Submitted to the Annals of Statistics
arXiv: math.ST/0812.3502v2

A DECONVOLUTION APPROACH TO ESTIMATION OF A COMMON SHAPE IN A SHIFTED CURVES MODEL

By Jérémie Bigot* and Sébastien Gadat*

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

This paper considers the problem of adaptive estimation of a mean pattern in a randomly shifted curve model. We show that this problem can be transformed into a linear inverse problem, where the density of the random shifts plays the role of a convolution operator. An adaptive estimator of the mean pattern, based on wavelet thresholding is proposed. We study its consistency for the quadratic risk as the number of observed curves tends to infinity, and this estimator is shown to achieve a near-minimax rate of convergence over a large class of Besov balls. This rate depends both on the smoothness of the common shape of the curves and on the decay of the Fourier coefficients of the density of the random shifts. Hence, this paper makes a connection between mean pattern estimation and the statistical analysis of linear inverse problems, which is a new point of view on curve registration and image warping problems. Some numerical experiments are given to illustrate the performances of our approach and to compare them with another algorithm existing in the literature.

1. Introduction.

1.1. *Model and objectives.* In many fields of interests including biology, medical imaging or chemistry, observations are coming from n individuals curves or graylevel images. Such observations are commonly referred to as functional data, and models involving such data have been recently extensively studied in statistics (see [40], [41] for a detailed introduction to functional data analysis). In such settings, it is reasonable to assume that the data at hand $Y_m, m = 1, \ldots, n$, satisfy the following white noise regression model:

$$dY_m(x) = f_m(x)dx + \epsilon_m dW_m(x),\ x \in \Omega,\ m = 1, \ldots, n \tag{1.1}$$

where Ω is a subset of $\mathbb{R}^d$, $f_m : \Omega \to \mathbb{R}$ are unknown regression functions, and W_m are independent standard Brownian motions on Ω with ϵ_m representing

AMS 2000 subject classifications: Primary 62G08; secondary 42C40
Keywords and phrases: Mean pattern estimation, Curve registration, Inverse problem, Deconvolution, Meyer wavelets, Adaptive estimation, Besov space, Minimax rate

different levels of additive noise. In many situations the individual curves or images have a certain common structure which may lead to the assumption that they are generated from some semi-parametric model of the form

$$f_m(x) = f(x, \tau_m), \quad \text{for } x \in \Omega \text{ and some } \tau_m \in \mathcal{T} \subset \mathbb{R}^p, \tag{1.2}$$

where $f : \Omega \times \mathcal{T} \to \mathbb{R}$ represents an unknown shape common to all the f_m's. This shape function (also called mean pattern) may depend on unknown individual *random* parameters $\tau_m, m = 1, \ldots, n$, belonging to a compact set $\mathcal{T}$ of $\mathbb{R}^p$, which model individual variations. Such a semi-parametric representation for the f_m's is the so-called *self-modeling regression framework* (SEMOR) introduced by [26]. *Shape invariant models* (SIM) are a special class of such models for which (see e.g. [26])

$$f_m(x) = f(\phi(x, \tau_m)), \tag{1.3}$$

where for any $\tau \in \mathcal{T}$, the function $x \mapsto \phi(x, \tau)$ is a smooth diffeomorphism of Ω and $\phi : \Omega \times \mathcal{T} \to \Omega$ is a *known* function. Models such as (1.3) are useful to account for shape variability in time between curves (see e.g [18], [29]) or in space between images, which is the well-known problem of curve or image warping (see [17] and the discussion therein for an overview, [5], [6] and references therein). SIM models (1.3) also represent a large class of statistical models to study the difficult problem of recovering a mean pattern from a set of similar curves or images in the presence of random deformations and additive noise, which corresponds to the general setting of Grenander's theory of shapes [20]. The overall objective of this paper is to discuss the fundamental problem of estimating of the mean pattern f which can then be used to learn non-linear modes of variations in time or shape between similar curves or images.

1.2. *Previous work on mean pattern estimation.* Very few results exist in the literature on nonparametric estimation of f for SIM models (1.3) based on noisy data from (1.1). The problem of estimating the common shape of a set of curves that differ only by a time transformation is usually referred to as the curve registration problem in statistics, and it has received a lot of attention over the last two decades, see e.g [4], [15], [16], [29], [39], [45]. However, in these papers, an asymptotic study as the number of curves n grows to infinity is generally not considered. Estimation of the shape function for SEMOR models related to (1.1) and (1.2) is studied in [26] with a double asymptotic in the number n of curves and the number of observed time points per curve. In the simplest case of shifted curves, various approaches have been developed. Based on a model with a fixed number n of curves, semiparametric estimation of deformation parameters and nonparametric estimation of the shape function is proposed in [31] and [44]. A generalization of this approach for the estimation of scaling, rotation and translation parameters for two-dimensional images is proposed in [6]. Estimation of a common shape for randomly shifted curves and asymptotic in n is also considered in [42]. There is also a huge literature in image analysis on mean pattern estimation, and some papers have recently addressed the problem of estimating the common shape of a set of similar images with asymptotic in the number of images, see e.g. [1], [5], [32] and references therein. However, in all the above cited papers rates of convergence and optimality of the proposed estimators for f have not been studied.

1.3. *A benchmark model for nonparametric estimation of a mean pattern.* The simplest SIM model is the case of randomly shifted curves, namely

$$f_m(x) = f(x - \tau_m), \qquad \text{for } x \in [0,1] \text{ and } \tau_m \in \mathbb{R},$$

that has recently received some attention in the statistical literature [30], [31], [42], [44]. In this paper it will thus be assumed that we observe realizations of n noisy and randomly shifted curves $Y_1, \ldots, Y_n$ coming from the following Gaussian white noise model

$$dY_m(x) = f(x - \tau_m)dx + \epsilon_m dW_m(x),\ x \in [0,1],\ m = 1, \ldots, n, \tag{1.4}$$

where f is the unknown mean pattern of the curves, W_m are independent standard Brownian motions on $[0,1]$, the ϵ_m's represent levels of noise which may vary from curve to curve, and the τ_m's are unknown random shifts independent of the W_m's. The aim of this paper is to study some statistical aspects related to the problem of estimating f, and to propose new methods of estimation.

Model (1.4) is realistic in many situations where it is reasonable to assume that the observed curves represent replications of almost the same process and when a large source of variation in the experiments is due to transformations of the time axis. Such a model is commonly used in many applied areas dealing with functional data such as neuroscience [24] or biology [42]. More generally, the model (1.4) represents a kind of benchmark model for studying the problem of recovering the mean pattern f in SIM models. The results derived in this paper show that the model (1.4), although simple, already provides some new insights on the statistical aspects of mean pattern estimation.

The function $f : \mathbb{R} \to \mathbb{R}$ is assumed to be periodic with period 1, and the shifts τ_m are supposed to be independent and identically distributed (iid) random variables with density $g : \mathbb{R} \to \mathbb{R}^+$ with respect to the Lebesgue measure dx on $\mathbb{R}$. Our goal is to estimate nonparametrically the shape function f on $[0,1]$ as the number of curves n goes to infinity.

Let $L^2([0,1])$ be the space of squared integrable functions on $[0,1]$ with respect to dx, and denote by $\|f\|^2 = \int_0^1 |f(x)|^2 dx$ the squared norm of a function f. Assume that $\mathcal{F} \subset L^2([0,1])$ represents some smoothness class of functions (e.g a Sobolev or a Besov ball), and let $\hat{f}_n \in L^2([0,1])$ be some estimator of the common shape f, i.e a measurable function of the random processes $Y_m,\ m = 1,\dots,n$. For some $f \in \mathcal{F}$, the risk of the estimator $\hat{f}_n$ is defined to be

$$\mathcal{R}(\hat{f}_n, f) = \mathbb{E}\|\hat{f}_n - f\|^2$$

where the above expectation $\mathbb{E}$ is taken with respect to the law of $\{Y_m, m = 1,\dots,n\}$. In this paper, we propose to investigate the optimality of an estimator by introducing the following minimax risk

$$\mathcal{R}_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f\in\mathcal{F}} \mathcal{R}(\hat{f}_n, f),$$

where the above infimum is taken over the set of all possible estimators in model (1.4). One of the main contributions of this paper is to derive asymptotic lower and upper bounds for $\mathcal{R}_n(\mathcal{F})$ which, to the best of our knowledge, has not been considered before.

Indeed, we show that there exists constants M_1, M_2, a sequence of reals $r_n = r_n(\mathcal{F})$ tending to infinity, and an estimator $\hat{f}_n^*$ such that

$$\lim_{n\to+\infty} r_n \mathcal{R}_n(\mathcal{F}) \geq M_1 \text{ and } \lim_{n\to+\infty} r_n \sup_{f\in\mathcal{F}} \mathcal{R}(\hat{f}_n^*, f) \leq M_2.$$

However, the construction of $\hat{f}_n^*$ may depend on unknown quantities such as the smoothness of f, and such estimates are therefore called non-adaptive. Since it is now recognized that wavelet decomposition is a powerful tool to derive adaptive estimators, see e.g [12], a second contribution of this paper is thus to propose wavelet-based estimators $\hat{f}_n$ that attain a near-minimax rate of convergence in the sense there exits a constant M_2 such that

$$\lim_{n\to+\infty} (\log n)^{-\beta} r_n \sup_{f\in\mathcal{F}} \mathcal{R}(\hat{f}_n, f) \leq M_2, \text{ for some } \beta > 0.$$

1.4. *Main result.* Minimax risks will be derived under particular smoothness assumptions on the density g. The main result of this paper is that the difficulty of estimating f is quantified by the decay to zero of the Fourier coefficients γ_ℓ of the density g of the shifts defined as

$$\gamma_\ell = \mathbb{E}\left(e^{-i2\pi\ell\tau}\right) = \int_{-\infty}^{+\infty} e^{-i2\pi\ell x} g(x)dx, \tag{1.5}$$

for $\ell \in \mathbb{Z}$. Depending how fast these Fourier coefficients tend to zero as $|\ell| \to +\infty$, the reconstruction of f will be more or less accurate. This comes from the fact that the expected value of each observed process $Y_m(x)$ is given by

$$\mathbb{E}Y_m(x) = \mathbb{E}f(x - \tau_m) = \int_{-\infty}^{+\infty} f(x - \tau)g(\tau)d\tau, \text{ for } x \in [0, 1].$$

This expected value is thus the convolution of f by the density g which makes the problem of estimating f an inverse problem whose degree of ill-posedness and associated minimax risk depend on the smoothness assumptions on g.

This phenomenon is a well-known fact in deconvolution problems, see e.g [25] [36], [37], and more generally for linear inverse problems as studied in [8]. In this paper, the following type of assumption on g is considered:

Assumption 1. *The Fourier coefficients of g have a polynomial decay i.e. for some real $\nu > 0$, there exist two constants $C_{max} \geq C_{min} > 0$ such that $C_{min}|\ell|^{-\nu} \leq |\gamma_\ell| \leq C_{max}|\ell|^{-\nu}$ for all $\ell \in \mathbb{Z}$.*

In standard inverse problems such as deconvolution, the optimal rate of convergence we can expect from an arbitrary estimator typically depends on such smoothness assumptions. The parameter ν is usually referred to as the degree of ill-posedness of the inverse problem, and it quantifies the difficult of inverting the convolution operator. The following theorem shows that a similar phenomenon holds for the minimax risk associated to model (1.4). Note that to simplify the presentation, all the theoretical results are given for the simple setting where the level of noise is the same for all curves i.e. $\epsilon_m = \epsilon$ for all $m = 1, \ldots, n$ and some $\epsilon > 0$.

Theorem 1. *Suppose that the smoothness class $\mathcal{F}$ is a Besov ball $B^s_{p,q}(A)$ of radius $A > 0$ with $p, q \geq 1$ and smoothness parameter $s > 0$ (a precise definition of Besov spaces will be given later on). Suppose that g satisfies Assumption 1, and is such that there exist two constants $C > 0$ and $\alpha > 1$ satisfying $g(x) \leq \frac{C}{1+|x|^\alpha}$ for all $x \in \mathbb{R}$. Let $p' = \min(2, p)$ and assume that $s \geq 1/p'$. If $s > 2\nu + 1$, then*

$$r_n(\mathcal{F}) = n^{\frac{2s}{2s+2\nu+1}}.$$

Hence, Theorem 1 shows that under Assumption 1 the minimax rate r_n is of polynomial order of the sample size n, and that this rate deteriorates as the degree of ill-posedness ν increases. Such a behavior is well known for standard periodic deconvolution in the white noise model [25], [36], and Theorem 1 shows that a similar phenomenon holds for the model (1.4). To the best of our knowledge, this is a new result which makes a connection between mean pattern estimation and the statistical analysis of deconvolution problems.

1.5. *Fourier Analysis and an inverse problem formulation.* Let us first remark that the model (1.4) exhibit some similarities with periodic deconvolution in the white noise model as described in [25]. For $x \in [0,1]$, let us define the following density function

$$G(x) = \sum_{k \in \mathbb{Z}} g(x+k). \tag{1.6}$$

Note that $G(x)$ exists for all $x \in [0,1]$ provided g has a sufficiently fast decay at infinity. In particular, the condition that $g(x) \leq \frac{C}{1+|x|^\alpha}$ for all $x \in \mathbb{R}$ and some $\alpha > 1$ (see Assumption 2 below) is sufficient to guarantee the existence of G. Since f is periodic with period 1, one has

$$\int_{-\infty}^{+\infty} f(x-\tau)g(\tau)d\tau = \int_0^1 f(x-\tau)G(\tau)d\tau,$$

and note that $\gamma_\ell = \int_{-\infty}^{+\infty} e^{-i2\pi\ell x} g(x)dx = \int_0^1 e^{-i2\pi\ell x} G(x)dx$. Hence, if one defines $\xi_m(x) = f(x-\tau_m) - \int_0^1 f(x-\tau)G(\tau)d\tau$ and $\xi(x) = \frac{1}{n}\sum_{m=1}^n \xi_m(x)$, then taking the mean of the n equations in (1.4) yields the model

$$dY(x) = \int_0^1 f(x-\tau)G(\tau)d\tau dx + \xi(x)dx + \frac{\epsilon}{\sqrt{n}} dW(x),\ x \in [0,1], \tag{1.7}$$

with $\epsilon^2 = \frac{1}{n}\sum_{m=1}^n \epsilon_m^2$ and where $W(x)$ is a standard Brownian motion $[0,1]$.

The model (1.7) differs from the periodic deconvolution model investigated in [25] by the error term $\xi(x)$. Asymptotically $\xi(x)$ is a Gaussian variable, so this suggests to use the wavelet thresholding procedures developed in [25] to derive upper bounds for the minimax risk. However, it should be noted that the additive error term $\xi(x)$ significantly complicates the estimation procedure as *the variance of* $\xi(x)$ *clearly depends on the unknown function* f. Moreover, deriving lower bounds for the minimax risk in models such as (1.7) is significantly more difficult than in the standard white noise model without the additive term $\xi(x)$.

Now let us formulate models (1.4) and (1.7) in the Fourier domain. Supposing that $f \in L^2([0,1])$, we denote by θ_ℓ its Fourier coefficients for $\ell \in \mathbb{Z}$, namely $\theta_\ell = \int_0^1 e^{-2i\ell\pi x} f(x)dx$. The model (1.4) can then be rewritten as

$$\begin{aligned} c_{m,\ell} := \int_0^1 e^{-2i\ell\pi x} dY_m(x) &= \theta_\ell e^{-i2\pi\ell\tau_m} + \epsilon_m z_{\ell,m} \\ &= \theta_\ell \gamma_\ell + \xi_{\ell,m} + \epsilon_m z_{\ell,m}, \end{aligned} \tag{1.8}$$

with $\xi_{\ell,m} = \theta_\ell e^{-i2\pi\ell\tau_m} - \theta_\ell\gamma_\ell$, and where $z_{\ell,m}$ are iid $N_{\mathbb{C}}(0,1)$ variables, i.e. complex Gaussian variables with zero mean and such that $\mathbb{E}|z_{\ell,n}|^2 = 1$.

Thus, we can compute the sample mean $\tilde{c}_\ell$ of the ℓ^{th} Fourier coefficient over the n curves as

$$\tilde{c}_\ell = \frac{1}{n}\sum_{m=1}^{n} c_{\ell,m} = \theta_\ell\tilde{\gamma}_\ell + \epsilon\eta_\ell = \theta_\ell\gamma_\ell + \xi_\ell + \epsilon\eta_\ell, \tag{1.9}$$

with $\tilde{\gamma}_\ell = \frac{1}{n}\sum_{m=1}^{n} e^{-i2\pi\ell\tau_m}$, $\xi_\ell = \frac{1}{n}\sum_{m=1}^{n}\xi_{\ell,m}$, $\epsilon^2 = \frac{1}{n}\sum_{m=1}^{n}\epsilon_m^2$, and where the η_ℓ's are iid complex Gaussian variables with zero mean and such that $\mathbb{E}|\eta_\ell|^2 = \frac{1}{n}$. The average Fourier coefficients $\tilde{c}_\ell$ in equation (1.9) can thus be viewed as a set of observations which is very close to a sequence space formulation of a statistical inverse problem as described e.g by [8]. As in model (1.7) the additive error term ξ_ℓ is asymptotically Gaussian, however its variance is $\frac{1}{n}|\theta_\ell|^2(1-|\gamma_\ell|^2)$ which is obviously unknown as it depends on f.

If we assume that the density g of the random shifts is known, one can perform a deconvolution step by taking

$$\hat{\theta}_\ell = \frac{\tilde{c}_\ell}{\gamma_\ell} = \theta_\ell\frac{\tilde{\gamma}_\ell}{\gamma_\ell} + \epsilon\frac{\eta_\ell}{\gamma_\ell}. \tag{1.10}$$

to estimate the Fourier coefficients of f since, for large n, $\frac{\tilde{\gamma}_\ell}{\gamma_\ell}$is close to 1 by the strong law of large numbers.

Based on the $\hat{\theta}_\ell$'s, two types of estimators are studied. The simplest one uses spectral cut-off with a cutting frequency depending on the smoothness assumptions on f, and is thus non-adaptive. The second estimator is based on wavelet thresholding and is shown to be adaptive using the procedure developed in [25]. Note that part of our results are presented for the case where the coefficients γ_ℓ are known. Such a framework is commonly used in nonparametric regression and inverse problems to obtain consistency results and to study asymptotic rates of convergence, where it is generally supposed that the law of the additive error is Gaussian with zero mean and *known* variance ϵ^2, see e.g [25] [36], [8]. In model (1.4), the random shifts may be viewed as a second source of noise and for the theoretical analysis of this problem the law of this other random noise is also supposed to be known.

1.6. *An inverse problem with unknown operator.* If the density g is unknown one can view the problem of estimating f in model (1.4) as a deconvolution problem with unknown eigenvalues which complicates significantly the estimation procedure. Such a framework corresponds to the general setting of an inverse problem with a partially unknown operator. Recently, some papers have addressed this problem, see e.g. [9], [13], [22], assuming that an independent sample of noisy eigenvalues is available which allows an estimation of the γ_ℓ's. However such an assumption is not applicable to our model (1.4). Therefore, we introduce a new method for estimating f is the case of an unknown density g which leads to a new class of estimators to recover a mean pattern.

1.7. *Organization of the paper.* In Section 2, we consider a linear but non-adaptive estimator based on spectral cut-off. In Section 3, a nonlinear and adaptive estimator based on wavelet thresholding is studied in the case of known density g, and upper bound for the minimax risk are studied over Besov balls. In Section 4, we derive lower bounds for the minimax risk. In Section 5, it is explained how one can estimate the mean pattern f when the density g is unknown. Finally, in Section 6, some numerical examples are proposed to illustrate the performances of our approach and to compare them with another algorithm proposed in the literature. All proofs are deferred to a technical Appendix at the end of the paper.

2. Linear estimation of the common shape and upper bounds for the risk for Sobolev balls.

2.1. *Risk decomposition.* For $\ell \in \mathbb{Z}$, a linear estimator of the θ_ℓ's is given by $\hat{\theta}_\ell^\lambda = \lambda_\ell \frac{\tilde{c}_\ell}{\gamma_\ell}$, where $\lambda = (\lambda_\ell)_{\ell \in \mathbb{Z}}$ is a sequence of nonrandom weights called a filter. An estimator $\hat{f}_{n,\lambda}$ of f is then obtained via the inverse Fourier transform $\hat{f}_{n,\lambda}(x) = \sum_{\ell \in \mathbb{Z}} \hat{\theta}_\ell^\lambda e^{-i2\pi\ell x}$, and thanks to the Parseval's relation, the risk of this estimator is given by $\mathcal{R}(\hat{f}_{n,\lambda}, f) = \mathbb{E} \sum_{\ell \in \mathbb{Z}} |\hat{\theta}_\ell - \theta_\ell|^2$. The problem is then to choose the sequence $(\lambda_\ell)_{\ell \in \mathbb{Z}}$ in an optimal way. The following proposition gives the bias-variance decomposition of $\mathcal{R}(\hat{f}_{n,\lambda}, f)$.

PROPOSITION 1. *For any given nonrandom filter λ, the risk of the estimator $\hat{f}_{n,\lambda}$ can be decomposed as*

$$\mathcal{R}(\hat{f}_{n,\lambda}, f) = \underbrace{\sum_{\ell \in \mathbb{Z}} (\lambda_\ell - 1)^2 |\theta_\ell|^2}_{Bias} + \underbrace{\frac{1}{n} \sum_{\ell \in \mathbb{Z}} \lambda_\ell^2 \left[|\theta_\ell|^2 \left(\frac{1}{|\gamma_\ell|^2} - 1 \right) + \frac{\epsilon^2}{|\gamma_\ell|^2} \right]}_{Variance}. \tag{2.1}$$

Note that the decomposition (2.1) does not correspond exactly to the classical bias-variance decomposition for linear inverse problems. Indeed, the variance term in (2.1) differs from the classical expression of the variance for linear estimator in statistical inverse problems which would be in our notations $\epsilon^2 \sum_{\ell \in \mathbb{Z}} \frac{\lambda_\ell^2}{|\gamma_\ell|^2}$. Hence, contrary to classical inverse problems, the variance term of the risk depends also on the Fourier coefficients θ_ℓ of the unknown function f to recover.

2.2. *Linear estimation.* Let us introduce the following smoothness class of functions which can be identified with a periodic Sobolev ball

$$H_s(A) = \left\{ f \in L^2([0,1]) \; ; \sum_{\ell \in \mathbb{Z}} (1 + |\ell|^{2s}) |\theta_\ell|^2 \leq A \right\}$$

for some constant A and some smoothness parameter $s > 0$, where $\theta_\ell = \int_0^1 e^{-2i\ell\pi x} f(x) dx$. Now consider a linear estimator obtained by spectral cut-off, i.e. for a projection filter of the form $\lambda_\ell^M = 1\!\!1_{|\ell| \leq M}$ for some integer M. For an appropriate choice of M, the following proposition gives the asymptotic behavior of the risk $\mathcal{R}(\hat{f}_{n,\lambda^M}, f)$.

PROPOSITION 2. *Assume that f belongs to $H_s(A)$ for some real $s > 1/2$ and $A > 0$, and that g satisfies (1) i.e. polynomial decay of the γ_ℓ's. Then, if $M = M_n$ is chosen such that $M_n \sim n^{\frac{1}{2s+2\nu+1}}$, then there exists a constant C not depending on n such that as $n \to +\infty$*

$$\sup_{f \in H_s(A)} \mathcal{R}(\hat{f}_{n,\lambda^M}, f) \leq C n^{-\frac{2s}{2s+2\nu+1}}.$$

The above choice for M_n depends on the smoothness s of the function f which is generally unknown in practice and such a spectral cut-off estimator is thus called non-adaptive. Moreover, the result is only suited for smooth functions since Sobolev balls $H_s(A)$ for $s > 1/2$ are not suited to model shape functions f which may have singularities such as points of discontinuity.

3. Nonlinear estimation with Meyer wavelets and upper bounds for the risk for Besov balls. Wavelets have been successfully used for various inverse problems [11], and for the specific case of deconvolution Meyer wavelets, a special class of band-limited functions introduced by [35], have recently received special attention in nonparametric regression, see [25] and [36].

3.1. *Wavelet decomposition and the periodized Meyer wavelet basis.* This wavelet basis is derived through the periodization of the Meyer wavelet basis of $L^2(\mathbb{R})$ (see [25] for further details on its construction). Denote by $\phi_{j,k}$ and $\psi_{j,k}$ the Meyer scaling and wavelet functions at scale $j \geq 0$ and location $0 \leq k \leq 2^j - 1$. For any function f of $L^2([0,1])$, its wavelet decomposition can be written as: $f = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{+\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}$, where $c_{j_0,k} = \int_0^1 f(x)\phi_{j_0,k}(x)dx$, $\beta_{j,k} = \int_0^1 f(x), \psi_{j,k}(x)dx$ and $j_0 \geq 0$ denotes the usual coarse level of resolution. Moreover, the squared norm of f is given by $\|f\|^2 = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}^2 + \sum_{j=j_0}^{+\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2$. It is well known that Besov spaces can be characterized in terms of wavelet coefficients (see e.g [25]). Let $s > 0$ denote the usual smoothness parameter, then for the Meyer wavelet basis and for a Besov ball $B_{p,q}^s(A)$ of radius $A > 0$ with $1 \leq p, q \leq \infty$, one has that $B_{p,q}^s(A) = \left\{ f \in L^2([0,1]) : \left(\sum_{k=0}^{2^{j_0}-1} |c_{j_0,k}|^p \right)^{\frac{1}{p}} + \left(\sum_{j=j_0}^{+\infty} 2^{j(s+1/2-1/p)q} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \leq A \right\}$ with the respective above sums replaced by maximum if $p = \infty$ or $q = \infty$.

Meyer wavelets can be used to efficiently compute the coefficients $c_{j,k}$ and $\beta_{j,k}$ by using the Fourier transform. Indeed, thanks to the Plancherel's identity, one obtains that

$$\beta_{j,k} = \sum_{\ell \in \Omega_j} \psi_\ell^{j,k} \theta_\ell, \tag{3.1}$$

where $\psi_\ell^{j,k} = \int_0^1 \psi_{j,k}(x) e^{-i2\pi\ell x} dx$ denote the Fourier coefficients of $\psi_{j,k}$ and $\Omega_j = \{\ell \in \mathbb{Z}; \psi_\ell^{j,k} \neq 0\}$. As Meyer wavelets $\psi_{j,k}$ are band-limited Ω_j is a finite subset set of $[-2^{j+2}c_0, -2^j c_0] \cup [2^j c_0, 2^{j+2} c_0]$ with $c_0 = 2\pi/3$ (see [25]), and fast algorithms for computing the above sum have been proposed by [27] and [38]. The coefficients $c_{j_0,k}$ can be computed analogously with ϕ instead of ψ and $\tilde{\Omega}_{j_0} = \{\ell \in \mathbb{Z}; \phi_\ell^{j_0,k} \neq 0\}$ instead of Ω_j.

Hence, the noisy Fourier coefficients $\hat{\theta}_\ell$ given by (1.10) can be used to quickly compute the following empirical wavelet coefficients of f as

$$\hat{\beta}_{j,k} = \sum_{\ell \in \Omega_j} \psi_\ell^{j,k} \hat{\theta}_\ell \text{ and } \hat{c}_{j_0,k} = \sum_{\ell \in \Omega_{j_0}} \phi_\ell^{j_0,k} \hat{\theta}_\ell. \tag{3.2}$$

3.2. *Nonlinear estimation via hard-thresholding.* It is well known that adaptivity can be obtained by using nonlinear estimators based on appropriate thresholding of the estimated wavelet coefficients (see e.g [12]) . A non-linear estimator by hard-thresholding is defined by

$$\hat{f}_n^h = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} \mathbb{1}_{\{|\hat{\beta}_{j,k}| \geqslant \lambda_{j,k}\}} \psi_{j,k} \tag{3.3}$$

where the $\lambda_{j,k}$'s are appropriate thresholds (positive numbers), and j_1 is the finest resolution level used for the estimation. As shown by [25], for periodic deconvolution the choice for j_1 and the thresholds $\lambda_{j,k}$ typically depends on the degree ν of ill-posedness of the problem. Following Theorem 1 in [25] to control moments of order 2 and 4 of $|\hat{\beta}_{j,k} - \beta_{j,k}|$ and the probability of deviation of $\hat{\beta}_{j,k}$ from $\beta_{j,k}$. For this, one needs the following assumption on the decay of the density $g(x)$ as $|x| \to +\infty$

Assumption 2. *There exists a constant $C > 0$ and a real $\alpha > 1$ such that the density g satisfies $g(x) \leq \frac{C}{1+|x|^{\alpha}}$ for all $x \in \mathbb{R}$.*

Note that Assumption 2 is not a very restrictive condition as g is supposed to be an integrable function on $\mathbb{R}$. This can also be viewed as a sufficient condition to ensure the existence of the density $G(x)$ introduced in (1.6).

Proposition 3. *Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $s - 1/p + 1/2 > 0$ and $A > 0$. Assume that g satisfies Assumptions 1 and 2. Then, there exist positive constants C_3 and C_4 such that for any $j \geq j_0 \geq 0$, $0 \leq k \leq 2^j - 1$ and all $f \in B^s_{p,q}(A)$, $\mathbb{E}|\hat{c}_{j_0,k} - c_{j_0,k}|^2 \leq C_3 \frac{2^{2j_0\nu}}{n}$, $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^2 \leq C_3 \frac{2^{2j\nu}}{n}$, and $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^4 \leq C_4 \left(\frac{2^{j4\nu}}{n^2} + \frac{2^{j(4\nu+1)}}{n^3} \right)$.*

Proposition 3 shows that the variance of the empirical wavelet coefficients is proportional to $\frac{2^{2j\nu}}{n}$ which comes from the amplification of the noise by the inversion of the convolution operator. The choice of the threshold $\lambda_{j,k}$ is done by controlling the probability of deviation of the empirical wavelet coefficients $\hat{\beta}_{j,k}$ from the true wavelet coefficient $\beta_{j,k}$, which is given by the following proposition:

Proposition 4. *Let $f \in L^2([0,1])$, $n \geq 1$ and $j \geq 0$. Suppose that g satisfies Assumption 2. For $0 \leq k \leq 2^j - 1$ and $\theta_\ell = \int_0^1 f(x) e^{-i2\pi\ell x} dx$, define*

$$\sigma_j^2 = 2^{-j}\epsilon^2 \sum_{\ell \in \Omega_j} |\gamma_\ell|^{-2}, V_j^2 = \|g\|_\infty 2^{-j} \sum_{\ell \in \Omega_j} \frac{|\theta_\ell|^2}{|\gamma_\ell|^2} \text{ and } \delta_j = 2^{-j/2} \sum_{\ell \in \Omega_j} \frac{|\theta_\ell|}{|\gamma_\ell|}.$$

with $\|g\|_\infty = \sup_{x \in \mathbb{R}}\{g(x)\}$. Let $t > 0$, then,

$$\mathbb{P}\left(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq 2 \max\left(\sigma_j \sqrt{\frac{2t}{n}}, \sqrt{\frac{2V_j^2 t}{n}} + \delta_j \frac{t}{3n} \right) \right) \leq 2\exp(-t).$$

Proposition 4 suggests to take level-dependent thresholds of the form

$$\lambda^*_{j,k} = \lambda^*_j = 2\max\left(\sigma_j \sqrt{\frac{2\eta \log(n)}{n}}, \sqrt{\frac{2\eta V_j^2 \log(n)}{n}} + \delta_j \frac{\eta \log(n)}{3n} \right) \tag{3.4}$$

where $\eta > 0$ is constant whose choice has to be discussed. The first term in the maximum (3.4) is the classical universal threshold with heteroscedastic variance σ_j^2 which corresponds to an upper bound of the variance of the Gaussian term $\epsilon \sum_{\ell \in \Omega_j} \frac{\eta_l}{\gamma_\ell}$ in the expression of $\hat{\beta}_{j,k}$. However, the second term in the maximum (3.4) depends on the modulus of the unknown Fourier coefficients θ_ℓ, and thus the thresholds λ_j^* cannot be used in practice.

As suggested by an anonymous referee, the computation of the threshold $\lambda_{j,k}^*$ can be simplified using the following arguments. Since there exists two constants C_1, C_2 such that for all $\ell \in \Omega_j$, $C_1 2^j \leq \ell \leq C_2 2^j$, and since $\lim_{|\ell| \to +\infty} \theta_\ell = 0$ uniformly for $f \in B_{p,q}^s(A)$ it follows that as $j \to +\infty$

$$V_j^2 = \|g\|_\infty 2^{-j} \sum_{\ell \in \Omega_j} \frac{|\theta_\ell|^2}{|\gamma_\ell|^2} = o\left(2^{-j} \sum_{\ell \in \Omega_j} |\gamma_\ell|^{-2}\right) = o\left(\sigma_j^2\right).$$

Also, if $f \in B_{p,q}^s(A)$ then $\sum_{\ell \in \Omega_j} |\theta_\ell|^2 = o(1)$ as $j \to +\infty$, and thus by Cauchy-Schwarz inequality, then as $j \to +\infty$

$$\delta_j = 2^{-j/2} \sum_{\ell \in \Omega_j} \frac{|\theta_\ell|}{|\gamma_\ell|} \leq 2^{-j/2} \left(\sum_{\ell \in \Omega_j} |\theta_\ell|^2\right)^{1/2} \left(\sum_{\ell \in \Omega_j} |\gamma_\ell|^{-2}\right)^{1/2} = o\left(\sigma_j\right),$$

which finally implies that $\lambda_j^* = o\left(\lambda_j\right)$ as $j \to +\infty$ where

$$\lambda_{j,k} = \lambda_j = \sigma_j \sqrt{\frac{2\eta \log(n)}{n}}, \tag{3.5}$$

which corresponds to the usual universal thresholds for deconvolution problem based on wavelet decomposition as in [25]. Hence if one chooses j_0 to be slowly growing with n (e.g. $j_0 = \log(\log(n))$), or avoid thresholding at very low resolution levels, then the threshold λ_j can be used instead of λ_j^* whose computation would require an estimation of the $|\theta_\ell|$'s.

Combining Propositions 3 and 4, and the above remarks on the thresholds λ_j and λ_j^*, then by arguing as in the proof of Theorem 1 in [25], we finally arrive at the following theorem which gives an upper bound for the minimax risk over a large class of Besov balls.

THEOREM 2. *Assume that g satisfies Assumptions 1 and 2. Let j_1 and j_0 be the largest integers such that $2^{j_1} \leq \left(\frac{n}{\log(n)}\right)^{\frac{1}{2\nu+1}}$ and $2^{j_0} \leq \log(\log(n))$. Let $\hat{f}_n^h$ be the non-linear estimator obtained by hard-thresholding with the above choice for j_1 and j_0, and using the thresholds λ_j defined by equation*

(3.5) with $\eta \geq 2$. Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Let $p' = \min(2, p)$, $s' = s + 1/2 - 1/p$, and assume that $s \geq 1/p'$.
If $s \geq (2\nu+1)(1/p - 1/2)$, then

$$\sup_{f \in B^s_{p,q}(A)} \|\hat{f}^h_n - f\|^2 = \mathcal{O}\left(n^{-\frac{2s}{2s+2\nu+1}} (\log n)^{\beta}\right) \text{ with } \beta = \frac{2s}{2s+2\nu+1}.$$

If $s < (2\nu+1)(1/p - 1/2)$, then

$$\sup_{f \in B^s_{p,q}(A)} \|\hat{f}^h_n - f\|^2 = \mathcal{O}\left(\left(\frac{n}{\log(n)}\right)^{-\frac{2s'}{2s'+2\nu}}\right).$$

In standard periodic deconvolution in the white noise model (see e.g. [25]), there exists two different upper bounds for the minimax rate which are usually referred to as the dense case ($s \geq (2\nu+1)(1/p-1/2)$) when the hardest functions to estimate are spread uniformly over $[0,1]$, and the sparse case ($s < (2\nu+1)(1/p-1/2)$) when the worst functions to estimate have only one non-vanishing wavelet coefficient. Theorem 2 shows that a similar phenomenon holds for the model (1.4), and to the best of our knowledge, this is a new result.

4. Minimax lower bound. The following theorem gives an asymptotic lower bound on the minimax risk $\mathcal{R}_n(B^s_{p,q}(A))$ for a large class of Besov balls.

THEOREM 3. *Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Suppose that g satisfies Assumption 1. Let $p' = \min(2,p)$. Assume that $s \geq 1/p'$ and $\nu > 1/2$.*
If $s \geq (2\nu+1)(1/p-1/2)$ and $s > 2\nu+1$ (dense case), there exits a constant M_1 depending only on A, s, p, q such that

$$\mathcal{R}_n(B^s_{p,q}(A)) \geq M_1 n^{-\frac{2s}{2s+2\nu+1}} \text{ as } n \to +\infty$$

In the dense case, the hardest functions to estimate are spread uniformly over the interval $[0,1]$, and the proof is based on an adaptation of Assouad's cube technique (see e.g Lemma 10.2 in [21]) to the specific setting of model (1.4). Lower bounds for minimax risk are classically derived by controlling the probability for the likelihood ratio (in the statistical model of interested) of being strictly greater than some constant uniformly over an appropriate set of test functions. To derive Theorem 3 , we show that one needs to control the expectation over the random shifts of the likelihood ratio associated to model (1.4), and not only the likelihood ratio itself. Hence, the proof of Theorem 3 is not a direct and straightforward adaptation of Assouad's cube technique or Lemma 10.1 in [21] as used classically in a standard white noise model to derive minimax risk in nonparametric deconvolution in the dense case. For more details, we refer to the proof of Theorem 3 in the Appendix.

Deriving minimax risk in the dense case for the model (1.4) is rather difficult and the proof is quite long and technical. In the sparse case, finding lower bounds for the minimax rate is also a difficult task. We believe that this could be done by adapting to model (1.4) a result by [28] which yields a lower bound for a specific problem of distinguishing between a finite number of hypotheses (see Lemma 10.1 in [21]). However, this is far beyond the scope of this paper and we leave this problem open for future wok.

5. Estimating f when the density g is unknown. Obviously, assuming that the density g of the shifts is known is not very realistic in practice. However, estimating f when the density g is unknown falls into the setting of inverse problems with an unknown operator which is a difficult problem. Recently some papers [9], [13], [22] have considered nonparametric estimation for inverse problem with a partially unknown operator, by assuming that an independent sample of noisy eigenvalues is available which allows an estimation of the γ_ℓ's. In the settings of these papers, the law of the noisy eigenvalues sample is supposed to be known (typically Gaussian), but in the model (1.4) such assumptions are not realistic, and a data-based estimator of g has to be found. For this, we propose to make a connection between estimation of a mean pattern in the randomly shifted curve model (1.4) and some well-known results in the shape analysis literature on Frechet mean for data lying in a nonlinear manifold (see e.g. [2], [3] and references therein).

5.1. *Frechet mean for functional data.* Suppose that $Z_1, \dots, Z_n$ denote iid random variables taking their values in a vector space V. As V is a linear space (with addition well defined), estimation of a mean pattern for the Z_m's is given by the usual linear average $\overline{Z}_n = \frac{1}{n}\sum_{m=1}^{n} Z_m$. However in many applications, some geometric and statistical considerations may lead to the assumption that two vectors Z, Z' in V are considered to be the same if they are equal up to certain transformations which are represented by the action of some group H on the space V. A well-known example (see [2], [3] and references therein) is the case where $V = \mathbb{R}^{2k}$, the space of k points in the plane $\mathbb{R}^2$, and H is generated by composition of scaling, rotations and translations of the plane, namely

$$h \cdot Z = a \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} Z + b,$$

for $h = h(a, \theta, b) \in H$, with $(a, \theta, b) \in \mathbb{R}^+ \times [0, 2\pi] \times \mathbb{R}^2$. In this setting, two vectors $Z, Z' \in \mathbb{R}^{2k}$ represent the same shape if $d_H(Z, Z') := \inf_{(a,\theta,b)\in\mathbb{R}^+\times[0,2\pi]\times\mathbb{R}^2} \|Z - h(a, \theta, b) \cdot Z'\|_{\mathbb{R}^{2k}} = 0$, which leads to the Kendall's shape space Σ_2^k consisting of the equivalent classes of shapes in $\mathbb{R}^{2k}$ under the action of scaling, rotations and translations (see e.g. [2], [3] and references therein). Since the space Σ_2^k is a nonlinear manifold, the usual linear average $\overline{Z}_n$ does not fall into Σ_2^k due to the fact that the Euclidean distance $\|\cdot\|_{\mathbb{R}^{2k}}$ is not meaningful to represent shape variations. A better notion of empirical mean $\tilde{Z}_n$ of n shapes in $\mathbb{R}^{2k}$ is given by (see e.g. [2]): $\tilde{Z}_n = \arg\min_{Z\in\Sigma_2^k} \frac{1}{n}\sum_{m=1}^{n} d_H^2(Z, Z_m)$. More generally, Fréchet [14] has extended the notion of averaging to general metric spaces via minimum mean squared error estimation in the following way: if $Z_1, \ldots, Z_n$ are iid random variables in a general metric space $\mathcal{M}$, with a distance $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^+$, then the Frechet mean of a collection of such data points is defined as the minimizer (not necessarily unique) of the sum-of-squared distances to each of the data points, that is

$$\tilde{Z}_n = \arg\min_{Z\in\mathcal{M}} \frac{1}{n}\sum_{m=1}^{n} d^2(Z, Z_m).$$

Now let us return to the randomly shifted curve model (1.4). Define $H = \mathbb{R}$ as the translation group acting on periodic functions $f \in L^2([0, 1])$ with period 1 by

$$\tau \cdot f(x) = f(x + \tau), \qquad \text{for } x \in [0, 1] \text{ and } \tau \in H.$$

Let $Y_1, \ldots, Y_n$ be n functions (possibly random) in $L^2([0, 1])$. Following the definition of Frechet mean, a notion of average for functional data taking into account the action of the translation group $H = \mathbb{R}$ would be

$$\tilde{f}_n = \arg\min_{f\in L^2([0,1])} \frac{1}{n}\sum_{m=1}^{n} \min_{\tau_m\in\mathbb{R}} \int_0^1 |f(x) - Y_m(x + \tau_m)|^2 dx.$$

If the Y_m's are noisy curves generated from the randomly shifted curve model (1.4), a pre-smoothing step of the observed curves seems natural to compute a consistent Frechet mean estimate. In the case of the translation group, this smoothing step and the definition of Frechet mean can be expressed in the Fourier domain as

$$(5.1)\quad (\hat{\theta}_{-\ell_0}, \ldots, \hat{\theta}_{\ell_0}) = \underset{(\theta_{-\ell_0}, \ldots, \theta_{\ell_0}) \in \mathbb{R}^{2\ell_0+1}}{\arg\min} \frac{1}{n} \sum_{m=1}^{n} \min_{\tau_m \in \mathbb{R}} \sum_{|\ell| \leq \ell_0} |c_{m,\ell} e^{2i\ell\pi\tau_m} - \theta_\ell|^2,$$

where $c_{m,\ell} = \int_0^1 e^{-2i\ell\pi x} dY_m(x)$ and ℓ_0 is some frequency cut-off parameter whose choice will be discussed later. A smoothed Frechet mean is then given by $\tilde{f}_{n,\ell_0} = \sum_{|\ell| \leq \ell_0} \hat{\theta}_\ell e^{-2i\ell\pi x}$. The computation of $\bar{f}_{n,\ell_0}$ can be made in two steps since it can be checked that $\hat{\theta}_\ell = \frac{1}{n} \sum_{m=1}^{n} c_{m,\ell} e^{2i\ell\pi\hat{\tau}_m}$, where

$$(5.2)\quad (\hat{\tau}_1, \ldots, \hat{\tau}_n) = \underset{(\tau_1, \ldots, \tau_n) \in \mathbb{R}^n}{\arg\min} \frac{1}{n} \sum_{m=1}^{n} \sum_{|\ell| \leq \ell_0} |c_{m,\ell} e^{2i\ell\pi\tau_m} - \frac{1}{n} \sum_{q=1}^{n} c_{q,\ell} e^{2i\ell\pi\tau_q}|^2.$$

Therefore, computing the Frechet mean of the smoothed curves $Y_1, \ldots, Y_n$ amounts to minimise the above criteria which automatically yields an estimation of the random shifts $\tau_1, \ldots, \tau_n$ in model (1.4). This allows an estimation of the common shape f by $\tilde{f}_{n,\ell_0}$ in the case of an unknown density g, and the estimates $(\hat{\tau}_1, \ldots, \hat{\tau}_n)$ of the random shifts can be used to estimate the density g itself and the eigenvalues γ_ℓ. The goal of this section is thus to study some statistical properties of such a two-step procedure which, to the best of our knowledge, has not been considered before in the setting of model (1.4) and in connection with Frechet mean for functional data. Moreover, it will be shown in our numerical experiments that the criterion (5.2) can be minimised using a standard gradient-descent algorithm which leads to a new and fast method for estimating f in the case of an unknown density g.

5.2. *Upper bound for the estimation of the shifts.* Recall that our model (1.4) in the Fourier domain is

$$(5.3)\qquad c_{m,\ell} = \theta_\ell e^{-i2\pi\ell\tau_m^*} + \epsilon z_{\ell,m},\ \ell \in \mathbb{Z} \text{ for } m = 1, \ldots, n,$$

where $z_{\ell,m}$ are iid $N_{\mathbb{C}}(0, 1)$ variables, the random shifts $\tau_m^*, m = 1, \ldots, n$ are i.i.d variables with density g, and $\theta_\ell = \int_0^1 f(x) e^{-i2\pi\ell x} dx$. Model (5.3) is clearly non-identifiable, as for any $\tau_0 \in \mathbb{R}$, one can replace the θ_ℓ's by $\theta_\ell e^{i2\pi\ell\tau_0}$ and the τ_m^*'s by $\tau_m^* - \tau_0$ without changing the formulation of the model. Let us thus introduce the following identifiability conditions:

ASSUMPTION 3. *The density g has a compact support included in the interval $\mathcal{T} = [-\frac{1}{4}, \frac{1}{4}]$ and has zero mean i.e. is such that $\int_{\mathcal{T}} \tau g(\tau) d\tau = 0$.*

ASSUMPTION 4. *The unknown shape function f is such that $\theta_1 \neq 0$.*

Then, define the following criterion for $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n) \in \mathcal{T}^n$

$$M_n(\boldsymbol{\tau}) \quad = \quad \frac{1}{n} \sum_{m=1}^{n} \sum_{|\ell| \leq \ell_0} |c_{m,\ell} e^{2i\ell\pi\tau_m} - \frac{1}{n} \sum_{q=1}^{n} c_{q,\ell} e^{2i\ell\pi\tau_q}|^2.$$

Let $\overline{\mathcal{T}}_n = \{(\tau_1, \ldots, \tau_n) \in \mathcal{T}^n$ such that $\sum_{m=1}^{n} \tau_m = 0\}$. Using the identifiabilty condition given by Assumption 3, it is natural to define an estimation of the true shifts $\tau_1^*, \ldots, \tau_n^*$ as

$$\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_n) = \arg\min_{\boldsymbol{\tau} \in \overline{\mathcal{T}}_n} M_n(\boldsymbol{\tau}),$$

i.e. by considering the estimators that minimize the empirical criterion $M_n(\boldsymbol{\tau})$ on the constrained set $\overline{\mathcal{T}}_n$. Then, the following theorem holds:

THEOREM 4. *Suppose that Assumptions 3 and 4 hold. Then, for any* $t > 0$

$$\mathbb{P}\left(\frac{1}{n} \sum_{m=2}^{n} (\hat{\tau}_m - \tau_m^*)^2 \geq C(f, \ell_0, \epsilon, n, t, g) \right) \leq 3\exp(-t), \tag{5.4}$$

with $C(f, \ell_0, \epsilon, n, t, g) = 4\max\left[C_1(f, \ell_0)\left(\sqrt{C_2(\epsilon, n, \ell_0, t)} + C_2(\epsilon, n, \ell_0, t) \right), C_3(t, n, g) \right]$, *where* $C_1(f, \ell_0)$ *is a positive constant depending only on the shape function* f *and the frequency cut-off parameter* ℓ_0,

$$C_2(\epsilon, n, \ell_0, t) = \epsilon^2(2\ell_0 + 1) + 2\epsilon^2 \sqrt{\frac{2\ell_0 + 1}{n} t} + 2\frac{\epsilon^2}{n} t,$$

and

$$C_3(t, n, g) = \left(\sqrt{2\sigma_g^2 \frac{t}{n}} + \frac{t}{12n} \right)^2 \text{ with } \sigma_g^2 = \int_{\mathcal{T}} \tau^2 g(\tau) d\tau.$$

Theorem 4 provides an upper bound (in probability) for the consistency of the estimators $\hat{\tau}_m$ of the true random shifts $\tau_m^*, m = 2, \ldots, n$ using the standard squared distance. Note that since the minimum of $M_n(\boldsymbol{\tau})$ is computed on the constrained set $\overline{\mathcal{T}}_n$, it follows that $\hat{\tau}_1 = -\sum_{m=2}^{n} \hat{\tau}_m$. However, one can remark that as $n \to +\infty$, the constant $C(f, \ell_0, \epsilon, n, t, g)$ in inequality (5.4) tends to $4C_1(f, \ell_0)\left(\epsilon^2(2\ell_0 + 1) + \epsilon\sqrt{2\ell_0 + 1}\right)$. This shows that $\hat{\tau}_m, m = 2, \ldots, n$ are not consistent estimators in the sense that inequality (5.4) cannot be used to prove that $\lim_{n \to +\infty} \frac{1}{n} \sum_{m=2}^{n} (\hat{\tau}_m - \tau_m^*)^2 = 0$ in probability. To the contrary, inequality (5.4) suggests that there exists a constant $C > 0$ such that $\frac{1}{n} \sum_{m=2}^{n} (\hat{\tau}_m - \tau_m^*)^2 > C\epsilon^2(2\ell_0 + 1)$ with positive probability. Therefore this suggests that the accuracy of the estimates $\hat{\tau}_m, m = 2, \ldots, n$ depends on the level of noise ϵ^2 and the frequency cut-off ℓ_0.

5.3. *Lower bound for the estimation of the shifts.* Let us now prove that the consistency of any estimate of the random shifts in model (5.3) is limited by the level of noise ϵ^2 in the observed curves. For this let us make the following smoothness assumptions:

ASSUMPTION 5. *The function f is such that $\sum_{\ell \in \mathbb{Z}} (2\pi\ell)^2 |\theta_\ell|^2 < +\infty$.*

ASSUMPTION 6. *The density g is compactly supported on a interval $\mathcal{T} = [\tau_{\min}, \tau_{\max}]$ such that $\lim_{\tau \to \tau_{\min}} g(\tau) = \lim_{\tau \to \tau_{\max}} g(\tau) = 0$.*

Then, the following theorem holds:

THEOREM 5. *Denote by $X = (c_{m,\ell})_{\ell \in \mathbb{Z}, m=1,\dots,n}$ the set of observations taking values in the set $\mathcal{X} = \mathbb{C}^{\infty \times n}$. Let $\hat{\tau}^n = \hat{\tau}^n(X)$ denote any estimator (a measurable function of the observations X) of the true shifts $(\tau_1, \dots, \tau_n)$. Then, under Assumptions 5 and 6*

$$\mathbb{E}\left(\frac{1}{n}\sum_{m=1}^{n}(\hat{\tau}_m^n - \tau_m^*)^2\right) \geq \frac{\epsilon^2}{\sum_{\ell \in \mathbb{Z}} (2\pi\ell)^2 |\theta_\ell|^2 + \epsilon^2 \int_{\mathcal{T}} \left(\frac{\partial}{\partial \tau} \log g(\tau)\right)^2 g(\tau) d\tau}.$$

Clearly, Theorem 5 shows that as $n \to +\infty$ then $\mathbb{E}\left(\frac{1}{n}\sum_{m=1}^{n}(\hat{\tau}_m^n - \tau_m^*)^2\right)$ does not converge to zero which explains the results obtained in Theorem 4 on the consistency of the estimators $\hat{\tau}_m, m = 2, \dots, n$ based on Frechet mean for functional data. Note that Assumption 5 can be avoided if one only considers estimators $\hat{\tau}^{n,\ell_0}$ of the shifts based on the observations $c_{m,\ell}$ for $m = 1, \dots, n$ and $|\ell| \leq \ell_0$ in model (5.3). In this case the lower bound in Theorem 5 becomes

$$\mathbb{E}\left(\frac{1}{n}\sum_{m=1}^{n}(\hat{\tau}_m^{n,\ell_0} - \tau_m^*)^2\right) \geq \frac{\epsilon^2}{\sum_{|\ell| \leq \ell_0} (2\pi\ell)^2 |\theta_\ell|^2 + \epsilon^2 \int_{\mathcal{T}} \left(\frac{\partial}{\partial \tau} \log g(\tau)\right)^2 g(\tau) d\tau}.$$

5.4. *Estimation of the mean pattern f and the density g.* An estimation of the eigenvalue γ_ℓ is given by $\hat{\gamma}_\ell = \frac{1}{n}\sum_{m=2}^{n} e^{-i2\pi\ell\hat{\tau}_m}$, for $|\ell| \leq \ell_0$ and an estimator for the density g is naturally given by $\hat{g}(x) = \sum_{|\ell| \leq \ell_0} \hat{\gamma}_\ell e^{-i2\pi\ell x}$. The mean pattern f can be estimated by the smoothed Frechet mean $\tilde{f}_{n,\ell_0}$ defined in Section 5.1, but following the results in Section on nonlinear wavelet-based estimation, two other estimators for f can be defined: the first one is given by

$$\hat{f}_{n,1} = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k,1} \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k,1} \mathbb{1}_{\{|\hat{\beta}_{j,k,1}| \geqslant \hat{\lambda}_{j,1}\}} \psi_{j,k} \tag{5.5}$$

where $\hat{\beta}_{j,k,1} = \sum_{\ell \in \Omega_j} \psi_\ell^{j,k} \hat{\theta}_{\ell,1}$ and $\hat{c}_{j_0,k,1} = \sum_{\ell \in \Omega_{j_0}} \phi_\ell^{j_0,k} \hat{\theta}_{\ell,1}$ with

$$\hat{\theta}_{\ell,1} = \frac{1}{\hat{\gamma}_\ell} \left(\frac{1}{n} \sum_{m=1}^{n} c_{\ell,m} \right),$$

and $\hat{\lambda}_{j,1} = 2\hat{\sigma}_j \sqrt{\frac{2\eta \log(n)}{n}}$ is the threshold suggested by the expression (3.5) of λ_j with $\hat{\sigma}_j^2 = 2^{-j}\epsilon^2 \sum_{\ell \in \Omega_j} |\hat{\gamma}_\ell|^{-2}$. A second estimator is given by first realigning the curves using the estimation of the shifts namely

$$\hat{f}_{n,2} = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k,2} \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k,2} \mathbb{1}_{\{|\hat{\beta}_{j,k,2}| \geqslant \hat{\lambda}_{j,2}\}} \psi_{j,k} \tag{5.6}$$

where $\hat{\beta}_{j,k,2} = \sum_{\ell \in \Omega_j} \psi_\ell^{j,k} \hat{\theta}_{\ell,2}$ and $\hat{c}_{j_0,k,2} = \sum_{\ell \in \Omega_{j_0}} \phi_\ell^{j_0,k} \hat{\theta}_{\ell,2}$ with

$$\hat{\theta}_{\ell,2} = \frac{1}{n} \sum_{m=2}^{n} c_{\ell,m} e^{i2\pi\ell\hat{\tau}_m},$$

and $\hat{\lambda}_{j,2}$ is a threshold whose choice would depend on the law of the $\hat{\beta}_{j,k,2}$'s. Studying the consistency and the rate of convergence of the estimators $\hat{f}_{n,1}$ and $\hat{f}_{n,2}$ is a difficult task. Indeed the results in Section 5.1 have been derived using the fact that the law of the wavelet coefficients $\hat{\beta}_{j,k}$ and $\hat{c}_{j_0,k}$ given by (3.2) is *known* which allows the calibration of the threshold λ_j in (3.5). Thus, we simply suggest to take $\hat{\lambda}_{j,2} = \hat{\lambda}_{j,1}$. Extending the asymptotic results of Section 5.1 remains a challenge that is beyond the scope of this paper. Moreover, the results of Theorems 4 and 5 suggest that the estimators $\hat{f}_{n,1}$ and $\hat{f}_{n,2}$ could be consistent by considering a double asymptotic setting with $n \to +\infty$ and $\epsilon \to 0$ which is an interesting point of view for future work that certainly leads to different minimax rates of convergence.

6. Numerical experiments. We compare our approach with the Procrustean mean which is a standard algorithm commonly used to extract a mean pattern. The Procrustean mean is based on an alternative scheme between estimation of the shifts and averaging of back-transformed curves given estimated values of the shifts parameters, see e.g [45], [26]. To be more precise it consists of an initialization step $\hat{f}_0 = \frac{1}{n}\sum_{m=1}^n Y_m$ which is the simple average of the observed curves that is taken as a first reference mean pattern. Then, at iteration $1 \leq i \leq i_{\max}$, it computes for all $1 \leq m \leq n$ an estimation $\hat{\tau}_{m,i}$ of the m-th shift as $\hat{\tau}_{m,i} = \arg\min_{\tau \in \mathbb{R}} \|Y_m(\cdot + \tau) - \hat{f}_{i-1}\|^2$ and then takes $\hat{f}_i(x) = \frac{1}{n}\sum_{m=1}^n Y_m(x + \hat{\tau}_{m,i})$ as a new reference mean pattern. This is repeated until the estimated reference curve does not change, and usually the algorithm converges in a few steps (we took $i_{\max} = 3$). In all simulations, we used the wavelet toolbox Wavelab [7] and the WaveD algorithm developed by [38] for fast deconvolution with Meyer wavelets.

6.1. *Shift estimation by gradient descent.* Let us denote by $\nabla M_n(\boldsymbol{\tau}) \in \mathbb{R}^n$ the gradient of $M_n(\boldsymbol{\tau})$ at $\boldsymbol{\tau} \in \mathbb{R}^n$. This gradient is simple to compute as for $m = 1, \ldots, n$:

$$\frac{\partial}{\partial \tau_m} M_n(\boldsymbol{\tau}) = -\frac{2}{n} \sum_{|\ell| \leq \ell_0} \Re \left[2i\pi\ell c_{\ell,m} e^{2i\ell\pi\tau_m} \left(\overline{\frac{1}{n} \sum_{q=1}^{n} c_{\ell,q} e^{2i\ell\pi\tau_q}} \right) \right]$$

In practice, to estimate the shifts, the criterion $M_n(\boldsymbol{\tau})$ is then minimized by the following gradient descent algorithm with the constraint that $\tau_1 = -\sum_{m=2}^{n} \tau_m$:

Initialization : let $\boldsymbol{\tau}^0 = 0 \in \mathbb{R}^n$, $\delta_0 = \frac{1}{\|\nabla M_n(\boldsymbol{\tau}^0)\|}$, $M(0) = M_n(\boldsymbol{\tau}^0)$, and set $p = 0$.

Step 2 : let $\boldsymbol{\tau}^{new} = \boldsymbol{\tau}^p - \delta_p \nabla M_n(\boldsymbol{\tau}^p)$ and $\tau_1^{new} = -\sum_{m=2}^{n} \tau_m^{new}$.
Let $M(p+1) = M_n(\boldsymbol{\tau}^{new})$.
While $M(p+1) > M(p)$ **do**

$$\delta_p = \delta_p/\kappa, \quad \text{and } \boldsymbol{\tau}^{new} = \boldsymbol{\tau}^p - \delta_p \nabla M_n(\boldsymbol{\tau}^m), \text{ with } \tau_1^{new} = -\sum_{m=2}^{n} \tau_m^{new},$$

and set $M(p+1) = M_n(\boldsymbol{\tau}^{new})$.
End while
Then, take $\boldsymbol{\tau}^{p+1} = \boldsymbol{\tau}^{new}$.

Step 3 : if $M(p) - M(p+1) \geq \rho(M(1) - M(p+1))$ then set $p = p+1$ and return to Step 2, else stop the iterations, and take $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^{p+1}$.

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 δ_p.

6.2. *Estimation with an unknown density g.* For the mean pattern f to recover, we consider the four tests functions shown in Figures 1a-4a. Then, we simulate $n = 200$ randomly shifted curves with shifts following a Laplace distribution $g(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\sqrt{2}\frac{|x|}{\sigma}\right)$ with $\sigma = 0.1$. Gaussian noise with a moderate variance (different to that used in the Laplace distribution) is then added to each curve. A subsample of 10 curves is shown in Figures 1b-4b for each test function, and the average of the observed curves, referred to as the direct mean in what follows, is displayed in Figures1c- 4c. Note this gives a poor estimation of the mean pattern.

The Fourier coefficients of the density g are given by $\gamma_\ell = \frac{1}{1+2\sigma^2\pi^2\ell^2}$ which corresponds to a degree of ill-posedness $\nu = 2$. An estimation of $\hat{\gamma}_\ell$ of γ_ℓ can be performed as explained in Section 5 using the gradient descent algorithm described in Section 6.1. The choice of the frequency cut-off ℓ_0 used to compute these estimators is a delicate model selection problem, and in our simulations we took the arbitrary choice $\ell_0 = 3$ which gives satisfactory results in the numerical experiments. Theorem 4 suggests that this choice should depend on n and the level of noise ϵ, but finding data-based values for ℓ_0 remains a challenge that we leave open for future work.

To compute the threshold $\hat{\lambda}_{j,1} = \hat{\lambda}_{j,2}$ used in the definition of $\hat{f}_{n,1}$ and $\hat{f}_{n,2}$ (see Section 5) one has to estimate ϵ^2. This is done by taking $\hat{\epsilon}^2 = \frac{1}{n}\sum_{m=1}^{n} \hat{\epsilon}_m^2$, where the variance ϵ_m^2 of the noise for the m-th curve is easily estimated using the wavelet coefficients at the finest resolution level. Note that such thresholds are quite simple to compute using the Fast Fourier Transform and the fact that the set of frequencies Ω_j can be easily obtained using WaveLab. Finally, we have found that choosing η between 1 and 2 to compute $\hat{\lambda}_{j,1}$ gives quite satisfactory results.

Then, we took $j_0 = 3 \approx \log_2(\log(n))$, but the choice $j_1 \approx \frac{1}{2\nu+1}\log_2(\frac{n}{\log(n)})$ is obviously too small. So in our simulations, j_1 is chosen to be the maximum resolution level allowed by the discretization i.e. $j_1 = \log_2(N) - 1 = 7$. For each test function, the estimators $\hat{f}_{n,1}$, $\hat{f}_{n,2}$ are displayed in Figures 1(d)(e) -4(d)(e). The Procrustean mean is displayed is Figures 1(f) -4(f). One can see that the results are rather satisfactory for $\hat{f}_{n,1}$ and the Procrustean mean. Clearly the best results are given by the estimator $\hat{f}_{n,2}$ which gives very good estimates of the function f particularly for functions with isolated singularites such as the Blocks and Bumps functions in Figures 3 and 4 . It should be noted that these results are obtained in the case of an unknown density g which shows the quality of the procedure proposed in Section 5 to estimate the shifts and the γ_ℓ's. For reasons of space a detailed simulation study is not given, but it has been found that the good performances of the wavelet-based estimator remain consistent across other standard test signals.

7. Conclusions and future work. This paper makes a connection between mean pattern estimation and the statistical analysis of inverse prob-

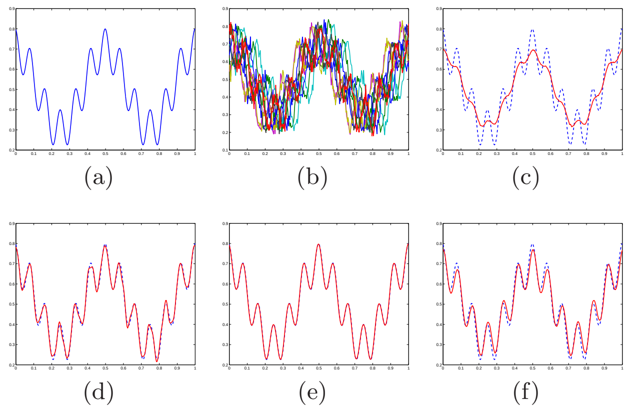

FIG 1. *Wave function. (a) Mean pattern f, (b) Sample of 10 curves out of $n = 200$, (c) Direct mean, Deconvolution by wavelet thresholding with (d) $\hat{f}_{n,1}$ and (e) $\hat{f}_{n,2}$, (f) Procrustean mean*

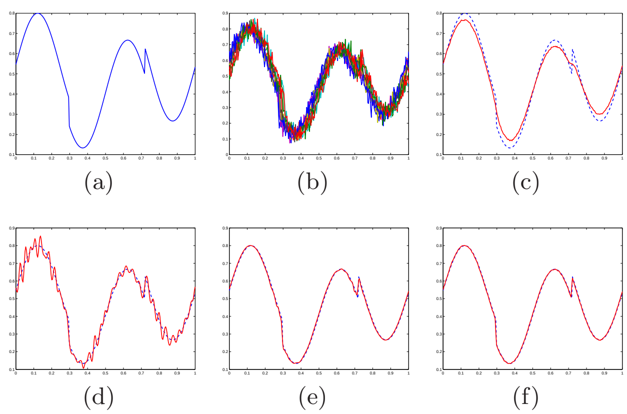

FIG 2. *HeaviSine function. (a) Mean pattern f, (b) Sample of 10 curves out of $n = 200$, (c) Direct mean, Deconvolution by wavelet thresholding with (d) $\hat{f}_{n,1}$ and (e) $\hat{f}_{n,2}$, (f) Procrustean mean*

lems for a very simple model with shifted curves. A natural extension would be to consider more complex deformations in SIM models such as the homothetic shifted regression model proposed in [44], or the rigid deformation model for images considered in [6]. Another promising approach would be to consider a double asymptotic setting with $n \to +\infty$ and $\epsilon \to 0$ to study the consistency and rate of convergence for estimators of the mean pattern in the case of an unknown density g.

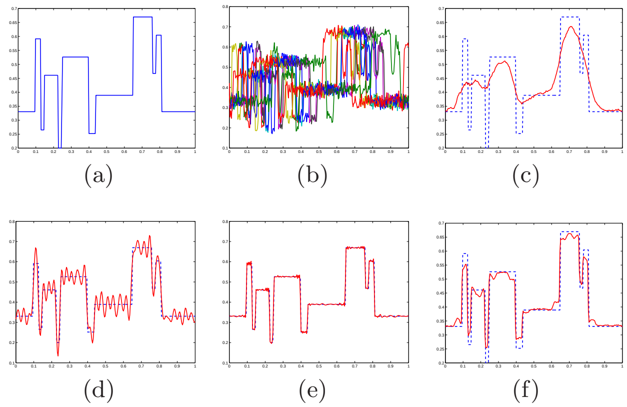

FIG 3. *Blocks function. (a) Mean pattern f, (b) Sample of 10 curves out of $n = 200$, (c) Direct mean, Deconvolution by wavelet thresholding with (d) $\hat{f}_{n,1}$ and (e) $\hat{f}_{n,2}$, (f) Procrustean mean*

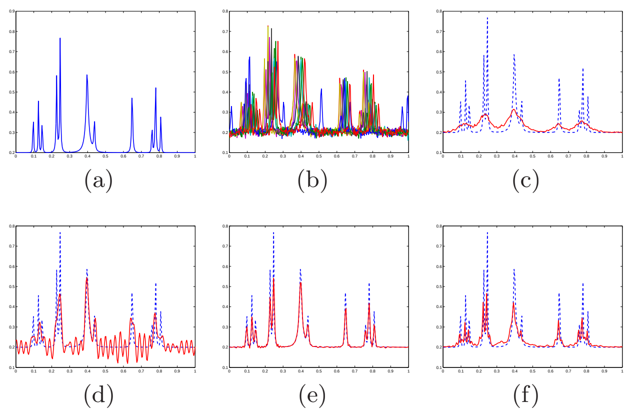

FIG 4. *Bumps function. (a) Mean pattern f, (b) Sample of 10 curves out of $n = 200$, (c) Direct mean, Deconvolution by wavelet thresholding with (d) $\hat{f}_{n,1}$ and (e) $\hat{f}_{n,2}$, (f) Procrustean mean*

APPENDIX A: APPENDIX SECTION

In what follows C will denote a generic constant whose value may change from line to line.

Proof of Theorem 1: it follows immediately from Theorem 2 and Theorem 3. $\square$

Proof of Proposition 1: let $\kappa_\ell = \left(\frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right)\theta_\ell$ and $\epsilon_{\ell,n} = \frac{\epsilon}{\gamma_\ell}\left(\frac{1}{n}\sum_{m=1}^{n} z_{\ell,m}\right)$

for all $\ell \in \mathbb{Z}$. Then, for a given filter λ, the risk $\mathcal{R}(\hat{f}_{n,\lambda}, f)$ can be written as

$$\begin{aligned}\mathcal{R}(\hat{f}_{n,\lambda}, f) &= \sum_{\ell\in\mathbb{Z}}(\lambda_\ell - 1)^2\theta_\ell^2 + \mathbb{E}\left[\lambda_\ell^2|\kappa_\ell|^2 + \lambda_\ell^2|\epsilon_{\ell,n}|^2\right] + \lambda_\ell(\lambda_\ell - 1)\mathbb{E}[\bar{\theta}_\ell\kappa_\ell + \theta_\ell\bar{\kappa}_\ell] \\ &\quad + \lambda_\ell(\lambda_\ell - 1)\mathbb{E}[\theta_\ell\bar{\epsilon}_{\ell,n} + \bar{\theta}_\ell\epsilon_{\ell,n}] + \lambda_\ell^2\mathbb{E}[\bar{\kappa}_\ell\epsilon_{\ell,n} + \kappa_\ell\bar{\epsilon}_{\ell,n}].\end{aligned}$$

Now using the fact that κ_ℓ and $\epsilon_{\ell,n}$ are independent and that $\mathbb{E}\epsilon_{\ell,n} = 0$, we obtain that

$$\begin{aligned}R(\hat{f}_{n,\lambda}, f) &= \sum_{k\in\mathbb{Z}}\left[(\lambda_\ell - 1)^2|\theta_\ell|^2 + \lambda_\ell^2|\theta_\ell|^2\mathbb{E}\left|\frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right|^2 + \frac{\lambda_\ell^2\epsilon^2}{n|\gamma_\ell|^2}\right] \\ &= \sum_{\ell\in\mathbb{Z}}(\lambda_\ell - 1)^2|\theta_\ell|^2 + \sum_{\ell\in\mathbb{Z}}\lambda_\ell^2|\theta_\ell|^2\left(\frac{1}{|\gamma_\ell|^2}\left(\frac{1}{n} + \frac{n-1}{n}\gamma_\ell\gamma_{-\ell}\right) - 1\right) \\ &\quad + \sum_{\ell\in\mathbb{Z}}\frac{\lambda_\ell^2\epsilon^2}{n|\gamma_\ell|^2} \\ &= \sum_{\ell\in\mathbb{Z}}(\lambda_\ell - 1)^2|\theta_\ell|^2 + \sum_{\ell\in\mathbb{Z}}\frac{\lambda_\ell^2}{n}\left[|\theta_\ell|^2\left(\frac{1}{|\gamma_\ell|^2} - 1\right) + \frac{\epsilon^2}{|\gamma_\ell|^2}\right],\end{aligned}$$

which completes the proof. □

Proof of Proposition 2: from Proposition 1 it follows that

$$\mathcal{R}(\hat{f}_{n,\lambda^M}, f) = \sum_{|\ell|>M_n}|\theta_\ell|^2 + \frac{1}{n}\sum_{|\ell|\leq M_n}\left(|\theta_\ell|^2\left(\frac{1}{|\gamma_\ell|^2} - 1\right) + \frac{\epsilon^2}{|\gamma_\ell|^2}\right)$$

By assumption $f \in H_s(A)$, which implies that there exists two positive constants C_1 and C_2 not depending on f and n such that for all sufficiently large n, $\sum_{|\ell|>M_n}|\theta_\ell|^2 \leq C_1 M_n^{-2s}$ and $\frac{1}{n}\sum_{|\ell|\leq M_n}|\theta_\ell|^2 \leq C_2 n^{-1}$. Now, given that g satisfies Assumption (1), it follows that there exists a positive constants C_3 not depending on f and n such that for all sufficiently large n, $\frac{1}{n}\sum_{|\ell|\leq M_n}\frac{|\theta_\ell|^2+\epsilon^2}{|\gamma_\ell|^2} \leq C_3 n^{-1}M_n^{2\nu+1}$. Hence the result immediately follows from the choice $M_n \sim n^{\frac{1}{2s+2\nu+1}}$, which completes the proof. □

For the proof of Propositions 3 and 4, let us remark that that $\hat{\beta}_{j,k} - \beta_{j,k} = Z_1 + Z_2$ with

$$Z_1 = \sum_{\ell\in\Omega_j}\psi_\ell^{j,k}\theta_\ell\left(\frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right) \quad \text{and} \quad Z_2 = \epsilon\sum_{\ell\in\Omega_j}\frac{\psi_\ell^{j,k}}{\gamma_\ell}\eta_\ell. \tag{A.1}$$

Under Assumption 2, $G(x) = \sum_{m\in\mathbb{Z}} g(x+m)$, exists for all $x \in [0,1]$ and is a bounded density. Throughout the proof we use the following lemma whose proof is straightforward:

LEMMA 1. *Let $h \in L^2([0,1])$ be a 1-periodic function on $\mathbb{R}$. Then, $\int_{\mathbb{R}} h(x)g(x)dx = \int_0^1 h(x)G(x)dx$.*

Proof of Proposition 3: first note that since $|\psi_\ell^{j,k}| \leq 2^{-j/2}$ and $\Omega_j \subset [-2^{j+2}c_0, -2^j c_0] \cup [2^j c_0, 2^{j+2}c_0]$, see [25], it follows that $\#\{\Omega_j\} \leq 4\pi 2^j$ and that under Assumption 1, $|\gamma_\ell|^{-2} \sim 2^{2j\nu}$ for all $\ell \in \Omega_j$. This implies that there exists a constant $C > 0$ such that

$$\sum_{\ell \in \Omega_j} |\frac{\psi_\ell^{j,k}}{\gamma_\ell}|^2 \leq C 2^{2j\nu} \quad \text{and} \quad \sum_{\ell \in \Omega_j} |\frac{\psi_\ell^{j,k}}{\gamma_\ell}| \leq C 2^{j(\nu+1/2)}. \tag{A.2}$$

Then, we need the following lemma which shows that the Fourier coefficients $\theta_\ell = \int_0^1 e^{-2i\ell\pi x} f(x)dx$ are uniformly bounded for all $f \in B_{p,q}^s(A)$.

LEMMA 2. *Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $s - 1/p + 1/2 > 0$ and $A > 0$. Then, there exists a constant $A' > 0$ such that for all $f \in B_{p,q}^s(A)$ and all $\ell \in \mathbb{Z}$, $|\theta_\ell| \leq A'$.*

Proof : since $|\phi_\ell^{j_0,k}| \leq 2^{-j_0/2}$ and $|\psi_\ell^{j,k}| \leq 2^{-j/2}$ one can remark using Cauchy-Schwarz inequality that for any $j_0 \geq 0$

$$\begin{aligned}
|\theta_\ell| &\leq \sum_{k=0}^{2^{j_0}-1} |c_{j_0,k}||\phi_\ell^{j_0,k}| + \sum_{j=j_0}^{+\infty} \sum_{k=0}^{2^j-1} |\beta_{j,k}||\psi_\ell^{j,k}| \\
&\leq \left(\sum_{k=0}^{2^{j_0}-1} |c_{j_0,k}|^2 \right)^{1/2} + \sum_{j=j_0}^{+\infty} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^2 \right)^{1/2}
\end{aligned}$$

Now using the inequality $\left(\sum_{r=1}^m |a_r|^2\right)^{1/2} \leq m^{(1/2-1/p)_+} \left(\sum_{r=1}^m |a_r|^p\right)^{1/p}$ for ℓ_p-norm in $\mathbb{R}^m$ it follows that $|\theta_\ell| \leq 2^{j_0(1/2-1/p)_+} \left(\sum_{k=0}^{2^{j_0}-1} |c_{j_0,k}|^p\right)^{1/p} + \sum_{j=j_0}^{+\infty} 2^{j(1/2-1/p)_+} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p\right)^{1/p}$.

Since $f \in B_{p,q}^s(A)$, one has that $\left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p\right)^{1/p} \leq A2^{-j(s+1/2-1/p)}$ and $\left(\sum_{k=0}^{2^{j_0}-1} |c_{j_0,k}|^p\right)^{1/p} \leq A$ which implies that $|\theta_\ell| \leq A2^{j_0(1/2-1/p)_+} + A\sum_{j=j_0}^{+\infty} 2^{-j(s+1/2-1/p-(1/2-1/p)_+)}$. Taking for instance $j_0 = 0$ completes the proof since by assumption $s + 1/2 - 1/p > 0$. □

Control of $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^2$ (the proof to control $\mathbb{E}|\hat{c}_{j_0,k} - c_{j_0,k}|^2$ follows from the same arguments): from the decomposition (A.1) it follows that $\mathbb{E}|\hat{\beta}_{j,k} -$

$\beta_{j,k}|^2 \leq 2\mathbb{E}|Z_1|^2 + 2\mathbb{E}|Z_2|^2$. Since η_ℓ are iid $N_{\mathbb{C}}(0, 1/n)$, the bound (A.2) implies that

$$\mathbb{E}|Z_2|^2 \leq C\frac{2^{2j\nu}}{n}. \tag{A.3}$$

Then, let us write $Z_1 = \frac{1}{n}\sum_{m=1}^{n}(W_m - \mathbb{E}W_m)$ with $W_m = h_{j,k}(\tau_m)$ and $h_{j,k}(\tau) = \sum_{\ell \in \Omega_j} \frac{\psi_\ell^{j,k}\theta_\ell}{\gamma_\ell} e^{-i2\pi\ell\tau}$ for $\tau \in \mathbb{R}$. By independence of the τ_m's, one has that $\mathbb{E}|Z_1|^2 \leq \frac{1}{n}\mathbb{E}|W_1|^2$. Applying Lemma 1 with $h = h_{j,k}$ and since the density G is bounded, it follows that

$$\begin{aligned} \mathbb{E}|W_1|^2 &= \int_{\mathbb{R}} |h_{j,k}(\tau)|^2 g(\tau) d\tau = \int_0^1 |h_{j,k}(\tau)|^2 G(\tau) d\tau \\ &\leq C\int_0^1 |h_{j,k}(\tau)|^2 d\tau \leq C \sum_{\ell \in \Omega_j} \frac{|\psi_\ell^{j,k}|^2 |\theta_\ell|^2}{|\gamma_\ell|^2} \end{aligned} \tag{A.4}$$

where the last inequality follows from Parseval's relation. Then, using the bound (A.2) and Lemma 2, inequality (A.4) implies that there exists a constant C such that for all $f \in B^s_{p,q}(A)$

$$\mathbb{E}Z_1^2 \leq C\frac{1}{n}\sum_{\ell \in \Omega_j} \frac{|\psi_\ell^{j,k}|^2}{|\gamma_\ell|^2} \leq C\frac{2^{2j\nu}}{n}, \tag{A.5}$$

Hence using the bounds (A.3) and (A.5), it follows that there exists a constant C such that for all $f \in B^s_{p,q}(A)$, $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^2 \leq C\frac{2^{2j\nu}}{n}$.

Control of $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^4$: from the decomposition (A.1) it follows that $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^4 \leq C(\mathbb{E}|Z_1|^4 + \mathbb{E}|Z_2|^4)$. As Z_2 is a centered Gaussian variable with variance $\frac{1}{n}\epsilon^2 \sum_{\ell \in \Omega_j} |\frac{\psi_\ell^{j,k}}{\gamma_\ell}|^2 \leq C\frac{2^{2j\nu}}{n}$, one has that

$$\mathbb{E}|Z_2|^4 \leq C\frac{2^{j4\nu}}{n^2}. \tag{A.6}$$

Then, remark that $Z_1 = \frac{1}{n}\sum_{m=1}^{n} Y_m$ with $Y_m = \sum_{\ell \in \Omega_j} \frac{\psi_\ell^{j,k}\theta_\ell}{\gamma_\ell}(e^{-i2\pi\ell\tau_m} - \gamma_\ell)$, and recall the so-called Rosenthal's inequality for moment bounds of iid variables [43]: if $X_1, \ldots, X_n$ are iid random variables such that $\mathbb{E}X_j = 0$, $\mathbb{E}X_j^2 \leqslant \sigma^2$, there exists a positive constant C such that $\mathbb{E}|\sum_{j=1}^{n} X_j/n|^4 \leqslant C(\sigma^4/n^2 + \mathbb{E}|X_1|^4/n^3)$.

Now remark that $\mathbb{E}Y_m = 0$, and arguing as previously for the control of $\mathbb{E}|W_1|^2$, see equation (A.4), it follows that $\mathbb{E}|Y_m|^2 \leq C2^{2j\nu}$ where C is constant not depending on f. Then, remark that

$$\mathbb{E}|Y_1|^4 \leq C\left(\int_{\mathbb{R}} |h_{j,k}(\tau)|^4 g(\tau)d\tau + |\beta_{jk}|^4\right) \text{ with } h_{j,k}(\tau) = \sum_{\ell\in\Omega_j} \frac{\psi_\ell^{j,k}\theta_\ell}{\gamma_\ell} e^{-i2\pi\ell\tau},$$

and that

$$\int_{\mathbb{R}} |h_{j,k}(\tau)|^4 g(\tau)d\tau \leq \sup_{\tau\in\mathbb{R}}\{|h_{j,k}(\tau)|^2\} \int_{\mathbb{R}} |h_{j,k}(\tau)|^2 g(\tau)d\tau$$

Note that using (A.2) and Lemma 2, it follows that $|h_{j,k}(\tau)| \leq \sum_{\ell\in\Omega_j} \frac{|\psi_\ell^{j,k}||\theta_\ell|}{|\gamma_\ell|}$ $\leq C\sum_{\ell\in\Omega_j} \frac{|\psi_\ell^{j,k}|}{|\gamma_\ell|} \leq C2^{j(\nu+1/2)}$ uniformly for $f \in B^s_{p,q}(A)$. Then, arguing as for the control of $\mathbb{E}|W_1|^2$, see equation (A.4), one has that $\int_{\mathbb{R}} |h_{j,k}(\tau)|^2 g(\tau)d\tau \leq C2^{2j\nu}$, which finally implies that $\mathbb{E}|Y_1|^4 \leq C2^{j(4\nu+1)}$, since $|\beta_{jk}| \leq C$ uniformly for $f \in B^s_{p,q}(A)$. Then, using Rosenthal's inequality, it follows that there exists a constant C such that for all $f \in B^s_{p,q}(A)$

$$\mathbb{E}|Z_1|^4 \leq C(\frac{2^{j4\nu}}{n^2} + \frac{2^{j(4\nu+1)}}{n^3}), \tag{A.7}$$

which completes the proof for the control of $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^4$ using (A.6) and (A.7). $\square$

Proof of Proposition 4: let $u > 0$, and remark that from the decomposition (A.1) it follows

$$\mathbb{P}\left(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq u\right) \leq \mathbb{P}(|Z_1| \geq u/2) + \mathbb{P}(|Z_2| \geq u/2)$$

Recall that the η_ℓ's are iid $N_{\mathbb{C}}(0, 1/n)$. Hence, Z_2 is a centered Gaussian variable with variance $\frac{1}{n}\epsilon^2 \sum_{\ell\in\Omega_j} |\frac{\psi_\ell^{j,k}}{\gamma_\ell}|^2 \leq \frac{1}{n}\sigma_j^2$, with $\sigma_j^2 = 2^{-j}\epsilon^2 \sum_{\ell\in\Omega_j} |\gamma_\ell|^{-2}$, which implies that (see e.g. [34]) for any $t > 0$

$$\mathbb{P}(|Z_2| \geq \sigma_j\sqrt{\frac{2t}{n}}) \leq 2\exp(-t) \tag{A.8}$$

By definition, $\tilde{\gamma}_\ell = \frac{1}{n}\sum_{m=1}^n e^{-i2\pi\ell\tau_m}$, and thus $Z_1 = \frac{1}{n}\sum_{m=1}^n (W_m - \mathbb{E}W_m)$ with $W_m = \sum_{\ell\in\Omega_j} \frac{\psi_\ell^{j,k}\theta_\ell}{\gamma_\ell} e^{-i2\pi\ell\tau_m}$. Remark that W_m are random variables bounded by $\delta_j = 2^{-j/2}\sum_{\ell\in\Omega_j} \frac{|\theta_\ell|}{|\gamma_\ell|}$. Moreover, using Lemma 1 with $h = h_{j,k}(\tau) = \sum_{\ell\in\Omega_j} \frac{\psi_\ell^{j,k}\theta_\ell}{\gamma_\ell} e^{-i2\pi\ell\tau}$ for $\tau \in \mathbb{R}$ it follows that

$$\mathbb{E}|W_1|^2 = \int_{\mathbb{R}} |h_{j,k}(\tau)|^2 g(\tau)d\tau \leq \|G\|_\infty \sum_{\ell\in C_j} \frac{|\psi_\ell^{j,k}|^2|\theta_\ell|^2}{|\gamma_\ell|^2} \leq V_j^2,$$

where $V_j^2 = \|g\|_\infty 2^{-j} \sum_{\ell \in C_j} \frac{|\theta_\ell|^2}{|\gamma_\ell|^2}$, since $|\psi_\ell^{j,k}| \leq 2^{-j/2}$ and $\|g\|_\infty = \|G\|_\infty$. Hence, from Bernstein's inequality it follows that for any $t > 0$ (see e.g Proposition 2.9 in [34])

$$\mathbb{P}\left(|Z_1| \geq \sqrt{\frac{2V_j^2 t}{n}} + \delta_j \frac{t}{3n} \right) \leq 2\exp(-t). \tag{A.9}$$

Taking $u = 2\max\left(\sigma_j \sqrt{\frac{2t}{n}},\ \sqrt{\frac{2V_j^2 t}{n}} + \delta_j \frac{t}{3n} \right)$ for $t > 0$ concludes the proof of Proposition 4. $\square$

A.1. Proof of Theorem 3. Let us fix a resolution $j \geq 0$ whose choice will be discussed later on, and consider for any $\eta = (\eta_i)_{i \in \{0 \ldots 2^j - 1\}} \in \{\pm 1\}^{2^j}$ the function $f_{j,\eta}$ defined as $f_{j,\eta} = \gamma_j \sum_{i=0}^{2^j - 1} \eta_k \psi_{j,k}$, where $\gamma_j = c 2^{-j(s+1/2)}$, and c is a positive constant satisfying $c \leq A$ which implies that $f_{j,\eta} \in B_{p,q}^s(A)$. For some $0 \leq i \leq 2^j - 1$ and $\eta \in \{\pm 1\}^{2^j}$, define also the vector $\eta^i \in \{\pm 1\}^{2^j}$ with components equal to those of η except the i^{th} one.

Let $\psi_{j,k} \star g(x) = \int_{\mathbb{R}} \psi_{j,k}(x-u) g(u) du$. By Parseval's relation, one has that $\|\psi_{j,k} \star g\|^2 = \sum_{\ell \in \Omega_j} |\psi_\ell^{j,k}|^2 |\gamma_\ell|^2$. Hence, under Assumption 1 of a polynomial decay for γ_ℓ and using the fact that $|\psi^{j,k}| \leq 2^{-j/2}$ for Meyer wavelets (see [25]) it follows that there exists a constant C such that $\|\psi_{j,k} \star g\|^2 \leq C 2^{-2j\nu}$.

A.1.1. *Algebraic settings.* We set the resolution $j = j(n)$ to be the largest integer satisfying $2^{j(n)} \leq n^{\frac{1}{2s+2\nu+1}}$. However, to simplify the presentation, the dependency of j on n is dropped in what follows. The definition of $f_{j,\eta}$, γ_j and the bound $\|\psi_{j,k} \star g\|^2 \leq C 2^{-2j\nu}$ thus imply that

$$\begin{aligned}
\gamma_j &= \mathcal{O}(n^{-\frac{s+1/2}{2s+2\nu+1}}) \text{ and } \|f_{j,\eta}\|^2 = \mathcal{O}(n^{-\frac{2s}{2s+2\nu+1}}), \\
\|f_{j,\eta} \star g\|^2 &= \|\gamma_j \sum_k \eta_k (\psi_{j,k} \star g)\|^2 = \mathcal{O}\left(n^{\frac{-2s-2\nu}{2s+2\nu+1}} \right), \\
\|(f_{j,\eta} - f_{j,\eta^i}) \star g\|^2 &= \|2\gamma_j \eta_i (\psi_{j,i} \star g)\|^2 = \mathcal{O}\left(\gamma_j^2 2^{-2j\nu} \right) = \mathcal{O}\left(n^{-1} \right).
\end{aligned}$$

From the above equations, we can thus conclude that $n\|(f_{j,\eta} - f_{j,\eta^i}) \star g\|^2 = \mathcal{O}(1)$, but note that the term $n\|f_{j,\eta} \star g\|^2$ does not converge to 0. At last, observe that by assumption $s > 2\nu + 1$ which implies that $n\|f_{j,\eta} \star g\|^3 \to 0$, $n\|(f_{j,\eta} - f_{j,\eta^i}) \star g\| \|f_{j,\eta}\| \|f_{j,\eta} \star g\| \to 0$ and $n\|f_{j,\eta}\|^3 \to 0$.

A.1.2. *Likelihood ratio.* Let $F(Y)$ be real valued and bounded measurable function of the n trajectories $Y = (Y_1, \dots, Y_n)$. Because of the independence of the τ_i's and the W_i's, we have that

$$\mathbb{E}_f\left[F(Y)\right] = \int_{\mathbb{R}^n} \mathbb{E}_{f,W}\left[F(Y)|\tau_1 = t_1, \dots, \tau_n = t_n\right] g(t_1)dt_1 \dots g(t_n)dt_n,$$

where $\mathbb{E}_f$ denotes the expectation with respect to the law of $Y = (Y_1, \dots, Y_n)$ when f is the true hypothesis, and $\mathbb{E}_{f,W}$ is used to denote expectation only with respect to law of the Brownian motions $W_1, \dots, W_n$ where the shifts are fixed and f is the true hypothesis. Now using the classical Girsanov formula it follows that for any function $h \in L^2([0,1])$

$$\begin{aligned}\mathbb{E}_f\left[F(Y)\right] &= \int_{\mathbb{R}^n} \mathbb{E}_{h,W}\left[F(Y)|\tau_1 = t_1, \dots, \tau_n = t_n\right] \Lambda_n(f,h) g(t_1)dt_1 \dots g(t_n)dt_n \\ &= \mathbb{E}_h\left[F(Y)\Lambda_n(f,h)\right]\end{aligned}$$

where $\Lambda_n(f,h)$ is the following likelihood ratio

$$\Lambda_n(f,h) = \prod_{i=1}^n \exp\left(\int_0^1 (f(x-\tau_i) - h(x-\tau_i))dY_i(x) + \frac{1}{2}\|h\|^2 - \frac{1}{2}\|f\|^2\right).$$

In what follows, f_0 is used to denote the hypothesis $f \equiv 0$.

A.1.3. *Technical Lemmas.* Given n arbitrary trajectories $Y_1, \dots Y_n$ from model (1.4), we define $\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta^i}, f_0)\right)$ as the expectation of the likelihood ratio with respect to the law of the random shifts, namely

$$\begin{aligned}\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta^i}, f_0)\right) &= \int_{\mathbb{R}^n} \prod_{i=1}^n e^{\int_0^1 (f(x-\tau_i) - h(x-\tau_i))dY_i(x) + \frac{1}{2}\|h\|^2 - \frac{1}{2}\|f\|^2} \\ &\quad g(\tau_1)\dots g(\tau_n)d\tau_1 \dots d\tau_n.\end{aligned}$$

LEMMA 3. *Suppose for some constants $\lambda > 0$ and $\pi_0 > 0$ and all sufficiently large n we have that*

$$\mathbb{P}_{f_{j,\eta}}\left(\frac{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta^i}, f_0)\right)}{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right)} \geq e^{-\lambda}\right) \geq \pi_0, \tag{A.10}$$

for all $f_{j,\eta}$ and all $i \in \{0 \dots 2^j - 1\}$. Then, there exists a positive constant C, such that for all sufficiently large n and any estimator $\hat{f}_n$

$$\max_{\eta \in \{\pm 1\}^{2^j}} \mathbb{E}_{f_{j,\eta}}\|\hat{f}_n - f_{j,\eta}\|^2 \geq C\pi_0 e^{-\lambda} 2^j \gamma_j^2$$

Proof of Lemma 3 : our proof is inspired by the proof of Lemma 2.10 in [21]. For this let $I_{jk} = [\frac{k}{2^j}, \frac{k+1}{2^j}]$ and arguing as in [21] it follows that for any estimator $\hat{f}_n$

$$\begin{aligned}\max_{\eta \in \{\pm 1\}^{2^j}} \mathbb{E}_{f_{j,\eta}} \|\hat{f}_n - f_{j,\eta}\|^2 \quad &\geq \quad 2^{-2^j} \sum_{k=0}^{2^j-1} \sum_{\eta | \eta_k = 1} \mathbb{E}_{f_{j,\eta}} \left[\int_{I_{j,k}} |\hat{f}_n - f_{j,\eta}|^2 \right. \\ &\left. + \Lambda_n(f_{j,\eta^k}, f_{j,\eta}) \int_{I_{j,k}} |\hat{f}_n - f_{j,\eta^k}|^2 \right].\end{aligned}$$

Let $Z(Y) = \left[\int_{I_{j,k}} |\hat{f}_n - f_{j,\eta}|^2 + \Lambda_n(f_{j,\eta^k}, f_{j,\eta}) \int_{I_{j,k}} |\hat{f}_n - f_{j,\eta^k}|^2 \right]$ and remark that

$$\begin{aligned}\mathbb{E}_{f_{j,\eta}}[Z(Y)] \quad &= \quad \mathbb{E}_{f_0,W} \int_{\mathbb{R}^n} \left[\Lambda_n(f_{j,\eta}, f_0) \int_{I_{j,k}} |\hat{f}_n - f_{j,\eta}|^2 \right. \\ &\left. + \Lambda_n(f_{j,\eta^k}, f_0) \int_{I_{j,k}} |\hat{f}_n - f_{j,\eta^k}|^2 \right] g(\tau_1) d\tau_1 \dots g(\tau_n) d\tau_n.\end{aligned}$$

Now, since under the hypothesis f_0 the trajectories $Y_1, \dots, Y_n$ do not depend on the random shifts $\tau_1, \dots, \tau_n$ it follows that $\hat{f}_n$ does not depend on the shifts $\tau_1, \dots, \tau_n$ as it is by definition a measurable function with respect to the sigma algebra generated by $Y_1, \dots, Y_n$. This implies that for any $\delta > 0$

$$\begin{aligned}\mathbb{E}_{f_{j,\eta}}[Z(Y)] \quad &= \quad \mathbb{E}_{f_0,W} \left[\mathbb{E}_\tau \left(\Lambda_n(f_{j,\eta}, f_0) \right) \int_{I_{j,k}} |\hat{f}_n - f_{j,\eta}|^2 \right. \\ &\left. + \mathbb{E}_\tau \left(\Lambda_n(f_{j,\eta^k}, f_0) \right) \int_{I_{j,k}} |\hat{f}_n - f_{j,\eta^k}|^2 \right] \\ &\geq \quad \mathbb{E}_{f_0,W} \left[\mathbb{E}_\tau \left(\Lambda_n(f_{j,\eta}, f_0) \right) \delta^2 \mathbb{1}_{\left\{ \int_{I_{j,k}} |\hat{f}_n - f_{j,\eta}|^2 > \delta^2 \right\}} \right. \\ &\left. + \mathbb{E}_\tau \left(\Lambda_n(f_{j,\eta^k}, f_0) \right) \delta^2 \mathbb{1}_{\left\{ \int_{I_{j,k}} |\hat{f}_n - f_{j,\eta^k}|^2 > \delta^2 \right\}} \right].\end{aligned}$$

Now, remark that

$$\begin{aligned}\left(\int_{I_{j,k}} |\hat{f}_n - f_{j,\eta}|^2 \right)^{1/2} + \left(\int_{I_{j,k}} |\hat{f}_n - f_{j,\eta^k}|^2 \right)^{1/2} \quad &\geq \quad \left(\int_{I_{j,k}} |f_{j,\eta} - f_{j,\eta^k}|^2 \right)^{1/2} \\ &\geq \quad 2\gamma_j \left(\int_{I_{j,k}} |\psi_{jk}|^2 \right)^{1/2},\end{aligned}$$

and let us argue as in the proof of Lemma 2 in [46] to a find a lower bound for $\int_{I_{j,k}} |\psi_{jk}|^2$. By definition, see Section 3, $\psi_{j,k}(x) = 2^{j/2} \sum_{i\in\mathbb{Z}} \psi^*(2^j(x+i)-k)$ where ψ^* is the Meyer wavelet over $\mathbb{R}$ used to construct ψ. A change of variable shows that $\int_{I_{j,k}} |\psi_{jk}(x)|^2 dx = \int_0^1 \left|\sum_{i\in\mathbb{Z}} \psi^*(x+2^j i)\right|^2 dx$ which implies that $\int_{I_{j,k}} |\psi_{jk}(x)|^2 dx \geq \int_0^1 |\psi^*(x)|^2 dx - \sum_{i\in\mathbb{Z}^*} \int_0^1 \left|\psi^*(x+2^j i)\right|^2 dx$. Now as ψ^* has a fast decay, it follows that there exists a constant $A > 0$ such that $|\psi^*(x)| \leq \frac{A}{1+x^2}$. Thus, $\int_{I_{j,k}} |\psi_{jk}(x)|^2 dx \geq \int_0^1 |\psi^*(x)|^2 dx - A^2 2^{-2j} \sum_{i\in\mathbb{Z}^*} i^{-2}$. Hence, it follows that there exists a constant $\rho > 0$ such that $\left(\int_{I_{j,k}} |\psi_{jk}|^2\right)^{1/2} \geq \rho$ for any k, and all j sufficiently large.

Hence if one takes $\delta = 2\rho\gamma_j$ it follows that

$$\mathbb{1}_{\left\{\int_{I_{j,k}} |\hat{f}_n - f_{j,\eta}|^2 > \delta^2\right\}} \geq \mathbb{1}_{\left\{\int_{I_{j,k}} |\hat{f}_n - f_{j,\eta^k}|^2 \leq \delta^2\right\}},$$

which yields

$$\begin{aligned}
\mathbb{E}_{f_{j,\eta}}[Z(Y)] &\geq \delta^2 \mathbb{E}_{f_0,W}\left[\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right) \min\left(1, \frac{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta^k}, f_0)\right)}{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right)}\right)\right] \\
&= \delta^2 \mathbb{E}_{f_0}\left[\Lambda_n(f_{j,\eta}, f_0) \min\left(1, \frac{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta^k}, f_0)\right)}{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right)}\right)\right] \\
&= \delta^2 \mathbb{E}_{f_{j,\eta}}\left[\min\left(1, \frac{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta^k}, f_0)\right)}{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right)}\right)\right],
\end{aligned}$$

and arguing as in the proof of Lemma 2.10 in [21] completes the proof. □

Now remark that under the hypothesis $f = f_{j,\eta}$, then each Y_i is given by $dY_i(x) = f_{j,\eta}(x-\alpha_i)dx + dW_i(x)$ where each α_i is the true random shift of the i^{th} trajectory. Thus, under this hypothesis, we obtain $\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right) = \prod_{i=1}^n \int_{\mathbb{R}} g(\tau_i) e^{\left[\int_0^1 f_{j,\eta}(x-\tau_i) f_{j,\eta}(x-\alpha_i)dx + f_{j,\eta}(x-\tau_i)dW_i(x) - \frac{1}{2}\|f_{j,\eta}\|^2\right]} d\tau_i$, and
$\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta^i}, f_0)\right) = \prod_{i=1}^n \int_{\mathbb{R}} g(\tau_i) e^{\left[\int_0^1 f_{j,\eta^i}(x-\tau_i) f_{j,\eta}(x-\alpha_i)dx + f_{j,\eta^i}(x-\tau_i)dW_i(x) - \frac{1}{2}\|f_{j,\eta^i}\|^2\right]} d\tau_i$.
Using the two expressions above, we now study the condition (A.10).

LEMMA 4. *Following the choices of $j(n)$ and $\gamma_{j(n)}$ given in our algebraic setting, there exists $\lambda > 0$ and $\pi_0 > 0$ such that for all sufficiently large n*

$$\mathbb{P}_{f_{j,\eta}}\left(\frac{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta^i}, f_0)\right)}{\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right)} \geq e^{-\lambda}\right) \geq \pi_0.$$

Proof of Lemma 4 : to obtain the required bound, we use several second order Taylor expansions. From the Cauchy-Schwarz inequality, we have

$$e^{\int_0^1 f_{j,\eta}(x-\tau_i) f_{j,\eta}(x-\alpha_i)dx} = 1 + \int_0^1 f_{j,\eta}(x-\tau_i) f_{j,\eta}(x-\alpha_i)dx + \mathcal{O}_p(\|f_{j,\eta}\|^4)$$

A similar argument yields $\mathbb{E}\left[\left|\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right|\right] \leq \|f_{j,\eta}\|$, and the Markov inequality used with a second order expansion implies $e^{\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)} = 1 + \int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x) + \frac{1}{2}\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 + \mathcal{O}_p(\|f_{j,\eta}\|^3)$. Looking now at the complete expression of $\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right)$, we obtain $\mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right) = \prod_{i=1}^n \int_{\mathbb{R}} g(\tau_i)\left[1 + \int_0^1 f_{j,\eta}(x-\tau_i) f_{j,\eta}(x-\alpha_i)dx + \int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right.$ $\left. + \frac{1}{2}\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 - \frac{1}{2}\|f_{j,\eta}\|^2 + \mathcal{O}_p(\|f_{j,\eta}\|^3)\right]$. The Fubini-type theorem for stochastic integrals (see for instance [23], chapter 3, lemma 4.1) enables to write $\log \mathbb{E}_\tau\left(\Lambda_n(f_{j,\eta}, f_0)\right) = \sum_{i=1}^n \log\left[1 + \int_0^1 (f_{j,\eta} \star g)(x) f_{j,\eta}(x-\alpha_i)dx\right.$ $+ \int_0^1 (f_{j,\eta} \star g)(x)dW_i(x)\frac{1}{2}\int_{\mathbb{R}}\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 g(\tau_i)d\tau_i - \frac{1}{2}\|f_{j,\eta}\|^2$ $\left. + \mathcal{O}_p(\|f_{j,\eta}\|^3)\right]$

Then, applying a classical expansion of the logarithm $\log(1+z) = z - \frac{z^2}{2} + \mathcal{O}(z^3)$, we obtain

$$\log \mathbb{E}_\tau \Lambda_n(f_{j,\eta}, f_0) = z - \frac{z^2}{2} + \mathcal{O}_p(z^3)$$

$$= \sum_{i=1}^n \int_0^1 (f_{j,\eta} \star g)(x) f_{j,\eta}(x-\alpha_i)dx + \int_0^1 (f_{j,\eta} \star g)(x)dW_i(x) \tag{A.11}$$

$$+ \frac{1}{2}\sum_{i=1}^n \int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 - \frac{n}{2}\|f_{j,\eta}\|^2 \tag{A.12}$$

$$- \frac{1}{2}\sum_{i=1}^n \left(\int_0^1 (f_{j,\eta} \star g)(x) f_{j,\eta}(x-\alpha_i)dx + \int_0^1 (f_{j,\eta} \star g)(x)dW_i(x)\right. \tag{A.13}$$

$$\left. + \frac{1}{2}\int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2\right)^2 + \mathcal{O}_p(n\|f_{j,\eta}\|^3). \tag{A.14}$$

We first discuss on the size of the terms in equations (A.13) and (A.14). The first term in (A.13) can be bounded using the Cauchy-Schwarz inequality

$$\sum_{i=1}^{n}\left(\int_0^1 (f_{j,\eta}\star g)(x)f_{j,\eta}(x-\alpha_i)dx\right)^2 \leq n\|f_{j,\eta}\star g\|^2\|f_{j,\eta}\|^2 = \mathcal{O}_p(n\|f_{j,\eta}\|^4).$$

But observe that $\sum_{i=1}^{n}\mathbb{E}_{W_i}\left(\int_0^1 (f_{j,\eta}\star g)(x)dW_i(x)\right)^2 = n\|f_{j,\eta}\star g\|^2$ which does not converge to 0. Then, the Jensen inequality implies

$$\begin{aligned}&\textstyle\sum_{i=1}^{n}\mathbb{E}_{W_i}\left(\int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 d\tau_i\right)^2\\ &\leq \textstyle\sum_{i=1}^{n}\int_{\mathbb{R}} g(\tau_i)\mathbb{E}_{W_i}\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^4 d\tau_i = \mathcal{O}_p(n\|f_{j,\eta}\|^4).\end{aligned}$$

Let us now study the terms derived from double products in equations (A.13) and (A.14), use first that $2|ab|\leq(a^2+b^2)$ to get $\sum_{i=1}^{n}\mathbb{E}_{\alpha_i,W_i}\left|\int_0^1 (f_{j,\eta}\star g)(x) f_{j,\eta}(x-\alpha_i)dx\right|\left|\int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 d\tau_i\right| = \mathcal{O}_p(n\|f_{j,\eta}\|^4)$.

The Cauchy-Schwarz and Jensen inequalities imply

$$\begin{aligned}&\textstyle\sum_{i=1}^{n}\mathbb{E}_{W_i}\left|\int_0^1 (f_{j,\eta}\star g)(x)dW_i(x)\right|\left|\int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 d\tau_i\right|\\ &= \mathcal{O}_p(n\|f_{j,\eta}\|^3).\end{aligned}$$

At last, the Cauchy-Schwarz and Jensen inequalities on the remaining double-product term imply also

$$\begin{aligned}&\mathbb{E}_{W,\alpha}\left|\textstyle\sum_{i=1}^{n}\int_0^1 (f_{j,\eta}\star g)(x)f_{j,\eta}(x-\alpha_i)dx\int_0^1 (f_{j,\eta}\star g)(x)dW_i(x)\right|\\ &= \mathcal{O}_p(n\|f_{j,\eta}\|^3).\end{aligned}$$

All the above bounds enables us to write

$$\begin{aligned}&L_1 := \log\Lambda_n(f_{j,\eta},f_0)\\ &= \textstyle\sum_{i=1}^{n}\int_0^1 (f_{j,\eta}\star g)(x)f_{j,\eta}(x-\alpha_i)dx + \int_0^1 (f_{j,\eta}\star g)(x)dW_i(x)\\ &+\tfrac{1}{2}\textstyle\sum_{i=1}^{n}\int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 - \tfrac{n}{2}\|f_{j,\eta}\|^2\\ &-\tfrac{1}{2}\textstyle\sum_{i=1}^{n}\left(\int_0^1 (f_{j,\eta}\star g)(x)dW_i(x)\right)^2 + \mathcal{O}_p(n\|f_{j,\eta}\|^3).\end{aligned}$$

In a similar way, we can also write

$$\begin{aligned}&L_2 := \log\Lambda_n(f_{j,\eta^i},f_0) = \textstyle\sum_{i=1}^{n}\int_0^1 (f_{j,\eta^i}\star g)(x)f_{j,\eta}(x-\alpha_i)dx+\\ &\textstyle\int_0^1 (f_{j,\eta^i}\star g)(x)dW_i(x) + \tfrac{1}{2}\sum_{i=1}^{n}\int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta^i}(x-\tau_i)dW_i(x)\right]^2 - \tfrac{n}{2}\|f_{j,\eta^i}\|^2\\ &-\tfrac{1}{2}\textstyle\sum_{i=1}^{n}\left(\int_0^1 (f_{j,\eta^i}\star g)(x)dW_i(x)\right)^2 + \mathcal{O}_p(n\|f_{j,\eta}\|^3).\end{aligned}$$

For sake of simplicity, let us write $h = f_{j,\eta^i} - f_{j,\eta} = 2\eta_i \psi_{j,i}$. The difference $L = L_2 - L_1$ can thus be decomposed as

$$\begin{aligned}
L &= \sum_{i=1}^{n} \int_0^1 (h \star g)(x)[f_{j,\eta}(x-\alpha_i) - f_{j,\eta} \star g(x)]dx && \text{(A.15)}\\
&+ \sum_{i=1}^{n} \int_0^1 (h \star g)(x)(f_{j,\eta} \star g)(x)dx + \int_0^1 (h \star g)(x)dW_i(x) && \text{(A.16)}\\
&+ \frac{1}{2}\sum_{i=1}^{n} \int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta^i}(x-\tau_i)dW_i(x)\right]^2 - \frac{n}{2}\|f_{j,\eta^i}\|^2 && \text{(A.17)}\\
&- \frac{1}{2}\sum_{i=1}^{n} \int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 + \frac{n}{2}\|f_{j,\eta}\|^2 && \text{(A.18)}\\
&- \frac{1}{2}\left[\sum_{i=1}^{n}\left(\int_0^1 (f_{j,\eta^i} \star g)(x)dW_i(x)\right)^2 - n\|f_{j,\eta^i} \star g\|^2\right] && \text{(A.19)}\\
&- \frac{n}{2}\|f_{j,\eta^i} \star g\|^2 && \text{(A.20)}
\end{aligned}$$

$$\begin{aligned}
&+ \frac{1}{2}\left[\sum_{i=1}^{n}\left(\int_0^1 (f_{j,\eta} \star g)(x)dW_i(x)\right)^2 - n\|f_{j,\eta} \star g\|^2\right] && \text{(A.21)}\\
&+ \frac{n}{2}\|f_{j,\eta} \star g\|^2 && \text{(A.22)}\\
&+ \mathcal{O}_p(n\|f_{j,\eta}\|^3). && \text{(A.23)}
\end{aligned}$$

Bound for (A.15) : we use the classical Bennett's inequality (see e.g [34]) for a sum of independent and bounded variables. Define $S = \sum_{i=1}^{n} \int_0^1 (h \star g)(x)[f_{j,\eta}(x-\alpha_i) - f_{j,\eta} \star g(x)]dx$. From Cauchy-Schwarz inequality, the random variables $\int_0^1 (h \star g)(x) f_{j,\eta}(x-\alpha_i)dx$ are bounded by a constant b such that $b = \|h \star g\|\|f_{j,\eta}\|$. Let v and c to be defined as

$$v = \sum_{i=1}^{n} \mathbb{E}\left[\int_0^1 (h \star g)(x) f_{j,\eta}(x-\alpha_i)dx\right]^2 \text{ and } c = b/3.$$

From the Cauchy Schwarz inequality, we have that $v \leq n\|f_{j,\eta}\|^2\|h \star g\|^2$ and as $h = f_{j,\eta^i} - f_{j,\eta}$, by using our algebraic settings in Section A.1.1, we observe that $v \to 0$. The Bennett's inequality therefore implies that for any $\kappa > 0$:

$$\mathbb{P}(|S| \geq \kappa) \leq 2e^{\dfrac{-\kappa^2}{2(n\|f_{j,\eta}\|^2\|h \star g\|^2 + \kappa\|h \star g\|\|f_{j,\eta}\|/3)}}.$$

From our algebraic settings in Section A.1.1, one has thus that as $n \to \infty$, the $\mathbb{P}(|S| \geq \kappa)$ converges to 0.

Bound for (A.17,A.18,A.19,A.21) : applying Lemma 5 (proved below) to the chi-square statistics in the expressions (A.17,A.18) yields that for any $\kappa > 0$ $\mathbb{P}\left(\left|\frac{1}{2}\sum_{i=1}^n \int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta^i}(x-\tau_i)dW_i(x)\right]^2 - \frac{n}{2}\|f_{j,\eta^i}\|^2\right| \geq \kappa\right) \leq 2e^{\frac{-\kappa^2}{n\|f_{j,\eta^i}\|^4 + 2\kappa\|f_{j,\eta^i}\|^2}}$, and $\mathbb{P}\left(\left|\frac{1}{2}\sum_{i=1}^n \int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f_{j,\eta}(x-\tau_i)dW_i(x)\right]^2 - \frac{n}{2}\|f_{j,\eta}\|^2\right| \geq \kappa\right) \leq 2e^{\frac{-\kappa^2}{n\|f_{j,\eta}\|^4+2\kappa\|f_{j,\eta}\|^2}}$. Similarly we obtain for the chi-square statistics in (A.19,A.21) that for any $\kappa > 0$ $\mathbb{P}\left(\left|\frac{1}{2}\left[\sum_{i=1}^n \left(\int_0^1 (f_{j,\eta}\star g)(x)dW_i(x)\right)^2 - n\|f_{j,\eta}\star g\|^2\right]\right| \geq \kappa\right) \leq 2e^{\frac{-\kappa^2}{n\|f_{j,\eta}\star g\|^4 + 2\kappa\|f_{j,\eta}\star g\|^2}}$, and $\mathbb{P}\left(\left|\frac{1}{2}\left[\sum_{i=1}^n \left(\int_0^1 (f_{j,\eta^i}\star g)(x)dW_i(x)\right)^2 - n\|f_{j,\eta^i}\star g\|^2\right]\right| \geq \kappa\right) \leq 2e^{\frac{-\kappa^2}{n\|f_{j,\eta^i}\star g\|^4 + 2\kappa\|f_{j,\eta^i}\star g\|^2}}$. It follows from the algebraic setting in Section A.1.1 that $n\|f_{j,\eta}\|^4 \to 0$ and $\|f_{j,\eta}\|^2 \to 0$, as well as $n\|f_{j,\eta}\star g\|^4 \to 0$ and $\|f_{j,\eta}\star g\|^2 \to 0$ and the above probabilities converge to zero as $n \to \infty$.

Bound for (A.16,A.20,A.22) : using the first term of (A.16), simple computation shows that yields $\sum_{i=1}^n \int_0^1((f_{j,\eta^i}-f_{j,\eta})\star g)(x)(f_{j,\eta}\star g)(x)dx - \frac{n}{2}\|f_{j,\eta^i}\star g\|^2 + \frac{n}{2}\|f_{j,\eta}\star g\|^2 = -\frac{n}{2}\|h\star g\|^2$, and we obtain from our algebraic settings that this term converges to 0 since $n\|h\star g\|^2 \to 0$. Moreover, the second term of (A.16) is the sum of n i.i.d centered normal variables and the Cirelson-Ibragimov-Sudakov's inequality [10] ensures that

$$\mathbb{P}\left(\left|\sum_{i=1}^n \int_0^1 (h\star g)(x)dW_i(x)\right| \geq \kappa\right) \leq 2e^{\frac{-\kappa^2}{2n\|h\star g\|^2}},$$

and thus the above probability goes to zero.

Bound for (A.23): from our algebraic settings in Section A.1.1, it follows immediately that $n\|f_{j,\eta}\|^3 \to 0$.

Hence, by combining all the above bounds, it follows that we have shown that $L_2 - L1$ is the sum of various terms which all converge to zero in probability or that are larger than some negative constant with probability tending to one as $n \to +\infty$, which completes the proof of Lemma 4. $\square$

LEMMA 5. *Let g be a density function on $\mathbb{R}$, and $(W_i)_{i\in\{1\dots n\}}$ be independent standard Brownian motions on $[0,1]$. Then, for any $f \in L^2([0,1])$ and $\alpha > 0$,*

$$P\left(\frac{1}{2}\sum_{i=1}^{n}\int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f(t-\tau_i)dW_i(t)\right]^2 d\tau_i - \frac{n}{2}\|f\|^2 \geq \alpha\right) \leq e^{\dfrac{-\alpha^2}{n\|f\|^4 + 2\alpha\|f\|^2}}.$$

Proof of Lemma 5 Consider $\zeta_n = \frac{1}{2}\sum_{i=1}^{n}\int_{\mathbb{R}} g(\tau_i)\left[\int_0^1 f(t-\tau_i)dW_i(t)\right]^2 d\tau_i - \frac{n}{2}\|f\|^2$. We use a Laplace transform technique to bound $\mathbb{P}(\zeta_n \geq \alpha)$. For any $\dfrac{1}{\|f\|^2} > t > 0$, we have by Markov's inequality

$$\mathbb{P}(\zeta_n \geq \alpha) \quad \leq \quad e^{-\alpha t - \dfrac{n}{2}\|f\|^2 t}\prod_{i=1}^{n}\mathbb{E}\left[e^{t/2\int_{\mathbb{R}} g(\tau_i)\left(\int_0^1 f(t-\tau_i)dW_i(t)\right)^2 d\tau_i}\right].$$

We apply now Jensen's inequality for the exponential function and the measure $g(\tau)d\tau$ to obtain

$$\mathbb{P}(\zeta_n \geq \alpha) \quad \leq \quad e^{-\alpha t - \dfrac{n}{2}\|f\|^2 t}\prod_{i=1}^{n}\int_{\mathbb{R}} g(\tau_i)\mathbb{E}\left[e^{t/2\left(\int_0^1 f(t-\tau_i)dW_i(t)\right)^2}\right] d\tau_i.$$

Remark that $\left(\int_0^1 f(t-\tau_i)dW_i(t)\right)^2$ follows a chi-square distribution whose Laplace transform does not depend on τ_i and thus

$$\mathbb{P}(\zeta_n \geq \alpha) \quad \leq \quad e^{-\alpha t - \dfrac{n}{2}\|f\|^2 t - \dfrac{n}{2}\log\left(1 - t\|f\|^2\right)}.$$

Let $\tilde{\alpha} = \dfrac{\alpha}{n/2\|f\|^2}$ and minimizing now the last bound with respect to t yield the optimal choice $t^\star = \frac{\tilde{\alpha}}{1+\tilde{\alpha}}$. With this choice, we obtain $\mathbb{P}(\zeta_n \geq \alpha) \leq \exp(\dfrac{n}{2}[\log(1+\tilde{\alpha}) - \tilde{\alpha}])$. Now use the classical bound $\log(1+u) - u \leq \frac{-u^2}{2(1+u)}$, valid for all $u \geq 0$, to get $\mathbb{P}(\zeta_n \geq \alpha) \leq \exp(\frac{n}{2} \times \frac{\tilde{\alpha}^2}{2(1+\tilde{\alpha})}) = \exp(\dfrac{-\alpha^2}{n\|f\|^4 + 2\alpha\|f\|^2})$, which completes the proof of the lemma. $\square$

A.1.4. *A lower bound for the minimax risk.* By Lemma 3 and Lemma 4, it follows that there exists a constant C_1 such that for all sufficiently large n,

$$\inf_{\hat{f}_n} \sup_{f \in B^s_{p,q}(A)} \mathbb{E} \left\| \hat{f}_n - f \right\|^2 \geq \inf_{\hat{f}_n} \max_{\eta \in \{\pm 1\}^{2^j}} \mathbb{E}_{f_{j,\eta}} \|\hat{f}_n - f_{j,\eta}\|^2 \geq C_1 n^{-\frac{2s}{2s+2\nu+1}},$$

which completes the proof of Theorem 3. □

A.2. Proof of Theorem 4. For $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n) \in \mathcal{T}^n$ define the criterion $M(\boldsymbol{\tau}) = \frac{1}{n}\sum_{m=1}^n \sum_{|\ell| \leq \ell_0} |\theta_\ell e^{2i\ell\pi(\tau_m - \tau_m^*)} - \frac{1}{n}\sum_{q=1}^n \theta_\ell e^{2i\ell\pi(\tau_q - \tau_q^*)}|^2$. Then let us first prove the following lemma:

Lemma 6. *Suppose that Assumption 4 hold. Then, the function $\boldsymbol{\tau} \mapsto M(\boldsymbol{\tau})$ has a unique minimum on $\overline{\mathcal{T}}_n$ at $\boldsymbol{\tau} = \tilde{\boldsymbol{\tau}}$ such that $M(\tilde{\boldsymbol{\tau}}) = 0$ given by $\tilde{\boldsymbol{\tau}} = (\tau_1^* - \bar{\tau}_n, \tau_2^* - \bar{\tau}_n, \ldots, \tau_n^* - \bar{\tau}_n)$, where $\bar{\tau}_n = \frac{1}{n}\sum_{m=1}^n \tau_m^*$.*

Proof of Lemma 6: by definition of $M(\boldsymbol{\tau})$ it follows immediately that $M(\tilde{\boldsymbol{\tau}}) = 0$ and thus $\tilde{\boldsymbol{\tau}}$ is a minimum since $M(\boldsymbol{\tau}) \geq 0$ for all $\boldsymbol{\tau} \in \mathcal{T}^n$. Now suppose that there exists $\boldsymbol{\tau} \in \overline{\mathcal{T}}_n$ such that $M(\boldsymbol{\tau}) = 0$. This implies that for all $m = 1, \ldots, n$ and all $-\ell_0 \leq \ell \leq \ell_0$ $|\theta_\ell|^2 |e^{2i\ell\pi(\tau_m - \tau_m^*)} - \frac{1}{n}\sum_{q=1}^n e^{2i\ell\pi(\tau_q - \tau_q^*)}|^2 = 0$. Since by assumption $\theta_1^* \neq 0$ it follows that for $\ell = 1$, $|e^{2i\pi(\tau_m - \tau_m^*)} - \frac{1}{n}\sum_{q=1}^n e^{2i\pi(\tau_q - \tau_q^*)}|^2 = 0$ for all $m = 1, \ldots, n$, which implies that $e^{2i\pi(\tau_m - \tau_m^*)} = e^{2i\pi(\tau_q - \tau_q^*)}$ for all $m, q = 1, \ldots, n$, since $\frac{1}{n}\sum_{q=1}^n e^{2i\pi(\tau_q - \tau_q^*)}$ does not depend on m. This implies that $\tau_m - \tau_m^* = \tau_0 \mod 1$ for $m = 2, \ldots, n$, where $\tau_0 = \tau_1 - \tau_1^*$. By assumption τ_1, τ_1^* belong to $\mathcal{T} = [-\frac{1}{4}, \frac{1}{4}]$ and thus $|\tau_0| \leq \frac{1}{2}$. Hence, $\tau_m = \tau_m^* + \tau_0$ for $m = 1, \ldots, n$. Since $\sum_{m=1}^n \tau_m = 0$ this implies that $\tau_0 = -\frac{1}{n}\sum_{m=1}^n \tau_m^*$ and thus $\tau_m = \tilde{\tau}_m$ for $m = 1, \ldots, n$ which completes the proof. □

Let $F : \mathbb{R}^{n-1} \to \mathbb{R}^n$ given by $F(\tau_2, \ldots, \tau_n) = (-\sum_{m=2}^n \tau_m, \tau_2, \ldots, \tau_n)^t$, and let $\tilde{M} : \mathcal{T}^{n-1} \to \mathbb{R}^+$ defined by $\tilde{M}(\tau_2, \ldots, \tau_n) = M(F(\tau_2, \ldots, \tau_n))$.

Lemma 7. *Let $\nabla^2 \tilde{M}(\tilde{\boldsymbol{\tau}}_{-1})$ denotes the Hessian of $\tilde{M}$ at $\tilde{\boldsymbol{\tau}}_{-1} = (\tilde{\tau}_2, \ldots, \tilde{\tau}_n)$, then $\nabla^2 \tilde{M}(\tilde{\boldsymbol{\tau}}_{-1}) = \left(\frac{2}{n}\sum_{|\ell| \leq \ell_0} |2\pi\ell|^2 |\theta_\ell|^2\right)\left(I_{n-1} + \mathbb{1}_{n-1}^t \mathbb{1}_{n-1}\right)$, where I_{n-1} is the $(n-1) \times (n-1)$ identity matrix and $\mathbb{1}_{n-1} = (1, \ldots, 1)^t$ is the vector of $\mathbb{R}^{n-1}$ with all entries equal to one. Moreover, $\lambda_{\min}(\nabla^2 \tilde{M}(\tilde{\boldsymbol{\tau}}_{-1})) = \frac{2}{n}\sum_{|\ell| \leq \ell_0} |2\pi\ell|^2 |\theta_\ell|^2$, where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of a symmetric matrix A.*

Proof of Lemma 7: first remark that for $\boldsymbol{\tau}_{-1} = (\tau_2, \ldots, \tau_n) \in \mathbb{R}^{n-1}$ then $\nabla^2 \tilde{M}(\boldsymbol{\tau}_{-1}) = \nabla F^t \nabla^2 M(F(\boldsymbol{\tau}_{-1})) \nabla F$ where $\nabla^2 M(F(\boldsymbol{\tau}_{-1}))$ denotes the Hessian of M at $F(\boldsymbol{\tau}_{-1})$ and ∇F is the gradient of F ($n \times (n-1)$ matrix not depending on τ). Now, since for any $\boldsymbol{\tau} \in \mathcal{T}^n$ $M(\boldsymbol{\tau}) = \sum_{|\ell|\leq \ell_0} |\theta_\ell|^2 \left(1 - \left|\frac{1}{n}\sum_{q=1}^n e^{2i\ell\pi(\tau_q - \tau_q^*)}\right|^2\right)$ it follows that for $m = 2, \ldots, n$

$$\frac{\partial}{\partial \tau_m} M(\boldsymbol{\tau}) = -\frac{2}{n^2} \sum_{|\ell|\leq \ell_0} |\theta_\ell|^2 \Re\left[2i\pi\ell e^{2i\ell\pi(\tau_m - \tau_m^*)} \left(\overline{\sum_{q=1}^n e^{2i\ell\pi(\tau_q - \tau_q^*)}}\right)\right]$$

where $\Re[z]$ denotes the real part of a complex number. Hence for $m_1 \neq m_2$

$$\frac{\partial^2}{\partial \tau_{m_2}\partial \tau_{m_1}} M(\boldsymbol{\tau}) = -\frac{2}{n^2} \sum_{|\ell|\leq \ell_0} |2\pi\ell|^2 |\theta_\ell|^2 \Re\left[e^{2i\ell\pi(\tau_{m_1} - \tau_{m_1}^* - \tau_{m_2} + \tau_{m_2}^*)}\right]$$

and for $m_1 = m_2$

$$\frac{\partial^2}{\partial \tau_{m_1}\partial \tau_{m_1}} M(\boldsymbol{\tau}) = -\frac{2}{n^2} \sum_{|\ell|\leq \ell_0} |2\pi\ell|^2 |\theta_\ell|^2 \Re\left[1 - e^{2i\ell\pi(\tau_{m_1} - \tau_{m_1}^*)} \left(\overline{\sum_{q=1}^n e^{2i\ell\pi(\tau_q - \tau_q^*)}}\right)\right].$$

Then, remark that $F(\tilde{\boldsymbol{\tau}}_{-1}) = \tilde{\boldsymbol{\tau}}$. Hence by taking $\tau_m = \tilde{\tau}_m$ for $m = 2, \ldots, n$ in the above formulas, it follows that

$$\nabla^2 M(\tilde{\boldsymbol{\tau}}) = \nabla^2 M(F(\tilde{\boldsymbol{\tau}}_{-1})) = \frac{2}{n} \sum_{|\ell|\leq \ell_0} |2\pi\ell|^2 |\theta_\ell|^2 \left(I_n - \frac{1}{n} \mathbb{1}_n^t \mathbb{1}_n\right), \tag{A.24}$$

where I_n is the $n \times n$ identity matrix and $\mathbb{1}_n = (1, \ldots, 1)^t$ is the vector of $\mathbb{R}^n$ with all entries equal to one. Hence the result follows from (A.24) and the equality $\nabla^2 \tilde{M}(\tilde{\boldsymbol{\tau}}_{-1}) = \nabla F^t \nabla^2 M(F(\tilde{\boldsymbol{\tau}}_{-1})) \nabla F$, and the fact that the eigenvalues of the matrix $A = I_{n-1} + \mathbb{1}_{n-1}^t \mathbb{1}_{n-1}$ are n (of multiplicity 1) and 1 (of multiplicity $n-2$). $\square$

LEMMA 8. *Suppose that Assumption 4 hold. Then, there exists a constant $\kappa(f) > 0$ (depending on the shape function f) such that for all $\boldsymbol{\tau} \in \overline{\mathcal{T}}_n$ $M(\boldsymbol{\tau}) - M(\tilde{\boldsymbol{\tau}}) \geq \kappa(f) \left(\sum_{|\ell|\leq \ell_0} |\theta_\ell|^2\right) \left(\frac{1}{n}\sum_{m=2}^n (\tau_m - \tilde{\tau}_m)^2\right)$*

Proof of Lemma 8: first remark that for any $\boldsymbol{\tau} \in \overline{\mathcal{T}}_n$ then $\tilde{M}(\boldsymbol{\tau}_{-1}) = M(F(\boldsymbol{\tau}))$ where $\boldsymbol{\tau}_{-1} = (\tau_2, \ldots, \tau_n)$. Since $\tilde{\boldsymbol{\tau}}$ is a minimum of $\boldsymbol{\tau} \mapsto M(\boldsymbol{\tau})$, a second order Taylor expansion implies that for all $\boldsymbol{\tau}_{-1}$ in neighborhood $\mathcal{V} \subset \mathcal{T}^{n-1}$ of $\tilde{\boldsymbol{\tau}}_{-1}$

$$\begin{aligned} M(\boldsymbol{\tau}) - M(\tilde{\boldsymbol{\tau}}) &= \tilde{M}(\boldsymbol{\tau}_{-1}) - \tilde{M}(\tilde{\boldsymbol{\tau}}_{-1}) \\ &= (\boldsymbol{\tau}_{-1} - \tilde{\boldsymbol{\tau}}_{-1})^t \nabla^2 \tilde{M}(\tilde{\boldsymbol{\tau}}_{-1})(\boldsymbol{\tau}_{-1} - \tilde{\boldsymbol{\tau}}_{-1}) + o(\|\boldsymbol{\tau}_{-1} - \tilde{\boldsymbol{\tau}}_{-1}\|^2). \end{aligned}$$

Using Lemma 7 and the above equation, it follows that there exists a universal constant $0 < c_1 < 1$ and an open neighborhood $\tilde{\mathcal{V}} \subset \mathcal{V}$ of $\tilde{\boldsymbol{\tau}}$ such that for all $\boldsymbol{\tau} \in \tilde{\mathcal{V}}$

$$M(\boldsymbol{\tau}) - M(\tilde{\boldsymbol{\tau}}) \geq 2c_1 \left(\sum_{|\ell| \leq \ell_0} |2\pi\ell|^2 |\theta_\ell|^2 \right) \left(\frac{1}{n} \sum_{m=2}^{n} (\tau_m - \tilde{\tau}_m)^2 \right).$$

Now remark that under Assumption 4, $M(\boldsymbol{\tau}) > M(\tilde{\boldsymbol{\tau}}) = 0$ for all $\boldsymbol{\tau} \in \overline{\mathcal{T}}_n \backslash \tilde{\mathcal{V}}$ by Lemma 6. Since $M(\boldsymbol{\tau}) = \sum_{|\ell| \leq \ell_0} |\theta_\ell|^2 \left(1 - \left| \frac{1}{n} \sum_{q=1}^{n} e^{2i\ell\pi(\tau_q - \tau_q^*)} \right|^2 \right)$ the compactness of $\overline{\mathcal{T}}_n$ and the continuity of $\tau \mapsto M(\boldsymbol{\tau})$ implies that there exists a constant $0 < c_2(f) < 1$ (depending on $\tilde{\mathcal{V}}$ and thus on f) such that for all $\boldsymbol{\tau} \in \overline{\mathcal{T}}_n \backslash \tilde{\mathcal{V}}$, $M(\boldsymbol{\tau}) \geq \sum_{|\ell| \leq \ell_0} |\theta_\ell|^2 (1 - c_2(f))$. Moreover since $\mathcal{T}$ is a compact set it follows that there exists a universal constant $c_3 > 0$ such that $(\tau_m - \tilde{\tau}_m)^2 \leq c_3$ for all $m = 2, \ldots, n$, which implies that for all $\boldsymbol{\tau} \in \overline{\mathcal{T}}_n$ $\frac{1}{n} \sum_{m=2}^{n} (\tau_m - \tilde{\tau}_m)^2 \leq c_3$. Therefore

$$M(\boldsymbol{\tau}) - M(\tilde{\boldsymbol{\tau}}) \geq \left(c_3^{-1} (1 - c_2(f)) \sum_{|\ell| \leq \ell_0} |\theta_\ell|^2 \right) \left(\frac{1}{n} \sum_{m=2}^{n} (\tau_m - \tilde{\tau}_m)^2 \right)$$

for all $\boldsymbol{\tau} \in \overline{\mathcal{T}}_n \backslash \tilde{\mathcal{V}}$. Then the result follows by taking $\kappa(f) = \min(2c_1, c_3^{-1} (1 - c_2(f)))$ and the fact that $\sum_{|\ell| \leq \ell_0} |2\pi\ell|^2 |\theta_\ell|^2 \geq \sum_{|\ell| \leq \ell_0} |\theta_\ell|^2$. $\square$

Now recall that $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_n) = \arg\min_{\boldsymbol{\tau} \in \overline{\mathcal{T}}_n} M_n(\boldsymbol{\tau})$. Since $\hat{\boldsymbol{\tau}}$ is a minimum of $\boldsymbol{\tau} \mapsto M_n(\boldsymbol{\tau})$ and $\tilde{\boldsymbol{\tau}}$ is a minimum of $\boldsymbol{\tau} \mapsto M(\boldsymbol{\tau})$ it follows that $M(\hat{\boldsymbol{\tau}}) - M(\tilde{\boldsymbol{\tau}}) \leq 2 \sup_{\boldsymbol{\tau} \in \mathcal{T}^n} |M_n(\boldsymbol{\tau}) - M(\boldsymbol{\tau})|$. Therefore Lemma 8 imply that

$$\text{(A.25)} \quad \frac{1}{n} \sum_{m=2}^{n} (\hat{\tau}_m - \tilde{\tau}_m)^2 \leq 2 \left(\kappa(f) \left(\sum_{|\ell| \leq \ell_0} |\theta_\ell|^2 \right) \right)^{-1} \sup_{\boldsymbol{\tau} \in \mathcal{T}^n} |M_n(\boldsymbol{\tau}) - M(\boldsymbol{\tau})|.$$

LEMMA 9. *Let $Z = \sup_{\boldsymbol{\tau} \in \mathcal{T}^n} |M_n(\boldsymbol{\tau}) - M(\boldsymbol{\tau})|$. Then for any $t > 0$*

$$\mathbb{P} \left(Z \leq \left(1 + 2 \left(\sum_{|\ell| \leq \ell_0} |\theta_\ell|^2 \right)^{1/2} \right) \left(\sqrt{C(\epsilon, n, \ell_0, t)} + C(\epsilon, n, \ell_0, t) \right) \right) \geq 1 - \exp(-t).$$

where $C(\epsilon, n, \ell_0, t) = \epsilon^2 (2\ell_0 + 1) + 2\epsilon^2 \sqrt{\frac{2\ell_0 + 1}{n} t} + 2\frac{\epsilon^2}{n} t$

Proof: remark that $M_n(\boldsymbol{\tau})$ can be decomposed as $M_n(\boldsymbol{\tau}) = M(\boldsymbol{\tau}) + L(\boldsymbol{\tau}) + Q(\boldsymbol{\tau})$, where

$$L(\boldsymbol{\tau}) = 2\frac{\epsilon}{n}\sum_{m=1}^{n}\sum_{|\ell|\leq \ell_0} \Re\left[\left(\theta_\ell e^{2i\ell\pi(\tau_m-\tau_m^*)} - \frac{1}{n}\sum_{q=1}^{n}\theta_\ell e^{2i\ell\pi(\tau_q-\tau_q^*)}\right)\left(\overline{z_{m,\ell}e^{2i\ell\pi\tau_m} - \frac{1}{n}\sum_{q=1}^{n} z_{q,\ell}e^{2i\ell\pi\tau_q}}\right)\right]$$

$$Q(\boldsymbol{\tau}) = \frac{\epsilon^2}{n}\sum_{m=1}^{n}\sum_{|\ell|\leq \ell_0}\left|z_{m,\ell}e^{2i\ell\pi\tau_m} - \frac{1}{n}\sum_{q=1}^{n} z_{q,\ell}e^{2i\ell\pi\tau_q}\right|^2$$

By Cauchy-Schwarz inequality $|L(\boldsymbol{\tau})| \leq 2\sqrt{M(\boldsymbol{\tau})}\sqrt{Q(\boldsymbol{\tau})}$. Since $M(\boldsymbol{\tau}) \leq \sum_{|\ell|\leq\ell_0}|\theta_\ell|^2$ for all $\boldsymbol{\tau}\in\mathcal{T}^n$ one has that $|L(\boldsymbol{\tau})| \leq 2\left(\sum_{|\ell|\leq\ell_0}|\theta_\ell|^2\right)^{1/2}\sqrt{Q(\boldsymbol{\tau})}$, Therefore

(A.26)

$$\sup_{\boldsymbol{\tau}\in\mathcal{T}^n}|M_n(\boldsymbol{\tau})-M(\boldsymbol{\tau})| \leq \left(1 + 2\left(\sum_{|\ell|\leq\ell_0}|\theta_\ell|^2\right)^{1/2}\right)\left(\sqrt{\sup_{\boldsymbol{\tau}\in\mathcal{T}^n} Q(\boldsymbol{\tau})} + \sup_{\boldsymbol{\tau}\in\mathcal{T}^n} Q(\boldsymbol{\tau})\right),$$

Thus it suffices to derive a concentration inequality for $\sup_{\boldsymbol{\tau}\in\mathcal{T}^n} Q(\boldsymbol{\tau})$. For this remark that $Q(\boldsymbol{\tau}) \leq W_1$ for all $\boldsymbol{\tau}\in\mathcal{T}^n$, where $W_1 = \sum_{|\ell|\leq\ell_0}\frac{\epsilon^2}{n}\sum_{m=1}^{n}|z_{m,\ell}|^2$. Then using a standard concentration inequality for sum of χ^2 variables (see e.g. [34]) one has that for any $t > 0$ $\mathbb{P}\left(\sup_{\boldsymbol{\tau}\in\mathcal{T}^n} Q(\boldsymbol{\tau}) \geq C(\epsilon,n,\ell_0,t)\right) \leq \mathbb{P}\left(W_1 \geq C(\epsilon,n,\ell_0,t)\right) \leq \exp(-t)$. where $C(\epsilon,n,\ell_0,t) = \epsilon^2(2\ell_0+1) + 2\epsilon^2\sqrt{\frac{2\ell_0+1}{n}t} + 2\frac{\epsilon^2}{n}t$. Therefore the result follows using inequality (A.26). □

From Lemma 9 and inequality (A.25) it follows that

(A.27)

$$\mathbb{P}\left(\frac{1}{n}\sum_{m=2}^{n}(\hat{\tau}_m - \tilde{\tau}_m)^2 \leq \frac{2 + 4\left(\sum_{|\ell|\leq\ell_0}|\theta_\ell|^2\right)^{1/2}}{\kappa(f)\left(\sum_{|\ell|\leq\ell_0}|\theta_\ell|^2\right)}\left(\sqrt{C(\epsilon,n,\ell_0,t)} + C(\epsilon,n,\ell_0,t)\right)\right) \geq 1-\exp(-t).$$

To complete the proof remark that $\frac{1}{n}\sum_{m=2}^{n}(\hat{\tau}_m-\tau_m^*)^2 \leq 2\left(\frac{1}{n}\sum_{m=2}^{n}(\hat{\tau}_m - \tilde{\tau}_m)^2 + \left(\frac{1}{n}\sum_{m=1}^{n}\tau_m^*\right)^2\right)$. Since the τ_m^* are i.i.d variables with zero mean and bounded by 1/4, Bernstein's inequality (see e.g. [34]) implies that for any $t > 0$ then

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{m=1}^{n}\tau_m^*\right| \geq \sqrt{2\sigma_g^2\frac{t}{n}} + \frac{t}{12n}\right) \leq 2\exp(-t), \tag{A.28}$$

where $\sigma_g^2 = \int_{\mathcal{T}}\tau^2 g(\tau)d\tau$. Then Theorem 4 follows from inequalities (A.27) and (A.28). □

A.3. Proof of Theorem 5. To simplify the notations we write $\tau_m = \tau_m^*$ to denote the true shifts. Part of the proof is inspired by general results on Van Tree inequalities in [19]. First let us considered the case where the shifts $\tau_m, m = 1, \ldots, n$ are fixed parameters to estimate and let $\tau^n = (\tau_1, \ldots, \tau_n)$. Recall that $X = (c_{m,\ell})_{\ell \in \mathbb{Z}, m=1,\ldots,n}$ denote the set of observations taking values in the set $\mathcal{X} = \mathbb{C}^{\infty \times n}$. Then, the likelihood of the random variable X is given by $p(x|\tau^n) = C \prod_{m=1}^n \prod_{\ell \in \mathbb{Z}} \exp\left\{ -\frac{1}{2\epsilon^2} \left| c_{m,\ell} - \theta_\ell e^{-i2\pi\ell\tau_m} \right|^2 \right\}$. Therefore for $m = 1, \ldots, n$

$$\mathbb{E}_\tau \left(\frac{\partial}{\partial \tau_m} \log p(x|\tau^n)) \right) = 0, \tag{A.29}$$

where for a function $h(X)$ of the random variable X, $\mathbb{E}_\tau h(X) = \int_{\mathcal{X}} h(x) p(x|\tau^n) dx$. Then, for $m_1 \neq m_2$ one has that $\mathbb{E}_\tau \left(\frac{\partial}{\partial \tau_{m_1}} \log p(x|\tau^n) \frac{\partial}{\partial \tau_{m_2}} \log p(x|\tau^n) \right) = 0$, and for $m_1 = m_2$ $\mathbb{E}_\tau \left(\frac{\partial}{\partial \tau_{m_1}} \log p(x|\tau^n) \right)^2 = \epsilon^{-2} \sum_{\ell \in \mathbb{Z}} (2\pi\ell)^2 |\theta_\ell|^2$.

Now assume that the shifts are i.i.d. random variables with density $g(\tau)$ satisfying Assumption (6). Let $\hat{\tau}^n = \hat{\tau}^n(X)$ denote any estimator of the shifts τ^n. Then define the following vectors U and $V = (V_1, \ldots, V_n)'$ in $\mathbb{R}^n$ as

$$U = \hat{\tau}^n - \tau^n \text{ and } V_m = \frac{\partial}{\partial \tau_m} [p(x|\tau^n) g_n(\tau^n)] \frac{1}{p(x|\tau^n) g_n(\tau^n)} \text{ for } m = 1, \ldots, n,$$

where $g_n(\tau^n) = \prod_{m=1}^n g(\tau_m)$. First remark that

$$\begin{aligned}
\mathbb{E}\left(U'V\right) &= \int_{\mathcal{X}} \int_{\mathcal{T}^n} \sum_{m=1}^n (\hat{\tau}_m^n - \tau_m) \frac{\partial}{\partial \tau_m} [p(x|\tau^n) g_n(\tau^n)] d\tau^n dx \\
&= \int_{\mathcal{X}} \sum_{m=1}^n \hat{\tau}_m^n \left(\int_{\mathcal{T}^n} \frac{\partial}{\partial \tau_m} [p(x|\tau^n) g_n(\tau^n)] d\tau^n \right) dx \\
&\quad - \int_{\mathcal{X}} \sum_{m=1}^n \left(\int_{\mathcal{T}^n} \tau_m \frac{\partial}{\partial \tau_m} [p(x|\tau^n) g_n(\tau^n)] d\tau^n \right) dx
\end{aligned}$$

An integration by part and the fact that $\lim_{\tau \to \tau_{\min}} g(\tau) = \lim_{\tau \to \tau_{\max}} g(\tau) = 0$ implies that $\int_{\mathcal{T}^n} \frac{\partial}{\partial \tau_m} [p(x|\tau^n) g_n(\tau^n)] d\tau^n = 0$. Using again an integration by part and Assumption 3 one has that $\int_{\mathcal{T}^n} \tau_m \frac{\partial}{\partial \tau_m} [p(x|\tau^n) g_n(\tau^n)] d\tau^n = -\int_{\mathcal{T}^n} p(x|\tau^n) g_n(\tau^n) d\tau^n$. Therefore $\mathbb{E}\left(U'V\right) = \sum_{m=1}^n \int_{\mathcal{T}^n} \int_{\mathcal{X}} p(x|\tau^n) g_n(\tau^n) d\tau = n$. Now using Cauchy-Schwartz inequality it follows that $n^2 = \left(\mathbb{E}\left(U'V\right)\right)^2 \leq \mathbb{E}\left(U'U\right) \mathbb{E}\left(V'V\right)$. Then remark that

$$\mathbb{E}\left(U'U\right) = \mathbb{E} \sum_{m=1}^n (\hat{\tau}_m^n - \tau_m)^2 = \int_X \int_{\mathcal{T}^n} (\hat{\tau}_m^n(x) - \tau_m)^2 p(x|\tau^n) g_n(\tau^n) dx d\tau,$$

and

$$\begin{aligned}\mathbb{E}\left(V'V\right) &= \mathbb{E}\sum_{m=1}^{n}\left(\frac{\partial}{\partial\tau_m}\left[\log p(x|\tau^n)+\log g_n(\tau^n)\right]\right)^2\\ &= \mathbb{E}\sum_{m=1}^{n}\left(\frac{\partial}{\partial\tau_m}\log p(x|\tau^n)\right)^2+\mathbb{E}\sum_{m=1}^{n}\left(\frac{\partial}{\partial\tau_m}\log g_n(\tau^n)\right)^2,\end{aligned}$$

since by using (A.29) it follows that $\mathbb{E}\left(\sum_{m=1}^{n}\frac{\partial}{\partial\tau_m}\log p(x|\tau^n)\frac{\partial}{\partial\tau_m}\log g_n(\tau^n)\right) = \sum_{m=1}^{n}\int_{\mathcal{T}^n}\left(\int_{\mathcal{X}}\left(\frac{\partial}{\partial\tau_m}\log p(x|\tau^n)\right)p(x|\tau^n)dx\right)\left(\frac{\partial}{\partial\tau_m}\log g_n(\tau^n)\right)g_n(\tau^n)d\tau^n = 0.$ Hence

$$\begin{aligned}\mathbb{E}\left(V'V\right) &= n\epsilon^{-2}\sum_{\ell\in\mathbb{Z}}(2\pi\ell)^2|\theta_\ell|^2+\mathbb{E}\sum_{m=1}^{n}\left(\frac{\partial}{\partial\tau_m}\log g(\tau_m)\right)^2\\ &= n\epsilon^{-2}\sum_{\ell\in\mathbb{Z}}(2\pi\ell)^2|\theta_\ell|^2+n\int_{\mathcal{T}}\left(\frac{\partial}{\partial\tau}\log g(\tau)\right)^2 g(\tau)d\tau,\end{aligned}$$

which completes the proof using that $n^2\leq\mathbb{E}\left(U'U\right)\mathbb{E}\left(V'V\right)$. □

ACKNOWLEDGEMENTS

We would like to thank Patrick Cattiaux for fruitful discussions on likelihood ratios and the Girsanov formula.

REFERENCES

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

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

[3] Rabi. Bhattacharya and Vic Patrangenaru. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds–ii. *Ann. Statist.*, 33:1225–1259.

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

[5] Bigot, J., Gadat, S and Loubes, J.M. (2007). *Statistical M-Estimation and Consistency in large deformable models for Image Warping*, Journal of Mathematical Imaging and Vision, to be published.

[6] Bigot, J. , Gamboa, F. and Vimond, M. (2008). *Estimation of translation, rotation and scaling between noisy images using the Fourier Mellin transform*, SIAM Journal on Imaging Sciences, to be published.

[7] Buckheit, J.B. , Chen, S. , Donoho, D.L. and Johnstone, I. (1995). *Wavelab reference manual*, Department of Statistics, Stanford University, `http://www-stat.stanford.edu/software/wavelab`.

[8] Cavalier, L. , Golubev, G. K. , Picard, D. and Tsybakov, A. B. (2002). *Oracle inequalities for inverse problems*, The Annals of Statistics, 30 (3), 843–874.

[9] Cavalier, L. and Raimondo, M. (2007). *Wavelet Deconvolution With Noisy Eigenvalues*, IEEE Trans. on Signal Processing, 55 (4), 2414-2424.

[10] Cirelson, B. S. , Ibragimov, I. A. and Sudakov, V. N. (1976). *Norms of Gaussian sample functions*, Lecture Notes in Mathematics, Berlin: Springer-Verlag, 550, 20-41.

[11] Donoho, D. L. (1995). *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Appl. Comput. Harmon. Anal., 2 (2), 101–126.

[12] Donoho, D. L. , Johnstone, I. M. , Kerkyacharian, G. and Picard, D. (1995). *Wavelet Shrinkage: Asymptopia?*, Journal of the Royal Statistical Society (B), 57, 301–369.

[13] Efromovich, S. and Koltchinskii, V. (2001). *On inverse problems with unknown operators*, IEEE Transactions on Information Theory, 47, 2876-2894.

[14] Frechet, M. (1948). *Les éléments aleatoires de nature quelconque dans un espace distancié*, Annales de L'Institut Henri Poincare, 10 (1948) 215-310.

[15] Gasser, T. and Kneip, A. (1995). *Searching for Structure in Curve Samples*, Journal of the American Statistical Association, 90 (432), 1179-1188.

[16] Gasser, T. and Kneip, A. (1992). *Statistical Tools to Analyze Data Representing a Sample of Curves*, The Annals of Statistics, 20 (3), 1266-1305.

[17] Glasbey, C. A. and Mardia, K. V. (2001). *A penalized likelihood approach to image warping* (with discussion), Journal of the Royal Statistical Society (B), 63 (3), 465–514.

[18] Gervini, D. and Gasser, T. (2004). Self-Modeling Warping Functions, *Journal of the Royal Statistical Society (Series B)* **66**, 2, 959–971.

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

[20] Grenander, U. (1993). *General Pattern Theory: A Mathematical Study of Regular Structures*, Oxford University Press, New York.

[21] Hardle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation and. Statistical Applications*, Lecture Notes in Statistics, New York: Spriner-Verlag.

[22] Hoffman, M. and Reiss, M. (2008). *Nonlinear estimation for linear inverse problems with error in the operator*, The Annals of Statistics, 36(1), 310-336.

[23] Ikeda, N. and Watanabe, S (1989). *Stochastic Differential Equations and Diffusion Processes*, Second Edition, North-Holland/Kodansya, Tokyo.

[24] Isserles, U., Ritov, Y. and Trigano, T. (2008). *Semiparametric density estimation of shifts between curves*, Technical Report.

[25] Johnstone, I. , Kerkyacharian, G. , Picard, D. and Raimondo, M. (2004). *Wavelet deconvolution in a periodic setting*, Journal of the Royal Statistical Society (B), 66, 547–573.

[26] Kneip, A. and Gasser, T. (1988). *Convergence and consistency results for self-modelling regression*, The Annals of Statistics, 16, 82-112.

[27] Kolaczyk, E.D. (1994). *Wavelet Methods for the Inversion of Certain Homogeneous Linear Operators in the Presence of Noisy Data*, Ph.D. thesis, Department of Statistics, Stanford University.

[28] Korostelëv, A. P. and Tsybakov, A. B. (1993). *Minimax theory of image reconstruction*, , Lecture Notes in Statistics, New York.

[29] Liu, X. and Müller, H.G. (2004). *Functional Convex Averaging and Synchronization for Time-Warped Random Curves*, Journal of the American Statistical Association, 99 (467), 687-699.

[30] Castillo, I. and Loubes, J.M. (2008). *Estimation of the distribution of random*

shifts deformation, Mathematical Methods of Statistics, to be published.

[31] Loubes, J.M., Maza, E. and Gamboa, F. (2007). *Semi-parametric estimation of shifts*, Electronic Journal of Statistics, 1, 616-640.

[32] Ma, J. , Miller M.I., Trouvé A. and Younes L. (2008). *Bayesian template estimation in computational anatomy*, NeuroImage, 42 (1), 252-261.

[33] Mallat, S. (1998). *A Wavelet Tour of Signal Processing*, Academic Press, New York.

[34] Massart, P. (2006). *Concentration Inequalities and Model Selection: Ecole d'été de Probabilités de Saint-Flour XXXIII - 2003*, Lecture Notes in Mathematics Springer.

[35] Meyer, Y. (1992). *Wavelets and operators*, Cambridge Studies in Advanced Mathematics, Cambridge University Press.

[36] Pensky, M. and Sapatinas, T. (2008). *Functional deconvolution in a periodic setting: uniform case*, The Annals of Statistics, to be published.

[37] Pensky, M. and Vidakovic, B. (1999). *Adaptive Wavelet Estimator for Nonparametric Density Deconvolution*, The Annals of Statistics, 27, 2033–2053.

[38] Raimondo, M. and Stewart, M. (2007). *The WaveD Transform in R: performs fast translation-invariant wavelet deconvolution*, Journal of Statistical Software, 21 (3), 1-27.

[39] Ramsay, J.O. and Li, X. (2001). *Curve Registration*, Journal of the Royal Statistical Society (B), 63, 243-259.

[40] Ramsay, J.O. and Silverman, B.W. (2005). Functional Data Analysis, Second edition, *New York: Springer Verlag*.

[41] Ramsay, J.O. and Silverman, B.W. (2002). Applied Functional Data Analysis, *New York: Springer Verlag*.

[42] Ronn, B. (1998). *Nonparametric Maximum Likelihood Estimation for Shifted Curves*, Journal of the Royal Statistical Society (B), 60, 351-363.

[43] Rosenthal, H. P. (1972). *On the span in L_p of sequences of independent random variables. II, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, 149–167.*

[44] Vimond, M. (2008). *Efficient estimation for homothetic shifted regression models*, The Annals of Statistics, to be published.

[45] Wang, K. and Gasser, T. (1997). *Alignment of Curves by Dynamic Time Warping*, The Annals of Statistics, 25 (3), 1251-1276.

[46] Willer, T. (2006). *Deconvolution in white noise with a random blurring function*, Technical report.

Institut de Mathématiques de Toulouse
Université de Toulouse et CNRS (UMR 5219)
31062 Toulouse, Cedex 9, France
E-mail: Jeremie.Bigot@math.univ-toulouse.fr
Sebastien.Gadat@math.univ-toulouse.fr