)$ where $r > 0$ et $\theta \in [0, 2\pi]$. We also define: $d = \text{dist}(\mathbb{E}(Y), L)$. We get:

$$\mathbb{E}(Z) = \mathbb{E}(Y) + \int_d^{+\infty} \frac{r^2 \exp(-\frac{r^2}{2\sigma^2})}{\pi\sigma^2} \sqrt{2} g\left(\frac{d}{r}\right) dr \times (-1, 1), \tag{28}$$

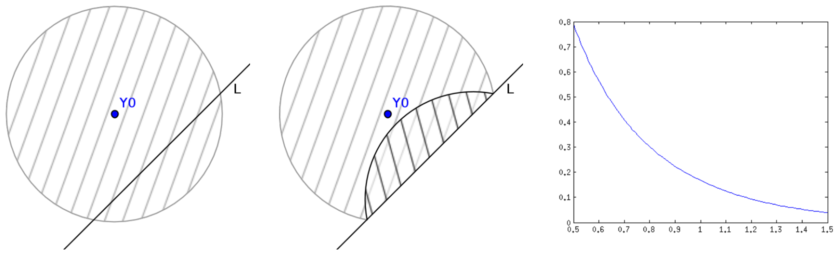
where g is a non-negative function on $[0, 1]$ defined by $g(x) = \sin(\arccos(x)) - x\arccos(x)$. Here we want to compute $\tilde{\rho} = \rho([\mathbb{E}(Y)], [\mathbb{E}(Z)])$ where ρ is the distance in the quotient space defined in (2). As we know that $\mathbb{E}(Y)$, $\mathbb{E}(Z)$ are in the same half-plane delimited by L, we have:

$$\tilde{\rho} = \rho([\mathbb{E}(Y)], [\mathbb{E}(Z)]) = ||\mathbb{E}(Y) - \mathbb{E}(Z)|| = 2\int_d^{+\infty} \frac{r^2 \exp(-\frac{r^2}{2\sigma^2})}{\pi\sigma^2} g\left(\frac{d}{r}\right) dr. \tag{29}$$

We can conclude that:

Theorem 2. *The Fréchet mean of $[Y]$ in the quotient space is an orbit of two points which are on the line passing through the expected value of Y and perpendicular to L and we compute the relative gap between the bias $\tilde{\rho}$ and $d = dist(\mathbb{E}(Y), L)$ by:*

$$\frac{\tilde{\rho}}{d} = \frac{\sigma}{d}\frac{2}{\pi}\int_{\frac{d}{\sigma}}^{+\infty} r^2 \exp\left(-\frac{r^2}{2}\right) g\left(\frac{d}{r\sigma}\right) dr. \tag{30}$$



(a) Graphic representation of Y, $y_0 = \mathbb{E}(Y)$ (b) Graphic representation of Z, $y_0 \neq \mathbb{E}(Z)$ (c) Variation of $\frac{\tilde{\rho}}{d}$ in function of $\frac{d}{\sigma}$.

Fig. 1. Z and Y have not the same mean, therefore there is a bias.

Remark 4. Here, contrarily to [8], it is not the ratio $||\mathbb{E}(Y)||$ over the noise which matters to estimate the bias, but the ratio $\text{dist}(\mathbb{E}(Y), L)$ over the noise which matters. But in fact, there it is not so different, in both case we measure the distance between the signal and the singularities (which is $\{0\}$ in [8] for the action of rotations, L in this case).

Discussion

In this article we have compared two notions of mean, one is the expectation of our random variable in our linear space, the other is the Fréchet mean in the space quotiented by translations. By differencing the variance in the quotient space we managed to show that when our random variable has a density whose support is large enough due to noise, the template in the total space mapped in the quotient space is not a Fréchet mean. But is the template mapped in the quotient space close to the Fréchet mean in the quotient space? We have answered to this question only in a special case of torus by computing the bias with a Gaussian noise. In this case, the bias depends on the scale of the noise and on the regularity of the signal, (measured here by how far our signal is from the set of singularity). In future work, we will generalise this estimation of the bias for a general torus $\mathbb{T} = (\mathbb{Z}/N\mathbb{Z})^D$, in order to see the influence of N (the number of pixels for each side of the picture) and D (the dimension of the picture) on the size of the bias. We have also seen the role played by the nature of the isotropy group for the presence of bias. This was already observed in [8], restricted to Gaussian noise.

In the section 1, we showed a bias for the Fréchet mean estimator with a particular group action defined in (1). But we have never used that definition. We have only used some properties of this group action: a finite group acts isometrically and effectively on a finite dimensional vector space. Therefore the theorem 1 generalises to any group with these properties.

What if the isotropy group is not reduced to $\{0\}$? Suppose now that G is a finite group acting isometrically and effectively on $\mathbb{R}^n$. Let suppose that $\mathbb{E}(Y)$ is singular, and that $\mathrm{Supp}(h)$ contains a neighbourhood of a point y such that y is strictly closer to $g_0 \cdot \mathbb{E}(Y)$ than $\mathbb{E}(Y)$ with some $g_0 \in G$. We make the extra assumption that it exists K a subgroup of G with $G = \{k \times i,\ k \in K,\ i \in Iso(\mathbb{E}(Y))\}$ and $K \cap \mathrm{Iso}(\mathbb{E}(Y)) = \{e_G\}$, then we have the same result: Indeed, by noting $y_0 = \mathbb{E}(Y)$, we have:

$$[y_0] = \{g \cdot y_0,\ g \in G\} = \{g \cdot y_0,\ g \in K\} \tag{31}$$

Therefore if we note:

$$\tilde{J}(\mu) = \int_{\mathbb{R}^n} \inf_{g \in K} ||y - g \cdot \mu||^2 h(y) dy \tag{32}$$

Then we have: $J(y_0) = \tilde{J}(y_0)$ and for all $\mu \in \mathbb{R}^n$ $\tilde{J}(\mu) \geq J(\mu)$, now if we consider the action of G but restricted to K (the action is still isometric, and effective: since no element leaves y_0, moreover by writing $g_0 = k \times i$ with $iy_0 = y_0$, and $k \in K$, y is strictly closer to $k \cdot \mathbb{E}(Y)$ than $\mathbb{E}(Y)$), then we know (by the theorem (1), with y_0 a regular point for the action of the group K) that it exists $\mu \in \mathbb{R}^n$ such that $\tilde{J}(\mu) < \tilde{J}(y_0)$, therefore $J(y_0) = J(\tilde{y}_0) > \tilde{J}(\mu) \geq J(\mu)$, then y_0 is still not a minimum of the variance J, therefore the expected value of Y mapped in the quotient space is not a Frechet mean of $[Y]$ even in this case.

In a more general case: when we take an infinite-dimensional vector space quotiented by a group action, for instance when the group is a subgroup of the group of smooth diffeomorphism, is there always a bias? And when it does, can we measure the bias in function of the scale of the noise and the distance between the template and the singularities? Figure 1(c) shows us that the bias is not so important in favourable cases: when the noise is low and the signal far from the singularities. Then we can hope that it will be also the case in a more general case. If so, one could keep using the Fréchet mean in the quotient space in order to estimate the template in the total space.

References

1. Allassonnière, S., Amit, Y., Trouvé, A.: 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)
2. Allassonnière, S., Kuhn, E., Trouvé, A.: Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. Bernoulli 16(3), 641–678 (2010)
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 1, 1–22 (1977)
4. Grenander, U., Miller, M.I.: Computational anatomy: An emerging discipline. Q. Appl. Math. LVI(4), 617–694 (Dec 1998)
5. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage 23, S151–S160 (2004)
6. Kendall, D.G.: A survey of the statistical theory of shape. Statistical Science pp. 87–99 (1989)
7. Kurtek, S.A., Srivastava, A., Wu, W.: Signal estimation under random time-warpings and nonlinear signal alignment. In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K. (eds.) Advances in Neural Information Processing Systems 24, pp. 675–683. Curran Associates, Inc. (2011)
8. Miolane, N., Pennec, X.: Biased estimators on quotient spaces. Geometric Science of Information (2015)
9. Sabuncu, M., Balci, S.K., Golland, P.: Discovering modes of an image population through mixture modeling. Proceeding of the MICCAI conference LNCS(5242), 381–389 (2008)
10. Zhang, M., Singh, N., Fletcher, P.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: Gee, J., Joshi, S., Pohl, K., Wells, W., Zöllei, L. (eds.) Information Processing in Medical Imaging, Lecture Notes in Computer Science, vol. 7917, pp. 37–48. Springer Berlin Heidelberg (2013)