# Sharp template estimation in a shifted curves model

## Jérémie Bigot, Sébastien Gadat, Clément Marteau

# Sharp template estimation in a shifted curves model

*Jérémie Bigot, Sébastien Gadat & Clément Marteau*
Institut de Mathématiques de Toulouse
Université de Toulouse et CNRS (UMR 5219)
31062 Toulouse, Cedex 9, France
{Jeremie.Bigot,Sebastien.Gadat,Clement.Marteau}@math.univ-toulouse.fr,

November 9, 2009

### 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 stochastic linear inverse problem. 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.

*Keywords:* Template estimation, Curve alignment, Stochastic inverse problem, Oracle inequality, Adaptive estimation.

## 1 Introduction

### 1.1 Model and objectives

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

$$dY_j(x) = f(x - \tau_j)dx + \epsilon dW_j(x), \ x \in [0,1], \ j = 1, \ldots, n \qquad (1.1)$$

where $W_j$ are independent standard Brownian motions on $[0,1]$, $\epsilon$ represents a level of noise common to all curves, the $\tau_j$'s are unknown random shifts, $f$ 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 $\tau_j$ 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 $\mathbb{R}$. Estimating $f$

can be seen as a stochastic inverse problem as this template is not observed directly, but through $n$ independent realizations of the stochastic operator $A_\tau : L_p^2([0,1]) \to L_p^2([0,1])$ defined by

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

where $L_p^2([0,1])$ denotes the space of squared integrable functions on $[0,1]$ with period 1, and $\tau$ is random variable with density $g$. The additive Gaussian noise makes this problem ill-posed, and [BG09] 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 $[0,1]$, [BG09] have derived the minimax rate of convergence for the estimation of $f$ over Besov balls as $n$ tends to infinfity. 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 deterministic deconvolution problem in statistics, see e.g. [Fan91], [Don95], but the results in [BG09] represent a novel contribution and a new point of view on template estimation in stochastic inverse problems such as (1.1).

However, the approach followed in [BG09] 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.

### 1.1.1 Deconvolution formulation

Let us first explain how the model (1.1) can be transformed into a deconvolution problem as the one studied in [DJKP95]. Denote $G$ the following density function defined on $[0;1]$ as

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

The density $G$ exists as soon as $g$ satisfies the weak condition $g(x) \le \frac{C}{1+|x|^\nu}$ for any $\nu > 1$ and suitable constant $C$. Note that the Fourier coefficients of G are given by

$$\int_0^1 G(t)e^{-i2\pi lt} dt = \int_{-\infty}^\infty g(t)e^{-i2\pi lt} dt = \gamma_l$$

Consider now the 1-periodization of $f$ extended to $\mathbb{R}$, one has

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

The observations $Y_j$ can be written as

$$dY_j(x) = f \star G(x)dx + \xi_j(x)dx + \epsilon dW_j(x), \tag{1.2}$$

where $\xi_j$ is a second noise term defined as $\xi_j(x) = f(x - \tau_j) - f \star G(x)$. Hence, our model can be seen as a deconvolution problem with a noisy operator $H : f \mapsto f \star G + \xi$ and a more classical independent additive noise $W$. Note also that the realizations $H_j : f \mapsto f \star G + \xi_j$ are unbiased realizations of the operator $H$ but presents a variance term which depends on the function $f$ we want to estimate. This appears to be a new setting in the field of inverse problem with unknown operators as considered in [CH05], [EK01], [HR05], [Mar06] and [CR07].

We will see in the sequel that the additive noise $\xi$ which depends on $f$ slightly modifies the quadratic risk and the way to estimate $f$ when compared to classical procedures used in standard inverse problems with a deterministic operator.

## 1.2 Fourier Analysis and an inverse problem formulation

Supposing that $f \in L_p^2([0,1])$, we denote by $\theta_k$ its $k^{th}$ 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

$$c_{j,k} := \int_0^1 e^{-2ik\pi x} dY_j(x) = \theta_k e^{-i2\pi k\tau_j} + \epsilon z_{k,j} \tag{1.3}$$

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 $k^{th}$ Fourier coefficient over the $n$ curves as

$$\tilde{c}_k := \frac{1}{n} \sum_{j=1}^n c_{k,j} = \theta_k \tilde{\gamma}_k + \frac{\epsilon}{\sqrt{n}} \xi_k, \tag{1.4}$$

where

$$\tilde{\gamma}_k := \frac{1}{n} \sum_{j=1}^n e^{-i2\pi k\tau_j}, \tag{1.5}$$

and the $\xi_k$'s are i.i.d. complex Gaussian variables with zero mean and variance 1. The Fourier coefficients $\tilde{c}_k$ in equation (1.4) 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)

$$c_k = \theta_k \gamma_k + \sigma z_k, \tag{1.6}$$

where the $\gamma_k$'s are eigenvalues of a known linear operator, $z_k$ are random noise variables and $\sigma$ is a level of noise which goes to zero for studying asymptotic properties. The issue in such models is to recover the coefficients $\theta_k$ from the observations $c_k$ under various conditions on the decay to zero of the $\gamma_k$'s as $|k| \to +\infty$. A large class of estimators for the problem (1.6) 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 (1.4) as an inverse problem (with $\sigma = \frac{\epsilon}{\sqrt{n}}$) where the eigenvalues of the linear operator are the Fourier coefficients of the density $g$ of the shifts i.e.

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

Indeed, let us assume that the density $g$ of the random shifts is known. In this case, to estimate the Fourier coefficients of $f$, one can perform a deconvolution step of the form

$$\hat{\theta}_k = \lambda_k \frac{\tilde{c}_k}{\gamma_k}, \tag{1.7}$$

where $\tilde{c}_k$ is defined in (1.4) 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 $\lambda$ will be presented in the case where the coefficients $\gamma_k$ are known. Such a framework is commonly used in inverse problems such as (1.6) to obtain consistency results and to study asymptotic rates of convergence, where it is

3

generally supposed that the law of the additive error is Gaussian with zero mean and *known* variance $\sigma^2$, 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:

$$c_k = \gamma_k \theta_k + \epsilon \xi_k, \ \tilde{\gamma}_k = \gamma_k + \sigma \eta_k, \ \forall k \in \mathbb{N}, \tag{1.8}$$

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 $\theta_k$ 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 [HR05].

In this paper, our framework is sligthly 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 $f$ 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 [GK95], [GK92], [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 $\epsilon$ 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 [LMG07] and [Vim08] as the number of observations per curve grows, but with a fixed number $n$ of curves. A generalisation of this approach for the estimation of scaling, rotation and translation parameters for two-dimensional images is also proposed in [BGV08], 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 [CL08] 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 $f$ based on spectral cut-off when the eigenvalues $\gamma_k$ are known. Based on the principle of unbiased risk minimization developed by [CGPT02], we derive an oracle inequality that is then used to derive an adaptive estimator of $f$ on Sobolev spaces. This estimator is based on the Fourier transform of the curves with a data-based choice of the frequency cut-off. In Section 3, we study asymptotic properties of this estimator in terms of minimax rates of converge over Sobolev balls. Finally in Section 4, a short simulation study is proposed to illustrate the numerical properties of the estimator. 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 $\gamma_k$ are known. In this situation it is possible to choose a data-dependent filter $\lambda^\star$ that mimic the performances of an optimal filter $\lambda^0$ 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 $\lambda^0$ 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 $\gamma_k$'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 $k \in \mathbb{Z}$*

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

Remark that the knowledge of the constants $C_{max}, C_{min}$ and $\beta$ will not be necessary for the construction of our estimator.

## 2.2 Risk decomposition

Assuming that $\gamma_k \neq 0$ for all $k \in \mathbb{Z}$, we recall that an estimator of the $\theta_k$'s is given by, see equation (1.7)

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

where $\lambda = (\lambda_k)_{k \in \mathbb{Z}}$ is a real sequence. Examples of commonly used filters include projection weights $\lambda_k = \mathbb{1}_{|k| \leq N}$ for some integer $N$, 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.

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 $\lambda$ we use the classical $\ell_2$-norm to define the risk of the estimator $\hat{\theta}(\lambda) = (\hat{\theta}_k)_{k \in \mathbb{Z}}$

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

Note that analyzing the above risk (2.2) is equivalent to analyze the mean integrated square risk $R(\hat{f}_\lambda, f) = \mathbb{E}\|\hat{f}_\lambda - f\|^2 = \mathbb{E}\left(\int_0^1 (\hat{f}_\lambda(x) - f(x))^2 dx\right)$ for the estimator $\hat{f}_\lambda(x) = \sum_{k \in \mathbb{Z}} \hat{\theta}_k e^{-2ik\pi x}$. The following lemma gives the bias-variance decomposition of $R(\lambda, \theta)$.

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

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

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 tradeoff 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 $\theta_k$ of the unknown function $f$ to recover. Indeed, our data $\gamma_k^{-1} \tilde{c}_k$ are noisy observations of $\theta_k$:

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

and we invert the problem using the sequence $(\gamma_k)_{k \in \mathbb{N}}$ instead of $(\tilde{\gamma}_k)_{k \in \mathbb{N}}$, which is involved in the construction of the coefficient $c_k$. It explains the presence of the second term $V_2$. In particular, the quadratic risk is expressed in its usual form in the case where $\tilde{\gamma}_k = \gamma_k$.

A similar phenomenon occurs with the model (1.8), although it is more difficult to quantify. Indeed, in this setting:

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

Hence, we also observe an additionnal term depending on $\theta$. This term is controled using a Taylor expension but the quadratic risk cannot be expressed in a simple form. We refer to [Mar09] for a discussion with some numerical simulation and to [CH05], [EK01], [HR05], [Mar06] and [CR07].

## 2.3 An oracle estimator and unbiased estimation of the risk

Suppose that one is given a finite set of possible candidate filters $\Lambda = (\lambda^N)_{N \in I}$, with $\lambda^N = (\lambda_k^N)_{k \in \mathbb{Z}}$, $N \in I \subset \mathbb{N}$ which satisfy some general conditions to be discussed later on. In the case of projection filters, $\Lambda$ can be for example the set of filters $\lambda_k^N = \mathbb{1}_{|k| \leq N}, k \in \mathbb{Z}$ for $N = 1, \ldots, m_0$.

Given a set of filters $\Lambda$, the best estimator corresponds to the filter $\lambda^0$, called oracle, which minimizes the risk $R(\lambda, \theta)$ over $\Lambda$ i.e.

$$\lambda^0 := \arg\min_{\lambda \in \Lambda} R(\lambda, \theta). \tag{2.4}$$

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

Now, suppose that it is possible to construct an unbiased estimator $\hat{\Theta}_k^2$ of $|\theta_k|^2$. For any nonrandom filter $\lambda$, using $\hat{\Theta}_k^2$, one can compute an estimator $\tilde{U}(\lambda, X)$ of the risk $R(\lambda, \theta)$. Then, for choosing a data-dependent filter, the principle of unbiased risk estimation (see [CGPT02] for further details) simply suggests to minimize the criterion $U(\lambda, X)$ over $\lambda \in \Lambda$ instead of the criterion $R(\lambda, \theta)$. Our data-dependent choice of $\lambda$ is thus

$$\lambda^\star := \arg\min_{\lambda \in \Lambda} \tilde{U}(\lambda, X). \tag{2.5}$$

Typically, in practice, all the filters $\lambda \in \Lambda$ are such that $\lambda_k = 0$ (or vanishingly small) for all $k$ large enough. Hence, for such choices of filters, numerical computation of the above expression is thus feasible since it only involves the computation of finite sums.

## 2.4 Oracle inequalities for projection filters

### 2.4.1 Unbiased Risk Estimation (URE)

For the sake of simplicity, we only consider spectral cut-off schemes in the following. In this case, $\Lambda$ corresponds to the set of filters $(\mathbb{1}_{|k|\leq N})_{k\in\mathbb{Z}}$ for $N \in \mathbb{N}$. All the results presented in this paper could be generalized to wider families of estimators (Tikhonov, Landweber, Pinsker,...). The price to pay is to get longer and more technical proofs.

From Lemma 2.1, the quadratic risk $R(\theta, \lambda) := R(\theta, N)$ of a projection filter can be written as:

$$
\begin{aligned}
R(\theta, N) &= \sum_{|k|>N} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k|\leq N} |\gamma_k|^{-2} + \frac{1}{n} \sum_{|k|\leq N} |\theta_k|^2 \left( \frac{1}{|\gamma_k|^2} - 1 \right) \\
&= \|\theta\|_2^2 - \sum_{|k|\leq N} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k|\leq N} |\gamma_k|^{-2} + \frac{1}{n} \sum_{|k|\leq N} |\theta_k|^2 \left( \frac{1}{|\gamma_k|^2} - 1 \right)
\end{aligned}
$$

We aim to minimize $R$ with respect to $N$ while $\theta$ is unknown. Using $\hat{\Theta}_k^2 = \gamma_k^{-2} \left[ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right]$ as an unbiased estimator of $|\theta_k|^2$, we minimize U defined as

$$U(Y, N) = -\left(1 - \frac{1}{n}\right) \sum_{|k|\leq N} |\gamma_k|^{-2} \left\{ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right\} + \frac{\epsilon^2}{n} \sum_{|k|\leq N} |\gamma_k|^{-2} + \frac{1}{n} \sum_{|k|\leq N} |\gamma_k|^{-4} | \left\{ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right\}, \tag{2.6}$$

which is an unbiased risk estimator of $R(\theta, N) - \|\theta\|_2^2$.

Unfortunately, such a criterion does not lead to satisfying results. Instead of the approach developed in [CH05], we take into account the error generated by the use of an approximation of the eigenvalues. The estimator related to the criterion (2.6) involves processes that require a specific treatment. In order to contain these processes, we will consider in the following the criterion

$$\bar{U}(Y, N) = -\sum_{|k|\leq N} |\gamma_k|^{-2} \left\{ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right\} + \frac{\epsilon^2}{n} \sum_{|k|\leq N} |\gamma_k|^{-2} + \frac{\log^2(n)}{n} \sum_{|k|\leq N} |\gamma_k|^{-2} \left\{ |\bar{c}_k| - \frac{\epsilon^2}{n} \right\}, \tag{2.7}$$

Remark that $\bar{U}(Y, N)$ can be written as $U(Y, N) + \text{pen}(N)$ where $(\text{pen}(N))_{N \in \mathbb{N}}$ denotes a penalty term. It appears from the proofs that this penalty is a natural candidate for the control of the processes involved in the behavior of the estimator constructed below. The associated data-based filter is defined as

$$N^\star = \arg \min_{N \leq m_0} \bar{U}(Y, N), \tag{2.8}$$

where

$$m_0 = \inf \left\{ k : |\gamma_k|^2 \leq \frac{\log^2 n}{n} \right\} - 1. \tag{2.9}$$

Remark that we do not minimize our criterion $\bar{U}(Y, N)$ over $\mathbb{N}$ but rather for $N \leq m_0$. Indeed, each coefficient $\theta_k$ 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 constant in $k$, it seems clear that large $k$ should be avoided.

Similar bounds on the resolution level 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 $f$.

### 2.4.2 Sharp estimator of the risk

We are now able to propose a first adaptive estimator. In the following, we denote by $\theta^\star$ the estimator related to the bandwidth $N^\star$ namely

$$\theta_k^\star = \frac{\tilde{c}_k}{\gamma_k} \mathbb{1}_{\{k \leq N^\star\}}. \tag{2.10}$$

The next theorem summarizes the performances of $\theta^\star$ through a simple oracle inequality. The proof is postponed to the Section 5.

**Theorem 2.1** *Let $\theta^\star$ defined by (2.10) and assume that the density $g$ satisfies Assumption 2.1. Then, there exists $0 < \gamma_1 < 1$ such that, for all $0 < \gamma < \gamma_1$,*

$$\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \leq (1 + h_1(\gamma)) \inf_{N \leq m_0} \bar{R}(\theta, N) + \frac{C_1 \epsilon^2}{n} \frac{1}{\gamma^{4\beta + 1}} + \frac{C_1}{n\gamma}, \tag{2.11}$$

*where*

$$\bar{R}(\theta, N) = \sum_{|k| > N} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k| \leq N} |\gamma_k|^{-2} + \frac{\log^2(n)}{n} \sum_{|k| \leq N} |\gamma_k|^{-2} |\theta_k|^2, \tag{2.12}$$

*$h_1(\gamma) \to 0$ as $\gamma \to 0$ and $C_1$ denotes a positive constant independent of $\epsilon$ and $n$.*

From Theorem 2.1, our estimator $\theta^\star$ presents a behavior similar to the minimizer of $\bar{R}(\theta, N)$. This term only differs from the quadratic risk by a log term. This result can be explained by the choice of the criterion (2.7). The two last terms in the right hand side of (2.11) are at least of order $1/n$ and may be thus considered as negligible in most cases.

In the next section, we prove that our estimator attains the minimax of convergence on many functional spaces. In particular, the log term and the bandwidth $m_0$ have no influence on the performances of our estimator from a minimax point of view.

### 2.4.3 Rough estimator

In the procedure described above, we have decided to take into account the error generated by the use of a the sequence $(\gamma_k)_{k \in \mathbb{N}}$ instead of $(\tilde{\gamma}_k)_{k \in \mathbb{N}}$. Although their setting is slightly different from ours, papers dealing with regularization with unknown operator consider implicitly this

error as negligible for the regularization. The goal is then to prove that the related estimator are not affected by the noise in the operator, i.e. this error is avoided in the oracle.

It is thus also possible to apply a similar scheme in our setting and consider the bias enlightened in Lemma 2.1 as negligible. We introduce

$$\tilde{R}(\theta, N) = \sum_{|k|>N} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k|\leq N} |\gamma_k|^{-2}, \tag{2.13}$$

that corresponds to the usual quadratic risk in an inverse problems setting.

From now on, our aim is to mimic the oracle for $\tilde{R}(\theta, N)$, i.e

$$\tilde{N}_0 = \arg \min_{N \in \mathbb{N}} \tilde{R}(\theta, N).$$

To this end, we use exactly the same scheme than for the construction of $\theta^\star$ starting from $\tilde{R}(\theta, N)$ instead of $R(\theta, N)$. Define

$$\tilde{U}(Y, N) = - \sum_{|k|\leq N} |\gamma_k|^{-2} \left\{ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right\} + \frac{\epsilon^2}{n} \sum_{|k|\leq N} |\gamma_k|^{-2}. \tag{2.14}$$

Then, we introduce

$$\tilde{N} = \arg \min_{N \leq m_0} \tilde{U}(Y, N) \text{ and } \tilde{\theta}_k = \frac{\tilde{c}_k}{\gamma_k} \mathbb{1}_{\{k \leq \tilde{N}\}}, \tag{2.15}$$

where $m_0$ has been introduced in (2.9). Hence, this estimator only differs from the previous one by the choice of the regularization parameter $\tilde{N}$. The performances of $\tilde{\theta}$ are detailed bellow.

**Theorem 2.2** *Let $\tilde{\theta}$ defined by (2.15) and assume that the density $g$ satisfies Assumption 2.1. Then, there exists $0 < \gamma_2 < 1$ such that, for all $0 < \gamma < \gamma_2$,*

$$\mathbb{E}_\theta \|\tilde{\theta} - \theta\|^2 \leq (1 + h_2(\gamma)) \inf_{N \leq m_0} R(\theta, N) + \frac{C_2 \epsilon^2}{n} \left( \frac{\|\theta\|^2 \log^2(n)}{\gamma^2} \right)^{2\beta} + \frac{C_2 \epsilon^2}{n} \frac{1}{\gamma^{4\beta+1}} + \frac{C_2}{n}, \tag{2.16}$$

*where $h_2(\gamma) \to 0$ as $\gamma \to 0$ and $C_2$ denote a positive constant independent of $\epsilon$ and $n$.*

We will see in Section 3 that the performances of $\theta^\star$ and $\tilde{\theta}$ are essentially the same from a minimax point of view. The existing differences may be revealed by the comparison of the oracle inequalities obtained in Theorems 2.1 and 2.2, although this is always a difficult task. Since $\bar{R}(\theta, N)$ only differs from $R(\theta, N)$ by a log term, we may be interested in the residual of order $\|\theta\|^2$. For fixed $\epsilon$ and $n$, this term may have importance compared to $R(\theta, N)$, in particular for large $\|\theta\|^2$. Hence, the second estimator may be incongruous when estimating function with large norm.

More carefully, $\tilde{\theta}$ is a pertinent choice as soon as $\tilde{R}(\theta, N)$ is close to $R(\theta, N)$. This can be strengthened by the study of the quadratic risk defined in Lemma 2.1. For instance, with a fixed $\epsilon$, this will be the case for function with 'small' Fourier coefficients (in particular small norms). On the other hand, as soon as $\epsilon$ becomes 'small', the behaviour of $\tilde{R}(\theta, N)$ and $R(\theta, N)$ may strongly differs. This may produce significant differences on the performances of both $\theta^\star$ and $\tilde{\theta}$.

## 3   Minimax rates of convergence for Sobolev balls

We provide in this section a short discussion about the performances of our estimator from the asymptotic minimax point of view. For this, let $1 \leq p, q \leq \infty$ and $A > 0$, and suppose that $f$ belongs to a Besov ball $\mathcal{B}_{p,q}^s(A)$ of radius $A$ (see e.g. [DJKP95] for a precise definition of Besov spaces). [BG09] 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:*

- *(Regularity condition on $f$) $f \in \mathcal{B}_{p,q}^s(A)$ and $s \geq p'$,*

- *(Regularity condition on $g$) $g$ satisfies the polynomial decay condition (2.1) at rate $\beta$ for its Fourier coefficients,*

- *(Dense case) $s \geq (2\beta + 1)(1/p - 1/2)$ and $s \geq 2\beta + 1$.*

*Then, there exists a universal constant $M_1$ depending on $A, s, p, q$ such that*

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

*where $\hat{f}_n \in L_p^2([0,1])$ 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 (1.2). Then, let us introduce the following smoothness class of functions which can be identified with a periodic Sobolev ball:

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

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 $\mathcal{B}_{2,2}^s(A')$. Assuming $f \in H_s(A)$ with $s > 0$, then the classical choice $N^\star \sim n^{\frac{1}{2s+2\beta+1}}$ yields that

$$R(\theta, N^\star) \sim \inf_{N \leq m_0} R(\theta, N) \sim n^{\frac{-2s}{2s+2\beta+1}}.$$

provided $N^\star \leq m_0$. It can be checked that the choice (2.9) implies that $m_0 \sim n^{\frac{1}{2\beta}}$ and thus for a sufficiently large $n$, we have that $N^\star < m_0$. Similarly the choice $N^\star \sim n^{\frac{1}{2s+2\beta+1}}$ yields that

$$\bar{R}(\theta, N^\star) \sim \inf_{N \leq m_0} \bar{R}(\theta, N^\star) \sim \log^2(n) n^{\frac{-2s}{2s+2\beta+1}},$$

Now, remark that for the two estimators $\theta^\star$ and $\tilde{\theta}$, both Theorems 2.1 and 2.2 yield that $\mathbb{E}_\theta \|\theta^\star - \theta\|^2 = \mathcal{O}\left(\inf_{N \leq m_0} \bar{R}(\theta, N)\right)$ and $\mathbb{E}_\theta \|\tilde{\theta} - \theta\|^2 = \mathcal{O}\left(\inf_{N \leq m_0} R(\theta, N)\right)$ as $n \to +\infty$, since additional terms in bounds (2.11) and (2.16) are of the order $\mathcal{O}(\frac{1}{n^{1-\zeta}})$ for a sufficiently small positive $\zeta$. Hence, combining the above arguments one finally obtains the following result:

**Corollary 1** *Suppose that the density $g$ satisfies the polynomial decay condition (2.1) at rate $\beta$ for its Fourier coefficients. Then, as $n \to +\infty$*

$$\sup_{f \in H_s(A)} \mathbb{E}_\theta \|\theta^\star - \theta\|^2 \sim \log^2(n) n^{\frac{-2s}{2s+2\beta+1}}$$

*and*

$$\sup_{f \in H_s(A)} \mathbb{E}_\theta \|\tilde{\theta} - \theta\|^2 \sim n^{\frac{-2s}{2s+2\beta+1}}.$$

From the lower bound obtained in Theorem 3.1 we conclude that, for $s \geq 2\beta + 1$, the performances of the estimator $\tilde{\theta}$ are asymptotically optimal from the minimax point of view, while the estimator $\theta^\star$ is near-optimal up to a $\log^2(n)$ factor. This near-optimal rate of convergence of $\theta^\star$ is due to the use of the penalised criterion $\bar{U}(Y, N)$, see (2.7), with a penalty term involving a $\frac{\log^2(n)}{n}$ factor used to eliminate the term $\frac{1}{n} \sum_{|k| \leq N} |\gamma_k|^{-4} \left\{ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right\}$ in the unbiased risk $U(Y, N)$, see (2.6). This shows that the performances of $\theta^\star$ and $\tilde{\theta}$ are essentially the same from a minimax point of view.

# 4 Numerical experiments

For the mean pattern $f$ to recover, we consider the smooth function shown in Figure 1(a). Then, we simulate $n = 100$ 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 Figure 1(b). 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 (2.9) thus leads to the choice $m_0 = 32$. Minimisation of the criterions (2.8) and (2.15) leads respectively to the choices $N^\star = 13$ and $\tilde{N} = 30$. An example of estimation by spectral cut-off using either the value of $N^\star$ or $\tilde{N}$ is displayed in Figure 1(c) and Figure 1(d). The estimator obtained with the frequency cut-off $N^\star = 13$ is very satisfactory, while the choice $\tilde{N} = 30$ seems to be too large as the resulting estimator in Figure 1(d) is not as smooth as the estimator with $N^\star = 13$.
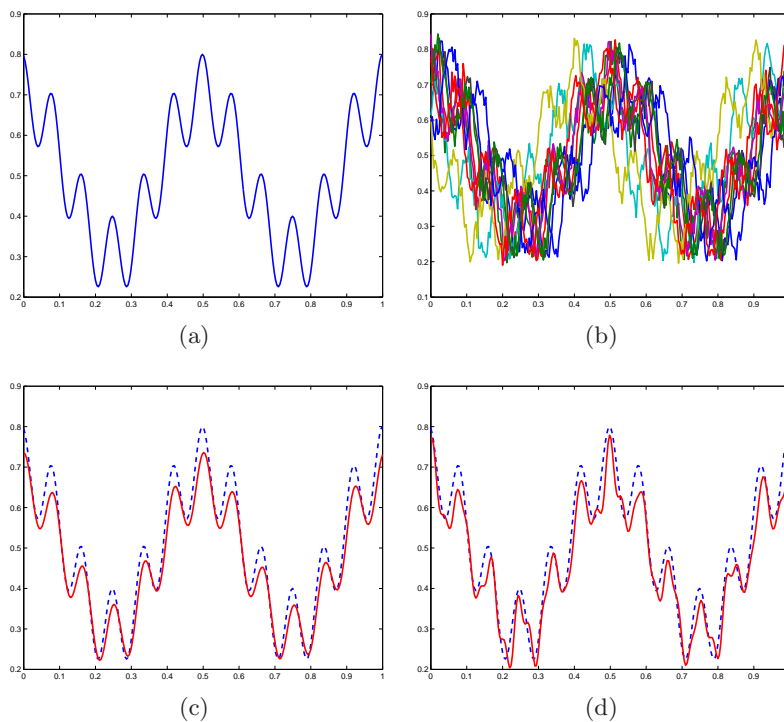


Figure 1: Wave function. (a) Mean pattern $f$, (b) Sample of 10 curves out of $n = 100$, (c) Estimation by spectral cut-off with $N^\star = 13$, (d) Estimation by spectral cut-off with $\tilde{N} = 30$. The dotted curve corresponds to the true mean pattern $f$.

This result tends to suggest that minimising $\bar{U}(Y, N)$ leads to a smaller choice for the frequency cut-off than the one obtained by the minimisation of the criterion $\tilde{U}(Y, N)$. This is confirmed by the results displayed in Figure 2 which gives the histogram of the selected values for $N^\star$ and $\tilde{N}$ over $M = 100$ independent replications of the above described simulations. Clearly the value of $N^\star$ is generally much smaller than $\tilde{N}$, and thus minimising (2.15) may lead to undersmoothing which illustrates numerically our discussion in Section 2 on the differences between $\theta^\star$ and $\tilde{\theta}$.

# 5 Proofs

**Proof of Theorem 2.1.** The proof uses the following scheme. In a first time, we compute the quadratic risk of $\theta^\star$ and we prove that it is close to $\bar{R}(\theta, N^\star)$. The aim of the second part is to
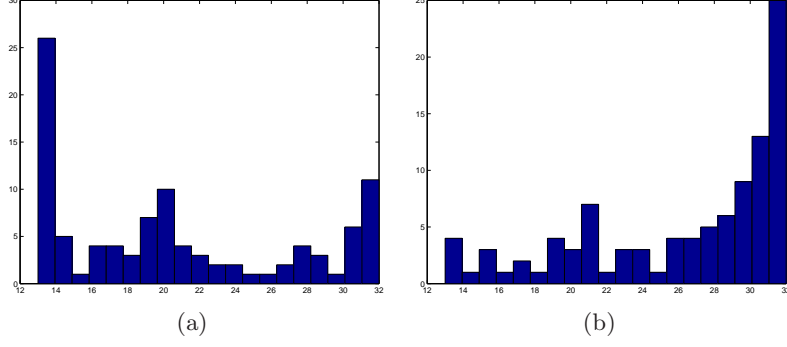
Figure 2: Selection of the frequency cut-off over $M = 100$ replications of the simulations (with $m_0 = 32$): (a) Histogram of the selected value for $N^\star$, (b) Histogram of the selected value for $\tilde{N}$.

prove that $\bar{U}(Y, N^\star)$ is close to $\bar{R}(\theta, N^\star)$, even for a random bandwidth $N^\star$. Then, we use the fact that $N^\star$ minimizes the criterion $\bar{U}(Y, N^\star)$ over the integer smaller than $m_0$ and we compute the expectation of $U(Y, N)$ for all deterministic $N$ in order to obtain an oracle inequality.

In a first time,

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

where for a given $z \in \mathbb{C}$, $Re(z)$ denotes the real part of $z$ and $\bar{z}$ the conjuguate. The last equality can be rewritten as

$$
\begin{aligned}
\mathbb{E}_\theta \|\theta^\star - \theta\|^2 &= \mathbb{E}_\theta \tilde{R}(\theta, N^\star) + \mathbb{E}_\theta \sum_{|k| \leq N^\star} \left| \frac{\tilde{\gamma}_k}{\gamma_k} - 1 \right|^2 |\theta_k|^2 + \frac{\epsilon^2}{n} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2} (|\xi_k|^2 - 1) \\
&\qquad + 2\mathbb{E}_\theta \sum_{|k| \leq N^\star} \frac{\epsilon}{\sqrt{n}} Re\left( (\gamma_k^{-1} \tilde{\gamma}_k - 1)\theta_k \times \bar{\gamma}_k^{-1} \bar{\xi}_k \right), \\
&= \mathbb{E}_\theta \tilde{R}(\theta, N^\star) + A_1 + A_2 + A_3, \quad (5.1)
\end{aligned}
$$

where $\tilde{R}(\theta, N)$ is defined in (2.13). Thanks to Lemma 5.1, setting $K = 1$,

$$
A_1 = \mathbb{E}_\theta \sum_{|k| \leq N^\star} \left| \frac{\tilde{\gamma}_k}{\gamma_k} - 1 \right|^2 |\theta_k|^2 \leq \frac{\log^2(n)}{n} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2} |\theta_k|^2 + \frac{C}{n}. \quad (5.2)
$$

Now, consider a bound for $A_2$. For all $N \in \mathbb{N}$ set $\Sigma_N = \sum_{|k| \leq N} |\gamma_k|^{-4}$. Then, for all $p \in ]1, 2[$

12

and $1 > \gamma > 0$:

$$
\begin{aligned}
A_2 &= \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-2}(|\xi_k|^2 - 1), \\
&= \frac{\epsilon^2}{n}\mathbb{E}_\theta \left[ \sum_{|k|\leq N^\star} |\gamma_k|^{-2}(|\xi_k|^2 - 1) - \gamma\sqrt{\Sigma_N^\star}^p \right] + \gamma\frac{\epsilon^2}{n}\mathbb{E}_\theta \sqrt{\Sigma_N^\star}^p, \\
&\leq \frac{\epsilon^2}{n}\mathbb{E}_\theta \sup_N \left[ \sum_{|k|\leq N} |\gamma_k|^{-2}(|\xi_k|^2 - 1) - \gamma\sqrt{\Sigma_N^\star}^p \right] + \gamma\frac{\epsilon^2}{n}\mathbb{E}_\theta \sqrt{\Sigma_N^\star}^p, \\
&\leq \gamma\frac{\epsilon^2}{n}\mathbb{E}_\theta \sqrt{\Sigma_N^\star}^p + \frac{C}{\gamma^{1/(1-p)}}\frac{\epsilon^2}{n}.
\end{aligned}
$$

The last step can be derived from a Doob inequality: see for instance [CG06]. Thanks to the polynomial Assumption 2.1 on the sequence $(\gamma_k)_k$ and setting $p = 2 \times (2\beta+1)/(4\beta+1)$, we obtain

$$
A_2 = \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-2}(|\xi_k|^2 - 1) \leq \gamma\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-2} + \frac{C}{\gamma^{4\beta+1}}\frac{\epsilon^2}{n}. \tag{5.3}
$$

Then, for all $1 > B > 0$, using the Cauchy-Schwarz and Young inequalities with the bounds (5.2) and (5.3)

$$
\begin{aligned}
A_3 &= 2\mathbb{E}_\theta \sum_{|k|\leq N^\star} \frac{\epsilon}{\sqrt{n}}Re\left((\gamma_k^{-1}\tilde{\gamma}_k - 1)\theta_k \times \bar{\gamma}_k^{-1}\bar{\xi}_k\right), \\
&\leq B\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-2}|\xi_k|^2 + B^{-1}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\theta_k|^2 \left|\frac{\tilde{\gamma}_k}{\gamma_k} - 1\right|^2,
\end{aligned}
$$

Thus, for any $K > 0$,

$$
A_3 \leq (B + B\gamma)\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-2} + B^{-1}K\frac{\log^2(n)}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-2}|\theta_k|^2 + \frac{C\epsilon^2}{n\gamma^{4\beta+1}} + \frac{C}{nK}. \tag{5.4}
$$

With $B = \sqrt{K} = \sqrt{\gamma}$, we obtain from (5.1)-(5.4)

$$
\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \leq (1 + \gamma + 2\sqrt{\gamma})\mathbb{E}_\theta \bar{R}(\theta, N^\star) + \frac{C\epsilon^2}{n\gamma^{4\beta+1}} + \frac{C}{n}, \tag{5.5}
$$

where $\bar{R}(\theta, N)$ is defined in (2.12). This concludes the first step of our proof. Now, we write $\bar{U}(Y, N^\star)$ in terms of $\bar{R}(\theta, N^\star)$. In the following, we define $x_n = (1 - n^{-1})$. We have

$$
\begin{aligned}
\bar{U}(Y, N^\star) &= -x_n \sum_{|k|\leq N^\star} |\gamma_k|^{-2}\left\{|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}\right\} + \frac{\epsilon^2}{n}\sum_{|k|\leq N^\star} |\gamma_k|^{-2} + \frac{\log^2(n)}{n}\sum_{|k|\leq N} |\gamma_k|^{-4}\left\{|\tilde{c}_k| - \frac{\epsilon^2}{n}\right\}, \\
&= \bar{R}(\theta, N^\star) - \left(1 - \frac{1}{n}\right)\sum_{|k|\leq N^\star} |\gamma_k|^{-2}\left\{|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}\right\} - \sum_{|k|\geq N^\star} |\theta_k|^2 \\
&\quad + \frac{\log^2(n)}{n}\sum_{|k|\leq N^\star} \left[|\gamma_k|^{-4}\left\{|\tilde{c}_k| - \frac{\epsilon^2}{n}\right\} - |\gamma_k|^{-2}|\theta_k|^2\right]
\end{aligned}
$$

This equality can be rewritten as

$$
\begin{aligned}
\bar{R}(\theta, N^\star) &= \bar{U}(Y, N^\star) + \|\theta\|^2 + x_n \sum_{|k|\leq N^\star} \left\{|\gamma_k|^{-2}|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}|\gamma_k|^{-2}\right\} - \sum_{|k|\leq N^\star} |\theta_k|^2 \\
&\quad + \frac{\log^2(n)}{n}\sum_{|k|\leq N^\star} \left[|\gamma_k|^{-2}|\theta_k|^2 - |\gamma_k|^{-4}\left\{|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}\right\}\right]. \tag{5.6}
\end{aligned}
$$

13

For all $k \in \mathbb{N}$
$$|\tilde{c}_k|^2 = |\theta_k \tilde{\gamma}_k|^2 + \frac{\epsilon^2}{n}|\xi_k|^2 + 2\epsilon n^{-1/2} Re(\theta_k \tilde{\gamma}_k \bar{\xi}_k),$$

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

Since $x_n < 1$

$$
\begin{aligned}
x_n \mathbb{E}_\theta & \sum_{|k| \leq N^\star} \left\{ |\gamma_k|^{-2}|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}|\gamma_k|^{-2} \right\} - \sum_{|k| \leq N^\star} |\theta_k|^2 \\
\leq \quad & \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 \left( \left|\frac{\tilde{\gamma}_k}{\gamma_k}\right|^2 - 1 \right) + \frac{\epsilon^2}{n} x_n \sum_{|k| \leq N^\star} |\gamma_k|^{-2}(|\xi_k|^2 - 1) + 2\frac{\epsilon}{\sqrt{n}} x_n \sum_{|k| \leq N^\star} |\gamma_k|^{-2} Re(\theta_k \tilde{\gamma}_k \bar{\xi}_k), \\
= \quad & E_1 + E_2 + E_3. \tag{5.7}
\end{aligned}
$$

First consider the bound of $E_1$. Thanks to Lemma 5.2 and some simple algebra

$$
\begin{aligned}
E_1 \quad = \quad & \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 \left( \left|\frac{\tilde{\gamma}_k}{\gamma_k}\right|^2 - 1 \right), \\
\leq \quad & 2\gamma \frac{\log^2(n)}{n} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 |\gamma_k|^{-2} + \gamma \mathbb{E}_\theta \sum_{|k| > N^\star} |\theta_k|^2 \\
& + \gamma \sum_{|k| > N_0} |\theta_k|^2 + \gamma^{-1} \sum_{|k| \leq N_0} |\theta_k|^2 |\gamma_k|^{-2}(1 - |\gamma_k|^2) + \frac{C}{n\gamma^2}, \\
\leq \quad & 2\gamma \mathbb{E}_\theta \bar{R}(\theta, N^\star) + \left( \gamma + \frac{\gamma^{-1}}{\log^2(n)} \right) \bar{R}(\theta, N_0) + \frac{C}{n\gamma^2},
\end{aligned}
$$

where
$$N_0 = \arg \min_{N \leq m_0} \bar{R}(\theta, N).$$

The terms $E_2$ and $E_3$ are bounded using respectively (5.3) and Lemma 5.3. We get

$$
\begin{aligned}
\mathbb{E}_\theta & \sum_{|k| \leq N^\star} \left\{ |\gamma_k|^{-2}|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}|\gamma_k|^{-2} - |\theta_k|^2 \right\} \\
& \leq \quad D\gamma \mathbb{E}_\theta \bar{R}(\theta, N^\star) + D\gamma \bar{R}(\theta, N_0) + \frac{\epsilon^2}{n}\frac{C}{\gamma^{4\beta+1}} + \frac{C}{n\gamma^2}. \tag{5.8}
\end{aligned}
$$

We are now interested in the second residual term of (5.6). Thanks to the definition of $\tilde{c}_k$:

$$
\begin{aligned}
\frac{\log^2 n}{n} \mathbb{E}_\theta & \sum_{|k| \leq N^\star} |\gamma_k|^{-2} \left\{ -|\gamma_k|^{-2}|\tilde{c}_k|^2 + \frac{\epsilon^2}{n}|\gamma_k|^{-2} + |\theta_k|^2 \right\} \\
= \quad & \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2}|\theta_k|^2 \left( 1 - \left|\frac{\tilde{\gamma}_k}{\gamma_k}\right|^2 \right) + \frac{\epsilon^2}{n} \sum_{|k| \leq N^\star} |\gamma_k|^{-4}(1 - |\xi_k|^2) - 2\frac{\epsilon}{\sqrt{n}} \sum_{|k| \leq N^\star} |\gamma_k|^{-4} Re(\theta_k \tilde{\gamma}_k \bar{\xi}_k), \\
\leq \quad & D\gamma \mathbb{E}_\theta \bar{R}(\theta, N^\star) + D\gamma \bar{R}(\theta, N_0) + \frac{\epsilon^2}{n}\frac{C}{\gamma^{4\beta+1}} + \frac{C}{n\gamma^2}, \tag{5.9}
\end{aligned}
$$

for some $D > 0$ independent of $\epsilon$ and $n$. Indeed, we can use essentialy the same algebra as for the bound of the terms $E_1, E_2$ and $E_3$ and the inequality

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

14

Hence, using (5.8) and (5.9)

$$(1 - D\gamma)\mathbb{E}_\theta \bar{R}(\theta, N^\star) \leq \mathbb{E}_\theta U(Y, N^\star) + \|\theta\|^2 + D\gamma \tilde{R}(\theta, N_0) + \frac{C}{n\gamma^2} + \frac{C\epsilon^2}{n}\frac{1}{\gamma^{4\beta+1}}. \qquad (5.10)$$

¿From the definition of $N^\star$, we immediatly get

$$(1 - D\gamma)\mathbb{E}_\theta \bar{R}(\theta, N^\star) \leq \mathbb{E}_\theta U(Y, N_0) + \|\theta\|^2 + D\gamma \tilde{R}(\theta, N_0) + \frac{C}{n\gamma^2} + \frac{C\epsilon^2}{n}\frac{1}{\gamma^{4\beta+1}},$$

where $N_0$ denotes the oracle bandwidth. Since

$$\mathbb{E}_\theta U(Y, N_0) = \tilde{R}(\theta, N_0) - \|\theta\|^2,$$

we obtain

$$(1 - D\gamma)\mathbb{E}_\theta \tilde{R}(\theta, N) \leq (1 + D\gamma)\tilde{R}(\theta, N_0) + \frac{C}{n\gamma^2} + \frac{C\epsilon^2}{n}\frac{1}{\gamma^{4\beta+1}}. \qquad (5.11)$$

Using (5.5) and (5.11), we get:

$$\begin{aligned}
\mathbb{E}_\theta \|\theta^\star - \theta\|^2 &\leq (1 + D\sqrt{\gamma})\mathbb{E}_\theta \tilde{R}(\theta, N^\star) + \frac{C\epsilon^2}{n}\frac{1}{\gamma^{4\beta+1}} + \frac{C}{n\gamma}, \\
&\leq \left(\frac{1 + D\sqrt{\gamma}}{1 - D\gamma}\right)\tilde{R}(\theta, N_0) + \frac{C\epsilon^2}{n}\frac{1}{\gamma^{4\beta+1}} + \frac{C}{n\gamma}.
\end{aligned}$$

This concludes the proof of Theorem 2.1.

$\square$

**Proof of Theorem 2.2**. The proof follows the same main lines as for Theorem 2.1. Inequality (5.1) provides:

$$\begin{aligned}
\mathbb{E}_\theta \|\theta^\star - \theta\|^2 &= \mathbb{E}_\theta \tilde{R}(\theta, N^\star) + \mathbb{E}_\theta \sum_{|k| \leq N^\star} \left|\frac{\tilde{\gamma}_k}{\gamma_k} - 1\right|^2 |\theta_k|^2 + \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2}(|\xi_k|^2 - 1) \\
&\quad + 2\mathbb{E}_\theta \sum_{|k| \leq N^\star} \frac{\epsilon}{\sqrt{n}}Re\left((\gamma_k^{-1}\tilde{\gamma}_k - 1)\theta_k \times \bar{\gamma}_k^{-1}\bar{\xi}_k\right), \\
&= \mathbb{E}_\theta \tilde{R}(\theta, N^\star) + A_1 + A_2 + A_3.
\end{aligned}$$

Thanks to Lemma 5.1 and an inequality of [CGPT02], we obtain for all $0 < \gamma < 1$:

$$\begin{aligned}
A_1 &\leq \log^2(n)\frac{\epsilon^2}{n}\mathbb{E}_\theta \sup_{|k| \leq N^\star} |\gamma_k|^{-2}|\theta_k|^2 + \frac{C}{n}, \\
&\leq \gamma\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2} + \frac{C\epsilon^2}{n}\left(\frac{\|\theta\|^2 \log^2(n)}{\gamma}\right)^{2\beta} + \frac{C}{n}. \qquad (5.12)
\end{aligned}$$

Then, for all $B > 0$, using the Cauchy-Schwarz and Young inequalities with the bounds (5.2),(5.3):

$$\begin{aligned}
A_3 &= 2\mathbb{E}_\theta \sum_{|k| \leq N^\star} \frac{\epsilon}{\sqrt{n}}Re\left((\gamma_k^{-1}\tilde{\gamma}_k - 1)\theta_k \times \bar{\gamma}_k^{-1}\bar{\xi}_k\right), \\
&\leq (B + B\gamma + B^{-1}\gamma)\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2} + \frac{C\epsilon^2}{n}\left(\frac{\|\theta\|^2 \log^2(n)}{\gamma}\right)^{2\beta} + \frac{C\epsilon^2}{n\gamma^{4\beta+1}} + \frac{C}{n}. \quad (5.13)
\end{aligned}$$

15

With the choice $B = \sqrt{\gamma}$, we obtain from (5.1)-(5.4):

$$\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \leq (1 + 3\gamma + 2\sqrt{\gamma})\mathbb{E}_\theta \tilde{R}(\theta, N^\star) + \frac{C\epsilon^2}{n}\left(\frac{\|\theta\|^2 \log^2(n)}{\gamma}\right)^{2\beta} + \frac{C\epsilon^2}{n\gamma^{4\beta+1}} + \frac{C}{n}. \quad (5.14)$$

Then,

$$
\begin{aligned}
U(Y, N^\star) &= -\sum_{|k|\leq N^\star}\left\{|\gamma_k|^{-2}|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}|\gamma_k|^{-2}\right\} + \frac{\epsilon^2}{n}\sum_{|k|\leq N^\star}|\gamma_k|^{-2}, \\
&= -\sum_{|k|\leq N^\star}\left\{|\gamma_k|^{-2}|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}|\gamma_k|^{-2}\right\} - \sum_{|k|\geq N^\star}|\theta_k|^2 + \sum_{|k|\geq N^\star}|\theta_k|^2 + \frac{\epsilon^2}{n}\sum_{|k|\leq N^\star}|\gamma_k|^{-2}, \\
&= \tilde{R}(\theta, N^\star) - \sum_{|k|\leq N^\star}\left\{|\gamma_k|^{-2}|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}|\gamma_k|^{-2}\right\} - \sum_{|k|\geq N^\star}|\theta_k|^2.
\end{aligned}
$$

This equality can be rewritten as

$$\tilde{R}(\theta, N^\star) = U(Y, N^\star) + \|\theta\|^2 + \sum_{|k|\leq N^\star}\left\{|\gamma_k|^{-2}|\tilde{c}_k|^2 - \frac{\epsilon^2}{n}|\gamma_k|^{-2} - |\theta_k|^2\right\}.$$

Hence,

$$
\begin{aligned}
\mathbb{E}_\theta \tilde{R}(\theta, N) &= \mathbb{E}_\theta U(Y, N^\star) + \|\theta\|^2 + \mathbb{E}_\theta \sum_{|k|\leq N^\star}|\theta_k|^2\left(\left|\frac{\tilde{\gamma}_k}{\gamma_k}\right|^2 - 1\right) \\
&\quad + \frac{\epsilon^2}{n}\sum_{|k|\leq N^\star}|\gamma_k|^{-2}(|\xi_k|^2 - 1) + 2\frac{\epsilon}{\sqrt{n}}\sum_{|k|\leq N^\star}|\gamma_k|^{-2}Re(\theta_k\tilde{\gamma}_k\bar{\xi}_k), \\
&= \mathbb{E}_\theta U(Y, N^\star) + \|\theta\|^2 + E_1 + E_2 + E_3. \quad (5.15)
\end{aligned}
$$

Using previous results:

$$
\begin{aligned}
E_1 &= \mathbb{E}_\theta \sum_{|k|\leq N^\star}|\theta_k|^2\left(\left|\frac{\tilde{\gamma}_k}{\gamma_k}\right|^2 - 1\right), \\
&\leq 2\gamma\mathbb{E}_\theta \tilde{R}(\theta, N^\star) + \gamma\tilde{R}(\theta, N_0) + \frac{C}{n} + \frac{C\epsilon^2}{n}\left(\frac{\|\theta\|^2 \log^2(n)}{\gamma^2}\right)^{2\beta}. \quad (5.16)
\end{aligned}
$$

The terms $E_2$ and $E_3$ are bounded using respectively (5.3) and Lemma 5.3. We get:

$$
\begin{aligned}
\mathbb{E}_\theta \tilde{R}(\theta, N^\star) &\leq \mathbb{E}_\theta U(Y, N^\star) + \|\theta\|^2 + D\gamma\mathbb{E}_\theta \tilde{R}(\theta, N^\star) + D\gamma\tilde{R}(\theta, N_0) \\
&\quad + \frac{C}{n\gamma^{2\beta}} + \frac{C\epsilon^2}{n}\left(\frac{\|\theta\|^2 \log^2(n)}{\gamma^2}\right)^{2\beta}.
\end{aligned}
$$

Hence,

$$(1 - D\gamma)\mathbb{E}_\theta \tilde{R}(\theta, N^\star) \leq \mathbb{E}_\theta U(Y, N^\star) + \|\theta\|^2 + D\gamma\tilde{R}(\theta, N_0) + \frac{C}{n\gamma^{2\beta}} + \frac{C\epsilon^2}{n}\left(\frac{\|\theta\|^2 \log^2(n)}{\gamma^2}\right)^{2\beta}. \quad (5.17)$$

¿From the definition of $N^\star$, we immediatly get:

$$(1 - D\gamma)\mathbb{E}_\theta \tilde{R}(\theta, N^\star) \leq \mathbb{E}_\theta U(Y, N_0) + \|\theta\|^2 + D\gamma\tilde{R}(\theta, N_0) + \frac{C}{n\gamma^{2\beta}} + \frac{C\epsilon^2}{n}\left(\frac{\|\theta\|^2 \log^2(n)}{\gamma^2}\right)^{2\beta}. \quad (5.18)$$

16

In order to conclude the proof, we prove that $\mathbb{E}_\theta U(Y, N_0)$ is close to $R(\theta, N_0)$. First remark that:

$$
\begin{aligned}
\mathbb{E}_\theta U(Y, N_0) &= \mathbb{E}_\theta \left[ -\sum_{k=1}^{N_0} |\gamma_k|^{-2} \left\{ |\tilde{c}_k|^2 - \frac{\epsilon^2}{n} \right\} + \frac{\epsilon^2}{n} \sum_{k=1}^{N_0} |\gamma_k|^{-2} \right], \\
&= \mathbb{E}_\theta \left[ -\sum_{k=1}^{N_0} \left\{ |\gamma_k|^{-2} |\tilde{c}_k|^2 - |\gamma_k|^{-2} \frac{\epsilon^2}{n} - |\theta_k|^2 \right\} \right] - \sum_{k=1}^{N_0} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{k=1}^{N_0} |\gamma_k|^{-2}.
\end{aligned}
$$

Since for all $k \in \mathbb{N}$:

$$
\mathbb{E}_\theta |\tilde{c}_k|^2 = |\theta_k|^2 \mathbb{E}_\theta |\tilde{\gamma}_k|^2 + \frac{\epsilon^2}{n} = |\theta_k|^2 \left( |\gamma_k|^2 + \frac{1}{n} \right) + \frac{\epsilon^2}{n},
$$

we obtain,

$$
\mathbb{E}_\theta U(Y, N_0) = -\sum_{k=1}^{N_0} |\theta_k|^2 \frac{|\gamma_k|^{-2}}{n} + \sum_{|k|>N_0} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{k=1}^{N_0} |\gamma_k|^{-2} - \|\theta\|^2.
$$

Therefore,

$$
\mathbb{E}_\theta U(Y, N_0) = -\sum_{k=1}^{N_0} |\theta_k|^2 \frac{|\gamma_k|^{-2}}{n} + \tilde{R}(\theta, N_0) - \|\theta\|^2 \leq \tilde{R}(\theta, N_0) - \|\theta\|^2,
$$

and

$$
(1 - D\gamma) \mathbb{E}_\theta \tilde{R}(\theta, N) \leq (1 + D\gamma) \tilde{R}(\theta, N_0) + \frac{C}{n\gamma^{2\beta}} + \frac{C\epsilon^2}{n} \left( \frac{\|\theta\|^2 \log^2(n)}{\gamma^2} \right)^{2\beta}. \tag{5.19}
$$

Using (5.5) and (5.19), we get:

$$
\begin{aligned}
\mathbb{E}_\theta \|\theta^\star - \theta\|^2 &\leq (1 + D\sqrt{\gamma}) \mathbb{E}_\theta \tilde{R}(\theta, N^\star) + \frac{C\epsilon^2}{n} \left( \frac{\|\theta\|^2 \log^2(n)}{\gamma^2} \right)^{2\beta} + \frac{C\epsilon^2}{n} \frac{1}{\gamma^{4\beta+1}} + \frac{1}{n}, \\
&\leq \left( \frac{1 + D\sqrt{\gamma}}{1 - D\gamma} \right) \tilde{R}(\theta, N_0) + \frac{C\epsilon^2}{n} \left( \frac{\|\theta\|^2 \log^2(n)}{\gamma^2} \right)^{2\beta} + \frac{C\epsilon^2}{n} \frac{1}{\gamma^{4\beta+1}} + \frac{1}{n}.
\end{aligned}
$$

Since $\tilde{R}(\theta, N) \leq R(\theta, N)$, we eventually get:

$$
\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \leq \left( \frac{1 + D\sqrt{\gamma}}{1 - D\gamma} \right) \inf_N R(\theta, N) + \frac{C\epsilon^2}{n} \left( \frac{\|\theta\|^2 \log^2(n)}{\gamma^2} \right)^{2\beta} + \frac{C\epsilon^2}{n} \frac{1}{\gamma^{4\beta+1}} + \frac{1}{n}.
$$

This concludes the proof of Theorem 2.2.

$\square$

# Appendix

**Lemma 5.1** *For all $K > 0$, we have*

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

*where $C$ denote a positive constant independent of $\epsilon$ and $n$.*

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

$$\mathbb{E}_\theta \sum_{|k| \le N^\star} \left| \frac{\tilde{\gamma}_k}{\gamma_k} - 1 \right|^2 |\theta_k|^2 = \mathbb{E}_\theta \sum_{|k| \le N^\star} |\theta_k|^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2,$$

$$\le Q \mathbb{E}_\theta \sum_{|k| \le N^\star} |\theta_k|^2 |\gamma_k|^{-2} + \mathbb{E}_\theta \sum_{|k| \le N^\star} |\theta_k|^2 |\gamma_k|^{-2} \left\{ |\tilde{\gamma}_k - \gamma_k|^2 - Q \right\} \mathbb{1}_{\{|\tilde{\gamma}_k - \gamma_k|^2 > Q\}}.$$

Thanks to (2.8) and (2.9)

$$\mathbb{E}_\theta \sum_{|k| \le N^\star} |\theta_k|^2 |\gamma_k|^{-2} \left\{ |\tilde{\gamma}_k - \gamma_k|^2 - Q \right\} \mathbb{1}_{\{|\tilde{\gamma}_k - \gamma_k|^2 > Q\}}$$

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

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

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

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

$$P(|\tilde{\gamma}_k - \gamma_k|^2 \ge x) = P \left( \left| \frac{1}{n} \sum_{l=1}^n \left\{ e^{-2i\pi k\tau_l} - \mathbb{E}[e^{-2i\pi k\tau_l}] \right\} \right| \ge \sqrt{x} \right),$$

$$\le 2 \exp \left\{ -\frac{(n\sqrt{x})^2}{2 \sum_{l=1}^n \mathrm{Var}(e^{-2i\pi k\tau_l}) + n\sqrt{x}/3} \right\},$$

$$\le 2 \exp \left\{ -\frac{(n\sqrt{x})^2}{2n + n\sqrt{x}/3} \right\}.$$

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

$$\mathbb{E}_\theta \left[ |\tilde{\gamma}_k - \gamma_k|^2 - Q \right] \mathbb{1}_{\{|\tilde{\gamma}_k - \gamma_k|^2 > Q\}} \le \int_Q^{+\infty} \exp \left\{ -\frac{nx}{2 + \sqrt{x}/3} \right\} dx,$$

$$\le \int_Q^{36} \exp \left\{ -\frac{nx}{4} \right\} dx + \int_{36}^{+\infty} \exp \left\{ -Cn\sqrt{x} \right\} dx \le \frac{C}{n} e^{-Qn/4},$$

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

$$\mathbb{E}_\theta \sum_{|k| \le N^\star} \left| \frac{\tilde{\gamma}_k}{\gamma_k} - 1 \right|^2 |\theta_k|^2 \le K \frac{\log^2(n)}{n} \mathbb{E}_\theta \sum_{|k| \le N^\star} |\gamma_k|^{-2} |\theta_k|^2 + \frac{Cnm_0}{\log^2(n)} e^{-K \log^2(n)/4},$$

$$\le K \frac{\log^2(n)}{n} \mathbb{E}_\theta \sum_{|k| \le N^\star} |\gamma_k|^{-2} |\theta_k|^2 + \frac{C}{nK},$$

where $C$ denotes a positive constant independent of $\epsilon$ and $n$. This concludes the proof of Lemma 5.1.

$\square$

**Lemma 5.2** *Let $N^\star$ defined in (2.8). For all deterministic bandwidth $N$ and $0 < \gamma < 1$, we have*

$$\mathbb{E}_\theta \sum_{|k| \le N^\star} |\theta_k|^2 \left( \left| \frac{\tilde{\gamma}_k}{\gamma_k} \right|^2 - 1 \right) \le 2\gamma \frac{\log^2(n)}{n} \mathbb{E}_\theta \sum_{|k| \le N^\star} |\theta_k|^2 |\gamma_k|^{-2} + \gamma \mathbb{E}_\theta \sum_{|k| > N^\star} |\theta_k|^2$$

$$+ \gamma \sum_{|k| > N} |\theta_k|^2 + \frac{\gamma^{-1}}{n} \sum_{|k| \le N} |\theta_k|^2 |\gamma_k|^{-2} (1 - |\gamma_k|^2) + \frac{C}{n\gamma^2}.$$

18

*where $C$ denotes a positive constant independent of $\epsilon$ and $n$.*

PROOF. In a first time, remark that

$$\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 \left( \left| \frac{\tilde{\gamma}_k}{\gamma_k} \right|^2 - 1 \right) = \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 |\gamma_k|^{-2} (|\tilde{\gamma}_k - \gamma_k + \gamma_k|^2 - |\gamma_k|^2),$$

$$= \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 |\gamma_k|^{-2} \left\{ |\tilde{\gamma}_k - \gamma_k|^2 + 2Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k) \right\} \quad (5.20)$$

Let $N \in \mathbb{N}$ be a deterministic bandwidth. Since $\mathbb{E}_\theta \tilde{\gamma}_k = \gamma_k$ for all $k \in \mathbb{N}$, we can write that

$$\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k)$$

$$= \mathbb{E}_\theta \sum_{|k| \in \{N...N^\star\}} |\theta_k|^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k),$$

$$\leq \mathbb{E}_\theta \left| \sum_{|k| \in \{N...N^\star\}} |\theta_k|^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k) \right|,$$

$$\leq \mathbb{E}_\theta \sum_{k \in \mathbb{Z}} \left| (\mathbb{1}_{\{|k| \leq N^\star\}} - \mathbb{1}_{\{|k| \leq N\}}) |\theta_k|^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k) \right|.$$

Using simple algebra

$$\left| \mathbb{1}_{\{|k| \leq N^\star\}} - \mathbb{1}_{\{|k| \leq N\}} \right| = \left| (\mathbb{1}_{\{|k| \leq N^\star\}} + \mathbb{1}_{\{|k| \leq N\}})(\mathbb{1}_{\{|k| \leq N^\star\}} - \mathbb{1}_{\{|k| \leq N\}}) \right|,$$

$$= (\mathbb{1}_{\{|k| \leq N^\star\}} + \mathbb{1}_{\{|k| \leq N\}}) \left| \mathbb{1}_{\{|k| > N^\star\}} - \mathbb{1}_{\{|k| > N\}} \right|,$$

$$\leq \mathbb{1}_{\{|k| > N^\star\}} \mathbb{1}_{\{|k| \leq N\}} + \mathbb{1}_{\{|k| \leq N^\star\}} \mathbb{1}_{\{|k| > N\}}.$$

For all $\gamma > 0$, using the Cauchy-Schwartz and Young inequalities, we obtain

$$\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k)$$

$$\leq \mathbb{E}_\theta \sum_{k \in \mathbb{Z}} \mathbb{1}_{\{|k| > N^\star\}} \mathbb{1}_{\{|k| \leq N\}} |\theta_k|^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k)$$

$$+ \mathbb{E}_\theta \sum_{k \in \mathbb{Z}} \mathbb{1}_{\{|k| \leq N^\star\}} \mathbb{1}_{\{|k| > N\}} |\theta_k|^2 |\gamma_k|^{-2} Re((\tilde{\gamma}_k - \gamma_k)\bar{\gamma}_k)$$

$$\leq \gamma \mathbb{E}_\theta \sum_{|k| > N^\star} |\theta_k|^2 + \gamma \sum_{|k| > N} |\theta_k|^2 + \gamma^{-1} \mathbb{E}_\theta \sum_{|k| \leq N} |\theta_k|^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2$$

$$+ \gamma^{-1} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2. \quad (5.21)$$

Hence, from (5.20) and (5.21)

$$\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 \left( \left| \frac{\tilde{\gamma}_k}{\gamma_k} \right|^2 - 1 \right) \leq (1 + \gamma^{-1}) \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\theta_k|^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2 + \gamma \mathbb{E}_\theta \sum_{|k| > N^\star} |\theta_k|^2$$

$$+ \gamma \sum_{|k| > N} |\theta_k|^2 + \gamma^{-1} \mathbb{E}_\theta \sum_{|k| \leq N} |\theta_k|^2 |\gamma_k|^{-2} |\tilde{\gamma}_k - \gamma_k|^2.$$

A direct application of Lemma 5.1 provides, for all $K > 0$

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

$$+ \gamma \sum_{|k| > N} |\theta_k|^2 + \frac{\gamma^{-1}}{n} \sum_{|k| \leq N} |\theta_k|^2 |\gamma_k|^{-2} (1 - |\gamma_k|^2) + \frac{C}{nK}.$$

Just set $K = \gamma^2$ in order to conclude the proof of Lemma 5.2.

$\square$

**Lemma 5.3** *Let $N^\star$ the bandwidth defined in (2.8). For all deterministic bandwidth $N$ and $0 < \gamma < 1$, we have*

$$\frac{2\epsilon}{\sqrt{n}}\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2} Re(\theta_k \tilde{\gamma}_k \bar{\xi}_k) \leq 3\gamma \left\{ \sum_{|k| > N_0} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k| \leq N_0} |\gamma_k|^{-2} \right\}$$

$$+ 3\gamma \log^2(n) \mathbb{E}_\theta \left\{ \sum_{|k| > N^\star} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k| \leq N^\star} |\gamma_k|^{-2} \right\} + \frac{C}{n} + \frac{C\epsilon^2}{\gamma^{4\beta+1}}.$$

PROOF. In the following, we will use the inequality:

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

for some $\tau > 0$, wich can be proved using a Bernstein type inequality. Then, for all $\gamma > 0$, using the above result and inequality (4.31) of [CG06], we obtain

$$\frac{2\epsilon}{\sqrt{n}}\mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2} Re(\theta_k \tilde{\gamma}_k \bar{\xi}_k) \leq \gamma \left\{ \sum_{|k| > N_0} |\theta_k|^2 + \frac{\epsilon^2}{n} \mathbb{E}_\theta \sum_{|k| \leq N_0} |\gamma_k|^{-4} |\tilde{\gamma}_k|^2 \right\}$$

$$+ \gamma \mathbb{E}_\theta \left\{ \sum_{|k| > N^\star} |\theta_k|^2 + \frac{\epsilon^2}{n} \sum_{|k| \leq N^\star} |\gamma_k|^{-4} |\tilde{\gamma}_k|^2 \right\} + \frac{C\epsilon^2}{n} \frac{1}{\gamma^{4\beta+1}}.$$

In order to prove the above inequality, we use the inequality (4.31) of [CG06] and Since $\mathbb{E}_\theta \tilde{\gamma}_k = \gamma_k$,

$$\frac{\epsilon^2}{n} \mathbb{E}_\theta \sum_{|k| \leq N} |\gamma_k|^{-4} |\tilde{\gamma}_k|^2 = \frac{\epsilon^2}{n} \sum_{|k| \leq N} |\gamma_k|^{-4} \left\{ \mathbb{E}_\theta |\tilde{\gamma}_k|^2 - |\gamma_k|^2 + |\gamma_k|^2 \right\},$$

$$= \frac{\epsilon^2}{n} \sum_{|k| \leq N} |\gamma_k|^{-4} \left\{ |\gamma_k|^2 + \text{Var}(\tilde{\gamma}_k) \right\},$$

$$= \frac{\epsilon^2}{n} \sum_{|k| \leq N} |\gamma_k|^{-4} \left\{ |\gamma_k|^2 + \frac{1}{n} \right\} \leq 2\frac{\epsilon^2}{n} \sum_{|k| \leq N_0} |\gamma_k|^{-2}.$$

The same kind of inequality can be obtained with the random bandwidth $N^\star$. Indeed,

$$\frac{\epsilon^2}{n} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-4} |\tilde{\gamma}_k|^2 = \frac{\epsilon^2}{n} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-4} |\tilde{\gamma}_k - \gamma_k + \gamma_k|^2,$$

$$\leq \frac{\epsilon^2}{n} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-4} \left\{ 2|\tilde{\gamma}_k - \gamma_k|^2 + 2|\gamma_k|^2 \right\},$$

$$\leq \frac{2\epsilon^2}{n} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-2} + \frac{2\epsilon^2}{n} \mathbb{E}_\theta \sum_{|k| \leq N^\star} |\gamma_k|^{-4} |\tilde{\gamma}_k - \gamma_k|^2.$$

Using the same algebra as in the proof of Lemma 5.1, we obtain, for all $Q > 0$:

$$\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k^{-4}||\tilde{\gamma}_k - \gamma_k|^2$$

$$= Q\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-4} + \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-4} \left\{|\tilde{\gamma}_k - \gamma_k|^2 - Q\right\} \mathbb{1}_{\{|\tilde{\gamma}_k - \gamma_k|^2 > Q\}},$$

$$\leq Q\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-4} + \frac{C\epsilon^2 n}{\log^2(n)}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\tilde{\gamma}_k - \gamma_k|^2 \mathbb{1}_{\{|\tilde{\gamma}_k - \gamma_k|^2 > Q\}},$$

$$\leq Q\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-4} + \frac{C\epsilon^2}{\log^4(n)}e^{-Qn/4}.$$

Setting $Q = n^{-1}\log^2(n)$, we obtain,

$$\frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k^{-4}||\tilde{\gamma}_k - \gamma_k|^2 \leq \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-2}\frac{|\gamma_k|^{-2}\log^2(n)}{n} + \frac{C\epsilon^2}{n},$$

$$\leq \frac{\epsilon^2}{n}\mathbb{E}_\theta \sum_{|k|\leq N^\star} |\gamma_k|^{-2} + \frac{C\epsilon^2}{n}.$$

This concludes the proof.

$\square$

# References

[BG09] J. Bigot and S. Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *preprint*, 2009.

[BGV08] 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*, page to be published, 2008.

[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.

[CG06] Laurent Cavalier and Yuri. Golubev. Risk hull method and regularization by projections of ill-posed inverse problems. *Annals of Statistics*, 34:1653–1677, 2006.

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

[CL08] I. Castillo and J.M. Loubes. Estimation of the distribution of random shifts deformation. *Mathematical Methods of Statistics*, to appear, 2008.

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

[Dal07]     A. Dalalyan. Penalized maximum likelihood and semiparametric second-order effi-
            ciency. *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]    David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard.
            Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57:301–369, 1995.

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

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

[Fan91]     Jianquin 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.

[HR05]      Marc Hoffmann and Markus Reiß. Nonlinear estimation for linear inverse problems
            with error in the operator. *Annals of Statistics*, 38:310–336, 2005.

[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.. Mller. Functional convex averaging and synchronization for time-
            warped random curves. *JASA*, 99(467):687–699, 2004.

[LMG07]     J.M. Loubes, E. Maza, and F. Gamboa. Semi-parametric estimation of shifts. *EJS*,
            1:616–640, 2007.

[Mar06]     Clément Marteau. Regularization of inverse problems with unknown operator. *Math-
            ematical Methods of Statistics*, 15:415–443, 2006.

[Mar09]     Clément 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. *JRSS 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: Spriner-Verlag, 2002.

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

[Vim08]     M. Vimond. Efficient estimation for homothetic shifted regression models. *Ann.
            Statist.*, page to be published, 2008.

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