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

Jérémie Bigot, Benjamin Charlier

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

Jérémie Bigot[1,2] and Benjamin Charlier[1]

Institut de Mathématiques de Toulouse[1]
Université de Toulouse et CNRS (UMR 5219)

Center for Mathematical Modelling[2],
Universidad de Chile

Jeremie.Bigot@math.univ-toulouse.fr, Benjamin.Charlier@math.univ-toulouse.fr

October 1, 2010

Abstract

A new class of statistical deformable models is introduced to study high-dimensional curves or images. These models are useful to analyze the geometric modes of variation of a data set around a common mean pattern. It is shown that an appropriate tool for statistical inference in such models is the notion of empirical Fréchet mean. This leads to a new procedure to construct a mean pattern from a set of curves or images, and to estimate the shape variability of such data. Using a non-asymptotic framework, we propose consistent estimators of the mean pattern and the deformation parameters modeling the geometric variability of curves or images. Numerical experiments are given to illustrate the finite sample performances of the procedure. An application to the analysis of the geometric variability of a set of images is also proposed.

Keywords: Mean pattern estimation, Fréchet mean, Shape analysis, Deformable models, Curve registration, Image warping, Geometric variability, Principal component analysis, High-dimensional data.
AMS classifications: Primary 62G08; secondary 42C40

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.

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. Let Ω be a subset of $\mathbb{R}^d$, with $d = 1$ for modeling curves and $d = 2, 3$ for modeling two or three dimensional images. 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), the t_ℓ's are non-random points in Ω (deterministic design), the error terms ε_j^ℓ are independent and identically distributed (i.i.d.) normal variables with zero mean and variance 1, and σ represents the level of additive noise.

Here, we suppose that the individual curves or images possess a certain common structure we want to recover. It may lead to the assumption that the observations are random elements which vary around a common mean pattern. Estimating such a mean pattern and characterizing the mode of variations of

the individuals around this common pattern is of fundamental interest in many applications. Principal components analysis (PCA) is widely used to estimate the linear variations in intensity of a data set of curves or images around the usual empirical mean $\bar{Y}_\ell = \frac{1}{J}\sum_{j=1}^{J} Y_j^\ell,\ \ell = 1,\dots,n$. However, in many situations, such data also exhibit a source of geometric variations in time or space which cannot be recovered by standard PCA. 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]

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

where $f^* : \Omega \longrightarrow \mathbb{R}$ is the common 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 such a model, the usual empirical mean is not a consistent estimator of the mean pattern f^* since by the law of large numbers

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

Therefore, in model (1.2), estimating by standard PCA the modes of variation of the data around the usual empirical mean $\bar{Y}_\ell$ is not very meaningful. In such settings, a possible approach is Grenander's pattern theory which considers that the curves or images $f_j, j =, 1,\dots,J$ are obtained through the deformation of a common mean pattern (also called template). More precisely, in Grenander's pattern theory [Gre93], [GM07], images are considered as points in an infinite dimensional manifold and the variations of the images are modeled by the action of Lie groups on the manifold. 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, and we refer to [HHM10a], [HHM10b] and the discussion therein for further details.

To describe more formally such a source of geometric variability, denote by $\mathcal{F}$ some subspace of $L^2(\Omega)$ (the set of square integrable real-valued functions on Ω with respect to the Lebesgue measure dt on $\mathbb{R}^d$) and by Θ a compact subset of $\mathbb{R}^p$. To the set Θ, we associate a parametric family of operators $(T_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ such that for each $\boldsymbol{\theta} \in \Theta$ 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:

- *Randomly shifted curves:* for every $(\boldsymbol{\theta}, f) \in \Theta \times \mathcal{F}$, define the deformation operators

$$T_{\boldsymbol{\theta}} f(t) := f(t - \boldsymbol{\theta}),$$

  with $d = 1$, $\Omega = [0,1]$, $\mathcal{F} \subset L^2_{per}([0,1])$ (the space of periodic functions in $L^2([0,1])$ with period 1) and Θ a compact interval of $\mathbb{R}$.

- *Rigid deformation of two-dimensional images:* for every $(\boldsymbol{\theta}, f) \in \Theta \times \mathcal{F}$, define the deformation operators

$$T_{\boldsymbol{\theta}} f(t) := f\left(e^a R_\alpha t - b\right), \quad \text{for } \boldsymbol{\theta} = (a, \alpha, b) \in \Theta \tag{1.3}$$

  with $d = 2$, $\Omega = \mathbb{R}^2$, $\mathcal{F} \subset L^2(\Omega)$, $\Theta = [-A, A] \times [0, 2\pi] \times [-B, B]^2$ for $A, B > 0$ 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 $\mathcal{F}$ (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 parametrize by a its Lie algebra $\mathcal{G}$ which is a vector space endowed with an extra structure, the Lie bracket. Indeed, there is a canonical local parametrization around the identity given by the exponential map at the identity $\exp : \mathcal{G} \to G$. If Θ is a compact subset of $\mathcal{G}$ such that the restriction of exp to Θ is a diffeomorphism, then for every $(\boldsymbol{\theta}, f) \in \Theta \times \mathcal{F}$ define the deformation operators

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

- *Non-rigid deformation of curves or images:* let us recall that a diffemorphism of Ω is a smooth map $\psi : \Omega \longrightarrow \Omega$ having a smooth inverse with $\psi(\Omega) = \psi^{-1}(\Omega) = \Omega$, and assume that one can construct a family $(\psi_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \Theta}$ of parametric diffeomorphisms of Ω (further details on how to build such a family will be given later on). Then, for every $(\boldsymbol{\theta}, f) \in \Theta \times \mathcal{F}$, define the deformation operators

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

Modeling the geometric variability of curves or images in $\mathcal{F} \subset L^2(\Omega)$ by a generic family of deformation operators $(T_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \Theta}$ thus includes the point of view of Grenander's pattern theory. However, our setting is more general as it is not required that the operation of $(T_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \Theta}$ on $L^2(\Omega)$ corresponds to a Lie group action. 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 $f^* \in \mathcal{F}$ such that

$$f_j(t) = T_{\boldsymbol{\theta}_j^*}\left[f^* + Z_j\right](t), \quad \text{for all } t \in \Omega, \text{ and } j = 1, \ldots, J, \tag{1.4}$$

where $\boldsymbol{\theta}_j^* \in \Theta$, $j = 1, \ldots, J$ are i.i.d. random variables (independent of the ε_j^ℓ's) with an unknown density g with compact support in Θ. The function f^* represents the unknown mean pattern common to all the f_j's. The Z_j's are supposed to be independent realizations of a second order centered Gaussian process Z taking its values in $L^2(\Omega)$ independent of the ε_j^ℓ's. The Z_j's represent the individual variations in amplitude/intensity around the mean pattern f^*. Thus, they correspond to a classical source of variability in intensity that could be analyzed by standard PCA. To the contrary, the random operators $T_{\boldsymbol{\theta}_j}$ model geometric deformations in time or space, and thus correspond to a source of variability in shape in the data. Therefore, if we assume that that the $T_{\boldsymbol{\theta}}$'s are linear operators, equation (1.4) leads to the following *statistical deformable model* for curve or image analysis

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

We emphasis that the error terms ε_j^ℓ and the zero-mean Gaussian processes Z_j are of different kind. The ε_j^ℓ is an additive noise modeling the errors in the measurements while the Z_j's model linear variations in intensity of the individuals around a common mean pattern. Note that a subclass of the deformable model (1.5) is the so-called shape invariant model (SIM)

$$Y_j^\ell = T_{\boldsymbol{\theta}_j^*} f^*(t_\ell) + \sigma \varepsilon_j^\ell, \quad j = 1, \ldots, J, \text{ and } \ell = 1, \ldots, n, \tag{1.6}$$

i.e. without incorporating in (1.5) the additive terms Z_j for $j = 1, \ldots, J$. However, the SIM is not very realistic in practice as observed curves or images generally exhibit a variation in intensity, and we thus focus on the more general deformable model (1.5).

The main goal of this paper is to 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 to use is the notion of empirical Fréchet mean of a data set [Fré48], [BP03], [BP05] that has been widely studied in shape analysis [LK00], [Le98], [Goo91], [KM97], and more recently in biomedical imaging [JDJB04]. In this setting, we derive non-asymptotic consistency results by keeping fixed the number J of observed curves (or images) and the number n of design points. The estimation of f^* and the parameters $\boldsymbol{\theta}_j^*$ in models such as (1.5) can then be used to study geometric modes of variation in a data set when standard PCA is not appropriate.

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 present in detail the case (1.2) of randomly shifted curves, and we give some consistency results for the Fréchet mean in this framework. In Section 4 and Section 5, we give general conditions to extend to the more general deformable model (1.5) the consistency results obtained in the setting of shifted curves. Section 6 contains some numerical experiments, and we present an application to geometric PCA of images using a parametric family of diffeomorphisms. All proofs are postponed to a technical Appendix.

2 The estimating procedure

2.1 Fréchet mean

Fréchet [Fré48] has extended the notion of averaging to general metric spaces via mean squared error minimization in the following way: if $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ denotes i.i.d. random variables with values in a metric space $\mathcal{M}$ with metric $d_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \longrightarrow \mathbb{R}^+$, then the empirical Fréchet mean $\overline{\mathbf{Y}}_{\mathcal{M}}$ of the sample $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ is defined as a minimizer (not necessarily unique) of

$$\overline{\mathbf{Y}}_{\mathcal{M}} \in \operatorname*{argmin}_{y \in \mathcal{M}} \sum_{j=1}^{J} d^2_{\mathcal{M}}(y, \mathbf{Y}_j). \tag{2.1}$$

If $\mathcal{M}$ is a Hilbert space endowed with inner product $\langle \cdot, \cdot \rangle$ and associated distance $d_{\mathcal{M}}(y, y')^2 = \langle y - y', y - y' \rangle$, then the empirical Fréchet mean is unique, and it coincides with the usual empirical mean $\overline{\mathbf{Y}}_{\mathcal{M}} = \frac{1}{J} \sum_{j=1}^{J} \mathbf{Y}_j$. When $\mathcal{M}$ is not a vector space but a nonlinear manifold, a well-known example is the computation of the mean of a set of planar shapes when $\mathcal{M} = \Sigma_2^n$ is the Kendall's shape space [Ken84]. Let $\mathbf{Y} \in \mathbb{R}^{n \times 2}$ be a set of n points in $\mathbb{R}^2$ representing a planar shape, and 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

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

Then, two shapes $\mathbf{Y}_1, \mathbf{Y}_2 \in \mathbb{R}^{n \times 2}$ are said to be equivalent if there exists $\boldsymbol{\theta} \in \Theta$ such that $\mathbf{Y}_1 = T_{\boldsymbol{\theta}} \mathbf{Y}_2$. The set of equivalent classes under this relation is by definition the Kendall's shape space Σ_2^n. Consistent estimation of a mean shape in Σ_2^n has been studied by various authors, see e.g. [LK00], [Le98], and in particular by [Goo91] and [KM97] when a set of random shapes $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ is drawn from the following model

$$\mathbf{Y}_j = T_{\boldsymbol{\theta}_j^*}(\mu + \boldsymbol{\zeta}_j), \ j = 1, \dots, J \tag{2.2}$$

which is very similar to the statistical deformable model (1.5), where $\mu \in \mathbb{R}^{n \times 2}$ is the unknown mean shape to estimate, the $\boldsymbol{\theta}_j^*$'s are i.i.d. random variables in Θ, and the $\boldsymbol{\zeta}_j$'s are i.i.d. centered Gaussian error terms.

More generally, a detailed study of some properties of the Fréchet mean (such as consistency and uniqueness) has been proposed in [BP03] and [BP05] when $\mathcal{M}$ is a finite dimensional Riemannian manifold and $d_{\mathcal{M}}$ is the geodesic distance (which includes the case of Kendall's shape space).

2.2 A dissimilarity measure based on deformation operators

However, the general framework in [BP03] and [BP05] is not adapted to the study of curves or images which are high-dimensional random variables. To define a notion of empirical Fréchet mean for such objects, let us suppose that the family of deformation operators $(T_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \Theta}$ is invertible in the sense that there exists a family of operators $(\tilde{T}_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \Theta}$ such that for any $(\boldsymbol{\theta}, f) \in \Theta \times L^2(\Omega)$

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

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

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

If $d_T^2(h,f) = 0$ then there exists $\boldsymbol{\theta} \in \Theta$ 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 empirical Fréchet mean of data from model (1.5). For this purpose, suppose that $\hat{f}_j$ are smooth functions in $L^2(\Omega)$ obtained from the data Y_j^ℓ, $\ell = 1, \ldots, n$ for $j = 1, \ldots, J$ (further details on this smoothing step will be given later on). Following the definition (2.1) of a Fréchet mean, define an estimator of the mean pattern f^* as

$$\hat{f} = \operatorname*{argmin}_{f\in L^2(\Omega)} \frac{1}{J} \sum_{j=1}^{J} d_T^2(\hat{f}_j, f). \tag{2.3}$$

Then, remark that the computation of $\hat{f}$ can be done in two steps: first minimize the following criterion

$$(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_J) = \operatorname*{argmin}_{(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_J)\in\Theta^J} M(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J), \tag{2.4}$$

where

$$M(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) = \frac{1}{J} \sum_{j=1}^{J} \int_\Omega \left(\tilde{T}_{\boldsymbol{\theta}_j} \hat{f}_j(t) - \frac{1}{J} \sum_{j'=1}^{J} \tilde{T}_{\boldsymbol{\theta}_{j'}} \hat{f}_{j'}(t)\right)^2 dt, \tag{2.5}$$

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

$$\hat{f}(t) = \frac{1}{J} \sum_{j=1}^{J} \tilde{T}_{\hat{\boldsymbol{\theta}}_j} \hat{f}_j(t), \qquad \text{for } t \in \Omega, \tag{2.6}$$

as an estimator of the mean pattern f^*.

Note that this two steps procedure does not require the use of a reference template to compute estimators $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_J$ of the deformation parameters. Thus, it differs from standard algorithms used to compute a mean pattern. Classical approaches are based on an iterative algorithm which is an alternative scheme between computation of deformation parameters $\hat{\boldsymbol{\theta}}_j$ and averaging of back-transformed curves or images using the inverse deformation operators $\tilde{T}_{\hat{\boldsymbol{\theta}}_j}$. Such iterative algorithms lead to the well known Procrustean method,see e.g. [DM98], [Goo91]. The formulation (2.4) is thus an alternative solution to Procrustean methods and, in this paper, we show that minimization of (2.4) can be done by a gradient-descent algorithm.

2.3 Previous work in mean pattern estimation and geometric variability analysis

The problem of estimating the mean pattern of a set of curves that differ by a time transformation is usually referred to as the curve registration problem. Among the various methods that have been proposed, one can distinguish between landmark-based approaches see e.g. [GK92], [Big06], and non-parametric time warping to align a set of curves see e.g. [RL01], [WG97], [LM04]. However, in these papers, studying consistent estimators of the common shape f^* as the number of curves J and design points n tend to infinity is not considered. For the SIM (1.6), 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 the estimation of scaling, rotation and translation parameters for two-dimensional images is also proposed in [BGV09], but also with a fixed number J of observed images. 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 supposed to be a parametric transformation of a common 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). This is a limitation of the SEMOR approach as, in a lot of applications, the functions f_j are varying locally in intensity from one individual to another. Estimation in the SEMOR model is done using a Procrustean algorithm, and [KG88] studied the consistency of such procedure in an asymptotic framework where both the number of curves J and the number n of design points grow to infinity.

The model (1.5) is also very much connected to the well-known problem of image warping. There is a wide literature on this subject. Recently, there has 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] and references therein. The deformable model (1.5) to account for variability in shape and intensity is also related to the theory of metamorphoses developed by [TY05] in which infinitesimal variations of curves or images are modeled as a combination of elastic deformations and photometric variations. However, the approach followed in [TY05] is purely deterministic in the sense that it is not focused on the analysis of random variations in shape and intensity or on the estimation of a mean pattern.

In this paper, a general theory is sought to unify previous work in statistics on curve registration and image warping. For this purpose, we propose to use the notion of empirical Fréchet mean, which, to the best of our knowledge, has not been considered before in a statistical setting using general deformable models such as (1.5).

3 The case of randomly shifted curves

Let us first consider the case of randomly shifted curves (1.2) with an equispaced design to give an idea of the consistency results that can be expected using an estimating procedure based on empirical Fréchet mean. In this setting, the model (1.5) can be written as

$$Y_j^\ell = f^*\left(\tfrac{\ell}{n} - \boldsymbol{\theta}_j^*\right) + Z_j\left(\tfrac{\ell}{n} - \boldsymbol{\theta}_j^*\right) + \sigma\varepsilon_j^\ell, \quad j = 1, \dots, J, \text{ and } \ell = 1, \dots, n. \tag{3.1}$$

3.1 The random perturbations Z_j

First, let us make the following assumption on the random perturbation Z of the mean pattern f^*.

Assumption 3.1. *The Z_j's in* (3.1) *are i.i.d. copies of a second order stationary Gaussian process Z taking its value in $L^2_{per}([0,1])$ with zero mean and covariance function $R : [0,1] \to \mathbb{R}$.*

Under Assumption 3.1, the law of Z is thus invariant by the action of a shift. This hypothesis is similar to the condition given in [KM97] to ensure consistency of Fréchet mean estimators in model (2.2), where, after a normalization step, the deformations considered are rotations of the plane. The authors in [KM97] study the case where the law of the error term $\boldsymbol{\zeta}_j$ in (2.2) is isotropic, that is to say, invariant by the action of rotations.

Throughout this paper $\mathbb{E}_{\boldsymbol{\theta}}[\cdot]$ denotes expectation conditionally to $\boldsymbol{\theta} \in \Theta^J$. For any random element $\boldsymbol{\theta}_j^* \in \Theta$, the vector $\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j = \left[Z_j(\frac{\ell}{n} - \boldsymbol{\theta}_j^*)\right]_{\ell=1}^n$ is, conditionally on $\boldsymbol{\theta}_j^*$, a centered Gaussian random variable and its covariance matrix is a Toeplitz matrix equals to

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

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

$$\gamma_{\max}\big(\boldsymbol{\Sigma}\big) \leq \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} \left|R\left(\tfrac{k}{n}\right)\right| = \gamma \tag{3.2}$$

where $\gamma = \int_0^1 |R(t)|\, dt$ is a positive constant independent of n. The constant γ can be interpreted as a measure of the "variance" of the random perturbations $Z_j, j = 1, \ldots, J$ in model (3.1).

3.2 Choice of the smoothing estimators $\hat{f}_j$

A convenient choice for the smoothing of the observed curves in (3.1) is to do low-pass Fourier filtering. Define the empirical Fourier coefficient at the frequency k by $\hat{c}_{j,k} = \frac{1}{n}\sum_{\ell=1}^n Y_j^\ell e^{-i2\pi k\frac{\ell}{n}}$ for $k = -(n-1)/2, \ldots, (n-1)/2$ (assuming for simplicity that n is odd). It gives for a fixed smoothing parameter $\lambda \in \mathbb{N}$ and $t \in [0,1]$ the following linear estimators

$$\hat{f}_j^\lambda(t) = \sum_{|k|\leq\lambda} \hat{c}_{j,k} e^{i2\pi kt} = \frac{1}{n}\sum_{|k|\leq\lambda}\sum_{\ell=1}^n e^{-i2\pi\frac{\ell}{n}k} e^{i2\pi kt} Y_j^\ell = \frac{1}{n}\sum_{\ell=1}^n \sum_{|k|\leq\lambda} e^{i2\pi k(t-\frac{\ell}{n})} Y_j^\ell = \langle S_\lambda(t), \mathbf{Y}_j\rangle\,, \quad (3.3)$$

where $S_\lambda(t) = \left[\frac{1}{n}\sum_{|k|\leq\lambda} e^{i2\pi k(t-\frac{\ell}{n})}\right]_{\ell=1}^n$. Then, define the following Sobolev ball $H_s(A)$ of radius $A > 0$ and regularity $s > 0$ using a characterization in term of the Fourier coefficients $c_k(f) = \int_0^1 f(t)e^{-i2\pi kt}dt$, $k \in \mathbb{Z}$ for a function $f \in L^2_{per}([0,1])$,

$$H_s(A) = \Big\{ f \in L^2_{per}([0,1]), \sum_{k\in\mathbb{Z}} (1+|k|^2)^s\, |c_k(f)|^2 < A \Big\}. \quad (3.4)$$

Then, as a possible choice for the smoothness class $\mathcal{F}$ to which the mean pattern f^* is supposed to belong, take $\mathcal{F} = H_s(A)$.

3.3 Identifiability conditions

Suppose that $\Theta = [-\rho, \rho]$ for some $\rho > 0$. Using low-pass filtering, the estimators of the random shifts $\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_J^*$ are given by

$$(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_J) = \underset{(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_J)\in\Theta^J}{\operatorname{argmin}} M_\lambda(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J), \quad (3.5)$$

where the criterion $M_\lambda(\boldsymbol{\theta}) = M_\lambda(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J)$ for $\boldsymbol{\theta} \in \Theta^J$, see equation (2.5), has a simple expression in the Fourier domain. Indeed, thanks to Parseval's relation and the fact that the translation of a function corresponds to a frequency modulation of its Fourier coefficients, it follows that

$$\begin{aligned} M_\lambda(\boldsymbol{\theta}) &= \frac{1}{J}\sum_{j=1}^J \int_\Omega \left(\hat{f}_j(t+\boldsymbol{\theta}_j) - \frac{1}{J}\sum_{j'=1}^J \hat{f}_{j'}(t+\boldsymbol{\theta}_{j'}) \right)^2 dt \\ &= \frac{1}{J}\sum_{j=1}^J \sum_{|k|\leq\lambda} \left| \hat{c}_{j,k} e^{i2\pi k\boldsymbol{\theta}_j} - \frac{1}{J}\sum_{j'=1}^J \hat{c}_{j',k} e^{i2\pi k\boldsymbol{\theta}_{j'}} \right|^2. \end{aligned}$$

However, the minimization (3.5) is not well defined. Indeed, if $\boldsymbol{\theta}_0 \in \Theta$ satisfies $\hat{\boldsymbol{\theta}}_j + \boldsymbol{\theta}_0 \in \Theta$, $j = 1, \ldots, J$, then $(\hat{\boldsymbol{\theta}}_1 + \boldsymbol{\theta}_0, \ldots, \hat{\boldsymbol{\theta}}_J + \boldsymbol{\theta}_0)$ is also a global minimum of $M_\lambda(\boldsymbol{\theta})$. This comes from the fact that, without any further assumptions on the set Θ^J and the density g of the random shifts, then the model (3.1) is not identifiable (to see this, simply replace in (3.1) $f^*(\cdot)$ by $f^*(\cdot - \boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}_j^*$ by $\boldsymbol{\theta}_j^* + \boldsymbol{\theta}_0$). Choosing identifiability conditions in model (3.1) amounts to impose constraints on the minimization of the criterion

$$D(\boldsymbol{\theta}) = \frac{1}{J}\sum_{j=1}^J \int_\Omega \left(f^*(t - \boldsymbol{\theta}_j^* + \boldsymbol{\theta}_j) - \frac{1}{J}\sum_{j'=1}^J f^*(t - \boldsymbol{\theta}_{j'}^* + \boldsymbol{\theta}_{j'}) \right)^2 dt. \quad (3.6)$$

This criterion $D(\boldsymbol{\theta})$ can be interpreted as a version without noise of the criterion $M_\lambda(\boldsymbol{\theta})$. Obviously, 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, the minimizer

of D on Θ^J is clearly not unique, and minimizing $M_\lambda(\boldsymbol{\theta})$ does not allow to recover the true shifts $\boldsymbol{\theta}^*$ nor the true mean pattern f^*. 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

$$\boldsymbol{\Theta}_0 = \{(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 + \dots + \boldsymbol{\theta}_J = 0\} \tag{3.7}$$

It is shown in [BG10] (Lemma 6) that if f^* 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 $\boldsymbol{\Theta}_0$ at

$$\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} = (\boldsymbol{\theta}_1^* - \bar{\boldsymbol{\theta}}^*, \dots, \boldsymbol{\theta}_J^* - \bar{\boldsymbol{\theta}}^*) \text{ where } \bar{\boldsymbol{\theta}}^* = \frac{1}{J} \sum_{j=1}^J \boldsymbol{\theta}_j^*. \tag{3.8}$$

Under such assumptions, it is thus natural to compute an estimator $\hat{\boldsymbol{\theta}}^\lambda$ of the random shifts over the constrained set $\boldsymbol{\Theta}_0$ defined as

$$\hat{\boldsymbol{\theta}}^\lambda = (\hat{\boldsymbol{\theta}}_1^\lambda, \dots, \hat{\boldsymbol{\theta}}_J^\lambda) = \underset{(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \boldsymbol{\Theta}_0}{\operatorname{argmin}} M_\lambda(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J). \tag{3.9}$$

Therefore, introduce the following identifiability conditions:

Assumption 3.2. *The support of the density g is included in $\Theta = [-\rho, \rho]$ for some $0 < \rho < 1/4$ and is such that $\int_\Theta \boldsymbol{\theta} g(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0$.*

Assumption 3.3. *The mean pattern f^* is such that $\int_0^1 f^*(t) e^{-i2\pi t} dt \neq 0$.*

Following the classical guidelines in M-estimation (see e.g. [vdV98]), a necessary condition to ensure the convergence of M-estimators such as (3.9) is that the local minima of $D(\boldsymbol{\theta})$ over $\boldsymbol{\Theta}_0$ are well separated from the global minimum of $D(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}$ (satisfying $D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}) = 0$). An example of such a separability condition is the following one:

Assumption 3.4. *For any $\delta > 0$, there exists a constant $C(\Theta, f^*, \delta) > 0$ (not depending on J) such that for any $\boldsymbol{\theta}^* \in \Theta^J$*

$$\min_{\boldsymbol{\theta} \in B^c(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}, \delta)} D(\boldsymbol{\theta}) \geq C(\Theta, f^*, \delta), \tag{3.10}$$

where $B^c(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}, \delta)$ denotes the complementary of $B(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}, \delta) = \left\{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0,\ \left|\boldsymbol{\theta}_j - [\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}]_j\right| \leq \delta,\ j = 1, \dots, J\right\}$, where $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} = \left([\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}]_1, \dots, [\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}]_J\right)$ is the vector defined in (3.8) and $\boldsymbol{\Theta}_0$ is the constrained set (3.7).

Assumption 3.2 implies that the criterion $D(\boldsymbol{\theta})$ has a unique minimum on $\boldsymbol{\Theta}_0$. The condition (3.10) guarantees that, outside a ball $B(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}, \delta)$ centered at $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}$, the local maxima of $D(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ are sufficiently separated from 0. In Assumption 3.4, the important fact is that the constant $C(\Theta, f^*, \delta)$ does not depend on the dimension J of the set $\boldsymbol{\Theta}_0$ of parameters onto which M_λ and D are minimized. In the Fourier domain, the criterion $D(\boldsymbol{\theta})$ can be written as

$$D(\boldsymbol{\theta}) = \sum_{k \in \mathbb{Z}} |c_k^*|^2 \left(1 - \left|\frac{1}{J} \sum_{j=1}^J e^{i2\pi k(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)}\right|^2\right), \quad \text{where } c_k^* = \int_0^1 f^*(t) e^{-i2\pi k t} dt.$$

Thus the value of the constant $C(\Theta, f^*, \delta)$ depends on a precise upper bound on the exponential sums $\frac{1}{J}\sum_{j=1}^J e^{i2\pi k(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)}$ for $k \in \mathbb{Z}$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in B^c(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}, \delta)$ which is a delicate problem. Nevertheless, if the size of $\boldsymbol{\Theta}$ is sufficiently small, the following proposition shows that Assumption 3.4 is satisfied.

Proposition 3.1. *Suppose that Assumptions 3.2 and 3.3 hold with $\rho < 1/16$. Then, Assumption 3.4 is satisfied with*

$$C(\Theta, f^*, \delta) = C(f^*, \rho)\delta^2,$$

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

A proof is given in the Appendix. The condition that $\rho < 1/16$ means 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 the concentration of the data around a common mean pattern has been used in various papers to prove the uniqueness and the consistency of Fréchet mean for random variables lying in a Riemannian manifold, see [BP03], [BP05], [LK00], [Le98].

3.4 Consistent estimation of the random shifts

Under such assumptions, the theorem below gives the deviation in probability between $\hat{\boldsymbol{\theta}}^{\lambda}$ and $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_J^*)$.

Theorem 3.1. *Consider the model* (3.1). *Let* $\hat{\boldsymbol{\theta}}^{\lambda}$ *be the estimator defined by (3.9) and assume that* $\mathcal{F} = H_s(A)$ *for some* $A > 0$ *and* $s \geq 3$. *Suppose that Assumptions 3.1 to 3.4 hold. Then, for any* $\lambda \geq 1$ *and* $x > 0$

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

with

$$A_1(x, J, n, \lambda, \sigma^2, \gamma) = (\sigma^2 + \gamma)\left(\sqrt{\upsilon(x, J, n, \lambda)} + \upsilon(x, J, n, \lambda)\right) + \left(\sqrt{B(\lambda, n)} + B(\lambda, n)\right)$$

and

$$A_2(x, J) = \left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2,$$

where $C_1(\Theta, \mathcal{F}, f^*) > 0$ *is constant depending only on* $\Theta, \mathcal{F}, f^*$, $\|.\|$ *is the standard Euclidean norm in* $\mathbb{R}^J$,

$$\upsilon(x, J, n, \lambda) = \frac{2\lambda + 1}{n}\left(1 + 4\frac{x}{J} + \sqrt{4\frac{x}{J}}\right) \text{ and } B(\lambda, n) = \frac{2\lambda + 1}{n} + \lambda^{-2s}.$$

Let us make some comments on the deviation inequality in Theorem 3.1. The result follows from the inequality,

$$\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*\|^2 \leq \frac{2}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2 + \frac{2}{J}\|\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} - \boldsymbol{\theta}^*\|^2.$$

The term $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ comes from the control of the quantity $\frac{2}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$, while the term $A_2(x, J)$ is derived from a Bernstein's inequality on $\frac{2}{J}\|\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} - \boldsymbol{\theta}^*\|^2 = 2\left|\frac{1}{J}\sum_{j=1}^J \boldsymbol{\theta}_j^*\right|^2$ (see the Appendix for further details). Remark that for fixed values of the number of design points n and the spectral cutoff λ, then $\lim_{J\to+\infty} A_2(x, J) = 0$. This means that for J sufficiently large then $\overline{\boldsymbol{\theta}}^* = \frac{1}{J}\sum_{j=1}^J \boldsymbol{\theta}_j^* \approx 0$ and thus $\boldsymbol{\theta}_{\boldsymbol{\Theta}_0} \approx \boldsymbol{\theta}^*$. Hence, the convergence in probability of $\boldsymbol{\theta}_{\boldsymbol{\Theta}_0}$ to $\boldsymbol{\theta}^*$ is ensured by an asymptotic only in J.

The term $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ is more complex. It depends on the spectral cutoff λ via the bias $B(\lambda, n)$ and the variance $\upsilon(x, J, n, \lambda)$ used to define the smoothing estimators $\hat{f}_j$. An optimal choice of the parameter λ depends on n. By choosing a sequence $\lambda = \lambda_n$ such that $\lim_{n\to+\infty} \lambda_n = +\infty$ and $\lim_{n\to+\infty} \frac{\lambda_n}{n} = 0$, we have a tradeoff between low variance and low bias. By using such a choice for λ, it follows that for fixed J and $x > 0$, then $\lim_{n\to+\infty} B(\lambda_n, n) = \lim_{n\to+\infty} \upsilon(x, J, n, \lambda_n) = 0$, which implies that $\lim_{n\to+\infty} A_1(x, J, n, \lambda_n, \sigma^2, \gamma) = 0$. However, if n remains fixed, then $\lim_{J\to+\infty} A_1(x, J, n, \lambda, \sigma^2, \gamma) = h(\frac{2\lambda+1}{n}) + h(\frac{2\lambda+1}{n} + \lambda^{-2s})$, where $h(u) = \sqrt{u} + u$, for all $u \geq 0$. Thus, an asymptotic only in J is not sufficient to ensure that $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ converges to 0.

Hence, under the assumptions of Theorem 3.1, one can only prove the convergence in probability of $\hat{\boldsymbol{\theta}}^{\lambda}$ to the true shifts $\boldsymbol{\theta}^*$ by taking a double asymptotic setting by letting n and J going to $+\infty$ (provided the smoothing parameter $\lambda = \lambda_n$ is well chosen)). This would suggest that, if the number

n of design points is fixed, then it is in general not possible to estimate $\boldsymbol{\theta}^*$ as the number of curves J grows to infinity. Nevertheless, if J is fixed, $\hat{\boldsymbol{\theta}}^{\lambda}$ converges in probability to $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}$ as n goes to infinity. This means that Theorem 3.1 gives an upper bound on $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*\|^2$ which suggests that an asymptotic only in J with n fixed is not sufficient to estimate $\boldsymbol{\theta}^*$. In fact, the next theorem shows that it is possible to derive a lower bound on $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*\|^2$.

Theorem 3.2. *Consider the model* (3.1). *Suppose that Assumption 3.1 holds, that f^* is continuously differentiable on $[0,1]$, and that the support of the density g is included in $\Theta = [-\rho, \rho]$ with $\lim_{\boldsymbol{\theta}\to\pm\rho} g(\boldsymbol{\theta}) = 0$ 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 $\boldsymbol{\theta}^*$, i.e. a measurable function of the data in model (3.1). Then, for any $n \geq 1$ and $J \geq 1$*

$$\mathbb{E}\left[\frac{1}{J}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2\right] \geq \frac{n^{-1}(\sigma^2 + \gamma)}{\|\partial_t f^*\|_{\infty}^2 + n^{-1}(\sigma^2+\gamma)\int_{\Theta}\left(\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta})\right)\right)^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{3.11}$$

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

The important fact is that the right hand side of inequality (3.11) does not depend on J. Therefore, if the number of design points n is fixed, then Theorem 3.2 shows that it is impossible to recover the true shifts in model (3.1) by letting only the number of curves J going to $+\infty$. On the other hand, the lower bound (3.11) tends to zero as $n \to +\infty$, which is consistent with the above discussion on the fact that $\hat{\boldsymbol{\theta}}^{\lambda}$ converges to $\boldsymbol{\theta}^*$ as $n \to +\infty$ and $J \to +\infty$. An illustration of these facts are given in Section 6 on numerical experiments.

3.5 Consistent estimation of the mean pattern

The Fréchet mean estimator of f^* defined by (2.6) can now be written as

$$\hat{f}^{\lambda}(t) = \frac{1}{J}\sum_{j=1}^{J} \hat{f}_j^{\lambda}(t + \hat{\boldsymbol{\theta}}_j^{\lambda}) = \frac{1}{J}\sum_{j=1}^{J}\left\langle S_{\lambda}(t + \hat{\boldsymbol{\theta}}_j^{\lambda}), \mathbf{Y}_j\right\rangle, \qquad \text{for all } t \in [0,1]. \tag{3.12}$$

The theorem below gives a deviation in probability between $\hat{f}^{\lambda}$ and f^*.

Theorem 3.3. *Under assumptions of Theorem 3.1, 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 3.1, $C_2(\Theta, \mathcal{F}, f^*)$ and $C_3(\Theta, f^*)$ are positive constants depending only on $\Theta, \mathcal{F}, f^*$, and $\|\hat{f}^{\lambda} - f^*\|_{L^2}^2 = \int_0^1 \left|f^{\lambda}(t) - f^*(t)\right|^2 dt$.

The proof is given in the Appendix. The terms appearing in the deviation inequality given in Theorem 3.3 are the same as those appearing in Theorem 3.1. Therefore, similar comments to those made on the consistency of the estimators of the shifts can be made. Consider the function $f^*_{\boldsymbol{\Theta}_0}(t) := \frac{1}{J}\sum_{j=1}^{J} f^*(t - \boldsymbol{\theta}_j^* + [\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}]_j) = f^*(t - \bar{\boldsymbol{\theta}}^*)$ for $t \in [0,1]$, where $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} = ([\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}]_1, \ldots, [\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}]_J)$ and the following inequality $\|\hat{f}^{\lambda} - f^*\|_{L^2}^2 \leq 2\|\hat{f}^{\lambda} - f^*_{\boldsymbol{\Theta}_0}\|_{L^2}^2 + 2\|f^*_{\boldsymbol{\Theta}_0} - f^*\|_{L^2}^2$. Again, $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ controls the quantity $\|\hat{f}^{\lambda} - f^*_{\boldsymbol{\Theta}_0}\|_{L^2}^2$ while $A_2(x, J)$ controls $\|f^*_{\boldsymbol{\Theta}_0} - f^*\|_{L^2}^2$. Thus, 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^*. If J remains fixed, the Fréchet mean converges to the shifted version $f^*_{\boldsymbol{\Theta}_0}$ of f^* as $n \to +\infty$.

When the number of design points n is fixed and we only let J going to infinity, Theorem 3.2 shows that it is not possible to recover the true shifts. It is thus expected that the Fréchet mean $\hat{f}^{\lambda}$ should

not converge to the mean pattern f^* in the setting n fixed and $J \to +\infty$. To support this argument, consider the following ideal estimator

$$\tilde{f}(t) = \frac{1}{J}\sum_{j=1}^{J} f_j^*(t + \hat{\boldsymbol{\theta}}_j^{\lambda}) = \frac{1}{J}\sum_{j=1}^{J} f^*(t - \boldsymbol{\theta}_j^* + \hat{\boldsymbol{\theta}}_j^{\lambda}), \qquad \text{for all } t \in [0,1], \tag{3.13}$$

where $f_j^*(t) = f^*(t - \boldsymbol{\theta}_j^*), j = 1, \dots, J$. This corresponds to the case of an ideal smoothing step from the data (3.1) that would yield $\hat{f}_j = f_j^*$ for all $j = 1, \dots, J$. Obviously, $\tilde{f}(t)$ is not an estimator since it depends on the unobserved quantities f^* and $\boldsymbol{\theta}_j^*$, but we can consider it as a benchmark to analyse the converge of the Fréchet mean $\hat{f}^{\lambda}$ to f^*.

Theorem 3.4. *Suppose that the assumptions of Theorem 3.2 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*

$$\mathbb{E}\|\tilde{f} - f^*\|_{L^2} \geq C(f^*, \rho) \frac{n^{-1}(\sigma^2 + \gamma)}{\|\partial_t f^*\|_{\infty}^2 + n^{-1}(\sigma^2 + \gamma)\int_{\Theta} \left(\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta})\right)\right)^2}, \tag{3.14}$$

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

As for Theorem 3.2 the right hand side of inequality (3.14) does not depend on J. Thus, in the setting n fixed and $J \to +\infty$, the ideal estimator $\tilde{f}$ does not converge to f^* for the expected quadratic risk. This supports the argument that, when using the Fréchet mean $\hat{f}^{\lambda}$, one cannot reconstruct the mean pattern f^* when the shifts are estimated in an asymptotic setting where the number of design points n is fixed. Indeed, in such a setting, the estimation of the shifts $\boldsymbol{\theta}_j^*, j = 1, \dots, J$ is limited by the "variance" γ of the random perturbations Z_j and the variance σ^2 of the additive Gaussian noise in model (3.1).

3.6 Discussion on the various asymptotic settings

To sum up the above discussion, we compile the results on the convergence (in probability or in expectation) of $\hat{\boldsymbol{\theta}}^{\lambda}$ and $\hat{f}^{\lambda}$ in the following diagram,

$$\hat{\boldsymbol{\theta}}^{\lambda} \underset{n,J}{\longrightarrow} \boldsymbol{\theta}^* \qquad \hat{\boldsymbol{\theta}}^{\lambda} \underset{J}{\not\longrightarrow} \boldsymbol{\theta}^* \ (n \text{ fixed}) \qquad \hat{\boldsymbol{\theta}}^{\lambda} \underset{n}{\longrightarrow} \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} \ (J \text{ fixed}) \qquad \hat{\boldsymbol{\theta}}^{\lambda} \underset{J}{\not\longrightarrow} \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0} \ (n \text{ fixed})$$

$$\hat{f}^{\lambda} \underset{n,J}{\longrightarrow} f^* \qquad \hat{f}^{\lambda} \underset{n}{\longrightarrow} f^*_{\boldsymbol{\Theta}_0} \ (J \text{ fixed})$$

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

In the above results, it has been assumed that the process Z is stationary, and we have restricted the analysis to the case of random shifts in dimension $d = 1$. The purpose of the next sections is to give sufficient conditions to generalize these results to the case of more complex deformation operators and non-stationary processes Z.

4 Notations and main assumptions in the general case

In this section, we define our main assumptions on the shape function f^*, the smoothing step, the family of deformation operators, and the random processes $Z_j, j = 1, \dots, J$. As an illustrative example, necessary conditions for which such assumptions hold are given for the randomly shifted curve model (1.2). In what follows, C, C_0, C_1, C_2, denote positive constants whose value may change from line to line. The notations $C(.)$ specify the dependency of C on some quantities. We use a bold symbol $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \Theta^J \subset \mathbb{R}^{pJ}$ to denote a J-ads of vectors in Θ, and the notation $\boldsymbol{\theta}_j = (\theta_j^1, \dots, \theta_j^p) \in \mathbb{R}^p$ to denote the components of the vector $\boldsymbol{\theta}_j \in \Theta \subset \mathbb{R}^p$. Note that, when context is clear, we also use the notation $\boldsymbol{\theta}$ to denote an element $\boldsymbol{\theta} = (\theta^1, \dots, \theta^p) \in \Theta$. The standard Euclidean norm of a vector $\mathbf{c}$ in $\mathbb{R}^k$ is denoted by $\|\mathbf{c}\|$. The L^2 norm of a function $f \in L^2(\Omega)$ is denoted by $\|f\|_{L^2}^2 = \int_{\Omega} |f(t)|^2 \, dt$.

4.1 Smoothness of the mean pattern and the deformation operators

The following assumptions on the deformation operators will be used throughout the paper. In this part, the notation $(\mathcal{L}_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ is used to denote either $(T_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ or their inverse $(\tilde{T}_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$.

Assumption 4.1 (Smoothness class $\mathcal{F}$ and regularity of the deformation operators)**.** *For all $\boldsymbol{\theta}\in\Theta$, $\mathcal{L}_{\boldsymbol{\theta}} : L^2(\Omega) \longrightarrow L^2(\Omega)$ is a linear operator satisfying $\mathcal{L}_{\boldsymbol{\theta}} f \in \mathcal{F}$ for all $f \in \mathcal{F}$. The smoothness class $\mathcal{F}$ is such that the functions $t \longmapsto \partial_{\theta^{p_1}} \mathcal{L}_{\boldsymbol{\theta}} f(t)$, $t \longmapsto \partial_{\theta^{p_2}} \partial_{\theta^{p_1}} \mathcal{L}_{\boldsymbol{\theta}} f(t)$ and $t \longmapsto \partial_{\theta^{p_3}} \partial_{\theta^{p_2}} \partial_{\theta^{p_1}} \mathcal{L}_{\boldsymbol{\theta}} f(t)$ belong to $L^2(\Omega)$ for any $f \in \mathcal{F}$ and all $p_1, p_2, p_3 = 1, \ldots, p$.*

Assumption 4.2 (Continuity of the deformation operators)**.** *There exists a constant $C(\Theta) > 0$ such that for any $f \in L^2(\Omega)$,*

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

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 4.3 (Differentiability of the deformation operators)**.** *There is a constant $C(\Theta, \mathcal{F}, f^*) > 0$ such that,*

$$\max\left\{ \|\partial_{\theta_2^{p_1}} \tilde{T}_{\boldsymbol{\theta}_2} T_{\boldsymbol{\theta}_1} f^*\|_{L^2}^2, \|\partial_{\theta_2^{p_2}} \partial_{\theta_2^{p_1}} \tilde{T}_{\boldsymbol{\theta}_2} T_{\boldsymbol{\theta}_1} f^*\|_{L^2}^2, \|\partial_{\theta_2^{p_3}} \partial_{\theta_2^{p_2}} \partial_{\theta_2^{p_1}} \tilde{T}_{\boldsymbol{\theta}_2} T_{\boldsymbol{\theta}_1} f^*\|_{L^2}^2 \right\} \leq C(\Theta, \mathcal{F}, f^*),$$

for all $p_1, p_2, p_3 = 1, \ldots, p$ and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$.

Assumption 4.2 can be interpreted as a Lipschitz condition on the mapping $(f, \boldsymbol{\theta}) \longmapsto \mathcal{L}_{\boldsymbol{\theta}} f$ and Assumption 4.1 ensures the differentiability of the functional $M(\boldsymbol{\theta})$ defined in (2.5). The first inequality of Assumption 4.2, 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 (2.4) to have non trivial solutions by avoiding the functional $M(\boldsymbol{\theta})$ in (2.5) being arbitrarily small. Indeed, consider as an illustration the case where $p = 1$, $\Omega = [0, 1]$ and the $T_{\boldsymbol{\theta}}$'s are the following scaling operators:

$$T_{\boldsymbol{\theta}} f(t) = e^{\boldsymbol{\theta}} f(t) \quad \text{and} \quad \tilde{T}_{\boldsymbol{\theta}} f(t) = e^{-\boldsymbol{\theta}} f(t)$$

for all $t \in \Omega$. In this case $\|T_{\boldsymbol{\theta}} f\|_{L^2}^2 = e^{2\boldsymbol{\theta}} \|f\|_{L^2}^2$ and $\|\tilde{T}_{\boldsymbol{\theta}} f\|_{L^2}^2 = e^{-2\boldsymbol{\theta}} \|f\|_{L^2}^2$. To satisfy Assumption 4.2, a convenient choice is $\Theta = [-\rho, \rho]$ for $\rho > 0$. Suppose now that we do not restrict to a compact set and we have $\Theta = \mathbb{R}$. Since $M(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \leq \frac{1}{J} \sum_{j=1}^J e^{-2\boldsymbol{\theta}_j} \|\hat{f}_j\|_{L^2}^2$, one could let the $\boldsymbol{\theta}_j$'s going to $+\infty$ and the optimization problem (2.5) would have a trivial solution at $\hat{\boldsymbol{\theta}}_j = +\infty, j = 1, \ldots, J$ such that $M(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) = 0$. A particular case where the first inequality of Assumption 4.2 holds is when the $\mathcal{L}_{\boldsymbol{\theta}}$'s are isometric, i.e $\|\mathcal{L}_{\boldsymbol{\theta}} f\|_{L^2}^2 = \|f\|_{L^2}^2$, for any $f \in L^2(\Omega)$. In [HHM10a] and [HHM10b], the authors study the case of compact Lie groups acting isometrically on a finite dimensional space (e.g. rigid deformation of k-ads of the plane) which allows them to derive nice geometric structure on the quotient spaces. However this condition is rather restrictive and in a lot of applications, the deformations operators are not isometric, e.g. in the non-rigid deformation of images or curves.

As an illustration we now check these assumptions in the case (1.2) of randomly shifted curves. One has that $p = 1$, $\Omega = [0, 1]$, $T_{\boldsymbol{\theta}}(f)(t) = f(t - \boldsymbol{\theta})$ and $\tilde{T}_{\boldsymbol{\theta}}(f)(t) = f(t + \boldsymbol{\theta})$. These operators act isometrically on $L^2_{per}([0, 1])$, as a change of variable implies immediately that $\|\mathcal{L}_{\boldsymbol{\theta}} f\|_{L^2}^2 = \|f\|_{L^2}^2$. To verify Assumptions 4.1, 4.2 and 4.3, take as in Section 3

$$\mathcal{F} = H_s(A) \text{ with } s \geq 3, \text{ see equation (3.4).}$$

This implies that a function $f \in \mathcal{F}$ is at least three times differentiable and that

$$\max\left\{ \|\partial_t f\|_{L^2}^2, \|\partial_t^2 f\|_{L^2}^2, \|\partial_t^3 f\|_{L^2}^2 \right\} \leq A.$$

Moreover, we have $\partial_\theta \mathcal{L}_{\boldsymbol{\theta}} f(t) = \pm \partial_t f(t \pm \boldsymbol{\theta})$, $\partial^2_\theta \mathcal{L}_{\boldsymbol{\theta}} f(t) = \partial^2_t f(t \pm \boldsymbol{\theta})$ and $\partial^3_\theta \mathcal{L}_{\boldsymbol{\theta}} f(t) = \pm \partial^3_t f(t \pm \boldsymbol{\theta})$, where for a function $f \in \mathcal{F}$, $\partial_t f$ denotes its first derivative, $\partial^2_t f$ its second derivative at point t and so on. Thus, Assumption 4.3 is verified with the constant $C(\Theta, \mathcal{F}) = A$. Then, remark that a change of variable gives for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, $\|\tilde{T}_{\boldsymbol{\theta}_1} f - \tilde{T}_{\boldsymbol{\theta}_2} f\|^2_{L^2} = \|f(\cdot + \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) - f\|^2_{L^2}$. Since $f \in H_s(A)$, the mean value theorem can be used to prove that $\|\tilde{T}_{\boldsymbol{\theta}_1} f - \tilde{T}_{\boldsymbol{\theta}_2} f\|^2_{L^2} \leq C(A, \Theta)\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^2$ for some constant $C(A, \Theta) > 0$, and thus Assumption 4.2 holds.

4.2 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 4.1. *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 $t \in \Omega$*

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

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

Assumption 4.4. *For all $\lambda \in \Lambda$ and all $\ell = 1, \ldots, n$, the function $t \longmapsto S^\ell_\lambda(t)$ belong to $L^2(\Omega)$, where $S^\ell_\lambda(t)$ 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 $\mathbf{T}_{\boldsymbol{\theta}} \mathbf{f} = \big(T_{\boldsymbol{\theta}} f(t_\ell)\big)_{\ell=1}^n$.*

Note that it follows from the above assumption, that the functions $t \longmapsto \hat{f}_j(t) = \langle S_\lambda(t), \mathbf{Y}_j \rangle$, $j = 1, \ldots, J$ belong to $L^2(\Omega)$. Typically the vector $S_\lambda(t)$ depends on the design points $(t_\ell)_{\ell=1}^n$. In a one-dimensional setting ($d = 1$) and for $\Omega = [0, 1]$ a typical example is low-pass Fourier filtering with an equi-spaced design as defined in Section 3. Another illustrative example is spline smoothing in the setting of Reproducing Kernel Hilbert Spaces (RKHS), for which $\Lambda = \mathbb{R}_+$ but not necessarily with an equi-spaced design (see [Wah90] for an introduction to spline smoothing and RKHS).

Let us now specify how the bias/variance behavior of the linear estimators $\hat{f}^\lambda_j$ 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 $\mathbf{Y} = (Y_\ell)_{\ell=1,\ldots,n}$, can be evaluated in term of the expected quadratic risk $R_\lambda(\hat{f}^\lambda, f)$ defined by

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

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 $V_\lambda(t) = \|S_\lambda(t)\|^2_{\mathbb{R}^n}$, 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 4.5. *There exist a constant $\kappa(\mathcal{F}) > 0$ and a real-valued functions $\lambda \longmapsto B(\lambda)$, such that for all $f \in \mathcal{F}$, $\|B_\lambda(f, \cdot)\|^2_{L^2} = \|\langle S_\lambda(\cdot), \mathbf{f} \rangle - f(\cdot)\|^2_{L^2} \leq \kappa(\mathcal{F}) B(\lambda)$. 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 4.5 in the case of the randomly shifted curves model (1.2) for which $\Omega = [0, 1]$. Assume that the design points are equi-spaced and that the smoothing step is obtained by low-pass Fourier filtering. Following Section 3, take $\mathcal{F} = H_s(A)$ defined in (3.4). In this setting, $V(\lambda) = \frac{2\lambda+1}{n}$ and $\|B_\lambda(f, \cdot)\|^2_{L^2} = \sum_{|k| \leq \lambda} |\tilde{c}_k(f) - c_k(f)|^2 + \sum_{|k| > \lambda} |c_k(f)|^2$ where $\tilde{c}_k(f) = \frac{1}{n} \sum_{\ell=1}^n f(\frac{\ell}{n}) e^{-i2\pi k \frac{\ell}{n}}$. Therefore, $\|B_\lambda(f, \cdot)\|^2_{L^2} \leq C(A) B(\lambda)$ for some positive constant $C(A)$ depending only on A, and $B(\lambda) = \frac{2\lambda+1}{n} + \lambda^{-2s}$. Thus, Assumption 4.5 holds with $\lambda_n = n^{\frac{1}{2s+1}}$.

4.3 Identifiability conditions

Let us now make the following assumptions on the set Θ and the density g of the random deformation parameters:

Assumption 4.6. *The support of the density g is included in Θ which is a compact set of $\mathbb{R}^p$ of the form*

$$\Theta = \left\{ \boldsymbol{\theta} = (\theta^1, \dots, \theta^p) \in \mathbb{R}^p, |\theta^k| \leq \rho_k, 1 \leq k \leq p \right\} \tag{4.1}$$

where $\rho_k > 0, k = 1, \dots, p$, are positive constants.

In the case of randomly shifted curves (1.2) one has $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}, |\boldsymbol{\theta}| \leq \rho_1\}$ with $0 < \rho_1 < 1$. Recall that it has been discussed in Section 3 that without any further assumptions on the set Θ^J and the density g of the random shifts, the model (1.2) is not identifiable. In the case of general deformation operators, choosing identifiability conditions in model (1.5) amounts to impose constraints on the minimization of the following criterion $\boldsymbol{\theta} \longmapsto D(\boldsymbol{\theta})$ defined for all $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \Theta^J$ by

$$D(\boldsymbol{\theta}) = \frac{1}{J} \sum_{j=1}^{J} \int_{\Omega} \left(\tilde{T}_{\boldsymbol{\theta}_j} T_{\boldsymbol{\theta}_j^*} f^*(t) - \frac{1}{J} \sum_{j'=1}^{J} \tilde{T}_{\boldsymbol{\theta}_{j'}} T_{\boldsymbol{\theta}_{j'}^*} f^*(t) \right)^2 dt. \tag{4.2}$$

This criterion $D(\boldsymbol{\theta})$ can be viewed as the deterministic version of $M(\boldsymbol{\theta})$ defined in (2.5). Obviously, using that $\tilde{T}_{\boldsymbol{\theta}} T_{\boldsymbol{\theta}} f^* = f^*$, the criterion $D(\boldsymbol{\theta})$ has a minimum at $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_J^*)$ such that $D(\boldsymbol{\theta}^*) = 0$. However, without any further restrictions the minimizer of $D(\boldsymbol{\theta})$ is not necessarily unique on Θ^J, and minimizing $M(\boldsymbol{\theta})$ does not allow to recover the true parameters of deformation $\boldsymbol{\theta}^*$ nor the true function f^*. Let us now explain how to define a constraint set $\boldsymbol{\Theta} \subset \Theta^J$ to impose identifiability conditions.

The case of randomly shifted curves

In the case of shifted curves, $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}, |\boldsymbol{\theta}| \leq \rho_1\}$ and for any $\boldsymbol{\theta}_0 \in \Theta$, we have $D(\boldsymbol{\theta}_1^* + \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_J^* + \boldsymbol{\theta}_0) = 0$. In other terms, the criterion D vanishes along the line $\{\boldsymbol{\theta}^* + \boldsymbol{\theta}_0 \mathbb{1}_J,\ \boldsymbol{\theta}_0 \in \mathbb{R}\} \subset \mathbb{R}^J$, where $\mathbb{1}_J = (1, \dots, 1)' \in \mathbb{R}^J$. Identifiability conditions in models such as (1.5) are discussed in [GLM07], [Vim10], [BG10] for shifted curves models. Recall that for shifted curves, it is shown in [BG10] that if f^* is such that $\int_0^1 f^*(t) e^{-i2\pi t} dt \neq 0$ and $\boldsymbol{\Theta}_0 = \{(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 + \ldots + \boldsymbol{\theta}_J = 0\}$ with $\Theta \subset [-\rho_1,\ \rho_1]$ and $0 < \rho_1 < \frac{1}{4}$, then the criterion $D(\boldsymbol{\theta})$ has a unique minimum on $\boldsymbol{\Theta}_0$ at $\boldsymbol{\theta}_{\boldsymbol{\Theta}_0}^* = (\boldsymbol{\theta}_1^* - \bar{\boldsymbol{\theta}}^*, \dots, \boldsymbol{\theta}_J^* - \bar{\boldsymbol{\theta}}^*)$ where $\bar{\boldsymbol{\theta}}^* = \frac{1}{J} \sum_{j=1}^{J} \boldsymbol{\theta}_j^*$. This amounts to add a linear constraint in the optimization of $D(\boldsymbol{\theta})$ by choosing $\boldsymbol{\Theta}_0 = \mathbb{1}_J{}^{\perp}$ (the orthogonal of $\mathbb{1}_J$ in $\mathbb{R}^J$). As shown in Section 3, this choice is well suited if it is assumed that the density g of the random shifts has zero mean. More generally, if the deformation parameters $\boldsymbol{\theta}_j, j = 1, \dots, J$ 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

$$\boldsymbol{\Theta}_1 = \{(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 = 0\}. \tag{4.3}$$

This is again a linear constraint by choosing $\boldsymbol{\Theta}_1 = e_1{}^{\perp}$ where $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^J$. If one supposes that $\boldsymbol{\theta}_1^* = 0$ and $\int_0^1 f^*(t) e^{-i2\pi t} dt \neq 0$, then it can be easily shown that the criterion $D(\boldsymbol{\theta})$ has a unique minimum on $\boldsymbol{\Theta}_1$ at $\boldsymbol{\theta}_{\boldsymbol{\Theta}_1} = (0, \boldsymbol{\theta}_2^*, \dots, \boldsymbol{\theta}_J^*)$. 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. A graphical illustration of such choices of identifiability conditions is given in Figure 1.

The case of a Lie group action

Let us now consider the case where the deformation operators possess a Lie group structure. Recall that it means that $T_{\boldsymbol{\theta}} f(t) = f\left(\exp(\boldsymbol{\theta}) \cdot t\right)$ where Θ denotes a compact subset of the Lie algebra $\mathcal{G}$ of

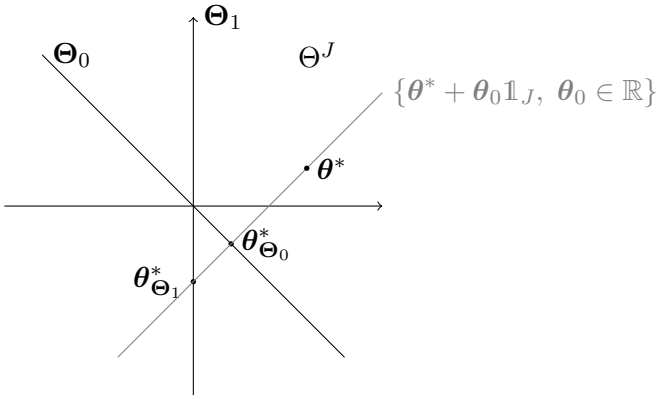


Figure 1: Choice of identifiability conditions for shifted curves in the case $J = 2$.

a connected Lie group G of dimension p acting on Ω, and $\exp : \mathcal{G} \to G$ denotes the exponential map (see [Hel01] for an introduction to Lie groups). In this framework, the shifted curve model (1.2) is a particular case of a commutative group with $G = (S^1, +)$ the torus in dimension one, $\mathcal{G} = \mathbb{R}$ with vanishing Lie brackets, and $\exp(\boldsymbol{\theta}) = \boldsymbol{\theta} \mod 1$. In this case the deformation operators satisfy the commutative relations

$$\tilde{T}_{\boldsymbol{\theta}_1} T_{\boldsymbol{\theta}_2} = T_{\boldsymbol{\theta}_2} \tilde{T}_{\boldsymbol{\theta}_1} = T_{\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1}. \tag{4.4}$$

If G is a commutative group then the deformation operators $T_{\boldsymbol{\theta}} f(t) = f\left(\exp(\boldsymbol{\theta}) \cdot t\right)$ and $\tilde{T}_{\boldsymbol{\theta}} f(t) = f\left(\exp(-\boldsymbol{\theta}) \cdot t\right)$ will also satisfy the relations (4.4). Using equation (4.4), one can see that the criterion D vanishes along the affine subspace $\{\boldsymbol{\theta}^* + \mathbb{1}_J \otimes \boldsymbol{\theta}_0,\ \boldsymbol{\theta}_0 \in \mathbb{R}^p\} \subset \mathbb{R}^{pJ}$ where $\otimes$ is the tensor product. In this situation, using the arguments given in the case of shifted curves, identifiability conditions can be given by restricting the optimization of $D(\boldsymbol{\theta})$ over the subsets $\boldsymbol{\Theta}_0 = \{(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 + \ldots + \boldsymbol{\theta}_J = 0\}$ or $\boldsymbol{\Theta}_1 = \{(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 = 0\}$.

When the group G is not commutative, then choosing identifiability conditions is more problematic. As an example consider the case (1.3) of rigid deformation of two dimensional images, for which there exists again a subset of Θ^J onto which $D(\boldsymbol{\theta})$ is 0. However, this subset is no longer a line. To see this, recall the notations of model (1.3). We have $\boldsymbol{\theta}_i = (a_i, \alpha_i, b_i)$ for $i = 1, 2$ and let $\boldsymbol{\theta}_1 \star \boldsymbol{\theta}_2 = (a_1 + a_2, \alpha_1 + \alpha_2, e^{a_2} R_{\alpha_2} b_1 + b_2)$ and $\boldsymbol{\theta}_2 \star \boldsymbol{\theta}_1 = (a_1 + a_2, \alpha_1 + \alpha_2, e^{a_1} R_{\alpha_1} b_2 + b_1)$. Then, contrary to relations (4.4) in the commutative case, one has that

$$T_{\boldsymbol{\theta}_1} T_{\boldsymbol{\theta}_2} = T_{\boldsymbol{\theta}_1 \star \boldsymbol{\theta}_2} \neq T_{\boldsymbol{\theta}_2 \star \boldsymbol{\theta}_1} = T_{\boldsymbol{\theta}_2} T_{\boldsymbol{\theta}_2} \tag{4.5}$$

In this example, for $\boldsymbol{\theta}_0 = (a_0, \alpha_0, b_0)$, we have that $D(\boldsymbol{\theta}_0 \star \boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_0 \star \boldsymbol{\theta}_J^*) = 0$. In other terms, the set

$$\Theta * \boldsymbol{\theta}^* = \{(\boldsymbol{\theta}_0 \star \boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_0 \star \boldsymbol{\theta}_J^*),\ \boldsymbol{\theta}_0 \in \Theta\} \subset \Theta^J$$

onto which D vanishes is no longer a line in Θ^J. A graphical illustration of this fact is displayed in Figure 2. In this case it is always possible to choose, says, the first observation as a reference image, and match all the other images onto it. It amounts to optimize $D(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}_1$.

The case of general deformation operators

Let us now discuss the more general case. Conditions to guarantee the uniqueness of a minimum of D over a restricted set $\boldsymbol{\Theta} \subset \Theta^J$ can also be given. First, recall that the criterion D has a minimum at $\boldsymbol{\theta}^*$ such that $D(\boldsymbol{\theta}^*) = 0$. If $\boldsymbol{\theta} \in \Theta^J$ with $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$ is such that $D(\boldsymbol{\theta}) = 0$ then (by definition of D)

$$\tilde{T}_{\boldsymbol{\theta}_j} T_{\boldsymbol{\theta}_j^*} f^* = \tilde{T}_{\boldsymbol{\theta}_1} T_{\boldsymbol{\theta}_1^*} f^* \text{ for all } 1 \leq j \leq J. \tag{4.6}$$

Now, let $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, and suppose that the function f^* is such that: if $\tilde{T}_{\boldsymbol{\theta}_2} T_{\boldsymbol{\theta}_1} f^* = f^*$ then necessarily $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. Such an assumption implies, among other things, that the function f^* cannot be invariant

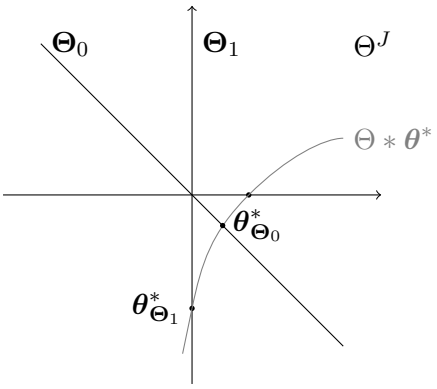


Figure 2: Choice of identifiability conditions for rigid deformation of images when $J = 2$.

by the action of the deformation operators. Under such a condition and if one further assumes that $\boldsymbol{\theta}^*_1 = 0$, it can be checked that equation (4.6) implies that the criterion $D(\boldsymbol{\theta})$ has a unique minimum on $\mathbf{\Theta}_1 = \{(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \Theta^J,\ \boldsymbol{\theta}_1 = 0\}$ at $\boldsymbol{\theta}_1 = (0, \boldsymbol{\theta}^*_2, \dots, \boldsymbol{\theta}^*_J)$. Deriving other identifiability conditions depends on the specific form of the deformation operators. Nevertheless, in what follows, we propose to use general assumptions that will be shown to guaranty identifiability in the model (1.5).

A pre-constrained optimization problem

In what follows, we will suppose that there exists a linear subset $\mathbf{\Theta}$ of Θ^J onto which D has a unique minimum. Throughout the paper we denote by A' the transpose of a matrix $A \in \mathbb{R}^{k \times k'}$, $k, k' \in \mathbb{N}$. Let us first define the set $\mathbf{\Theta}$.

Definition 4.2. *Let $I \leq J$ be a positive integer and $\mathcal{T}$ a compact subset of $\mathbb{R}^{pI}$. Let $\phi : \mathbb{R}^{pI} \longrightarrow \mathbb{R}^{pJ}$ be an orthogonal linear mapping (i.e. $\phi'\phi = Id_{pI}$ where Id_{pI} is the $pI \times pI$ identity matrix). Then,*

$$\mathbf{\Theta} := \phi\mathcal{T} = \{\phi\tau\ ;\ \tau \in \mathcal{T}\}\,.$$

Using the above definition, our first identifiability condition is the following assumption:

Assumption 4.7. *The criterion D has a unique minimum on $\mathbf{\Theta}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*_{\mathbf{\Theta}}$ that is $\boldsymbol{\theta}^*_{\mathbf{\Theta}} = argmin_{\boldsymbol{\theta} \in \mathbf{\Theta}} D(\boldsymbol{\theta})$. Equivalently, it means that there exist a unique $\boldsymbol{\tau}^*_{\mathbf{\Theta}} \in \mathcal{T}$ such that $\boldsymbol{\tau}^*_{\mathbf{\Theta}} = \underset{\boldsymbol{\tau} \in \mathcal{T}}{argmin}\, \tilde{D}(\tau)$, where $\tilde{D}(\tau) := D(\phi\tau)$ and $\boldsymbol{\theta}^*_{\mathbf{\Theta}} = \phi\boldsymbol{\tau}^*_{\mathbf{\Theta}}$.*

In the examples discussed above $I = J - 1$. For the constrained set $\mathbf{\Theta}_1 = {e_1}^\perp$ a possible choice for ϕ is

$$\begin{array}{rcl} \phi_1 : \mathcal{T}_1 \subset \mathbb{R}^{p(J-1)} & \longrightarrow & \mathbf{\Theta}_1 \subset \mathbb{R}^{pJ} \\ \boldsymbol{\tau} = (\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_{J-1}) & \longmapsto & \boldsymbol{\theta} = (0, \boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_{J-1}). \end{array}$$

with $\mathcal{T}_1 = \Theta^{J-1}$ (remark that $\phi_1'\phi_1 = Id_{p(J-1)}$). Recall that in the case of shifted curves, if one supposes that $\boldsymbol{\theta}^*_1 = 0$ and $\int_0^1 f^*(t)e^{-i2\pi t}dt \neq 0$, then Assumption 4.7 is satisfied with $\boldsymbol{\theta}^*_{\mathbf{\Theta}} = (0, \boldsymbol{\theta}^*_2, \dots, \boldsymbol{\theta}^*_J)$. For more general deformation operators, Assumption 4.7 is satisfied when using the mapping ϕ_1 and the condition that $\tilde{T}_{\boldsymbol{\theta}_2} T_{\boldsymbol{\theta}_1} f^* = f^*$ implies that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$.

We detail also the case of the constraint $\mathbf{\Theta}_0$ (3.7) for $p = 1$. Define,

$$\begin{array}{rcl} \phi_0 : \mathcal{T}_0 \subset \mathbb{R}^{J-1} & \longrightarrow & \mathbf{\Theta}_0 \subset \mathbb{R}^J \\ \boldsymbol{\tau} = (\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_{J-1}) & \longmapsto & H\boldsymbol{\tau} \end{array}$$

where $H \in J \times (J-1)$ is a so-called Helmert sub-matrix such that the j-th column of H, for $1 \leq j \leq J-1$, is given by $(\underbrace{h_j, \ldots, h_j}_{j \text{ times}}, -jh_j, \underbrace{0, \ldots, 0}_{J-j-1 \text{ times}})'$, where $h_j = (j(j+1))^{-\frac{1}{2}}$. It simply means that the column of H define an orthonormal basis of $\mathbb{1}_J{}^{\perp}$. For more details on Helmert matrices in shape analysis, we refer to Definition 2.5 in [DM98]. By construction one has $H'H = Id_{J-1}$ and $\phi \mathcal{T}_0 = \mathbf{\Theta}_0$. Thus, in the case of shifted curves, if one supposes that $\int_0^1 f^*(t)e^{-i2\pi t}dt \neq 0$ then Assumption 4.7 is satisfied with $\boldsymbol{\theta}^*_{\mathbf{\Theta}_0} = (\boldsymbol{\theta}^*_1 - \bar{\boldsymbol{\theta}}^*, \ldots, \boldsymbol{\theta}^*_J - \bar{\boldsymbol{\theta}}^*)$.

Uniqueness of global minimum

It is important to control the Hessian of the criterion D on the the compact set $\mathcal{T} = \phi^{-1}\mathbf{\Theta}$ around its minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}^*_{\mathbf{\Theta}}$. Let us suppose that Assumption 4.7 is satisfied and introduce the following definition:

Definition 4.3. *Let $\gamma_{\min}(J, f^*, \boldsymbol{\theta}^*, \mathbf{\Theta})$ be the smallest eigenvalue of the $pI \times pI$ matrix $\phi'[\nabla^2 D(\boldsymbol{\theta}^*_{\mathbf{\Theta}})]\phi$ that is*

$$\gamma_{\min}(J, f^*, \boldsymbol{\theta}^*, \mathbf{\Theta}) = \min_{\boldsymbol{\tau} \in \mathbb{R}^{pI}} \frac{\boldsymbol{\tau}'\phi'[\nabla^2 D(\boldsymbol{\theta}^*_{\mathbf{\Theta}})]\phi\boldsymbol{\tau}}{\boldsymbol{\tau}'\boldsymbol{\tau}} = \min_{\boldsymbol{\tau} \in \mathbb{R}^{pI}} \frac{\boldsymbol{\tau}'[\nabla^2 \tilde{D}(\boldsymbol{\tau}^*_{\mathbf{\Theta}})]\boldsymbol{\tau}}{\boldsymbol{\tau}'\boldsymbol{\tau}},$$

where $\nabla^2 D$ denotes the Hessian of D and $\nabla^2 \tilde{D}$ denotes the Hessian of $\tilde{D}$.

We give here the expression of $\nabla^2 D$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*_{\mathbf{\Theta}}$. The Hessian of D can be decomposed into a sum of a block diagonal and a block symmetric $pJ \times pJ$ matrix,

$$\nabla^2 D(\boldsymbol{\theta}^*_{\mathbf{\Theta}}) = \frac{2}{J^2}\Big(J \operatorname*{Diag}_{j_1=1,\ldots,J}(A_{j_1,j_1}) - \big[A_{j_1,j_2}\big]_{j_1,j_2=1}^{J}\Big),$$

where $A_{j_1,j_2} = \left[\left\langle \partial_{\theta_{j_1}^{p_1}} f^*_{\boldsymbol{\theta}^*_{j_1}}, \partial_{\theta_{j_2}^{p_2}} f^*_{\boldsymbol{\theta}^*_{j_2}} \right\rangle_{L^2}\right]_{p_1,p_2=1}^{p} \in \mathbb{R}^{p\times p}$ and $\partial_{\theta_{j_1}^{p_1}} f^*_{\boldsymbol{\theta}^*_{j_1}} = \partial_{\theta_{j_1}^{p_1}} \big[\tilde{T}_{\boldsymbol{\theta}} T_{\boldsymbol{\theta}^*_{j_1}} f^*\big]_{|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*_{\mathbf{\Theta}}}}$. The Diag notation stands for the block diagonal matrix. For the general expression of the gradient and the Hessian of D at $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*_{\mathbf{\Theta}}$, refer to the Appendix C.

Finally, the minimum of D on $\mathbf{\Theta}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*_{\mathbf{\Theta}}$ should be sufficiently well separated from other local minima of D in the sense of the following assumption:

Assumption 4.8. *There is a constant $C_1(\Theta, f^*) > 0$ (not depending on J) such that for any $\boldsymbol{\theta}^* \in \Theta^J$*

$$J\gamma_{\min}(J, f^*, \boldsymbol{\theta}^*, \mathbf{\Theta}) \geq C_1(\Theta, f^*). \tag{4.7}$$

For any $\delta > 0$, there exists a constant $C_2(\Theta, f^*, \delta) > 0$ (not depending on J) such that for any $\boldsymbol{\theta}^* \in \Theta^J$

$$\min_{\boldsymbol{\theta} \in B^c(\boldsymbol{\theta}^*_{\mathbf{\Theta}}, \delta)} D(\boldsymbol{\theta}) - D(\boldsymbol{\theta}^*_{\mathbf{\Theta}}) \geq C_2(\Theta, f^*, \delta), \tag{4.8}$$

where $B^c(\boldsymbol{\theta}^*_{\mathbf{\Theta}}, \delta)$ denotes the complementary of the set

$$B(\boldsymbol{\theta}^*_{\mathbf{\Theta}}, \delta) = \left\{\boldsymbol{\theta} \in \mathbf{\Theta},\ \left|\boldsymbol{\theta}_j^{p_1} - [\boldsymbol{\theta}^*_{\mathbf{\Theta}}]_j^{p_1}\right| \leq \delta,\ j = 1, \ldots, J,\ p_1 = 1, \ldots, p\right\}$$

Let us discuss the above assumptions. The first assertion (4.7) of Assumption 4.8 is used in the proofs to derive a lower quadratic bound in a neighborhood of $\boldsymbol{\theta}^*_{\mathbf{\Theta}}$ of the criterion $D(\boldsymbol{\theta})$ of the form $D(\boldsymbol{\theta}) - D(\boldsymbol{\theta}^*_{\mathbf{\Theta}}) \geq C(\Theta, \mathcal{F}, f^*)\frac{1}{J}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*_{\mathbf{\Theta}}\|^2$. In the case of randomly shifted curves, if $\mathbf{\Theta} = \mathbf{\Theta}_0$ or $\mathbf{\Theta} = \mathbf{\Theta}_1$, we have that $\nabla^2 D(\boldsymbol{\theta}^*_{\mathbf{\Theta}}) = \frac{2}{J^2}\|\nabla f^*\|^2_{L^2}(JId_J - \mathbb{1}_{J\times J})$. for any value of $\boldsymbol{\theta}^*$, where ∇f^* denotes the first derivative of f^*. The eigenvalues of $\nabla^2 D(\boldsymbol{\theta}^*_{\mathbf{\Theta}})$ are 0 and $\frac{2}{J}\|\nabla f^*\|^2_{L^2}$ with multiplicity 1 and $J-1$ respectively. Thus, for $\mathbf{\Theta} = \mathbf{\Theta}_0$ $\nabla^2 \tilde{D}(\boldsymbol{\tau}^*_{\mathbf{\Theta}_0}) = \phi_0' \nabla^2 D(\boldsymbol{\theta}^*_{\mathbf{\Theta}_0})\phi_0 = \frac{2}{J}\|\nabla f^*\|^2_{L^2} Id_{J-1}$ and

$$J\gamma_{\min}(J, f^*, \boldsymbol{\theta}^*, \mathbf{\Theta}_0) = 2\|\nabla f^*\|^2_{L^2},$$

which shows that the first assertion of Assumption 4.8 is satisfied. If we choose a constraint of the type $\mathbf{\Theta}_1 = \{\boldsymbol{\theta} \in \Theta,\ \boldsymbol{\theta}_1 = 0\}$, we have that $\phi_1' \nabla^2 D(\boldsymbol{\theta}^*_{\mathbf{\Theta}_1}) \phi_1 = \frac{2}{J^2} \|\nabla f^*\|^2_{L^2} (JId_{J-1} - \mathbb{1}_{(J-1)\times(J-1)})$. Therefore, $J\gamma_{\min}(J, f^*, \boldsymbol{\theta}^*, \mathbf{\Theta}_1) = \frac{2}{J} \|\nabla f^*\|^2_{L^2}$, and Assumption 4.8 is not satisfied. Hence, the choice of $\mathbf{\Theta} = \mathbf{\Theta}_0$ improves by a factor J the conditioning of the Hessian of the criterion and leads to a constrained set satisfying the first assertion of Assumption 4.8.

The second assertion (4.8) of Assumption 4.8 guarantees that the local minima of $D(\boldsymbol{\theta})$ are well separated from its global minimum which is a standard assumption in M-estimaton (see e.g [vdV98]). We have discussed such an assumption in the case of shifted curves in Section 3 with $\mathbf{\Theta} = \mathbf{\Theta}_0$, and Proposition 3.1 gives simple conditions to verify it. However, deriving simple conditions such that assertion (4.8) is satisfied in the case of more general operators depends on the problem of interest.

4.4 Random perturbation of the mean pattern f^* by the Z_j's

Let us recall that the Z_j's in model (1.5) are independent realizations of a second order Gaussian process Z taking its values in $L^2(\Omega)$. These random processes model the linear variations in intensity of the curves/images around the mean pattern f^*. However, there are observed through the action of the random deformation operators $T_{\boldsymbol{\theta}_j}, j = 1, \dots, J$. Therefore, to study the consistency of the estimators, it is important to specify how the action of $T_{\boldsymbol{\theta}_j}$ modifies the law of the process Z_j. In particular, the action of the deformation operator $T_{\boldsymbol{\theta}_j} Z_j(t_\ell)$ and $T_{\boldsymbol{\theta}_j} Z_j(t_{\ell'})$ modifies the structure of the correlations between $Z_j(t_\ell)$ and $Z_j(t_{\ell'})$ for $\ell \neq \ell'$. It is thus important to give conditions that guarantee that the smoothing step control these correlations. For this purpose, let us make the following assumption on the Gaussian process Z:

Assumption 4.9. *The process Z is a second order Gaussian process taking its values in $L^2(\Omega)$ with zero mean. 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,

$$\lim_{n\to\infty} \gamma_n(\Theta) \sqrt{V(\lambda_n)} = 0, \tag{4.9}$$

where $V(\lambda_n)$ is the variance defined in Assumption 4.5.

Intuitively, the condition (4.9) 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 $\ell, \ell' = 1, \dots, n$ and all $\boldsymbol{\theta} \in \Theta$. In the case of randomly shifted curves with an equi-spaced design, a simple condition for which Assumption 4.9 holds is the case where Z is stationary process. Recall that in Section 3, we have introduced the covariance matrix $\mathbf{\Sigma_n} = \big[\mathbb{E}[Z(t_\ell)Z(t_{\ell'})]\big]_{\ell,\ell'=1}^n = \big[R(|t_\ell - t_{\ell'}|)\big]_{\ell,\ell'=1}^n$, see Assumption 3.1. If the design is equi-spaced, then the first assertion of Assumption 4.9 holds with $\gamma_n(\Theta) = \gamma = \int_0^1 |R(t)|\, dt$. Thus, $\lim_{n\to\infty} \gamma_n \sqrt{V(\lambda_n)} = \lim_{n\to\infty} \sqrt{V(\lambda_n)} = 0$, which proves that Assumption 4.9 is satisfied.

5 Consistency in the general case

5.1 Consistent estimation of the deformation parameters

In this section, we use the notations introduced in Section 4, and we assume that identifiability conditions have been chosen. With the notations introduced in Section 4.3, consider for $\lambda \in \Lambda$ the following estimator of the deformation parameters

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

where

$$M_\lambda(\boldsymbol{\theta}) = \frac{1}{J}\sum_{j=1}^{J}\int_\Omega \left(\tilde{T}_{\boldsymbol{\theta}_j}\langle S_\lambda(t), \mathbf{Y}_j\rangle - \frac{1}{J}\sum_{j'=1}^{J}\tilde{T}_{\boldsymbol{\theta}_{j'}}\langle S_\lambda(t), \mathbf{Y}_{j'}\rangle\right)^2 dt, \tag{5.1}$$

and $\boldsymbol{\Theta}$ is the constrained set introduced in Definition 4.2. 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 4.7. 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 (3.7) and if the density g of the shifts has zero mean, then $\boldsymbol{\theta}_{\boldsymbol{\Theta}_0} = (\boldsymbol{\theta}^*_1 - \bar{\boldsymbol{\theta}}^*, \dots, \boldsymbol{\theta}^*_J - \bar{\boldsymbol{\theta}}^*)$ with $\bar{\boldsymbol{\theta}}^* = \frac{1}{J}\sum_{j=1}^J \boldsymbol{\theta}^*_j$ can be shown to be close to $\boldsymbol{\theta}^*$ (see Lemma A.1 in the Appendix). This allows to show the consistency of $\hat{\boldsymbol{\theta}}^\lambda$ to $\boldsymbol{\theta}^*$ as formulated in Theorem 3.1. Therefore, the next result only bounds the distance between $\hat{\boldsymbol{\theta}}^\lambda$ and $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$.

Theorem 5.1. *Consider the model* (1.5) *and suppose that Assumptions 4.1 to 4.9 hold with* $n \geq 1$ *and* $J \geq 2$. *Then, for any* $\lambda \in \Lambda$ *and* $x > 0$

$$\mathbb{P}\left(\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\|^2 \geq C_1(\Theta, \mathcal{F}, f^*)\Big[(\gamma_n(\Theta) + \sigma^2)\Big(\sqrt{v(x,J,\lambda)} + v(x,J,\lambda)\Big) + \Big(\sqrt{B(\lambda)} + B(\lambda)\Big)\Big]\right) \leq 2e^{-x}, \tag{5.2}$$

with $C_1(\Theta, \mathcal{F}, f^*) > 0$, $v(x,J,\lambda) := V(\lambda)\left(1 + 4\frac{x}{J} + \sqrt{4\frac{x}{J}}\right)$ *and where* $\|.\|$ *is the Euclidean norm in* $\mathbb{R}^{pJ}$.

Let us comment the inequality (5.2). The term $v(x,J,\lambda)$ depends on the variance $V(\lambda)$ of the linear estimators $\hat{f}^\lambda_j$, $j = 1, \dots, J$, and recall that $B(\lambda)$ is a uniform control on the bias of these estimators (see Section 4.2). Using a classical tradeoff between variance and bias, Assumption 4.5 implies that there exists a sequence of smoothing parameters λ_n such that $\lim_{n\to+\infty} V(\lambda_n) = \lim_{n\to+\infty} B(\lambda_n) = 0$. Moreover, using Assumption 4.9, it follows that $\lim_{n\to+\infty} \gamma_n(\Theta)\Big(\sqrt{v(x,J,\lambda_n)} + v(x,J,\lambda_n)\Big) = 0$ for any $x > 0$ and $J \geq 2$. If the number of curves J remains fixed, Theorem 5.1 implies that $\hat{\boldsymbol{\theta}}^\lambda$ converges in probability to $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$ as the number of design points n increases. Such an an asymptotic in n has been considered in [GLM07] and in [Vim10] to estimate deterministic shifts in a SIM model using a semiparametric framework.

To the contrary let us fix the number of design points n, and consider an asymptotic setting where the number J of curves/images is let going to infinity. For any $x > 0$ and $\lambda \in \Lambda$, $\lim_{J\to+\infty} v(x,J,\lambda) = V(\lambda)$. Therefore, Theorem 5.1 cannot be used to prove that $\hat{\boldsymbol{\theta}}^\lambda$ converges to $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$ as J tends to infinity. To the contrary one has that

$$\lim_{J\to+\infty}(\gamma_n(\Theta) + \sigma^2)\Big(\sqrt{v(x,J,\lambda)} + v(x,J,\lambda)\Big) + \Big(\sqrt{B(\lambda)} + B(\lambda)\Big) = (\gamma_n(\Theta) + \sigma^2)h(V(\lambda)) + h(B(\lambda))$$

where $h(u) = \sqrt{u} + u$ for $u \geq 0$. This would suggest 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. This fact has already been noticed in Section 3, since the lower bound given in Theorem 3.2 shows that $\hat{\boldsymbol{\theta}}^\lambda$ is not a consistent estimator of $\boldsymbol{\theta}^*$ as $J \to +\infty$ when the process Z is stationary. The following theorem shows that a similar phenomenon arises under the following assumption on the process Z,

Assumption 5.1. *The process* Z *is a second order Gaussian process taking its values in* $L^2(\Omega)$ *with zero mean. Moreover, there exists a positive semi-definite symmetric* $n \times n$ *matrix* $\Sigma(\Theta)$ *such that for any random vector* $\boldsymbol{\theta} \in \Theta$ *the covariance matrix of* $\mathbf{Z} = [Z(t_\ell)]_{\ell=1}^n$ *satisfies* $\mathbb{E}_{\boldsymbol{\theta}}\big[\mathbf{T}_{\boldsymbol{\theta}}\mathbf{Z}(\mathbf{T}_{\boldsymbol{\theta}}\mathbf{Z})'\big] = \boldsymbol{\Sigma}(\Theta)$.

This assumption means that the law of the random process Z is invariant by the deformation operators $T_{\boldsymbol{\theta}}$ for $\boldsymbol{\theta} \in \Theta$.

Theorem 5.2. *Consider the model* (1.5)*. Suppose that Assumptions 4.1, 4.2, 4.3 , 4.6 and 5.1 hold. Assume the density g is continuously differentiable, that its support strictly include in Θ, and that $\int_\Theta \left(\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta})\right)\right)^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta} < +\infty$. Let $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{pJ}$ be any estimator (a measurable function of the data) of $\boldsymbol{\theta}^*$. Then,*

$$\mathbb{E}\left[\frac{1}{J}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2\right] \geq \frac{n^{-1}(\sigma^2 + \tilde{\gamma}_n(\Theta))}{C(\Theta, \mathcal{F}, f^*) + n^{-1}(\sigma^2 + \tilde{\gamma}_n(\Theta)) \int_\Theta \left(\partial_{\boldsymbol{\theta}} \log\left(g(\boldsymbol{\theta})\right)\right)^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{5.3}$$

where $C(\Theta, \mathcal{F}, f^*)$ is a strictly positive constant and $\tilde{\gamma}_n(\Theta)$ denotes the largest eigenvalue of $\Sigma(\Theta)$.

The right hand side of the inequality (5.3) gives a lower bound on $\mathbb{E}[\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\|^2]$. This bound is independent of J, and thus if the number of design points n is fixed, increasing the number of curves/images does not improve the quality of the estimation of the deformation parameters for any estimators $\hat{\boldsymbol{\theta}}$. Nevertheless, this lower bound is going to 0 as $n \to +\infty$ (provided that $\lim_{n\to+\infty} n^{-1}\tilde{\gamma}_n(\Theta)$) which is consistent with the results given by Theorem 5.1.

5.2 Consistent estimation of the mean pattern

From equation (2.6), recall that an estimator $\hat{f}^\lambda$ of the mean pattern f^* is defined as

$$\hat{f}^\lambda = \frac{1}{J}\sum_{j=1}^J \tilde{T}_{\hat{\boldsymbol{\theta}}_j^\lambda} \hat{f}_j^\lambda = \frac{1}{J}\sum_{j=1}^J \tilde{T}_{\hat{\boldsymbol{\theta}}_j^\lambda} \left\langle S_\lambda(.), \mathbf{Y}_j \right\rangle. \tag{5.4}$$

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 (3.7), 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

$$f^*_{\boldsymbol{\Theta}_0}(t) := \frac{1}{J}\sum_{j=1}^J f^*(t - \boldsymbol{\theta}_j^* + [\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}]_j) = f^*(t - \bar{\boldsymbol{\theta}}^*) \text{ for all } t \in [0,1].$$

Hence, under the condition that $\int_\Theta \boldsymbol{\theta} g(\boldsymbol{\theta}) d\boldsymbol{\theta} = 0$ (the shifts have zero mean), 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 3.3.

Theorem 5.3. *Consider the model* (1.5) *and suppose that Assumptions 4.1 to 4.9 hold. Then, for any $\lambda \in \Lambda$ and $x > 0$*

$$\mathbb{P}\left(\|\hat{f}^\lambda - f^*_{\boldsymbol{\Theta}}\|^2_{L^2} \geq C_2(\Theta, \mathcal{F}, f^*)\Big[(\gamma_n(\Theta) + \sigma^2)\Big(\sqrt{\upsilon(x, J, \lambda)} + \upsilon(x, J, \lambda)\Big) + \Big(\sqrt{B(\lambda)} + B(\lambda)\Big)\Big]\right) \leq 2e^{-x}, \tag{5.5}$$

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

Similar comments that those made for the consistency of the deformation parameters can be given as the same terms appear in the deviation inequalities (5.2) and (5.5). The consistency of $\hat{f}^\lambda$ to $f^*_{\boldsymbol{\Theta}}$ is 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 5.3 cannot be used to prove the convergence of $\hat{f}^\lambda$ to $f^*_{\boldsymbol{\Theta}}$.

6 Numerical experiments

6.1 A general gradient descent algorithm

To compute the estimator $\hat{\boldsymbol{\theta}}^{\lambda}$ one has to minimize the criterion $M_\lambda(\boldsymbol{\theta})$ defined in (5.1). Generally, the expression of the gradient of $M_\lambda(\boldsymbol{\theta})$ is available in a closed form, and thus a gradient descent algorithm with an adaptive step can be easily implemented. More precisely the algorithm is composed of the following steps:

Initialization: let $\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}$ (if $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ a possible choice is $\boldsymbol{\theta}^0 = 0$), $\gamma_0 = \frac{1}{\|\nabla M_\lambda(\boldsymbol{\theta}^0)\|}$, $F(0) = M_\lambda(u^0)$, and set $m = 0$.

Step 2: let $\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^m - \gamma_m \nabla M_\lambda(\boldsymbol{\theta}^m)$ and $F(m+1) = M_\lambda(\boldsymbol{\theta}^{new})$.
While $F(m+1) > F(m)$ **do**
$\gamma_m = \gamma_m/\kappa$, and $\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^m - \gamma_m \nabla M_\lambda(\boldsymbol{\theta}^m)$, and let $F(m+1) = M_\lambda(\boldsymbol{\theta}^{new})$.
End while
Then, take $\boldsymbol{\theta}^{m+1} = \boldsymbol{\theta}^{new}$. Set $m = m+1$

Step 3: if $F(m) - F(m+1) \geq \rho(F(1) - F(m+1))$ then return to Step 2, else stop the iterations, and take $\hat{\boldsymbol{\theta}}^{\lambda} = \boldsymbol{\theta}^{m+1}$.

In the above algorithm, $\rho > 0$ is a small stopping parameter and $\kappa > 1$ is a parameter to control the choice of the adaptive step γ_m. In practice, one does not known the compact set Θ to which the $\boldsymbol{\theta}_j^*$ are supposed to belong. So it is difficult to constraint $\boldsymbol{\theta}^m$ to belong to a specific compact set. Nevertheless, depending on the choice of the constrained set $\boldsymbol{\Theta}$, it is possible to impose that $\boldsymbol{\theta}^m$ satisfies the same constraints. For instance, if $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ one can modify $\boldsymbol{\theta}^m = (\boldsymbol{\theta}_j^m)_{j=1,\dots,J}$ at each iteration such that $\sum_{j=1}^J \boldsymbol{\theta}_j^m = 0$.

6.2 Randomly shifted curves

We present some simulations to illustrate the results given in Section 3. The random shift framework was used as an example throughout this paper to illustrate the estimating procedure. In this setting, $d = 1$, $\Omega = [0,1]$ and $\mathcal{F} = H_3(A) \subset L^2_{per}([0,1])$. In the simulations, we took random shifts $\boldsymbol{\theta}_j$ having a uniform density g with compact support equal to $[-\frac{1}{5}, \frac{1}{5}]$.The model (3.1) is

$$Y_j^\ell = f^*(\tfrac{\ell}{n} - \boldsymbol{\theta}_j^*) + Z_j(\tfrac{\ell}{n} - \boldsymbol{\theta}_j^*) + \sigma\varepsilon_j^\ell, \quad \text{and } j = 1,\dots,J,\ \ell = 1,\dots,n \tag{6.1}$$

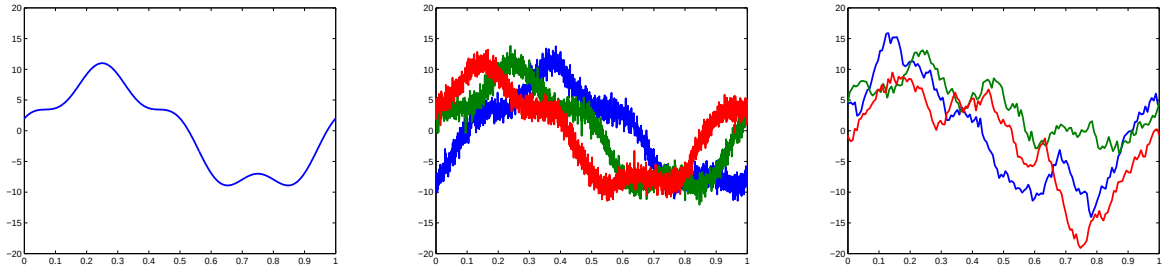
and we took $f^*(t) = 9\sin(2\pi t) + 2\cos(8\pi t)$ for $t \in [0,1]$ as a mean pattern, see Figure 3(a). For the constrained set we took

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

We consider the case where the linear smoothing is a Fourier low pass filter. In the following, we fix the spectral cut-off to $\lambda = 7$ which is reasonable value to reconstruct f^* representing a good tradeoff between bias and variance. Choosing λ could also be done by a cross-validation procedure. However, studying the theoretical properties of a data-based choice of λ is beyond the scope of the paper. In the following, 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, i.e without the addtive term Z. Figure 3(b) gives a sample of the data used with small level noise $\sigma = 2$. In the simulations, we took a larger level of noise $\sigma = 8$. In fact, the low pass filters efficiently smooth the white noise

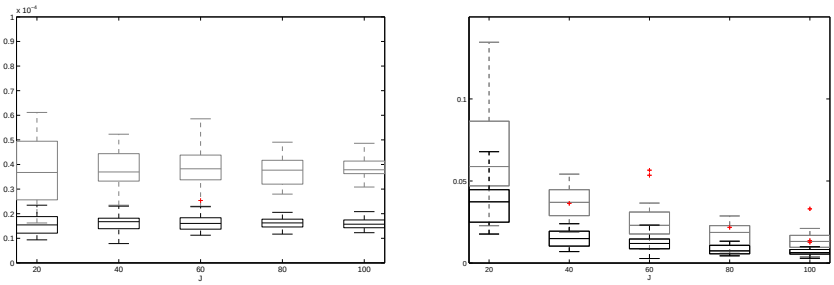


(a) (b) (c)

Figure 3: (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 $\varsigma = 4$.

and the results remain good even with a relatively large σ. 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 (6.1) with $Z_j = 0, j = 1, \ldots, J$. For each repetition we computed $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*\|^2$ and $\|\hat{f}^{\lambda} - f^*\|^2_{L^2}$. Boxplot of these quantities are displayed in Figure 4(a) and 4(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 $\hat{f}^{\lambda}_j$. Recall that the lower bound given in Theorem 3.2 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 4. Indeed, the quantity $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*\|^2$ does not become smaller when J grows, and it is noticeably smaller when n increases.

Similarly, the quantity $\|\hat{f}^{\lambda} - f^*\|^2_{L^2}$ is clearly smaller with $n = 1024$ than $n = 512$. One can also see that the variance and the mean of $\|\hat{f}^{\lambda} - f^*\|^2_{L^2}$ over the simulations tend to decrease as J grows but n remains fixed. This could be interpreted as a surprising fact since the quality of $\hat{f}^{\lambda}$ depends on the distance $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*\|^2$ which is not a decreasing function of J. However, note that the value of $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*\|^2$ is of order 10^{-4} for all values of n and J. This mean that the shifts are very well estimated and that $\hat{f}^{\lambda}$ is close to the ideal estimator $\tilde{f}^{\lambda}$ that would be obtained if we knew the true values of the shifts that is $\tilde{f}^{\lambda}(t) = \frac{1}{J}\sum_{j=1}^{J} \tilde{T}_{\boldsymbol{\theta}^*_j} \hat{f}^{\lambda}_j(t) = \frac{1}{J}\sum_{j=1}^{J} \hat{f}^{\lambda}_j(t + \boldsymbol{\theta}^*_j)$. It is clear from model (6.1) that $\tilde{f}^{\lambda}$ can be shown to be consistent estimator of f^* at the design points as J tends to infinity and n remains fixed, provided that $\lambda = \lambda_J \to +\infty$ is chosen in an appropriate way. This explains why, in numerical experiments, the quantity $\|\hat{f}^{\lambda} - f^*\|^2_{L^2}$ is decreasing with J.



(a) (b)

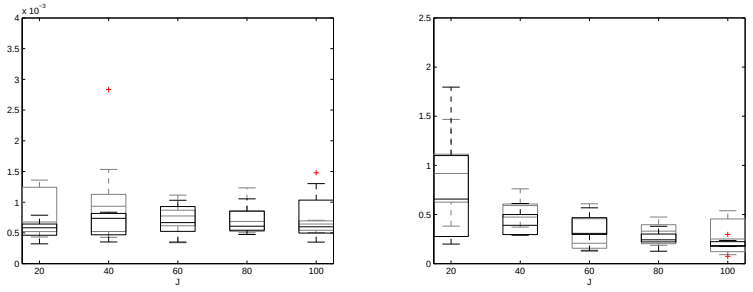
Figure 4: Boxplot of $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ (a) and $\|\hat{f}^{\lambda} - f^*_{\boldsymbol{\Theta}_0}\|^2_{L^2}$ (b) over $M = 20$ repetitions from a SIM model of shifted curves. Boxplot in gray correspond to $n = 512$, and in black to $n = 1024$.

Complete model

We now add the terms Z_j in (6.1) to model linear variations in amplitude of the curves around the template f^*. First, we generate a stationary periodic Gaussian process. To do this, the covariance matrix must be a particular Toeplitz matrix. As suggested in [Gre93] one possibility is to choose

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

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 3(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 5(a) and 5(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 5(a)) and $\|\hat{f}^{\lambda} - f^*_{\boldsymbol{\Theta}_0}\|^2$ has a smaller mean and variance as n increases. Nevertheless, increasing n does not really change a lot the results. Indeed, the level of noise ($\sigma^2 + \varsigma^2$) is rather high, and thus the low pass filter fails to smooth efficiently the low frequency part of the additive noise Z and ε.



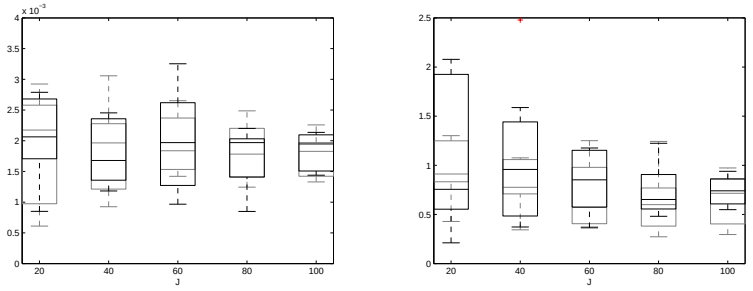
(a) (b)

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 (3.1) with a stationnary error term Z. Boxplot in gray correspond to $n = 512$, and in black to $n = 1024$.

We finally run the same simulations with a non stationary noise $Z_j(t) = \alpha_j \psi(t)$ where ψ is a positive periodic smooth deterministic function such that $\|\psi\|_{L^2} = 1$ and $\alpha_j \sim \mathcal{N}(0, \varsigma^2)$ with $\varsigma = 4$. Note that, in this case, the sequence $\gamma_n(\Theta)$ is of order n and Assumption 4.9 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 6(a) for $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^{\lambda} - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2$ and in Figure 6(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 6(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 6(b).

6.3 Geometric PCA of images

Let us consider images of handwritten digits taken from the Mnist database (see [LBBH98] for more details on this data set). For each digit, one observes a set of $J = 100$ graylevel images of size 28×28 pixels. Each image can thus be viewed as a noisy random function observed $n = 784$ points on a regular grid of $\Omega = [0, 1]^2$. We only report results for digit 1, 3 and 5. In Figure 7 to Figure 9, we display for each digit a sample of 8 images out of $J = 100$ to give an idea of the variability of such data sets. Clearly, a large source of variability in such images is due to local geometric deformations.



(a) (b)

Figure 6: 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 (3.1) with a non-stationnary error term Z. Boxplot in gray correspond to $n = 512$, and in black to $n = 1024$.

Computing parametric diffeomorphisms

It is appropriate to model the geometric variability of such data by a local deformation operator $T_{\boldsymbol{\theta}} f(t) := f(\psi_{\boldsymbol{\theta}}(t))$ where $\psi_{\boldsymbol{\theta}} : \Omega \to \Omega$ is a diffeomorphism. To build a family $(\psi_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ of parametric diffeomorphisms of Ω, we follow the approach proposed in [BGL09]. Let $v : \Omega \to \mathbb{R}^2$ be a smooth parametric vector field given by a linear combination of known basis functions $\{h_k : \Omega \to \mathbb{R},\ k = 1, \ldots, K\}$, such that

$$v(t) = \begin{pmatrix} \sum_{k=1}^K \theta_k^{(1)} h_k(t) \\ \sum_{k=1}^K \theta_k^{(2)} h_k(t) \end{pmatrix} \in \mathbb{R}^2 \text{ for } t \in \Omega.$$

where the $\theta_k^{(i)}$'s are reals coefficients. The function v is thus parametrized by the set of coefficients $\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \ldots, \theta_K^{(1)})' \in \mathbb{R}^K$ and $\boldsymbol{\theta}^{(2)} = (\theta_1^{(2)}, \ldots, \theta_K^{(2)})' \in \mathbb{R}^K$, and we write $v = v_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) \in \mathbb{R}^p$ with $p = 2K$ to stress this dependency. In what follows, it will be assumed that the basis functions are continuously differentiable on Ω and such that h_j and $\partial_t h_j$ vanish at the boundaries of Ω. For the h_j's we took in our simulations a set of $K = 9$ two-dimensional B-spline functions obtained by the tensor product of one-dimensional B-spline of degree $p = 3$ using equally-spaced knots on $[0, 1]$. Then, let $t \in \Omega$ and for $u \in [0, 1]$ consider the following ordinary differential equation (ODE)

$$\frac{\partial}{\partial u}\psi(u, t) = v_{\boldsymbol{\theta}}(\psi(u, t)) \tag{6.2}$$

with initial condition $\psi(0, t) = t$. Then, it can be shown (see e.g. [BGL09]) that for any $u \in [0, 1]$ the solution of the above ODE is unique and such that $t \mapsto \psi(u, t)$ is a diffeomorphism of Ω such that $\psi(u, \Omega) = \Omega$. Then, denote by $\psi_{\boldsymbol{\theta}}(t) = \psi(1, t)$ the solution at $u = 1$. In this way, we finally obtain a diffeomorphism $\psi_{\boldsymbol{\theta}}$ that is parametrized by the set of coefficients $\boldsymbol{\theta} \in \mathbb{R}^p$.

Fréchet mean versus the standard empirical mean

To smooth the images, we use the two-dimensional discrete Fourier transform of an $N \times N$ image (with $N = 28$) and low-pass filtering (the frequency cutoff is chosen to $\lambda = 10$ in the horizontal and vertical directions). Using the above described family of parametric diffeomorphisms, we compute the Fréchet mean $\hat{f}^{\lambda}$ of the smoothed images using the constraint that $\sum_{j=1}^J \boldsymbol{\theta}_j = 0$ (details on how to compute the gradient of the criterion $M(\boldsymbol{\theta})$ using such deformation operators can be found in [BGL09]). The results displayed in Figure 7 to Figure 9 clearly show the improvements obtained when using the Fréchet mean instead of the usual empirical mean $\bar{\mathbf{Y}}$. It can be seen that the Fréchet mean is an image with sharper edges than the empirical mean which is a very blurred image.

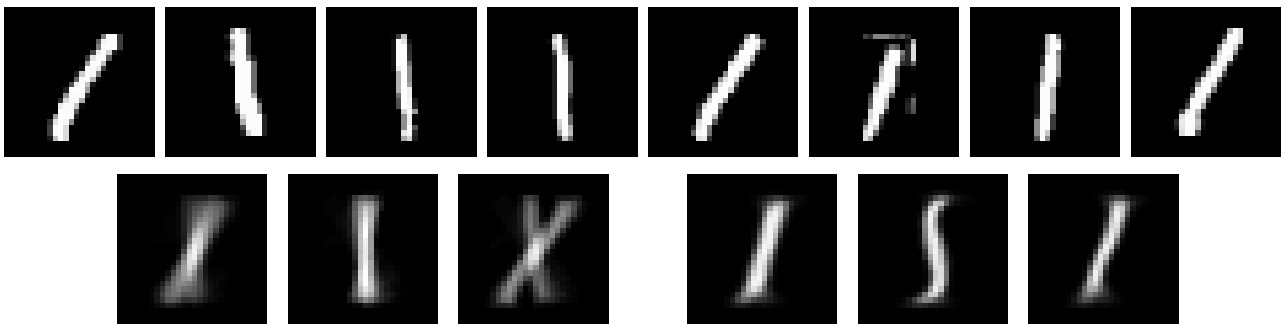
Geometric PCA

For each digit, one can perform a standard linear PCA by transforming the $J = 100$ images into vectors of $\mathbb{R}^n$ with $n = 748$. Let us denote by $\hat{e}_1 \in \mathbb{R}^n$ and $\hat{e}_2 \in \mathbb{R}^n$ the first two principal component vectors with associated eigenvalues $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ (by a slight abuse of notations these vectors can also be viewed as images). Then, we define the first two linear modes of variations of the data as the images $\bar{\mathbf{Y}} + \alpha_1 \hat{e}_1$ and $\bar{\mathbf{Y}} + \alpha_2 \hat{e}_2$ with $\alpha_1 = 1.5\hat{\sigma}_1$ and $\alpha_2 = 1.5\hat{\sigma}_2$. For the digit 1 and 5, it can be seen that linear PCA is not appropriate to represent the shape variability of the data, as the second mode of variation show geometric artifacts. The images in Figure 7(c) and Figure 9(c) are no longer a single digit but rather the superposition of two digits in different orientations.

Denote by $\hat{\boldsymbol{\theta}}_1^{(i)}, \ldots, \hat{\boldsymbol{\theta}}_J^{(i)}, i = 1, 2$ be the estimated coefficients obtained by the minimization of $M(\boldsymbol{\theta})$. A geometric PCA can be obtained by performing a standard linear PCA of the vectors $\hat{\boldsymbol{\theta}}_j^{(i)}, j = 1, \ldots, J$ for each direction $i = 1, 2$. Let $\hat{e}_1^{(i)} \in \mathbb{R}^K$ and $\hat{e}_2^{(i)} \in \mathbb{R}^K$ be the first two principal component vectors with associated eigenvalues $\hat{\sigma}_{i,1}^2$ and $\hat{\sigma}_{i,2}^2$ of the "data" $\hat{\boldsymbol{\theta}}_1^{(i)}, \ldots, \hat{\boldsymbol{\theta}}_J^{(i)}$ for $i = 1, 2$. Then, define the first two principal vector fields

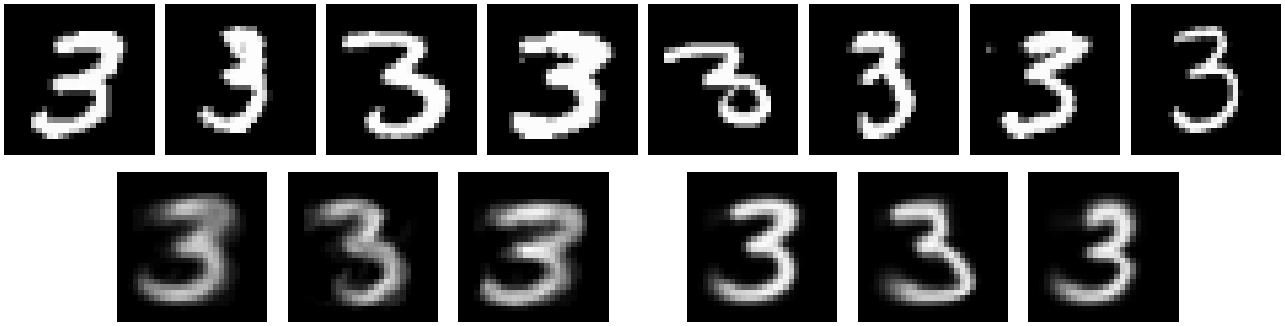
$$\hat{v}_1(t) = 1.5 \begin{pmatrix} \hat{\sigma}_{1,1} \sum_{k=1}^K \hat{e}_{1,k}^{(1)} h_k(t) \\ \hat{\sigma}_{2,1} \sum_{k=1}^K \hat{e}_{1,k}^{(2)} h_k(t) \end{pmatrix} \text{ and } \hat{v}_2(t) = 1.5 \begin{pmatrix} \hat{\sigma}_{1,2} \sum_{k=1}^K \hat{e}_{2,k}^{(1)} h_k(t) \\ \hat{\sigma}_{2,2} \sum_{k=1}^K \hat{e}_{2,k}^{(2)} h_k(t) \end{pmatrix} \text{ for } t \in \Omega,$$

and the two first modes of deformations $\hat{\psi}_1$ and $\hat{\psi}_2$ as the solution at $u = 1$ of the ODE (6.2) governed by the vector field $\hat{v}_1$ and $\hat{v}_2$ respectively. Then, the first and second geometric mode of variation of the data are defined as the images $\hat{f}^\lambda \circ \hat{\psi}_1$ and $\hat{f}^\lambda \circ \hat{\psi}_2$. The results of such a geometric PCA procedure is displayed in Figure 7(c) and Figure 9(c). For the digit 3, it gives results similar to standard linear PCA but with images with sharper images. For the digits 1 and 5, it gives results of a different flavor. Such a geometric PCA overcomes the artefacts observed with standard linear PCA, and clearly better reflects the geometric variability of the observed images.



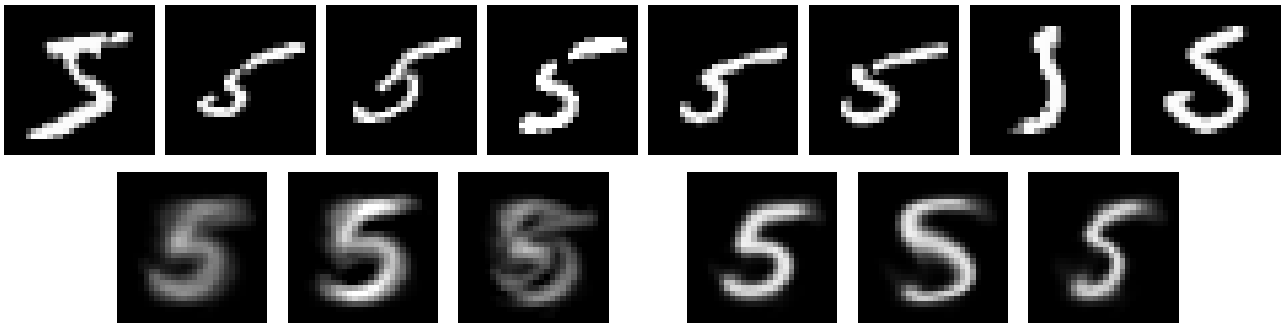
(a) $\bar{\mathbf{Y}}$ (b) $\bar{\mathbf{Y}} + \alpha_1 \hat{e}_1$ (c) $\bar{\mathbf{Y}} + \alpha_2 \hat{e}_2$ (d) $\hat{f}^\lambda$ (e) $\hat{f}^\lambda \circ \hat{\psi}_1$ (f) $\hat{f}^\lambda \circ \hat{\psi}_2$

Figure 7: Digit 1 (Mnist database). First row: sample of 8 digits out of $J = 100$. Second row: (a) empirical mean $\bar{\mathbf{Y}}$, (b) $\bar{\mathbf{Y}} + \alpha_1 \hat{e}_1$ (first linear mode of variation), (c) $\bar{\mathbf{Y}} + \alpha_2 \hat{e}_2$ (second linear mode of variation), (d) Fréchet mean $\hat{f}^\lambda$, (e) $\hat{f}^\lambda \circ \hat{\psi}_1$ (first geometric mode of variation), (f) $\hat{f}^\lambda \circ \hat{\psi}_2$ (second geometric mode of variation).



(a) $\bar{\mathbf{Y}}$ (b) $\bar{\mathbf{Y}} + \alpha_1 \hat{e}_1$ (c) $\bar{\mathbf{Y}} + \alpha_2 \hat{e}_2$ (d) $\hat{f}^{\lambda}$ (e) $\hat{f}^{\lambda} \circ \hat{\psi}_1$ (f) $\hat{f}^{\lambda} \circ \hat{\psi}_2$

Figure 8: Digit 3 (Mnist database). First row: sample of 8 digits out of $J = 100$. Second row: (a) empirical mean $\bar{\mathbf{Y}}$, (b) $\bar{\mathbf{Y}} + \alpha_1 \hat{e}_1$ (first linear mode of variation), (c) $\bar{\mathbf{Y}} + \alpha_2 \hat{e}_2$ (second linear mode of variation), (d) Fréchet mean $\hat{f}^{\lambda}$, (e) $\hat{f}^{\lambda} \circ \hat{\psi}_1$ (first geometric mode of variation), (f) $\hat{f}^{\lambda} \circ \hat{\psi}_2$ (second geometric mode of variation).



(a) $\bar{\mathbf{Y}}$ (b) $\bar{\mathbf{Y}} + \alpha_1 \hat{e}_1$ (c) $\bar{\mathbf{Y}} + \alpha_2 \hat{e}_2$ (d) $\hat{f}^{\lambda}$ (e) $\hat{f}^{\lambda} \circ \hat{\psi}_1$ (f) $\hat{f}^{\lambda} \circ \hat{\psi}_2$

Figure 9: Digit 5 (Mnist database). First row: sample of 8 digits out of $J = 100$. Second row: (a) empirical mean $\bar{\mathbf{Y}}$, (b) $\bar{\mathbf{Y}} + \alpha_1 \hat{e}_1$ (first linear mode of variation), (c) $\bar{\mathbf{Y}} + \alpha_2 \hat{e}_2$ (second linear mode of variation), (d) Fréchet mean $\hat{f}^{\lambda}$, (e) $\hat{f}^{\lambda} \circ \hat{\psi}_1$ (first geometric mode of variation), (f) $\hat{f}^{\lambda} \circ \hat{\psi}_2$ (second geometric mode of variation).

Appendices

A Proof of the consistency in the setting of randomly shifted curves

A.1 Proof of Theorem 3.1

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 A.1. *Suppose that Assumption 3.2 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 3.1 is a consequence of Lemma A.1 and Theorem 5.1. Indeed, it can be easily checked that, under the assumptions of Theorem 3.1, Assumptions 4.1 to 4.9 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 4). □

A.2 Proof of Proposition 3.1

The result of Proposition 3.1 is a direct consequence of the following lemma:

Lemma A.2. *Suppose that Assumptions 3.2 and 3.3 holds with $\rho < 1/16$. Then, for any $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \in \boldsymbol{\Theta}_0$, one has that $D(\boldsymbol{\theta}) \geq C(f^*, \rho)\frac{1}{J}\sum_{j=1}^{J} \left|\boldsymbol{\theta}_j - [\boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}]_j\right|^2$, where $C(f^*, \rho) > 0$ is a constant depending only on f^* and ρ.*

Proof. By definition of $D(\boldsymbol{\theta})$ and thanks to Assumption 3.3, it follows that for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

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

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 3.2 and the hypothesis that $\rho < 1/16$. Hence, using the lower bound (A.1), it follows that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

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

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

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

The above equality together with the lower bound (A.2) completes the proof. □

A.3 Proof of Theorem 3.2

The random variable $\boldsymbol{\theta}_j^*$'s are i.i.d. with values in Θ and with density $g : \Theta \longrightarrow \mathbb{R}$. Denote by $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_J) \in \mathbb{R}^{n\times J}$ the observations generated by the model (3.1). Under Assumption 3.1, the log-likelihood of $\mathbf{Y}$ conditionally to $\boldsymbol{\theta}^*$ is equal to

$$\log(p(\mathbf{Y}|\boldsymbol{\theta}^*)) = -\frac{Jn}{2}\log(2\pi) + \frac{J}{2}\log(\det(\Lambda)) - \frac{1}{2}\sum_{j=1}^{J}(\mathbf{Y}_j - \mathbf{f}^*(\cdot - \boldsymbol{\theta}_j^*))'\Lambda(\mathbf{Y}_j - \mathbf{f}^*(\cdot - \boldsymbol{\theta}_j^*)),$$

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

$$\mathbb{E}_{\boldsymbol{\theta}^*}\left[\partial_{\boldsymbol{\theta}^*_{j_1}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*))\partial_{\boldsymbol{\theta}^*_{j_2}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*))\right] = \begin{cases} 0 & \text{if } j_1 \neq j_2, \\ \partial_t \mathbf{f}^*(\cdot - \boldsymbol{\theta}^*_{j_1}) \ \Lambda \ \partial_t \mathbf{f}^*(\cdot - \boldsymbol{\theta}^*_{j_1}) & \text{if } j_1 = j_2, \end{cases}$$

where $\partial_t \mathbf{f}^*(\cdot - \boldsymbol{\theta}^*_{j_1}) = \left[\partial_t f^*(t_\ell - \boldsymbol{\theta}^*_{j_1})\right]_{\ell=1}^n$. Hence, under the assumptions of Theorem 3.2 on g, we can apply the multivariate Van Tree inequality (see Theorem 1 in [GL95]), to obtain that for any estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{Y})$

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{J}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2\right] &\geq \frac{J}{\sum_{j=1}^J \int_{\Theta^J} \partial_t \mathbf{f}^*(\cdot - \boldsymbol{\theta}^*_j)\Lambda \partial_t \mathbf{f}^*(\cdot - \boldsymbol{\theta}^*_j) g^J(\boldsymbol{\theta}^*) d\boldsymbol{\theta}^* + J \int_\Theta \left(\partial_{\boldsymbol{\theta}^*_1} \log\left(g(\boldsymbol{\theta}^*_1)\right)\right)^2 g(\boldsymbol{\theta}^*_1) d\boldsymbol{\theta}^*_1} \\
&\geq \frac{J}{(\sigma^2+\gamma)^{-1} \sum_{j=1}^J \int_{\Theta^J} \left\|\partial_t \mathbf{f}^*(\cdot - \boldsymbol{\theta}^*_j)\right\|^2 g^J(\boldsymbol{\theta}^*) d\boldsymbol{\theta}^* + J \int_\Theta \left(\partial_{\boldsymbol{\theta}^*_1} \log\left(g(\boldsymbol{\theta}^*_1)\right)\right)^2 g(\boldsymbol{\theta}^*_1) d\boldsymbol{\theta}^*_1} \\
&\geq \frac{(\sigma^2+\gamma)}{n\|\partial_t f^*\|_\infty^2 + (\sigma^2+\gamma) \int_\Theta \left(\partial_{\boldsymbol{\theta}^*_1} \log\left(g(\boldsymbol{\theta}^*_1)\right)\right)^2 g(\boldsymbol{\theta}^*_1) d\boldsymbol{\theta}^*_1},
\end{aligned}$$

where $g^J(\boldsymbol{\theta}^*) = g(\boldsymbol{\theta}^*_1) \dots g(\boldsymbol{\theta}^*_J)$. □

A.4 Proof of Theorem 3.3

Consider the following inequality $\|\hat{f}^\lambda - f^*\|^2 \leq 2\|\hat{f}^\lambda - f^*_{\mathbf{\Theta}_0}\|^2 + 2\|f^*_{\mathbf{\Theta}_0} - f^*\|^2$, where $f^*_{\mathbf{\Theta}_0}(t) = f^*(t - \bar{\boldsymbol{\theta}}^*)$ 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 $\|f^*_{\mathbf{\Theta}_0} - f^*\|^2_{L^2} \leq C(\Theta, f^*)|\bar{\boldsymbol{\theta}}^*|^2 = C(\Theta, f^*)\frac{1}{J}\|\boldsymbol{\theta}^*_{\mathbf{\Theta}_0} - \boldsymbol{\theta}^*\|^2$. As explained in part A.1 the assumptions of Theorem 3.3 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 5.3. □

A.5 Proof of Theorem 3.4

Let $n \geq 1$. We have that

$$\mathbb{E}[\|\tilde{f} - f^*\|_{L^2}] = \mathbb{E}\|\tilde{f} - f^*_{\mathbf{\Theta}_0} + f^*_{\mathbf{\Theta}_0} - f^*\|_{L^2} \geq \Big| \underbrace{\mathbb{E}\|\tilde{f} - f^*_{\mathbf{\Theta}_0}\|_{L^2}}_{\mathbf{A}} - \underbrace{\mathbb{E}\|f^*_{\mathbf{\Theta}_0} - f^*\|_{L^2}}_{\mathbf{B}} \Big| \tag{A.3}$$

where for all $t \in [0,1]$, $\tilde{f}(t) = \frac{1}{J}\sum_{j=1}^J f^*(t - \boldsymbol{\theta}^*_j + \hat{\boldsymbol{\theta}}^\lambda_j)$, and $f^*_{\mathbf{\Theta}_0}(t) = f^*(t + \bar{\boldsymbol{\theta}}^*)$. 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 $C_0(f^*, g, n, \sigma^2, \gamma, \rho) = C(f^*, \rho)\frac{n^{-1}(\sigma^2+\gamma)}{\|\partial_t f^*\|^2_\infty + n^{-1}(\sigma^2+\gamma)\int_\Theta (\partial_{\boldsymbol{\theta}} \log(g(\boldsymbol{\theta})))^2}$ 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 $\mathbb{E}\|\tilde{f} - \tilde{f}^*\|_{L^2} \geq \frac{1}{2}C_0(f^*, g, n, \sigma^2, \gamma, \rho)$, 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

$$\begin{aligned}
\|\tilde{f} - f^*_{\mathbf{\Theta}_0}\|_{L^2} = \|\frac{1}{J}\sum_{j=1}^J f^*(\cdot - \boldsymbol{\theta}^*_j + \hat{\boldsymbol{\theta}}^\lambda_j) - f^*(\cdot + \bar{\boldsymbol{\theta}}^*)\|_{L^2} &= \left(\sum_{k\in\mathbb{Z}} \left|\frac{1}{J}\sum_{j=1}^J \left(e^{i2\pi k(-\boldsymbol{\theta}^*_j + \hat{\boldsymbol{\theta}}^\lambda_j)} - e^{i2\pi k\bar{\boldsymbol{\theta}}^*}\right)c^*_k\right|^2\right)^{\frac{1}{2}}, \\
&\geq |c^*_1|\left|\frac{1}{J}\sum_{j=1}^J (e^{i2\pi(\hat{\boldsymbol{\theta}}^\lambda_j - [\boldsymbol{\theta}^*_{\mathbf{\Theta}_0}]_j)} - 1)\right|,
\end{aligned}$$

where $\boldsymbol{\theta}^*_{\mathbf{\Theta}_0} = (\boldsymbol{\theta}^*_1 - \bar{\boldsymbol{\theta}}^*, \dots, \boldsymbol{\theta}^*_J - \bar{\boldsymbol{\theta}}^*)$, the right hand side of the preceding inequality being positive since Assumption 3.2 ensures that $c^*_1 \neq 0$ for all $j = 1, \dots, J$. Let $u_j = 2\pi(\hat{\boldsymbol{\theta}}^\lambda_j - [\boldsymbol{\theta}^*_{\mathbf{\Theta}_0}]_j), j = 1, \dots, J$

. Since $\sum_{j=1}^{J} u_j = 0$ and $|u_j| \leq 4\pi\rho < 3, j = 1, \ldots, J$ (by our assumption on ρ), Lemma A.3 implies that

$$\|\tilde{f} - f^*_{\boldsymbol{\Theta}_0}\|_{L^2} \geq C(f^*, \rho)\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*_{\boldsymbol{\Theta}_0}\|^2. \tag{A.4}$$

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 $\mathbf{C} = 2\mathbb{E}\big[\,|\bar{\boldsymbol{\theta}}^*|\,\frac{1}{J}\sum_{j=1}^{J}|\hat{\boldsymbol{\theta}}^\lambda_j - \boldsymbol{\theta}^*_j|\big]$. By applying Theorem 3.2 we get that

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

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 (A.4), 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

$$\mathbf{A} = \mathbb{E}\big[\|\tilde{f}^\lambda - \tilde{f}^*\|_{L^2}\big] \geq C_0(f^*, g, n, \sigma^2, \gamma, \rho). \tag{A.5}$$

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

$$\mathbf{B} = \mathbb{E}[\|\tilde{f}^*_{\boldsymbol{\Theta}_0} - \tilde{f}^*\|_{L^2}] \leq \frac{1}{2}C_0(f^*, g, n, \sigma^2, \gamma, \rho). \tag{A.6}$$

To conclude the proof, equations (A.3), (A.5) and (A.6) imply that there exists a $J_0 \in \mathbb{N}$ such that $J \geq J_0$ implies $\mathbb{E}\|\hat{f}^\lambda - \tilde{f}^*\|_{L^2} \geq |\mathbf{A} - \mathbf{B}| \geq \frac{1}{2}C_0(f^*, g, n, \sigma^2, \gamma, \rho)$. □

Lemma A.3. *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 $\|u\|^2 = u_1^2 + \ldots + u_J^2$.*

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 $\delta < 3$. □

B Proof of the consistency in general case

B.1 Proof of Theorem 5.1

We explain here the main arguments of the proof of Theorem 5.1. Technical Lemmas are given in the second part of the Appendix. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) = (\theta_1^1, \dots, \theta_1^p, \dots, \theta_J^1, \dots, \theta_J^p) \in \mathbb{R}^{pJ}$ and decompose the criterion (5.1) as follows,

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

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

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

and the others terms contain the Z_j's and $\boldsymbol{\varepsilon}_j$'s error terms. Let $\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j = \left(T_{\boldsymbol{\theta}_j^*} Z_j(t_\ell)\right)_{\ell=1}^n$ and $\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{f}^* = \left(T_{\boldsymbol{\theta}_j^*} f^*(t_\ell)\right)_{\ell=1}^n$, 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, \\ 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) \\ &\qquad \times \left(\tilde{T}_{\boldsymbol{\theta}_j}\left\langle S_\lambda(t), \boldsymbol{\varepsilon}_j\right\rangle - \frac{1}{J}\sum_{j'=1}^{J}\tilde{T}_{\boldsymbol{\theta}_{j'}}\left\langle S_\lambda(t), \boldsymbol{\varepsilon}_{j'}\right\rangle\right) dt \\ Q_\lambda^\varepsilon(\boldsymbol{\theta}) &= \frac{\sigma^2}{J}\sum_{j=1}^{J}\int_\Omega \left(\tilde{T}_{\boldsymbol{\theta}_j}\left\langle S_\lambda(t), \boldsymbol{\varepsilon}_j\right\rangle - \frac{1}{J}\sum_{j'=1}^{J}\tilde{T}_{\boldsymbol{\theta}_{j'}}\left\langle S_\lambda(t), \boldsymbol{\varepsilon}_{j'}\right\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) \\ &\qquad \times \left(\tilde{T}_{\boldsymbol{\theta}_j}\left\langle S_\lambda(t), \boldsymbol{\varepsilon}_j\right\rangle - \frac{1}{J}\sum_{j'=1}^{J}\tilde{T}_{\boldsymbol{\theta}_{j'}}\left\langle S_\lambda(t), \boldsymbol{\varepsilon}_{j'}\right\rangle\right) dt. \end{aligned}$$

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. Lemma C.2 ensures that there is a constant $C(\Theta, \mathcal{F}, f^*) > 0$ such that,

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

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

$$\begin{aligned}\left|D(\hat{\boldsymbol{\theta}}^{\lambda}) - D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}})\right| &\leq 2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|D(\boldsymbol{\theta}) - M_\lambda(\boldsymbol{\theta})| \qquad \text{(B.2)}\\ &= \underbrace{2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{R_\lambda(\boldsymbol{\theta}) + Q_\lambda(\boldsymbol{\theta})\right\}}_{\mathbf{B}} + \underbrace{2\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{Q^Z_\lambda(\boldsymbol{\theta}) + R^Z_\lambda(\boldsymbol{\theta}) + R^{Z,\varepsilon}_\lambda(\boldsymbol{\theta}) + Q^\varepsilon_\lambda(\boldsymbol{\theta}) + R^\varepsilon_\lambda(\boldsymbol{\theta})\right\}}_{\mathbf{V}}\end{aligned}$$

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

Control of B. Using Assumption 4.5 and 4.2, we have that $Q_\lambda(\boldsymbol{\theta}) \leq \frac{C(\Theta)}{J}\sum_{j=1}^J \left\|B_\lambda(T_{\boldsymbol{\theta}^*_j} f, t)\right\|^2_{L^2} \leq C(\Theta, \mathcal{F})B(\lambda)$. 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 4.2, there exists a constant such $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\{D(\boldsymbol{\theta})\} \leq C(\Theta, \mathcal{F}, f^*)$ and thus

$$\mathbf{B} \leq C(\Theta, \mathcal{F}, f^*)\big(B(\lambda) + \sqrt{B(\lambda)}\big). \tag{B.3}$$

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

$$\left|Q^Z_\lambda(\boldsymbol{\theta}) + R^Z_\lambda(\boldsymbol{\theta}) + R^{Z,\varepsilon}_\lambda(\boldsymbol{\theta}) + Q^\varepsilon_\lambda(\boldsymbol{\theta}) + R^\varepsilon_\lambda(\boldsymbol{\theta})\right| \leq C(\Theta, \mathcal{F}, f^*)\left(\sqrt{Q^Z_\lambda(\boldsymbol{\theta})} + Q^Z_\lambda(\boldsymbol{\theta}) + Q^\varepsilon_\lambda(\boldsymbol{\theta}) + \sqrt{Q^\varepsilon_\lambda(\boldsymbol{\theta})}\right),$$

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

$$\mathbb{P}\left(\mathbf{V} \geq C(\Theta, \mathcal{F}, f^*)(\gamma_{\max}(n) + \sigma^2)\big(\upsilon(x, J, \lambda) + \sqrt{\upsilon(x, J, \lambda)}\big)\right) \leq 2e^{-x}, \tag{B.4}$$

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

Putting together equations (B.1), (B.2), (B.3) and (B.4), we have

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

which completes the proof of Theorem 5.1. □

B.2 Proof of Theorem 5.2

The proof is very similar to the proof of Theorem 3.2. The random variable $\boldsymbol{\theta}^*_j = (\theta^{*,1}_j, \ldots, \theta^{*,p}_j)$'s are i.i.d. with values in $\Theta \subset \mathbb{R}^p$ and with density $g : \Theta \longrightarrow \mathbb{R}$. Consider the model (1.5) and let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_J) \in \mathbb{R}^{n\times J}$. Under Assumption 5.1 the log-likelihood of $\mathbf{Y}$ conditionally to $\boldsymbol{\theta}^*$ is equal to

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

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

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

where $\partial_{\theta^{*,p_1}_{j_1}} \mathbf{T}_{\boldsymbol{\theta}^*_{j_1}} \mathbf{f}^* = \left[\partial_{\theta^{*,p_1}_{j_1}} T_{\boldsymbol{\theta}^*_{j_1}} f^*(t_\ell)\right]_{\ell=1}^{n}$. Then, remark that for each $j_1 = 1, \dots, J$ and $p_1, p_2 = 1, \dots, p$ we have that

$$(\partial_{\theta^{*,p_1}_{j_1}} \mathbf{T}_{\boldsymbol{\theta}^*_{j_1}} \mathbf{f}^*)' \, \Lambda \, (\partial_{\theta^{*,p_2}_{j_1}} \mathbf{T}_{\boldsymbol{\theta}^*_{j_1}} \mathbf{f}^*) \leq (\sigma^2 + \gamma_n(\Theta))^{-1} \left\| \partial_{\theta^{*,p_1}_{j_1}} \mathbf{T}_{\boldsymbol{\theta}^*_{j_1}} \mathbf{f}^* \right\|^2 \leq nC(\Theta, \mathcal{F}, f^*)(\sigma^2 + \gamma_n(\Theta))^{-1},$$

where the last inequality is a consequence of Assumption 4.3. Under the assumptions of Theorem 5.2 on the density g, the result follows from the multivariate Van Tree's inequality (see Theorem 1 in [GL95]). □

B.3 Proof of Theorem 5.3

In this part, we use the notations introduced in the proof of Theorem 5.1. We have,

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

Again, the first term above depends on the bias, and the second term (stochastic) can be controlled in probability. Under Assumptions 4.2 and 4.5 we have that

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

and

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

The stochastic term $\frac{1}{J} \sum_{j=1}^{J} \left\| \left\langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}^*_j} \mathbf{Z}_j + \boldsymbol{\varepsilon}_j \right\rangle \right\|_{L^2}^2$ in the above inequality can be been controlled using Lemma C.3 and the arguments in the proof of Corollaries C.1 and C.2 to obtain that for any $x > 0$

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

Then, from Theorem 5.1 it follows that

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

which completes the proof. □

C Technical Lemmas

Assumption 4.1 and 4.3 allow us to derive an expression of the gradient and the Hessian of the criterion. Let $f^*_{\boldsymbol{\theta}_j}(t) = \tilde{T}_{\boldsymbol{\theta}_j} T_{\boldsymbol{\theta}^*_j} f^*(t)$ for all $t \in \Omega$, and remark that for all $j_1 = 1, \dots, J$ and $p_1 = 1, \dots, p$,

$$\partial_{\theta^{p_1}_{j_1}} D(\boldsymbol{\theta}) = \frac{2}{J} \left\langle \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, f^*_{\boldsymbol{\theta}_{j_1}} \right\rangle_{L^2} - \frac{2}{J^2} \left\langle \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \sum_{j'=1}^{J} f^*_{\boldsymbol{\theta}_{j'}} \right\rangle_{L^2}. \tag{C.1}$$

The second derivatives are

$$\partial_{\theta^{p_2}_{j_2}} \partial_{\theta^{p_1}_{j_1}} D(\boldsymbol{\theta}) = -\frac{2}{J^2} \left\langle \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \partial_{\theta^{p_2}_{j_2}} f^*_{\boldsymbol{\theta}_{j_2}} \right\rangle_{L^2} \text{ if } j_1 \neq j_2, \tag{C.2}$$

$$\partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} D(\boldsymbol{\theta}) = \frac{2}{J} \left\langle \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \left(f^*_{\boldsymbol{\theta}_{j_1}} - \frac{1}{J} \sum_{j'=1}^{J} f^*_{\boldsymbol{\theta}_{j'}} \right) \right\rangle_{L^2} + \left(\frac{2}{J} - \frac{2}{J^2} \right) \left\langle \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \partial_{\theta^{p_2}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}} \right\rangle_{L^2} \tag{C.3}$$

For $\alpha = (\alpha_1, \dots, \alpha_{pJ}) \in \mathbb{N}^{pJ}$, denote by $|\alpha| = \alpha_1 + \ldots + \alpha_{pJ}$ and

$$(\partial_{\boldsymbol{\theta}})^\alpha = (\partial_{\theta^1_1})^{\alpha_1} \dots (\partial_{\theta^p_1})^{\alpha_p} \dots (\partial_{\theta^1_J})^{\alpha_{p(J-1)+1}} \dots (\partial_{\theta^p_J})^{\alpha_{pJ}}.$$

Then, the differential of order 3 of D at $\mathbf{c} \in \boldsymbol{\Theta}$ applied at $\boldsymbol{\theta} \in \mathbb{R}^{pJ}$ writes as $\left[\nabla^3 D(\mathbf{c})\right](\boldsymbol{\theta}) = \sum_{|\alpha|=3} (\partial_{\boldsymbol{\theta}})^\alpha D(\mathbf{c}) \boldsymbol{\theta}^\alpha$ where $\boldsymbol{\theta}^\alpha = (\theta^1_1)^{\alpha_1} \dots (\theta^p_1)^{\alpha_p} \dots (\theta^1_J)^{\alpha_{p(J-1)+1}} \dots (\theta^p_J)^{\alpha_{pJ}}$.

Lemma C.1. *There is a constant $C(\Theta, \mathcal{F}, f^*)$ independent of J and n such that for all $\mathbf{c} \in \Theta^J$ and $\boldsymbol{\theta} \in \Theta^J, \delta > 0$ satisfying $\left|\boldsymbol{\theta}^{p_1}_j\right| \leq \delta$ for each $j = 1, \dots, J$ and $p_1 = 1, \dots, p$, $\left|\left[\nabla^3 D(\mathbf{c})\right](\boldsymbol{\theta})\right| \leq C(\Theta, \mathcal{F}, f^*) \frac{\delta}{J} \|\boldsymbol{\theta}\|^2$.*

Proof. Recall that $f^*_{\boldsymbol{\theta}_j}(t) = \tilde{T}_{\boldsymbol{\theta}_j} T_{\boldsymbol{\theta}^*_j} f^*(t)$ for all $t \in \Omega$. Then, from equations (C.2) and (C.3), it follows that for all $j_1, j_2, j_3 = 1, \dots, J$ and $p_1, p_2, p_3 = 1, \dots, p$,

$$\partial_{\theta^{p_3}_{j_3}} \partial_{\theta^{p_2}_{j_2}} \partial_{\theta^{p_1}_{j_1}} D(\boldsymbol{\theta}) = 0, \qquad \text{if } j_1 \neq j_2 \text{ and } j_2 \neq j_3 \text{ and } j_1 \neq j_3,$$

$$\partial_{\theta^{p_3}_{j_2}} \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} D(\boldsymbol{\theta}) = -\frac{2}{J^2} \left\langle \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \partial_{\theta^{p_3}_{j_2}} f^*_{\boldsymbol{\theta}_{j_2}} \right\rangle_{L^2}, \qquad \text{if } j_1 \neq j_2,$$

$$\partial_{\theta^{p_3}_{j_1}} \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} D(\boldsymbol{\theta}) = \frac{2}{J} \left\langle \partial_{\theta^{p_3}_{j_1}} \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \left(f^*_{\boldsymbol{\theta}_{j_1}} - \frac{1}{J} \sum_{j'=1}^{J} f^*_{\boldsymbol{\theta}_{j'}} \right) \right\rangle_{L^2} + \left(\frac{2}{J} - \frac{2}{J^2} \right) \left(\left\langle \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \partial_{\theta^{p_3}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}} \right\rangle_{L^2} + \left\langle \partial_{\theta^{p_3}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \partial_{\theta^{p_2}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}} \right\rangle_{L^2} + \left\langle \partial_{\theta^{p_3}_{j_1}} \partial_{\theta^{p_2}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}} \right\rangle_{L^2} \right)$$

By Cauchy-Schwarz inequality

$$\left| \left\langle \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \partial_{\theta^{p_3}_{j_2}} f^*_{\boldsymbol{\theta}_{j_2}} \right\rangle_{L^2} \right| \leq \|\partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}\|_{L^2} \|\partial_{\theta^{p_3}_{j_2}} f^*_{\boldsymbol{\theta}_{j_2}}\|_{L^2} \leq C(\Theta, \mathcal{F}, f^*),$$

where the last inequality is a consequence of Assumption 4.3. Next

$$\begin{aligned} \left| \left\langle \partial_{\theta^{p_3}_{j_1}} \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}}, \left(f^*_{\boldsymbol{\theta}_{j_1}} - \frac{1}{J} \sum_{j'=1}^{J} f^*_{\boldsymbol{\theta}_{j'}} \right) \right\rangle_{L^2} \right| &\leq \left\| \partial_{\theta^{p_3}_{j_1}} \partial_{\theta^{p_2}_{j_1}} \partial_{\theta^{p_1}_{j_1}} f^*_{\boldsymbol{\theta}_{j_1}} \right\|_{L^2} \left\| f^*_{\boldsymbol{\theta}_{j_1}} - \frac{1}{J} \sum_{j'=1}^{J} f^*_{\boldsymbol{\theta}_{j'}} \right\|_{L^2} \\ &\leq 2C(\Theta, \mathcal{F}, f^*) \|f^*\|_{L^2} \leq C(\Theta, \mathcal{F}, f^*) \end{aligned}$$

where we also use Assumption 4.3 and then Assumption 4.2 to derive the last bound. Using the above bounds it follows that

$$\begin{aligned}
\left|\left[\nabla^3 D(\mathbf{c})\right](\boldsymbol{\theta})\right| &= \left|\sum_{p_1,p_2,p_3=1}^{p}\sum_{j_1=1}^{J}\partial_{\theta_{j_1}^{p_3}}\partial_{\theta_{j_1}^{p_2}}\partial_{\theta_{j_1}^{p_1}}D(\mathbf{c})\boldsymbol{\theta}_{j_1}^{p_1}\boldsymbol{\theta}_{j_1}^{p_2}\boldsymbol{\theta}_{j_1}^{p_3} + 3\sum_{j_1\neq j_2=1}^{J}\partial_{\theta_{j_2}^{p_3}}\partial_{\theta_{j_1}^{p_2}}\partial_{\theta_{j_1}^{p_1}}D(\mathbf{c})\boldsymbol{\theta}_{j_1}^{p_1}\boldsymbol{\theta}_{j_1}^{p_2}\boldsymbol{\theta}_{j_2}^{p_3}\right| \\
&\leq \delta C(\Theta,\mathcal{F},f^*)p\sum_{p_1,p_2=1}^{p}\left(\frac{4}{J}\sum_{j_1=1}^{J}\left|\boldsymbol{\theta}_{j_1}^{p_1}\boldsymbol{\theta}_{j_1}^{p_2}\right| + \frac{6(J-1)}{J^2}\sum_{j_1=1}^{J}\left|\boldsymbol{\theta}_{j_1}^{p_1}\boldsymbol{\theta}_{j_1}^{p_2}\right|\right) \\
&\leq \delta C(\Theta,\mathcal{F},f^*)\frac{1}{J}\sum_{j=1}^{J}\sum_{p_1,p_2=1}^{p}\left|\boldsymbol{\theta}_{j}^{p_1}\boldsymbol{\theta}_{j}^{p_2}\right| = \delta C(\Theta,\mathcal{F},f^*)\frac{1}{J}\sum_{j=1}^{J}\left(\sum_{p_1=1}^{p}\left|\boldsymbol{\theta}_{j}^{p_1}\right|\right)^2 \\
&\leq \delta C(\Theta,\mathcal{F},f^*)p\frac{1}{J}\sum_{j=1}^{J}\sum_{p_1=1}^{p}\left|\boldsymbol{\theta}_{j}^{p_1}\right|^2 \leq \delta C(\Theta,\mathcal{F},f^*)\frac{1}{J}\left\|\boldsymbol{\theta}\right\|^2.
\end{aligned}$$

□

Lemma C.2. *Under Assumptions 4.2 to 4.8, there exists a constant $C(\Theta,\mathcal{F},f^*)>0$ such that*

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

Proof. Let us split the compact set $\boldsymbol{\Theta}$ is in two parts : a ball of radius $\delta>0$ centered at $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$, namely, $\mathcal{N}_{\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}}(\delta)=\{\boldsymbol{\theta}\in\boldsymbol{\Theta},\ \left|\boldsymbol{\theta}_j^{p_1}-[\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}]_j^{p_1}\right|\leq\delta, j=1,\dots,J, p_1=1,\dots,p\}$ and his complementary. The radius δ will be chosen so that D can be approximated in $\mathcal{N}_{\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}}(\delta)$ by a quadratic function.

First, using the fact that $\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$ is a minimum of $D(\boldsymbol{\theta})$, a second order Taylor expansion and Lemma C.1 imply that for any $\boldsymbol{\theta}\in\mathcal{N}_{\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}}(\delta)$, there exists a $\mathbf{c}=\mathbf{c}(\boldsymbol{\theta})\in\mathcal{N}_{\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}}(\delta)$ such that

$$\begin{aligned}
D(\boldsymbol{\theta})-D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}) &= \frac{1}{2}(\boldsymbol{\tau}-\boldsymbol{\tau}^*_{\boldsymbol{\Theta}})'\left[\phi'\ \nabla^2 D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}})\ \phi\right](\boldsymbol{\tau}-\boldsymbol{\tau}^*_{\boldsymbol{\Theta}}) + \frac{1}{6}[\nabla^3 D(\mathbf{c})](\boldsymbol{\theta}-\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}) \\
&\geq \frac{1}{2}(\boldsymbol{\tau}-\boldsymbol{\tau}^*_{\boldsymbol{\Theta}})'\left[\phi'\ \nabla^2 D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}})\ \phi\right](\boldsymbol{\tau}-\boldsymbol{\tau}^*_{\boldsymbol{\Theta}}) - \delta C(\Theta,\mathcal{F},f^*)\frac{1}{J}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\right\|^2 \\
&\geq \frac{1}{2}J\gamma_{\min}(J)\frac{1}{J}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\right\|^2\left(1-\frac{\delta C(\Theta,\mathcal{F},f^*)}{J\gamma_{\min}(J)}\right),
\end{aligned}$$

where $\boldsymbol{\tau}^*_{\boldsymbol{\Theta}}$ and $\boldsymbol{\tau}$ are the vectors such that $\phi\boldsymbol{\tau}^*_{\boldsymbol{\Theta}}=\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}$, $\phi\boldsymbol{\tau}=\boldsymbol{\theta}$, and using the fact that $\|\boldsymbol{\tau}-\boldsymbol{\tau}^*_{\boldsymbol{\Theta}}\|=\|\boldsymbol{\theta}-\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\|$ (since $\phi'\phi=Id_{pI}$). Assumption 4.8 implies that one can choose $\delta=\delta(\Theta,f^*,\delta)>0$ sufficiently small such that for all J and n the constant $C_2(\Theta,\mathcal{F},f^*)=\frac{1}{2}J\gamma_{\min}(J)\left(1-\frac{\delta C(\Theta,\mathcal{F},f^*)}{J\gamma_{\min}(J)}\right)$ is strictly positive. Then, using such a δ it follows that for all $\boldsymbol{\theta}\in\mathcal{N}_{\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}}(\delta)$, $|D(\boldsymbol{\theta})-D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}})|\geq C_2(\Theta,\mathcal{F},f^*)\frac{1}{J}\left\|\boldsymbol{\tau}-\boldsymbol{\tau}^*_{\boldsymbol{\Theta}}\right\|^2$. Using Assumption 4.7 there exists a constant $C_2(\Theta,f^*,\delta)>0$ (not depending on J) such that for any $\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\mathcal{N}_{\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}}(\delta)$ $D(\boldsymbol{\theta})-D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}})\geq C_2(\Theta,f^*,\delta)$. Moreover, there is $C_4(\Theta)>0$ such that for all $\boldsymbol{\tau}_j$, $j=1,\dots,J$, we have $\|\boldsymbol{\theta}_j-[\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}]_j\|^2\leq C_4(\Theta)$. Then $\frac{1}{J}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\right\|^2\leq C_4(\Theta)$ which implies that for $\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\mathcal{N}_{\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}}$, $|D(\boldsymbol{\theta})-D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}})|\geq\frac{C_2(\Theta,f^*,\delta)}{C_4(\Theta)}\frac{1}{J}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\right\|^2$. Finally, let $C(\Theta,\mathcal{F},f^*)=\min\left\{C_2(\Theta,\mathcal{F},f^*),\frac{C_4(\Theta)}{C_2(\Theta,f^*,\delta)}\right\}$ and thus for $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{\lambda}$, $\frac{1}{J}\|\boldsymbol{\theta}^{\lambda}-\boldsymbol{\theta}^*_{\boldsymbol{\Theta}}\|^2\leq C(\Theta,\mathcal{F},f^*)|D(\hat{\boldsymbol{\theta}}^{\lambda})-D(\boldsymbol{\theta}^*_{\boldsymbol{\Theta}})|$, which completes the proof. □

Lemma C.3. *Let $\xi_{\lambda,J}(A_1,\dots,A_J)=\frac{1}{J}\sum_{j=1}^{J}\|\langle S_\lambda(\cdot),A_j\boldsymbol{\varepsilon}_j\rangle\|_{L^2}^2$, where $\boldsymbol{\varepsilon}_j\sim\mathcal{N}(0,I_n)$ and the A_j's are nonrandom non-negative $n\times n$ symmetric matrices. Then, for all $x>0$ and all $n\geq 1$,*

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

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 $r_{j,1} \geq \ldots \geq r_{j,n}$ the eigenvalues of $\mathbf{A}_j$ with $r_{j,1} \geq \ldots \geq r_{j,n} \geq 0$ 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 $\log\left(\mathbb{E}\left[e^{u\tilde{\varphi}_{\lambda,j}}\right]\right) \leq -\sum_{\ell=1}^{n}\frac{u^2r_{j,\ell}^2}{1-2ur_{j,\ell}} \leq \frac{u^2\|r_j\|^2}{1-2ur_1}$. where $\|r_j\|^2 = r_{j,1}^2 + \ldots + r_{n,j}^2$. Finally, we have

$$\mathbb{P}\left(\tilde{\varphi}_{\lambda,J} \geq \alpha\right) \leq \exp\left(-\left(u\alpha - \sum_{j=1}^{J}\frac{\|r_j\|^2 u^2}{1-2r_1u}\right)\right) = \exp\left(-\left(u\alpha - \frac{\|r\|^2 u^2}{1-2r_1u}\right)\right), \tag{C.4}$$

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 (C.4) 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

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

by setting $x = \frac{\alpha^2}{4r_1\alpha + 4\|r\|^2}$. We derive the following concentration inequality for $\xi_{\lambda,J} = \frac{1}{J}\tilde{\xi}_{\lambda,J} + \frac{1}{J}\sum_{j=1}^{J}\operatorname{tr}(\mathbf{A}_j)$, $\mathbb{P}\left(\xi_{\lambda,J} \geq \frac{1}{J}\sum_{j=1}^{J}\sum_{\ell=1}^{n} r_{j,\ell} + 4\frac{r_1}{J}x + 4\frac{\|r\|}{J}\sqrt{x}\right) \leq e^{-x}$. To finish the proof, the Cauchy-Schwarz inequality gives $\|r\|^2 = \sum_{j=1}^{J}\sum_{\ell=1}^{n} r_{j,\ell}^2 \leq \left(\sum_{j=1}^{J}\sum_{\ell=1}^{n} r_{j,\ell}\right)^2$ since all the $r_{j,\ell}$'s are positive. □

Corollary C.1. *Under Assumptions 4.2 to 4.8, there exists a 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 4.2 gives the uniform bound

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

where $\xi_{\lambda,J}(\sigma Id_n, \ldots, \sigma Id_n)$ is defined in Lemma C.3 and does not depend on $\boldsymbol{\theta}$. Thus, the result immediately follows from Lemma C.3. □

Corollary C.2. *Under Assumptions 4.2 to 4.9, there exists a 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 4.2 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 C.3 with $A_j = \mathbb{E}_{\boldsymbol{\theta}^*}\big[\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j(\mathbf{T}_{\boldsymbol{\theta}_j^*}\mathbf{Z}_j)'\big]^{\frac{1}{2}}$. 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 $\mathbf{S}_\lambda = \big[\langle S_\lambda^\ell, S_\lambda^{\ell'}\rangle_{L^2}\big]_{\ell,\ell'=1}^n$. Under Assumption 4.9 we have that $\mathrm{tr}(\mathbf{A}_j) \leq \gamma_{\max}(A_j)\,\mathrm{tr}\,\mathbf{S}_\lambda \leq \gamma_n(\Theta)V(\lambda)$, 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 C.3 and by taking expectation with respect to $\boldsymbol{\theta}^*$. □

References

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

[AKT09] Stéphanie Allassonière, Estelle Kuhn, and Alain 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.

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

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

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

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

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

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

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

[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.

[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.

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

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

[GLM07] Fabrice Gamboa, Jean-Michel Loubes, and Elie Maza. Semi-parametric estimation of shifts. *Electron. J. Stat.*, 1:616–640, 2007.

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

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

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

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

[Hel01] Sigurdur Helgason. *Differential geometry, Lie groups, and symmetric spaces*, volume 34 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. Corrected reprint of the 1978 original.

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

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

[HJ90] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[JDJB04] S. Joshi, Brad Davis, B Matthieu Jomier, and Guido Gerig B. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23:151–160, 2004.

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

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

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

[LBBH98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

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

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

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

[Mas07] Pascal Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

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

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

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

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

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

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

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

[Wah90] Grace Wahba. *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

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