Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Sharp template estimation in a shifted curves model

Jérémie Bigot, Sébastien Gadat and Clément Marteau

Institut de Mathématiques de Toulouse
Université de Toulouse et CNRS (UMR 5219)
31062 Toulouse, France
e-mail: {Jeremie.Bigot,Sebastien.Gadat,Clement.Marteau}@math.univ-toulouse.fr

Abstract: This paper considers the problem of adaptive estimation of a template in a randomly shifted curve model. Using the Fourier transform of the data, we show that this problem can be transformed into a linear inverse problem with a random operator. Our aim is to approach the estimator that has the smallest risk on the true template over a finite set of linear estimators defined in the Fourier domain. Based on the principle of unbiased empirical risk minimization, we derive a nonasymptotic oracle inequality in the case where the law of the random shifts is known. This inequality can then be used to obtain adaptive results on Sobolev spaces as the number of observed curves tend to infinity. Some numerical experiments are given to illustrate the performances of our approach.

AMS 2000 subject classifications: Primary 62G07, ; secondary 42C40, 41A29.
Keywords and phrases: Template estimation, Curve alignment, Inverse problem, Oracle inequality, Adaptive estimation, Random operator.

1. Introduction

1.1. Model and objectives

The goal of this paper is to study a special class of linear inverse problems with a random operator. We consider the problem of estimating a curve f, called template or shape function, from the observations of n noisy and randomly shifted curves $Y_1, \dots Y_n$ coming from the following Gaussian white noise model:

$$dY_j(x) = f(x - \tau_j)dx + \epsilon dW_j(x),\ x \in [0,1],\ j = 1, \dots, n \tag{1.1}$$

where W_j are independent standard Brownian motions on $[0,1]$, ϵ represents a level of noise common to all curves, the τ_j's are unknown random shifts independent of the W_j's, f is the unknown template to recover, and n is the number of observed curves that may be let going to infinity to study asymptotic properties. This model 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 (see e.g. [IRT08]) or biology (see e.g.

[Ron98]). A well known problem in functional data analysis is the alignment of similar curves that differ by a time transformation to extract their common features, and (1.1) is a simple model where f represents such common features (see [RS02], [RS05] for a detailed introduction to curve alignment problems in statistics).

The function $f : \mathbb{R} \to \mathbb{R}$ is assumed to be of period 1 so that the model (1.1) is well defined, and the shifts τ_j are supposed to be independent and identically distributed (i.i.d.) random variables with density $g : \mathbb{R} \to \mathbb{R}$ with respect to the Lebesgue measure dx on $\mathbb{R}$. Throughout the paper, it is supposed that the density g is known. Estimating f can be seen as an inverse problem with a random operator. Indeed, the template f is not observed directly, but through n independent realizations of the random operator $A_\tau : L^2_{per}([0,1]) \to L^2_{per}([0,1])$ defined by

$$A_\tau(f)(x) = f(x-\tau),\ x \in [0,1],$$

where $L^2_{per}([0,1])$ denotes the space of squared integrable functions on $[0,1]$ with period 1, and τ is random variable with density g. The additive Gaussian noise makes this problem ill-posed, and [BG10] have shown that estimating f in such models is in fact a deconvolution problem where the density g of the random shifts plays the role of the convolution operator. For the L^2 risk on $[0,1]$, [BG10] have derived the minimax rate of convergence for the estimation of f over Besov balls as n tends to infinity. This minimax rate depends both on the smoothness of the template and on the decay of the Fourier coefficients of the density g. This is a well known fact for standard deconvolution problem in statistics, see e.g. [Fan91], [Don95], but the results in [BG10] represent a novel contribution and a new point of view on template estimation in inverse problems with a random operator such as (1.1). This appears also to be a new setting in the field of inverse problem with partially known operators as considered in [CH05], [EK01], [HR08], [Mar06] and [CR07].

However, the approach followed in [BG10] is only asymptotic, and the main goal of this paper is to derive non-asymptotic results to study the estimation of f by keeping fixed the number n of observed curves.

1.2. Fourier Analysis and an inverse problem formulation

Supposing that $f \in L^2_{per}([0,1])$, we denote by θ_k its kth Fourier coefficient, namely:

$$\theta_k = \int_0^1 e^{-2ik\pi x} f(x)dx.$$

In the Fourier domain, the model (1.1) can be rewritten as

$$c_{j,k} := \int_0^1 e^{-2ik\pi x} dY_j(x) = \theta_k e^{-i2\pi k\tau_j} + \epsilon z_{k,j} \tag{1.2}$$

where $z_{k,j}$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ variables, i.e. complex Gaussian variables with zero mean and such that $\mathbb{E}|z_{k,j}|^2 = 1$. This means that the real and imaginary parts

of the $z_{k,j}$'s are Gaussian variables with zero mean and variance $1/2$. Thus, we can compute the sample mean of the kth Fourier coefficient over the n curves as

$$\tilde{c}_k := \frac{1}{n}\sum_{j=1}^{n} c_{k,j} = \theta_k \tilde{\gamma}_k + \frac{\epsilon}{\sqrt{n}}\xi_k, \tag{1.3}$$

where

$$\tilde{\gamma}_k := \frac{1}{n}\sum_{j=1}^{n} e^{-i2\pi k\tau_j}, \tag{1.4}$$

and the ξ_k's are i.i.d. complex Gaussian variables with zero mean and variance 1. The Fourier coefficients $\tilde{c}_k$ in equation (1.3) can be viewed as observations coming from a statistical inverse problem. Indeed, the standard sequence space model of an ill-posed statistical inverse problem is (see [CGPT02] and the references therein)

$$c_k = \theta_k \gamma_k + \sigma z_k, \tag{1.5}$$

where the γ_k's are eigenvalues of a known linear operator, z_k are random noise variables and σ is a level of noise which goes to zero for studying asymptotic properties. The issue in such models is to recover the coefficients θ_k from the observations c_k under various conditions on the decay to zero of the γ_k's as $|k| \to +\infty$. A large class of estimators for the problem (1.5) can be written as

$$\hat{\theta}_k = \lambda_k \frac{c_k}{\gamma_k},$$

where $\lambda = (\lambda_k)_{k\in\mathbb{Z}}$ is a sequence of reals called filter. Various estimators of this form have been studied in a number of papers, and we refer to [CGPT02] for more details.

In a sense, we can view equation (1.3) as a linear inverse problem (with $\sigma = \frac{\epsilon}{\sqrt{n}}$) with a stochastic operator whose eigenvalues $\tilde{\gamma}_k = \frac{1}{n}\sum_{j=1}^{n} e^{-i2\pi k\tau_j}$ are random variables that are not observed. Nevertheless, it is supposed that the density g of the shifts is known. Therefore, one can compute the expectation γ_k of the random eigenvalues $\tilde{\gamma}_k$ given by

$$\gamma_k := \mathbb{E}\tilde{\gamma}_k = \mathbb{E}\left(e^{-i2\pi k\tau}\right) = \int_{-\infty}^{+\infty} e^{-i2\pi kx} g(x)dx.$$

Hence, if we assume that the density g of the random shifts is known, estimation of the Fourier coefficients of f can be obtained by a deconvolution step of the form

$$\hat{\theta}_k = \lambda_k \frac{\tilde{c}_k}{\gamma_k}, \tag{1.6}$$

where $\tilde{c}_k$ is defined in (1.3) and $\lambda = (\lambda_k)_{k\in\mathbb{Z}}$ is a filter whose choice will be discussed later on. Theoretical properties and optimal choices for the filter λ are presented in the case where the coefficients γ_k are known. Such a framework is commonly used in inverse problems such as (1.5) 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 σ^2, see e.g [CGPT02]. In model (1.1), 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.

Recently, some papers have addressed the problem of regularization with partially known operator. For instance, [CH05] consider the case where the eigenvalues are unknown but independently observed. They deal with the model:

$$c_k = \gamma_k \theta_k + \epsilon \xi_k, \ \tilde{\gamma}_k = \gamma_k + \sigma \eta_k, \ \forall k \in \mathbb{N}, \tag{1.7}$$

where $(\xi_k)_{k\in\mathbb{N}}$ and $(\eta_k)_{k\in\mathbb{N}}$ denote i.i.d standard gaussian variables. In this case, each coefficient θ_k can be estimated by $\tilde{\gamma}_k^{-1} c_k$. Similar models have been considered in [CR07], [Mar06] or [Mar09]. In a more general setting, we may refer to [EK01] and [HR08].

In this paper, our framework is slightly different in the sense that the operator is stochastic, but the regularization is operated using deterministic eigenvalues. Hence the approach followed in the previous papers is no directly applicable to model (1.1). We believe that estimating f and deriving convergence rates in model (1.1) without the knowledge of g remains a difficult task, and this paper is a first step to address this issue.

1.3. Previous work in template estimation and shift recovery

The problem of estimating the common shape of a set of curves that differ by a time transformation is usually referred to as the curve registration problem, and it has received a lot of attention in the literature over the last two decades. Among the various methods that have been proposed, one can distinguish between landmark-based approaches which aim at aligning common structural points of the curves (typically locations of extrema) see e.g [GK92], [GK95], [Big06], and nonparametric modeling of the warping functions to align a set of curves see e.g [RL01], [WG97], [LM04]. However, in these papers, studying consistent estimates of the common shape f as the number of curves n tends to infinity is generally not considered.

In the simplest case of shifted curves, various approaches have been developed. Self-modelling regression methods proposed by [KG88] are semiparametric models where each observed curve is a parametric transformation of a common regression function. Such models are usually referred to as shape invariant models and estimation in this setting is usually done by iterating the following two steps: estimation of the parameters of the transformations (here the shifts) given a reference curve, and nonparametric estimation of a template by aligning the observed curves given a set of known transformation parameters. [KG88] studied the consistency of such a two steps procedure in an asymptotic framework where both the number of functions n and the number of observed points per curves grows to infinity. Due to the asymptotic equivalence between the white noise model and nonparametric regression with an equi-spaced design (see [BL96]),

such an asymptotic framework in our setting would correspond to the case where both n tends to infinity and ϵ is let going to zero. In this paper we prefer to focus only on the case where n may be let going to infinity, and to leave fixed the level of additive noise in each observed curve.

Based on a model with curves observed at discrete time points, semiparametric estimation of the shifts and the shape function is proposed in [GLM07] and [Vim10] as the number of observations per curve grows, but with a fixed number n of curves. A generalization of this approach for the estimation of scaling, rotation and translation parameters for two-dimensional images is also proposed in [BGV09], but also with a fixed number of observed images. Semiparametric and adaptive estimation of a shift parameter in the case of a single observed curve in a white noise model is also considered by [DGT06] and [Dal07]. Estimation of a common shape for randomly shifted curves and asymptotic in n is considered in [Ron98] from the point of view of semiparametric estimation when the parameter of interest is infinite dimensional.

However, in all the above cited papers rates of convergence or oracle inequalities for the estimation of the template are generally not studied. Moreover, our procedure differs from the approaches classically used in curve registration as our estimator is obtained in only one very simple step, and it is not based on an alternative scheme between estimation of the shifts and averaging of back-transformed curves given estimated values of the shifts parameters.

Finally, note that [CL09] and [IRT08] consider a model similar to (1.1), but they rather focus on the the estimation of the density g of the shifts as n tends to infinity. Using such an approach could be a good start for studying the estimation of the template f without the knowledge of g. However, we believe that this is far beyond the scope of this paper, and we prefer to leave this problem open for future work.

1.4. Organization of the paper

In Section 2, we consider an estimator of the shape function f using monotone filters when the eigenvalues γ_k are known. Based on the principle of unbiased risk minimization developed by [CGPT02], we propose a data-based choice for the filter λ in (1.6). Then, we derive an oracle inequality showing that the resulting estimator has a risk close to an ideal one when choosing λ over a class of monotone filters. In Section 3, as an example, we study the case of projection filters. This gives an estimator based on the Fourier transform of the curves with a data-based choice of the frequency cut-off. We study its asymptotic properties in terms of minimax rates of converge over Sobolev balls. Finally in Section 4, a detailed simulation study is proposed to illustrate the numerical properties of such estimators. All proofs are deferred to a technical section at the end of the paper.

2. Estimation of the common shape

In the following, we assume that the Fourier coefficients γ_k are known. In this situation it is possible to choose a data-dependent filter $\lambda^\star$ that mimics the performances of an optimal filter λ^0 called oracle that would be obtained if we knew the true template f. The performances of this filter are related to the performances of the filter λ^0 via an oracle inequality. In this section, most of our results are non-asymptotic and are thus related to the approach proposed in [CGPT02] to study standard statistical inverse problems via oracle inequalities.

2.1. Smoothness assumptions for the density g

In a deconvolution problem, it is well known that the difficulty of estimating f is quantified by the decay to zero of the γ_k's as $|k| \to +\infty$. Depending how fast these Fourier coefficients tend to zero as $|k| \to +\infty$, the reconstruction of f will be more or less accurate. This phenomenon was systematically studied by [Fan91] in the context of density deconvolution. In this paper, the following type of assumption on g is considered:

Assumption 2.1. *The Fourier coefficients of g have a polynomial decay i.e. for some real $\beta \geq 0$, there exists two constants $C_{max} \geq C_{min} > 0$ such that for all $k \in \mathbb{Z}$*

$$C_{min}|k|^{-\beta} \leq |\gamma_k| \leq C_{max}|k|^{-\beta}. \tag{2.1}$$

2.2. Risk decomposition

Recall that an estimator of the θ_k's is given by $\hat{\theta}_k = \lambda_k \gamma_k^{-1} \tilde{c}_k,\ k \in \mathbb{Z}$, see equation (1.6), where $\lambda = (\lambda_k)_{k\in\mathbb{Z}}$ is a real sequence. Examples of commonly used filters include projection weights $\lambda_k = \mathbb{1}_{|k|\leq j}$ for some integer j, and the Tikhonov weights $\lambda_k = 1/(1 + (|k|/\nu_2)^{\nu_1})$ for some parameters $\nu_1 > 0$ and $\nu_2 > 0$. Based on the $\hat{\theta}_k$'s, one can estimate the signal f using the Fourier reconstruction formula

$$\hat{f}_\lambda(x) = \sum_{k\in\mathbb{Z}} \hat{\theta}_k e^{-2ik\pi x}.$$

The problem is then to choose the sequence $(\lambda_k)_{k\in\mathbb{Z}}$ in an optimal way with respect to an appropriate risk. For a given filter λ we use the classical ℓ_2-norm to define the risk of the estimator $\hat{\theta}(\lambda) = (\hat{\theta}_k)_{k\in\mathbb{Z}}$

$$R(\theta, \lambda) = \mathbb{E}_\theta \|\hat{\theta}(\lambda) - \theta\|_2^2 = \mathbb{E}_\theta \sum_{k\in\mathbb{Z}} |\hat{\theta}_k - \theta_k|^2 \tag{2.2}$$

Note that analyzing the above risk (2.2) is equivalent to analyze the mean integrated square risk $R(\hat{f}_\lambda, f) = \mathbb{E}\|\hat{f}_\lambda - f\|^2 = \mathbb{E}\left(\int_0^1 (\hat{f}_\lambda(x) - f(x))^2 dx\right)$. The following lemma gives the bias-variance decomposition of $R(\lambda, \theta)$. A detailed proof can be found in [BG10].

Lemma 2.1. *For any given nonrandom filter λ, the risk of the estimator $\hat{\theta}(\lambda)$ can be decomposed as*

$$R(\theta,\lambda) = \underbrace{\sum_{k\in\mathbb{Z}}(\lambda_k-1)^2|\theta_k|^2}_{Bias} + \underbrace{\frac{1}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2\frac{\epsilon^2}{|\gamma_k|^2}}_{V_1} + \underbrace{\frac{1}{n}\sum_{k\in\mathbb{Z}}\left[\lambda_k^2|\theta_k|^2\left(\frac{1}{|\gamma_k|^2}-1\right)\right]}_{V_2} \tag{2.3}$$

For a fixed number of curves n and a given shape function f, the problem of choosing an optimal filter in a set of possible candidates is to find the best trade-off between low bias and low variance in the above expression. However, this decomposition does not correspond exactly to the classical bias-variance decomposition for linear inverse problems. Indeed, the variance term in (2.3) is the sum of two terms and differs from the classical expression of the variance for linear estimator in statistical inverse problems. Using our notations, the classical variance term is $V_1 = \frac{\epsilon^2}{n}\sum_{k\in\mathbb{Z}}\frac{\lambda_k^2}{|\gamma_k|^2}$ and appears in most of linear inverse problems. However, contrary to standard inverse problems, the variance term of the risk also depends on the Fourier coefficients θ_k of the unknown function f to recover. Indeed, our data $\gamma_k^{-1}\tilde{c}_k$ are noisy observations of θ_k:

$$\gamma_k^{-1}\tilde{c}_k = \theta_k + \left(\frac{\tilde{\gamma}_k}{\gamma_k}-1\right)\theta_k + \frac{\epsilon}{\sqrt{n}}\gamma_k^{-1}\xi_k. \tag{2.4}$$

Hence, using the sequence $(\gamma_k)_{k\in\mathbb{N}}$ instead of $(\tilde{\gamma}_k)_{k\in\mathbb{N}}$ introduces an additional error. This explains the presence of the second term V_2.

A similar phenomenon occurs with the model (1.7), although it is more difficult to quantify it. Indeed, in this setting:

$$\tilde{\gamma}_k^{-1}c_k = \theta_k + \left(\frac{\gamma_k}{\tilde{\gamma}_k}-1\right)\theta_k + \epsilon\tilde{\gamma}_k^{-1}\xi_k,\ \forall k\in\mathbb{N}. \tag{2.5}$$

Hence, we also observe an additional term depending on θ. This term is controlled using a Taylor expansion but the quadratic risk cannot be expressed in a simple form. Let us stress that the difficulty of studying problem (2.5), when compared to our estimator (2.4), comes from the fact that in (2.5) there is a random term in the denominator. We refer to [Mar09] for a discussion with some numerical simulation and to [CH05], [EK01], [HR08], [Mar06] and [CR07].

2.3. An oracle estimator and unbiased estimation of the risk (URE)

Suppose that one is given a set Λ of cardinality $N \geq 1$ of possible candidate filters, that is $\Lambda = (\lambda^j)_{j\in\{1,\dots,N\}}$, with $\lambda^j = (\lambda_k^j)_{k\in\mathbb{Z}},\ j = 1,\dots,N$ which satisfy some general conditions to be discussed later on. In the case of projection filters, Λ can be for example the set of filters $\lambda_k^j = \mathbb{1}_{|k|\leq j}, k\in\mathbb{Z}$ for $j = 1,\dots,N$.

Given a set of filters Λ, the best estimator is defined as the filter λ^0 (called oracle) which has the smallest risk $R(\theta, \lambda)$ over Λ, that is

$$\lambda^0 := \arg\min_{\lambda \in \Lambda} R(\theta, \lambda). \tag{2.6}$$

This filter is an ideal one because it cannot be computed in practice as the sequence of coefficients θ is unknown. However, the oracle λ^0 can be used as a benchmark to evaluate the quality of a data-dependent filter $\lambda^\star$ chosen in the set Λ. This is the main interpretation of the oracle inequality that we will develop in the next section.

2.4. Oracle inequalities for monotone filters

2.4.1. Definitions

First, let us introduce the following class of monotone filters:

$$\Lambda_{mon} := \left\{ \lambda = (\lambda_k)_{k \in \mathbb{Z}} : \lambda_k = \lambda_{-k}, \sum_{k \in \mathbb{Z}} \lambda_k^2 < +\infty,\ 1 \geq \lambda_0 \geq \ldots \geq \lambda_m \geq \ldots \geq 0 \right\},$$

In practice, the filters λ in the set Λ are such that $\lambda_k = 0$ (or vanishingly small) for all k large enough. Hence, for such choices of filters, numerical minimization of criterions such as (2.6) is feasible, since it only involves the computation of finite sums. Let us thus define the following threshold m_0 beyond which all values of the filters λ in Λ vanish

$$m_0 = \inf \left\{ k : |\gamma_k|^2 \leq \frac{\log^2 n}{n} \right\} - 1. \tag{2.7}$$

Then, Λ is supposed to be a finite set of cardinality N of monotone filters λ which satisfies $\lambda_k = 0$ as soon as $|k| \geq m_0$, that is

Assumption 2.2. *For $N \geq 1$, $\Lambda = (\lambda^j)_{j \in \{1,\ldots,N\}} \subset \Lambda_{mon}$ with $\lambda_k^j = 0$ for $|k| \geq m_0$ and $j = 1, \ldots, N$.*

The choice of the filter $\lambda^\star$ will be obtained by minimization of a data-based criterion whose derivation is guided by the unbiased risk estimate (URE) minimization principle developed by [CGPT02]. Typically, one cannot minimize such a criterion over filters $(\lambda_k)_{k \in \mathbb{Z}}$ of infinite length. Indeed, each coefficient θ_k is estimated by $\gamma_k^{-1} \tilde{c}_k$ where $\gamma_k = \mathbb{E}\tilde{\gamma}_k$. Hence, the ratio $\gamma_k^{-1}\tilde{\gamma}_k$ should be as close as possible to 1. Since $\gamma_k \to 0$ as $k \to +\infty$ and the variance of $\tilde{\gamma}_k$ is equal to $\frac{1}{n} + (1 - \frac{1}{n})|\gamma_k|^2$, it is clear that large values of k should be discarded. Bounds similar to (2.7) on the maximum number of non-vanishing values for the filters are used in papers related to partially known operator, see for instance [CH05] or [EK01]. This bounds have to be carefully chosen but are not of first importance. In general, estimating the operator is easier than estimating the function f.

In this paper, we have chosen to present an adaptive estimator based on the URE principle. Given the finite family Λ, our aim is to select the best possible filter among this family. We are aware that different adaptive schemes are available in the literature. For instance, the penalized blockwise Stein's rule (see [CT02] and references therein) provides a filter for the model (1.5) leading to an oracle inequality among all monotone filters. In some sense, the generalization of such kind of result to our model would be more powerful. Nevertheless, we think that our approach is also interesting in this setting since it does not impose a particular regularization scheme. Moreover, the differences between model (1.5) and model (1.3) are easier to underline with our method.

2.4.2. Adaptive regularization scheme

Let us now explain how to compute an estimator $U(Y,\lambda)$ of the risk $R(\theta,\lambda)$. First, recall that Lemma 2.1 yields the following expression of the quadratic risk $R(\theta,\lambda)$

$$R(\theta,\lambda) \quad = \quad \underbrace{\sum_{k\in\mathbb{Z}}(1-\lambda_k)^2|\theta_k|^2}_{:=Bias} + \underbrace{\frac{\epsilon^2}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2|\gamma_k|^{-2}}_{:=V_1} + \underbrace{\frac{1}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2|\theta_k|^2\left(\frac{1}{|\gamma_k|^2}-1\right)}_{:=V_2},$$

and suppose that it is possible to construct an estimator $\hat{\Theta}_k^2$ of $|\theta_k|^2$ from the observations of the shifted curves $Y=(Y_i)_{i=1\dots n}$. For any non-random filter λ satisfying Assumption 2.2, by replacing $|\theta_k|^2$ in (2.3) by $\hat{\Theta}_k^2$, the above decomposition of the risk $R(\theta,\lambda)$ suggests to compute a data-based criterion $U(Y,\lambda)$ (depending only on (Y,λ)) of the form

$$U(Y,\lambda) := \sum_{|k|\le m_0}(\lambda_k^2-2\lambda_k)\hat{\Theta}_k^2+\frac{\epsilon^2}{n}\sum_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2}+\frac{1}{n}\sum_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2}\hat{\Theta}_k^2. \quad (2.8)$$

The criterion $U(Y,\lambda)$ is thus an approximation of $R(\theta,\lambda)-\|\theta\|_2^2$. Then, for choosing a data-dependent filter $\lambda^\star$, the principle of URE, see [CGPT02] for further details, simply suggests to minimize the criterion $U(Y,\lambda)$ over $\lambda\in\Lambda$. Following the principle of URE, a data-dependent choice of λ would thus be given by

$$\lambda^\star := \arg\min_{\lambda\in\Lambda} U(Y,\lambda). \quad (2.9)$$

In the following, we use $\hat{\Theta}_k^2=\gamma_k^{-2}\left[|\tilde{c}_k|^2-\frac{\epsilon^2}{n}\right]$ as an estimator of $|\theta_k|^2$. Remark that $\mathbb{E}_\theta\hat{\Theta}_k^2\neq|\theta_k|^2$. Hence, the criterion $U(Y,\lambda)$ is not an unbiased estimation of $R(\theta,\lambda)-\|\theta\|_2^2$, meaning that we rather use the principle of minimization of a risk estimate. Nevertheless, we will prove that this bias can be controlled, and that it is in some sense negligible compared to $R(\theta,\lambda)$.

Note that for the computation of $U(Y,\lambda)$, we have taken into account all the terms (Bias, V_1 and V_2) in the decomposition of the risk $R(\theta,\lambda)$. Unfortunately, when using $\hat{\Theta}_k^2=\gamma_k^{-2}\left[|\tilde{c}_k|^2-\frac{\epsilon^2}{n}\right]$ as an estimator of $|\theta_k|^2$, minimization

of such a criterion does not lead to satisfactory results. This is due the term $\frac{1}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2|\gamma_k|^{-4}|\left\{|\tilde{c}_k|^2-\frac{\epsilon^2}{n}\right\}$ in (2.8) which is an estimation of the variance term V_2 in the decomposition (2.3) of the risk $R(\theta,\lambda)$. The main issue is that the study of this term requires a control of $|\gamma_k|^{-4}$, and not only $|\gamma_k|^{-2}$ as for the study of the classical variance term $V_1=\frac{\epsilon^2}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2|\gamma_k|^{-2}$ in standard inverse problem. Nevertheless, by definition (2.7), one has that $\frac{\log^2(n)}{n}\gamma_k^{-2}\leq 1$ for all $|k|\leq m_0$. Therefore, this suggests to rather consider filters minimizing a criterion of the form

$$\begin{aligned} U_1(Y,\lambda) \quad := \quad & \sum_{|k|\leq m_0}(\lambda_k^2-2\lambda_k)|\gamma_k|^{-2}\left\{|\tilde{c}_k|^2-\frac{\epsilon^2}{n}\right\}+\frac{\epsilon^2}{n}\sum_{|k|\leq m_0}\lambda_k^2|\gamma_k|^{-2} \\ & +\frac{\log^2(n)}{n}\sum_{|k|\leq m_0}\lambda_k^2|\gamma_k|^{-4}\left\{|\tilde{c}_k|^2-\frac{\epsilon^2}{n}\right\}. \end{aligned} \tag{2.10}$$

Alternatively, following [CH05], it is sometimes possible to neglect the error generated by the use of an approximation of the unknown random eigenvalues $\tilde{\gamma}_k$ by γ_k which yet corresponds to the term V_2. Indeed, remark that one may find $\rho>0$ such that

$$V_2\leq\frac{1}{n}\sum_{k\in\mathbb{Z}}\frac{\lambda_k^2|\theta_k|^2}{|\gamma_k|^2}\leq\frac{1}{n}\|\theta\|^2\sup_{k\in\mathbb{Z}}\lambda_k^2|\gamma_k|^{-2}\leq\rho\|\theta\|^2\frac{1}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2|\gamma_k|^{-2}=\rho\|\theta\|^2\frac{V_1}{\epsilon^2}.$$

Hence, depending on the values of ρ, ϵ^2 and $\|\theta\|^2$, the variance term V_2 may be negligible compared to V_1. In this case, one could rather consider the following criterion $U_0(Y,\lambda)$ derived from the decomposition on the classic quadratic risk (*i.e.* Bias + V_1), and defined as

$$U_0(Y,\lambda):=\sum_{|k|\leq m_0}(\lambda_k^2-2\lambda_k)|\gamma_k|^{-2}\left\{|\tilde{c}_k|^2-\frac{\epsilon^2}{n}\right\}+\frac{\epsilon^2}{n}\sum_{|k|\leq m_0}\lambda_k^2|\gamma_k|^{-2}. \tag{2.11}$$

In the sequel, we summarize these two approaches by considering the more general criterion $U_\alpha(Y,\lambda)$ given by

$$\begin{aligned} U_\alpha(Y,\lambda) \quad := \quad & \sum_{k\in\mathbb{Z}}(\lambda_k^2-2\lambda_k)|\gamma_k|^{-2}\left\{|\tilde{c}_k|^2-\frac{\epsilon^2}{n}\right\}+\frac{\epsilon^2}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2|\gamma_k|^{-2} \\ & +\alpha\frac{\log^2 n}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2|\gamma_k|^{-4}\left\{|\tilde{c}_k|^2-\frac{\epsilon^2}{n}\right\}. \end{aligned} \tag{2.12}$$

where $0\leq\alpha\leq 1$ is a parameter to be discussed. All the following results of the paper are given for any value of the parameter α in $[0,1]$. Following the URE principle, we will study the theoretical properties of the filters $\lambda_\alpha^\star\in\Lambda$ defined as

$$\lambda_\alpha^\star=\arg\min_{\lambda\in\Lambda}U_\alpha(Y,\lambda). \tag{2.13}$$

for $0 \leq \alpha \leq 1$. Note that $U_\alpha(Y,\lambda)$ can be written as a penalized version of the empirical risk $U_\alpha(Y,\lambda)$ defined in (2.11) and α can hence be considered as a penality constant. The choice of this term has no real influence on the rate of convergence (see Section 3 below) although this is a delicate problem in nonparametric statistic from a non-asymptotic point of view. Nevertheless, the following heuristic arguments can be given. The presence of the additional penalized term V_2 is due to the variability along the time axis (random translation) of the template f. When ϵ is small compared to $\|\theta\|_2$, the white noise deconvolution may be considered as negligible comparing to the alignment issue of the observed curves. The mean error will be larger when the signal to reconstruct possesses a large number of modes. Thus, in a framework with a small ϵ and a large $\|\theta\|_2$, it may be reasonable to choose $\alpha \neq 0$. To the contrary, if the level of noise ϵ is large, the model (1.1) can certainly be considered as being close to the standard white noise deconvolution problem. In this setting, setting $\alpha = 0$ may be recommended. Moreover, an optimal choice of α is certainly related to the number of observed curves n. The problem of choosing α is thus discussed in detail in Section 4 on numerical experiments.

2.4.3. Sharp estimator of the oracle risk

We are now able to propose an adaptive estimator of θ. In the following, α will belong to $[0,1]$ and we denote by $\theta^\star_\alpha$ the estimator related to the filters $\lambda^\star_\alpha$ defined in (2.13) that is

$$\theta^\star_{k,\alpha} = \frac{\tilde{c}_k}{\gamma_k}\lambda^\star_{k,\alpha} \text{ for } \theta^\star_\alpha = (\theta^\star_{k,\alpha})_{k\in\mathbb{Z}} \text{ and } \lambda^\star_\alpha = (\lambda^\star_{k,\alpha})_{k\in\mathbb{Z}}. \tag{2.14}$$

To simplify the notations, we omit the dependency of $\theta^\star_\alpha$ and $\lambda^\star_\alpha$ on α, and write $\theta^\star = \theta^\star_\alpha$ and $\lambda^\star = \lambda^\star_\alpha$. Through a simple oracle inequality, the next theorem relates the performances of $\theta^\star$ to the ideal filter λ^0 minimizing the risk $R(\theta,\lambda)$ over $\lambda \in \Lambda$. We denote by L_Λ the term introduced in [CGPT02] which in some sense measure the complexity of the family Λ. The proof of the theorem and a complete definition of L_Λ are given in the Appendix.

Theorem 2.1. *Suppose that Assumption 2.2 holds and that the density g satisfies Assumption 2.1. Let $\theta^\star$ defined by (2.14). Then, there exists $0 < \gamma_1 < 1$ such that, for all $0 < \gamma < \gamma_1$,*

$$\begin{aligned}\mathbb{E}_\theta\|\theta^\star - \theta\|^2 \quad &\leq \quad (1+h_{\gamma,n}) \inf_{\lambda\in\Lambda}\left[R(\theta,\lambda) + \alpha\frac{\log^2 n}{n}\sum_{k\in\mathbb{Z}}\lambda_k^2|\gamma_k|^{-2}|\theta_k|^2\right] + \Gamma(\theta) \qquad (2.15)\\ &+C_1\frac{\epsilon^2}{n}L_\Lambda\omega\left(\|\theta\|^2 L_\Lambda\gamma^{-1}\right) + C_2\frac{\log^2 n}{n}\omega\left((1-\alpha)\|\theta\|^2\log^2(n)\gamma^{-1}\right) + Ce^{-\gamma^2\log^{1+\tau} n},\end{aligned}$$

where $h_{\gamma,n)} \to 0$ as $\gamma \to 0$ and $n \to +\infty$, C_1, C_2 and $\tau > 0$ are suitable constants independent of n,

$$\Gamma(\theta) = \left|\sum_{|k|>m_0} \epsilon^2\{\lambda_k^0\}^2|\gamma_k|^{-2} + (1-(1-\lambda_k^0)^2)\theta_k^2\right|,$$

and $\omega(x) = \max_{\lambda\in\Lambda} \sup_k \lambda_k^2 |\gamma_k|^{-2} \mathbb{1}_{\left\{\sum_i \lambda_i^2|\gamma_i|^{-2} \leq x \sup_i \lambda_i^2|\gamma_i|^{-2}\right\}} \quad \forall x > 0.$

Theorem 2.1 proves that the quadratic risk is comparable to the risk of the oracle up to some residual terms. Before explaining these terms, just a few words on the quantities in the infimum. First, if $\alpha = 0$, then $\mathbb{E}_\theta\|\theta^\star - \theta\|^2$ is comparable to $R(\theta, \lambda^0)$ but the price to pay is a residual term of order $\log^2(n)/n$. In the case where $\alpha = 1$, we reach the quadratic risk up to a log term. This lack of precision can be explained by the processes involved in $U_1(Y, \lambda)$, which are hardly controllable due to the dependency between the $\tilde{\gamma}_k$. Previously, we have only given some heuristic arguments on the way α could be chosen. Theorem 2.1 presents results for all possible values of α between 0 and 1. Therefore, the above theorem can give some hints on how choosing α, but keeping in mind that theses results are derived from successive upper bounds that could certainly be enhanced. Let us recall that the choice of α has no real influence on the minimax rates of convergence (see Section 3). It is strongly related to the choice of a good penalty in our criteria which is a classical issue in many statistical problems.

The function ω was initially introduced in [CGPT02]. Under Assumption 2.1, it is of order $x^{2\beta}$ for many kind of filters (spectral cut-off, Tikhonov, Landweber, etc...). Hence, the two terms of (2.15) depending on ω are respectively of order ϵ^2/n and $\log^2(n)/n$. They can be reasonably considered as negligible compared to $R(\theta, \lambda^0)$ in many situations (see for instance Section 3 bellow). The same remark hold for the term $Ce^{-\gamma^2 \log^{1+\tau} n}$, which tends to 0 faster than n^{-1}.

We conclude this discussion with the term $\Gamma(\theta)$. This term measures the error associated to the truncation of the estimation at the order m_0. Consider for instance the particular case of a projection (or spectral cut-off) family: $\lambda_k^j = \mathbf{1}_{\{|k|\leq j\}}$ for all $j = 1, \ldots, N$. Denote by $\lambda_{j_0} = \lambda^0$ the oracle filter. Then, $\Gamma(\theta) = 0$ as soon as the oracle bandwidth j_0 is smaller than m_0. In some sense, the control of $(\tilde{\gamma}_k)_{k\in\mathbb{Z}}$ is easier than the estimation of $(\theta_k)_{k\in\mathbb{Z}}$ (no inversion to perform). Hence, in many cases, $\Gamma(\theta) = 0$. A similar discussion holds for other kind of filters.

3. Minimax rates of convergence for Sobolev balls

Let us now study the special case of projection filters. In this section, we prove that such estimators attain the minimax rate of convergence on many functional spaces. In particular, the term $\log^2(n)$ added in (2.12) to control the estimation of the variance term V_2, and the maximal bandwidth m_0 (2.7) have no influence on the performances of our estimator from a minimax point of view.

Let $1 \leq p, q \leq \infty$ and $A > 0$, and suppose that f belongs to a Besov ball $\mathcal{B}^s_{p,q}(A)$ of radius A (see e.g. [DJKP95] for a precise definition of Besov spaces). [BG10] have derived the following asymptotic minimax lower bound for the quadratic risk over a large class of Besov balls.

Theorem 3.1. *Let $1 \leq p, q \leq \infty$ and $A > 0$, let $p' = p \wedge 2$ and assume that: $f \in \mathcal{B}^s_{p,q}(A)$ and $s \geq p'$ (Regularity condition on f), g satisfies the polynomial*

decreasing condition (2.1) at rate β on its Fourier coefficients (Regularity condition on g), $s \geq (2\beta+1)(1/p-1/2)$ and $s \geq 2\beta+1$ (Dense case). Then, there exists a universal constant M_1 depending on A, s, p, q such that

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{B}^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n - f\|^2 \geq M_1 n^{\frac{-2s}{2s+2\beta+1}}, \quad as \quad n \to \infty,$$

where $\hat{f}_n \in L^2_{per}([0,1])$ denotes any estimator of the common shape f, i.e a measurable function of the random processes $Y_j,\ j = 1, \ldots, n$

Therefore, Theorem 3.1 extends the lower bound $n^{\frac{-2s}{2s+2\beta+1}}$ usually obtained in a classical deconvolution model to the more complicated model of deconvolution with a random operator derived from equation (1.1). Then, let us introduce the following smoothness class of functions which can be identified with a periodic Sobolev ball:

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

for some constant $A > 0$ and some smoothness parameter $s > 0$, where $\theta_k = \int_0^1 e^{-2ik\pi x} f(x)dx$. It is known (see e.g. [DJKP95]) that if s is not an integer then $H_s(A)$ can be identified with a Besov ball $\mathcal{B}^s_{2,2}(A')$.

Let $\Lambda = (\lambda^j)_{j \in \{1,\ldots,N\}}$, with $\lambda^j_k = \mathbb{1}_{|k| \leq j}, k \in \mathbb{Z}$ for $j = 1, \ldots, N$ and $N \leq m_0$ be a set of projection filters. In this case, the decomposition of the quadratic risk for the filter $\lambda^j \in \Lambda$ is

$$R(\theta, \lambda^j) = \sum_{|k| \geq j} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k| \leq j} |\gamma_k|^{-2} + \frac{1}{n} \sum_{|k| \leq j} |\theta_k|^2 \left(\frac{1}{|\gamma_k|^2} - 1 \right),$$

Assuming that $s \geq 2\beta + 1$ and $f \in H_s(A)$, then the classical choice $\lambda^\star_k = 1_{k \leq j^\star}$ where $j^\star \sim n^{\frac{1}{2s+2\beta+1}}$ yields that

$$R(\theta, \lambda^\star) \sim \inf_{\lambda \in \Lambda} R(\theta, \lambda) \sim n^{\frac{-2s}{2s+2\beta+1}},$$

provided that $j^\star \leq m_0$. It can be checked that the choice (2.7) implies that $m_0 \sim n^{\frac{1}{2\beta}}$ and thus for a sufficiently large n,the condition $j^\star \leq m_0$ is satisfied since $n^{\frac{1}{2s+2\beta+1}} << n^{\frac{1}{2\beta}}$. From the lower bound obtained in Theorem 3.1 we conclude that the quadratic risk $\inf_{\lambda \in \Lambda} R(\theta, \lambda)$ decays asymptotically at the optimal (in the minimax sense) rate of convergence:

$$\sup_{f \in H_s(A)} \inf_{\lambda \in \Lambda} R(\theta, \lambda) \sim \sup_{f \in H_s(A)} \inf_{\lambda \in \Lambda} R(\theta, \lambda) \sim n^{\frac{-2s}{2s+2\beta+1}}.$$

Now, remark that for the estimator $\theta^\star_\alpha$ defined by (2.14), Theorems 2.1 yields that

$$\mathbb{E}_\theta \|\theta^\star_\alpha - \theta\|^2 = \mathcal{O}\left(\inf_{\lambda \in \Lambda} R(\theta, \lambda) \right) \text{ as } n \to +\infty, \text{ for any } 0 \leq \alpha \leq 1,$$

since it can be checked that in the case of projection filters, the additional terms in the upper bound (2.15) are of the order $\mathcal{O}(\frac{1}{n^{1-\zeta}})$ for a sufficiently small positive ζ. Thus, for any $0 \leq \alpha \leq 1$, the performances of the estimator $\theta^\star_\alpha$ is asymptotically optimal from the minimax convergence point of view.

4. Numerical experiments

The goal of this section on numerical experiments is to study the influence of the term α used in the definition (2.12) of the criterion $U_\alpha(Y, \lambda)$. For sake of simplicity, we study the case of projection filters $\Lambda = (\lambda^j)_{j\in\{1,\ldots,N\}}$, with $\lambda^j_k = \mathbb{1}_{|k|\leq j}, k \in \mathbb{Z}$ for $j = 1, \ldots, N$ and $N = m_0$ even if our experiments could be extended to more complex filters. In this case the choice of a filter amounts to choose a frequency cut-off level $1 \leq j \leq m_0$. For $\lambda^j \in \Lambda$ and $0 \leq \alpha \leq 1$, the criterion to minimize over $1 \leq j \leq m_0$ is

$$\begin{aligned} U_\alpha(Y, j) \quad := \quad & -\sum_{|k|\leq j} |\gamma_k|^{-2} \left\{ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right\} + \frac{\epsilon^2}{n} \sum_{|k|\leq j} |\gamma_k|^{-2} \\ & + \alpha \frac{\log^2 n}{n} \sum_{|k|\leq j} |\gamma_k|^{-4} \left\{ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right\}. \end{aligned}$$

For the mean pattern f to recover, we consider the three test functions shown in Figure 1. Then, for each test function, we simulate $n = 20$ 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 is then added to each curve. The level of the additive Gaussian noise is measured as the root of the signal-to-noise ratio ($rsnr$) defined as

$$rsnr = \left(\frac{\int_0^1 (f(x) - \bar{f})^2 dx}{\epsilon^2} \right)^{1/2} \quad \text{where } \bar{f} = \int_0^1 f(x)dx.$$

A sub-sample of 10 curves for $rsnr = 7$ is shown in Figure 1 for each test function. The Fourier coefficients of the density g are given by $\gamma_k = \frac{1}{1+2\sigma^2\pi^2k^2}$ which corresponds to a degree of ill-posedness $\beta = 2$. The condition (2.7) leads to the choice $m_0 = 27$. An example of estimation by spectral cut-off by minimizing the criterion $U_\alpha(Y, j)$ with $\alpha = 0$ is displayed in Figure 2. One can see that the obtained estimators are rather oscillatory suggesting that the selected frequency cut-off is somewhat too large when taking $\alpha = 0$.

These results illustrate the problem of choosing the value of α. To better understand the influence of this parameter, we present a short simulation study. The factors are the number of curves $n = 20, 50, 100$ and the signal-to-noise ratio $rsnr = 3, 7$. For each combination of these two factors, we generate $m = 1, \ldots, M$ (with $M = 100$) independent replications of the above described simulations. For each replication m we compute the estimator $\theta^*_{\alpha,m}$ for α ranging on a fine grid of $[0, 1]$. Then, since the template f and its Fourier coefficients

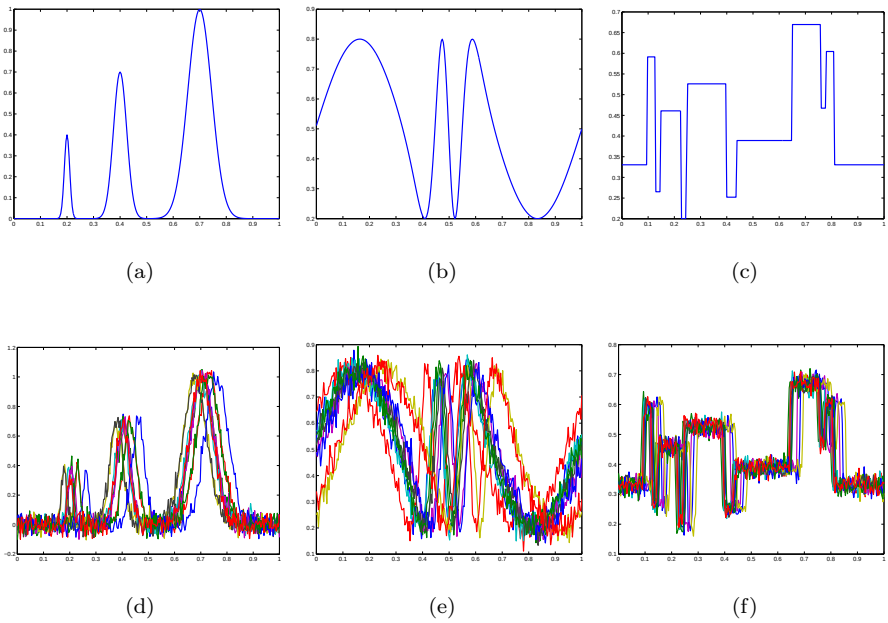


FIG 1. *Test functions and an example of randomly shifted curves. First line: (a) Bumps, (b) Sine, (c) Blocks. Second line: sample of 10 curves out of* $n = 20$ *for each test function.*

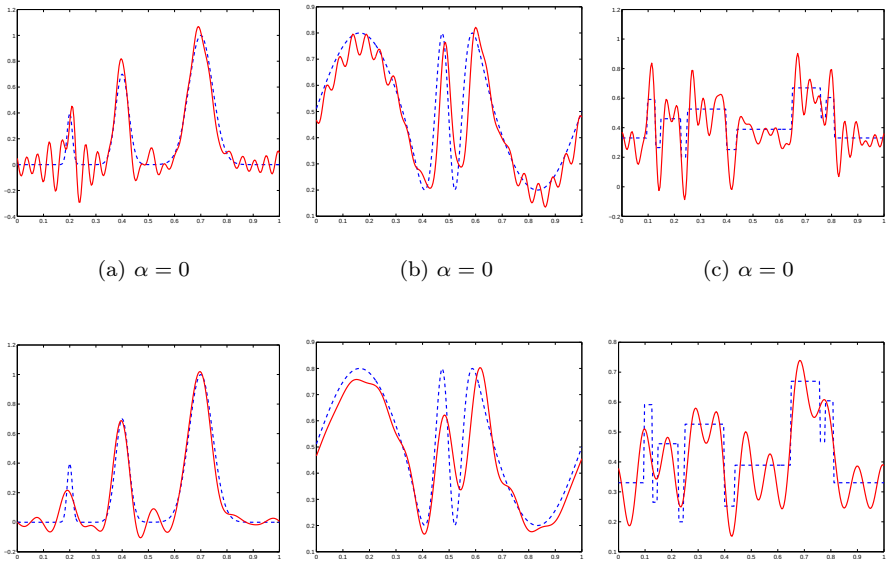
θ are known, one can compute for each value of α the following empirical mean squared error (MSE)

$$MSE(\alpha) = \frac{1}{M} \sum_{m=1}^{M} \|\theta^*_{\alpha,m} - \theta\|_2^2.$$

For each test function and each combination of the factors, we display in Figure 3 the curve $\alpha \to MSE(\alpha)$. The value α^* minimizing $\alpha \to MSE(\alpha)$ depends on the template to recover. These simulations show that α^* tends to be smaller as the number n of curves grows. The value of α^* is also closer to zero when the signal-to-noise ratio decreases (which corresponds to high values of ϵ). This confirms the heuristic arguments developed in Section 2. If the level of noise ϵ is large compared to $\|\theta\|_2^2$ (case of a low signal-to-noise ratio), then the model (1.1) is close to the standard white noise deconvolution problem. In this case, setting $\alpha = 0$ leads to satisfactory results which corresponds to taking the classical decomposition of the risk in standard inverse problems to do the estimation.

To conclude this section, let us consider the estimation by spectral cut-off by minimizing the criterion $U_\alpha(Y, j)$ with $\alpha = \alpha^*$ in the case $n = 20$ and $rsnr = 7$. This example is displayed in Figure 1. One can see that the obtained estimators are much smoother than those obtained with the choice $\alpha = 0$. This confirms

the importance of the choice of α. However, finding a data-based value for α is clearly challenging and is an interesting topic for future work.



(a) $\alpha = 0$ (b) $\alpha = 0$ (c) $\alpha = 0$

(d) $\alpha_* = 0.1212$ (e) $\alpha_* = 0.1919$ (f) $\alpha_* = 0.0808$

FIG 2. *An example of template estimation ($n = 20$ and $rsnr = 7$) with $\alpha = 0$ and $\alpha = \alpha_*$ for each test function.*

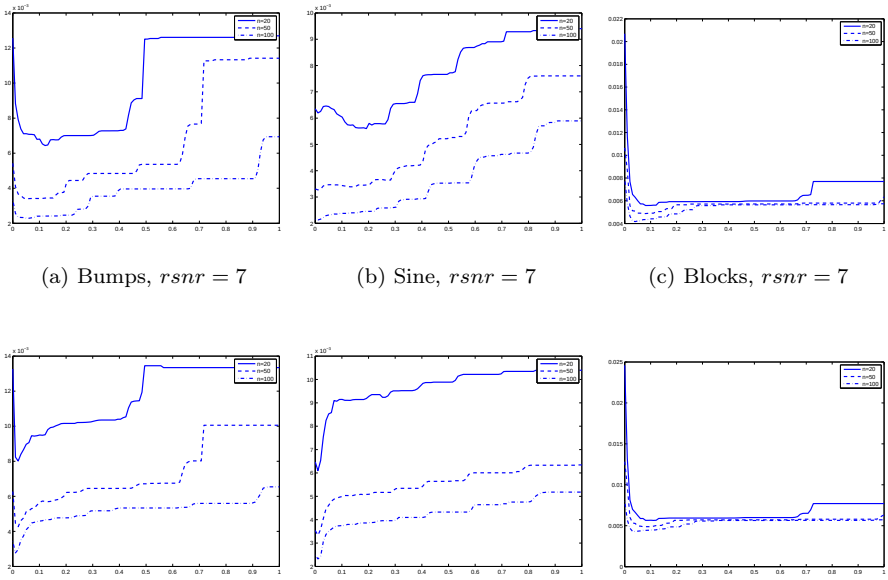
Appendix

This Appendix is divided in two parts. In the first part, we detail the scheme used for the proof of Theorem 2.1. The second part contains some technical lemmas. Throughout the proof, C denote a generic positive constant whose value may change from line to line. We provide first some short definitions which will be used in the sequel. In some sense, these terms measure the complexity associated to the set of filters Λ using the notations in [CGPT02].

Definition 4.1. *For each $\lambda \in \Lambda$, define*

$$\rho(\lambda) = \sup_{|k| \leq m_0} \frac{|\gamma_k|^{-2}\lambda_k}{\sqrt{\sum_{|i|\leq m_0} |\gamma_i|^{-4}\lambda_i^4}}, \quad \rho_\Lambda = \max_{\lambda\in\Lambda} \rho(\lambda),$$

$$S = \frac{\max_{\lambda\in\Lambda} \sup_{|i|\leq m_0} |\gamma_i|^{-2}\lambda_i^2}{\min_{\lambda\in\Lambda} \sup_{|i|\leq m_0} |\gamma_i|^{-2}\lambda_i^2}, \quad M = \sum_{\lambda\in\Lambda} e^{-1/\rho(\lambda)}, \text{ and } L_\Lambda = \log(NS) + \rho_\Lambda^2 \log^2(MS). \tag{4.1}$$



(a) Bumps, $rsnr = 7$ (b) Sine, $rsnr = 7$ (c) Blocks, $rsnr = 7$

(d) Bumps, $rsnr = 3$ (e) Sine, $rsnr = 3$ (f) Blocks, $rsnr = 3$

FIG 3. *Empirical Mean Square Error (MSE) over $M = 100$ simulations as a function of $\alpha \in [0, 1]$ for $n = 20, 50, 100$ and $rsnr = 7, 3$ for each test function: Bumps (first column), Sine (second Column), Blocks (third column).*

For a brief discussion on these quantities, we refer to [CGPT02]. For all $\lambda \in \Lambda$, we also introduce $R_\alpha(\theta, \lambda)$ as

$$R_\alpha(\theta,\lambda) = \sum_{|k|\le m_0} (1-\lambda_k)^2\theta_k^2 + \sum_{|k|\le m_0} \lambda_k^2|\gamma_k|^{-2} + \sum_{|k|>m_0} |\theta_k|^2 + \alpha\frac{\log^2(n)}{n}\sum_{|k|\le m_0} \lambda_k^2|\theta_k|^2|\gamma_k|^{-2}, \tag{4.2}$$

which corresponds to an approximation of the quadratic risk.

4.1. Proof of Theorem 2.1

The proof uses the following scheme. The first step consists in computing the quadratic risk of $\theta^\star$ and proving that it is close to $R_\alpha(\theta, \lambda^\star)$. The aim of the second part is to show that $U_\alpha(Y, \lambda^\star)$ is close to $R_\alpha(\theta, \lambda^\star)$, even for a random filter $\lambda^\star$. Then, we use the fact that $\lambda^\star$ minimizes the criterion $U_\alpha(Y, \lambda^\star)$ over the filters in Λ and we compute the expectation of $U_\alpha(Y, \lambda)$ for all deterministic λ in order to obtain an oracle inequality.

Step 1: remark that

$$\begin{aligned}
&\mathbb{E}_\theta\|\theta^\star-\theta\|^2\\
&\quad= \mathbb{E}_\theta\sum_{k\in\mathbb{Z}}|\theta_k^\star-\theta_k|^2,\\
&\quad= \mathbb{E}_\theta\sum_{|k|\le m_0}|\lambda_k^\star\gamma_k^{-1}\tilde{c}_k-\theta_k|^2+\sum_{|k|>m_0}|\theta_k|^2,\\
&\quad= \mathbb{E}_\theta\sum_{|k|\le m_0}\left|\left(\lambda_k^\star\frac{\tilde{\gamma}_k}{\gamma_k}-1\right)\theta_k+\lambda_k^\star\gamma_k^{-1}\frac{\epsilon}{\sqrt{n}}\xi_k\right|^2+\sum_{|k|>m_0}|\theta_k|^2,\\
&\quad= \mathbb{E}_\theta\sum_{|k|\le m_0}\left|\lambda_k^\star\frac{\tilde{\gamma}_k}{\gamma_k}-1\right|^2|\theta_k|^2+\frac{\epsilon^2}{n}\mathbb{E}_\theta\sum_{|k|\le m_0}\{\lambda_k^\star\}^2|\xi_k|^2|\gamma_k|^{-2}+\sum_{|k|>m_0}|\theta_k|^2\\
&\qquad\qquad +2\mathbb{E}_\theta\sum_{|k|\le m_0}\frac{\epsilon}{\sqrt{n}}Re\left[(\lambda_k^\star\gamma_k^{-1}\tilde{\gamma}_k-1)\theta_k\times\lambda_k^\star\bar{\gamma}_k^{-1}\bar{\xi}_k\right],
\end{aligned}$$

where for a given $z\in\mathbb{C}$, $Re(z)$ denotes the real part of z and $\bar{z}$ the conjugate of z. In the following, we denote by $\tilde{R}(\theta,\lambda)$ the commonly used risk in inverse problems, i.e.

$$\tilde{R}(\theta,\lambda):=\sum_{|k|\le m_0}\left|1-\lambda_k\frac{\tilde{\gamma}_k}{\gamma_k}\right||\theta_k|^2+\frac{\epsilon^2}{n}\sum_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2}+\sum_{|k|>m_0}|\theta_k|^2,\ \forall\lambda\in\Lambda.$$

Then $\mathbb{E}_\theta\|\theta^\star-\theta\|^2$ can be rewritten as

$$\begin{aligned}
\mathbb{E}_\theta\|\theta^\star-\theta\|^2 &= \mathbb{E}_\theta\tilde{R}(\theta,\lambda^\star)+\frac{\epsilon^2}{n}\mathbb{E}_\theta\sum_{|k|\le m_0}|\gamma_k|^{-2}\{\lambda_k^\star\}^2(|\xi_k|^2-1)\\
&\qquad +2\mathbb{E}_\theta\sum_{|k|\le m_0}\frac{\epsilon}{\sqrt{n}}Re\left((\lambda_k^\star\gamma_k^{-1}\tilde{\gamma}_k-1)\theta_k\times\lambda_k^\star\bar{\gamma}_k^{-1}\bar{\xi}_k\right),\\
&= \mathbb{E}_\theta\tilde{R}(\theta,\lambda^\star)+A_1+A_2. \qquad (4.3)
\end{aligned}$$

In order to bound A_1, we follow the notations of [CGPT02]. Let us define

$$\Delta(\lambda)=L_\Lambda\frac{\epsilon^2}{n}\sup_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2}\text{ and }\bar{\Delta}(\lambda)=\frac{\log^2(n)}{n}\sup_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2}\text{ for all }\lambda\in\Lambda, \qquad (4.4)$$

where L_Λ has been introduced in (4.1). Then, we apply the inequality (32) of [CGPT02]: there exists a universal constant C such that for any $\gamma>0$

$$\begin{aligned}
A_1 &:= \frac{\epsilon^2}{n}\mathbb{E}_\theta\sum_{|k|\le m_0}|\gamma_k|^{-2}\{\lambda_k^\star\}^2(|\xi_k|^2-1)\\
&\le \gamma\frac{\epsilon^2}{n}\mathbb{E}_\theta\sum_{|k|\le m_0}\{\lambda_k^\star\}^2|\gamma_k|^{-2}+C\gamma^{-1}\mathbb{E}_\theta\Delta(\lambda^\star). \qquad (4.5)
\end{aligned}$$

Now, consider a bound for A_2 defined as

$$A_2 := 2\mathbb{E}_\theta \sum_{|k|\leq m_0} \frac{\epsilon}{\sqrt{n}} Re\left((\lambda_k^\star \gamma_k^{-1}\tilde{\gamma}_k - 1)\theta_k \times \lambda_k^\star \bar{\gamma}_k^{-1}\bar{\xi}_k\right).$$

We apply inequality (31) of [CGPT02] to obtain for any $\gamma > 0$

$$A_2 \leq \gamma \mathbb{E}_\theta \sum_{|k|\leq m_0} \left|1 - \lambda_k^\star \tilde{\gamma}_k \gamma_k^{-1}\right|^2 |\theta_k|^2 + C\gamma^{-1}\mathbb{E}_\theta \Delta(\lambda^\star). \tag{4.6}$$

Now, for all $\gamma > 0$, inequalities (4.3), (4.5) and (4.6) yield

$$\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \leq (1+\gamma)\,\mathbb{E}_\theta \tilde{R}(\theta, \lambda^\star) + C\gamma^{-1}\mathbb{E}_\theta \Delta(\lambda^\star). \tag{4.7}$$

for some positive constant C. At last, we show that $\tilde{R}(\theta, \lambda^\star)$ is close to $R_\alpha(\theta, \lambda^\star)$ defined in (4.2). Remark that

$$\begin{aligned}
&\mathbb{E}_\theta \tilde{R}(\theta, \lambda^\star) \\
&= \mathbb{E}_\theta \sum_{|k|\leq m_0} (1-\lambda_k^\star)^2 |\theta_k|^2 + \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\gamma_k|^{-2} + \mathbb{E}_\theta \sum_{|k|>m_0} |\theta_k|^2 \\
&\qquad + \mathbb{E}_\theta \sum_{|k|\leq m_0} \left[|1-\lambda_k^\star \tilde{\gamma}_k \gamma_k^{-1}|^2 - (1-\lambda_k^\star)^2\right] |\theta_k|^2, \\
&= \mathbb{E}_\theta R_0(\theta, \lambda^\star) + \underbrace{\mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 \left|\frac{\tilde{\gamma}_k}{\gamma_k} - 1\right|^2}_{:=B_1} \\
&\qquad \underbrace{+2\,\mathbb{E}_\theta \sum_{|k|\leq m_0} \lambda_k^\star (1-\lambda_k^\star) Re\left(1 - \frac{\tilde{\gamma}_k}{\gamma_k}\right) |\theta_k|^2}_{:=B_2}.
\end{aligned}$$

First, we apply the Lemma 4.1 with $K = \gamma$ in order to bound B_1. We obtain

$$B_1 \leq \gamma \frac{\log^2 n}{n} \mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\gamma_k|^{-2} |\theta_k|^2 + Ce^{-\gamma \log^{1+\tau}(n)},$$

for some $\tau > 0$. Concerning B_2, we use the inequality $2ab \leq \gamma a + \gamma^{-1} b$ for all $\gamma > 0$ and Lemma 4.1 with $K = \gamma^2$ in order to obtain

$$\begin{aligned}
B_2 &= 2\mathbb{E}_\theta \sum_{|k|\leq m_0} \lambda_k^\star (1-\lambda_k^\star) Re\left(1 - \gamma_k^{-1}\tilde{\gamma}_k\right) |\theta_k|^2, \\
&\leq \gamma \mathbb{E}_\theta \sum_{|k|\leq m_0} (1-\lambda_k^\star)^2 |\theta_k|^2 + \gamma^{-1}\mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\theta_k|^2 |\gamma_k|^{-2} |\gamma_k - \tilde{\gamma}_k|^2, \\
&\leq \gamma \mathbb{E}_\theta \sum_{|k|\leq m_0} (1-\lambda_k^\star)^2 |\theta_k|^2 + \gamma \frac{\log^2(n)}{n} \mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\theta_k|^2 |\gamma_k|^{-2} + Ce^{-\gamma^2 \log^{1+\tau}(n)}.
\end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
&\mathbb{E}_\theta \tilde{R}(\theta,\lambda^\star)\\
&\leq (1+\gamma)\mathbb{E}_\theta R_0(\theta,\lambda^\star) + 2\gamma\frac{\log^2(n)}{n}\mathbb{E}_\theta \sum_{|k|\leq m_0}\{\lambda_k^\star\}^2|\theta_k|^2|\gamma_k|^{-2} + Ce^{-\gamma^2\log^{1+\tau}(n)},\\
&\leq (1+2\gamma)\mathbb{E}_\theta R_\alpha(\theta,\lambda^\star) + (1-\alpha)\gamma\|\theta\|^2\mathbb{E}_\theta\bar{\Delta}(\lambda^\star) + Ce^{-\gamma^2\log^{1+\tau}(n)}.
\end{aligned}$$

Using (4.7), we get

$$\begin{aligned}
\mathbb{E}_\theta\|\theta^\star-\theta\|^2 \leq\ & (1+2\gamma)^2\mathbb{E}_\theta R_\alpha(\theta,\lambda^\star) + C\gamma^{-1}\mathbb{E}_\theta\Delta(\lambda^\star)\\
& + C(1-\alpha)\gamma\|\theta\|^2\mathbb{E}_\theta\bar{\Delta}(\lambda^\star) + Ce^{-\gamma^2\log^{1+\tau}(n)}. \qquad (4.8)
\end{aligned}$$

This concludes the Step 1.

<u>Step 2:</u> First, we write $U_\alpha(Y,\lambda^\star)$ in terms of $R_\alpha(\theta,\lambda^\star)$. Remark that

$$\begin{aligned}
&U_\alpha(Y,\lambda^\star)\\
&= \sum_{|k|\leq m_0}(\{\lambda_k^\star\}^2 - 2\lambda_k^\star)|\gamma_k|^{-2}\left\{|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}\right\} + \frac{\epsilon^2}{n}\sum_{|k|\leq m_0}\{\lambda_k^\star\}^2|\gamma_k|^{-2}\\
&\qquad + \alpha\frac{\log^2(n)}{n}\sum_{|k|\leq m_0}\{\lambda_k^\star\}^2|\gamma_k|^{-4}\left\{|\tilde{c}_k| - \frac{\epsilon^2}{n}\right\},\\
&= R_\alpha(\theta,\lambda^\star) + \sum_{|k|\leq m_0}\left[(\{\lambda_k^\star\}^2 - 2\lambda_k^\star)|\gamma_k|^{-2}\left\{|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}\right\} - (1-\lambda_k^\star)^2\theta_k^2\right]\\
&\quad - \sum_{|k|\geq m_0}|\theta_k|^2 + \alpha\frac{\log^2(n)}{n}\sum_{|k|\leq m_0}\{\lambda_k^\star\}^2\left[|\gamma_k|^{-4}\left\{|\tilde{c}_k| - \frac{\epsilon^2}{n}\right\} - |\gamma_k|^{-2}|\theta_k|^2\right] \qquad (4.9)
\end{aligned}$$

Recall that for all $k\in\mathbb{N}$

$$|\tilde{c}_k|^2 = |\theta_k\tilde{\gamma}_k|^2 + \frac{\epsilon^2}{n}|\xi_k|^2 + 2\frac{\epsilon}{\sqrt{n}}Re(\theta_k\tilde{\gamma}_k\bar{\xi}_k),$$

and

$$|\gamma_k|^{-2}|\tilde{c}_k|^2 = |\theta_k|^2\left|\frac{\tilde{\gamma}_k}{\gamma_k}\right|^2 + \frac{\epsilon^2}{n}|\gamma_k|^{-2}|\xi_k|^2 + 2\frac{\epsilon}{\sqrt{n}}|\gamma_k|^{-2}Re(\theta_k\tilde{\gamma}_k\bar{\xi}_k).$$

Hence, equality (4.9) can be rewritten as

$$
\begin{aligned}
&R_\alpha(\theta,\lambda^\star)\\
&\quad= U_\alpha(Y,\lambda^\star)+\|\theta\|^2+\frac{\epsilon^2}{n}\sum_{|k|\le m_0}(2\lambda_k^\star-\{\lambda_k^\star\}^2)|\gamma_k|^{-2}(|\xi_k|^2-1)\\
&+\sum_{|k|\le m_0}(2\lambda_k^\star-\{\lambda_k^\star\}^2)\theta_k^2\left(\left|\frac{\tilde\gamma_k}{\gamma_k}\right|^2-1\right)+2\frac{\epsilon}{\sqrt{n}}\sum_{|k|\le m_0}(2\lambda_k^\star-\{\lambda_k^\star\}^2)|\gamma_k|^{-2}\mathrm{Re}(\tilde\gamma_k\theta_k\bar\xi_k)\\
&\qquad+\alpha\frac{\log^2(n)}{n}\sum_{|k|\le m_0}\{\lambda_k^\star\}^2\left[|\gamma_k|^{-2}|\theta_k|^2-|\gamma_k|^{-4}\left\{|\tilde c_k|^2-\frac{\epsilon^2}{n}\right\}\right],\\
&\quad= U_\alpha(Y,\lambda^\star)+\|\theta\|^2+C_1+C_2+C_3+C_4. \qquad (4.10)
\end{aligned}
$$

First consider the bound of C_1. Thanks to Lemma 4.2

$$
\begin{aligned}
C_1 &:= \sum_{|k|\le m_0}(2\lambda_k^\star-\{\lambda_k^\star\}^2)\theta_k^2\left(\left|\frac{\tilde\gamma_k}{\gamma_k}\right|^2-1\right),\\
&\le \gamma\mathbb{E}_\theta R_\alpha(\theta,\lambda^\star)+\left(\gamma+\frac{\gamma^{-1}}{\log^2(n)}\right)R_\alpha(\theta,\lambda^0)+(1-\alpha)\gamma\|\theta\|^2\mathbb{E}_\theta\Delta(\lambda^\star)\\
&\qquad+(1-\alpha)\gamma^{-1}\|\theta\|^2\bar\Delta(\lambda^0)+Ce^{-\gamma^2\log^{1+\tau}(n)}.
\end{aligned}
$$

Concerning C_2, we use the inequality (36) of [CGPT02]. We get

$$
\begin{aligned}
C_2 &:= \frac{\epsilon^2}{n}\sum_{|k|\le m_0}(2\lambda_k^\star-\{\lambda_k^\star\}^2)|\gamma_k|^{-2}(|\xi_k|^2-1),\\
&\le 2\gamma\frac{\epsilon^2}{n}\sum_{|k|\le m_0}\{\lambda_k^\star\}^2|\gamma_k|^{-2}+C\gamma^{-1}\mathbb{E}_\theta\Delta(\lambda^\star). \qquad (4.11)
\end{aligned}
$$

Then, using Lemma 4.3

$$
\begin{aligned}
C_3 &:= \frac{\epsilon}{\sqrt{n}}\sum_{|k|\le m_0}(2\lambda_k^\star-\{\lambda_k^\star\}^2)|\gamma_k|^{-2}\mathrm{Re}(\tilde\gamma_k\theta_k\bar\xi_k)\\
&\le 3\gamma\mathbb{E}_\theta R(\theta,\lambda^\star)+2\gamma R(\lambda^0,\theta)+C\gamma^{-1}\mathbb{E}_\theta\Delta(\lambda^\star)+C\gamma^{-1}\mathbb{E}_\theta\Delta(\lambda^0) \qquad (4.12)
\end{aligned}
$$

We are now interested in C_4, the last residual term of (4.10). Thanks to the

definition of $\tilde{c}_k$:

$$\begin{aligned}
C_4 &:= \frac{\log^2 n}{n}\mathbb{E}_\theta \sum_{|k|\le m_0} \{\lambda_k^\star\}^2|\gamma_k|^{-2}\left\{-|\gamma_k|^{-2}|\tilde{c}_k|^2 + \frac{\epsilon^2}{n}|\gamma_k|^{-2} + |\theta_k|^2\right\} \\
&= \frac{\log^2 n}{n}\mathbb{E}_\theta \sum_{|k|\le m_0} \{\lambda_k^\star\}^2|\gamma_k|^{-2}|\theta_k|^2\left(1-\left|\frac{\tilde{\gamma}_k}{\gamma_k}\right|^2\right) \\
&+\frac{\epsilon^2}{n}\frac{\log^2 n}{n}\mathbb{E}_\theta \sum_{|k|\le m_0} \{\lambda_k^\star\}^2|\gamma_k|^{-4}(1-|\xi_k|^2) - 2\frac{\log^2 n}{n}\frac{\epsilon}{\sqrt{n}}\mathbb{E}_\theta \sum_{|k|\le m_0} \{\lambda_k^\star\}^2|\gamma_k|^{-4}Re(\theta_k\tilde{\gamma}_k\bar{\xi}_k), \\
&\le D\gamma\mathbb{E}_\theta R_\alpha(\theta,\lambda^\star) + D\gamma R_\alpha(\theta,\lambda_0) \\
&\qquad +(1-\alpha)\gamma\|\theta\|^2\mathbb{E}_\theta\Delta(\lambda^\star) + (1-\alpha)\gamma^{-1}\|\theta\|^2\bar{\Delta}(\lambda^0) + Ce^{-\log^{1+\tau}(n)}, \qquad (4.13)
\end{aligned}$$

for some $D, C > 0$ independent of ϵ and n. Indeed, we can use essentially the same algebra as for the bound of the terms C_1, C_2 and C_3 and the inequality

$$|\gamma_k|^{-2} \le \frac{n}{\log^2 n},\ \forall k \le m_0.$$

Now, we are interested in the terms $\Delta(.)$ and $\bar{\Delta}(.)$ introduced in (4.4). For all $\lambda \in \Lambda$ and $x > 0$, we have

$$\begin{aligned}
&\sup_{|k|\le m_0} \lambda_k^2|\gamma_k|^{-2} \\
&\le \frac{1}{x}\sum_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2} + \sup_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2}\mathbf{1}_{\left\{x\sup_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2}\ge\sum_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2}\right\}} \\
&\le \frac{1}{x}\sum_{|k|\le m_0}\lambda_k^2|\gamma_k|^{-2} + \omega(x), \qquad (4.14)
\end{aligned}$$

where the function ω is introduced in Theorem 2.1. Hence, using (4.10)-(4.14) with a suitable choice for x,

$$\begin{aligned}
&(1-D\gamma)\mathbb{E}_\theta R_\alpha(\theta,\lambda^\star) \\
&\le \mathbb{E}_\theta U_\alpha(Y,\lambda^\star) + \|\theta\|^2 + D\left(\gamma + \frac{\gamma^{-1}}{\log^2(n)}\right)R_\alpha(\theta,\lambda^0) + Ce^{-\log^{1+\tau}(n)} \\
&\quad +C_1\frac{\epsilon^2}{n}L_\Lambda\omega\left((1-\alpha)\|\theta\|^2L_\Lambda\gamma^{-1}\right) + C_2\frac{\log^2 n}{n}\omega\left((1-\alpha)\|\theta\|^2\log^2(n)\gamma^{-1}\right)
\end{aligned}$$

<u>Step 3:</u> From the definition of $\lambda^\star$, we immediately get

$$\begin{aligned}
&(1-D\gamma)\mathbb{E}_\theta R_\alpha(\theta,\lambda^\star) \\
&\le \mathbb{E}_\theta U_\alpha(Y,\lambda^0) + \|\theta\|^2 + D\left(\gamma + \frac{\gamma^{-1}}{\log^2(n)}\right)R_\alpha(\theta,\lambda^0) + Ce^{-\log^{1+\tau}(n)} \\
&\quad +C_1\frac{\epsilon^2}{n}L_\Lambda\omega\left((1-\alpha)\|\theta\|^2L_\Lambda\gamma^{-1}\right) + C_2\frac{\log^2 n}{n}\omega\left((1-\alpha)\|\theta\|^2\log^2(n)\gamma^{-1}\right),
\end{aligned}$$

where λ_0 denotes the oracle bandwidth.

Step 4: Using that for all $|k| \leq m_0$,

$$\mathbb{E}_\theta\left(\hat{\Theta}_k^2\right) = |\theta_k|^2\left(1 - \frac{1}{n} + \frac{1}{n\gamma_k^2}\right) \leq |\theta_k|^2\left(1 - \frac{1}{n} + \frac{1}{\log^2(n)}\right)$$

it follows that

$$\mathbb{E}_\theta U_\alpha(Y, \lambda^0) \leq \left(1 + \frac{1}{\log^2(n)}\right)\left(R_\alpha(\theta, \lambda^0) - \|\theta\|^2\right).$$

Using (4.8) and step 3 of the proof, we get for γ small enough the results of Theorem 2.1. □

4.2. Technical Lemmas

Lemma 4.1. *For all $K > 0$, we have*

$$\mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 \left|\frac{\tilde{\gamma}_k}{\gamma_k} - 1\right|^2 |\theta_k|^2 \leq K\frac{\log^2(n)}{n}\mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\gamma_k|^{-2}|\theta_k|^2 + Ce^{-K\log^{1+\tau}(n)},$$

where C, τ denote positive constants independent of ϵ and n.

PROOF. Let $Q > 0$ a deterministic term which will be chosen later.

$$\begin{aligned}\mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 \left|\frac{\tilde{\gamma}_k}{\gamma_k} - 1\right|^2 |\theta_k|^2 &= \mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\theta_k|^2|\gamma_k|^{-2}|\tilde{\gamma}_k - \gamma_k|^2, \\ \leq\ & Q\mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\theta_k|^2|\gamma_k|^{-2} + \mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\theta_k|^2|\gamma_k|^{-2}\left\{|\tilde{\gamma}_k - \gamma_k|^2 - Q\right\}\mathbb{1}_{\{|\tilde{\gamma}_k-\gamma_k|^2>Q\}}.\end{aligned}$$

Thanks to (2.7) and the monotonicity of λ, we have

$$\begin{aligned}&\mathbb{E}_\theta \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\theta_k|^2|\gamma_k|^{-2}\left\{|\tilde{\gamma}_k - \gamma_k|^2 - Q\right\}\mathbb{1}_{\{|\tilde{\gamma}_k-\gamma_k|^2>Q\}} \\ &\leq\ C\frac{n}{\log^2(n)} \sum_{|k|\leq m_0} |\theta_k|^2\mathbb{E}_\theta\left\{|\tilde{\gamma}_k - \gamma_k|^2 - Q\right\}\mathbb{1}_{\{|\tilde{\gamma}_k-\gamma_k|^2>Q\}}.\end{aligned}$$

For all $|k| \leq m_0$, using an integration by part

$$\mathbb{E}_\theta\left[|\tilde{\gamma}_k - \gamma_k|^2 - Q\right]\mathbb{1}_{\{|\tilde{\gamma}_k-\gamma_k|^2>Q\}} = \int_Q^{+\infty} P(|\tilde{\gamma}_k - \gamma_k|^2 \geq x)dx.$$

Let $x \geq Q$. A Bernstein type inequality provides

$$\begin{aligned} P(|\tilde{\gamma}_k - \gamma_k|^2 \geq x) &= P\left(\left|\frac{1}{n}\sum_{l=1}^{n}\left\{e^{-2i\pi k\tau_l} - \mathbb{E}[e^{-2i\pi k\tau_l}]\right\}\right| \geq \sqrt{x}\right), \\ &\leq 2\exp\left\{-\frac{(n\sqrt{x})^2}{2\sum_{l=1}^{n}\mathrm{Var}(e^{-2i\pi k\tau_l}) + n\sqrt{x}/3}\right\}, \\ &\leq 2\exp\left\{-\frac{n^2 x}{2n + n\sqrt{x}/3}\right\}. \end{aligned}$$

Hence, for all $|k| \leq m_0$,

$$\begin{aligned} &\mathbb{E}_\theta\left[|\tilde{\gamma}_k - \gamma_k|^2 - Q\right]\mathbb{1}_{\{|\tilde{\gamma}_k - \gamma_k|^2 > Q\}} \\ &\leq \int_Q^{+\infty}\exp\left\{-\frac{nx}{2+\sqrt{x}/3}\right\}dx, \\ &\leq \int_Q^{36}\exp\left\{-\frac{nx}{4}\right\}dx + \int_{36}^{+\infty}\exp\left\{-Cn\sqrt{x}\right\}dx \leq \frac{C}{n}e^{-Qn/4}, \end{aligned}$$

where C denotes a positive constant independent of Q. Let $K > 0$. Choosing for instance $Q = n^{-1}K\log^2(n)$, we obtain

$$\begin{aligned} &\mathbb{E}_\theta\sum_{|k|\leq m_0}\{\lambda_k^\star\}^2\left|\frac{\tilde{\gamma}_k}{\gamma_k} - 1\right|^2|\theta_k|^2 \\ &\leq K\frac{\log^2(n)}{n}\mathbb{E}_\theta\sum_{|k|\leq m_0}\{\lambda_k^\star\}^2|\gamma_k|^{-2}|\theta_k|^2 + C\frac{nm_0}{\log^2(n)}e^{-K\log^2(n)/4}, \\ &\leq K\frac{\log^2(n)}{n}\mathbb{E}_\theta\sum_{|k|\leq |m_0|}\{\lambda_k^\star\}^2|\gamma_k|^{-2}|\theta_k|^2 + Ce^{-K\log^{1+\tau}(n)}, \end{aligned}$$

where C, τ denote positive constants independent of ϵ and n. This ends the proof of Lemma 4.1.

□

Lemma 4.2. *Let $\lambda^\star$ defined in (2.13). For all deterministic filter λ and $0 < \gamma < 1$, we have*

$$\begin{aligned} &\mathbb{E}_\theta\sum_{|k|\leq m_0}|\theta_k|^2\{\lambda_k^\star\}^2\left(\left|\frac{\tilde{\gamma}_k}{\gamma_k}\right|^2 - 1\right) \\ &\leq \gamma\mathbb{E}_\theta R_\alpha(\theta,\lambda^\star) + \left(\gamma + \frac{\gamma^{-1}}{\log^2(n)}\right)R_\alpha(\theta,\lambda^0) \\ &\quad + (1-\alpha)\gamma^{-1}\|\theta\|^2\bar{\Delta}(\lambda^0) + (1-\alpha)\gamma\|\theta\|^2\mathbb{E}_\theta\bar{\Delta}(\lambda^\star) + Ce^{-\gamma^2\log^{1+\tau}(n)}. \end{aligned}$$

where C, τ denote positive constants independent of ϵ and n.

PROOF. First, remark that

$$\begin{aligned}
&\mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 \left(\left| \frac{\tilde{\gamma}_k}{\gamma_k} \right|^2 - 1 \right) \\
&= \mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 |\gamma_k|^{-2} (|\tilde{\gamma}_k - \gamma_k + \gamma_k|^2 - |\gamma_k|^2), \\
&= \mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2 \\
&\quad + 2\mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k) \qquad (4.15)
\end{aligned}$$

Let $\lambda \in \Lambda$ be a deterministic filter, since $\mathbb{E}_\theta \tilde{\gamma}_k = \gamma_k$ for all $k \in \mathbb{N}$, we can write that

$$\begin{aligned}
&\mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k) \\
&= \mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 (\{\lambda_k^\star\}^2 - \lambda_k^2) |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k),
\end{aligned}$$

and simple algebra yields

$$|\{\lambda_k^\star\}^2 - \lambda_k^2| \leq \lambda_k^\star |1 - \lambda_k^\star| + \lambda_k |1 - \lambda_k| + \lambda_k^\star |1 - \lambda_k| + \lambda_k |1 - \lambda_k^\star|, \forall k \in \mathbb{N}.$$

Next, the Young inequality implies that for all $\gamma \in]0;1]$:

$$\begin{aligned}
&\mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \left| \{\lambda_k^\star\}^2 - \lambda_k^2 \right| |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k) \qquad (4.16) \\
&\leq \gamma^{-1} \sum_{|k|\leq m_0} \left[[\{\lambda_k^\star\}^2 + \{\lambda_k\}^2] \right] |\theta_k|^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2 \\
&\qquad + \gamma \sum_{|k|\leq m_0} \left[|1 - \lambda_k|^2 + |1 - \lambda_k^\star|^2 \right] |\theta_k|^2.
\end{aligned}$$

Hence, from equations (4.15) and (4.16), we obtain

$$\begin{aligned}
&\mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 \left(\left| \frac{\tilde{\gamma}_k}{\gamma_k} \right|^2 - 1 \right) \\
&\leq (1 + \gamma^{-1}) \mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2 \\
&+ \gamma^{-1} \mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \lambda_k^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2 + \gamma \mathbb{E}_\theta \sum_{|k|\leq m_0} \left[|1 - \lambda_k|^2 + |1 - \lambda_k^\star|^2 \right] |\theta_k|^2.
\end{aligned}$$

A direct application of Lemma 4.1 provides, for all $K > 0$

$$\mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \left(\left| \frac{\tilde{\gamma}_k}{\gamma_k} \right|^2 - 1 \right) \leq (1+\gamma^{-1}) K \frac{\log^2(n)}{n} \mathbb{E}_\theta \sum_{|k|\leq m_0} |\theta_k|^2 \{\lambda_k^\star\}^2 |\gamma_k|^{-2}$$
$$+ \frac{\gamma^{-1}}{n} \sum_{|k|\leq m_0} \lambda_k^2 |\theta_k|^2 |\gamma_k|^{-2} (1 - |\gamma_k|^2)$$
$$+ 2\gamma \mathbb{E}_\theta \sum_{|k|\leq m_0} [|1-\lambda_k|^2 + |1-\lambda_k^\star|^2] |\theta_k|^2 + Ce^{-K \log^{1+\tau}(n)}.$$

Fix $K = \gamma^2$ and remark that

$$\sum_{|k|\leq m_0} |\gamma_k|^{-2} |\theta_k|^2 \lambda_k^2 \leq \|\theta\|^2 \sup_{|k|\leq m_0} \lambda_k^2 |\gamma_k|^{-2}, \forall \lambda \in \Lambda,$$

in order to conclude the proof of Lemma 4.2.

□

Lemma 4.3. *Let $\lambda^\star$ the filter defined in (2.13). For all deterministic filter λ and $0 < \gamma < 1$, we have*

$$\frac{2\epsilon}{\sqrt{n}} \mathbb{E}_\theta \sum_{|k|\leq m_0} \left(2\lambda_k^\star - \{\lambda_k^\star\}^2\right) |\gamma_k|^{-2} Re(\theta_k \tilde{\gamma}_k \bar{\xi}_k) \leq 2\gamma R(\theta, \lambda)$$
$$+ 3\gamma \mathbb{E}_\theta R(\theta, \lambda^\star) + C\gamma^{-1} \mathbb{E}_\theta \Delta(\lambda^\star) + C\gamma^{-1} \mathbb{E}_\theta \Delta(\lambda) + Ce^{-\log^{1+\tau}(n)},$$

for some $\tau > 0$.

PROOF. In the following, we first state the inequality

$$P\left(\bigcap_{k=1}^{m_0} \left\{ \left| \frac{\tilde{\gamma}_k}{\gamma_k} \right| \leq 2 \right\} \right) \geq 1 - Cm_0 \exp(-\log^2 n),$$

for some $\tau > 0$. Indeed

$$\begin{aligned} P\left(\bigcup_{k=1}^{m_0} \left\{ \left| \frac{\tilde{\gamma}_k}{\gamma_k} \right| > 2 \right\} \right) &\leq \sum_{k=1}^{m_0} P\left(\left| \frac{\tilde{\gamma}_k}{\gamma_k} \right| > 2 \right) \\ &\leq \sum_{k=1}^{m_0} P\left(|\tilde{\gamma}_k - \gamma_k| > |\gamma_k| \right), \\ &\leq \sum_{k=1}^{m_0} P\left(|\tilde{\gamma}_k - \gamma_k| > \frac{\log^2(n)}{n} \right), \\ &\leq Cm_0 \exp(-\log^2 n). \end{aligned}$$

Then, for all $\gamma > 0$, using the above result and inequality (35) of [CGPT02], we obtain

$$\begin{aligned}
&\frac{2\epsilon}{\sqrt{n}}\mathbb{E}_\theta \sum_{|k|\leq m_0} \left(2\lambda_k^\star - \{\lambda_k^\star\}^2\right)|\gamma_k|^{-2} Re(\theta_k \tilde{\gamma}_k \bar{\xi}_k) \\
&\leq \gamma \left\{ \sum_{|k|\leq m_0} (1-\lambda_k)^2 |\theta_k|^2 + \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq m_0} \lambda_k^2 |\gamma_k|^{-4} |\tilde{\gamma}_k|^2 \right\} + C\gamma^{-1}\mathbb{E}_\theta \Delta(\lambda) \\
&\quad + \gamma \mathbb{E}_\theta \left\{ \sum_{|k|\leq m_0} (1-\lambda_k^\star)^2 |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\gamma_k|^{-4} |\tilde{\gamma}_k|^2 \right\} + C\gamma^{-1}\mathbb{E}_\theta \Delta(\lambda^\star), \\
&\leq 4\gamma \left\{ \sum_{|k|\leq m_0} (1-\lambda_k)^2 |\theta_k|^2 + \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq m_0} \lambda_k^2 |\gamma_k|^{-2} \right\} + Ce^{-\log^{1+\tau}(n)} \\
&+ 4\gamma \mathbb{E}_\theta \left\{ \sum_{|k|\leq m_0} (1-\lambda_k^\star)^2 |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k|\leq m_0} \{\lambda_k^\star\}^2 |\gamma_k|^{-2} \right\} + C\gamma^{-1}\mathbb{E}_\theta \Delta(\lambda^\star) + C\gamma^{-1}\mathbb{E}_\theta \Delta(\lambda).
\end{aligned}$$

This concludes the proof.

□

References

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

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

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

[BL96] Lawrence D. Brown and Mark G. Low. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6):2384–2398, 1996.

[CGPT02] L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov. Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3):843–874, 2002. Dedicated to the memory of Lucien Le Cam.

[CH05] L. Cavalier and N.W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21(4):1345–1361, 2005.

[CL09] I. Castillo and J.-M. Loubes. Estimation of the distribution of random shifts deformation. *Math. Methods Statist.*, 18(1):21–42, 2009.

[CR07] L. Cavalier and M. Raimondo. Wavelet deconvolution with noisy eigenvalues. *IEEE Trans. on Signal Processing*, 55:2414–2424, 2007.

[CT02] L. Cavalier and A.B. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, (123):323–354, 2002.

[Dal07] A. Dalalyan. Penalized maximum likelihood and semiparametric second-order efficiency. *Math. Methods of Statist.*, 16(1):43–63, 2007.

[DGT06] A. Dalalyan, G.K. Golubev, and A. B. Tsybakov. Penalized maximum likelihood and semiparametric second-order efficiency. *Ann. Statist.*, 34(1):169–201, 2006.

[DJKP95] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57:301–369, 1995.

[Don95] D.L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2(2):101–126, 1995.

[EK01] S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.

[Fan91] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19:1257–1272, 1991.

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

[GK95] T. Gasser and A. Kneip. Searching for structure in curve samples. *JASA*, 90(432):1179–1188, 1995.

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

[HR08] M. Hoffmann and M. Reiß. Nonlinear estimation for linear inverse problems with error in the operator. *Annals of Statistics*, 38:310–336, 2008.

[IRT08] U. Isserles, Y. Ritov, and T. Trigano. Semiparametric density estimation of shifts between curves. *preprint*, 2008.

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

[LM04] X. Liu and H.G.. M§ller. Functional convex averaging and synchronization for time-warped random curves. *JASA*, 99(467):687–699, 2004.

[Mar06] C. Marteau. Regularization of inverse problems with unknown operator. *Mathematical Methods of Statistics*, 15:415–443, 2006.

[Mar09] C. Marteau. On the stability of the risk hull method for projection estimator. *Journal of Statistical Planning and Inference*, 139:1821–1835, 2009.

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

[Ron98] Birgitte B. Ronn. Nonparametric maximum likelihood estimation for shifted curves. *JRSS B*, 60:351–363, 1998.

[RS02] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Lecture Notes in Statistics, New York: Springer-Verlag, 2002.

[RS05] J.O. Ramsay and B.W. Silverman. *Applied Functional Data Analysis*. Lecture Notes in Statistics, New York: Springer-Verlag, 2005.

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

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