



HAL
open science

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

Jérémie Bigot, Benjamin Charlier

► **To cite this version:**

Jérémie Bigot, Benjamin Charlier. On the consistency of Fréchet mean in deformable models for curve and image analysis. 2010. hal-00522831v2

HAL Id: hal-00522831

<https://hal.science/hal-00522831v2>

Preprint submitted on 15 Mar 2011 (v2), last revised 22 Aug 2011 (v4)

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

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

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

Jérémie Bigot and Benjamin Charlier

Institut de Mathématiques de Toulouse
Université de Toulouse et CNRS (UMR 5219)

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

March 15, 2011

Abstract

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

Keywords: Mean pattern estimation, Fréchet mean, Shape analysis, Deformable models, Curve registration, Image warping, Geometric variability, 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:

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

where $f_j : \Omega \rightarrow \mathbb{R}$ 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 are random elements which vary around a common mean pattern. The goal of this paper is to estimate such a mean pattern and to study the consistency of the proposed estimators in various asymptotic setting: either when both the number n of design points and the number J of curves (or images) tend to infinity, or when n (resp. J) remains fixed while J (resp. n) tends to infinity.

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

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

where $f^* : \Omega \rightarrow \mathbb{R}$ 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

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

Therefore, in model (1.2), the usual empirical mean \bar{Y}^ℓ 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 $L^2(\Omega)$ the set of square integrable real-valued functions on Ω with respect to the Lebesgue measure dt on \mathbb{R}^d and by \mathcal{P} an open subset of \mathbb{R}^p . To the set \mathcal{P} , we associate a parametric family of operators $(T_\boldsymbol{\theta})_{\boldsymbol{\theta} \in \mathcal{P}}$ such that for each $\boldsymbol{\theta} \in \mathcal{P}$ the operator $T_\boldsymbol{\theta} : L^2(\Omega) \rightarrow L^2(\Omega)$ represents a geometric deformation (parametrized by $\boldsymbol{\theta}$) of a curve or an image. Examples of such deformation operators include the cases of:

- *Shifted curves*: $T_\boldsymbol{\theta} f(t) := f(t - \boldsymbol{\theta})$, with $d = 1$, $\Omega = [0, 1]$, $f \in L_{per}^2([0, 1])$ (the space of periodic functions in $L^2([0, 1])$ with period 1) and \mathcal{P} an open set of \mathbb{R} .
- *Rigid deformation of two-dimensional images*: $T_\boldsymbol{\theta} f(t) := f(e^a R_\alpha t - b)$, for $\boldsymbol{\theta} = (a, \alpha, b) \in \mathcal{P}$ with $d = 2$, $\Omega = \mathbb{R}^2$, $f \in L^2(\Omega)$, $\mathcal{P} \subset \mathbb{R} \times \mathbb{R} \times \mathbb{R}^2$ where $R_\alpha = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$ is a rotation matrix in \mathbb{R}^2 , e^a is an isotropic scaling and b a translation in \mathbb{R}^2 .
- *Deformation by a Lie group action*: the two above cases are examples of a Lie group action on the space $L^2(\Omega)$ (see [Hel01] for an introduction to Lie groups). More generally, assume that G is a connected Lie group of dimension p acting on Ω , meaning that for any $(g, t) \in G \times \Omega$ the action \cdot of G onto Ω is such that $g \cdot t \in \Omega$. In general, G is not a linear space but can be locally parametrized by its Lie algebra $\mathcal{G} \simeq \mathbb{R}^p$ 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 $\exp : \mathcal{G} \rightarrow G$. If $\mathcal{P} \subset \mathbb{R}^p$, then for every $(\boldsymbol{\theta}, f) \in \mathcal{P} \times L^2(\Omega)$ define the deformation operators

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

- *Non-rigid deformation of curves or images*: let us recall that a diffeomorphism of Ω is a smooth map $\psi : \Omega \rightarrow \Omega$ having a smooth inverse with $\psi(\Omega) = \psi^{-1}(\Omega) = \Omega$, and assume that one

can construct a family $(\psi_{\theta})_{\theta \in \mathcal{P}}$ of parametric diffeomorphisms of Ω (further details on how to build such a family can be found in [BGL09]). Then, for every $(\theta, f) \in \mathcal{P} \times L^2(\Omega)$, define the deformation operators

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

Modeling the geometric variability of curves or images by a generic family of deformation operators $(T_{\theta})_{\theta \in \mathcal{P}}$ 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 $(T_{\theta})_{\theta \in \mathcal{P}}$ 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 $f^* \in L^2(\Omega)$ such that

$$f_j(t) = T_{\theta_j^*}[f^* + Z_j](t), \quad \text{for all } t \in \Omega, \text{ and } j = 1, \dots, J, \quad (1.3)$$

where $\theta_j^* \in \mathcal{P}$, $j = 1, \dots, J$ are i.i.d. random variables (independent of the ε_j^ℓ 's) with an unknown density g with compact support Θ included in \mathcal{P} . 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^* . To the contrary, the random operators T_{θ_j} model geometric deformations in time or space, and thus correspond to a source of variability in shape in the data. In what follows, we make the following assumption

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

$$\Theta = \{\theta = (\theta^1, \dots, \theta^p) \in \mathbb{R}^p, |\theta^{p_1}| \leq \rho, 1 \leq p_1 \leq p\} \quad (1.4)$$

where $\rho > 0$.

Therefore, if we assume that the T_{θ} 's are linear operators, equation (1.3) leads to the following *statistical deformable model* for curve or image analysis

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

The two error terms ε_j^ℓ and 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 individual variations in intensity around the mean pattern. Note that a subclass of the deformable model (1.5) is the so-called shape invariant model (SIM)

$$Y_j^\ell = T_{\theta_j^*}f^*(t_\ell) + \sigma\varepsilon_j^\ell, \quad j = 1, \dots, J, \text{ and } \ell = 1, \dots, n, \quad (1.6)$$

i.e. without incorporating in (1.5) the additive terms Z_j for $j = 1, \dots, J$. 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 goal of this paper is twofold. First, we propose a general methodology for estimating f^* and the θ_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], [Zie77], [OC95], [BP03] that has been widely studied in shape analysis [Goo91], [KM97], [Le98], [LK00], and more recently in biomedical imaging [JDJB04], [Pen06] among many others. Secondly, we study the consistency of the resulting estimators in various asymptotic settings: either when n and J both tend to infinity, or when n is fixed and $J \rightarrow +\infty$, or when J is fixed and $n \rightarrow +\infty$.

1.2 Organization of the paper

Section 2 contains a description of our estimating procedure and a review of previous work in mean pattern estimation. In Section 3, we derive a lower bound on the variance of estimators of the

deformation parameters that shows the impossibility of consistently estimating the θ_j^* 's in the setting where n is fixed and only $J \rightarrow +\infty$. In Section 4 we derive consistency results for the Fréchet mean in the case (1.2) of randomly shifted curves when n and J tend to infinity. In Section 5 and Section 6, 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 7 contains some numerical experiments. 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 $d_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^+$, then the empirical Fréchet mean $\bar{\mathbf{Y}}_{\mathcal{M}}$ of the sample $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ is defined as a minimizer (not necessarily unique) of

$$\bar{\mathbf{Y}}_{\mathcal{M}} \in \operatorname{argmin}_{y \in \mathcal{M}} \sum_{j=1}^J d_{\mathcal{M}}^2(y, \mathbf{Y}_j). \quad (2.1)$$

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 $\bar{\mathbf{Y}}_{\mathcal{M}} = \frac{1}{J} \sum_{j=1}^J \mathbf{Y}_j$. 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]. Consistent estimation of a mean shape in Σ_2^n has been studied by various authors, see e.g. [Le98], [KBCL99], [LK00], and in particular by [Goo91] and [KM97] using a model similar to (1.5) where the level of noise $\sigma = 0$.

More generally, a detailed study of some properties of the Fréchet mean (such as consistency and uniqueness) has been proposed in [Zie77], [OC95] and [BP03], [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 $(T_{\theta})_{\theta \in \mathcal{P}}$ is invertible in the sense that there exists a family of operators $(\tilde{T}_{\theta})_{\theta \in \mathcal{P}}$ such that for any $(\theta, f) \in \mathcal{P} \times L^2(\Omega)$

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

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

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

If $d_T^2(h, f) = 0$ then there exists $\theta \in \mathcal{P}$ such that $f = \tilde{T}_{\theta} h$ 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 let \mathcal{F} denote a subspace of $L^2(\Omega)$ and suppose that \hat{f}_j are smooth functions in $\mathcal{F} \subset L^2(\Omega)$ obtained from the data Y_j^{ℓ} , $\ell = 1, \dots, n$ for $j = 1, \dots, J$ (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

$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \frac{1}{J} \sum_{j=1}^J d_T^2(\hat{f}_j, f). \quad (2.2)$$

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

$$(\hat{\theta}_1, \dots, \hat{\theta}_J) = \underset{(\theta_1, \dots, \theta_J) \in \Theta^J}{\operatorname{argmin}} M(\theta_1, \dots, \theta_J), \quad (2.3)$$

where

$$M(\theta_1, \dots, \theta_J) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \hat{f}_j(t) - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \hat{f}_{j'}(t) \right)^2 dt, \quad (2.4)$$

which gives an estimation of the deformation parameters $\theta_1^*, \dots, \theta_J^*$, and then in a second step take

$$\hat{f}(t) = \frac{1}{J} \sum_{j=1}^J \tilde{T}_{\hat{\theta}_j} \hat{f}_j(t), \quad \text{for } t \in \Omega, \quad (2.5)$$

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 $\hat{\theta}_1, \dots, \hat{\theta}_J$ of the deformation parameters. The formulation (2.3) is thus an alternative solution to Procrustean methods and minimization of (2.3) can be done by a gradient-descent algorithm. The Generalize Procrustes Algorithms (GPA) used to compute a mean pattern (see e.g [DM98], [Goo91]) is based on an iterative algorithm which is an alternative scheme between computation of deformation parameters $\hat{\theta}_j$ and averaging of back-transformed curves or images using the inverse deformation operators $\tilde{T}_{\hat{\theta}_j}$. A similar approach to the template estimation problem that we propose in this paper has been developed by [JDJB04] in the context of biomedical images using diffeomorphic deformation operators

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 Lower bounds for the estimation of the deformation parameters

In this section, we derive non-asymptotic lower bounds for the quadratic risk of an arbitrary estimator of the deformation parameters. In what follows, C, C_0, C_1, C_2 , denote positive constants whose value may change from line to line. The notations $C(\cdot)$ specify the dependency of C on some quantities. Recall that the parameter space \mathcal{P} is a subset of \mathbb{R}^p and Θ is a compact subset of \mathcal{P} . We use a bold symbol $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \mathcal{P}^J \subset \mathbb{R}^{pJ}$ to denote a J -ads of vectors in \mathcal{P} , and the notation $\boldsymbol{\theta}_j = (\theta_j^1, \dots, \theta_j^p) \in \mathbb{R}^p$ to denote the components of the vector $\boldsymbol{\theta}_j \in \mathcal{P} \subset \mathbb{R}^p$. Note that, when context is clear, we also use the notation $\boldsymbol{\theta}$ to denote an element $\boldsymbol{\theta} = (\theta^1, \dots, \theta^p) \in \mathcal{P}$. The standard Euclidean norm of a vector \mathbf{c} in \mathbb{R}^k is denoted by $\|\mathbf{c}\|$, regardless of $k \in \mathbb{N}$. The L^2 -norm of a function $f \in L^2(\Omega)$ is denoted by $\|f\|_{L^2}^2 = \int_{\Omega} |f(t)|^2 dt$. Let us make the following assumption,

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

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

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

This assumption requires differentiability of the mapping $(\boldsymbol{\theta}, t) \mapsto T_{\boldsymbol{\theta}} f^*(t)$ with respect to each component θ^{p_1} , $p_1 = 1, \dots, p$ of the deformation parameters.

3.1 Shape Invariant Model

Theorem 3.1. *Consider the SIM (1.6) and suppose that Assumption 3.1 holds. Assume that the density g satisfies Assumption 1.1, and that $\int_{\Theta} \|\partial_{\boldsymbol{\theta}} \log(g(\boldsymbol{\theta}))\|^2 g(\boldsymbol{\theta}) d\boldsymbol{\theta} < +\infty$. Let $\hat{\boldsymbol{\theta}} \in \mathcal{P}^J$ be any estimator (a measurable function of the data) of $\boldsymbol{\theta}^*$. Then, for any $n \geq 1$ and $J \geq 1$,*

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

where $C(\Theta, f^*)$ is the constant defined in Assumption 3.1.

The right hand side of inequality (3.1) gives a lower bound on $\mathbb{E}[\frac{1}{J} \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2]$ that is independent of J . This bound does not decrease as J increases. Thus, if the number n of design points is fixed, increasing the number J of curves (or images) does not improve the quality of the estimation of the deformation parameters for any estimator $\hat{\boldsymbol{\theta}}$. Nevertheless, this lower bound is going to 0 as $n \rightarrow +\infty$. In Section 6.1, we will show that under suitable assumptions, the estimator $\hat{\boldsymbol{\theta}}$ defined equation 2.3 is consistent using a double asymptotic $n \rightarrow \infty$ and $J \rightarrow \infty$.

The fact that it is required to have an asymptotic setting in the number n of design points means that in model (1.6) a preliminary smoothing step is necessary to eliminate the additive error terms.

3.2 General model

Let us now consider the more general model (1.5) when $(T_{\theta})_{\theta \in \mathcal{P}}$ are linear operators from $L^2(\Omega)$ to $L^2(\Omega)$ satisfying Assumption 3.1. The main difference with (1.6) is the extra error terms $T_{\theta_j^*} Z_j$, $j = 1, \dots, J$. Recall that the Z_j 's are supposed to be i.i.d realization of a random process Z with value in $L^2(\Omega)$. In what follows, $\mathbb{E}_{\theta}[\cdot]$ denotes expectation conditionally to $\theta \in \Theta^J$. Since the random processes Z_j 's are observed through the action of the random deformation operators $T_{\theta_j^*}$ it is necessary to specify how the $T_{\theta_j^*}$'s modify the law of the process Z .

Assumption 3.2. Z is a second order Gaussian process with zero mean taking its values in $L^2(\Omega)$. Moreover, there exists a positive semi-definite symmetric $n \times n$ matrix $\Sigma_n(\Theta)$ such that the covariance matrix of $\mathbf{Z} = [Z(t_\ell)]_{\ell=1}^n$ satisfies $\mathbb{E}_{\theta}[\mathbf{T}_{\theta} \mathbf{Z} (\mathbf{T}_{\theta} \mathbf{Z})'] = \Sigma_n(\Theta)$. The smallest eigenvalue of $\Sigma_n(\Theta)$ is denoted by $s_n^2(\Theta)$.

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

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

$$\mathbb{E} \left[\frac{1}{J} \|\hat{\theta} - \theta^*\|^2 \right] \geq \frac{(\sigma^2 + s_n^2(\Theta))n^{-1}}{C(\Theta, f^*) + (\sigma^2 + s_n^2(\Theta))n^{-1} \int_{\Theta} \|\partial_{\theta_1} \log(g(\theta))\|^2 g(\theta) d\theta}. \quad (3.2)$$

where $C(\Theta, f^*)$ is the constant defined in Assumption 3.1.

Comments similar to those made after Theorem 3.1 can be made. Again, the main fact is that the lower bound (3.2) does not depends on J . Thus, increasing the number J of observations does not decrease the variance of any estimator of the deformations parameters.

3.3 Application to the shifted curves model

In this part, we apply Theorem 3.1 to the shifted curves model (1.2) with $t_\ell = \frac{\ell}{n}$ for all $\ell = 1, \dots, n$ (equispaced design). In this setting, the model (1.5) can be written as

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

As a corollary of Theorem 3.2, we have the following result,

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

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

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^* .

As remarked previously, the right hand side of inequality (3.4) does not depend on J . Therefore, if the number of design points n is fixed, then Theorem 3.3 shows that it is impossible to recover the true shifts in model (3.3) by letting only the number of curves J going to $+\infty$. On the other hand, the lower bound (3.4) tends to zero as $n \rightarrow +\infty$. This is consistent with the results presented in the next section where it is shown that, under extra assumptions, estimator (2.3) converges to $\boldsymbol{\theta}^*$ as $n \rightarrow +\infty$ and $J \rightarrow +\infty$ (see Theorem 4.1). An illustration of these facts is given in Section 7 on numerical experiments.

Since it is not possible to recover the true shifts, whatever is the estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}^*$, it is thus expected that the estimator \hat{f} (2.5) of the mean pattern f^* cannot be consistent. A result in this direction is given Theorem 4.3.

4 Consistent estimation in the shifted curves model

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

4.1 The random perturbations Z_j

Following the assumptions of Theorem 3.3, the random perturbation Z of the mean pattern f^* is supposed to be a second order stationary Gaussian process Z taking its value in $L_{per}^2([0, 1])$ with zero mean and covariance function $R : [0, 1] \rightarrow \mathbb{R}$. The law of Z is thus invariant by the action of a shift. For any random element $\boldsymbol{\theta}_j^* \in \Theta$, the vector $\mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j = [Z_j(\frac{\ell}{n} - \boldsymbol{\theta}_j^*)]_{\ell=1}^n$ is, conditionally to $\boldsymbol{\theta}_j^*$, a centered Gaussian random variable and its covariance matrix is a Toeplitz matrix equals to

$$\boldsymbol{\Sigma}_n = \mathbb{E}_{\boldsymbol{\theta}_j^*} [\mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j (\mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j)'] = \left[\mathbb{E} \left[Z \left(\frac{\ell}{n} \right) Z \left(\frac{\ell'}{n} \right) \right] \right]_{\ell, \ell'=1}^n = \left[R \left(\frac{|\ell - \ell'|}{n} \right) \right]_{\ell, \ell'=1}^n. \quad (4.1)$$

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

$$\gamma_{\max}(\boldsymbol{\Sigma}_n) \leq \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n |R \left(\frac{k}{n} \right)| = \gamma \quad (4.2)$$

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 $Z_j, j = 1, \dots, J$ in model (3.3).

4.2 Choice of the smoothing estimators \hat{f}_j

A convenient choice for the smoothing of the observed curves in (3.3) 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 $k = -(n-1)/2, \dots, (n-1)/2$ (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

$$\hat{f}_j^\lambda(t) = \sum_{|k| \leq \lambda} \hat{c}_{j,k} e^{i2\pi kt} = \frac{1}{n} \sum_{|k| \leq \lambda} \sum_{\ell=1}^n e^{-i2\pi \frac{\ell}{n} k} e^{i2\pi kt} Y_j^\ell = \frac{1}{n} \sum_{\ell=1}^n \sum_{|k| \leq \lambda} e^{i2\pi k(t - \frac{\ell}{n})} Y_j^\ell = \langle S_\lambda(t), \mathbf{Y}_j \rangle, \quad (4.3)$$

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 $f \in L_{per}^2([0, 1])$,

$$H_s(A) = \left\{ f \in L_{per}^2([0, 1]), \sum_{k \in \mathbb{Z}} (1 + |k|^2)^s |c_k(f)|^2 < A \right\}. \quad (4.4)$$

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

4.3 Identifiability conditions

Recall that Θ is a compact set of \mathbb{R} of the form $[-\rho, \rho]$ for some $\rho > 0$. Using low-pass filtering, the estimators of the random shifts $\theta_1^*, \dots, \theta_J^*$ are given by

$$(\hat{\theta}_1, \dots, \hat{\theta}_J) = \underset{(\theta_1, \dots, \theta_J) \in \Theta^J}{\operatorname{argmin}} M_\lambda(\theta_1, \dots, \theta_J), \quad (4.5)$$

where the criterion $M_\lambda(\theta) = M_\lambda(\theta_1, \dots, \theta_J)$ for $\theta \in \Theta^J$, see equation (2.4), 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

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

However, the minimization (4.5) is not well defined. Indeed, if $\theta_0 \in \Theta$ satisfies $\hat{\theta}_j + \theta_0 \in \Theta$, $j = 1, \dots, J$, then $(\hat{\theta}_1 + \theta_0, \dots, \hat{\theta}_J + \theta_0)$ is also a global minimum of $M_\lambda(\theta)$. 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 (3.3) is not identifiable (to see this, simply replace in (3.3) $f^*(\cdot)$ by $f^*(\cdot - \theta_0)$ and θ_j^* by $\theta_j^* + \theta_0$). Choosing identifiability conditions in model (3.3) amounts to impose constraints on the minimization of the criterion

$$D(\theta) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(f^*(t - \theta_j^* + \theta_j) - \frac{1}{J} \sum_{j'=1}^J f^*(t - \theta_{j'}^* + \theta_{j'}) \right)^2 dt. \quad (4.6)$$

This criterion $D(\theta)$ can be interpreted as a version without noise of the criterion $M_\lambda(\theta)$. Obviously, the criterion $D(\theta)$ has a minimum at $\theta^* = (\theta_1^*, \dots, \theta_J^*)$ such that $D(\theta^*) = 0$. However, the minimizer of D on Θ^J is clearly not unique, and minimizing $M_\lambda(\theta)$ does not allow to recover the true shifts θ^* nor the true mean pattern f^* . If the true shifts are supposed to have zero mean (i.e. $\int_{\Theta} \theta g(\theta) d\theta = 0$) it is natural to introduce the constrained set

$$\Theta_0 = \{(\theta_1, \dots, \theta_J) \in \Theta^J, \theta_1 + \dots + \theta_J = 0\} \quad (4.7)$$

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 $D(\theta)$ has a unique minimum on Θ_0 at

$$\theta_{\Theta_0}^* = (\theta_1^* - \bar{\theta}^*, \dots, \theta_J^* - \bar{\theta}^*) \text{ where } \bar{\theta}^* = \frac{1}{J} \sum_{j=1}^J \theta_j^*. \quad (4.8)$$

Under such assumptions, it is thus natural to compute an estimator $\hat{\theta}^\lambda$ of the random shifts over the constrained set Θ_0 defined as

$$\hat{\theta}^\lambda = (\hat{\theta}_1^\lambda, \dots, \hat{\theta}_J^\lambda) = \underset{(\theta_1, \dots, \theta_J) \in \Theta_0}{\operatorname{argmin}} M_\lambda(\theta_1, \dots, \theta_J). \quad (4.9)$$

Therefore, introduce the following identifiability conditions:

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

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

We then have the following result:

Proposition 4.1. *Suppose that Assumptions 4.2 and 4.1 hold with $\rho < 1/16$. Then, for any $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \in \Theta_0$, one has that*

$$D(\boldsymbol{\theta}) - D(\boldsymbol{\theta}_{\Theta_0}^*) \geq C(f^*, \rho) \frac{1}{J} \|\boldsymbol{\theta} - \boldsymbol{\theta}_{\Theta_0}^*\|^2,$$

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

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 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].

4.4 Consistent estimation of the random shifts

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

Theorem 4.1. *Consider the model (3.3) and let $\hat{\boldsymbol{\theta}}^\lambda$ be the estimator defined by (4.9). Assume that $\mathcal{F} = H_s(A)$ for some $A > 0$ and $s \geq 1$ and that Z is a centered stationary process with value in $L_{per}^2([0, 1])$ and covariance function $R : [0, 1] \rightarrow \mathbb{R}$. Suppose that Assumptions 4.2 and 4.1 hold with $\rho < 1/16$. Then, for any $\lambda \geq 1$ and $x > 0$*

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

with

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

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 Θ, \mathcal{F}, f^* , $\|\cdot\|$ is the standard Euclidean norm in \mathbb{R}^J ,

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

Let us make some comments on the deviation inequality in Theorem 4.1. The result follows from the inequality,

$$\frac{1}{J} \|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*\|^2 \leq \frac{2}{J} \|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_{\Theta_0}^*\|^2 + \frac{2}{J} \|\boldsymbol{\theta}_{\Theta_0}^* - \boldsymbol{\theta}^*\|^2.$$

The term $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ comes from the control of the quantity $\frac{2}{J} \|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_{\Theta_0}^*\|^2$, while the term $A_2(x, J)$ is derived from a Bernstein's inequality on $\frac{2}{J} \|\boldsymbol{\theta}_{\Theta_0}^* - \boldsymbol{\theta}^*\|^2 = 2 \left| \frac{1}{J} \sum_{j=1}^J \boldsymbol{\theta}_j^* \right|^2$ (see the Appendix for further details). Remark that for fixed values of the number of design points n and the spectral cutoff λ , then $\lim_{J \rightarrow +\infty} A_2(x, J) = 0$. This means that for J sufficiently large then $\bar{\boldsymbol{\theta}}^* = \frac{1}{J} \sum_{j=1}^J \boldsymbol{\theta}_j^* \approx 0$ and thus $\boldsymbol{\theta}_{\Theta_0}^* \approx \boldsymbol{\theta}^*$. Hence, the convergence in probability of $\boldsymbol{\theta}_{\Theta_0}^*$ 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 $v(x, J, n, \lambda)$ 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 $\lim_{n \rightarrow +\infty} \lambda_n = +\infty$ and

$\lim_{n \rightarrow +\infty} \frac{\lambda_n}{n} = 0$, 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 $\lim_{n \rightarrow +\infty} B(\lambda_n, n) = \lim_{n \rightarrow +\infty} v(x, J, n, \lambda_n) = 0$, which implies that $\lim_{n \rightarrow +\infty} A_1(x, J, n, \lambda_n, \sigma^2, \gamma) = 0$. However, if n remains fixed, then $\lim_{J \rightarrow +\infty} A_1(x, J, n, \lambda, \sigma^2, \gamma) = h(\frac{2\lambda+1}{n}) + h(\frac{2\lambda+1}{n} + \lambda^{-2s})$, 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. Thus Theorem 4.1 is consistent with conclusions of Theorem 3.3, that is, if the number n of design points is fixed, then it is in general not possible to estimate θ^* as the number of curves J grows to infinity.

Hence, under the assumptions of Theorem 4.1, one can only prove the convergence in probability of $\hat{\theta}^\lambda$ to the true shifts θ^* by taking a double asymptotic setting by letting n and J going to $+\infty$ (provided the smoothing parameter $\lambda = \lambda_n$ is well chosen). Nevertheless, if J is fixed, $\hat{\theta}^\lambda$ converges in probability to $\theta_{\Theta_0}^*$ as n goes to infinity.

4.5 Consistent estimation of the mean pattern

The Fréchet mean estimator of f^* defined by (2.5) can now be written as

$$\hat{f}^\lambda(t) = \frac{1}{J} \sum_{j=1}^J \hat{f}_j^\lambda(t + \hat{\theta}_j^\lambda) = \frac{1}{J} \sum_{j=1}^J \langle S_\lambda(t + \hat{\theta}_j^\lambda), \mathbf{Y}_j \rangle, \quad \text{for all } t \in [0, 1]. \quad (4.10)$$

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

Theorem 4.2. *Under the assumptions of Theorem 4.1, 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 4.1, $C_2(\Theta, \mathcal{F}, f^*)$ and $C_3(\Theta, f^*)$ are positive constants depending only on Θ, \mathcal{F}, f^* , and $\|\hat{f}^\lambda - f^*\|_{L^2}^2 = \int_0^1 |f^\lambda(t) - f^*(t)|^2 dt$.

The proof is given in the Appendix. The terms appearing in the deviation inequality given in Theorem 4.2 are the same as those appearing in Theorem 4.1. Therefore, similar comments to those made on the consistency of the estimators of the shifts can be made. Consider the function $f_{\Theta_0}^*(t) := \frac{1}{J} \sum_{j=1}^J f^*(t - \theta_j^* + [\theta_{\Theta_0}^*]_j) = f^*(t - \bar{\theta}^*)$ for $t \in [0, 1]$, where $\theta_{\Theta_0}^* = ([\theta_{\Theta_0}^*]_1, \dots, [\theta_{\Theta_0}^*]_J)$ and the following inequality $\|\hat{f}^\lambda - f^*\|_{L^2}^2 \leq 2\|\hat{f}^\lambda - f_{\Theta_0}^*\|_{L^2}^2 + 2\|f_{\Theta_0}^* - f^*\|_{L^2}^2$. Again, $A_1(x, J, n, \lambda, \sigma^2, \gamma)$ controls the quantity $\|\hat{f}^\lambda - f_{\Theta_0}^*\|_{L^2}^2$ while $A_2(x, J)$ controls $\|f_{\Theta_0}^* - f^*\|_{L^2}^2$. Thus, a double asymptotic in n and J is needed to show that the Fréchet mean \hat{f}^λ converges in probability to the true mean pattern f^* . If J remains fixed, the Fréchet mean converges to the shifted version $f_{\Theta_0}^*$ of f^* as $n \rightarrow +\infty$.

4.6 A lower bound for the variance of the Fréchet mean

When the number of design points n is fixed and we only let J going to infinity, Theorem 3.3 shows that it is not possible to recover the true shifts. It is thus expected that the Fréchet mean \hat{f}^λ should not converge to the mean pattern f^* in the setting n fixed and $J \rightarrow +\infty$. To support this argument, consider the following ideal estimator

$$\tilde{f}(t) = \frac{1}{J} \sum_{j=1}^J f_j^*(t + \hat{\theta}_j^\lambda) = \frac{1}{J} \sum_{j=1}^J f^*(t - \theta_j^* + \hat{\theta}_j^\lambda), \quad \text{for all } t \in [0, 1], \quad (4.11)$$

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

Theorem 4.3. *Suppose that the assumptions of Theorem 3.3 are satisfied with, in Assumption 1.1, $\rho < \frac{3}{4\pi}$. Then, for any $n \geq 1$, there exists $J_0 \in \mathbb{N}$ such that $J \geq J_0$ implies*

$$\mathbb{E}[\|\tilde{f} - f^*\|_{L^2}] \geq C(f^*, \rho) \frac{n^{-1}\sigma^2}{\|\partial_t f^*\|_\infty^2 + n^{-1}\sigma^2 \int_\Theta (\partial_\theta \log(g(\theta)))^2 g(\theta) d\theta}, \quad (4.12)$$

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

The proof is postponed part 8.4. As for Theorem 3.3 the right hand side of inequality (4.12) does not depend on J . Thus, in the setting n fixed and $J \rightarrow +\infty$, 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}^λ , 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 θ_j^* , $j = 1, \dots, J$ is limited by the variance σ^2 of the additive Gaussian noise in model (3.3).

4.7 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 $\hat{\theta}^\lambda$ and \hat{f}^λ in the following diagram,

$$\begin{array}{cccc} \hat{\theta}^\lambda \xrightarrow[n, J]{} \theta^* & \hat{\theta}^\lambda \not\xrightarrow[J]{} \theta^* \text{ (} n \text{ fixed)} & \hat{\theta}^\lambda \xrightarrow[n]{} \theta_{\Theta_0}^* \text{ (} J \text{ fixed)} & \hat{\theta}^\lambda \not\xrightarrow[J]{} \theta_{\Theta_0}^* \text{ (} n \text{ fixed)} \\ \hat{f}^\lambda \xrightarrow[n, J]{} f^* & \hat{f}^\lambda \xrightarrow[n]{} f_{\Theta_0}^* \text{ (} J \text{ fixed)} & & \end{array}$$

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 .

5 Notations and main assumptions in the general case

In this section, we define a rather general framework to ensure the consistency of the empirical Fréchet mean. First, 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).

5.1 Smoothness of the mean pattern and the deformation operators

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

Assumption 5.1 (Continuity of the deformation operators). *For all $\theta \in \mathcal{P}$, $\mathcal{L}_\theta : L^2(\Omega) \rightarrow L^2(\Omega)$ is a linear operator satisfying $\mathcal{L}_\theta f \in \mathcal{F}$ for all $f \in \mathcal{F}$. There exists a constant $C(\theta) > 0$ such that for any $f \in L^2(\Omega)$ and $\theta \in \Theta$*

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

and a constant $C(\mathcal{F}, \Theta) > 0$ such that for any $f \in \mathcal{F}$ and $\theta_1, \theta_2 \in \Theta$,

$$\|\tilde{T}_{\theta_1} f - \tilde{T}_{\theta_2} f\|_{L^2}^2 \leq C(\mathcal{F}, \Theta) \|\theta_1 - \theta_2\|^2.$$

Assumption 5.1 can be interpreted as a Lipschitz condition on the mapping $(f, \boldsymbol{\theta}) \mapsto \mathcal{L}_{\boldsymbol{\theta}}f$. The first inequality, that is $\|\mathcal{L}_{\boldsymbol{\theta}}f\|_{L^2}^2 \leq C(\boldsymbol{\Theta}) \|f\|_{L^2}^2$, 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 (2.3) to have non trivial solutions by avoiding the functional $M(\boldsymbol{\theta})$ in (2.4) 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 $t \in \Omega$ and $\boldsymbol{\theta} \in \mathcal{P} = \mathbb{R}$. 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 5.1, a convenient choice is $\Theta = [-\rho, \rho]$ for $\rho > 0$ as in Assumption 1.1. Suppose now that we do not restrict to a compact set and we have $\Theta = \mathbb{R}$. Since $M(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) \leq \frac{1}{J} \sum_{j=1}^J e^{-2\boldsymbol{\theta}_j} \|\hat{f}_j\|_{L^2}^2$, one could let the $\boldsymbol{\theta}_j$'s going to $+\infty$ and the optimization problem (2.4) would have a trivial solution at $\hat{\boldsymbol{\theta}}_j = +\infty, j = 1, \dots, J$ such that $M(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J) = 0$. A particular case where the first inequality of Assumption 5.1 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 this assumption 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 $L_{per}^2([0, 1])$, as $\|\mathcal{L}_{\boldsymbol{\theta}}f\|_{L^2}^2 = \|f\|_{L^2}^2$ for all $\boldsymbol{\theta} \in \mathbb{R}$. To verify the second inequality of Assumption 5.1, take as in Section 4

$$\mathcal{F} = H_s(A) \text{ with } s \geq 1, \text{ see equation (4.4).}$$

Then, remark that a change of variable gives for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, $\|\tilde{T}_{\boldsymbol{\theta}_1}f - \tilde{T}_{\boldsymbol{\theta}_2}f\|_{L^2}^2 = \|f(\cdot + \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) - f\|_{L^2}^2$. Since $f \in H_s(A)$, the mean value theorem can be used to prove that $\|\tilde{T}_{\boldsymbol{\theta}_1}f - \tilde{T}_{\boldsymbol{\theta}_2}f\|_{L^2}^2 \leq C(A, \Theta)\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^2$ for some constant $C(A, \Theta) > 0$, and thus Assumption 5.1 holds.

5.2 The preliminary smoothing step

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

Definition 5.1. 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 $S_{\lambda} : \Omega \rightarrow \mathbb{R}^n$ such that for all $j = 1, \dots, J$ and all $t \in \Omega$

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

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

Assumption 5.2. For all $\lambda \in \Lambda$ and all $\ell = 1, \dots, n$, the function $t \mapsto S_{\lambda}^{\ell}(t)$ belong to $L^2(\Omega)$, where $S_{\lambda}^{\ell}(t)$ denotes the ℓ -th component of the vector $S_{\lambda}(t)$. Moreover, for all $\lambda \in \Lambda$, $f \in \mathcal{F}$ and $\boldsymbol{\theta} \in \Theta$, the function $t \mapsto \langle S_{\lambda}(t), \mathbf{T}_{\boldsymbol{\theta}}\mathbf{f} \rangle$ belongs to \mathcal{F} where $\mathbf{T}_{\boldsymbol{\theta}}\mathbf{f} = (T_{\boldsymbol{\theta}}f(t_{\ell}))_{\ell=1}^n$.

Note that it follows from the above assumption, that the functions $t \mapsto \hat{f}_j(t) = \langle S_{\lambda}(t), \mathbf{Y}_j \rangle$, $j = 1, \dots, J$ 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 4. 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}_j^λ depends on the smoothing parameter λ . For this, consider for some function $f \in \mathcal{F}$ the following regression model

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

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

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

where B_λ and V_λ denote the usual bias and variance of \hat{f}^λ given by $B_\lambda(f, t) = \langle S_\lambda(t), \mathbf{f} \rangle - f(t)$ and $V_\lambda(t) = \|S_\lambda(t)\|_{\mathbb{R}^n}^2$, for $t \in \Omega$, where $\mathbf{f} = (f(t_\ell))_{\ell=1}^n$. 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}^λ :

Assumption 5.3. *There exist a constant $\kappa(\mathcal{F}) > 0$ and a real-valued functions $\lambda \mapsto B(\lambda)$, such that for all $f \in \mathcal{F}$, $\|B_\lambda(f, \cdot)\|_{L^2}^2 = \|\langle S_\lambda(\cdot), \mathbf{f} \rangle - f(\cdot)\|_{L^2}^2 \leq \kappa(\mathcal{F})B(\lambda)$. Moreover there exists a sequence of smoothing parameters $(\lambda_n)_{n \in \mathbb{N}} \in \Lambda^{\mathbb{N}}$ with $\lim_{n \rightarrow +\infty} \lambda_n = +\infty$ such that $\lim_{n \rightarrow +\infty} B(\lambda_n) = 0$ and $\lim_{n \rightarrow +\infty} V(\lambda_n) = 0$.*

Let us illustrate Assumption 5.3 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 4, take $\mathcal{F} = H_s(A)$ defined in (4.4). In this setting, $V(\lambda) = \frac{2\lambda+1}{n}$ and $\|B_\lambda(f, \cdot)\|_{L^2}^2 = \sum_{|k| \leq \lambda} |\tilde{c}_k(f) - c_k(f)|^2 + \sum_{|k| > \lambda} |c_k(f)|^2$ where $\tilde{c}_k(f) = \frac{1}{n} \sum_{\ell=1}^n f(\frac{\ell}{n}) e^{-i2\pi k \frac{\ell}{n}}$. Therefore, $\|B_\lambda(f, \cdot)\|_{L^2}^2 \leq C(A)B(\lambda)$ for some positive constant $C(A)$ depending only on A , and $B(\lambda) = \frac{2\lambda+1}{n} + \lambda^{-2s}$. Thus, Assumption 5.3 holds with $\lambda_n = n^{\frac{1}{2s+1}}$.

5.3 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 $T_{\theta_j}, j = 1, \dots, J$. Therefore, to study the consistency of the estimators, it is important to specify how the action of T_{θ_j} modifies the law of the process Z_j . In particular, the action of the deformation operator $T_{\theta_j} Z_j(t_\ell)$ and $T_{\theta_j} Z_j(t_{\ell'})$ 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 5.4. *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 $\theta \in \Theta$*

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

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

$$\lim_{n \rightarrow \infty} \gamma_n(\Theta) \sqrt{V(\lambda_n)} = 0, \quad (5.1)$$

where $V(\lambda_n)$ is the variance defined in Assumption 5.3.

Intuitively, the condition (5.1) means that the variance of the linear smoother $S_\lambda(\cdot)$ has to be asymptotically smaller than the maximal correlations (measured by $\gamma_n(\Theta)$) between $T_\theta Z(t_\ell)$ and $T_\theta Z(t_{\ell'})$ for $\ell, \ell' = 1, \dots, n$ and all $\theta \in \Theta$. In the case of randomly shifted curves with an equi-spaced

design, a simple condition for which Assumption 5.4 holds is the case where Z is stationary process. Recall that in Section 4.1, we have introduced the covariance matrix $\Sigma_n = [\mathbb{E}[Z(t_\ell)Z(t_{\ell'})]]_{\ell, \ell'=1}^n = [R(|t_\ell - t_{\ell'}|)]_{\ell, \ell'=1}^n$. If the design is equi-spaced, then the first assertion of Assumption 5.4 holds with $\gamma_n(\Theta) = \gamma = \int_0^1 |R(t)| dt$. Thus, $\lim_{n \rightarrow \infty} \gamma_n \sqrt{V(\lambda_n)} = \lim_{n \rightarrow \infty} \sqrt{V(\lambda_n)} = 0$, which proves that Assumption 5.4 is satisfied.

5.4 Identifiability conditions

Recall that it has been discussed in Section 4.3 that without any further assumptions on the smoothness class \mathcal{F} and on the density g of the random shifts, the model (3.3) is not identifiable. Under Assumption 4.1 and 4.2, we had to constrain the minimization problem (4.5) on a subset Θ_0 of Θ^J to ensure the uniqueness of the minimum of the criterion $D(\theta)$ (4.6) on Θ_0 . In the case of general deformation operators, the criterion (4.6) can be defined for all $\theta = (\theta_1, \dots, \theta_J) \in \Theta^J$ by

$$D(\theta) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} T_{\theta_j^*} f^*(t) - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} T_{\theta_{j'}^*} f^*(t) \right)^2 dt. \quad (5.2)$$

The criterion $D(\theta)$ is a version of $M(\theta)$ defined in (2.4) with the ideal estimators $\hat{f}_j = T_{\theta_j^*} f^*$, $j = 1, \dots, J$. Obviously, using that for all $\theta \in \Theta$, $\tilde{T}_{\theta} T_{\theta} f^* = f^*$, the criterion $D(\theta)$ has a minimum at $\theta^* = (\theta_1^*, \dots, \theta_J^*)$ such that $D(\theta^*) = 0$. However, without any further restrictions the minimizer of $D(\theta)$ is not necessarily unique on Θ^J . Let us make the following Assumption,

Assumption 5.5. *Let $\Theta \subset \Theta^J$ such that there exists a unique $\theta_{\Theta}^* \in \Theta$ satisfying $D(\theta_{\Theta}^*) = 0$.*

Then, Θ is the set onto which we will carry the minimization of the criterion $M(\theta)$ (2.4). Note that we can eventually have $\Theta = \Theta^J$.

In the case of shifted curves with Assumption 4.1 and 4.2, it is shown in [BG10] (proof of Lemma 6) that the only set where criterion D vanishes is the line $\{\theta^* + \theta_0 \mathbf{1}_J, \theta_0 \in \mathbb{R}\} \subset \mathbb{R}^J$, where $\mathbf{1}_J = (1, \dots, 1)' \in \mathbb{R}^J$. An easy way to choose the set Θ is to take a linear subset of Θ^J , see Figure 1 for an illustration. By considering the subset

$$\Theta_0 = \Theta^J \cap \mathbf{1}_J^\perp = \{(\theta_1, \dots, \theta_J) \in \Theta^J, \theta_1 + \dots + \theta_J = 0\},$$

where $\mathbf{1}_J^\perp$ is the orthogonal of $\mathbf{1}_J$ in \mathbb{R}^J , then Assumption 5.5 is satisfied with $\theta_{\Theta_0}^*$ given in (4.8). More generally, if the deformation parameters θ_j , $j = 1, \dots, J$ are supposed to be random variables with zero mean, then optimizing $D(\theta)$ on Θ_0 is a natural choice. Another identifiability condition for shifted curves is proposed in [GLM07] and [Vim10] by taking

$$\Theta_1 = \Theta^J \cap e_1^\perp = \{(\theta_1, \dots, \theta_J) \in \Theta^J, \theta_1 = 0\}. \quad (5.3)$$

where $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^J$. In this case, $\theta_{\Theta_1}^* = (0, \theta_2^* - \theta_1^*, \dots, \theta_J^* - \theta_1^*)$. Choosing to minimize $D(\theta)$ on Θ_1 amounts to choose the first curve as a reference onto which all the others curves are aligned, meaning that the first shift θ_1^* is not random.

Following the classical guidelines in M-estimation (see e.g. [vdV98]), a necessary condition to ensure the convergence of M-estimators such as (2.3) is that the local minima of $D(\theta)$ over Θ are well separated from the global minimum of $D(\theta)$ at $\theta = \theta_{\Theta}^*$ (satisfying $D(\theta_{\Theta}^*) = 0$). The following assumption can be interpreted in this sense.

Assumption 5.6. *For all $\theta \in \Theta$ we have*

$$D(\theta) - D(\theta_{\Theta}^*) \geq C(\Theta, \mathcal{F}) \frac{1}{J} \|\theta - \theta_{\Theta}^*\|^2 \quad (5.4)$$

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

In the shifted curve model, Assumption 5.6 is verified if Assumption 4.1 and 4.2 hold (see Proposition 4.1).

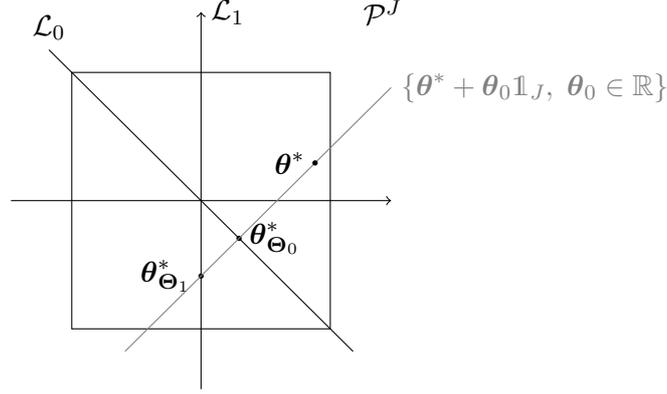


Figure 1: Choice of identifiability conditions for shifted curves in the case $J = 2$.

6 Consistency in the general case

In this section, we give the main results of consistency in the general model (1.5). All the assumptions and notations used in this section have been defined in Section 5.

6.1 Consistent estimation of the deformation parameters

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

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

where

$$M_\lambda(\theta) = \frac{1}{J} \sum_{j=1}^J \int_{\Omega} \left(\tilde{T}_{\theta_j} \langle S_\lambda(t), \mathbf{Y}_j \rangle - \frac{1}{J} \sum_{j'=1}^J \tilde{T}_{\theta_{j'}} \langle S_\lambda(t), \mathbf{Y}_{j'} \rangle \right)^2 dt, \quad (6.1)$$

and Θ is the constrained set introduced in Assumption 5.5. The estimator $\hat{\theta}^\lambda$ thus depends on the choice of Θ , and it will be shown that $\hat{\theta}^\lambda$ is a consistent estimator of the vector $\theta_{\Theta}^* \in \mathbb{R}^{pJ}$ defined in Assumption 5.5. Note that depending on the problem at hand and the choice of the constrained set Θ , it can be shown that θ_{Θ}^* is close to the true deformation parameters θ^* . For example, in the case of shifted curves, if $\Theta = \Theta_0$ defined in (4.7) and if the density g of the shifts has zero mean, then $\theta_{\Theta_0}^* = (\theta_1^* - \bar{\theta}^*, \dots, \theta_J^* - \bar{\theta}^*)$ with $\bar{\theta}^* = \frac{1}{J} \sum_{j=1}^J \theta_j^*$ can be shown to be close to θ^* (see Lemma 9.1 in the Appendix). This allows to show the consistency of $\hat{\theta}^\lambda$ to θ^* as formulated in Theorem 4.1. Therefore, the next result only bounds the distance between $\hat{\theta}^\lambda$ and θ_{Θ}^* .

Theorem 6.1. *Consider the model (1.5) and suppose that Assumptions 1.1 and 5.1 to 5.6 hold with $n \geq 1$ and $J \geq 2$. Then, for any $\lambda \in \Lambda$ and $x > 0$*

$$\mathbb{P} \left(\frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta}^*\|^2 \geq C_1(\Theta, \Theta, \mathcal{F}, f^*) \left[(\gamma_n(\Theta) + \sigma^2) \left(\sqrt{v(x, J, \lambda)} + v(x, J, \lambda) \right) + \left(\sqrt{B(\lambda)} + B(\lambda) \right) \right] \right) \leq 2e^{-x}, \quad (6.2)$$

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

Let us comment the inequality (6.2). The term $v(x, J, \lambda)$ depends on the variance $V(\lambda)$ of the linear estimators \hat{f}_j^λ , $j = 1, \dots, J$, and recall that $B(\lambda)$ is a uniform control on the bias of these estimators

(see Section 5.2). Using a classical tradeoff between variance and bias, Assumption 5.3 implies that there exists a sequence of smoothing parameters λ_n such that $\lim_{n \rightarrow +\infty} V(\lambda_n) = \lim_{n \rightarrow +\infty} B(\lambda_n) = 0$. Moreover, using Assumption 5.4, it follows that $\lim_{n \rightarrow +\infty} \gamma_n(\Theta) \left(\sqrt{v(x, J, \lambda_n)} + v(x, J, \lambda_n) \right) = 0$ for any $x > 0$ and $J \geq 2$. If the number of curves J remains fixed, Theorem 6.1 implies that $\hat{\theta}^\lambda$ converges in probability to θ_{Θ}^* as the number of design points n increases. Such 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 $\lambda \in \Lambda$, $\lim_{J \rightarrow +\infty} v(x, J, \lambda) = V(\lambda)$. Therefore, Theorem 6.1 cannot be used to prove that $\hat{\theta}^\lambda$ converges to θ_{Θ}^* as J tends to infinity. To the contrary one has that

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

where $h(u) = \sqrt{u} + u$ for $u \geq 0$. This confirms that $\hat{\theta}^\lambda$ is not a consistent estimator of θ_{Θ}^* (and thus of θ^*) as n remains fixed and J tends to infinity. This fact has already been noticed in Section 4 (see Theorem 3.3).

6.2 Consistent estimation of the mean pattern

From equation (2.5), recall that an estimator \hat{f}^λ of the mean pattern f^* is defined as

$$\hat{f}^\lambda = \frac{1}{J} \sum_{j=1}^J \tilde{T}_{\hat{\theta}_j^\lambda} \hat{f}_j^\lambda = \frac{1}{J} \sum_{j=1}^J \tilde{T}_{\hat{\theta}_j^\lambda} \langle S_\lambda(\cdot), \mathbf{Y}_j \rangle. \quad (6.3)$$

We study the consistency of \hat{f}^λ with respect to the shape function $f_{\Theta}^* := \frac{1}{J} \sum_{j=1}^J \tilde{T}_{[\theta_{\Theta}^*]_j} T_{\theta_j^*} f^*$, defined for $\theta_{\Theta}^* = ([\theta_{\Theta}^*]_1, \dots, [\theta_{\Theta}^*]_J)$. Again, depending on the problem at hand and the choice of the constrained set Θ , it can be shown that f_{Θ}^* is close to the true mean pattern f^* . For example, in the case of shifted curves with $\Theta = \Theta_0$ defined in (4.7), then $\theta_{\Theta_0} = (\theta_1^* - \bar{\theta}^*, \dots, \theta_J^* - \bar{\theta}^*)$ with $\bar{\theta}^* = \frac{1}{J} \sum_{j=1}^J \theta_j^*$. In this case

$$f_{\Theta_0}^*(t) := \frac{1}{J} \sum_{j=1}^J f^*(t - \theta_j^* + [\theta_{\Theta_0}^*]_j) = f^*(t - \bar{\theta}^*) \text{ for all } t \in [0, 1].$$

Hence, under the condition that $\int_{\Theta} \theta g(\theta) d\theta = 0$ (the shifts have zero mean), then $\bar{\theta}^* \approx 0$ for J sufficiently large, and thus $f_{\Theta}^*(t)$ is close to f^* which allows to show the consistency of \hat{f}^λ to f^* as formulated in Theorem 4.2.

Theorem 6.2. *Consider the model (1.5) and suppose that Assumptions 1.1 and 5.1 to 5.6 hold. Then, for any $\lambda \in \Lambda$ and $x > 0$*

$$\mathbb{P} \left(\|\hat{f}^\lambda - f_{\Theta}^*\|_{L^2}^2 \geq C_2(\Theta, \Theta, \mathcal{F}, f^*) \left[(\gamma_n(\Theta) + \sigma^2) \left(\sqrt{v(x, J, \lambda)} + v(x, J, \lambda) \right) + \left(\sqrt{B(\lambda)} + B(\lambda) \right) \right] \right) \leq 2e^{-x}, \quad (6.4)$$

where $C_2(\Theta, \Theta, \mathcal{F}, f^*) > 0$ is a constant depending only Θ , 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 (6.2) and (6.4). The consistency of \hat{f}^λ to f_{Θ}^* is guaranteed when n goes to infinity provided the level of smoothing $\lambda = \lambda_n$ is chosen so that $\lim_{n \rightarrow +\infty} V(\lambda_n) = \lim_{n \rightarrow +\infty} B(\lambda_n) = 0$. Again, if n remains fixed and only J is let going to infinity then Theorem 6.2 cannot be used to prove the convergence of \hat{f}^λ to f_{Θ}^* .

7 Numerical experiments

7.1 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 (6.1). 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 $\boldsymbol{\theta}^0 \in \Theta$ (if $\Theta = \Theta_0$ 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 know 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 Θ , it is possible to impose that $\boldsymbol{\theta}^m$ satisfies the same constraints. For instance, if $\Theta = \Theta_0$ one can modify $\boldsymbol{\theta}^m = (\boldsymbol{\theta}_j^m)_{j=1, \dots, J}$ at each iteration such that $\sum_{j=1}^J \boldsymbol{\theta}_j^m = 0$.

7.2 Randomly shifted curves

We present some simulations to illustrate the results given in Section 4. 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 $\mathcal{F} = H_3(A) \subset L_{per}^2([0, 1])$. 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 (3.3) is

$$Y_j^\ell = f^*(\frac{\ell}{n} - \boldsymbol{\theta}_j^*) + Z_j(\frac{\ell}{n} - \boldsymbol{\theta}_j^*) + \sigma \varepsilon_j^\ell, \quad \text{and } j = 1, \dots, J, \ell = 1, \dots, n \quad (7.1)$$

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 2(a). For the constrained set we took

$$\Theta_0 = \left\{ \boldsymbol{\theta} \in \left[-\frac{1}{2}, \frac{1}{2}\right]^J, \boldsymbol{\theta}_1 + \dots + \boldsymbol{\theta}_J = 0 \right\}.$$

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 additive term Z . Figure 2(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

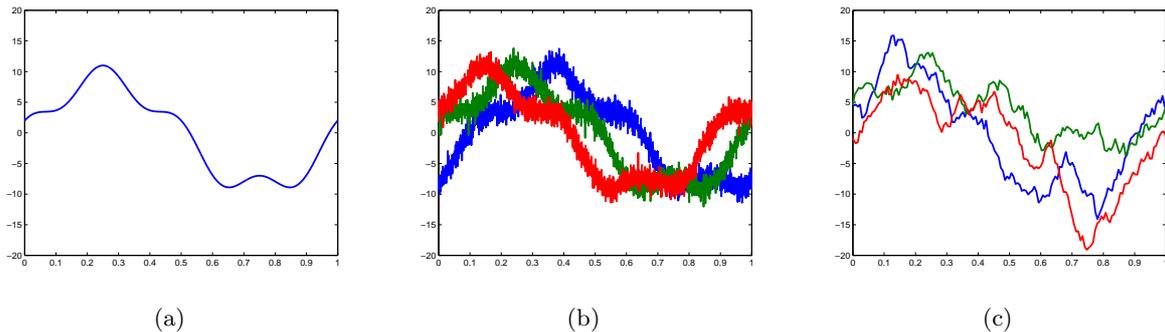


Figure 2: (a) mean pattern f^* . (b) $J = 3$ noisy curves in the SIM with $\sigma = 2$. (c) $J = 3$ noisy curves with $\sigma = 0$ and a stationary process Z with $\zeta = 4$.

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 (7.1) with $Z_j = 0, j = 1, \dots, J$. For each repetition we computed $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*\|^2$ and $\|\hat{f}^\lambda - f^*\|_{L^2}^2$. Boxplot of these quantities are displayed in Figure 3(a) and 3(b) respectively, for $J = 20, 40, \dots, 100$ and $n = 512$ (in gray) and $n = 1024$ (in black). As the smoothing parameter is fixed to $\lambda = 7$, increasing n simply reduces the variance of the linear smoothers \hat{f}_j^λ . Recall that the lower bound given in Theorem 3.3 shows that $\frac{1}{J}\mathbb{E}[\|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}^\lambda\|^2]$ does not decrease as J increases but should be smaller when the number of point n increases. This is exactly what we observe in Figure 3. 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 $\|\hat{f}^\lambda - f^*\|_{L^2}^2$ is clearly smaller with $n = 1024$ than $n = 512$. One can also see that the variance and the mean of $\|\hat{f}^\lambda - f^*\|_{L^2}^2$ 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}^λ 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}^λ is close to the ideal estimator \tilde{f}^λ that would be obtained if we knew the true values of the shifts that is $\tilde{f}^\lambda(t) = \frac{1}{J}\sum_{j=1}^J \tilde{T}_{\boldsymbol{\theta}_j^*} \hat{f}_j^\lambda(t) = \frac{1}{J}\sum_{j=1}^J \hat{f}_j^\lambda(t + \boldsymbol{\theta}_j^*)$. It is clear from model (7.1) that \tilde{f}^λ can be shown to be consistent estimator of f^* at the design points as J tends to infinity and n remains fixed, provided that $\lambda = \lambda_J \rightarrow +\infty$ is chosen in an appropriate way. This explains why, in numerical experiments, the quantity $\|\hat{f}^\lambda - f^*\|_{L^2}^2$ is decreasing with J .

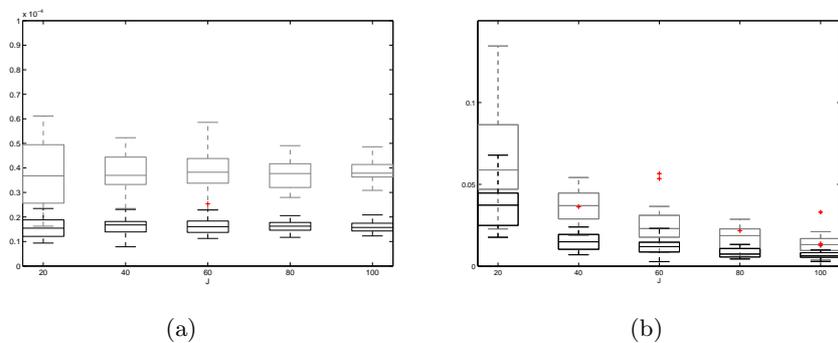


Figure 3: Boxplot of $\frac{1}{J}\|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}^*\|^2$ (a) and $\|\hat{f}^\lambda - f^*\|_{L^2}^2$ (b) over $M = 20$ repetitions from a SIM model of shifted curves. Boxplot in gray correspond to $n = 512$, and in black to $n = 1024$.

Complete model

We now add the terms Z_j in (7.1) 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 2(c) we plot $f^* + Z_j$, $j = 1, 2, 3$ with $\varsigma = \phi = 4$. Over $M = 20$ repetitions, we have computed the values of $\frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ and $\|\hat{f}^\lambda - f_{\Theta_0}^*\|_{L^2}^2$ for J is varying from 20 to 100 and $n = 512, 1024$. The results are displayed in Figure 4(a) and 4(b). We observe the same behaviors than in the simulations with the SIM model: the variance of $\frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ does not decrease as J increases (see Figure 4(a)) and $\|\hat{f}^\lambda - f_{\Theta_0}^*\|^2$ 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 .

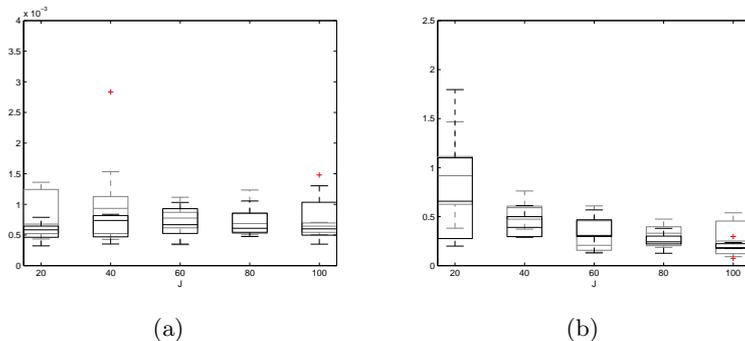


Figure 4: Boxplot of $\frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ (a) and $\frac{1}{J} \|\hat{f}^\lambda - f_{\Theta_0}^*\|^2$ (b) in model (3.3) with a stationary error term Z . Boxplot in gray correspond to $n = 512$, and in black to $n = 1024$.

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 5.4 is not verified. The levels of noise (σ and ς) are the same than in the stationary case in order to make things comparable. The results are presented in the same manner in Figure 5(a) for $\frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ and in Figure 5(b) for $\|\hat{f}^\lambda - f_{\Theta_0}^*\|_{L^2}^2$. 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 $\frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2$ remains rather high even when n or J increases (see Figure 5(a)). The convergence to zero of $\|\hat{f}^\lambda - f_{\Theta_0}^*\|_{L^2}^2$ which was clear in the stationary case, is now not so obvious in view of the numerical results displayed in Figure 5(b).

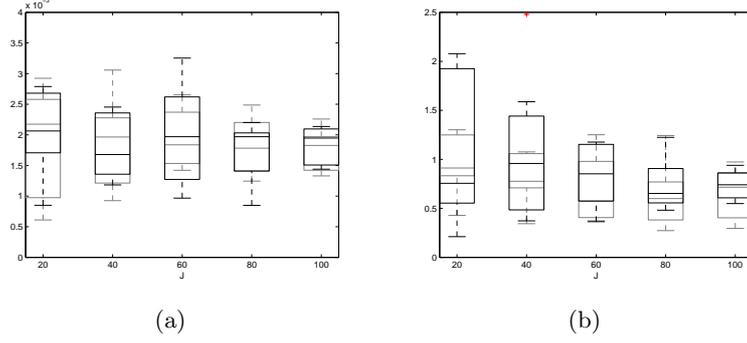


Figure 5: Boxplot of $\frac{1}{J} \|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_{\Theta_0}^*\|^2$ (a) and $\frac{1}{J} \|\hat{f}^\lambda - f_{\Theta_0}^*\|^2$ (b) in model (3.3) with a non-stationary error term Z . Boxplot in grey correspond to $n = 512$, and in black to $n = 1024$.

Appendices

8 Proof of the lower bounds

8.1 Proof of Theorem 3.1

The random variable $\boldsymbol{\theta}_j^* = ([\theta_j^*]_1^1, \dots, [\theta_j^*]_j^p)$'s are i.i.d. with values in $\Theta \subset \mathbb{R}^p$ and with density $g: \mathcal{P} \rightarrow \mathbb{R}$ with compact support Θ . Let $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_J) \in \mathbb{R}^{nJ}$ be the column vector of the observations generated by model (1.6). Conditionally to $\boldsymbol{\theta}^*$, \mathbf{Y} is a Gaussian vectors and the its log-likelihood is equal to

$$\log(p(\mathbf{Y}|\boldsymbol{\theta}^*)) = -\frac{Jn}{2} \log(2\pi) + \frac{J}{2} \log(\det(\Lambda)) - \frac{1}{2} \sum_{j=1}^J (\mathbf{Y}_j - \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f}^*)' \Lambda (\mathbf{Y}_j - \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{f}^*), \quad (8.1)$$

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

$$\mathbb{E}_{\boldsymbol{\theta}^*} [\partial_{[\theta^*]_{j_1}^{p_1}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*)) \partial_{[\theta^*]_{j_2}^{p_2}} \log(p(\mathbf{Y}|\boldsymbol{\theta}^*))] = \begin{cases} 0 & \text{if } j_1 \neq j_2, \\ -[(\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}^*)' \Lambda (\partial_{[\theta^*]_{j_1}^{p_2}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}^*)]_{p_1, p_2=1}^p & \text{if } j_1 = j_2, \end{cases} \quad (8.2)$$

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

$$(\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}^*)' \Lambda (\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}^*) \leq \sigma^{-2} \|\partial_{[\theta^*]_{j_1}^{p_1}} \mathbf{T}_{\boldsymbol{\theta}_{j_1}^*} \mathbf{f}^*\|^2 \leq C(\Theta, f^*) n \sigma^{-2}, \quad (8.3)$$

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

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

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

$$(\mathbb{E}[U^t V])^2 \leq \mathbb{E}[U^t U] \mathbb{E}[V^t V]. \quad (8.4)$$

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

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

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

$$\mathbb{E}[U^tV] = pJ.$$

Now, using that the expected score is zero and equation 8.2 we have

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

where $\partial_{\boldsymbol{\theta}_1} \log(g(\boldsymbol{\theta}_1)) = [\partial_{[\theta_1]_1} \log(g(\boldsymbol{\theta}_1)), \dots, \partial_{[\theta_1]_p} \log(g(\boldsymbol{\theta}_1))] \in \mathbb{R}^p$. Then, using inequality 8.3, it gives

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

Hence, using Equation 8.4 for any estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{Y})$ (see Theorem 1 in [GL95])

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

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

8.2 Proof of Theorem 3.2

The proof follows exactly the same guideline as the proof of Theorem 3.1. As above, $\mathbf{Y} \in \mathbb{R}^{nJ}$ is the column vector generated by model (1.5) and the random variable $\boldsymbol{\theta}_j^* = ([\theta_j^*]_1^1, \dots, [\theta_j^*]_j^p)$'s are i.i.d. with values in $\Theta \subset \mathbb{R}^p$ and density g . Then, conditionally to $\boldsymbol{\theta}^*$, \mathbf{Y} is a Gaussian vectors and Assumption 3.2 ensures that its log-likelihood has the same expression as in equation 8.1 but with

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

As the matrix $\boldsymbol{\Sigma}_n(\Theta)$ is positive semi definite with its smallest eigenvalue denoted by $s_n^2(\Theta)$ (see Assumption 3.2), the uniform bound (8.3) becomes

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

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

8.3 Proof of Theorem 3.3

For all $\theta \in \mathbb{R}$ the operators $T_\theta f(\cdot) = f(\cdot - \theta)$ are isometric from $L^2([0, 1])$ to $L^2([0, 1])$ as a change of variable implies immediately that $\|T_\theta f\|_{L^2}^2 = \|f\|_{L^2}^2$. For all continuously differentiable function f , we have

$$\partial_\theta T_\theta f(t) = -\text{sign}(\theta) \partial_t f(t - \theta),$$

where $\text{sign}(\cdot)$ is the sign function. Then, for all $\theta \in \Theta$, $\|\partial_\theta T_\theta f^*\|_{L^2}^2 = \|\partial_t f^*\|_{L^2}^2 \leq \|\partial_t f^*\|_\infty^2$ and Assumption 3.1 is satisfied with $C(\Theta, f^*) = \|\partial_t f^*\|_\infty^2$. Finally, as the error terms Z_j 's are i.i.d stationary random process the covariance function is invariant by the action of the shifts and Assumption 3.2 is satisfied with $\Sigma_n(\Theta) = \Sigma_n$ defined in (4.1) and $s_n^2(\Theta) = \gamma(\Sigma_n)$ defined in (4.2) (see Section 4.1 for further details). Then, the result of Theorem 3.3 follows as an application of Theorem 3.2. \square

8.4 Proof of Theorem 4.3

Let $n \geq 1$. We have that

$$\mathbb{E}[\|\tilde{f} - f^*\|_{L^2}] = \mathbb{E}\|\tilde{f} - f_{\Theta_0}^* + f_{\Theta_0}^* - f^*\|_{L^2} \geq \left| \underbrace{\mathbb{E}\|\tilde{f} - f_{\Theta_0}^*\|_{L^2}}_{\mathbf{A}} - \underbrace{\mathbb{E}\|f_{\Theta_0}^* - f^*\|_{L^2}}_{\mathbf{B}} \right| \quad (8.5)$$

where for all $t \in [0, 1]$, $\tilde{f}(t) = \frac{1}{J} \sum_{j=1}^J f^*(t - \theta_j^* + \hat{\theta}_j^\lambda)$, and $f_{\Theta_0}^*(t) = f^*(t + \bar{\theta}^*)$, with $\bar{\theta}^* = \frac{1}{J} \sum_{j=1}^J \theta_j^*$. In the rest of the proof, we show that \mathbf{A} is bounded from below by a quantity $C_0(f^*, g, n, \sigma^2, \rho) = C(f^*, \rho) \frac{n^{-1}\sigma^2}{\|\partial_t f^*\|_\infty^2 + n^{-1}\sigma^2 \int_\Theta (\partial_\theta \log(g(\theta)))^2}$ 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 $\mathbb{E}\|\tilde{f} - f^*\|_{L^2} \geq \frac{1}{2}C_0(f^*, g, n, \sigma^2, \rho)$, which will yield the desired result.

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

$$\begin{aligned} \|\tilde{f} - f_{\Theta_0}^*\|_{L^2} &= \left\| \frac{1}{J} \sum_{j=1}^J f^*(\cdot - \theta_j^* + \hat{\theta}_j^\lambda) - f^*(\cdot + \bar{\theta}^*) \right\|_{L^2} = \left(\sum_{k \in \mathbb{Z}} \left| \frac{1}{J} \sum_{j=1}^J \left(e^{i2\pi k(-\theta_j^* + \hat{\theta}_j^\lambda)} - e^{i2\pi k \bar{\theta}^*} \right) c_k^* \right|^2 \right)^{\frac{1}{2}}, \\ &\geq |c_1^*| \left| \frac{1}{J} \sum_{j=1}^J \left(e^{i2\pi(\hat{\theta}_j^\lambda - [\theta_{\Theta_0}^*]_j)} - 1 \right) \right|, \end{aligned}$$

where $\theta_{\Theta_0}^* = (\theta_1^* - \bar{\theta}^*, \dots, \theta_J^* - \bar{\theta}^*)$, the right hand side of the preceding inequality being positive since Assumption 4.2 ensures that $c_1^* \neq 0$ for all $j = 1, \dots, J$. Let $u_j = 2\pi(\hat{\theta}_j^\lambda - [\theta_{\Theta_0}^*]_j)$, $j = 1, \dots, J$. Since $\sum_{j=1}^J u_j = 0$ and $|u_j| \leq 4\pi\rho < 3$, $j = 1, \dots, J$ (by our assumption on ρ), Lemma 10.1 implies that

$$\|\tilde{f} - f_{\Theta_0}^*\|_{L^2} \geq C(f^*, \rho) \frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2. \quad (8.6)$$

Now, remark that $\mathbb{E}[\frac{1}{J} \|\hat{\theta}^\lambda - \theta_{\Theta_0}^*\|^2] \geq \mathbb{E}[\frac{1}{J} \|\hat{\theta}^\lambda - \theta^*\|^2] - \mathbf{C}$ with $\mathbf{C} = 2\mathbb{E}[|\bar{\theta}^*| \frac{1}{J} \sum_{j=1}^J |\hat{\theta}_j^\lambda - \theta_j^*|]$. By applying Theorem 3.3 we get that

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

Then, remark that $\mathbf{C} \leq 4\rho \sqrt{\mathbb{E}|\bar{\theta}^*|^2} \leq C(\rho, g) J^{-1/2}$. Hence \mathbf{C} tends to 0 as J goes to infinity. Therefore, using equation (8.6), 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

$$\mathbf{A} = \mathbb{E}[\|\tilde{f}^\lambda - \tilde{f}^*\|_{L^2}] \geq C_0(f^*, g, n, \sigma^2, \rho). \quad (8.7)$$

Upper bound on \mathbf{B} . By assumption, f^* is continuously differentiable on $[0, 1]$ implying that $\|f_{\Theta_0}^* - f^*\|_{L^2} = \|f^*(\cdot + \bar{\theta}^*) - f^*\|_{L^2} \leq \|\partial_t f\|_\infty |\bar{\theta}^*|$. Therefore, $\mathbb{E}\|f_{\Theta_0}^* - f^*\|_{L^2} \leq \|\partial_t f\|_\infty \sqrt{\mathbb{E}|\bar{\theta}^*|^2} \leq C(f^*, g)J^{-1/2}$. Hence, there exists a $J_2 \in \mathbb{N}$ such that $J \geq J_2$ implies

$$\mathbf{B} = \mathbb{E}[\|\tilde{f}_{\Theta_0}^* - \tilde{f}^*\|_{L^2}] \leq \frac{1}{2}C_0(f^*, g, n, \sigma^2, \rho). \quad (8.8)$$

To conclude the proof, equations (8.5), (8.7) and (8.8) imply that there exists a $J_0 \in \mathbb{N}$ such that $J \geq J_0$ implies $\mathbb{E}\|\hat{f}^\lambda - \tilde{f}^*\|_{L^2} \geq |\mathbf{A} - \mathbf{B}| \geq \frac{1}{2}C_0(f^*, g, n, \sigma^2, \rho)$. \square

9 Proof of the upper bounds

9.1 Proof of Proposition 4.1

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

$$D(\boldsymbol{\theta}) \geq |c_1^*|^2 \left(1 - \left|\frac{1}{J} \sum_{j=1}^J e^{i2\pi(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)}\right|^2\right) \quad (9.1)$$

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

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

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 $\boldsymbol{\theta} \in \Theta$

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

since $|(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*)| \leq 4\rho < 1/4$ for all $m, q = 1, \dots, n$ by Assumption 4.2 and the hypothesis that $\rho < 1/16$. Hence, using the lower bound (9.1), it follows that for all $\boldsymbol{\theta} \in \Theta$

$$D(\boldsymbol{\theta}) \geq C(f, \rho) \frac{1}{J^2} \sum_{j=1}^{J-1} \sum_{j'=j+1}^J |(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) - (\boldsymbol{\theta}_{j'} - \boldsymbol{\theta}_{j'}^*)|^2 \quad (9.2)$$

with $C(f, \rho) = 2|c_1^*|^2 C(\rho)$. Now assume that $\boldsymbol{\theta} \in \Theta_0$. Using the properties that $\sum_{j=1}^J \boldsymbol{\theta}_j = 0$ and $\sum_{j=1}^J (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) = -\sum_{j=1}^J \boldsymbol{\theta}_j^* = J\bar{\theta}^*$, it follows from elementary algebra that

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

The above equality together with the lower bound (9.2) completes the proof. \square

9.2 Proof of Theorem 4.1

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 9.1. *Suppose that Assumption 4.2 holds. Then, for any $x > 0$*

$$\mathbb{P}\left(\frac{1}{J}\|\boldsymbol{\theta}_{\Theta_0}^* - \boldsymbol{\theta}^*\|^2 \geq \rho^2 \left(\sqrt{\frac{2x}{J}} + \frac{x}{3J}\right)^2\right) \leq 2e^{-x}.$$

Using the inequality $\frac{1}{J}\|\hat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}^*\|^2 \leq \frac{2}{J}\|\hat{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_{\Theta_0}^*\|^2 + \frac{2}{J}\|\boldsymbol{\theta}_{\Theta_0}^* - \boldsymbol{\theta}^*\|^2$, it follows that Theorem 4.1 is a consequence of Lemma 9.1 and Theorem 6.1. Indeed, it can be easily checked that, under the assumptions of Theorem 4.1, Assumptions 5.1 to 5.4 are satisfied in the case of randomly shifted curves with an equi-spaced design and low-pass Fourier filtering, see the various arguments given in Section 5). The identifiability condition of Assumption 5.6 is given by Proposition 4.1. \square

9.3 Proof of Theorem 4.2

Consider the following inequality

$$\|\hat{f}^\lambda - f^*\|^2 \leq 2\|\hat{f}^\lambda - f_{\Theta_0}^*\|^2 + 2\|f_{\Theta_0}^* - f^*\|^2,$$

where $f_{\Theta_0}^*(t) = f^*(t - \bar{\boldsymbol{\theta}}^*)$ 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 $\|f_{\Theta_0}^* - f^*\|_{L^2}^2 \leq C(\Theta, f^*)|\bar{\boldsymbol{\theta}}^*|^2 = C(\Theta, f^*)\frac{1}{J}\|\boldsymbol{\theta}_{\Theta_0}^* - \boldsymbol{\theta}^*\|^2$. As explained in part 9.2 the assumptions of Theorem 4.2 are satisfied in the case of randomly shifted curves with an equi-spaced design and low-pass Fourier filtering. The result then follows from Theorem 6.2. \square

9.4 Proof of Theorem 6.1

We explain here the main arguments of the proof of Theorem 6.1. 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 (6.1) as follows,

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

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,

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

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

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

At this stage, recall that $\boldsymbol{\theta}_\Theta^* = \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} D(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}^\lambda = \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} M_\lambda(\boldsymbol{\theta})$. The proof follows a classical guideline in M-estimation: we show that the uniform (over Θ) convergence in probability of the criterion M_λ to D , yielding the convergence in probability of their argmins $\boldsymbol{\theta}_\Theta^*$ and $\hat{\boldsymbol{\theta}}^\lambda$ respectively. Assumption 5.6 ensures that there is a constant $C(\Theta, \mathcal{F}, f^*) > 0$ such that,

$$\frac{1}{J} \|\hat{\boldsymbol{\theta}}^\lambda - \boldsymbol{\theta}_\Theta^*\|^2 \leq C(\Theta, \Theta, \mathcal{F}, f^*) \left| D(\hat{\boldsymbol{\theta}}^\lambda) - D(\boldsymbol{\theta}_\Theta^*) \right| \quad (9.3)$$

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

$$\begin{aligned}
\left| D(\hat{\boldsymbol{\theta}}^\lambda) - D(\boldsymbol{\theta}_\Theta^*) \right| &\leq 2 \sup_{\boldsymbol{\theta} \in \Theta} |D(\boldsymbol{\theta}) - M_\lambda(\boldsymbol{\theta})| \\
&= \underbrace{2 \sup_{\boldsymbol{\theta} \in \Theta} \left\{ R_\lambda(\boldsymbol{\theta}) + Q_\lambda(\boldsymbol{\theta}) \right\}}_{\mathbf{B}} + \underbrace{2 \sup_{\boldsymbol{\theta} \in \Theta} \left\{ Q_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^Z(\boldsymbol{\theta}) + R_\lambda^{Z,\varepsilon}(\boldsymbol{\theta}) + Q_\lambda^\varepsilon(\boldsymbol{\theta}) + R_\lambda^\varepsilon(\boldsymbol{\theta}) \right\}}_{\mathbf{V}}
\end{aligned} \quad (9.4)$$

The rest of the proof is devoted to control the \mathbf{B} and \mathbf{V} terms.

Control of \mathbf{B} . Using Assumption 5.3 and 5.1, we have that $Q_\lambda(\boldsymbol{\theta}) \leq \frac{C(\Theta)}{J} \sum_{j=1}^J \left\| B_\lambda(T_{\theta_j^*} f, t) \right\|_{L^2}^2 \leq C(\Theta, \mathcal{F}) B(\lambda)$. Now by applying the Cauchy-Schwarz inequality, $|R_\lambda(\boldsymbol{\theta})| \leq \sup_{\boldsymbol{\theta} \in \Theta} \{ \sqrt{D(\boldsymbol{\theta})} \} \sqrt{Q_\lambda(\boldsymbol{\theta})}$. By Assumption 5.1, there exists a constant such $\sup_{\boldsymbol{\theta} \in \Theta} \{ D(\boldsymbol{\theta}) \} \leq C(\Theta, \mathcal{F}, f^*)$ and thus

$$\mathbf{B} \leq C(\Theta, \mathcal{F}, f^*) (B(\lambda) + \sqrt{B(\lambda)}). \quad (9.5)$$

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

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

where we have used the inequality $2ab \leq a^2 + b^2$, valid for any $a, b > 0$ to control the term $R_\lambda^{Z,\varepsilon}$. The quadratic terms Q_λ^Z and Q_λ^ε are controlled by Corollaries 10.1 and 10.2 respectively. It yields immediately to

$$\mathbb{P} \left(\mathbf{V} \geq C(\Theta, \mathcal{F}, f^*) (\gamma_{\max}(n) + \sigma^2) (v(x, J, \lambda) + \sqrt{v(x, J, \lambda)}) \right) \leq 2e^{-x}, \quad (9.6)$$

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

Putting together equations (9.3), (9.4), (9.5) and (9.6), we have

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

which completes the proof of Theorem 6.1. \square

9.5 Proof of Theorem 6.2

In this part, we use the notations introduced in the proof of Theorem 6.1. We have,

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

Again, the first term above depends on the bias, and the second term (stochastic) can be controlled in probability. Under Assumptions 5.1 and 5.3 we have that

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

and

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

The stochastic term $\frac{1}{J} \sum_{j=1}^J \left\| \langle S_\lambda(\cdot), \mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j + \boldsymbol{\varepsilon}_j \rangle \right\|_{L^2}^2$ in the above inequality can be controlled using Lemma 10.2 and the arguments in the proof of Corollaries 10.1 and 10.2 to obtain that for any $x > 0$

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

Then, from Theorem 6.1 it follows that

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

which completes the proof. \square

10 Technical Lemmas

Lemma 10.1. *Let $u = (u_1, \dots, u_J)$ such that $\sum_{j=1}^J u_j = 0$ with $|u_j| \leq \delta$ for some $0 \leq \delta < 3$ for all $j = 1, \dots, J$. 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 $\|u\|^2 = u_1^2 + \dots + u_J^2$.*

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

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

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

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

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 $\delta < 3$. \square

Lemma 10.2. *Let $\xi_{\lambda, J}(A_1, \dots, A_J) = \frac{1}{J} \sum_{j=1}^J \|\langle S_\lambda(\cdot), A_j \varepsilon_j \rangle\|_{L^2}^2$, where $\varepsilon_j \sim \mathcal{N}(0, I_n)$ 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

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

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 $r_{j, 1} \geq \dots \geq r_{j, n}$ the eigenvalues of \mathbf{A}_j with $r_{j, 1} \geq \dots \geq r_{j, n} \geq 0$ 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 $\tilde{\xi}_{\lambda, J} = J \xi_{\lambda, J} - J \mathbb{E} \xi_{\lambda, J} = \sum_{j=1}^J (\varepsilon_j' \mathbf{A}_j \varepsilon_j - \text{tr} \mathbf{A}_j)$. Let $\alpha > 0$, by Markov's inequality it follows that for all $u \in \left(0, \frac{1}{2r_1}\right)$, $\mathbb{P}\left(\tilde{\xi}_{\lambda, J} \geq \alpha\right) = \mathbb{P}\left(e^{u \tilde{\xi}_{\lambda, J}} \geq e^{u \alpha}\right) \leq e^{-u \alpha} \prod_{j=1}^J \mathbb{E} \left[e^{u \varepsilon_j' \mathbf{A}_j \varepsilon_j - u \text{tr} \mathbf{A}_j} \right]$, since the ε_j 's are independent. The log-Laplace transform of $\tilde{\varphi}_{\lambda, j} = \varepsilon_j' \mathbf{A}_j \varepsilon_j - \text{tr} \mathbf{A}_j$ is $\log(\mathbb{E} [e^{u \tilde{\varphi}_{\lambda, j}}]) = \sum_{\ell=1}^n -u r_{j, \ell} - \frac{1}{2} \log(1 - 2u r_{j, \ell})$. 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 $\log(\mathbb{E}[e^{u\tilde{\varphi}_{\lambda,j}}]) \leq -\sum_{\ell=1}^n \frac{u^2 r_{j,\ell}^2}{1-2ur_{j,\ell}} \leq \frac{u^2 \|r_j\|^2}{1-2ur_1}$, where $\|r_j\|^2 = r_{j,1}^2 + \dots + r_{n,j}^2$. Finally, we have

$$\mathbb{P}(\tilde{\varphi}_{\lambda,J} \geq \alpha) \leq \exp\left(-\left(u\alpha - \sum_{j=1}^J \frac{\|r_j\|^2 u^2}{1-2r_1 u}\right)\right) = \exp\left(-\left(u\alpha - \frac{\|r\|^2 u^2}{1-2r_1 u}\right)\right), \quad (10.1)$$

where $\|r\|^2 = \sum_{j=1}^J \sum_{\ell=1}^n r_{j,\ell}^2$. The right hand side of the above inequality achieves its minimum at $u = \frac{1}{2r_1} \left(1 - \frac{\|r\|}{\sqrt{2\alpha r_1 + \|r\|^2}}\right)$. Evaluating (10.1) at this point and using the inequality $(1+x)^{1/2} \leq 1 + \frac{x}{2}$, valid for all $x \geq -1$, one has that

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

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

Corollary 10.1. *Under Assumptions 5.1 to 5.3, there exists a constant $C(\Theta, \mathcal{F}) > 0$ such that for all $x > 0$,*

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

Proof. Assumption 5.1 gives the uniform bound

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

where $\xi_{\lambda,J}(\sigma Id_n, \dots, \sigma Id_n)$ is defined in Lemma 10.2 and does not depend on $\boldsymbol{\theta}$. Thus, the result immediately follows from Lemma 10.2. \square

Corollary 10.2. *Under Assumptions 5.1 to 5.4, there exists a constant $C(\Theta, \mathcal{F}) > 0$ such that for all $x \geq 0$,*

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

Proof. Assumption 5.1 gives the uniform bound

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

Hence, conditionally on $\boldsymbol{\theta}^*$ we have that $\sup_{\boldsymbol{\theta} \in \Theta^J} Q_{\lambda}^Z(\boldsymbol{\theta}) \leq C(\Theta, \mathcal{F})\xi_{\lambda,J}(A_1, \dots, A_J)$, where $\xi_{\lambda,J}(A_1, \dots, A_J)$ is defined in Lemma 10.2 with $A_j = \mathbb{E}_{\boldsymbol{\theta}^*} [\mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j (\mathbf{T}_{\boldsymbol{\theta}_j^*} \mathbf{Z}_j)']^{\frac{1}{2}}$. 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 $\mathbf{S}_{\lambda} = [\langle S_{\lambda}^{\ell}, S_{\lambda}^{\ell'} \rangle_{L^2}]_{\ell, \ell'=1}^n$. Under Assumption 5.4 we have that $\text{tr}(\mathbf{A}_j) \leq \gamma_{\max}(A_j) \text{tr} \mathbf{S}_{\lambda} \leq \gamma_n(\Theta)V(\lambda)$, for all $j = 1, \dots, J$ and any $\boldsymbol{\theta}^* \in \Theta^J$. Thus, the result follows by arguing as in the proof of Lemma 10.2 and by taking expectation with respect to $\boldsymbol{\theta}^*$. \square

References

- [AAT07] Stéphanie Allasonnière, 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 Allasonnière, 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.
- [KBCL99] D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and shape theory*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1999.
- [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.
- [Le98] H. Le. On the consistency 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.
- [OC95] J. M. Oller and J. M. Corcuera. Intrinsic analysis of statistical estimation. *Ann. Statist.*, 23(5):1562–1581, 1995.
- [Pen06] Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.*, 25:127–154, July 2006.

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