Electronic Journal of Statistics
Vol. 5 (2011) 1054–1089
ISSN: 1935-7524
DOI: 10.1214/11-EJS633

On the consistency of Fréchet means in deformable models for curve and image analysis

Jérémie Bigot and Benjamin Charlier

Institut de Mathématiques de Toulouse
Université de Toulouse et CNRS (UMR 5219), France
e-mail:
Jeremie.Bigot@math.univ-toulouse.fr; Benjamin.Charlier@math.univ-toulouse.fr

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

AMS 2000 subject classifications: Primary 62G08; secondary 42C40.
Keywords and phrases: Mean pattern estimation, Fréchet mean, shape analysis, deformable models, curve registration, image warping, geometric variability, high-dimensional data.

Received March 2011.

1. Introduction

1.1. A statistical deformable model for curve and image analysis

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

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

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

settings: either when both the number n of design points and the number J of curves (or images) tend to infinity, or when n (resp. J) remains fixed while J (resp. n) tends to infinity.

In many situations, data sets of curves or images exhibit a source of geometric variations in time or shape. In such settings, the usual Euclidean mean $\bar{Y}^\ell = \frac{1}{J}\sum_{j=1}^{J} Y_j^\ell$ in model (1.1) cannot be used to recover a meaningful mean pattern. Indeed, consider the following simple model of randomly shifted curves (with $d = 1$) which is commonly used in many applied areas such as neuroscience [TIR10] or biology [Røn01],

$$f_j(t_\ell) = f(t_\ell - \boldsymbol{\theta}_j^*), \quad j = 1, \ldots, J, \text{ and } \ell = 1, \ldots, n, \tag{1.2}$$

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

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

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

The main goal of this paper is to introduce statistical deformable models for curve and image analysis that are analogue to Goodall's perturbation models [Goo91], and to build consistent estimators of a mean pattern in such models. Our approach is inspired by Grenander's pattern theory which considers that the curves or images f_j in model (1.1) are obtained through the deformation of a mean pattern by a Lie group action [Gre93, GM07]. In the last decade, there has been a growing interest in transformation Lie groups to model the geometric variability of images, and the study of the properties of such deformation groups is now an active field of research (see e.g. [MY01, TY05] and

references therein). There is also currently a growing interest in statistics on the use of Lie group actions to analyze geometric modes of variability of a data set [HHM10a, HHM10b].

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

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

$$T_{\boldsymbol{\theta}} f(t) := f\left(e^a R_\alpha t - b\right), \qquad \text{for } \boldsymbol{\theta} = (a, \alpha, b) \in \mathcal{P},$$

  with $\Omega = \mathbb{R}^2$, $\mathcal{P} \subset \mathbb{R} \times \mathbb{R} \times \mathbb{R}^2$ where $R_\alpha = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$ is a rotation matrix in $\mathbb{R}^2$, e^a is an isotropic scaling and b a translation in $\mathbb{R}^2$.
- *Deformation by a Lie group action:* the two above cases are examples of a Lie group action on the space $L^2(\Omega)$ (see [Hel01] for an introduction to Lie groups). More generally, assume that G is a connected Lie group of dimension p acting on Ω, meaning that for any $(g, t) \in G \times \Omega$ the action $\cdot$ of G onto Ω is such that $g \cdot t \in \Omega$. In general, G is not a linear space but can be locally parametrized by a its Lie algebra $\mathcal{G} \simeq \mathbb{R}^p$ using the exponential map $\exp : \mathcal{G} \to G$. If $\mathcal{P} \subset \mathbb{R}^p$. This leads for $(\boldsymbol{\theta}, f) \in \mathcal{P} \times L^2(\Omega)$ to define the deformation operators

$$T_{\boldsymbol{\theta}} f(t) := f\left(\exp(\boldsymbol{\theta}) \cdot t\right).$$

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

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

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

$$f_j(t) = T_{\boldsymbol{\theta}^*_j}\left[f + Z_j\right](t), \qquad \text{for all } t \in \Omega, \text{ and } j = 1, \ldots, J, \tag{1.3}$$

where $\boldsymbol{\theta}^*_j \in \mathcal{P}$, $j = 1, \ldots, J$ are i.i.d. random variables (independent of the ε^ℓ_j's) with an unknown density g with compact support Θ included in $\mathcal{P}$ satisfying:

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

$$\Theta = \left\{\boldsymbol{\theta} = (\theta^1, \ldots, \theta^p) \in \mathbb{R}^p,\ |\theta^{p_1}| \le \rho,\ 1 \le p_1 \le p\right\} \tag{1.4}$$

where $\rho > 0$.

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

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

where ε_j^ℓ are i.i.d. normal variables with zero mean and variance 1.

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

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

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

$$\mathbf{Y}_j = T_{\boldsymbol{\theta}_j^*}(\mu + \boldsymbol{\zeta}_j), \ j = 1, \dots, J. \tag{1.6}$$

Model (1.6) is similar to the statistical deformable model (1.5), where $\mu \in \mathbb{R}^{n\times 2}$ is the unknown perturbation mean to estimate, and $\boldsymbol{\zeta}_j$ are i.i.d. random vectors in $\mathbb{R}^{n\times 2}$ with zero mean. Nevertheless, there exists major differences between our approach and the one in [Goo91]. First, in model (1.5), the deformations parameters $\boldsymbol{\theta}_j^*$ are assumed to be random variables following an unknown distribution, whereas they are just nuisance parameters in model (1.6) for shape analysis, see [Goo91, KM97]. In some applications (e.g. in biomedical imaging [JDJG04]), it is of interest to reconstruct the unobserved parameters $\boldsymbol{\theta}_j^*$ and to estimate their distribution. One of the main contribution of this paper is then to construct upper and lower bounds for the estimation of such deformation parameters. Moreover, in model (1.5), they are too additive error terms, whereas the model (1.6) only include the error term $\boldsymbol{\zeta}_j$. In model (1.5), the ε_j^ℓ is an additive noise modeling the errors in the measurements, while the Z_j's model (possibly smooth) variations in intensity of the individuals around the mean pattern f.

In [KM97], the authors studied the relationship between isotropicity of the additive noise $\boldsymbol{\zeta}_j$ and the convergence of Procrustean procedures to the perturbation mean μ as $J \to +\infty$. It is shown in [KM97] that, for isotropic errors, Procrustean means are consistent, but that, for non-isotropic errors, they may not converge to μ. For a recent discussion on the issues of consistency of sample Procrustes means in perturbation models and extension to non-metrical Fréchet means, we refer to [Huc10] and [Huc11]. In this paper, we carefully analyze the

role of the dimension n and the number of samples J on the consistency of Procrustean means in model (1.5). To obtain consistent procedures, we show that it is not required to impose very restrictive conditions on the error terms Z_j such as isotropicity for the $\boldsymbol{\zeta}_j$ in (1.6) for shape analysis. Here, the key quantity is the dimension n of the data (number of design points) which plays the central role to guarantee the converge of our estimators. This point is another major difference with the approach of statistical shape analysis [Goo91] that does not take into account the dimensionality of the shape space to analyze the consistency of Procrustean estimators.

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

$$Y_j^\ell = T_{\boldsymbol{\theta}_j^*} f(t_\ell) + \sigma \varepsilon_j^\ell, \quad j = 1, \ldots, J, \quad \text{and} \quad \ell = 1, \ldots, n, \tag{1.7}$$

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

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

1.2. Organization of the paper

Section 2 contains a description of our estimating procedure and a review of previous work in mean pattern estimation. In Section 3, we derive a lower bound for the quadratic risk of estimators of the deformation parameters. In Section 4, we discuss some identifiability issues in model (1.5). In Section 5 we derive consistency results for the Fréchet mean in the case (1.2) of randomly shifted curves. In Section 6 and Section 7, we give general conditions to extend these results to the more general deformable model (1.5). Section 8 contains some numerical experiments. A small conclusion with some perspectives are given in Section 9. All proofs are postponed to a technical Appendix.

2. The estimating procedure

2.1. A dissimilarity measure based on deformation operators

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

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

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

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

If $d_T^2(h,f) = 0$ then there exists $\boldsymbol{\theta} \in \mathcal{P}$ such that $f = \tilde{T}_{\boldsymbol{\theta}} h$ meaning that the functions f and h are equal up to a geometric deformation. Note that d_T is not necessarily a distance on $L^2(\Omega)$, but it can be used to define a notion of sample Fréchet mean of data from model (1.5). For this purpose let $\mathcal{F}$ denote a subspace of $L^2(\Omega)$ and suppose that $\hat{f}_j$ are smooth functions in $\mathcal{F} \subset L^2(\Omega)$ obtained from the data Y_j^ℓ, $\ell = 1, \ldots, n$ for $j = 1, \ldots, J$, see Section 5.2 and Section 6.2 for precise definitions. Following the definition of a Fréchet mean in general metric space [Fré48], define an estimator of the mean pattern f as

$$\hat{f} = \operatorname*{argmin}_{f\in\mathcal{F}} \frac{1}{J} \sum_{j=1}^{J} d_T^2(\hat{f}_j, f). \tag{2.1}$$

Note that $\hat{f}$ falls into the category of non-metrical sample Fréchet means whose definitions and asymptotic properties are discussed in [Huc10] for random variables belonging to Riemannian manifolds. However, unlike the usual approach in shape analysis, the Fréchet mean (2.1) is based on smoothed data. In what follows, we show that smoothing is a key preliminary step to obtain the convergence of $\hat{f}$ to the mean pattern f in the deformable model (1.5). It can be easily shown that the computation of $\hat{f}$ can be done in two steps: first minimize the following criterion

$$(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_J) = \operatorname*{argmin}_{(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_J)\in\Theta^J} M(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J), \tag{2.2}$$

where

$$M(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) = \frac{1}{J} \sum_{j=1}^{J} \int_\Omega \left(\tilde{T}_{\boldsymbol{\theta}_j} \hat{f}_j(t) - \frac{1}{J} \sum_{j'=1}^{J} \tilde{T}_{\boldsymbol{\theta}_{j'}} \hat{f}_{j'}(t)\right)^2 dt, \tag{2.3}$$

which gives an estimation of the deformation parameters $\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_J^*$, and then in a second step take

$$\hat{f}(t) = \frac{1}{J} \sum_{j=1}^{J} \tilde{T}_{\hat{\boldsymbol{\theta}}_j} \hat{f}_j(t), \qquad \text{for } t \in \Omega, \tag{2.4}$$

as an estimator of the mean pattern f.

Note that this two steps procedure belongs to the category of Procrustean methods (see e.g [DM98, Goo91]). A similar approach to (2.2) has been developed by [JDJG04] in the context of biomedical images using diffeomorphic deformation operators.

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

Estimating the mean pattern of a set of curves that differ by a time transformation is usually referred to as the curves registration problem, see e.g. [GK92, Big06, RL01, WG97, LM04]. However, in these papers, studying consistent estimators of the mean pattern f as the number of curves J and design points n tend to infinity is not considered. For the SIM (1.7), a semiparametric point of view has been proposed in [GLM07] and [Vim10] to estimate non-random deformation parameters (such as shifts and amplitudes) as the number n of observations per curve grows, but with a fixed number J of curves. A generalisation of this semiparametric approach for two-dimensional images is proposed in [BGV09]. The case of image deformations by a Lie group action is also investigated in [BLV10] from a semiparametric point of view using a SIM.

In the simplest case of randomly shifted curves in a SIM, [BG10] have studied minimax estimation of the mean pattern f by letting only the number J of curves going to infinity. Self-modelling regression (SEMOR) methods proposed by [KG88] are semiparametric models where each observed curve is a parametric transformation of the same regression function. However, the SEMOR approach does not incorporate a random fluctuations in intensity of the individuals around a mean pattern f through an unknown process Z_j as in model (1.5). The authors in [KG88] studied the consistency of the SEMOR approach using a Procrustean algorithm. Recently, there has also been a growing interest on the development of statistical deformable models for image analysis and the construction of consistent estimators of a mean pattern, see [GM01, BGV09, BGL09, AAT07, AKT09].

3. Lower bounds for the estimation of the deformation parameters

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

Assumption 3.1. *For all $\boldsymbol{\theta} = (\theta^1, \ldots, \theta^p) \in \mathcal{P}$, $T_{\boldsymbol{\theta}} : L^2(\Omega) \longrightarrow L^2(\Omega)$ is a linear operator such that the function $t \longmapsto \partial_{\theta^{p_1}} T_{\boldsymbol{\theta}} f(t)$ exists and belongs to $L^2(\Omega)$ for any $p_1 = 1, \ldots, p$. Moreover, there exists a constant $C(\Theta, f) > 0$ such that*

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

for all $p_1 = 1, \ldots, p$ and $\boldsymbol{\theta} \in \Theta$.

3.1. Shape invariant model

Theorem 3.1. *Consider the SIM (1.7) and suppose that Assumption 3.1 holds. Assume that g satisfies Assumption 1.1, and that $\int_\Theta \|\partial_{\boldsymbol{\theta}} \log(g(\boldsymbol{\theta}))\|^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta} <$*

$+\infty$. *Let* $\hat{\boldsymbol{\theta}} \in \mathcal{P}^J$ *be any estimator (a measurable function of the data) of* $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_J^*)$. *Then, for any* $n \geq 1$ *and* $J \geq 1$,

$$\mathbb{E}\left[\frac{1}{J}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_{\mathbb{R}^{pJ}}^2\right] \geq \frac{\sigma^2 n^{-1}}{C(\Theta, f) + \sigma^2 n^{-1} \int_\Theta \|\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta})\right)\|^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \tag{3.1}$$

where $C(\Theta, f)$ *is the constant defined in Assumption 3.1, and* $\|\cdot\|_{\mathbb{R}^{pJ}}$ *is the standard Euclidean norm in* $\mathbb{R}^{pJ}$.

The lower bound given in inequality (3.1) does not decrease as J increases. Thus, if the number n of design points is fixed, increasing the number J of curves (or images) does not improve the quality of the estimation of the deformation parameters for any estimator $\hat{\boldsymbol{\theta}}$. Nevertheless, this lower bound is going to 0 as the dimension $n \to +\infty$.

3.2. General model

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

Assumption 3.2. *There exists a positive semi-definite symmetric* $n \times n$ *matrix* $\boldsymbol{\Sigma}_n(\Theta)$ *such that the covariance matrix of* $\mathbf{Z} = [Z(t_\ell)]_{\ell=1}^n$ *satisfies*

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\mathbf{T}_{\boldsymbol{\theta}}\mathbf{Z}(\mathbf{T}_{\boldsymbol{\theta}}\mathbf{Z})'\right] = \boldsymbol{\Sigma}_n(\Theta).$$

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

Theorem 3.2. *Consider the general model* (1.5). *Suppose that Assumption 3.1 and 3.2 hold. Assume that the density* g *satisfies Assumption 1.1 and that* $\int_\Theta \|\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta})\right)\|^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta} < +\infty$. *Let* $\hat{\boldsymbol{\theta}} \in \mathcal{P}^J$ *be any estimator (a measurable function of the data) of* $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_J^*)$. *Then, for any* $n \geq 1$ *and* $J \geq 1$, *we have*

$$\mathbb{E}\left[\frac{1}{J}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_{\mathbb{R}^{pJ}}^2\right] \geq \frac{(\sigma^2 + s_n^2(\Theta)) n^{-1}}{C(\Theta, f) + (\sigma^2 + s_n^2(\Theta)) n^{-1} \int_\Theta \|\partial_{\boldsymbol{\theta}_1} \log\left(g(\boldsymbol{\theta})\right)\|^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \tag{3.2}$$

where $C(\Theta, f)$ *is the constant defined in Assumption 3.1, and* $s_n^2(\Theta)$ *denotes the smallest eigenvalue of* $\boldsymbol{\Sigma}_n(\Theta)$.

Again, the lower bound (3.2) does not depends on J. Thus, increasing the number J of observations does not decrease the quadratic risk of any estimator of the deformations parameters. Moreover, the lower bound (3.2) tends to zero as $n \to +\infty$ only if $\lim_{n\to+\infty} n^{-1}s_n^2(\Theta) = 0$.

3.3. Application to the shifted curves model

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

$$Y_j^\ell = f\left(\tfrac{\ell}{n} - \boldsymbol{\theta}_j^*\right) + Z_j\left(\tfrac{\ell}{n} - \boldsymbol{\theta}_j^*\right) + \sigma\varepsilon_j^\ell, \quad j = 1,\dots,J, \quad \text{and} \quad \ell = 1,\dots,n. \tag{3.3}$$

Theorem 3.3. *Consider the model (3.3). Assume that f is continuously differentiable on $[0,1]$ and that Z is a centered stationary process with value in $L^2_{per}([0,1])$. Suppose that $\Theta = [-\rho,\rho]$ with $\rho < \frac{1}{2}$ and $\int_\Theta \left(\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta})\right)\right)^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta} < +\infty$. Let $\hat{\boldsymbol{\theta}} \in \mathbb{R}^J$ be any estimator of the true random shifts $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*,\dots,\boldsymbol{\theta}_J^*)$, i.e. a measurable function of the data in model (3.3). Then, for any $n \geq 1$ and $J \geq 1$*

$$\mathbb{E}\left[\frac{1}{J}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2_{\mathbb{R}^J}\right] \geq \frac{n^{-1}\sigma^2}{\|\partial_t f\|^2_\infty + n^{-1}\sigma^2 \int_\Theta \left(\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta})\right)\right)^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta}}, \tag{3.4}$$

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

4. Identifiability conditions

4.1. The shifted curves model

Without any further assumptions, the randomly shifted curves model (3.3) is not identifiable. Indeed, if $\boldsymbol{\theta}_0 \in \Theta$ satisfies $\boldsymbol{\theta}_j^* + \boldsymbol{\theta}_0 \in \Theta$, $j = 1,\dots,J$, then replacing $f(\cdot)$ by $f(\cdot - \boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}_j^*$ by $\boldsymbol{\theta}_j^* + \boldsymbol{\theta}_0$ does not change the formulation of model (3.3). Choosing identifiability conditions amounts to impose constraints on the minimization of the criterion

$$D(\boldsymbol{\theta}) = \frac{1}{J}\sum_{j=1}^J \int_\Omega \left(f(t - \boldsymbol{\theta}_j^* + \boldsymbol{\theta}_j) - \frac{1}{J}\sum_{j'=1}^J f(t - \boldsymbol{\theta}_{j'}^* + \boldsymbol{\theta}_{j'})\right)^2 dt, \tag{4.1}$$

for $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_J) \in \Theta^J$, which can be interpreted as a version without noise of the criterion (2.2) using the ideal smoothers $\hat{f}_j(\cdot) = f(\cdot - \boldsymbol{\theta}_j^*)$. Obviously, the criterion $D(\boldsymbol{\theta})$ has a minimum at $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*,\dots,\boldsymbol{\theta}_J^*)$ such that $D(\boldsymbol{\theta}^*) = 0$, but this minimizer of D on Θ^J is clearly not unique. If the true shifts are supposed to have zero mean (i.e. $\int_\Theta \boldsymbol{\theta} g(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0$) it is natural to introduce the constrained set

$$\boldsymbol{\Theta}_0 = \{(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 + \cdots + \boldsymbol{\theta}_J = 0\}. \tag{4.2}$$

It is shown in [BG10] Lemma 6, that if $f \in L^2([0,1])$ is such that $\int_0^1 f(t)e^{-i2\pi t}dt \neq 0$ and if $\rho < 1/4$ (recall that $\Theta = [-\rho,\rho]$), then the criterion $D(\boldsymbol{\theta})$ has a unique

minimum on $\boldsymbol{\Theta}_0$ in the sense that $D(\boldsymbol{\theta}) > D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ with $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}$ where

$$\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} = (\boldsymbol{\theta}^*_1 - \bar{\boldsymbol{\theta}}^*, \ldots, \boldsymbol{\theta}^*_J - \bar{\boldsymbol{\theta}}^*) \text{ with } \bar{\boldsymbol{\theta}}^* = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\theta}^*_j. \tag{4.3}$$

Under such assumptions, we will compute estimators of the random shifts by minimizing the criterion (2.2) over the constrained set $\boldsymbol{\Theta}_0$ and not directly on Θ^J. Consistency of such constrained estimators will then be studied under the following identifiability conditions:

Assumption 4.1. *The mean pattern f is such that $\int_0^1 f(t)e^{-i2\pi t}dt \neq 0$.*

Assumption 4.2. *The support of the density g is included in $[-\rho', \rho']$ for some $0 < \rho' \leq \frac{\rho}{2} < 1/4$ and is such that $\int_\Theta \boldsymbol{\theta} g(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0$.*

Under such assumptions, $D(\boldsymbol{\theta})$ can be bounded from below by the quadratic function $\frac{1}{J}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ which will be an important property to derive consistent estimators.

Proposition 4.1. *Suppose that Assumptions 4.1 and 4.2 hold with $\rho < 1/16$. Then, for any $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \in \boldsymbol{\Theta}_0$, one has that*

$$D(\boldsymbol{\theta}) - D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}) \geq C(f, \rho)\frac{1}{J}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2,$$

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

Assumption 4.2 and the condition that $\rho < 1/16$ in Proposition 4.1 mean that the support of the density g of the shifts is sufficiently small, and that the shifted curves $f_j(t) = f(t - \boldsymbol{\theta}^*_j)$ are in some sense concentrated around the mean pattern f. Such an assumption of concentration of the data around the same mean pattern has been used in various papers to prove the uniqueness and the consistency of Fréchet means for random variables lying in a Riemannian manifold, see [Kar77, Le98, BP03, Afs11, Ken90].

4.2. The general case

In the case of general deformation operators, define for $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \in \Theta^J$ the criterion

$$D(\boldsymbol{\theta}) = \frac{1}{J}\sum_{j=1}^{J}\int_\Omega \left(\tilde{T}_{\boldsymbol{\theta}_j} T_{\boldsymbol{\theta}^*_j} f(t) - \frac{1}{J}\sum_{j'=1}^{J} \tilde{T}_{\boldsymbol{\theta}_{j'}} T_{\boldsymbol{\theta}^*_{j'}} f(t)\right)^2 dt. \tag{4.4}$$

Obviously, using that for all $\boldsymbol{\theta} \in \Theta$, $\tilde{T}_{\boldsymbol{\theta}} T_{\boldsymbol{\theta}} f = f$, the criterion $D(\boldsymbol{\theta})$ has a minimum at $\boldsymbol{\theta}^* = (\boldsymbol{\theta}^*_1, \ldots, \boldsymbol{\theta}^*_J)$ such that $D(\boldsymbol{\theta}^*) = 0$. However, without any further restrictions the minimizer of $D(\boldsymbol{\theta})$ is not necessarily unique on Θ^J.

Assumption 4.3. *Let $\boldsymbol{\Theta} \subset \Theta^J$ such that there exists a unique $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}} \in \boldsymbol{\Theta}$ satisfying $D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}) = 0$.*

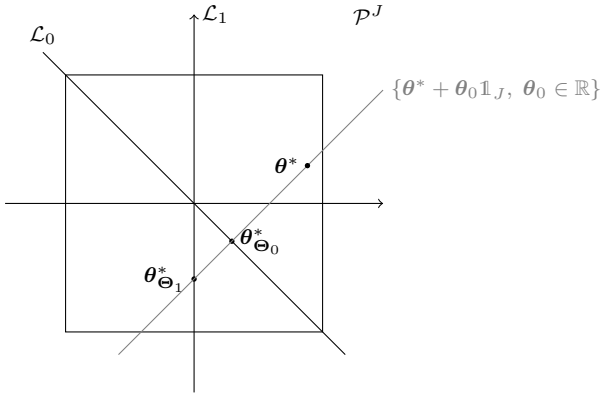


FIGURE 1. *Choice of identifiability conditions for shifted curves in the case* $J = 2$.

Then, $\boldsymbol{\Theta}$ is the set onto which we will carry the minimization of the criterion $M(\boldsymbol{\theta})$ (2.3). In the case of shifted curves and under Assumption 4.1 and 4.2, the only set onto which the criterion D vanishes is the line $\{\boldsymbol{\theta}^* + \boldsymbol{\theta}_0 \mathbb{1}_J,\ \boldsymbol{\theta}_0 \in \mathbb{R}\} \subset \mathbb{R}^J$, where $\mathbb{1}_J = (1, \ldots, 1)' \in \mathbb{R}^J$. An easy way to choose the set $\boldsymbol{\Theta}$ is to take a linear subset of Θ^J, see Figure 1 for an illustration. By considering the subset

$$\boldsymbol{\Theta}_0 = \Theta^J \cap \mathbb{1}_J^{\perp} = \{(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 + \cdots + \boldsymbol{\theta}_J = 0\},$$

where ${\mathbb{1}_J}^{\perp}$ is the orthogonal of $\mathbb{1}_J$ in $\mathbb{R}^J$, then Assumption 4.3 is satisfied with $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}$ given in (4.3). More generally, if the deformation parameters $\boldsymbol{\theta}_j$, $j = 1, \ldots, J$ are supposed to be random variables with zero mean, then optimizing $D(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}_0$ is a natural choice. Another identifiability condition for shifted curves is proposed in [GLM07] and [Vim10] by taking

$$\boldsymbol{\Theta}_1 = \Theta^J \cap {e_1}^{\perp} = \{(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 = 0\}. \tag{4.5}$$

where $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^J$. In this case, $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_1} = (0, \boldsymbol{\theta}^*_2 - \boldsymbol{\theta}^*_1, \ldots, \boldsymbol{\theta}^*_J - \boldsymbol{\theta}^*_1)$. Choosing to minimize $D(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}_1$ amounts to choose the first curve as a reference onto which all the others curves are aligned, meaning that the first shift $\boldsymbol{\theta}^*_1$ is not random, see Figure 1.

Following the classical guidelines in M-estimation (see e.g. [vdV98]), a necessary condition to ensure the convergence of M -estimators such as (2.2) is that the local minima of $D(\boldsymbol{\theta})$ over $\boldsymbol{\Theta}$ are well separated from the global minimum of $D(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$ (satisfying $D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}) = 0$). The following assumption can be interpreted in this sense.

Assumption 4.4. *For all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *we have*

$$D(\boldsymbol{\theta}) - D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}) \geq C(\boldsymbol{\Theta}, \mathcal{F}) \frac{1}{J} \|\boldsymbol{\theta} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\|^2 \tag{4.6}$$

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

In the shifted curve model, Assumption 4.4 is verified if Assumption 4.1 and 4.2 hold (see Proposition 4.1).

5. Consistent estimation in the shifted curves model

In this section, we give conditions to ensure consistency of the estimators defined in Section 2 in the shifted curves model (3.3) with an equi-spaced design.

5.1. The random perturbations Z_j

Following the assummtions of Theorem 3.3, Z will be supposed to be a stationary process Z with covariance function $R : [0,1] \longrightarrow \mathbb{R}$. The law of Z is thus invariant by the action of a shift. Conditionally to $\boldsymbol{\theta}_j^* \in \Theta$, the covariance of the vector $\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j = \left[Z_j(\frac{\ell}{n} - \boldsymbol{\theta}_j^*)\right]_{\ell=1}^n$ is a Toeplitz matrix equals to

$$\boldsymbol{\Sigma}_n = \mathbb{E}_{\boldsymbol{\theta}_j^*}\left[\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j(\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j)'\right] = \left[\mathbb{E}\left[Z\left(\tfrac{\ell}{n}\right)Z\left(\tfrac{\ell'}{n}\right)\right]\right]_{\ell,\ell'=1}^n = \left[R\left(\tfrac{|\ell-\ell'|}{n}\right)\right]_{\ell,\ell'=1}^n. \tag{5.1}$$

Let $\gamma_{\max}(\boldsymbol{\Sigma}_n)$ be the largest eigenvalue of the matrix $\boldsymbol{\Sigma}_n$. It follows from standard results on Toeplitz matrices (see e.g. [HJ90]) that

$$\gamma_{\max}\left(\boldsymbol{\Sigma}_n\right) \leq \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n}\left|R\left(\tfrac{k}{n}\right)\right| = \gamma \tag{5.2}$$

where $\gamma = \int_0^1 |R(t)|\,dt$ is a positive constant independent of n representing an upper bound of the variance of Z.

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

A convenient choice for the smoothing of the observed curves in (3.3) is to do low-pass Fourier filtering. Let $\hat{c}_{j,k} = \frac{1}{n}\sum_{\ell=1}^n Y_j^\ell e^{-i2\pi k\frac{\ell}{n}}$ for $k = -(n-1)/2, \ldots, (n-1)/2$ (assuming for simplicity that n is odd), and define for a spectral cut-off parameter $\lambda \in \mathbb{N}$ and $t \in [0,1]$ the linear estimators

$$\hat{f}_j^\lambda(t) = \sum_{|k|\leq\lambda} \hat{c}_{j,k}e^{i2\pi kt}. \tag{5.3}$$

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

$$H_s(A) = \left\{f \in L^2_{per}([0,1]), \sum_{k\in\mathbb{Z}}(1+|k|^2)^s\,|c_k(f)|^2 < A\right\}. \tag{5.4}$$

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

5.3. Consistent estimation of the random shifts

Using low-pass filtering, and following the discussion in Section 4.1 on identifiability issues, the estimators of the random shifts $\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_J^*$ are given by

$$\hat{\boldsymbol{\theta}}^\lambda = (\hat{\boldsymbol{\theta}}_1^\lambda, \dots, \hat{\boldsymbol{\theta}}_J^\lambda) = \underset{(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_J)\in\boldsymbol{\Theta}_0}{\operatorname{argmin}} M_\lambda(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J). \tag{5.5}$$

where the criterion $M_\lambda(\boldsymbol{\theta}) = M_\lambda(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J)$ for $\boldsymbol{\theta} \in \Theta^J$ is

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

and $\boldsymbol{\Theta}_0$ is the constrained set defined in (4.2).

Theorem 5.1. *Consider the model (3.3) and let $\hat{\boldsymbol{\theta}}^\lambda$ be the estimator defined by (5.5). Assume that $\mathcal{F} = H_s(A)$ for some $A > 0$ and $s \geq 1$, and that Z is a centered stationary process with value in $L^2_{per}([0,1])$ and covariance function $R : [0,1] \to \mathbb{R}$. Suppose that Assumptions 4.1 and 4.2 hold with $\rho < 1/16$. Then, for any $\lambda \geq 1$ and $x > 0$*

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

with $A_1(x, J, n, \lambda, \sigma^2, \gamma) = (\sigma^2 + \gamma)\big(\sqrt{\upsilon(x, J, n, \lambda)} + \upsilon(x, J, n, \lambda)\big) + \big(\sqrt{B(\lambda, n)} + B(\lambda, n)\big)$ and $A_2(x, J) = \big(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\big)^2$, where $C_1(\Theta, \mathcal{F}, f) > 0$ is constant depending only on $\Theta, \mathcal{F}, f$, $\upsilon(x, J, n, \lambda) = \frac{2\lambda+1}{n}\left(1 + 4\frac{x}{J} + \sqrt{4\frac{x}{J}}\right)$, $B(\lambda, n) = \frac{2\lambda+1}{n} + \lambda^{-2s}$. and $\gamma = \int_0^1 |R(t)|\, dt$.

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

Thus, Theorem 5.1 is consistent with the conclusions of Theorem 3.3, that is, if n is fixed, then it is not possible to estimate $\boldsymbol{\theta}^*$ by letting only J grows to infinity. Hence, under the assumptions of Theorem 5.1, one can only prove the convergence in probability of $\hat{\boldsymbol{\theta}}^\lambda$ to the true shifts $\boldsymbol{\theta}^*$ by taking the double asymptotic $n \to +\infty$ and $J \to +\infty$, provided the smoothing parameter $\lambda = \lambda_n$ is well chosen.

5.4. Consistent estimation of the mean pattern

In the case of randomly shifted curves, the Fréchet mean estimator (2.1) of f is $\hat{f}^\lambda(t) = \frac{1}{J}\sum_{j=1}^J \hat{f}_j^\lambda(t + \hat{\boldsymbol{\theta}}_j^\lambda)$.

Theorem 5.2. *Under the assumptions of Theorem 5.1, for any* $\lambda \geq 1$ *and* $x > 0$

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

where $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ *and* $A_2(x, J)$ *are defined in Theorem 5.1,* $C_2(\Theta, \mathcal{F}, f)$ *and* $C_3(\Theta, f)$ *are positive constants depending only on* $\Theta, \mathcal{F}, f$*, and* $\|\hat{f}^{\lambda} - f\|_{L^2}^2 = \int_0^1 \left| f^{\lambda}(t) - f(t) \right|^2 dt$.

Similar comments to those made on the consistency of the estimators of the shifts can be made. A double asymptotic in n and J is needed to show that the Fréchet mean $\hat{f}^{\lambda}$ converges in probability to the true mean pattern f. Moreover, if λ_n is too large (e.g. such that $\lim_{n \to +\infty} \frac{\lambda_n}{n} \neq 0$, which correspond to under-smoothing), then Theorem 5.2 cannot be used to prove that $\hat{f}^{\lambda}$ converges to f in probability. This illustrates the fact that, to achieve consistency, a sufficient amount of pre-smoothing is necessary before computing the Fréchet mean (2.1).

5.5. A lower bound for the Fréchet mean

From the results of Theorem 3.3, it is expected that the Fréchet mean $\hat{f}^{\lambda}$ does not converge to f in the setting n fixed and $J \to +\infty$. To support this argument, consider the following ideal estimator

$$\tilde{f}(t) = \frac{1}{J} \sum_{j=1}^{J} f_j(t + \hat{\boldsymbol{\theta}}_j^{\lambda}) = \frac{1}{J} \sum_{j=1}^{J} f(t - \boldsymbol{\theta}_j^* + \hat{\boldsymbol{\theta}}_j^{\lambda}), \quad \text{for all } t \in [0, 1], \tag{5.6}$$

where $f_j(t) = f(t - \boldsymbol{\theta}_j^*), j = 1, \ldots, J$. This corresponds to the case of an ideal smoothing step from the data (3.3) that would yield $\hat{f}_j = f_j$ for all $j = 1, \ldots, J$. Obviously, $\tilde{f}(t)$ is not an estimator since it depends on the unobserved quantities f and $\boldsymbol{\theta}_j^*$, but we can consider it as a benchmark to analyse the converge of the Fréchet mean $\hat{f}^{\lambda}$ to f.

Theorem 5.3. *Suppose that the assumptions of Theorem 3.3 are satisfied with* $\rho < \frac{3}{4\pi}$*. Then, for any* $n \geq 1$*, there exists* $J_0 \in \mathbb{N}$ *such that* $J \geq J_0$ *implies*

$$\mathbb{E}[\|\tilde{f} - f\|_{L^2}] \geq C(f, \rho) \frac{n^{-1}\sigma^2}{\|\partial_t f\|_{\infty}^2 + n^{-1}\sigma^2 \int_{\Theta} \left(\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta}) \right) \right)^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta}}, \tag{5.7}$$

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

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

6. Notations and main assumptions in the general case

6.1. Smoothness of the mean pattern and the deformation operators

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

Assumption 6.1. *For all $\boldsymbol{\theta}\in\mathcal{P}$, $\mathcal{L}_{\boldsymbol{\theta}} : L^2(\Omega) \longrightarrow L^2(\Omega)$ is a linear operator satisfying $\mathcal{L}_{\boldsymbol{\theta}} f \in \mathcal{F}$ for all $f\in\mathcal{F}$. There exists a constant $C(\Theta) > 0$ such that for any $f\in L^2(\Omega)$ and $\boldsymbol{\theta}\in\Theta$*

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

and a constant $C(\mathcal{F},\Theta) > 0$ such that for any $f\in\mathcal{F}$ and $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta$,

$$\|\tilde{T}_{\boldsymbol{\theta}_1} f - \tilde{T}_{\boldsymbol{\theta}_2} f\|_{L^2}^2 \leq C(\mathcal{F},\Theta)\,\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|^2\,.$$

Assumption 6.1 can be interpreted as a Lipschitz condition on the mapping $(f,\boldsymbol{\theta})\longmapsto\mathcal{L}_{\boldsymbol{\theta}} f$. The first inequality, that is $\|\mathcal{L}_{\boldsymbol{\theta}} f\|_{L^2}^2 \leq C(\Theta)\,\|f\|_{L^2}^2$, means that the action of the operator $\mathcal{L}_{\boldsymbol{\theta}}$ does not change too much the norm of f when $\boldsymbol{\theta}$ varies in Θ. Such an assumption on $T_{\boldsymbol{\theta}}$ and its inverse $\tilde{T}_{\boldsymbol{\theta}}$ forces the optimization problem (2.2) to have non trivial solutions by avoiding the functional $M(\boldsymbol{\theta})$ in (2.3) being arbitrarily small. It can be easily checked that Assumption 6.1 is satisfied in the case (1.2) of shifted curves with $\mathcal{F} = H_s(A)$ and $s\geq 1$.

6.2. The preliminary smoothing step

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

Definition 6.1. *Let Λ denote either $\mathbb{N}$ or $\mathbb{R}_+$ (set of smoothing parameters). To every $\lambda\in\Lambda$ is associated a non-random vector valued function $S_\lambda : \Omega\longrightarrow\mathbb{R}^n$ such that for all $j=1,\ldots,J$ and all $t\in\Omega$*

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

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

Assumption 6.2. *For all $\lambda\in\Lambda$ and all $\ell=1,\ldots,n$, the function $t\longmapsto S_\lambda^\ell(t)$ belong to $L^2(\Omega)$, where $S_\lambda^\ell(t)$ denotes the ℓ-th component of the vector $S_\lambda(t)$. Moreover, for all $\lambda\in\Lambda$, $f\in\mathcal{F}$ and $\boldsymbol{\theta}\in\Theta$, the function $t\longmapsto\langle S_\lambda(t), \mathbf{T}_{\boldsymbol{\theta}}\mathbf{f}\rangle$ belongs to $\mathcal{F}$ where $\mathbf{T}_{\boldsymbol{\theta}}\mathbf{f} = \left(T_{\boldsymbol{\theta}} f(t_\ell)\right)_{\ell=1}^n$.*

In the case (1.2) of randomly shifted curves with an equi-spaced design, then Assumption 6.2 holds with $S_\lambda(t) = \left[\frac{1}{n}\sum_{|k|\leq\lambda} e^{i2\pi k(t-\frac{\ell}{n})}\right]_{\ell=1}^n$. Let us now specify how the bias/variance behavior of the linear estimators $\hat{f}_j^\lambda$ depends on the smoothing parameter λ. For this, consider for some function $f\in\mathcal{F}$ the following regression model

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

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

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

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

Assumption 6.3. *There exist a constant $\kappa(\mathcal{F}) > 0$ and a real-valued function $\lambda \longmapsto B(\lambda)$, such that for all $f \in \mathcal{F}$,*

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

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

Let us illustrate Assumption 6.3 in the case of shifted curves with an equi-spaced design, and a smoothing step obtained by low-pass Fourier filtering. As in Section 5, take $\mathcal{F} = H_s(A)$ defined in (5.4). In this setting, $V(\lambda) = \frac{2\lambda+1}{n}$. It can be also checked that $\|B_\lambda(f,\cdot)\|_{L^2}^2 \leq C(A)B(\lambda)$ for some positive constant $C(A)$ depending only on A, and $B(\lambda) = \frac{2\lambda+1}{n} + \lambda^{-2s}$. Thus, Assumption 6.3 holds with $\lambda_n = n^{\frac{1}{2s+1}}$.

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

Assumption 6.4. *For any $n \geq 1$, there exists a real $\gamma_n(\Theta) > 0$ such that for any $\boldsymbol{\theta} \in \Theta$*

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

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

$$\lim_{n\to\infty} \gamma_n(\Theta)\sqrt{V(\lambda_n)} = 0, \tag{6.1}$$

where $V(\lambda_n)$ is the variance defined in Assumption 6.3.

Intuitively, the condition (6.1) means that the variance of the linear smoother $S_\lambda(\cdot)$ has to be asymptotically smaller that the maximal correlations (measured by $\gamma_n(\Theta)$) between $T_{\boldsymbol{\theta}} Z(t_\ell)$ and $T_{\boldsymbol{\theta}} Z(t_{\ell'})$ for $\ell, \ell' = 1, \ldots, n$ and all $\boldsymbol{\theta} \in \Theta$. In the case of randomly shifted curves with an equi-spaced design, a simple condition for which Assumption 6.4 holds is the case where Z is stationary process (see the arguments in Section 5.1).

7. Consistency in the general case

7.1. Consistent estimation of the deformation parameters

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

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

where

$$M_{\lambda}(\boldsymbol{\theta}) = \frac{1}{J} \sum_{j=1}^{J} \int_{\Omega} \left(\tilde{T}_{\boldsymbol{\theta}_j} \left\langle S_{\lambda}(t), \mathbf{Y}_j \right\rangle - \frac{1}{J} \sum_{j'=1}^{J} \tilde{T}_{\boldsymbol{\theta}_{j'}} \left\langle S_{\lambda}(t), \mathbf{Y}_{j'} \right\rangle \right)^2 dt, \tag{7.1}$$

and $\boldsymbol{\Theta}$ is the constrained set introduced in Assumption 4.3. The estimator $\hat{\boldsymbol{\theta}}^{\lambda}$ thus depends on the choice of $\boldsymbol{\Theta}$, and it will be shown that $\hat{\boldsymbol{\theta}}^{\lambda}$ is a consistent estimator of the vector $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}} \in \mathbb{R}^{pJ}$ defined in Assumption 4.3. Note that depending on the problem at hand and the choice of the constrained set $\boldsymbol{\Theta}$, it can be shown that $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$ is close to the true deformation parameters $\boldsymbol{\theta}^*$. For example, in the case of shifted curves, if $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ defined in (4.2) and if the density g of the shifts has zero mean, then $\boldsymbol{\theta}_{\boldsymbol{\Theta}_0} = (\boldsymbol{\theta}^*_1 - \bar{\boldsymbol{\theta}}^*, \ldots, \boldsymbol{\theta}^*_J - \bar{\boldsymbol{\theta}}^*)$ with $\bar{\boldsymbol{\theta}}^* = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\theta}^*_j$ can be shown to be close to $\boldsymbol{\theta}^*$ (see Lemma C.1 in the Appendix). This allows to show the consistency of $\hat{\boldsymbol{\theta}}^{\lambda}$ to $\boldsymbol{\theta}^*$ as formulated in Theorem 5.1. Therefore, the next result only bounds the distance between $\hat{\boldsymbol{\theta}}^{\lambda}$ and $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$.

Theorem 7.1. *Consider the model (1.5) and suppose that Assumptions 1.1, 4.3, 4.4 and 6.1 to 6.4 hold with $n \geq 1$ and $J \geq 2$. Then, for any $\lambda \in \Lambda$ and $x > 0$*

$$\mathbb{P}\Bigg(\frac{1}{J} \|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\|^2_{\mathbb{R}^{pJ}} \geq C_1(\Theta, \boldsymbol{\Theta}, \mathcal{F}, f) \Big[(\gamma_n(\Theta) + \sigma^2) \Big(\sqrt{\upsilon(x, J, \lambda)} + \upsilon(x, J, \lambda) \Big) + \Big(\sqrt{B(\lambda)} + B(\lambda) \Big) \Big] \Bigg) \leq 2e^{-x}, \tag{7.2}$$

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

Using Assumptions 6.3 and 6.4, it follows that $\lim_{n \to +\infty} \gamma_n(\Theta)\big(\sqrt{\upsilon(x, J, \lambda_n)} + \upsilon(x, J, \lambda_n)\big) = 0$ for any $x > 0$ and $J \geq 2$. If J remains fixed, Theorem 7.1 thus implies that $\hat{\boldsymbol{\theta}}^{\lambda}$ converges in probability to $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$ as $n \to +\infty$. To the contrary, let us fix n, and consider an asymptotic setting where only $J \to +\infty$. For any $x > 0$ and $\lambda \in \Lambda$, $\lim_{J \to +\infty} \upsilon(x, J, \lambda) = V(\lambda)$. Therefore, Theorem 7.1 cannot be used to prove that $\hat{\boldsymbol{\theta}}^{\lambda}$ converges to $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$ as $J \to +\infty$. This confirms that $\hat{\boldsymbol{\theta}}^{\lambda}$ is not a consistent estimator of $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$ (and thus of $\boldsymbol{\theta}^*$) as n remains fixed and J tends to infinity.

7.2. Consistent estimation of the mean pattern

Recall that the estimator $\hat{f}^\lambda$ of the mean pattern f is defined as

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

We study the consistency of $\hat{f}^\lambda$ with respect to the shape function

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

defined for $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}} = ([\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}]_1, \ldots, [\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}]_J)$. Again, depending on the problem at hand and the choice of the constrained set $\boldsymbol{\Theta}$, it can be shown that $f^*_{\boldsymbol{\Theta}}$ is close to the true mean pattern f. For example, in the case of shifted curves with $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ defined in (4.2), then $\boldsymbol{\theta}_{\boldsymbol{\Theta}_0} = (\boldsymbol{\theta}^*_1 - \bar{\boldsymbol{\theta}}^*, \ldots, \boldsymbol{\theta}^*_J - \bar{\boldsymbol{\theta}}^*)$ with $\bar{\boldsymbol{\theta}}^* = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\theta}^*_j$. In this case $f^*_{\boldsymbol{\Theta}_0}(t) := \frac{1}{J}\sum_{j=1}^{J} f(t - \boldsymbol{\theta}^*_j + [\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}]_j) = f(t - \bar{\boldsymbol{\theta}}^*)$. Hence, under the condition that $\int_\Theta \boldsymbol{\theta} g(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0$, then $\bar{\boldsymbol{\theta}}^* \approx 0$ for J sufficiently large, and thus $f^*_{\boldsymbol{\Theta}}(t)$ is close to f which allows to show the consistency of $\hat{f}^\lambda$ to f as formulated in Theorem 5.2.

Theorem 7.2. *Consider the model* (1.5) *and suppose that Assumptions 1.1, 4.3, 4.4 and 6.1 to 6.4 hold. Then, for any* $\lambda \in \Lambda$ *and* $x > 0$

$$\mathbb{P}\Big(\|\hat{f}^\lambda - f^*_{\boldsymbol{\Theta}}\|^2_{L^2} \geq C_2(\Theta, \boldsymbol{\Theta}, \mathcal{F}, f)\Big[(\gamma_n(\Theta) + \sigma^2)\Big(\sqrt{\upsilon(x, J, \lambda)} + \upsilon(x, J, \lambda)\Big) + \Big(\sqrt{B(\lambda)} + B(\lambda)\Big)\Big]\Big) \leq 2e^{-x}, \tag{7.3}$$

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

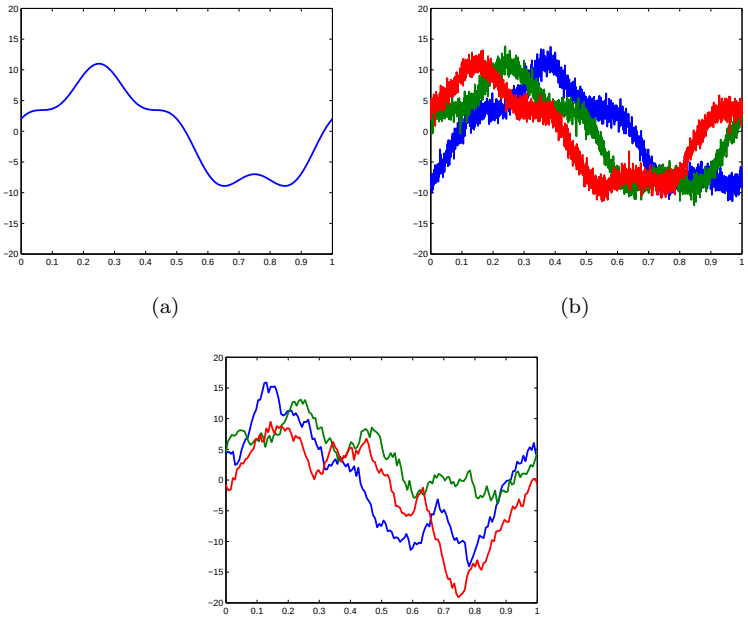
The consistency of $\hat{f}^\lambda$ to $f^*_{\boldsymbol{\Theta}}$ is thus guaranteed when n goes to infinity provided the level of smoothing $\lambda = \lambda_n$ is chosen so that $\lim_{n\to+\infty} V(\lambda_n) = \lim_{n\to+\infty} B(\lambda_n) = 0$. Again, if n remains fixed and only J is let going to infinity then Theorem 7.2 cannot be used to prove the convergence of $\hat{f}^\lambda$ to $f^*_{\boldsymbol{\Theta}}$.

8. Numerical experiments for randomly shifted curves

Consider the model (3.3) with random shifts $\boldsymbol{\theta}_j$ having a uniform density g with compact support equal to $[-\frac{1}{5}, \frac{1}{5}]$, and $f(t) = 9\sin(2\pi t) + 2\cos(8\pi t)$ for $t \in [0, 1]$ as a mean pattern, see Figure 2(a). For the constrained set we took

$$\boldsymbol{\Theta}_0 = \Big\{ \boldsymbol{\theta} \in \big[-\tfrac{1}{2}, \tfrac{1}{2}\big]^J,\ \boldsymbol{\theta}_1 + \cdots + \boldsymbol{\theta}_J = 0 \Big\}.$$

We use Fourier low pass filtering with spectral cut-off to $\lambda = 7$ which is reasonable value to reconstruct f representing a good tradeoff between bias and



(a) (b)

(c)

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

variance. We present some results of simulations under various assumptions of the process Z and the level σ of additive noise in the measurements.

Shape invariant model (SIM). The first numerical applications illustrate the role of n and J in the SIM model. Figure 2(b) gives a sample of the data used with $\sigma = 2$. The factors in the simulations are the number J of curves and the number of design points n. For each combination of these two factors, we simulate $M = 20$ repetitions of model (3.3). For each repetition we computed $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*\|^2$ and $\|\hat{f}^{\lambda} - f\|_{L^2}^2$. Boxplot of these quantities are displayed in Figure 3(a) and 3(b) respectively, for $J = 20, 40, \ldots, 100$ and $n = 512$ (in gray) and $n = 1024$ (in black). As the smoothing parameter is fixed to $\lambda = 7$, increasing n simply reduces the variance of the linear smoothers $\hat{f}_j^{\lambda}$. Recall that the lower bound given in Theorem 3.3 shows that $\frac{1}{J}\mathbb{E}[\|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}^{\lambda}\|^2]$ does not decrease as J increases but should be smaller when the number of point n increases. This is exactly what we observe in Figure 3. Similarly, the quantity $\|\hat{f}^{\lambda} - f\|_{L^2}^2$ is clearly smaller with $n = 1024$ than with $n = 512$.

Complete model. We now add the terms Z_j in (3.3) to model linear variations in amplitude of the curves around the template f. First, we generate a stationary periodic Gaussian process. To do this, the covariance matrix must be a particular

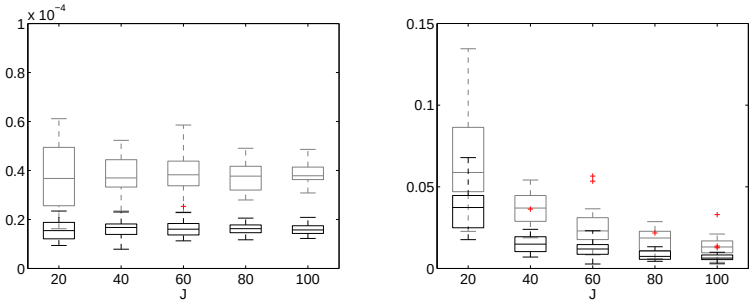


FIGURE 3. *Boxplot of* $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ *(a) and* $\|\hat{f}^{\lambda} - f_{\boldsymbol{\Theta}_0}\|^2_{L^2}$ *(b) over* $M = 20$ *repetitions from a SIM model of shifted curves. Boxplot in gray correspond to* $n = 512$*, and in black to* $n = 1024$.

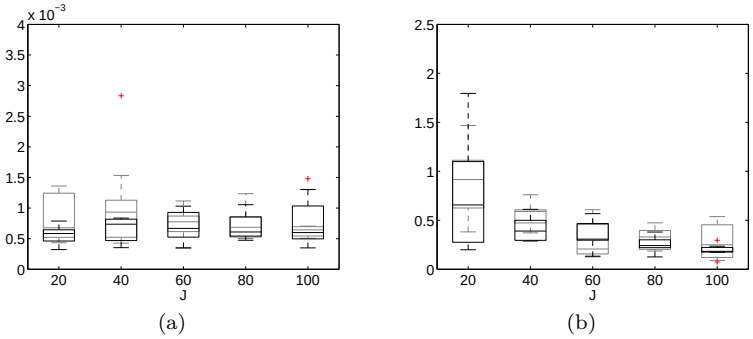


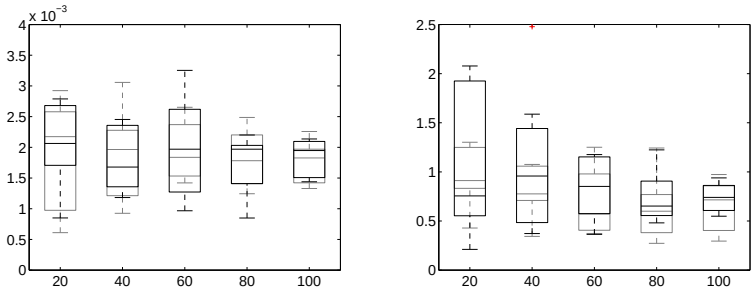
FIGURE 4. *Boxplot of* $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ *(a) and* $\frac{1}{J}\|\hat{f}^{\lambda} - f_{\boldsymbol{\Theta}_0}\|^2$ *(b) in model* (3.3) *with a stationnary error term* Z*. Boxplot in gray correspond to* $n = 512$*, and in black to* $n = 1024$.

Toeplitz matrix. As suggested in [Gre93] one possibility is to choose

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

where ϕ is a strictly positive parameter (we took $\phi = 4$) and ς a variance parameter. The level of additive noise is $\sigma = 8$, and we took $\varsigma = 4$. As an illustration, in Figure 2(c) we plot $f + Z_j$, $j = 1, 2, 3$ with $\varsigma = \phi = 4$. Over $M = 20$ repetitions, we have computed the values of $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ and $\|\hat{f}^{\lambda} - f_{\boldsymbol{\Theta}_0}\|^2_{L^2}$ for J is varying from 20 to 100 and $n = 512, 1024$. The results are displayed in Figure 4(a) and 4(b). We observe the same behaviors than in the simulations with the SIM model: the variance of $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ does not decrease as J increases (see Figure 4(a)) and $\|\hat{f}^{\lambda} - f_{\boldsymbol{\Theta}_0}\|^2$ has a smaller mean and variance as n increases.

We finally run the same simulations with a non stationary noise $Z_j(t) = \alpha_j \psi(t)$ where ψ is a positive periodic smooth deterministic function such that



(a) (b)

FIGURE 5. *Boxplot of* $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ *(a) and* $\frac{1}{J}\|\hat{f}^{\lambda} - f_{\boldsymbol{\Theta}_0}\|^2$ *(b) in model* (3.3) *with a non-stationnary error term* Z*. Boxplot in gray correspond to* $n = 512$*, and in black to* $n = 1024$*.*

$\|\psi\|_{L^2} = 1$ and $\alpha_j \sim \mathcal{N}(0, \varsigma^2)$ with $\varsigma = 4$. Note that, in this case, the sequence $\gamma_n(\Theta)$ is of order n and Assumption 6.4 is not verified. The levels of noise (σ and ς) are the same than in the stationary case in order to make things comparable. The results are presented in the same manner in Figure 5(a) for $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ and in Figure 5(b) for $\|\hat{f}^{\lambda} - f_{\boldsymbol{\Theta}_0}\|^2_{L^2}$. One can see that the results are very different. The estimators of the shifts have a much larger mean and variance, and the variance of$\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ remains rather high even when n or J increases (see Figure 5(a)). The convergence to zero of $\|\hat{f}^{\lambda} - f_{\boldsymbol{\Theta}_0}\|^2_{L^2}$ which was clear in the stationary case, is now not so obvious in view of the numerical results displayed in Figure 5(b).

9. Conclusion and perspectives

We have proposed to use a Fréchet mean of smoothed data to estimate a mean pattern of curves or images satisfying a non-parametric regression model including random deformations. Upper and lower bounds (in probability and expectation) for the estimation of the deformation parameters and the mean pattern have been derived. Our main result is that these bounds go to zero as the dimension n of the data (the number of sample points) goes to infinity, but that an asymptotic setting only in J (the number of observed curves or images) is not sufficient to obtain consistent estimators. An interesting topic for future investigation would be to study the rate of convergence of such estimators and to analyze their optimality (e.g. from a minimax point of view).

Appendix A: Proof of the results in Section 3

A.1. Proof of Theorem 3.1

Write $\boldsymbol{\theta}^*_j = ([\theta^*]^1_j, \ldots, [\theta^*]^p_j)$, and let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_J) \in \mathbb{R}^{nJ}$ be the column vector of the observations generated by model (1.7). Conditionally to $\boldsymbol{\theta}^*$, $\mathbf{Y}$ is

a Gaussian vector and its log-likelihood is equal to

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

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

$$\mathbb{E}_{\boldsymbol{\theta}^*}\left[\partial_{[\theta^*]_{j_1}^{p_1}}\log(p(\mathbf{Y}|\boldsymbol{\theta}^*))\partial_{[\theta^*]_{j_2}^{p_2}}\log(p(\mathbf{Y}|\boldsymbol{\theta}^*))\right] = \\ \begin{cases} 0 & \text{if } j_1 \neq j_2, \\ -\left[(\partial_{[\theta^*]_{j_1}^{p_1}}\mathbf{T}_{\boldsymbol{\theta}_{j_1}^*}\mathbf{f})'\ \Lambda\ (\partial_{[\theta^*]_{j_1}^{p_2}}\mathbf{T}_{\boldsymbol{\theta}_{j_1}^*}\mathbf{f})\right]_{p_1,p_2=1}^{p} & \text{if } j_1 = j_2, \end{cases} \tag{A.2}$$

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

$$(\partial_{[\theta^*]_{j_1}^{p_1}}\mathbf{T}_{\boldsymbol{\theta}_{j_1}^*}\mathbf{f})'\ \Lambda\ (\partial_{[\theta^*]_{j_1}^{p_1}}\mathbf{T}_{\boldsymbol{\theta}_{j_1}^*}\mathbf{f}) \leq \sigma^{-2}\|\partial_{[\theta^*]_{j_1}^{p_1}}\mathbf{T}_{\boldsymbol{\theta}_{j_1}^*}\mathbf{f}\|^2 \leq C(\Theta,f)n\sigma^{-2}, \tag{A.3}$$

where the last inequality is a consequence of Assumption 3.1. From now on, $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{Y}) = (\hat{\boldsymbol{\theta}}_1(\mathbf{Y}),\dots,\hat{\boldsymbol{\theta}}_1(\mathbf{Y}))$ is an arbitrary estimator (i.e any measurable function of $\mathbf{Y}$) of the true parameter $\boldsymbol{\theta}^*$. Let also $U = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$ and

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

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

$$(\mathbb{E}[U'V])^2 \leq \mathbb{E}[U'U]\mathbb{E}[V'V]. \tag{A.4}$$

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

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

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

Now, using that the expected score is zero and equation (A.2) we have

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

where $\partial_{\boldsymbol{\theta}_1}\log(g(\boldsymbol{\theta}_1)) = [\partial_{[\boldsymbol{\theta}]_1^1}\log(g(\boldsymbol{\theta}_1)),\ldots,\partial_{[\boldsymbol{\theta}]_1^p}\log(g(\boldsymbol{\theta}_1))] \in \mathbb{R}^p$. Then, using inequality A.3, it gives

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

Hence, using equation (A.4) for any estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{Y})$ (see Theorem 1 in [GL95])

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

And since $p \geq 1$, the claim in Theorem 3.1 is proved. □

A.2. Proof of Theorem 3.2

As above, let $\mathbf{Y} \in \mathbb{R}^{nJ}$ is the column vector generated by model (1.5). Then, conditionally to $\boldsymbol{\theta}^*$, $\mathbf{Y}$ is a Gaussian vectors and Assumption 3.2 ensures that its log-likelihood has the same expression as in equation (A.1) but with

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

As the matrix $\boldsymbol{\Sigma}_n(\Theta)$ is positive semi definite with it smallest eigenvalue denoted by $s_n^2(\Theta)$ (see Assumption 3.2), the uniform bound (A.3) becomes

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

for all $p_1 = 1,\ldots,p$ and $j = 1,\ldots,J$. As above the last inequality is a consequence of Assumption 3.1 and the rest of the proof is identical to the proof of Theorem 3.1. □

A.3. Proof of Theorem 3.3

For all $\boldsymbol{\theta} \in \mathbb{R}$ the operators $T_{\boldsymbol{\theta}} f(\cdot) = f(\cdot - \boldsymbol{\theta})$ are isometric from $L^2([0,1])$ to $L^2([0,1])$ as a change of variable implies immediately that $\|T_{\boldsymbol{\theta}} f\|_{L^2}^2 = \|f\|_{L^2}^2$. For all continuously differentiable function f, we have $\partial_\theta T_{\boldsymbol{\theta}} f(t) = -sign(\boldsymbol{\theta}) \partial_t f(t - \boldsymbol{\theta})$, where $sign(\cdot)$ is the sign function. Then, for all $\boldsymbol{\theta} \in \Theta$, $\|\partial_\theta T_{\boldsymbol{\theta}} f\|_{L^2}^2 = \|\partial_t f\|_{L^2}^2 \leq \|\partial_t f\|_\infty^2$ and Assumption 3.1 is satisfied with $C(\Theta, f) = \|\partial_t f\|_\infty^2$. Finally, as the error terms Z_j's are i.i.d stationary random process the covariance function is invariant by the action of the shifts and Assumption 3.2 is satisfied with $\Sigma_n(\Theta) = \boldsymbol{\Sigma}_n$ defined in (5.1) (see Section 5.1 for further details). Then, the result of Theorem 3.3 follows as an application of Theorem 3.2. □

Appendix B: Proof of the results in Section 4

B.1. Proof of Proposition 4.1

Remark that $D(\boldsymbol{\theta}) = \sum_{k \in \mathbb{Z}} |c_k^*|^2 \left(1 - \left| \frac{1}{J} \sum_{j=1}^J e^{i2\pi k(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)} \right|^2 \right)$, where $c_k^* = \int_0^1 f(t) e^{-i2\pi k t} dt$. Thanks to Assumption 4.1, it follows that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

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

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

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

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

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

since $|(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*)| \leq 4\rho < 1/4$ for all $m, q = 1, \ldots, n$ by Assumption 4.2 and the hypothesis that $\rho < 1/16$. Hence, using the lower bound (B.1), it follows that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

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

with $C(f,\rho) = 2|c_1^*|^2 C(\rho)$. Now assume that $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$. Using the properties that $\sum_{j=1}^J \boldsymbol{\theta}_j = 0$ and $\sum_{j=1}^J (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) = -\sum_{j=1}^J \boldsymbol{\theta}_j^* = J\bar{\boldsymbol{\theta}}^*$, it follows from elementary algebra that $\frac{1}{J}\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J} \left|(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*)\right|^2 = \sum_{j=1}^J (\boldsymbol{\theta}_j - (\boldsymbol{\theta}_j^* - \bar{\boldsymbol{\theta}}^*))^2$. This equality together with the lower bound (B.2) completes the proof. □

Appendix C: Proof of the results in Section 5

C.1. Proof of Theorem 5.1

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

Lemma C.1. *Suppose that Assumption 4.2 holds. Then, for any $x > 0$*

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

Using the inequality $\frac{1}{J}\|\hat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*\|^2 \leq \frac{2}{J}\|\hat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2 + \frac{2}{J}\|\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} - \boldsymbol{\theta}^*\|^2$, it follows that Theorem 5.1 is a consequence of Lemma C.1 and Theorem 7.1. Indeed, it can be easily checked that, under the assumptions of Theorem 5.1, Assumptions 6.1 to 6.4 are satisfied in the case of randomly shifted curves with an equi-spaced design and low-pass Fourier filtering, see the various arguments given in Section 6). The identifiability condition of Assumption 4.4 is given by Proposition 4.1. □

C.2. Proof of Theorem 5.2

Consider the following inequality $\|\hat{f}^\lambda - f\|^2 \leq 2\|\hat{f}^\lambda - f_{\boldsymbol{\Theta}_0}\|^2 + 2\|f_{\boldsymbol{\Theta}_0} - f\|^2$, where $f_{\boldsymbol{\Theta}_0}(t) = f(t - \bar{\boldsymbol{\theta}}^*)$ and $\bar{\boldsymbol{\theta}}^* = \frac{1}{J}\sum_{j=1}^J \boldsymbol{\theta}_j^* \in \Theta$. As f is assumed to be in $H_s(A)$, there exists a constant $C(\Theta, f) > 0$ such that $\|f_{\boldsymbol{\Theta}_0} - f\|^2_{L^2} \leq C(\Theta, f)|\bar{\boldsymbol{\theta}}^*|^2 = C(\Theta, f)\frac{1}{J}\|\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} - \boldsymbol{\theta}^*\|^2$. As explained in part C.1 the assumptions of Theorem 5.2 are satisfied in the case of randomly shifted curves with an equi-spaced design and low-pass Fourier filtering. The result then follows from Theorem 7.2. □

C.3. Proof of Theorem 5.3

Let $n \geq 1$. We have that

$$\mathbb{E}[\|\tilde{f} - f\|_{L^2}] = \mathbb{E}\|\tilde{f} - f_{\boldsymbol{\Theta}_0} + f_{\boldsymbol{\Theta}_0} - f\|_{L^2} \geq \Big|\underbrace{\mathbb{E}\|\tilde{f} - f_{\boldsymbol{\Theta}_0}\|_{L^2}}_{\mathbf{A}} - \underbrace{\mathbb{E}\|f_{\boldsymbol{\Theta}_0} - f\|_{L^2}}_{\mathbf{B}}\Big| \tag{C.1}$$

where for all $t \in [0,1]$, $\tilde{f}(t) = \frac{1}{J}\sum_{j=1}^J f(t - \boldsymbol{\theta}_j^* + \hat{\boldsymbol{\theta}}_j^\lambda)$, and $f_{\boldsymbol{\Theta}_0}(t) = f(t + \bar{\boldsymbol{\theta}}^*)$, with $\bar{\boldsymbol{\theta}}^* = \frac{1}{J}\sum_{j=1}^J \boldsymbol{\theta}_j^*$. In the rest of the proof, we show that $\mathbf{A}$ is bounded

from below by a quantity $C_0(f,g,n,\sigma^2,\rho) = C(f,\rho)\frac{n^{-1}\sigma^2}{\|\partial_t f\|_\infty^2 + n^{-1}\sigma^2 \int_\Theta (\partial_{\boldsymbol{\theta}} \log(g(\boldsymbol{\theta})))^2}$ independent of J (this statement is made precise later) and that $\mathbf{B}$ goes to zero as J goes to infinity. Then, these two facts imply that there exists a $J_0 \in \mathbb{N}$ such that $J \geq J_0$ implies that $\mathbb{E}\|\tilde{f} - \tilde{f}\|_{L^2} \geq \frac{1}{2} C_0(f,g,n,\sigma^2,\rho)$, which will yield the desired result.

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

$$\begin{aligned}
\|\tilde{f} - f_{\boldsymbol{\Theta}_0}\|_{L^2} &= \|\frac{1}{J}\sum_{j=1}^{J} f(\cdot - \boldsymbol{\theta}_j^* + \hat{\boldsymbol{\theta}}_j^\lambda) - f(\cdot + \bar{\boldsymbol{\theta}}^*)\|_{L^2} \\
&= \left(\sum_{k\in\mathbb{Z}} \left| \frac{1}{J} \sum_{j=1}^{J} \left(e^{i2\pi k(-\boldsymbol{\theta}_j^* + \hat{\boldsymbol{\theta}}_j^\lambda)} - e^{i2\pi k \bar{\boldsymbol{\theta}}^*} \right) c_k^* \right|^2 \right)^{\frac{1}{2}} \\
&\geq |c_1^*| \left| \frac{1}{J} \sum_{j=1}^{J} (e^{i2\pi(\hat{\boldsymbol{\theta}}_j^\lambda - [\boldsymbol{\theta}_{\boldsymbol{\Theta}_0}^*]_j)} - 1) \right|,
\end{aligned}$$

where $\boldsymbol{\theta}_{\boldsymbol{\Theta}_0}^* = (\boldsymbol{\theta}_1^* - \bar{\boldsymbol{\theta}}^*, \ldots, \boldsymbol{\theta}_J^* - \bar{\boldsymbol{\theta}}^*)$, the right hand side of the preceding inequality being positive since Assumption 4.2 ensures that $c_1^* \neq 0$ for all $j = 1, \ldots, J$. Let $u_j = 2\pi(\hat{\boldsymbol{\theta}}_j^\lambda - [\boldsymbol{\theta}_{\boldsymbol{\Theta}_0}^*]_j), j = 1, \ldots, J$. Since $\sum_{j=1}^J u_j = 0$ and $|u_j| \leq 4\pi\rho < 3,\ j = 1, \ldots, J$ (by our assumption on ρ), Lemma E.1 implies that

$$\|\tilde{f} - f_{\boldsymbol{\Theta}_0}\|_{L^2} \geq C(f,\rho) \frac{1}{J} \|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_{\boldsymbol{\Theta}_0}^*\|^2. \tag{C.2}$$

Now, remark that $\mathbb{E}\big[\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_{\boldsymbol{\Theta}_0}^*\|^2\big] \geq \mathbb{E}\big[\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*\|^2\big] - \mathbf{C}$ with

$$\mathbf{C} = 2\mathbb{E}\big[\,|\bar{\boldsymbol{\theta}}^*| \,\frac{1}{J}\sum_{j=1}^{J} |\hat{\boldsymbol{\theta}}_j^\lambda - \boldsymbol{\theta}_j^*|\big].$$

By applying Theorem 3.3 we get that

$$\mathbb{E}\big[\tfrac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*\|^2\big] \geq C(f,g,n,\sigma^2),$$

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

$$\mathbf{A} = \mathbb{E}\big[\|\tilde{f}^\lambda - \tilde{f}\|_{L^2}\big] \geq C_0(f,g,n,\sigma^2,\rho). \tag{C.3}$$

Upper bound on B. By assumption, f is continuously differentiable on $[0,1]$ implying that $\|f_{\boldsymbol{\Theta}_0} - f\|_{L^2} = \|f(\cdot + \bar{\boldsymbol{\theta}}^*) - f\|_{L^2} \leq \|\partial_t f\|_\infty |\bar{\boldsymbol{\theta}}^*|$. Therefore, $\mathbb{E}\|f_{\boldsymbol{\Theta}_0} - f\|_{L^2} \leq \|\partial_t f\|_\infty \sqrt{\mathbb{E}\,|\bar{\boldsymbol{\theta}}^*|^2} \leq C(f,g) J^{-1/2}$. Hence, there exists a $J_2 \in \mathbb{N}$

such that $J \geq J_2$ implies

$$\mathbf{B} = \mathbb{E}[\|\tilde{f}_{\boldsymbol{\Theta}_0} - \tilde{f}\|_{L^2}] \leq \frac{1}{2} C_0(f, g, n, \sigma^2, \rho). \tag{C.4}$$

To conclude the proof, equations (C.1), (C.3) and (C.4) imply that there exists a $J_0 \in \mathbb{N}$ such that $J \geq J_0$ implies $\mathbb{E}\|\hat{f}^{\lambda} - \tilde{f}\|_{L^2} \geq |\mathbf{A} - \mathbf{B}| \geq \frac{1}{2} C_0(f, g, n, \sigma^2, \rho)$. □

Appendix D: Proof of the results in Section 7

D.1. Proof of Theorem 7.1

We explain here the main arguments of the proof of Theorem 7.1. Technical Lemmas are given in the second part of the Appendix. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) = (\theta_1^1, \ldots, \theta_1^p, \ldots, \theta_J^1, \ldots, \theta_J^p) \in \mathbb{R}^{pJ}$ and decompose the criterion (7.1) as follows,

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

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

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

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

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

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

At this stage, recall that $\boldsymbol{\theta}_{\boldsymbol{\Theta}}^* = \operatorname{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} D(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}^\lambda = \operatorname{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} M_\lambda(\boldsymbol{\theta})$. The proof follows a classical guideline in M-estimation: we show that the uniform (over $\boldsymbol{\Theta}$) convergence in probability of the criterion M_λ to D, yielding the convergence in probability of their argmins $\boldsymbol{\theta}_{\boldsymbol{\Theta}}^*$ and $\hat{\boldsymbol{\theta}}^\lambda$ respectively. Assumption 4.4 ensures that there is a constant $C(\Theta, \mathcal{F}, f) > 0$ such that,

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

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

$$\left|D(\hat{\boldsymbol{\theta}}^\lambda) - D(\boldsymbol{\theta}_{\boldsymbol{\Theta}}^*)\right| \leq 2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|D(\boldsymbol{\theta}) - M_\lambda(\boldsymbol{\theta})| \tag{D.2}$$
$$= \underbrace{2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{R_\lambda(\boldsymbol{\theta}) + Q_\lambda(\boldsymbol{\theta})\right\}}_{\mathbf{B}} \tag{D.3}$$
$$+ \underbrace{2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{Q_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^{Z,\varepsilon}(\boldsymbol{\theta}) + Q_\lambda^\varepsilon(\boldsymbol{\theta}) + R_\lambda^\varepsilon(\boldsymbol{\theta})\right\}}_{\mathbf{V}}$$

The rest of the proof is devoted to control the **B** and **V** terms.

Control of B. Using Assumption 6.3 and 6.1, we have that

$$Q_\lambda(\boldsymbol{\theta}) \leq \frac{C(\Theta)}{J}\sum_{j=1}^{J}\left\|B_\lambda(T_{\boldsymbol{\theta}_j^*}f, t)\right\|_{L^2}^2 \leq C(\Theta, \mathcal{F})B(\lambda).$$

Now by applying the Cauchy-Schwarz inequality,

$$|R_\lambda(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\{\sqrt{D(\boldsymbol{\theta})}\}\sqrt{Q_\lambda(\boldsymbol{\theta})}.$$

By Assumption 6.1, there exists a constant such $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\{D(\boldsymbol{\theta})\} \leq C(\Theta,\mathcal{F},f)$ and thus

$$\mathbf{B} \leq C(\Theta,\mathcal{F},f)\big(B(\lambda)+\sqrt{B(\lambda)}\big). \tag{D.4}$$

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

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

where we have used the inequality $2ab\leq a^2+b^2$, valid for any $a,b>0$ to control the term $R_\lambda^{Z,\varepsilon}$. The quadratic terms Q_λ^Z and Q_λ^ε are controlled by Corollaries E.1 and E.2 respectively. It yields immediately to

$$\mathbb{P}\left(\mathbf{V}\geq C(\Theta,\mathcal{F},f)(\gamma_{\max}(n)+\sigma^2)\big(\upsilon(x,J,\lambda)+\sqrt{\upsilon(x,J,\lambda)}\big)\right)\leq 2e^{-x}, \tag{D.5}$$

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

Putting together equations (D.1), (D.2), (D.4) and (D.5), we have

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

which completes the proof of Theorem 7.1. □

D.2. Proof of Theorem 7.2

In this part, we use the notations introduced in the proof of Theorem 7.1. We have,

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

Again, the first term above depends on the bias, and the second term (stochastic) can be controlled in probability. Under Assumptions 6.1 and 6.3 we have that

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

and

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

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

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

Then, from Theorem 7.1 it follows that

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

which completes the proof. □

Appendix E: Technical Lemmas

Lemma E.1. *Let $u = (u_1,\ldots,u_J)$ such that $\sum_{j=1}^{J} u_j = 0$ with $|u_j| \leq \delta$ for some $0 \leq \delta < 3$ for all $j = 1,\ldots,J$. Then, there exists a constant $C(\delta) > 0$ such that $\left|\frac{1}{J}\sum_{j=1}^{J}(e^{iu_j}-1)\right| \geq \frac{C(\delta)}{J}\|u\|^2$ where $\|u\|^2 = u_1^2 + \cdots + u_J^2$.*

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

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

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

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

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

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

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

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

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

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

where $\mathbf{A}_j = A_j\mathbf{S}_\lambda A_j$ with $\mathbf{S}_\lambda = \left[\langle S_\lambda^\ell, S_\lambda^{\ell'}\rangle_{L^2}\right]_{\ell,\ell'=1}^n$. Now, denote by $r_{j,1} \geq \cdots \geq r_{j,n}$ the eigenvalues of $\mathbf{A}_j$ with $r_{j,1} \geq \cdots \geq r_{j,n} \geq 0$ and $r_1 = \max_{j,\ell}\{r_{j,\ell}\}$. We can write $\mathbf{A}_j = (\mathbf{S}_\lambda^{\frac{1}{2}}A_j)'(\mathbf{S}_\lambda^{\frac{1}{2}}A_j)$ and is positive semi-definite. Then, let $\tilde{\xi}_{\lambda,J} = J\xi_{\lambda,J} - J\mathbb{E}\xi_{\lambda,J} = \sum_{j=1}^J(\boldsymbol{\varepsilon}_j'\mathbf{A}_j\boldsymbol{\varepsilon}_j - \operatorname{tr}\mathbf{A}_j)$. Let $\alpha > 0$, by Markov's inequality it follows that for all $u \in \left(0, \frac{1}{2r_1}\right)$, $\mathbb{P}\left(\tilde{\xi}_{\lambda,J} \geq \alpha\right) = \mathbb{P}\left(e^{u\tilde{\xi}_{\lambda,J}} \geq e^{u\alpha}\right) \leq e^{-u\alpha}\prod_{j=1}^J \mathbb{E}\left[e^{u\boldsymbol{\varepsilon}_j'\mathbf{A}_j\boldsymbol{\varepsilon}_j - u\operatorname{tr}\mathbf{A}_j}\right]$, since the $\boldsymbol{\varepsilon}_j$'s are independent. The log-Laplace transform of $\tilde{\varphi}_{\lambda,j} = \boldsymbol{\varepsilon}_j'\mathbf{A}_j\boldsymbol{\varepsilon}_j - \operatorname{tr}\mathbf{A}_j$ is

$$\log\left(\mathbb{E}\left[e^{u\tilde{\varphi}_{\lambda,j}}\right]\right) = \sum_{\ell=1}^n -ur_{j,\ell} - \frac{1}{2}\log\left(1 - 2ur_{j,\ell}\right).$$

We now use the inequality $-x - \frac{1}{2}\log(1-2x) \leq \frac{x^2}{1-2x}$ for all $0 < x < \frac{1}{2}$ which holds since $u \in \left(0, \frac{1}{2r_1}\right)$. This implies that $\log\left(\mathbb{E}\left[e^{u\tilde{\varphi}_{\lambda,j}}\right]\right) \leq \sum_{\ell=1}^n \frac{u^2 {r_{j,\ell}}^2}{1-2ur_{j,\ell}} \leq \frac{u^2\|r_j\|^2}{1-2ur_1}$, where $\|r_j\|^2 = r_{j,1}^2 + \cdots + r_{n,j}^2$. Finally, we have

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

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

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

by setting $x = \frac{\alpha^2}{4r_1\alpha + 4\|r\|^2}$. We derive the following concentration inequality for $\xi_{\lambda,J} = \frac{1}{J}\tilde{\xi}_{\lambda,J} + \frac{1}{J}\sum_{j=1}^{J} \mathrm{tr}(\mathbf{A}_j)$, $\mathbb{P}\big(\xi_{\lambda,J} \geq \frac{1}{J}\sum_{j=1}^{J}\sum_{\ell=1}^{n} r_{j,\ell} + 4\frac{r_1}{J}x + \frac{\|r\|}{J}\sqrt{4x}\big) \leq e^{-x}$. To finish the proof, remark that

$$\|r\|^2 = \sum_{j=1}^{J}\sum_{\ell=1} r_{j,\ell}^2 \leq \left(\sum_{j=1}^{J}\sum_{\ell=1}^{n} r_{j,\ell}\right)^2$$

since all the $r_{j,\ell}$'s are positive. □

Corollary E.1. *Under Assumptions 6.1 to 6.3, there exists a positive constant $C(\Theta, \mathcal{F}) > 0$ such that for all $x > 0$,*

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

Proof. Assumption 6.1 gives the uniform bound

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

where $\xi_{\lambda,J}(\sigma Id_n, \ldots, \sigma Id_n)$ is defined in Lemma E.2 and does not depend on $\boldsymbol{\theta}$. Thus, the result immediately follows from Lemma E.2. □

Corollary E.2. *Under Assumptions 6.1 to 6.4, there exists a positive constant $C(\Theta, \mathcal{F}) > 0$ such that for all $x \geq 0$,*

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

Proof. Assumption 6.1 gives the uniform bound

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

Hence, conditionally on $\boldsymbol{\theta}^*$ we have that

$$\sup_{\boldsymbol{\theta}\in\Theta^J} Q_\lambda^Z(\boldsymbol{\theta}) \leq C(\Theta,\mathcal{F})\xi_{\lambda,J}\big(A_1,\ldots,A_J\big),$$

where $\xi_{\lambda,J}\big(A_1,\ldots,A_J\big)$ is defined in Lemma E.2 with

$$A_j = \mathbb{E}_{\boldsymbol{\theta}^*}\left[\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j(\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j)'\right]^{\frac{1}{2}}.$$

Let us now give an upper bound on the largest eigenvalues of the matrices $\mathbf{A}_j = A_j\mathbf{S}_\lambda A_j$ with $\mathbf{S}_\lambda = \left[\langle S_\lambda^\ell, S_\lambda^{\ell'}\rangle_{L^2}\right]_{\ell,\ell'=1}^n$. Under Assumption 6.4 we have that $\operatorname{tr}(\mathbf{A}_j) \leq \gamma_{\max}(A_j)\operatorname{tr}\mathbf{S}_\lambda \leq \gamma_n(\Theta)V(\lambda)$, for all $j=1,\ldots,J$ and any $\boldsymbol{\theta}^* \in \Theta^J$. Thus, the result follows by arguing as in the proof of Lemma E.2 and by taking expectation with respect to $\boldsymbol{\theta}^*$. □

Acknowledgements

We would like to thank Dominique Bakry for fruitful discussions. Both authors would like to thank the Center for Mathematical Modeling and the CNRS for financial support and excellent hospitality while visiting Santiago where part of this work was carried out. We are very much indebted to the referees and the Associate Editor for their constructive comments and remarks that resulted in a major revision of the original manuscript.

References

[AAT07] S. Allassonière, Y. Amit, and A. Trouvé. Toward a coherent statistical framework for dense deformable template estimation. *Journal of the Statistical Royal Society (B)*, 69:3–29, 2007. MR2301497

[Afs11] Bijan Afsari. Riemannian l^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011. MR2736346

[AKT09] S. Allassonière, E. Kuhn, and A. Trouvé. Bayesian deformable models building via stochastic approximation algorithm: a convergence study. *Bernoulli*, To be published, 2009.

[BG10] J. Bigot and S. Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Annals of statistics*, to be published, 2010. MR2676894

[BGL09] J. Bigot, S. Gadat, and J.-M. Loubes. Statistical M-estimation and consistency in large deformable models for image warping. *J. Math. Imaging Vision*, 34(3):270–290, 2009. MR2515449

[BGV09] J. Bigot, F. Gamboa, and M. Vimond. Estimation of translation, rotation and scaling between noisy images using the fourier mellin transform. *SIAM Journal on Imaging Sciences*, 2(2):614–645, 2009. MR2519925

[Big06] J. Bigot. Landmark-based registration of curves via the continuous wavelet transform. *Journal of Computational and Graphical Statistics*, 15(3):542–564, 2006. MR2291263

[BLV10] J. Bigot, J.-M. Loubes, and M. Vimond. Semiparametric estimation of shifts on compact lie groups for image registration. *Probability Theory and Related Fields*, to be published, 2010.

[BM98] L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998. MR1653272

[BP03] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds (i). *Annals of statistics*, 31(1):1–29, 2003. MR1962498

[BP05] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds (ii). *Annals of statistics*, 33:1225–1259, 2005. MR2195634

[DM98] I. L. Dryden and K. V. Mardia. *Statistical shape analysis*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1998. MR1646114

[Fré48] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H.Poincaré, Sect. B, Prob. et Stat.*, 10:235–310, 1948. MR0027464

[GK92] T. Gasser and A. Kneip. Statistical tools to analyze data representing a sample of curves. *Annals of Statistics*, 20(3):1266–1305, 1992. MR1186250

[GL95] R. D. Gill and B. Y. Levit. Applications of the Van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2):59–79, 1995. MR1354456

[GLM07] F. Gamboa, J.-M. Loubes, and E. Maza. Semi-parametric estimation of shifts. *Electron. J. Stat.*, 1:616–640, 2007. MR2369028

[GM01] C. A. Glasbey and K. V. Mardia. A penalized likelihood approach to image warping. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(3):465–514, 2001. MR1858399

[GM07] U. Grenander and M. Miller. *Pattern Theory: From Representation to Inference*. Oxford Univ. Press, Oxford, 2007. MR2285439

[Goo91] C. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B*, 53(2):285–339, 1991. MR1108330

[Gre93] U. Grenander. *General pattern theory - A mathematical study of regular structures*. Clarendon Press, Oxford, 1993. MR1270904

[Hel01] S. Helgason. *Differential geometry, Lie groups, and symmetric spaces*, volume 34 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. MR1834454

[HHM10a] S. Huckemann, T. Hotz, and A. Munk. Intrinsic manova for riemannian manifolds with an application to kendalls spaces of planar shapes. *IEEE Trans. Patt. Anal. Mach. Intell. Special Section on Shape Analysis and its Applications in Image Understanding*, 32(4):593–603, 2010.

[HHM10b] S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: Geodesic principal component analysis for riemannian manifolds modulo lie group actions. discussion paper with rejoinder. *Statistica Sinica*, 20:1–100, 2010. MR2640651

[HJ90] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. MR1084815

[Huc10] S Huckemann. Inference on 3d procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scand. J. Statist.*, To appear, 2010.

[Huc11] S Huckemann. Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.*, 39(2):1098–1124, 2011.

[JDJG04] S. Joshi, B. Davis, B. Jomier, and Guido G. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23:151–160, 2004.

[Kar77] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5):509–541, 1977. MR0442975

[KBCL99] D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and shape theory*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1999. MR1891212

[Ken84] D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math Soc.*, 16:81–121, 1984. MR0737237

[Ken90] W.S. Kendall. Probability, convexity and harmonic maps with small image i: uniqueness and fine existence. *Proc. London Math. Soc.*, 3(61):371–406, 1990. MR1063050

[KG88] A. Kneip and T. Gasser. Convergence and consistency results for self-modelling regression. *Annals of Statistics*, 16:82–112, 1988. MR0924858

[KM97] J. T. Kent and K. V. Mardia. Consistency of Procrustes estimators. *J. Roy. Statist. Soc. Ser. B*, 59(1):281–290, 1997. MR1436569

[Le98] H. Le. On the consitency of procrustean mean shapes. *Advances in Applied Probability*, 30:53–63, 1998. MR1618880

[LK00] H. Le and A. Kume. The fréchet mean shape and the shape of the means. *Advances in Applied Probability*, 32:101–113, 2000. MR1765168

[LM04] X. Liu and H.G. Muller. Functional convex averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association*, 99(467):687–699, 2004. MR2090903

[Mas07] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. MR2319879

[MY01] M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41:61–84, 2001.

[OC95] J. M. Oller and J. M. Corcuera. Intrinsic analysis of statistical estimation. *Ann. Statist.*, 23(5):1562–1581, 1995. MR1370296

[Pen06] X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.*, 25:127–154, July 2006. MR2254442

[RL01] J.O. Ramsay and X. Li. Curve registration. *Journal of the Royal Statistical Society (B)*, 63:243–259, 2001.

[Røn01] B. B. Rønn. Nonparametric maximum likelihood estimation for shifted curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):243–259, 2001. MR1841413

[TIR10] T. Trigano, U. Isserles, and Y. Ritov. Semiparametric curve alignment and shift density estimation for biological data. *Preprint*, 2010.

[TY05] A. Trouvé and L. Younes. Metamorphoses through lie group action. *Foundations of Computational Mathematics*, 5(2):173–198, 2005. MR2149415

[vdV98] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. MR1652247

[Vim10] M. Vimond. Efficient estimation for a subclass of shape invariant models. *Annals of statistics*, 38(3):1885–1912, 2010. MR2662362

[WG97] K. Wang and T. Gasser. Alignment of curves by dynamic time warping. *Annals of Statistics*, 25(3):1251–1276, 1997. MR1447750

[Zie77] H. Ziezold. On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A*, pages 591–602. Reidel, Dordrecht, 1977. MR0501230