HAL open science

Minimax properties of Fréchet means of discretely sampled curves

Jérémie Bigot, Xavier Gendre

To cite this version:

Jérémie Bigot, Xavier Gendre. Minimax properties of Fréchet means of discretely sampled curves. Annals of Statistics, 2013, 41 (2), pp.923-956. 10.1214/13-AOS1104 . hal-00737560v3

HAL Id: hal-00737560

https://hal.science/hal-00737560v3

Submitted on 11 Jun 2013

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.

The Annals of Statistics
2013, Vol. 41, No. 2, 923–956
DOI: 10.1214/13-AOS1104
© Institute of Mathematical Statistics, 2013

MINIMAX PROPERTIES OF FRÉCHET MEANS OF DISCRETELY SAMPLED CURVES

By Jérémie Bigot[1] and Xavier Gendre

DMIA–ISAE and Institut de Mathématiques de Toulouse

We study the problem of estimating a mean pattern from a set of similar curves in the setting where the variability in the data is due to random geometric deformations and additive noise. We propose an estimator based on the notion of Fréchet mean that is a generalization of the standard notion of averaging to non-Euclidean spaces. We derive a minimax rate for this estimation problem, and we show that our estimator achieves this optimal rate under the asymptotics where both the number of curves and the number of sampling points go to infinity.

1. Introduction.

1.1. *Fréchet means.* The Fréchet mean [10] is an extension of the usual Euclidean mean to nonlinear spaces endowed with non-Euclidean metrics. If $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ denote i.i.d. random variables with values in a metric space $\mathcal{M}$ with metric $d_{\mathcal{M}}$, then the empirical Fréchet mean $\overline{\mathbf{Y}}_{\mathcal{M}}$ of the sample $\mathbf{Y}_1, \dots, \mathbf{Y}_J$ is defined as a minimizer (not necessarily unique) of

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

For random variables belonging to a nonlinear manifold, a well-known example is the computation of the mean of a set of planar shapes in Kendall's shape space [18] that leads to the Procrustean means studied in [13]. A detailed study of some properties of the Fréchet mean in finite dimensional

Received October 2012; revised February 2013.

[1]Supported by the French Agence Nationale de la Recherche (ANR) under reference ANR-JCJC-SIMI1 DEMOS.

AMS 2000 subject classifications. Primary 62G08; secondary 62G20.

Key words and phrases. Fréchet mean, non-Euclidean metric, deformable models, lie group action, curve registration, minimax rate of convergence, Sobolev balls, functional data analysis.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in *The Annals of Statistics*, 2013, Vol. 41, No. 2, 923–956. This reprint differs from the original in pagination and typographic detail.

Riemannian manifolds (such as consistency and uniqueness) has been performed in [1–3, 17]. However, there is not so much work on the properties of the Fréchet mean in infinite dimensional and non-Euclidean spaces of curves or images. In this paper, we are concerned with the nonparametric estimation of a mean pattern (belonging to a nonlinear space) from a set of similar curves in the setting where the variability in the data is due to random geometric deformations and additive noise.

More precisely, let us consider noisy realizations of J curves $f_1, \dots, f_J : [0,1] \to \mathbb{R}$ sampled at n equispaced points $t_\ell = \frac{\ell}{n}, \ell = 1, \dots, n$,

$$Y_{\ell,j} = f_j(t_\ell) + \varepsilon_{\ell,j}, \qquad \ell = 1, \dots, n \text{ and } j = 1, \dots, J, \tag{1}$$

where the $\varepsilon_{\ell,j}$'s are independent and identically distributed (i.i.d.) Gaussian variables with zero expectation and known variance $\sigma^2 > 0$. In many applications, the observed curves have a similar structure that may lead to the assumption that the f_j's are random elements varying around the same mean pattern $f : [0,1] \to \mathbb{R}$ (also called reference template). However, due to additive noise and geometric variability in the data, this mean pattern is typically unknown and has to be estimated. In this setting, a widely used approach is Grenander's pattern theory [14, 15, 28, 29] that models geometric variability by the action of a Lie group on an infinite dimensional space of curves (or images).

When the curves f_j in (1) exhibit a large source of geometric variation in time, this may significantly complicates the construction of a consistent estimator of a mean pattern. In what follows, we consider the simple model of randomly shifted curves that is commonly used in many applied areas such as neurosciences [27] or biology [25]. In such a framework, we have

$$f_j(t) = f(t - \theta_j^*) \qquad \text{for all } t \in [0,1] \text{ and } j = 1, \dots, J, \tag{2}$$

where $f : [0,1] \to \mathbb{R}$ is an unknown curve that can be extended outside $[0,1]$ by 1-periodicity. In a similar way, we could consider a function f defined on the circle $\mathbb{R}/\mathbb{Z}$. The shifts θ_j^*'s are supposed to be i.i.d. real random variables (independent of the $\varepsilon_{\ell,j}$'s) that are sampled from an unknown distribution $\mathbb{P}_*$ on $\mathbb{R}$. In model (2), the shifts θ_j^* represent a source of geometric variability in time.

In functional data analysis, the problem of estimating a mean pattern from a set of curves that differ by a time transformation is usually referred to as the curve registration problem; see, for example, [24]. Registering functional data has received a lot of attention in the literature over the two last decades; see, for example, [4, 20, 24, 26, 32] and references therein. Nevertheless, in these papers, constructing consistent estimators of the mean pattern f as the number J of curves tends to infinity is generally not considered. Self-modeling regression methods proposed in [19] are semiparametric models for

curve registration that are similar to the shifted curves model, where each observed curve is a parametric transformation of an unknown mean pattern. Constructing a consistent estimator of the mean pattern in such models has been investigated in [19] in an asymptotic framework where both the number J of curves and the number n of design points grow toward infinity. However, deriving optimal estimators in the minimax sense has not been considered in [19]. Moreover, a novel contribution of this paper is to make a connection between the curve registration problem and the notion of Fréchet mean in non-Euclidean spaces which has not been investigated so far.

1.2. *Model and objectives.* The main goal of this paper is to construct nonparametric estimators of the mean pattern f from the data

$$Y_{\ell,j} = f(t_\ell - \theta_j^*) + \varepsilon_{\ell,j}, \qquad \ell = 1, \dots, n \text{ and } j = 1, \dots, J, \tag{3}$$

in the setting where both the number J of curves and the number n of design points are allowed to vary and to tend toward infinity.

In the sequel to this paper, it will be assumed that the random shifts are sampled from an unknown density g with respect to the Lebesgue measure $d\theta$ [namely $d\mathbb{P}_*(\theta) = g(\theta)\, d\theta$]. Note that since f is assumed to be 1-periodic, one may restrict to the case where the density g has a compact support included in the interval $[-\frac{1}{2}, \frac{1}{2}]$. Under assumption (2), the (standard) Euclidean mean $\bar{Y}_\ell = \frac{1}{J}\sum_{j=1}^J Y_{\ell,j}$ of the data is generally not a consistent estimator of the mean pattern f at $t = t_\ell$. Indeed, the law of large numbers implies that

$$\lim_{J\to\infty} \bar{Y}^\ell = \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^J f(t_\ell - \theta_j^*) = \int_{\mathbb{R}} f(t_\ell - \theta) g(\theta)\, d\theta \qquad \text{a.s.}$$

Thus, under mild assumptions on f and g, we have

$$\lim_{J\to\infty} \bar{Y}^\ell = f \star g(t_\ell) \neq f(t_\ell) \qquad \text{a.s.},$$

where $f \star g$ is the convolution product between f and g.

To build a consistent estimator of f in model (3), we propose to use a notion of empirical Fréchet mean in an infinite dimensional space. Recently, some properties of Fréchet means in randomly shifted curves models have been investigated in [6] and [5]. However, studying the rate of convergence and the minimax properties of such estimators in the double asymptotic setting $\min(n, J) \to +\infty$ has not been considered so far.

Note that model (3) is clearly not identifiable, as for any $\tilde{\theta} \in [-\frac{1}{2}, \frac{1}{2}]$, one may replace $f(\cdot)$ by $\tilde{f}(\cdot) = f(\cdot - \tilde{\theta})$ and θ_j^* by $\tilde{\theta}_j = \theta_j^* - \tilde{\theta}$ without changing model (3). Therefore, estimation of f is only feasible up to a time shift. Thus, we propose to consider the problem of estimating its equivalence class $[f]$ (or orbit) under the action of shifts. More precisely, let $L^2_{\text{per}}([0,1])$ be the

space of squared integrable functions on $[0,1]$ that can be extended outside $[0,1]$ by 1-periodicity. Let $\mathbb{S}_1$ be the one-dimensional torus. We recall that any element $\tau = \tau(\theta) \in \mathbb{S}_1$ can be identified with an element θ in the interval $[-\frac{1}{2}, \frac{1}{2}]$. For $f \in L^2_{\text{per}}([0,1])$, we define its equivalence class by the action of a time shift as

$$[f] := \{f^\tau, \tau \in \mathbb{S}_1\},$$

where for $\tau = \tau(\theta) \in \mathbb{S}_1$ (with $\theta \in [-\frac{1}{2}, \frac{1}{2}]$), $f^\tau(t) = f(t-\theta)$ for all $t \in [0,1]$. Let $f, h \in L^2_{\text{per}}([0,1])$, and we define the distance between $[f], [h] \in L^2_{\text{per}}([0,1])/\mathbb{S}_1$ as

$$d([f],[h]) = \inf_{\theta \in [-1/2,1/2]} \left(\int_0^1 |f(t-\theta) - h(t)|^2 \, dt \right)^{1/2}. \tag{4}$$

In the setting of Grenander's pattern theory, $(L^2_{\text{per}}([0,1])/\mathbb{S}_1, d)$ represents an infinite dimensional and nonlinear set of curves, and $\mathbb{S}_1$ is a Lie group modeling geometric variability in the data.

1.3. *Main contributions.* Let us assume that $\mathcal{F} \subset L^2_{\text{per}}([0,1])$ represents some smoothness class of functions (e.g., a Sobolev ball). Suppose also that the unknown density g of the random shifts in (2) belongs to some set $\mathcal{G}$ of probability density functions on $[-\frac{1}{2}, \frac{1}{2}]$. Let $\hat{f}_{n,J}$ be some estimator of f based on the random variables $Y_{\ell,j}$ given by (3) taking its values in $L^2_{\text{per}}([0,1])$. For some $f \in \mathcal{F}$, the risk of the estimator $\hat{f}_{n,J}$ is defined by

$$\mathcal{R}_g(\hat{f}_{n,J}, f) = \mathbb{E}^g(d^2([\hat{f}_{n,J}], [f])),$$

where the above expectation $\mathbb{E}^g$ is taken with respect to the distribution of the $Y_{\ell,j}$'s in (3) and under the assumption that the shifts are i.i.d. random variables sampled from the density g. We propose to investigate the optimality of an estimator by introducing the following minimax risk:

$$\mathcal{R}_{n,J}(\mathcal{F}, \mathcal{G}) = \inf_{\hat{f}_{n,J}} \sup_{g \in \mathcal{G}} \sup_{f \in \mathcal{F}} \mathcal{R}_g(\hat{f}_{n,J}, f),$$

where the above infimum is taken over the set of all possible estimators in model (3).

For $f \in L^2_{\text{per}}([0,1])$, let us denote its Fourier coefficients by

$$c_k = \int_0^1 f(t) e^{-i2\pi kt} \, dt, \qquad k \in \mathbb{Z}.$$

Suppose that $\mathcal{F} = \tilde{W}_s(A, c_*)$ is the following bounded set of nonconstant functions with degree of smoothness $s > 1/2$:

$$\tilde{W}_s(A, c_*) = \left\{ f \in L^2_{\text{per}}([0,1]); \sum_{k \in \mathbb{Z}} (1 + |k|^{2s}) |c_k|^2 \le A^2 \text{ with } |c_1| \ge c_* \right\}$$

for some positive reals A and c_*. The introduction of the above set is motivated by the definition of Sobolev balls. The additional assumption $|c_1| \geq c_*$ is needed to ensure identifiability of $[f]$ in model (3) with respect to the distance (4).

Moreover, let $\mathcal{G}^\kappa$ be a set of probability densities having a compact support of size smaller than κ with $0 < \kappa < 1/8$ defined as

$$\mathcal{G}^\kappa = \left\{ g : \left[-\frac{1}{2}, \frac{1}{2}\right] \to \mathbb{R}^+; \int_{-1/2}^{1/2} g(\theta)\, d\theta = 1 \text{ and } \operatorname{supp}(g) \subseteq [-\kappa/2, \kappa/2] \right\}.$$

Suppose also that the following condition holds:

$$J \asymp n^\alpha \qquad \text{for some } 0 < \alpha < 1/6,$$

where the notation $J \asymp n^\alpha$ means that there exist two positive constants $c_2 \geq c_1 > 0$ such that $c_1 n^\alpha \leq J \leq c_2 n^\alpha$ for any choices of J and n.

Then, under such assumptions, the main contribution of the paper is to show that one can construct an estimator $\hat{f}_{n,J}$ based on a smoothed Fréchet mean of discretely sampled curves that satisfies

$$\limsup_{\min(n,J)\to+\infty} r_{n,J}^{-1} \sup_{g\in\mathcal{G}^\kappa} \sup_{f\in\tilde{W}_s(A,c_*)} \mathcal{R}_g(\hat{f}_{n,J}, f) \leq C_0,$$

where $C_0 > 0$ is a constant that only depends on A, s, κ, c_* and σ^2. The rate of convergence $r_{n,J}$ is given by

$$r_{n,J} = n^{-1} + (nJ)^{-2s/(2s+1)}.$$

The two terms in the rate $r_{n,J}$ have different interpretations. The second term $(nJ)^{-2s/(2s+1)}$ is the usual nonparametric rate for estimating the function f (over a Sobolev ball) in model (3) that we would obtain if the true shifts $\theta_1^*, \dots, \theta_J^*$ were known. Moreover, under some additional assumptions, we will show that this rate is optimal in the minimax sense and that our estimator achieves it.

The first term n^{-1} in the rate $r_{n,J}$ can be interpreted as follows. As shown later in the paper, the computation of a Fréchet mean of curves is a two-step procedure. It consists of building estimators $\hat{\theta}_j^0$ of the unknown shifts and then aligning the observed curves. For $\theta = (\theta_1, \dots, \theta_J) \in \mathbb{R}^J$, let us define the Euclidean norm $\|\theta\| = (\sum_{j=1}^J |\theta_j|^2)^{1/2}$. One of the contributions of this paper is to show that estimation of the vector

$$\theta^0 = (\theta_1^* - \bar{\theta}_J, \dots, \theta_J^* - \bar{\theta}_J)' \in \mathbb{R}^J \qquad \text{where } \bar{\theta}_J = \frac{1}{J}\sum_{j=1}^J \theta_j^*,$$

is feasible at the rate n^{-1} for the normalized quadratic risk $\frac{1}{J}\mathbb{E}\|\hat{\theta}^0 - \theta^0\|^2$ and that this allows us to build a consistent Fréchet mean. If the number J of

curves was fixed, n^{-1} would correspond to the usual semi-parametric rate for estimating the shifts in model (3) in the setting where the θ_j^* are nonrandom parameters; see [7, 11, 31] for further details. Here, this rate of convergence has been obtained in the double asymptotic setting $\min(n, J) \to +\infty$. This setting significantly complicates the estimation of the vector $\theta^0 \in \mathbb{R}^J$ since its dimension J is increasing with the sample size nJ. Hence, in the case of $\min(n, J) \to +\infty$, estimating the shifts at the rate n^{-1} is not a standard semi-parametric problem, and we have to impose the constraint $J \asymp n^\alpha$ (with $0 < \alpha < 1/6$) to obtain this result. The term n^{-1} in the rate $r_{n,J}$ is thus the price to pay for not knowing the random shifts in (3) that need to be estimated to compute a Fréchet mean.

1.4. *Organization of the paper.* In Section 2, we introduce a notion of smoothed Fréchet means of curves. We also discuss the connection between this approach and the well-known problems of curve registration and image warping. In Section 3, we discuss the rate of convergence of the estimators of the shifts. We also build a Fréchet mean using model selection techniques, and we derive an upper bound on its rate of convergence. In Section 4, we derive a lower bound on the minimax risk, and we give some sufficient conditions to obtain a smoothed Fréchet mean converging at an optimal rate in the minimax sense. In Section 5, we discuss the main results of the paper and their connections with the nonparametric literature on deformable models. Some numerical experiments on simulated data are presented in Section 6. The proofs of the main results are gathered in a technical Appendix.

2. Smoothed Fréchet means of curves. Let $f_1, \dots, f_J$ be a set of functions in $L^2_{\text{per}}([0,1])$. We define the Fréchet mean $[\bar{f}]$ of $[f_1], \dots, [f_J]$ as

$$[\bar{f}] \in \underset{[f] \in L^2_{\text{per}}([0,1])/\mathbb{S}_1}{\operatorname{argmin}} \frac{1}{J} \sum_{j=1}^{J} d^2([f], [f_j]).$$

It can be easily checked that a representation $\bar{f} \in L^2_{\text{per}}([0,1])$ of the class $[\bar{f}]$ is given by the following two-step procedure:

(1) Computation of shifts to align the curves

$$(\tilde{\theta}_1, \dots, \tilde{\theta}_J) \in \underset{(\theta_1,\dots,\theta_J) \in [-1/2,1/2]^J}{\operatorname{argmin}} \frac{1}{J} \sum_{j=1}^{J} \int_0^1 \left| f_j(t + \theta_j) - \frac{1}{J} \sum_{j'=1}^{J} f_{j'}(t + \theta_{j'}) \right|^2 dt. \tag{5}$$

(2) Averaging after an alignment step: $\bar{f}(t) = \frac{1}{J} \sum_{j=1}^{J} f_j(t + \tilde{\theta}_j), t \in [0,1]$.

Let us now explain how the above two-step procedure can be used to define an estimator of f in model (3). Let

$$\{\phi_k(t) = e^{i2\pi kt}, t \in [0,1]\}_{k\in\mathbb{Z}}$$

be the standard Fourier basis. For legibility, we assume that $n = 2N \geq 4$ is even, and we split the data into two samples as follows:

$$Y_{q,j}^{(0)} = Y_{2q,j} \quad \text{and} \quad Y_{q,j}^{(1)} = Y_{2q-1,j}, \qquad q = 1, \ldots, N$$

for $j = 1, \ldots, J$, and

$$t_q^{(0)} = t_{2q} \quad \text{and} \quad t_q^{(1)} = t_{2q-1}, \qquad q = 1, \ldots, N.$$

For any $z \in \mathbb{C}$, we denote by $\overline{z}$ its complex conjugate. Then we define the following empirical Fourier coefficients, for any $j \in \{1, \ldots, J\}$:

$$\hat{c}_{k,j}^{(0)} = \frac{1}{N}\sum_{q=1}^{N} Y_{q,j}^{(0)} \overline{\phi_k(t_q^{(0)})} = \bar{c}_{k,j}^{(0)} + \frac{1}{\sqrt{N}} z_{k,j}^{(0)}, \qquad -\frac{N}{2} \leq k < \frac{N}{2},$$

$$\hat{c}_{k,j}^{(1)} = \frac{1}{N}\sum_{q=1}^{N} Y_{q,j}^{(1)} \overline{\phi_k(t_q^{(1)})} = \bar{c}_{k,j}^{(1)} + \frac{1}{\sqrt{N}} z_{k,j}^{(1)}, \qquad -\frac{N}{2} \leq k < \frac{N}{2},$$

where

$$\bar{c}_{k,j}^{(p)} = \frac{1}{N}\sum_{q=1}^{N} f(t_q^{(p)} - \theta_j^*)\overline{\phi_k(t_q^{(p)})}, \qquad p \in \{0,1\},$$

and the $z_{k,j}^{(p)}$'s are i.i.d. complex Gaussian variables with zero expectation and variance σ^2.

Then we define estimators of the unknown random shifts θ_j^* as

$$(\hat{\theta}_1, \ldots, \hat{\theta}_J) \in \underset{(\theta_1,\ldots,\theta_J)\in[-1/2,1/2]^J}{\operatorname{argmin}} M_n(\theta_1, \ldots, \theta_J), \tag{6}$$

where

$$M_n(\theta_1, \ldots, \theta_J) = \frac{1}{J}\sum_{j=1}^{J}\sum_{|k|\leq k_0}\left|\hat{c}_{k,j}^{(0)} e^{i2\pi k\theta_j} - \frac{1}{J}\sum_{j'=1}^{J}\hat{c}_{k,j'}^{(0)} e^{i2\pi k\theta_{j'}}\right|^2, \tag{7}$$

with some positive integer k_0 that will be discussed later. The smoothed Fréchet mean of f is then defined as

$$\hat{f}_{n,J}^{(m)}(t) = \sum_{|k|\leq m}\left(\frac{1}{J}\sum_{j=1}^{J}\hat{c}_{k,j}^{(1)} e^{i2\pi k\hat{\theta}_j}\right)\phi_k(t) = \frac{1}{J}\sum_{j=1}^{J}\hat{f}_j^{(m)}(t + \hat{\theta}_j), \qquad t \in [0,1],$$

where the integer $m \in \{1, \ldots, N/2\}$ is a frequency cut-off parameter that will be discussed later and $\hat{f}_j^{(m)}(t) = \sum_{|k|\leq m}\hat{c}_{k,j}^{(1)}\phi_k(t)$. Note that the estimators

$\hat{\theta}_j$ of the shifts have been computed using only half of the data and that the curves $\hat{f}_j^{(m)}$ are calculated using the other half of the data. By splitting the data in such a way, the random variables $\hat{\theta}_j$ and $\hat{f}_j^{(m)}$ are independent conditionally to $(\theta_1^*, \ldots, \theta_J^*)$. The computation of the $\hat{\theta}_j$'s can be performed by using a gradient descent algorithm to minimize the criterion (7); for further details, see [6].

Note also that this two-step procedure does not require the use of a reference template to compute estimators $\hat{\theta}_1, \ldots, \hat{\theta}_J$ of the random shifts. Indeed, one can interpret the term $\frac{1}{J}\sum_{j'=1}^{J} \hat{f}_{j'}(t+\theta_{j'})$ in (5) as a template that is automatically estimated. In statistics, estimating a mean pattern from set of curves that differ by a time transformation is usually referred to as the curve registration problem. It has received a lot of attention over the last two decades; see, for example, [4, 25, 27] and references therein. Hence, there exists a connection between our approach and the well-known problems of curve registration and its generalization to higher dimensions (image warping); see, for example, [12]. However, studying the minimax properties of an estimator of a mean pattern in curve registration models has not been investigated so far.

3. Upper bound on the risk.

3.1. *Consistent estimation of the unknown shifts.* Note that, due to identifiability issues in model (3), the minimization (6) is not well defined. Indeed, for any $(\hat{\theta}_1, \ldots, \hat{\theta}_J)$ that minimizes (7), one has that for any $\tilde{\theta}$ such that $(\hat{\theta}_1 + \tilde{\theta}, \ldots, \hat{\theta}_J + \tilde{\theta}) \in [-\frac{1}{2}, \frac{1}{2}]^J$, this vector is also a minimizer of M_n. Choosing identifiability conditions amounts to imposing constraints on the minimization of the criterion

$$M(\theta_1, \ldots, \theta_J) = \frac{1}{J}\sum_{j=1}^{J}\sum_{|k|\le k_0}\left| c_k e^{i2\pi k(\theta_j - \theta_j^*)} - \frac{1}{J}\sum_{j'=1}^{J} c_k e^{i2\pi k(\theta_{j'} - \theta_{j'}^*)}\right|^2, \tag{8}$$

where c_k, $k \in \mathbb{Z}$, are the Fourier coefficients of the mean pattern f. Criterion (8) can be interpreted as a version without noise criterion (7) when replacing $\hat{c}_{k,j}^{(0)}$ by $c_k e^{-i2\pi k\theta_j^*}$. Obviously, criterion (8) admits a minimum at $\theta^* = (\theta_1^*, \ldots, \theta_J^*)$ such that $M(\theta^*) = 0$. However, this minimizer over $[-\frac{1}{2}, \frac{1}{2}]^J$ is clearly not unique. To impose uniqueness of some minimum of M over a restricted set, let us introduce the following identifiability conditions:

Assumption 1. The distribution g of the random shifts has a compact support included in $[-\kappa/2, \kappa/2]$ for some $0 < \kappa < 1/8$.

Assumption 2. The mean pattern f in model (3) is such that $c_1 = \int_0^1 f(x)e^{-i2\pi x}\,dx \neq 0$.

Assumption 1 means that the support of the density g of the random shifts should be sufficiently small. This implies that the shifted curves $f(t-\theta_j^*)$ are somehow concentrated around the unknown mean pattern f. Such an assumption of concentration of the data around a reference shape has been used in various papers to prove the uniqueness and the consistency of Fréchet means for random variables lying in a Riemannian manifold; see [1–3, 17]. Assumption 1 could certainly be weakened by dealing with a basis other than the Fourier polynomials. However, this is not the main point of this paper. Recent studies in this direction have been made to derive asymptotic on the Fréchet mean of a distribution on the circle without restriction on its support; see [8, 16] or [22], for instance. Assumption 2 is an identifiability condition to avoid the case where the function f is constant over $[0,1]$ which would make impossible the estimation of the unobserved random shifts.

For $0<\kappa<1/8$, let us define the constrained set

$$\Theta_\kappa = \left\{ (\theta_1,\dots,\theta_J) \in [-\kappa/2,\kappa/2]^J, \sum_{j=1}^J \theta_j = 0 \right\}.$$

Let

$$\theta_j^0 = \theta_j^* - \frac{1}{J}\sum_{j'=1}^J \theta_{j'}^*, \qquad j=1,\dots,J \text{ and } \theta^0 = (\theta_1^0,\dots,\theta_J^0).$$

Thanks to Proposition 4.1 in [5], we have:

PROPOSITION 3.1. *Suppose that Assumptions 1 and 2 hold. Then, for any* $(\theta_1,\dots,\theta_J)\in\Theta_\kappa$,

$$M(\theta_1,\dots,\theta_J) - M(\theta_1^0,\dots,\theta_J^0) \ge C(f,\kappa)\frac{1}{J}\sum_{j=1}^J |\theta_j-\theta_j^0|^2,$$

where $C(f,\kappa) = 4\pi^2|c_1|^2\cos(4\pi\kappa) > 0$.

Therefore, over the constrained set Θ_κ, criterion (8) has a unique minimum at θ^0 such that $M(\theta^0)=0$. Let us now consider the estimators

$$\hat\theta^0 = (\hat\theta_1^0,\dots,\hat\theta_J^0) \in \operatorname*{argmin}_{(\theta_1,\dots,\theta_J)\in\Theta_\kappa} M_n(\theta_1,\dots,\theta_J). \tag{9}$$

The following theorem shows that, under appropriate assumptions, the vector $\hat\theta^0$ is a consistent estimator of θ^0.

THEOREM 3.1. *Suppose that Assumptions 1 and 2 hold. Let* $J\ge 2$ *and* $s\ge 2$. *Then there exists a constant* $C>0$ *that only depends on* A, s, κ, c_* *and* σ^2 *such that, for any* $f\in\tilde W_s(A,c_*)$, *we have*

$$\frac{1}{J}\mathbb{E}^g\|\hat\theta^0-\theta^0\|^2 \le \frac{C}{n}\left(1+\frac{k_0^5}{n^{1/2}}\right)\left(1+\frac{k_0^{3/2}J^3}{n^{1/2}}\right). \tag{10}$$

The hypothesis $s \geq 2$ in Theorem 3.1 is related to the need of handling the Hessian matrix associated to the criterion M_n. Inequality (10) shows that the quality of the estimation of the random shifts depends on the ratio between n and J. In particular, it suggests that the quality of this estimation should deteriorate if the number J of curves increases and n remains fixed. This shows that estimating the vector $\hat{\theta}^0 \in \mathbb{R}^J$ is not a standard parametric problem, since the dimension J is is allowed to grow to infinity in our setting. To the contrary, if J is not too large with respect to n, then an estimation of the shifts is feasible at the usual parametric rate n^{-1}. More precisely, by Theorem 3.1, we immediately have the following result:

COROLLARY 3.1. *Suppose that the assumptions of Theorem 3.1 are satisfied. If $k_0 \geq 1$ is a fixed integer and $J \asymp n^\alpha$ for some $0 < \alpha \leq 1/6$, then there exists $C_1 > 0$ that only depends on A, s, σ^2, κ, c_* and k_0 such that*

$$\frac{1}{J}\mathbb{E}^g\|\hat{\theta}^0 - \theta^0\|^2 \leq \frac{C_1}{n}.$$

Therefore, under the additional assumption that $J \asymp n^\alpha$, for some $0 < \alpha \leq 1/6$, the vector $\hat{\theta}^0$ converges to θ^0 at the rate n^{-1} for the normalized Euclidean norm. The assumption $\alpha \leq 1/6$ illustrates the fact that the number J of curves should not be too large with respect to the size n of the design. Such a condition appears to be sufficient, but we do not claim about the existence of an optimal rate at which the number n of design points should increase for a given increase in J.

3.2. *Estimation of the mean pattern.* For $p \in \{0,1\}$, let $Y^{(p)}$ be given by $(Y^{(p)}_{q,j})_{1\leq q\leq N, 1\leq j\leq J}$. Thanks to the estimator $\hat{\theta}^0$ of the random shifts, we can align the data $Y^{(1)}$ in order to estimate the mean pattern f in (3). Let $m_1 < N/2$ be some positive integer. For any $m \in \{1, \dots, m_1\}$, we recall that the estimator $\hat{f}^{(m)}_{n,J}$ is given by

$$\hat{f}^{(m)}_{n,J}(t) = \frac{1}{J}\sum_{j=1}^{J} \hat{f}^{(m)}_j(t + \hat{\theta}^0_j), \qquad t \in [0,1].$$

To simplify the notation, we omit the dependency on k_0, n and J of the above estimators, and we write $\hat{f}^{(m)} = \hat{f}^{(m)}_{n,J}$. We denote by $\mathbb{E}^{(1)}$ the expectation according to the distribution of $Y^{(1)}$. By construction, we recall that $Y^{(0)}$ and $Y^{(1)}$ are independent. Thus, we obtain

$$\mathbb{E}^{(1)}[\hat{f}^{(m)}(t)] = \bar{f}^{(m)}(t) = \frac{1}{J}\sum_{j=1}^{J} \bar{f}^{(m)}_j(t + \hat{\theta}^0_j), \qquad t \in [0,1],$$

where we have set

$$\bar{f}_j^{(m)}(t) = \sum_{|k|\le m} \bar{c}_{k,j}^{(1)} \phi_k(t), \qquad j \in \{1, \dots, J\}.$$

Therefore, $\hat{f}^{(m)}$ is a biased estimator of f with respect to $\mathbb{E}^{(1)}$. The idea of the procedure is that if the estimators $\hat{\theta}_j^0$ of the shifts behave well then $d^2([f], [\bar{f}^{(m_1)}])$ is small and estimating f amounts to estimate $\bar{f}^{(m_1)}$.

To choose an estimator of $\bar{f}^{(m_1)}$ among the $\hat{f}^{(m)}$'s, we take a model selection approach. Before describing the procedure, let us compute the quadratic risk of an estimator $\hat{f}^{(m)}$,

$$\begin{aligned}\mathbb{E}^{(1)}&\left[\int_0^1 |\bar{f}^{(m_1)}(t) - \hat{f}^{(m)}(t)|^2\, dt\right] \\ &= \int_0^1 |\bar{f}^{(m_1)}(t) - \bar{f}^{(m)}(t)|^2\, dt + \frac{(2m+1)\sigma^2}{NJ}.\end{aligned}$$

This risk is a sum of two nonnegative terms. The first one is a bias term that is small when m is close to m_1 while the second one is a variance term that is small when m is close to zero. The aim is to find a trade-off between these two terms thanks to the data only. More precisely, we choose some $\hat{m} \in \{1, \dots, m_1\}$ such that

$$\hat{m} \in \operatorname*{argmin}_{m\in\{1,\dots,m_1\}} \left\{ \int_0^1 |\hat{f}^{(m_1)}(t) - \hat{f}^{(m)}(t)|^2\, dt + \eta \frac{(2m+1)\sigma^2}{NJ} \right\}, \tag{11}$$

where $\eta > 1$ is some constant. In the sequel, the estimator that we finally consider is $\hat{f}_{n,J} = \hat{f}^{(\hat{m})}$.

Such a procedure is well known, and we refer to Chapter 4 of [21] for more details. In particular, the estimator $\hat{f}_{n,J}$ satisfies the following inequality:

$$\begin{aligned}&\mathbb{E}^{(1)}\left[\int_0^1 |\bar{f}^{(m_1)}(t) - \hat{f}_{n,J}(t)|^2\, dt\right] \\ &\quad \le C(\eta)\left\{ \min_{m\in\{1,\dots,m_1\}} \mathbb{E}^{(1)}\left[\int_0^1 |\bar{f}^{(m_1)}(t) - \hat{f}^{(m)}(t)|^2\, dt\right] + \frac{\sigma^2}{NJ} \right\} \\ &\quad \le C(\eta) \min_{m\in\{1,\dots,m_1\}} \left\{ \int_0^1 |\bar{f}^{(m_1)}(t) - \bar{f}^{(m)}(t)|^2\, dt + \frac{2(m+1)\sigma^2}{NJ} \right\},\end{aligned} \tag{12}$$

where $C(\eta) > 0$ only depends on η. It is known that an optimal choice for η is a difficult problem from a theoretical point of view. However, in practice, taking some η slightly greater than 2 leads to a procedure that behaves well as we discuss in Section 6. Moreover, for real data analysis, we often have to estimate the variance σ^2.

3.3. *Convergence rates over Sobolev balls.* Let us denote by $\lfloor x \rfloor$ the largest integer smaller than $x \in \mathbb{R}$. We now focus on the performances of our estimation procedure from the minimax point of view and with respect to the distance d defined in (4). Note that, in Section 3.2, we only use truncated Fourier series expansions for building the estimators $\hat{f}^{(m)}$. In practice, we could use other bases of $L^2_{\text{per}}([0,1])$, and we would still have a result like (12). In particular, the following theorem would remain true by combining model selection techniques with bases like piecewise polynomials or orthonormal wavelets to approximate a function.

THEOREM 3.2. *Assume that $nJ \geq \max\{21J, (4\sigma^2)^{2s+1}/c^{2s}\}$ where $0 < c < 1$ is such that $J \leq cn^\alpha$ for some $\alpha > 0$. Take $m_1 = \lfloor N/2 \rfloor - 1$ and let $s > 3/2$ and $A > 0$. Then the estimator $\hat{f}_{n,J}$ defined by procedure (11) is such that, for any $g \in \mathcal{G}^\kappa$,*

$$\sup_{f \in \tilde{W}_s(A,c_*)} \mathcal{R}_g(\hat{f}_{n,J}, f) \leq C\left(|m_1|^{-2s} + m_1 n^{-2s+1} + \frac{1}{J}\mathbb{E}^g(\|\hat{\theta}^0 - \theta^0\|^2) + (nJ)^{-2s/(2s+1)}\right)$$

for some $C > 0$ that only depends on A, s, σ^2, κ, k_0, η and c.

Therefore, using the results of Corollary 3.1 on the convergence rate of $\hat{\theta}^0$ to θ^0, we finally obtain the following result.

COROLLARY 3.2. *Suppose that the assumptions of Theorems 3.1 and 3.2 are satisfied. If $k_0 \geq 1$ is a fixed integer and $J \asymp n^\alpha$ for some $0 < \alpha \leq 1/6$, then there exists $C' > 0$ that only depends on A, s, σ^2, κ, k_0, η, c_* and c such that*

$$\sup_{g \in \mathcal{G}^\kappa} \sup_{f \in \tilde{W}_s(A,c_*)} \mathcal{R}_g(\hat{f}_{n,J}, f) \leq C'(n^{-1} + (nJ)^{-2s/(2s+1)}).$$

4. A lower bound on the risk. The following theorem gives a lower bound on the risk over the Sobolev ball $\tilde{W}_s(A, c_*)$.

THEOREM 4.1. *Let us recall that*

$$\mathcal{R}_{n,J}(\tilde{W}_s(A,c_*), \mathcal{G}^\kappa) = \inf_{\hat{f}_{n,J}} \sup_{g \in \mathcal{G}^\kappa} \sup_{f \in \tilde{W}_s(A,c_*)} \mathcal{R}_g(\hat{f}_{n,J}, f).$$

There exists a constant $C > 0$ that only depends on A, s, c_* and σ^2 such that

$$\liminf_{\min(n,J) \to +\infty} (nJ)^{2s/(2s+1)} \mathcal{R}_{n,J}(\tilde{W}_s(A,c_*), \mathcal{G}^\kappa) \geq C.$$

From the results of the previous sections, we also easily obtain the following upperbound on the risk.

COROLLARY 4.1. *Suppose that the assumptions of Corollary 3.2 hold, and assume that $2\alpha s \leq 1$. Then, there exists a constant $C' > 0$ that only depends on A, s, σ^2, κ, k_0, η, c_* and c such that*

$$\sup_{g \in \mathcal{G}^\kappa} \sup_{f \in \tilde{W}_s(A, c_*)} \mathbb{E}^g(d^2([\hat{f}_{n,J}], [f])) \leq C'(nJ)^{-2s/(2s+1)}. \tag{13}$$

Note that inequality (13) is a direct consequence of Corollary 3.2 and the fact that $n^{-1} = ((nJ)^{-2s/(2s+1)})$ in the settings $2\alpha s \leq 1$ and $J \asymp n^\alpha$. Therefore, under the assumption that $2\alpha s \leq 1$, the smoothed Fréchet mean converges at the optimal rate $(nJ)^{-2s/(2s+1)}$.

5. Discussion. As explained previously, the rate of convergence $r_{n,J} = n^{-1} + (nJ)^{-2s/(2s+1)}$ of the estimator $\hat{f}_{n,J}$ is the sum of two terms having different interpretations. The term $(nJ)^{-2s/(2s+1)}$ is the usual nonparametric rate that would be obtained if the random shifts $\theta_1^*, \dots, \theta_J^*$ were known. To interpret the second term n^{-1}, let us mention the following result that has been obtained in [5].

PROPOSITION 5.1. *Suppose that the function f is continuously differentiable. Assume that the density $g \in \mathcal{G}^\kappa$ with $g(-\kappa/2) = g(\kappa/2) = 0$ and that $\mathcal{I}_g^2 = \int_{-1/2}^{1/2} (\frac{\partial}{\partial\theta} \log g(\theta))^2 g(\theta)\, d\theta < +\infty$. Let $(\hat{\theta}_1, \dots, \hat{\theta}_J)$ denote any estimator of the true shifts $(\theta_1^*, \dots, \theta_J^*)$ computed from the $Y_{\ell,j}$'s in model (3). Then*

$$\mathbb{E}^g\left(\frac{1}{J}\sum_{j=1}^J (\hat{\theta}_j - \theta_j^*)^2\right) \geq \frac{\sigma^2}{n\int_0^1 |f'(t)|^2\, dt + \sigma^2 \mathcal{I}_g^2}. \tag{14}$$

Proposition (5.1) shows that it is not possible to build consistent estimators of the shifts by considering only the asymptotic setting where the number of curves J tends toward infinity. Indeed inequality (14) implies that $\liminf_{J\to+\infty} \mathbb{E}^g(\frac{1}{J}\sum_{j=1}^J (\hat{\theta}_j - \theta_j^*)^2) > 0$ for any estimators $(\hat{\theta}_1, \dots, \hat{\theta}_J)$. We recall that, under the assumptions of Corollary 3.1, one has

$$\mathbb{E}^g\left(\frac{1}{J}\sum_{j=1}^J |\hat{\theta}_j^0 - \theta_j^0|^2\right) \leq \frac{C_1}{n}.$$

The above inequality shows that, in the setting where n and J are both allowed to increase, the estimation of the unknown shifts $\hat{\theta}_j^0 = \hat{\theta}_j^* - \frac{1}{J}\sum_{m=1}^J \hat{\theta}_m^*$ is feasible at the rate n^{-1}. By Proposition 5.1, this rate of convergence cannot be improved. We thus interpret the term n^{-1} appearing in the rate $r_{n,J}$ of the smoothed Fréchet mean $[\hat{f}_{n,J}]$ as the price to pay for having to estimate the shifts to compute such estimators.

To conclude this discussion, we would like to mention the results that have been obtained in [6] in an asymptotic setting where only the number J of curves is let going to infinity. Consider the following model of randomly shifted curves with additive white noise:

$$dY_j(t) = f(t-\theta_j^*)\,dt + \varepsilon\,dW_j(t), \qquad t\in[0,1], j=1,\dots,J \text{ with } \theta_j^* \sim_{\text{i.i.d.}} g, \tag{15}$$

where the W_j's are independent Brownian motions with $\varepsilon>0$ being the level of additive noise. In model (15), the expectation of each observed curve dY_j is equal to the convolution of f by the density g since

$$\mathbb{E}^g[f(t-\theta_j^*)] = \int f(t-\theta)g(\theta)\,d\theta = f\star g(t).$$

Therefore, in the ideal situation where g is assumed to be known, it has been shown in [6] that estimating f in the asymptotic setting $J\to+\infty$ (with $\varepsilon>0$ being fixed) is a deconvolution problem. Indeed, suppose that, for some $\nu>1/2$,

$$\gamma_k = \int_0^1 g(\theta)e^{-i2\pi k\theta}\,d\theta \asymp |k|^{-\nu}, \qquad k\in\mathbb{Z},$$

with g being known. Then, one can construct an estimator $\hat{f}_J^*$ by a deconvolution procedure such that

$$\sup_{f\in W_s(A)} \mathbb{E}^g \int_0^1 |\hat{f}_J^*(t)-f(t)|^2\,dt \le CJ^{-2s/(2s+\nu+1)}$$

for some $C>0$ that only depends on A, s and ε and where $W_s(A)$ is Sobolev ball of degree $s>1/2$. Moreover, this rate of convergence is optimal since the results in [6] show that if $s>2\nu+1$, then there exists a constant $C'>0$ that only depends on A, s and ε such that

$$\liminf_{J\to+\infty} J^{2s/(2s+2\nu+1)} \inf_{\hat{f}_J} \sup_{f\in W_s(A)} \mathcal{R}(\hat{f}_J, f) \ge C',$$

where the above infimum is taken over the set of all estimators $\hat{f}_J$ of f in model (15). Hence, $r_J = J^{-2s/(2s+2\nu+1)}$ is the minimax rate of convergence over Sobolev balls in model (15) in the case of known g. This rate is of polynomial order of the number of curves J, and it deteriorates as the smoothness ν of the convolution kernel g increases. This phenomenon is a well-known fact in deconvolution problems; see, for example, [9, 23]. Hence, depending on g being known or not and the choice of the asymptotic setting, there exists a significant difference in the rates of convergence that can be achieved in a randomly shifted curves model. Our setting yields the rate $r_{n,J} = n^{-1} + (nJ)^{-2s/(2s+1)}$ (in the case where $J\asymp n^\alpha$ with $\alpha<1/6$)

that is clearly faster than the rate $r_J = J^{-2s/(2s+2\nu+1)}$. Nevertheless, the arguments in [6] also suggest that a smoothed Fréchet mean in (15) is not a consistent estimator of f if one only lets J going to infinity. Therefore, the number n of design points is clearly of primary importance to obtain consistent estimators of a mean pattern when using Fréchet means of curves.

6. Numerical experiments. The goal of this section is to study the performance of the estimator $\hat{f}_{n,J}$. The factors in the simulations are the number J of curves and the number n of design points. As a mean pattern f to recover, we consider the two test functions displayed in Figure 1. Then, for each combination of n and J, we generate $M = 100$ repetitions of model (3) of J curves with shifts sampled from the uniform distribution on $[-\kappa, \kappa]$ with $\kappa = 1/16$. The level of the additive Gaussian noise is measured as the root of the signal-to-noise ratio (rsnr) defined as

$$\text{rsnr} = \left(\frac{1}{\sigma^2} \int_0^1 (f(t) - \bar{f})^2 \, dt \right)^{1/2} \qquad \text{where } \bar{f} = \int_0^1 f(t) \, dt,$$

that is fixed to $\text{rsnr} = 0.5$ in all the simulations. Samples of noisy randomly shifted curves are displayed in Figure 1. For each repetition $p \in \{1, \dots, M\}$, we compute the estimator $\hat{f}_{n,J,p}$ using a gradient descent algorithm to minimize criterion (9) for estimating the shifts. For all values of n and J, we took $k_0 = 5$ in (7). The frequency cut-off $\hat{m}$ is chosen using (11) with $\eta = 2.5$.

To analyze the numerical performance of this estimator, we have considered the following ideal estimator that uses the knowledge of the true random shifts $\theta^*_{j,p}$ (sampled from the pth replication):

$$\tilde{f}^{(m)}_{n,J,p}(t) = \sum_{|k| \le m} \left(\frac{1}{J} \sum_{j=1}^{J} \hat{c}^{(1)}_{k,j,p} e^{i2\pi k \theta^*_{j,p}} \right) \phi_k(t), \qquad t \in [0,1].$$

The frequency cut-off $\hat{m}_*$ for the above ideal estimator is chosen using a model selection procedure based on the knowledge of the true shifts, that is,

$$\hat{m}_* \in \operatorname*{argmin}_{m \in \{1, \dots, m_1\}} \left\{ \int_0^1 |\tilde{f}^{(m_1)}_{n,J,p}(t) - \tilde{f}^{(m)}_{n,J,p}(t)|^2 \, dt + \eta \frac{(2m+1)\sigma^2}{NJ} \right\}$$

with $\eta = 2.5$.

Then, we define the relative empirical error between the two estimators as

$$R(n, J) = \frac{1/M \sum_{p=1}^{M} d^2([\hat{f}_{n,J,p}], [f])}{1/M \sum_{p=1}^{M} d^2([\tilde{f}^{(\hat{m}_*)}_{n,J,p}], [f])}.$$

In Figure 2, we display the ratio $R(n, J)$ for various values of n and J and for the two test functions displayed in Figure 1. It can be seen that the function

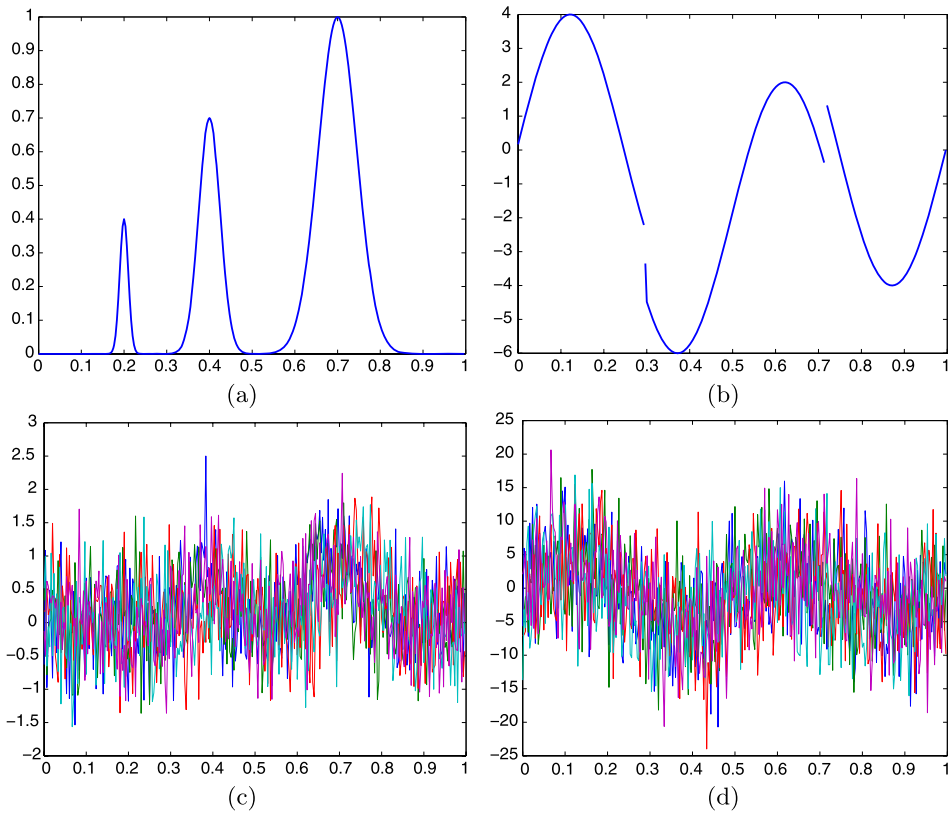
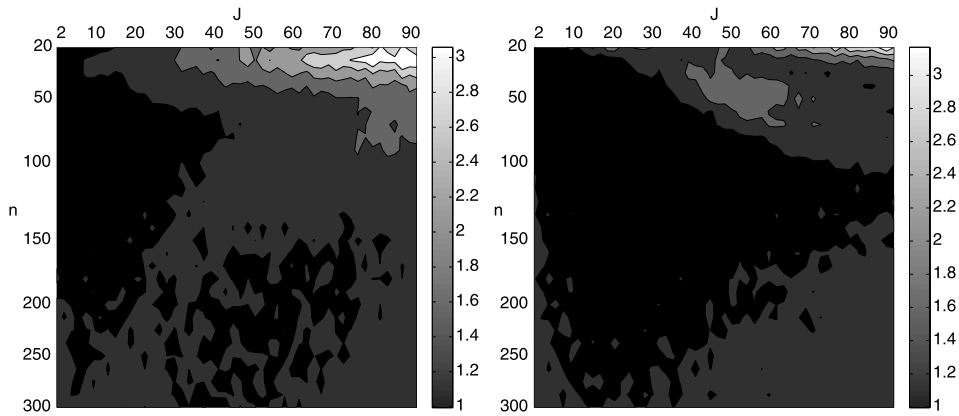


FIG. 1. *Two test functions f.* (a) *MixtGauss: a mixture of three Gaussians.* (b) *HeaviSine: a piecewise smooth curve with a discontinuity. Sample of* 5 *noisy randomly shifted curves with* $n = 300$ *for* (c) *MixtGauss and* (d) *HeaviSine.*

$J \mapsto R(n, J)$ is increasing. This means that the numerical performance of the estimator $\hat{f}_{n,J}$ deteriorates as the number J of curves increases and the number n remains fixed. This is clearly due to the fact that the estimation of the shifts becomes less precise when the dimension J increases. These numerical results are thus consistent with inequality (10) in Theorem 3.1 and our discussion on the rate of convergence of $\hat{f}_{n,J}$ in Section 3. On the other hand, the function $n \mapsto R(n, J)$ is decreasing, and it confirms that the number n of design points is a key parameter to obtain consistent estimators of a mean pattern f with Fréchet means of curves.

APPENDIX: PROOF OF THE MAIN RESULTS

Throughout the proofs, we repeatedly use the following lemma which follows immediately from Lemma 1.8 in [30].



(a) (b)

FIG. 2. *Relative empirical error $R(n, J)$ for various values of n (vertical axis) and J (horizontal axis) over $M = 100$ replications:* (a) *MixtGauss,* (b) *HeaviSine.*

LEMMA A.1. *If $f \in \tilde{W}_s(A, c_*)$, then there exists a constant $A_0 > 0$ only depending on A and s such that*

$$\max_{-N/2 \le k < N/2} |\bar{c}_{k,j}^{(p)} - c_k e^{-i2\pi k \theta_j^*}| \le A_0 N^{-s+1/2}, \qquad p \in \{0, 1\}$$

for all $1 \le j \le J$.

A.1. Proof of Theorem 3.1. For legibility, we will write $\mathbb{E} = \mathbb{E}^g$, that is, we omit the dependency on g of the expectation. The proof is divided in several lemmas. Let $\|\cdot\|$ denote the standard Euclidean norm in $\mathbb{R}^J$. First, we derive upper bounds on the second, fourth and sixth moments of $\|\hat{\theta}^0 - \theta^0\|$. The following upper bound on the second moment is weaker than the result that we plan to prove. It only gives the consistency of $\hat{\theta}^0$, and we will need some additional arguments to get the announced rate of convergence (10).

LEMMA A.2. *Let $N \ge 2$, $J \ge 1$ and $1 \le k_0 \le N/2$. We assume that Assumptions 1 and 2 are satisfied, and we suppose that $s > 3/2$. Then, we have the following upper bounds, for any $f \in \tilde{W}_s(A, c_*)$:*

$$\frac{1}{J}\mathbb{E}(\|\hat{\theta}^0 - \theta^0\|^2) \le C_1 \frac{k_0^{1/2}}{n^{1/2}}, \tag{16}$$

$$\frac{1}{J}\mathbb{E}(\|\hat{\theta}^0 - \theta^0\|^4) \le C_2 \frac{k_0 J}{n} \tag{17}$$

and

$$\frac{1}{J}\mathbb{E}(\|\hat{\theta}^0 - \theta^0\|^6) \le C_3 \frac{k_0^{3/2} J^2}{n^{3/2}}, \tag{18}$$

where C_1, C_2 and C_3 are positive constants that only depend on A, s, c_*, σ^2 and κ.

PROOF. Let $f \in \tilde{W}_s(A, c_*)$. Since $\hat{\theta}^0 = (\hat{\theta}^0_1, \dots, \hat{\theta}^0_J)$ is a minimizer of M_n, it follows that

$$M(\hat{\theta}^0) - M(\theta^0) \leq 2 \sup_{\theta \in \Theta_\kappa} |M_n(\theta) - M(\theta)|.$$

Therefore, by Proposition 3.1, we get

$$\frac{1}{J}\mathbb{E}\|\hat{\theta}^0 - \theta^0\|^2 \leq 2C^{-1}(c_*, \kappa)\mathbb{E}\Big(\sup_{\theta \in \Theta_\kappa} |M_n(\theta) - M(\theta)|\Big), \tag{19}$$

$$\frac{1}{J}\mathbb{E}\|\hat{\theta}^0 - \theta^0\|^4 \leq 4C^{-2}(c_*, \kappa)J\mathbb{E}\Big(\sup_{\theta \in \Theta_\kappa} |M_n(\theta) - M(\theta)|^2\Big) \tag{20}$$

and

$$\frac{1}{J}\mathbb{E}\|\hat{\theta}^0 - \theta^0\|^6 \leq 8C^{-3}(c_*, \kappa)J^2\mathbb{E}\Big(\sup_{\theta \in \Theta_\kappa} |M_n(\theta) - M(\theta)|^3\Big), \tag{21}$$

where we have set $C(c_*, \kappa) = 4\pi^2 c_*^2 \cos(8\pi\kappa)$.

Let $\theta \in \Theta_\kappa$ and note that $M_n(\theta)$ can be decomposed as

$$M_n(\theta) = \bar{M}(\theta) + Q(\theta) + L(\theta), \tag{22}$$

where

$$\bar{M}(\theta) = \frac{1}{J}\sum_{j=1}^{J}\sum_{|k| \leq k_0}\left|\bar{c}^{(0)}_{k,j}e^{2ik\pi\theta_j} - \frac{1}{J}\sum_{j'=1}^{J}\bar{c}^{(0)}_{k,j'}e^{2ik\pi\theta_{j'}}\right|^2,$$

$$Q(\theta) = \frac{1}{NJ}\sum_{j=1}^{J}\sum_{|k| \leq k_0}\left|z^{(0)}_{k,j}e^{2ik\pi\theta_j} - \frac{1}{J}\sum_{j'=1}^{J}z^{(0)}_{k,j'}e^{2ik\pi\theta_{j'}}\right|^2$$

and

$$L(\theta) = \frac{2}{J\sqrt{N}}\sum_{j=1}^{J}\sum_{|k| \leq k_0}\Re\Bigg[\left(\bar{c}^{(0)}_{k,j}e^{2ik\pi\theta_j} - \frac{1}{J}\sum_{j'=1}^{J}\bar{c}^{(0)}_{k,j'}e^{2ik\pi\theta_{j'}}\right) \times \overline{\left(z^{(0)}_{k,j}e^{2ik\pi\theta_j} - \frac{1}{J}\sum_{j'=1}^{J}z^{(0)}_{k,j'}e^{2ik\pi\theta_{j'}}\right)}\Bigg].$$

Using Lemma A.1, it follows that, for any $\theta \in \Theta_\kappa$,

$$|\bar{M}(\theta) - M(\theta)| \leq \frac{1}{J}\sum_{j=1}^{J}\sum_{|k| \leq k_0}\left||\bar{c}^{(0)}_{k,j}| + \frac{1}{J}\sum_{j'=1}^{J}|\bar{c}^{(0)}_{k,j'}| + 2|c_k|\right|$$

$$\times \Bigg| (\bar{c}_{k,j}^{(0)} - c_k e^{-2ik\pi\theta_j^*}) e^{2ik\pi\theta_j} - \frac{1}{J}\sum_{j'=1}^{J} (\bar{c}_{k,j'}^{(0)} - c_k e^{-2ik\pi\theta_{j'}^*}) e^{2ik\pi\theta_{j'}} \Bigg|$$

$$\leq 2A_0 N^{-s+1/2} \sum_{|k|\leq k_0} (4|c_k| + 2A_0 N^{-s+1/2})$$

$$\leq 8A_0 N^{-s+1/2}(2k_0+1)^{1/2} \sqrt{\sum_{|k|\leq k_0} |c_k|^2} + 4A_0^2(2k_0+1)N^{-2s+1}.$$

Hence, there exists a positive constant C that only depends on A and s such that

$$\sup_{\theta\in\Theta_\kappa} |\bar{M}(\theta) - M(\theta)| \leq C k_0^{1/2} N^{-s+1/2}. \tag{23}$$

Now, note that $Q(\theta) \leq \frac{\sigma^2}{NJ} Z$ with $Z = \sum_{|k|\leq k_0} \sum_{j=1}^{J} |z_{k,j}^{(0)}/\sigma|^2$ for any $\theta \in \Theta_\kappa$. Thus, it follows that

$$\mathbb{E} \sup_{\theta\in\Theta_\kappa} |Q(\theta)| \leq \sigma^2 (2k_0+1) N^{-1} \tag{24}$$

and

$$\mathbb{E} \sup_{\theta\in\Theta_\kappa} |Q(\theta)|^2 \leq 2\sigma^4 (2k_0+1)^2 N^{-2}. \tag{25}$$

By Jensen's inequality, we get

$$\mathbb{E} Z^{3/2} \leq (\mathbb{E} Z^2)^{3/4},$$

and since $\mathbb{E} Z^2 \leq 2J^2(2k_0+1)^2$, we obtain

$$\mathbb{E} \sup_{\theta\in\Theta_\kappa} |Q(\theta)|^{3/2} \leq 8^{1/4} \frac{\sigma^3}{N^{3/2}} (2k_0+1)^{3/2}. \tag{26}$$

Finally, using $\mathbb{E} Z^3 \leq 6J^3(2k_0+1)^3$, we have

$$\mathbb{E} \sup_{\theta\in\Theta_\kappa} |Q(\theta)|^3 \leq 6 \frac{\sigma^6}{N^3} (2k_0+1)^3. \tag{27}$$

By Cauchy–Schwarz's inequality,

$$L(\theta) \leq 2\sqrt{\bar{M}(\theta)} \sqrt{Q(\theta)}. \tag{28}$$

Thanks to Lemma A.1, we get

$$\bar{M}(\theta) \leq \frac{1}{J} \sum_{|k|\leq k_0} \sum_{j=1}^{J} |\bar{c}_{k,j}^{(0)}|^2 \leq \sum_{|k|\leq k_0} |c_k|^2 + A_0^2 N^{-2s+1}(2k_0+1).$$

Thus, it follows from (24), (25), (26) and (28) that there exists a positive constant C', only depending on A, s and σ^2, such that

$$\mathbb{E} \sup_{\theta \in \Theta_\kappa} |L(\theta)| \leq C' k_0^{1/2} N^{-1/2}, \tag{29}$$

$$\mathbb{E} \sup_{\theta \in \Theta_\kappa} |L(\theta)|^2 \leq C'^2 k_0 N^{-1} \tag{30}$$

and

$$\mathbb{E} \sup_{\theta \in \Theta_\kappa} |L(\theta)|^3 \leq C'^3 k_0^{3/2} N^{-3/2}. \tag{31}$$

Since $s > 3/2$, we obtain, by inequalities (23), (24) and (29),

$$\mathbb{E}\Big(\sup_{\theta \in \Theta_\kappa} |M_n(\theta) - M(\theta)| \Big) \leq C_1' k_0^{1/2} N^{-1/2},$$

by inequalities (23), (25) and (30),

$$\mathbb{E}\Big(\sup_{\theta \in \Theta_\kappa} |M_n(\theta) - M(\theta)|^2 \Big) \leq C_2' k_0 N^{-1},$$

and by inequalities (23), (27) and (31),

$$\mathbb{E}\Big(\sup_{\theta \in \Theta_\kappa} |M_n(\theta) - M(\theta)|^3 \Big) \leq C_3' k_0^{3/2} N^{-3/2},$$

where C_1', C_2' and C_3' are positive constants that only depend on A, s and σ^2. Combined with inequalities (19), (20) and (21), the announced result follows from the above upper bounds. □

In order to prove Theorem 3.1, we divide the rest of the proof in the three following steps. In the sequel of this section, we always assume that the hypotheses of Theorem 3.1 are satisfied, and we use the decomposition of $M_n(\theta)$ as defined in (22).

Step 1: there exists some positive constant C_1 that only depends on c_* such that

$$\begin{aligned} &\frac{n}{J} \|\hat{\theta}^0 - \theta^0\|^2 \\ &\quad \leq C_1 \Big(nJ \|\nabla M_n(\theta^0)\|^2 \\ &\qquad + nJ \sup_{\theta \in \mathcal{U}_\kappa} \|\nabla^2 M_n(\theta) - \nabla^2 M(\theta^0)\|_{\mathrm{op}}^2 \|\hat{\theta}^0 - \theta^0\|^2 \Big), \end{aligned} \tag{32}$$

where ∇ and ∇^2 denote the gradient and the Hessian operators, respectively, and where we have set

$$\mathcal{U}_\kappa = \{\theta \in \Theta_\kappa \text{ such that } \|\theta - \theta^0\| \leq \|\hat{\theta}^0 - \theta^0\|\}$$

and, for any $J \times J$ matrix B, the operator norm $\|B\|_{\mathrm{op}}$ is defined by

$$\|B\|_{\mathrm{op}} = \sup_{\theta \in \mathbb{R}^J \setminus \{0\}} \frac{\|B\theta\|}{\|\theta\|}.$$

Step 2: there exists some positive constant C_2 that only depends on A, s and σ^2 such that

$$nJ\mathbb{E}\|\nabla M_n(\theta^0)\|^2 \le C_2\left(1 + \frac{k_0^3}{n}\right). \tag{33}$$

Step 3: there exists some positive constant C_3 that only depends on A, s, κ, c_* and σ^2 such that

$$\begin{aligned} &nJ\mathbb{E}\Big(\sup_{\theta \in \mathcal{U}_\kappa} \|\nabla^2 M_n(\theta) - \nabla^2 M(\theta^0)\|_{\mathrm{op}}^2 \|\hat{\theta}^0 - \theta^0\|^2\Big) \\ &\qquad \le C_3\left(1 + \frac{k_0^5}{n^{1/2}}\right)\frac{k_0^{3/2} J^3}{n^{1/2}}. \end{aligned} \tag{34}$$

The result announced in Theorem 3.1 follows from inequalities (32), (33) and (34).

A.1.1. *Proof of Step 1.* The gradients of $\bar{M}(\theta), Q(\theta)$ and $L(\theta)$ follow from easy computations. We have, for any $1 \le \ell \le J$,

$$\frac{\partial}{\partial \theta_\ell}\bar{M}(\theta) = \frac{4\pi}{J^2}\sum_{|k| \le k_0} k\Re\left[\overline{i\bar{c}_{k,\ell}^{(0)} e^{2ik\pi\theta_\ell}}\left(\sum_{j=1}^J \bar{c}_{k,j}^{(0)} e^{2ik\pi\theta_j}\right)\right], \tag{35}$$

$$\frac{\partial}{\partial \theta_\ell}Q(\theta) = \frac{4\pi}{NJ^2}\sum_{|k| \le k_0} k\Re\left[\overline{i z_{k,\ell}^{(0)} e^{2ik\pi\theta_\ell}}\left(\sum_{j=1, j\ne \ell}^J z_{k,j}^{(0)} e^{2ik\pi\theta_j}\right)\right] \tag{36}$$

and

$$\begin{aligned} \frac{\partial}{\partial \theta_\ell}L(\theta) = -\frac{4\pi}{J^2\sqrt{N}}\sum_{|k| \le k_0} k\Im\Bigg[&\overline{z_{k,\ell}^{(0)} e^{2ik\pi\theta_\ell}}\left(\sum_{j=1, j\ne \ell}^J \bar{c}_{k,j}^{(0)} e^{2ik\pi\theta_j}\right) \\ &- \bar{c}_{k,\ell}^{(0)} e^{2ik\pi\theta_\ell}\overline{\left(\sum_{j=1, j\ne \ell}^J z_{k,j}^{(0)} e^{2ik\pi\theta_j}\right)}\Bigg]. \end{aligned} \tag{37}$$

Similarly, we can compute the Hessians of these functions as follows, for $1 \le \ell, \ell' \le J$, if $\ell \ne \ell'$,

$$\frac{\partial^2}{\partial \theta_{\ell'}\,\partial \theta_\ell}\bar{M}(\theta) = -\frac{8\pi^2}{J^2}\sum_{|k| \le k_0} k^2\Re[\overline{\bar{c}_{k,\ell}^{(0)}}\bar{c}_{k,\ell'}^{(0)} e^{2ik\pi(\theta_{\ell'} - \theta_\ell)}], \tag{38}$$

$$\frac{\partial^2}{\partial\theta_{\ell'}\,\partial\theta_\ell}Q(\theta) = -\frac{8\pi^2}{NJ^2}\sum_{|k|\le k_0} k^2\Re[\overline{z_{k,\ell}^{(0)}}z_{k,\ell'}^{(0)}e^{2ik\pi(\theta_{\ell'}-\theta_\ell)}] \tag{39}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial\theta_{\ell'}\,\partial\theta_\ell}L(\theta) = -\frac{8\pi^2}{J^2\sqrt{N}}\sum_{|k|\le k_0} k^2\Re[&\bar{c}_{k,\ell'}^{(0)}\overline{z_{k,\ell}^{(0)}}e^{2ik\pi(\theta_{\ell'}-\theta_\ell)}\\ &+\bar{c}_{k,\ell}^{(0)}\overline{z_{k,\ell'}^{(0)}}e^{2ik\pi(\theta_\ell-\theta_{\ell'})}],\end{aligned} \tag{40}$$

and if $\ell=\ell'$,

$$\frac{\partial^2}{\partial\theta_\ell\,\partial\theta_\ell}\bar{M}(\theta) = \frac{8\pi^2}{J^2}\sum_{|k|\le k_0} k^2\Re\left[\overline{\bar{c}_{k,\ell}^{(0)}e^{2ik\pi\theta_\ell}}\left(\sum_{j=1,j\neq\ell}^{J}\bar{c}_{k,j}^{(0)}e^{2ik\pi\theta_j}\right)\right], \tag{41}$$

$$\frac{\partial^2}{\partial\theta_\ell\,\partial\theta_\ell}Q(\theta) = \frac{8\pi^2}{NJ^2}\sum_{|k|\le k_0} k^2\Re\left[\overline{z_{k,\ell}^{(0)}e^{2ik\pi\theta_\ell}}\left(\sum_{j=1,j\neq\ell}^{J}z_{k,j}^{(0)}e^{2ik\pi\theta_j}\right)\right] \tag{42}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial\theta_\ell\,\partial\theta_\ell}L(\theta) = \frac{8\pi^2}{J^2\sqrt{N}}\sum_{|k|\le k_0} k^2\Re\Bigg[&\bar{c}_{k,\ell}^{(0)}e^{2ik\pi\theta_\ell}\overline{\left(\sum_{j=1,j\neq\ell}^{J}z_{k,j}^{(0)}e^{2ik\pi\theta_j}\right)}\\ &+\overline{z_{k,\ell}^{(0)}e^{2ik\pi\theta_\ell}}\left(\sum_{j=1,j\neq\ell}^{J}\bar{c}_{k,j}^{(0)}e^{2ik\pi\theta_j}\right)\Bigg].\end{aligned} \tag{43}$$

Using the fact that $\hat{\theta}^0\in\Theta_\kappa$ is a minimizer of M_n, so $\nabla M_n(\hat{\theta}^0)=0$, a Taylor expansion of $\theta\mapsto\nabla M_n(\theta)$ with an integral form of the remainder term leads to

$$0 = \nabla M_n(\theta^0) + \int_0^1 \nabla^2 M_n(\bar{\theta}(t))(\hat{\theta}^0-\theta^0)\,dt, \tag{44}$$

where, for any $t\in[0,1]$, we have set

$$\bar{\theta}(t) = \theta^0 + t(\hat{\theta}^0-\theta^0)\in\mathcal{U}_\kappa.$$

Thus, we have

$$\begin{aligned}&\nabla^2 M(\theta^0)(\hat{\theta}^0-\theta^0)\\ &\quad= -\nabla M_n(\theta^0) - \int_0^1(\nabla^2 M_n(\bar{\theta}(t)) - \nabla^2 M(\theta^0))(\hat{\theta}^0-\theta^0)\,dt.\end{aligned} \tag{45}$$

It follows from similar computations as we did for $\bar{M}$ that

$$\nabla^2 M(\theta^0) = \frac{8\pi^2}{J} \sum_{|k|\le k_0} k^2 |c_k|^2 \left(I_J - \frac{1}{J} \mathbb{1}_J \right),$$

where I_J is the $J \times J$ identity matrix and $\mathbb{1}_J$ denotes the $J \times J$ matrix with all entries equal to one. Therefore, using the fact that $\sum_{j=1}^J (\hat{\theta}_j^0 - \theta_j^0) = 0$, we obtain

$$\|\nabla^2 M(\theta^0)(\hat{\theta}^0 - \theta^0)\|^2 = \frac{64\pi^4}{J^2} \left(\sum_{|k|\le k_0} k^2 |c_k|^2 \right)^2 \|\hat{\theta}^0 - \theta^0\|^2,$$

and it shows that there exists a constant $C > 0$ that only depends on c_* such that

$$\|\nabla^2 M(\theta^0)(\hat{\theta}^0 - \theta^0)\|^2 \ge C \frac{1}{J^2} \|\hat{\theta}^0 - \theta^0\|^2. \tag{46}$$

Then, inequality (32) follows from (45) and (46).

A.1.2. *Proof of Step 2.* By using Lemma A.1, for any $1 \le k \le k_0$ and $1 \le \ell \le J$, we can expand

$$\bar{c}_{k,\ell}^{(0)} e^{2ik\pi\theta_\ell^0} = c_k e^{2ik\pi(\theta_\ell^0 - \theta_\ell^*)} + \alpha_{k,\ell} e^{2ik\pi\theta_\ell^0}$$

with $|\alpha_{k,\ell}| \le A_0 N^{-s+1/2}$. Because, for any j, $\theta_j^0 - \theta_j^* = \bar{\theta}_J$ does not depend on j, we have

$$\left| \Re \left[i \overline{\bar{c}_{k,\ell}^{(0)} e^{2ik\pi\theta_\ell}} \left(\sum_{j=1}^J \bar{c}_{k,j}^{(0)} e^{2ik\pi\theta_j} \right) \right] \right| \le A_0 J N^{-s+1/2} (2|c_k| + A_0 N^{-s+1/2}).$$

Thus, by equation (35) and using Cauchy–Schwarz's inequality, we obtain

$$\begin{aligned} \left| \frac{\partial}{\partial \theta_\ell} \bar{M}(\theta^0) \right|^2 &\le \frac{32\pi^2}{J^2} (2k_0 + 1) A_0^2 N^{-2s+1} \sum_{|k|\le k_0} k^2 (4|c_k|^2 + A_0^2 N^{-2s+1}) \\ &\le \frac{64\pi^2}{J^2} (2k_0 + 1) A_0^2 N^{-2s+1} (A^2 + A_0^2 k_0^3 N^{-2s+1}). \end{aligned}$$

Thus, there exists a positive constant C that only depends on A and s such that

$$nJ \|\nabla \bar{M}(\theta^0)\|^2 \le C k_0 (1 + k_0^3 n^{-2s+1}) n^{-2s+2}. \tag{47}$$

We now focus on ∇Q and, by (36), we can obtain

$$\begin{aligned} \sum_{\ell=1}^J \mathbb{E} \left| \frac{\partial}{\partial \theta_\ell} Q(\theta^0) \right|^2 &\le \frac{16\pi^2 (J-1)^2}{N^2 J^4} \sum_{\ell=1}^J \sum_{|k|\le k_0} k^2 \frac{1}{J-1} \sum_{j=1, j\ne\ell}^J \mathbb{E} |z_{k,\ell}^{(0)}|^2 |z_{k,j}^{(0)}|^2 \\ &\le \frac{32\pi^2 \sigma^4 k_0^3}{N^2 J}. \end{aligned}$$

Hence, we get

$$nJ\mathbb{E}\|\nabla Q(\theta^0)\|^2 \leq \frac{128\pi^2\sigma^4 k_0^3}{n}. \tag{48}$$

Finally, we deal with ∇L. Equation (37) and Lemma A.1 imply

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial}{\partial\theta_\ell}L(\theta)\right|^2 &\leq \frac{16\pi^2\sigma^2}{J^4N}\sum_{|k|\leq k_0} k^2\left(\left|\sum_{j=1,j\neq\ell}^{J}\bar{c}_{k,j}^{(0)}e^{2ik\pi\theta_j}\right|^2 + J|\bar{c}_{k,\ell}^{(0)}|^2\right)\\
&\leq \frac{16\pi^2\sigma^2}{J^4N}\sum_{|k|\leq k_0} k^2\left((J-1)\sum_{j=1,j\neq\ell}^{J}|\bar{c}_{k,j}^{(0)}|^2 + J|\bar{c}_{k,\ell}^{(0)}|^2\right)\\
&\leq \frac{16\pi^2\sigma^2}{J^2N}\sum_{|k|\leq k_0} k^2\left(\frac{1}{J}\sum_{j=1}^{J}|\bar{c}_{k,j}^{(0)}|^2\right)\\
&\leq \frac{32\pi^2\sigma^2}{J^2N}\sum_{|k|\leq k_0} k^2(|c_k|^2 + A_0^2N^{-2s+1}).
\end{aligned}$$

This last inequality leads to

$$nJ\mathbb{E}\|\nabla L(\theta^0)\|^2 \leq 64\pi^2\sigma^2(A^2+2A_0^2)(1+k_0^3n^{-2s+1}). \tag{49}$$

By combining inequalities (47), (48) and (49), we then obtain inequality (33).

A.1.3. *Proof of Step 3.* We introduce the Frobenius norm $\|B\|_F$ defined, for any $J\times J$ matrix $B=[B_{\ell,\ell'}]_{1\leq\ell,\ell'\leq J}$, as

$$\|B\|_F = \sqrt{\sum_{\ell,\ell'=1}^{J} B_{\ell,\ell'}^2}.$$

Moreover, for a self-adjoint matrix B, we will use the classical inequalities

$$\|B\|_{\mathrm{op}} \leq \|B\|_F \quad \text{and} \quad \|B\|_{\mathrm{op}} \leq \max_{1\leq\ell'\leq J}\sum_{\ell=1}^{J}|B_{\ell,\ell'}|. \tag{50}$$

In order to prove inequality (34), we use the following decomposition:

$$\begin{aligned}
\|\nabla^2 M_n(\theta) - \nabla^2 M(\theta^0)\|_{\mathrm{op}}^2 \leq 4(&\|\nabla^2\bar{M}(\theta) - \nabla^2 M(\theta)\|_F^2\\
&+ \|\nabla^2 M(\theta) - \nabla^2 M(\theta^0)\|_{\mathrm{op}}^2\\
&+ \|\nabla^2 Q(\theta)\|_F^2 + \|\nabla^2 L(\theta)\|_F^2).
\end{aligned} \tag{51}$$

We now deal with each term in the above inequality. Hereafter, ℓ and ℓ' always denote two integers in $\{1,\dots,J\}$.

First, let us consider $\ell \neq \ell'$, by (38) and Lemma A.1, we get

$$\begin{aligned}
&\left|\frac{\partial^2}{\partial\theta_{\ell'}\,\partial\theta_\ell}\bar{M}(\theta) - \frac{\partial^2}{\partial\theta_{\ell'}\,\partial\theta_\ell}M(\theta)\right|^2 \\
&\quad = \frac{64\pi^4}{J^4}\left(\sum_{|k|\le k_0} k^2 \Re[(\bar{c}_{k,\ell}^{(0)}\overline{\bar{c}_{k,\ell'}^{(0)}} - |c_k|^2 e^{2ik\pi(\theta_{\ell'}^*-\theta_\ell^*)})e^{i2\pi k(\theta_\ell-\theta_{\ell'})}]\right)^2 \\
&\quad \le \frac{64\pi^4}{J^4}\left(\sum_{|k|\le k_0} k^2 |\bar{c}_{k,\ell}^{(0)}\overline{\bar{c}_{k,\ell'}^{(0)}} - |c_k|^2 e^{2ik\pi(\theta_{\ell'}^*-\theta_\ell^*)}|\right)^2 \\
&\quad \le \frac{256\pi^4}{J^4}\left(\sum_{|k|\le k_0} k^2 (A_0|c_k|N^{-s+1/2} + A_0^2 N^{-2s+1})\right)^2.
\end{aligned}$$

In the case of $\ell = \ell'$, (41) and Lemma A.1 lead to

$$\begin{aligned}
&\left|\frac{\partial^2}{\partial\theta_{\ell}\,\partial\theta_\ell}\bar{M}(\theta) - \frac{\partial^2}{\partial\theta_{\ell}\,\partial\theta_\ell}M(\theta)\right|^2 \\
&\quad \le \frac{64\pi^4}{J^4}\left(\sum_{|k|\le k_0} k^2 \sum_{j=1, j\neq\ell}^{J} |\bar{c}_{k,\ell}^{(0)}\overline{\bar{c}_{k,j}^{(0)}} - |c_k|^2 e^{2ik\pi(\theta_j^*-\theta_\ell^*)}|\right)^2 \\
&\quad \le \frac{256\pi^4}{J^2}\left(\sum_{|k|\le k_0} k^2 (A_0|c_k|N^{-s+1/2} + A_0^2 N^{-2s+1})\right)^2.
\end{aligned}$$

Therefore, the above inequalities and (16) imply that there exists some positive constant C_M that only depends on A, s, κ and σ^2 such that

$$nJ\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2\bar{M}(\theta) - \nabla^2 M(\theta)\|_F^2\|\hat{\theta}^0 - \theta^0\|^2\Big) \le C_M\frac{k_0^{6+1/2}J}{n^{2s-3/2}}. \tag{52}$$

Second, using the fact that $2(1-\cos(t)) \le t^2$ for any $t \in \mathbb{R}$, if $\ell \neq \ell'$, then we have

$$\begin{aligned}
&\left|\frac{\partial^2}{\partial\theta_{\ell'}\,\partial\theta_\ell}M(\theta) - \frac{\partial^2}{\partial\theta_{\ell'}\,\partial\theta_\ell}M(\theta^0)\right| \\
&\quad = \frac{8\pi^2}{J^2}\left|\sum_{|k|\le k_0} k^2|c_k|^2\Re[e^{2ik\pi(\theta_\ell-\theta_\ell^0+\theta_{\ell'}^0-\theta_{\ell'})} - 1]\right| \\
&\quad \le \frac{16\pi^4}{J^2}\left(\sum_{|k|\le k_0} k^4|c_k|^2\right)|\theta_\ell - \theta_\ell^0 + \theta_{\ell'}^0 - \theta_{\ell'}|^2,
\end{aligned}$$

and if $\ell = \ell'$ then we obtain

$$\begin{aligned}
&\left|\frac{\partial^2}{\partial\theta_\ell\,\partial\theta_\ell}M(\theta) - \frac{\partial^2}{\partial\theta_\ell\,\partial\theta_\ell}M(\theta^0)\right| \\
&\quad = \frac{8\pi^2}{J^2}\left|\sum_{|k|\le k_0} k^2|c_k|^2\Re\left[\left(\sum_{j=1,j\neq\ell}^{J}(e^{2ik\pi(\theta_\ell-\theta_\ell^0-\theta_j+\theta_j^0)}-1)\right)\right]\right| \\
&\quad \le \frac{16\pi^4}{J^2}\left(\sum_{|k|\le k_0} k^4|c_k|^2\right)\sum_{j=1,j\neq\ell}^{J}|\theta_\ell-\theta_\ell^0-\theta_j+\theta_j^0|^2.
\end{aligned}$$

Therefore, by (50) and under the condition $s \ge 2$, we get

$$\begin{aligned}
\|\nabla^2 M(\theta) - \nabla^2 M(\theta^0)\|_{\mathrm{op}} &\le \frac{32\pi^4 A^2}{J^2}\max_{1\le\ell\le J}\sum_{j=1,j\neq\ell}^{J}|\theta_\ell-\theta_\ell^0-\theta_j+\theta_j^0|^2 \\
&\le \frac{64\pi^4 A^2}{J}\|\theta-\theta^0\|^2.
\end{aligned}$$

Thus, by definition of $\mathcal{U}_\kappa$ and by (18), we know that there exists some $C'_M > 0$ that only depends on A, s, σ^2 and κ such that

$$\begin{aligned}
&nJ\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2 M(\theta)-\nabla^2 M(\theta^0)\|_{\mathrm{op}}^2\|\hat\theta^0-\theta^0\|^2\Big) \\
&\qquad \le \frac{n(64\pi^4A^2)^2}{J}\mathbb{E}(\|\hat\theta^0-\theta^0\|^6) \\
&\qquad \le C'_M\frac{k_0^{3/2}J^2}{n^{1/2}}.
\end{aligned} \tag{53}$$

Next, we deal with the term relative to $\|\nabla^2 Q(\theta)\|_F^2$. Let us begin by noting that $\|\hat\theta^0-\theta^0\|^2 \le 4J\kappa^2$. Thus, we have

$$\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2 Q(\theta)\|_F^2\|\hat\theta^0-\theta^0\|^2\Big) \le 4J\kappa^2\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2 Q(\theta)\|_F^2\Big). \tag{54}$$

If we take $\ell \neq \ell'$, then using (39), we get

$$\left|\frac{\partial^2}{\partial\theta_{\ell'}\,\partial\theta_\ell}Q(\theta)\right|^2 \le \frac{64\pi^4}{N^2J^4}\left(\sum_{|k|\le k_0} k^2|z_{k,\ell}^{(0)}||z_{k,\ell'}^{(0)}|\right)^2$$

and if $\ell = \ell'$, then by (42), we have

$$\begin{aligned}
|\frac{\partial^2}{\partial\theta_\ell^2}Q(\theta)\Big|^2 &\le \frac{64\pi^4}{N^2J^4}\left(\sum_{|k|\le k_0} k^2|z_{k,\ell}^{(0)}|\left|\sum_{j=1,j\neq\ell}^{J} z_{k,j}^{(0)}e^{i2\pi k\theta_j}\right|\right)^2 \\
&\le \frac{64\pi^4}{N^2J^3}\sum_{j=1,j\neq\ell}^{J}\left(\sum_{|k|\le k_0} k^2|z_{k,\ell}^{(0)}||z_{k,j}^{(0)}|\right)^2.
\end{aligned}$$

Hence, the Cauchy–Schwarz inequality leads to the following upper bound:

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2 Q(\theta)\|_F^2\Big)\\
&\quad\le \frac{128\pi^4}{N^2J^3}\mathbb{E}\sum_{\ell=1}^{J}\sum_{\ell'=1,\ell'\neq\ell}^{J}\Big(\sum_{|k|\le k_0}k^2|z_{k,\ell}^{(0)}||z_{k,\ell'}^{(0)}|\Big)^2\\
&\quad\le \frac{128\pi^4}{N^2J^3}\sum_{\ell=1}^{J}\sum_{\ell'=1,\ell'\neq\ell}^{J}\mathbb{E}\Big(\sum_{|k|\le k_0}k^2|z_{k,\ell}^{(0)}|^2\Big)\mathbb{E}\Big(\sum_{|k|\le k_0}k^2|z_{k,\ell'}^{(0)}|^2\Big)\\
&\quad\le \frac{512\pi^4\sigma^4k_0^6}{N^2J}.
\end{aligned}$$

Combining this bound with (54) gives us some $C_Q>0$ that only depends on κ and σ^2 such that

$$nJ\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2 Q(\theta)\|_F^2\|\hat\theta^0-\theta^0\|^2\Big)\le C_Q\frac{k_0^6J}{n}. \tag{55}$$

Finally, we focus on the term concerning $\|\nabla^2 L(\theta)\|_F^2$. By the Cauchy–Schwarz inequality and (17), we have

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2 L(\theta)\|_F^2\|\hat\theta^0-\theta^0\|^2\Big)\\
&\quad\le\sqrt{\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2 L(\theta)\|_F^4\Big)}\sqrt{\mathbb{E}(\|\hat\theta^0-\theta^0\|^4)}\\
&\quad\le\sqrt{\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa}\|\nabla^2 L(\theta)\|_F^4\Big)}\sqrt{C_2\frac{k_0J^2}{n}}.
\end{aligned} \tag{56}$$

Using Lemma A.1 and (40), if $\ell\neq\ell'$, we obtain

$$\begin{aligned}
\Big|\frac{\partial^2}{\partial\theta_{\ell'}\,\partial\theta_\ell}L(\theta)\Big|^2 &\le \frac{64\pi^4}{J^4N}\Big(\sum_{|k|\le k_0}k^2(|\bar c_{k,\ell}^{(0)}\overline{z_{k,\ell'}^{(0)}}|+|\bar c_{k,\ell'}^{(0)}\overline{z_{k,\ell}^{(0)}}|)\Big)^2\\
&\le \frac{64\pi^4}{J^4N}\Big(\sum_{|k|\le k_0}k^2(|c_k|+A_0N^{-s+1/2})(|z_{k,\ell}^{(0)}|+|z_{k,\ell'}^{(0)}|)\Big)^2
\end{aligned}$$

and, by (43), if $\ell=\ell'$, we get

$$\begin{aligned}
&\Big|\frac{\partial^2}{\partial\theta_\ell\,\partial\theta_\ell}L(\theta)\Big|^2\\
&\quad\le\frac{64\pi^4}{J^4N}\Bigg(\sum_{|k|\le k_0}k^2\Bigg(|\bar c_{k,\ell}^{(0)}|\Big(\sum_{j=1,j\neq\ell}^{J}|z_{k,j}^{(0)}|\Big)+|z_{k,\ell}^{(0)}|\Big(\sum_{j=1,j\neq\ell}^{J}|\bar c_{k,j}^{(0)}|\Big)\Bigg)\Bigg)^2
\end{aligned}$$

$$\leq \frac{64\pi^4}{J^4 N}\left(\sum_{|k|\leq k_0} k^2\left((|c_k| + A_0 N^{-s+1/2})\left(J|z_{k,\ell}^{(0)}| + \sum_{j=1, j\neq \ell}^{J} |z_{k,j}^{(0)}|\right)\right)\right)^2.$$

Hence, we bound the expectation in (56) from above,

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa} \|\nabla^2 L(\theta)\|_F^4\Big) \\
&\quad\leq \frac{4096\pi^8}{J^4N^2}\mathbb{E}\Bigg(\Bigg[J\Bigg(\sum_{|k|\leq k_0} k^2(|c_k| + A_0N^{-s+1/2})\sum_{j=1}^{J}|z_{k,j}^{(0)}|\Bigg)^2 \\
&\qquad\qquad + \frac{4}{J}\sum_{\ell=1}^{J}\Bigg(\sum_{|k|\leq k_0} k^2(|c_k| + A_0N^{-s+1/2})|z_{k,\ell}^{(0)}|\Bigg)^2\Bigg]^2\Bigg) \\
&\quad\leq \frac{4\times 4096\pi^8}{J^2N^2}\mathbb{E}\Bigg(\Bigg(\sum_{|k|\leq k_0} k^2(|c_k| + A_0N^{-s+1/2})\sum_{j=1}^{J}|z_{k,j}^{(0)}|\Bigg)^4\Bigg) \\
&\quad\leq \frac{4\times 4096\pi^8}{N^2}\Bigg(\sum_{|k|\leq k_0} k^4(|c_k| + A_0N^{-s+1/2})^2\Bigg)^2\mathbb{E}\Bigg(\Bigg(\sum_{|k|\leq k_0}\sum_{j=1}^{J}|z_{k,j}^{(0)}|^2\Bigg)^2\Bigg) \\
&\quad\leq \frac{168\times 4096\pi^8\sigma^4J^2k_0^2}{N^2}\Bigg(\sum_{|k|\leq k_0} k^4(|c_k| + A_0N^{-s+1/2})^2\Bigg)^2.
\end{aligned}$$

Therefore, there exists some constant $C_L' > 0$ that only depends on A, s and σ^2 such that

$$\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa} \|\nabla^2 L(\theta)\|_F^4\Big) \leq C_L'\frac{k_0^2J^2}{n^2}\left(1 + \frac{k_0^5}{n^{2s-1}}\right)^2. \tag{57}$$

Using (56) and (57), we know that there exists some constant $C_L > 0$ that only depends on c_*, κ, A, s and σ^2 such that

$$nJ\mathbb{E}\Big(\sup_{\theta\in\mathcal{U}_\kappa} \|\nabla^2 L(\theta)\|_F^2\|\hat{\theta}^0 - \theta^0\|^2\Big) \leq C_L\frac{k_0^{3/2}J^3}{\sqrt{n}}\left(1 + \frac{k_0^5}{n^{2s-1}}\right). \tag{58}$$

Finally, we use (52), (53), (55) and (58) with (51) to get (34).

A.2. Proof of Theorem 3.2. Let us assume that $f \in \tilde{W}_s(A, c_*)$. We bound the distance between f and $\hat{f}_{n,J}$ from above,

$$\begin{aligned}
d^2([f],[\hat{f}_{n,J}]) &= \inf_{\theta\in[-1/2,1/2]}\int_0^1 |f(t-\theta) - \hat{f}_{n,J}(t)|^2\,dt \\
&\leq 2d^2([f],[\bar{f}^{(m_1)}]) + 2\int_0^1 |\bar{f}^{(m_1)}(t) - \hat{f}_{n,J}(t)|^2\,dt.
\end{aligned}$$

Taking the expectation according to the distribution of $Y^{(1)}$ on both sides and using (12) leads to

$$\begin{aligned}
&\mathbb{E}^{(1)}[d^2([f],[\hat{f}_{n,J}])] \\
&\quad \le 2d^2([f],[\bar{f}^{(m_1)}]) \\
&\quad\quad + 2C(\eta)\min_{m\in\{1,\dots,m_1\}}\left\{\int_0^1 |\bar{f}^{(m_1)}(t)-\bar{f}^{(m)}(t)|^2\,dt + \frac{2(m+1)\sigma^2}{NJ}\right\} \qquad (59)\\
&\quad \le 2d^2([f],[\bar{f}^{(m_1)}]) \\
&\quad\quad + 2C(\eta)\min_{m\in\{1,\dots,m_1\}}\left\{\sum_{m<|k|\le m_1}\left|\frac{1}{J}\sum_{j=1}^J \bar{c}_{k,j}^{(1)} e^{i2\pi k\hat{\theta}_j^0}\right|^2 + \frac{2(m+1)\sigma^2}{NJ}\right\}.
\end{aligned}$$

Let $\bar{\theta}_J = (\theta_1^* + \cdots + \theta_J^*)/J$, we recall that $\theta^0 = \theta^* - \bar{\theta}_J$. We begin by upper bounding the first term. Thanks to Jensen's inequality, we obtain

$$\begin{aligned}
d^2([f],[\bar{f}^{(m_1)}]) &\le \int_0^1 |f(t-\bar{\theta}_J) - \bar{f}^{(m_1)}(t)|^2\,dt \\
&\le \sum_{|k|>m_1}|c_k|^2 + \sum_{|k|\le m_1}\left|c_k e^{-i2\pi k\bar{\theta}_J} - \frac{1}{J}\sum_{j=1}^J \bar{c}_{k,j}^{(1)} e^{i2\pi k\hat{\theta}_j^0}\right|^2 \qquad (60)\\
&\le \sum_{|k|>m_1}|c_k|^2 + \sum_{|k|\le m_1}\frac{1}{J}\sum_{j=1}^J |c_k e^{-i2\pi k\bar{\theta}_J} - \bar{c}_{k,j}^{(1)} e^{i2\pi k\hat{\theta}_j^0}|^2.
\end{aligned}$$

Since $f \in \tilde{W}_s(A, c_*)$, we easily upper bound the bias part

$$\sum_{|k|>m_1}|c_k|^2 \le A|m_1|^{-2s}. \qquad (61)$$

To deal with the other part, we split it into two sums,

$$\begin{aligned}
&\frac{1}{J}\sum_{j=1}^J |c_k e^{-i2\pi k\bar{\theta}_J} - \bar{c}_{k,j}^{(1)} e^{i2\pi k\hat{\theta}_j^0}|^2 \\
&\quad \le \frac{2}{J}\sum_{j=1}^J |c_k e^{-i2\pi k\bar{\theta}_J} - c_k e^{i2\pi k(\hat{\theta}_j^0-\theta_j^*)}|^2 \\
&\quad\quad + \frac{2}{J}\sum_{j=1}^J |c_k e^{i2\pi k(\hat{\theta}_j^0-\theta_j^*)} - \bar{c}_{k,j}^{(1)} e^{i2\pi k\hat{\theta}_j^0}|^2 \qquad (62)\\
&\quad \le \frac{2|c_k|^2}{J}\sum_{j=1}^J |1 - e^{i2\pi k(\hat{\theta}_j^0-\theta_j^0)}|^2 + \frac{2}{J}\sum_{j=1}^J |\bar{c}_{k,j}^{(1)} - c_k e^{-i2\pi k\theta_j^*}|^2
\end{aligned}$$

$$\leq \frac{8\pi^2 k^2 |c_k|^2}{J} \sum_{j=1}^{J} (\hat{\theta}_j^0 - \theta_j^0)^2 + 2A_0^2 N^{-2s+1},$$

where the last inequality follows from Lemma A.1 and from $2(1-\cos t) \leq t^2$, $t \in \mathbb{R}$. Combining (60), (61) and (62), we get, for any $g \in \mathcal{G}^\kappa$,

$$\begin{aligned} \mathbb{E}^g[d^2([f],[\bar{f}^{(m_1)}])] &\leq A|m_1|^{-2s} + 2A_0^2(2m_1+1)N^{-2s+1} \\ &\quad + 8\pi^2 \Bigg(\sum_{|k|\leq m_1} k^2 |c_k|^2 \Bigg) \mathbb{E}^g \Bigg[\frac{1}{J} \sum_{j=1}^{J} (\hat{\theta}_j^0 - \theta_j^0)^2 \Bigg]. \end{aligned} \tag{63}$$

We now focus on the second term in (59). Let $\alpha_{k,j} = c_k e^{-i2\pi k \theta_j^*} - \bar{c}_{k,j}^{(1)}$, using Jensen's inequality and Lemma A.1, for any $m \in \{1, \dots, m_1\}$, we have

$$\begin{aligned} \sum_{m<|k|\leq m_1} \left| \frac{1}{J} \sum_{j=1}^{J} \bar{c}_{k,j}^{(1)} e^{i2\pi k \hat{\theta}_j^0} \right|^2 &\leq \sum_{m<|k|\leq m_1} \frac{1}{J} \sum_{j=1}^{J} |c_k e^{-i2\pi k \theta_j^*} + \alpha_{k,j}|^2 \\ &\leq 2 \sum_{|k|>m} |c_k|^2 + \frac{2}{J} \sum_{m<|k|\leq m_1} \sum_{j=1}^{J} |\alpha_{k,j}|^2 \\ &\leq 2Am^{-2s} + 4A_0^2 m_1 N^{-2s+1}. \end{aligned}$$

Let us consider m_* such that

$$m_* = \left\lfloor \left(\frac{nJ}{c} \right)^{1/(2s+1)} \right\rfloor,$$

where c is the constant such that $J \leq cn^\alpha$. Note that such a choice is allowed because it is such that $m_* \in \{1, \dots, m_1\}$ since $n \geq 21$, $s > 3/2$, $\alpha \in (0, 1/6]$ and $c \in (0,1)$. In particular, such a choice leads to the following upper bound:

$$\begin{aligned} &\min_{m \in \{1,\dots,m_1\}} \Bigg\{ \sum_{m<|k|\leq m_1} \left| \frac{1}{J} \sum_{j=1}^{J} \bar{c}_{k,j}^{(1)} e^{i2\pi k \hat{\theta}_j^0} \right|^2 + \frac{2(m+1)\sigma^2}{NJ} \Bigg\} \\ &\quad \leq 4A_0^2 m_1 N^{-2s+1} + \frac{2\sigma^2}{NJ} + 2 \min_{m \in \{1,\dots,m_1\}} \Bigg\{ Am^{-2s} + \frac{m\sigma^2}{NJ} \Bigg\} \\ &\quad \leq 4A_0^2 m_1 N^{-2s+1} + \frac{2\sigma^2}{NJ} \\ &\qquad + 2\Bigg(\frac{A}{2} c^{2s/(2s+1)} + 2\sigma^2 c^{2s/(2s+1)} \Bigg) (nJ)^{-2s/(2s+1)} \\ &\quad \leq 4A_0^2 m_1 N^{-2s+1} + (1 + (A + 4\sigma^2) c^{2s/(2s+1)}) (nJ)^{-2s/(2s+1)}. \end{aligned} \tag{64}$$

Putting (63) and (64) in (59) leads to, for any $g \in \mathcal{G}^\kappa$,

$$\begin{aligned}\mathcal{R}_g(\hat{f}_{n,J}, f) &\le A|m_1|^{-2s} + 4A_0^2((1+C(\eta))m_1+1)N^{-2s+1} \\ &\quad + 8\pi^2\Bigg(\sum_{|k|\le m_1} k^2|c_k|^2\Bigg)\mathbb{E}^g\left[\frac{1}{J}\sum_{j=1}^J(\hat{\theta}_j^0-\theta_j^0)^2\right] \\ &\quad + C(\eta)(1+(A+4\sigma^2)c^{2s/(2s+1)})(nJ)^{-2s/(2s+1)},\end{aligned}$$

that completes the proof using the fact that $m_1 = \lfloor N/2 \rfloor - 1$.

A.3. Proof of Theorem 4.1. The arguments that we use to derive this result are based on Assouad's cube lemma; see, for example, [30]. This lemma is classically used in nonparametric statistics to derive lower bounds on a risk. We will show that one can construct a set of functions $\mathcal{F}_0 \subset \tilde{W}_s(A, c_*)$ such that there exists a constant $C > 0$ (only depending on A, s, c_* and σ^2) such that, for any large enough n and J,

$$\mathcal{R}_{n,J}(\tilde{W}_s(A,c_*), \mathcal{G}^\kappa) \ge \inf_{\hat{f}_{n,J}} \sup_{g\in\mathcal{G}^\kappa} \sup_{f\in\mathcal{F}_0} \mathcal{R}_g(\hat{f}_{n,J}, f) \ge C(nJ)^{-2s/(2s+1)}, \tag{65}$$

where $\hat{f}_{n,J}$ denote some estimator of f. For the sake of legibility, we assume in the sequel that $c_* = 1$. Let

$$\mathcal{F}_0 = \left\{ f_w : t \in [0,1] \mapsto \sqrt{\mu_{n,J}} \sum_{k\in K_{n,J}} w_k\phi_k(t), w_k \in \{-1,1\}, w_{-k} = w_k \right\},$$

where $K_{n,J} = \{k \in \mathbb{Z}, 0 < |k| \le D_{n,J}\}$, $\mu_{n,J}$ is a positive real and $D_{n,J}$ is a positive integer that will be specified below. Let us introduce the notation $\Omega = \{-1,1\}^{D_{n,J}}$ and note that any function $f_w \in \mathcal{F}_0$ is parametrized by a unique element $w \in \Omega$. Under the condition

$$\mu_{n,J} = cD_{n,J}^{-2s-1} \qquad \text{with } c \le A, \tag{66}$$

it can easily be checked that $\mathcal{F}_0 \subset \tilde{W}_s(A, c_*)$. In what follows, $D_{n,J}$ is chosen as the largest integer smaller that $(nJ)^{1/(2s+1)}$. Hereafter, $\mathbb{E}_w^g$ will denote the expectation with respect to the distribution $\mathbb{P}_w^g$ of the random vector $(Y_{\ell,j})_{1\le\ell\le n, 1\le j\le J} \in \mathbb{R}^{nJ}$ in model (3) under the hypothesis that $f = f_w$ and the assumption that the shifts are i.i.d. random variables with density $g \in \mathcal{G}^\kappa$. Note that for any $g \in \mathcal{G}^\kappa$

$$\begin{aligned}\sup_{f\in\mathcal{F}_0} \mathcal{R}_g(\hat{f}_{n,J}, f) &= \sup_{f\in\mathcal{F}_0} \mathbb{E}\left[\inf_{\theta\in[0,1]}\left(\int_0^1 |\hat{f}_{n,J}(t-\theta) - f(t)|^2\,dt\right)\right] \\ &\ge \frac{1}{|\Omega|}\sum_{w\in\Omega} \mathbb{E}_w^g\left[\inf_{\theta\in[0,1]}\left(\int_0^1 |\hat{f}_{n,J}(t-\theta) - f_w(t)|^2\,dt\right)\right]\end{aligned}$$

$$\geq \frac{1}{|\Omega|}\sum_{w\in\Omega}\mathbb{E}_w^g\left[\inf_{\theta\in[0,1]}\sum_{k\in K_{n,J}}|\hat{c}_k e^{-2ik\pi\theta}-\sqrt{\mu_{n,J}}w_k|^2\right],$$

where $\hat{c}_k=\int_0^1\hat{f}_{n,J}(t)\overline{\phi_k(t)}\,dt$ is the kth Fourier coefficient of $\hat{f}_{n,J}$. Now, we consider, for $k\in K_{n,J}$ and $\theta\in[0,1]$,

$$\hat{w}_{k,\theta}\in\operatorname*{argmin}_{v\in\{-1,1\}}|\hat{c}_k e^{-2ik\pi\theta}-\sqrt{\mu_{n,J}}v|.$$

We have the inequality

$$\begin{aligned}|\sqrt{\mu_{n,J}}\hat{w}_{k,\theta}-\sqrt{\mu_{n,J}}w_k| &\leq |\hat{c}_k e^{-2ik\pi\theta}-\sqrt{\mu_{n,J}}\hat{w}_{k,\theta}|\\ &\quad+|\hat{c}_k e^{-2ik\pi\theta}-\sqrt{\mu_{n,J}}w_k|\\ &\leq 2|\hat{c}_k e^{-2ik\pi\theta}-\sqrt{\mu_{n,J}}w_k|\end{aligned}$$

that implies

$$\sup_{f\in\mathcal{F}_0}\mathcal{R}_g(\hat{f}_{n,J},f)\geq\frac{\mu_{n,J}}{4|\Omega|}\sum_{k\in K_{n,J}}\sum_{w\in\Omega}\mathbb{E}_w^g\left[\inf_{\theta\in[0,1]}(|\hat{w}_{k,\theta}-w_k|^2)\right].$$

For $w\in\Omega$ and $k\in K_{n,J}$, we define $w^{(k)}\in\Omega$ such that, for any $\ell\neq k$, $w_\ell^{(k)}=w_\ell$ and $w_k^{(k)}=-w_k$. Then it follows that

$$\sup_{f\in\mathcal{F}_0}\mathcal{R}_g(\hat{f}_{n,J},f)\geq\frac{\mu_{n,J}}{4|\Omega|}\sum_{k\in K_{n,J}}\sum_{w\in\Omega|w_k=1}R_k, \tag{67}$$

where we have set

$$R_k=\mathbb{E}_w^g\left[\inf_{\theta\in[0,1]}(|\hat{w}_{k,\theta}-w_k|^2)\right]+\mathbb{E}_{w^{(k)}}^g\left[\inf_{\theta\in[0,1]}(|\hat{w}_{k,\theta}+w_k|^2)\right].$$

Let $\theta^*=(\theta_1^*,\dots,\theta_J^*)$, we introduce the notation $\mathbb{E}_w^{\theta^*}$ to denote expectation with respect to the distribution $\mathbb{P}_w^{\theta^*}$ of the random vector $(Y_{\ell,j})_{1\leq\ell\leq n,1\leq j\leq J}\in\mathbb{R}^{nJ}$ in model (1) conditionally to $\theta_1^*,\dots,\theta_J^*$. Hence, using this notation, we have

$$R_k=\int_{[-1/2,1/2]^J}R_k(\theta^*)g(\theta_1^*)\cdots g(\theta_J^*)\,d\theta_1^*\cdots d\theta_J^*, \tag{68}$$

where

$$R_k(\theta^*)=\mathbb{E}_w^{\theta^*}\left[\inf_{\theta\in[0,1]}(|\hat{w}_{k,\theta}-w_k|^2)\right]+\mathbb{E}_{w^{(k)}}^{\theta^*}\left[\inf_{\theta\in[0,1]}(|\hat{w}_{k,\theta}+w_k|^2)\right].$$

Now, note that for any $0<\delta<1$,

$$\begin{aligned}R_k(\theta^*)&=\mathbb{E}_w^{\theta^*}\left[\inf_{\theta\in[0,1]}(|\hat{w}_{k,\theta}-w_k|^2)+\inf_{\theta\in[0,1]}(|\hat{w}_{k,\theta}+w_k|^2)\frac{d\mathbb{P}_{w^{(k)}}^{\theta^*}}{d\mathbb{P}_w^{\theta^*}}(Y)\right]\\ &\geq 4\mathbb{E}_w^g\min\left(1,\frac{d\mathbb{P}_{w^{(k)}}^{\theta^*}}{d\mathbb{P}_w^{\theta^*}}(Y)\right)\geq 4\delta\mathbb{P}_w^{\theta^*}\left(\frac{d\mathbb{P}_{w^{(k)}}^{\theta^*}}{d\mathbb{P}_w^{\theta^*}}(Y)\geq\delta\right),\end{aligned} \tag{69}$$

where $Y \in \mathbb{R}^{nJ}$ is the random vector obtained from the concatenation of the observations from model (3) under the hypothesis $f = f_w$ and conditionally to $\theta_1^*, \dots, \theta_J^*$. Because $w_k = 1$, we know that

$$\log \frac{d\mathbb{P}^{\theta^*}_{w^{(k)}}}{d\mathbb{P}^{\theta^*}_{w}}(Y) = -\frac{2}{\sigma^2} \sum_{j=1}^{J} \sum_{\ell=1}^{n} \mu_{n,J} |\phi_k(t_\ell - \theta_j^*)|^2 + \sqrt{\mu_{n,J}} \varepsilon_{\ell,j} \phi_k(t_\ell - \theta_j^*).$$

Therefore, $\log \frac{d\mathbb{P}^{\theta^*}_{w^{(k)}}}{d\mathbb{P}^{\theta^*}_{w}}(Y)$ is a random variable that is normally distributed with mean $-\frac{2}{\sigma^2} nJ\mu_{n,J}$ and variance $\frac{4}{\sigma^2} nJ\mu_{n,J}$. Now, since D_{nJ} is the largest integer smaller than $(nJ)^{1/(2s+1)}$, it follows from equation (66) that, for any n and J large enough,

$$0 \le nJ\mu_{n,J} \le 2A.$$

Thus, there exists $0 < \delta < 1$ and a constant $c_\delta > 0$ (only depending on A, σ^2 and δ) such that

$$\mathbb{P}^{\theta^*}_{w} \left(\frac{d\mathbb{P}^{\theta^*}_{w^{(k)}}}{d\mathbb{P}^{\theta^*}_{w}}(Y) \ge \delta \right) \ge c_\delta.$$

Combining this inequality with (67), (68) and (69) leads to

$$\sup_{f \in \mathcal{F}_0} \mathcal{R}_g(\hat{f}_{n,J}, f) \ge \frac{4\delta \mu_{n,J}}{|\Omega|} \sum_{k \in K_{n,J}} \sum_{w \in \Omega | w_k = 1} c_\delta \ge \delta c_\delta \mu_{n,J} D_{n,J}. \tag{70}$$

Since $\mu_{n,J} = cD_{n,J}^{-2s-1}$ and $D_{n,J} \le (nJ)^{1/(2s+1)}$, it follows that

$$\mu_{n,J} D_{n,J} = cD_{n,J}^{-2s} \ge c(nJ)^{-2s/(2s+1)},$$

which combined with (70) proves inequality (65) and completes the proof of Theorem 4.1.

Acknowledgments. We are very much indebted to the referees and the Associate Editor for their constructive comments and remarks that helped us to improve the presentation of the original manuscript.

REFERENCES

[1] Afsari, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346

[2] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498

[3] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634

[4] Bigot, J. (2006). Landmark-based registration of curves via the continuous wavelet transform. *J. Comput. Graph. Statist.* **15** 542–564. MR2291263

[5] Bigot, J. and Charlier, B. (2011). On the consistency of Fréchet means in deformable models for curve and image analysis. *Electron. J. Stat.* **5** 1054–1089. MR2836769

[6] Bigot, J. and Gadat, S. (2010). A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.* **38** 2422–2464. MR2676894

[7] Bigot, J., Loubes, J.-M. and Vimond, M. (2012). Semiparametric estimation of shifts on compact Lie groups for image registration. *Probab. Theory Related Fields* **152** 425–473. MR2892953

[8] Charlier, B. (2013). Necessary and sufficient condition for the existence of a Fréchet mean on the circle. *ESAIM Probab. Stat.* To appear.

[9] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324

[10] Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10** 215–310. MR0027464

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

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

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

[14] Grenander, U. (1993). *General Pattern Theory: A Mathematical Study of Regular Structures*. The Clarendon Press Oxford Univ. Press, New York. MR1270904

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

[16] Hotz, T. and Huckemann, S. (2011). Intrinsic means on the circle: Uniqueness, locus and asymptotics. Available at arXiv:1108.2141.

[17] Huckemann, S. F. (2011). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.* **39** 1098–1124. MR2816349

[18] Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. Lond. Math. Soc.* **16** 81–121. MR0737237

[19] Kneip, A. and Gasser, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.* **16** 82–112. MR0924858

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

[21] Massart, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Springer, Berlin. MR2319879

[22] McKilliam, R. G., Quinn, B. G. and Clarkson, I. V. L. (2012). Direction estimation by minimum squared arc length. *IEEE Trans. Signal Process.* **60** 2115–2124. MR2954196

[23] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** 2033–2053. MR1765627

[24] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York. MR1910407

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

[26] Tang, R. and Müller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95** 875–889. MR2461217

[27] Trigano, T., Isserles, U. and Ritov, Y. (2011). Semiparametric curve alignment and shift density estimation for biological data. *IEEE Trans. Signal Process.* **59** 1970–1984. MR2816476

[28] Trouvé, A. and Younes, L. (2005). Local geometry of deformable templates. *SIAM J. Math. Anal.* **37** 17–59 (electronic). MR2176922

[29] Trouvé, A. and Younes, L. (2011). Shape spaces. In *Handbook of Mathematical Methods in Imaging*. Springer, Berlin.

[30] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

[31] Vimond, M. (2010). Efficient estimation for a subclass of shape invariant models. *Ann. Statist.* **38** 1885–1912. MR2662362

[32] Wang, K. and Gasser, T. (1997). Alignment of curves by dynamic time warping. *Ann. Statist.* **25** 1251–1276. MR1447750

DMIA–ISAE
10 avenue Édouard Belin
BP 54032
31055 Toulouse Cedex 4
France
E-mail: Jeremie.Bigot@isae.fr

Institut de Mathématiques de Toulouse
IMT—Université Paul Sabatier
118 route de Narbonne
31062 Toulouse Cedex 4
France
E-mail: Xavier.Gendre@math.univ-toulouse.fr