



HAL
open science

Nonparametric Estimation for Functional Data by Wavelet Thresholding

Christophe Chesneau, Bertrand Maillot

► **To cite this version:**

Christophe Chesneau, Bertrand Maillot. Nonparametric Estimation for Functional Data by Wavelet Thresholding. 2011. hal-00634800v1

HAL Id: hal-00634800

<https://hal.science/hal-00634800v1>

Preprint submitted on 23 Oct 2011 (v1), last revised 5 Jun 2012 (v2)

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

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

Nonparametric Estimation for Functional Data by Wavelet Thresholding

Christophe Chesneau and Bertrand Maillot

Abstract: This paper deals with the density and regression estimation problems for functional data. Using wavelet bases for Hilbert spaces of functions, we develop a new adaptive procedure based on wavelet thresholding. Its asymptotic performances are explored via the maxiset approach.

Key words and phrases: Functional data, Density estimation, Non-parametric regression, Wavelets, Hard thresholding.

AMS 2000 Subject Classifications: 62G07, 60B11.

1 Introduction

Due to technological progress, in particular the enlarged capacity of computer memory and the increasing efficiency of data collection devices, there is a growing number of applied sciences (biometrics, chemometrics, meteorology, medical sciences...) where collected data are curves which require appropriate statistical tools. Because of this, functional data analysis has known a quite important development in the last fifteen years (see e.g. Ramsay and Silverman (1997), Ramsay and Silverman (2002), Ferraty and Vieu (2006), Dabo-Niang and Ferraty (2008), Ferraty (2010), Ferraty and Romain (2010) and Ferraty (2011) for monographs and collective books on this specific subject). However, whereas there has been quantities of works on the nonparametric estimation of the probability density function for univariate and multivariate random variables since the papers of Parzen (1962) and Rosenblatt (1956), much less attention has been paid to

Christophe Chesneau and Bertrand Maillot, Université de Caen, LMNO, Campus II, Science 3, 14032, Caen, France

the infinite-dimensional case. The extension of the results from the multivariate framework to the infinite dimensional one is not direct since there is no equivalent of the Lebesgue measure on an infinite dimensional Hilbert space. In fact, the only locally finite and translation invariant measure on an infinite dimensional Hilbert space is the null measure and any locally finite measure μ is even very irregular: denoting by $\mathcal{B}(x, r)$ the ball of center x and radius r , we have that, for any point x , any arbitrary large M and any arbitrary small r such that $\mu(\mathcal{B}(x, r)) < \infty$, there exist $(x_1, x_2) \in \mathcal{B}(x, r)^2$ such that $\mu(\mathcal{B}(x_1, r/4)) < M \times \mu(\mathcal{B}(x_2, r/4))$. For a coverage of the theme of measures on infinite dimension spaces, we refer to Xia (1972), Yamasaki (1985), Dalecky and Fomin (1991) and Uglanov (2000).

The first consistency result for a kernel estimator of the density function for infinite dimensional random variables has been obtained in Dabo-Niang (2002) where a rate is given in the special case when the kernel is an indicator function and the density is defined with respect to the Wiener measure. Later, different estimators of the density, based on orthogonal series (Dabo-Niang (2004)), delta sequences (Prakasa Rao (2010b)) or wavelets (Prakasa Rao (2010a)), have been proposed but none of them is adaptive. Note that the estimation of the density probability function is nonetheless itself of intrinsic interest but it also has a key role in mode estimation and curve clustering (see Dabo-Niang (2006)).

Contrary to the chronology of studies in the multivariate case, in the functional framework, estimators of the regression function have been proposed before those of the density. Ferraty and Vieu introduced the first fully nonparametric estimator of the regression function, at first under the hypothesis that the underlying measure has a fractal dimension in Ferraty and Vieu (2000) and then using only probabilities of small balls in Ferraty and Vieu (2004). However, since these pioneering works, no adaptive estimator has been proposed.

Considering the density estimation problem from functional data, Prakasa Rao (2010a) has recently developed a new procedure based on the multiresolution approach on a separable Hilbert space introduced by Goh (2007). This procedure belongs to the family of the linear wavelet estimators. As proved in (Prakasa Rao, 2010a, Theorem 3.1), it enjoys powerful asymptotic properties. However, such a linear wavelet estimator has two drawbacks: it is not adaptive (i.e. its perfor-

mances are deeply associated to the smoothness of the unknown function) and it is not efficient to estimate functions with complex singularities (the sparsity nature of the wavelet decomposition of the unknown function is not captured). For these reasons, (Prakasa Rao, 2010a, Page 2 lines 14-16) “*it would be interesting to investigate the advantage of these wavelet estimators for functional data by using wavelet thresholding suggested by Donoho et al. (1996)*”. This perspective motivates our study.

Adopting the multiresolution approach on a separable Hilbert space H of Goh (2007), we construct a new adaptive wavelet procedure using the hard thresholding rule of Donoho *et al.* (1996). Since H remains an abstract space, we propose to evaluate its asymptotic properties via the maxiset approach introduced by Cohen *et al.* (2001). The idea is to determine the maximal space of functions (or *maxiset*) where the procedure attains a given rate of convergence. Further details concerning this approach can be found in Kerkyacharian and Picard (2000) and Autin (2004). Using the general maxiset result of Kerkyacharian and Picard (2000), we prove that the maxiset of our procedure is the intersection of two different kinds of Besov spaces (defined in Section 2). The interest of such spaces is to contain a wide variety of unknown functions, complex or not.

A part of the paper is devoted to the regression model from functional data. We adapt the construction of our wavelet hard thresholding estimator to this problem and prove similar maxiset results.

The paper is structured as follows. In Section 2, we briefly describe the wavelet bases on H and we define some decomposition spaces. The density estimation problem for functional data via wavelet thresholding is considered in Section 3. The regression one is developed in Section 4. The proofs are gathered in Section 5.

2 Wavelet Bases on H and Decomposition Spaces

2.1 Wavelet Bases on H

Let us briefly describe the construction of wavelet bases on H introduced by Goh (2007). Let H be a separable Hilbert space of real- or complex-valued functions

defined on a set S . Since H is separable, it has an orthonormal basis $\mathcal{E} = \{e_j; j \in \Lambda\}$ for some countable index set Λ . As usual, we denote $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ the inner product and corresponding norm that H is equipped with.

Let $\{\mathcal{I}_k; k \geq 0\}$ be an increasing sequence of finite subsets of Λ such that $\bigcup_{k \geq 0} \mathcal{I}_k = \Lambda$ and, for any $k \geq 0$, $\mathcal{J}_k = \mathcal{I}_{k+1}/\mathcal{I}_k$. For any $k \geq 0$, we suppose that there exist $\zeta_{k,\ell} \in S$, $\ell \in \mathcal{I}_k$ and $\eta_{k,\ell} \in S$, $\ell \in \mathcal{J}_k$, such that the two matrices

$$A_k = (e_j(\zeta_{k,\ell}))_{(j,\ell) \in \mathcal{I}_k^2}, \quad B_k = (e_j(\eta_{k,\ell}))_{(j,\ell) \in \mathcal{J}_k^2},$$

satisfy one of the two following conditions:

(A1) $A_k^* A_k = \text{diag}(c_{k,\ell})_{\ell \in \mathcal{I}_k}$ and $B_k^* B_k = \text{diag}(s_{k,\ell'})_{\ell' \in \mathcal{J}_k}$, where $c_{k,\ell}$, $s_{k,\ell'}$, for $\ell \in \mathcal{I}_k$ and $\ell' \in \mathcal{J}_k$, are positive constants,

(A2) $A_k A_k^* = \text{diag}(d_{k,j})_{j \in \mathcal{I}_k}$ and $B_k B_k^* = \text{diag}(t_{k,j'})_{j' \in \mathcal{J}_k}$, where $d_{k,j}$, $t_{k,j'}$ for $j \in \mathcal{I}_k$ and $j' \in \mathcal{J}_k$, are positive constants.

For any $x \in S$, we set

$$\begin{cases} \phi_k(x; \zeta_{k,\ell}) = \sum_{j \in \mathcal{I}_k} \frac{1}{\sqrt{g_{j,k,\ell}}} \overline{e_j(\zeta_{k,\ell})} e_j(x), \\ \psi_k(x; \eta_{k,\ell}) = \sum_{j \in \mathcal{J}_k} \frac{1}{\sqrt{h_{j,k,\ell}}} \overline{e_j(\eta_{k,\ell})} e_j(x), \end{cases}$$

where

$$g_{j,k,\ell} = \begin{cases} c_{k,\ell} & \text{if (A1),} \\ d_{k,j} & \text{if (A2),} \end{cases} \quad h_{j,k,\ell} = \begin{cases} s_{k,\ell} & \text{if (A1),} \\ t_{k,j} & \text{if (A2).} \end{cases}$$

Then the collection

$$\mathcal{B} = \{\phi_0(x; \zeta_{0,\ell}), \ell \in \mathcal{I}_0; \psi_k(x; \eta_{k,\ell}), k \geq 0, \ell \in \mathcal{J}_k\}$$

is an orthonormal basis for H (see (Goh, 2007, Theorem 2 (a))).

Consequently, any $f \in H$ can be expressed on \mathcal{B} as

$$f(x) = \sum_{\ell \in \mathcal{I}_0} \alpha_{0,\ell} \phi_0(x; \zeta_{0,\ell}) + \sum_{k \geq 0} \sum_{\ell \in \mathcal{J}_k} \beta_{k,\ell} \psi_k(x; \eta_{k,\ell}), \quad x \in S,$$

where

$$\alpha_{0,\ell} = \langle f, \phi_0(\cdot; \zeta_{0,\ell}) \rangle, \quad \beta_{k,\ell} = \langle f, \psi_k(\cdot; \eta_{k,\ell}) \rangle. \quad (2.1)$$

We formulate the two following assumptions on \mathcal{E} :

- there exists a constant $C_1 > 0$ such that, for any integer $k \geq 0$,

$$\sum_{j \in \mathcal{I}_k} \frac{1}{g_{j,k,\ell}} |e_j(\zeta_{k,\ell})|^2 \leq C_1, \quad \sum_{j \in \mathcal{J}_k} \frac{1}{h_{j,k,\ell}} |e_j(\eta_{k,\ell})|^2 \leq C_1. \quad (2.2)$$

This assumption is obviously satisfied under **(A1)** with $C_1 = 1$. Remark also that the second example in (Goh, 2007, Section 4) satisfy both **(A2)** and (2.2).

- there exists a constant $C_2 > 0$ such that, for any integer $k \geq 0$,

$$\sup_{x \in S} \sum_{j \in \mathcal{J}_k} |e_j(x)|^2 \leq C_2 |\mathcal{J}_k|. \quad (2.3)$$

This assumption is satisfied by the three examples in Goh (2007) (we have $\sup_{x \in S} \sup_{j \in \mathcal{J}_k} |e_j(x)| \leq 1$). Remark that it contains (Prakasa Rao, 2010a, (3.16)).

2.2 Decomposition Spaces

Let $s > 0$ and $r > 0$. From the wavelet coefficients (2.1) of a function $f \in H$, we define the Besov spaces $\mathcal{B}_\infty^s(H)$ by

$$\mathcal{B}_\infty^s(H) = \left\{ f \in H; \sup_{m \geq 0} |\mathcal{J}_m|^{2s} \sum_{k \geq m} \sum_{\ell \in \mathcal{J}_k} |\beta_{k,\ell}|^2 < \infty \right\} \quad (2.4)$$

and the “weak Besov spaces” $\mathcal{W}^r(H)$ by

$$\mathcal{W}^r(H) = \left\{ f \in H; \sup_{\lambda > 0} \lambda^r \sum_{k \geq 0} \sum_{\ell \in \mathcal{J}_k} \mathbb{I}_{\{|\beta_{k,\ell}| \geq \lambda\}} < \infty \right\}, \quad (2.5)$$

where $\mathbb{I}_{\mathcal{A}}$ is the indicator function on \mathcal{A} .

Such kinds of function spaces are extensively used in approximation theory for the study of non linear procedures such as thresholding and greedy algorithms. See e.g. DeVore (1998) and Temlyakov (1998). From a statistical point of view, they are connected to the maxiset approach. See e.g. Cohen *et al.* (2001), Kerkycharian and Picard (2000) and Autin (2004).

3 Density Estimation for Functional Data

3.1 Problem statement

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space and $\{X_i; i \geq 1\}$ be i.i.d. random variables defined on $\{\Omega, \mathcal{F}, P\}$ and taking values in a separable Hilbert space \mathcal{H} associated with the corresponding Borel σ -algebra \mathcal{B} . Let P_X be the probability measure induced by X_1 on $(\mathcal{H}, \mathcal{B})$. Suppose that there exists a σ -finite measure ν on the measurable space $(\mathcal{H}, \mathcal{B})$ such that P_X is dominated by ν . As a consequence of the Radon-Nikodym theorem, there exists a nonnegative measurable function f such that

$$P_X(B) = \int_B f(x)\nu(dx), \quad B \in \mathcal{B}.$$

We aim to estimate f based on n observed functional data X_1, \dots, X_n .

We suppose that $f \in H$ where H is a separable Hilbert space or real-valued functions defined on \mathcal{H} and square integrable with respect to the σ -finite measure ν .

Moreover, we suppose that there exists a known constant $C_f > 0$ such that

$$\sup_{x \in S} f(x) \leq C_f. \quad (3.1)$$

3.2 Estimator

Adopting the notations of Section 2, we define the wavelet hard thresholding estimator \hat{f} by

$$\hat{f}(x) = \sum_{\ell \in \mathcal{L}_0} \hat{\alpha}_{0,\ell} \phi_0(x; \zeta_{0,\ell}) + \sum_{k=0}^{m_n} \sum_{\ell \in \mathcal{J}_k} \hat{\beta}_{k,\ell} \mathbb{I}_{\left\{|\hat{\beta}_{k,\ell}| \geq \kappa \sqrt{\frac{\ln n}{n}}\right\}} \psi_k(x; \eta_{k,\ell}), \quad (3.2)$$

$x \in S$, where

$$\hat{\alpha}_{k,\ell} = \frac{1}{n} \sum_{i=1}^n \phi_k(X_i; \zeta_{k,\ell}), \quad \hat{\beta}_{k,\ell} = \frac{1}{n} \sum_{i=1}^n \psi_k(X_i; \eta_{k,\ell}), \quad (3.3)$$

κ is a large enough constant and m_n is the integer satisfying

$$\frac{1}{2} \frac{n}{\ln n} < |\mathcal{J}_{m_n}| \leq \frac{n}{\ln n}.$$

The construction of \hat{f} consists in three steps: firstly, we estimate the unknown wavelet coefficients (2.1) of f by (3.3), secondly, we select only the “greatest” $\hat{\beta}_{k,\ell}$ via a hard thresholding (the “universal threshold” $\kappa(\ln n/n)^{1/2}$ is considered) and thirdly we reconstruct the selected ones on the initial wavelet basis. Details on the wavelet hard thresholding estimator for $H = \mathbb{L}_p([a, b])$ and the standard nonparametric models can be found in Donoho *et al.* (1996), Delyon and Juditsky (1996), Härdle *et al.* (1998) and Vidakovic (1999).

Note that our wavelet hard thresholding procedure is adaptive i.e. it does not depend on the knowledge of the smoothness of f .

3.3 Results

To evaluate the performance of \hat{f} , since we work in a general separable Hilbert space H , it seems interesting to adopt the maxiset approach introduced by Cohen *et al.* (2001). This is developed in the following theorem.

Theorem 3.1 *Consider the density estimation problem described in Subsection 3.1. We suppose that \mathcal{E} satisfies (2.2) and (2.3), and f satisfies (3.1). Let \hat{f} be (3.2). Then, for any $\theta \in (0, 1)$, we have the following equivalence*

$$\lim_{n \rightarrow \infty} \left(\frac{n}{\ln n} \right)^\theta E(\|\hat{f} - f\|^2) < \infty \Leftrightarrow f \in \mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H),$$

where $\mathcal{B}_\infty^{\theta/2}(H)$ is (2.4) with $s = \theta/2$ and $\mathcal{W}^{2(1-\theta)}(H)$ (2.5) with $r = 2(1 - \theta)$.

In other words, the maxiset associated to \hat{f} at the rate $(\ln n/n)^\theta$ is $\mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$.

An immediate consequence is the following upper bound result: if $f \in \mathcal{B}_\infty^{s/(2s+1)}(H) \cap \mathcal{W}^{2/(2s+1)}(H)$ for $s > 0$, then there exists a constant $C > 0$ such that

$$E(\|\hat{f} - f\|^2) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+1)}.$$

This rate of convergence corresponds to the near optimal one in the ‘‘standard’’ minimax setting (see e.g. Härdle *et al.* (1998)).

4 A Note on the Regression Estimation for Functional Data

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space and $\{(X_i, Y_i); i \geq 1\}$ be i.i.d. replica of a couple of random variables (X, Y) defined on $\{\Omega, \mathcal{F}, P\}$, where Y is real valued and X takes values in a separable Hilbert space \mathcal{H} associated with the corresponding Borel σ -algebra \mathcal{B} . Denote by ρ a given real measurable function. We want to estimate the regression function:

$$r(x) = E(\rho(Y)|X = x), \quad x \in S,$$

from $(X_1, Y_1), \dots, (X_n, Y_n)$.

Note that the kernel estimator of the regression function for functional data has been proposed by Ferraty and Vieu (2004).

Here, we suppose that there exists a known constant $C_\rho > 0$ such that

$$\sup_{y \in \mathbb{R}} \rho(y) \leq C_\rho. \quad (4.1)$$

Theorem 4.1 *Consider the regression estimation problem described above. We suppose that \mathcal{E} satisfies (2.2) and (2.3), and ρ satisfies (4.1). Let \hat{f} be as in (3.2) with*

$$\hat{\alpha}_{k,\ell} = \frac{1}{n} \sum_{i=1}^n \rho(Y_i) \phi_k(X_i; \zeta_{k,\ell}), \quad \hat{\beta}_{k,\ell} = \frac{1}{n} \sum_{i=1}^n \rho(Y_i) \psi_k(X_i; \eta_{k,\ell}).$$

Then, for any $\theta \in (0, 1)$, we have the following equivalence

$$\lim_{n \rightarrow \infty} \left(\frac{n}{\ln n} \right)^\theta E(\|\hat{f} - f\|^2) < \infty \Leftrightarrow f \in \mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H),$$

where $\mathcal{B}_\infty^{\theta/2}(H)$ is (2.4) with $s = \theta/2$ and $\mathcal{W}^{2(1-\theta)}(H)$ (2.5) with $r = 2(1 - \theta)$.

The proof of Theorem 4.1 is identical to the one of Theorem 3.1; thanks to (4.1), it is enough to replace $\phi_k(X_i; \zeta_{k,\ell})$ by $\rho(Y_i)\phi_k(X_i; \zeta_{k,\ell})$ and $\psi_k(X_i; \eta_{k,\ell})$ by $\rho(Y_i)\psi_k(X_i; \eta_{k,\ell})$.

5 Proofs

In this section, C denotes any constant that does not depend on j , k and n . Its value may change from one term to another and may depend on ϕ or ψ .

Proof of Theorem 3.1. The proof of Theorem 3.1 is a consequence of the general maxiset result (Kerkycharian and Picard, 2000, Theorems 3.1 and 3.2) with $c(n) = (\ln n/n)^{1/2}$, $\sigma_i = 1$, $r = 2$ and the following proposition.

Proposition 5.1 *For any $k \in \{0, \dots, m_n\}$ and any $\ell \in \mathcal{I}_k$ or $\ell \in \mathcal{J}_k$, let $\alpha_{k,\ell}$ and $\beta_{k,\ell}$ be (2.1), and $\hat{\alpha}_{k,\ell}$ and $\hat{\beta}_{k,\ell}$ be (3.3). Then*

(i) *there exists a constant $C > 0$ such that*

$$E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) \leq C \frac{\ln n}{n}.$$

(ii) *there exists a constant $C > 0$ such that*

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) \leq C \left(\frac{\ln n}{n} \right)^2.$$

(iii) *for $\kappa > 0$ large enough, we have*

$$P \left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}} \right) \leq 2 \left(\frac{\ln n}{n} \right)^2.$$

Let us now prove (i), (ii) and (iii) of Proposition 5.1 (which corresponds to (Kerkycharian and Picard, 2000, (3.1) and (3.2) of Theorem 3.1)).

(i) We have

$$E(\hat{\alpha}_{k,\ell}) = E(\phi_k(X_1; \zeta_{k,\ell})) = \int_{\mathcal{H}} f(x) \phi_k(x; \zeta_{k,\ell}) \nu(dx) = \alpha_{k,\ell}. \quad (5.1)$$

So

$$E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) = V(\hat{\alpha}_{k,\ell}) = \frac{1}{n}V(\phi_k(X_1; \zeta_{k,\ell})) \leq \frac{1}{n}E(|\phi_k(X_1; \zeta_{k,\ell})|^2).$$

It follows from (3.1), the fact that \mathcal{E} is an orthonormal basis of H and (2.2) that

$$\begin{aligned} E(|\phi_k(X_1; \zeta_{k,\ell})|^2) &= \int_{\mathcal{H}} |\phi_k(x; \zeta_{k,\ell})|^2 f(x) \nu(dx) \\ &\leq C_f \int_{\mathcal{H}} |\phi_k(x; \zeta_{k,\ell})|^2 \nu(dx) \\ &= C_f \int_{\mathcal{H}} \left| \sum_{j \in \mathcal{I}_k} \frac{1}{\sqrt{g_{j,k,\ell}}} e_j(\zeta_{k,\ell}) e_j(x) \right|^2 \nu(dx) \\ &= C_f \sum_{j \in \mathcal{I}_k} \frac{1}{g_{j,k,\ell}} |e_j(\zeta_{k,\ell})|^2 \leq C_f C_1. \end{aligned} \quad (5.2)$$

Therefore there exists a constant $C > 0$ such that

$$E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) \leq C \frac{1}{n} \leq C \frac{\ln n}{n}.$$

(ii) For the sake of simplicity, set

$$U_{i,k,\ell} = \psi_k(X_i; \eta_{k,\ell}) - E(\psi_k(X_i; \eta_{k,\ell})).$$

Then

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) = \frac{1}{n^4} E \left(\left| \sum_{i=1}^n U_{i,k,\ell} \right|^4 \right). \quad (5.3)$$

We will bound this last term via the Rosenthal inequality (recalled in the Appendix).

Proceeding as in (5.1), we obtain $E(U_{1,k,\ell}) = 0$.

By the Hölder inequality and (5.2) with $\psi_k(X_1; \eta_{k,\ell})$ instead of $\phi_k(X_1; \zeta_{k,\ell})$, we have

$$E(|U_{1,k,\ell}|^2) \leq CE(|\psi_k(X_1; \eta_{k,\ell})|^2) \leq C. \quad (5.4)$$

Let us now investigate the bound of $E(|U_{1,k,\ell}|^4)$. Observe that, thanks to the Cauchy-Schwarz inequality, (2.2) and (2.3), we have

$$\begin{aligned}
\sup_{x \in S} |\psi_k(x; \eta_{k,\ell})| &\leq \sup_{x \in S} \sum_{j \in \mathcal{J}_k} \frac{1}{\sqrt{h_{j,k,\ell}}} |e_j(\eta_{k,\ell})| |e_j(x)| \\
&\leq \left(\sum_{j \in \mathcal{J}_k} \frac{1}{h_{j,k,\ell}} |e_j(\eta_{k,\ell})|^2 \right)^{1/2} \left(\sup_{x \in S} \sum_{j \in \mathcal{J}_k} |e_j(x)|^2 \right)^{1/2} \\
&\leq C_1^{1/2} C_2^{1/2} \sqrt{|\mathcal{J}_k|} \leq C \sqrt{|\mathcal{J}_{m_n}|} \leq C \sqrt{\frac{n}{\ln n}}. \tag{5.5}
\end{aligned}$$

The Hölder inequality, (5.5) and (5.4) yield

$$\begin{aligned}
E(|U_{1,k,\ell}|^4) &\leq CE (|\psi_k(X_1; \eta_{k,\ell})|^4) \leq C \frac{n}{\ln n} E (|\psi_k(X_1; \eta_{k,\ell})|^2) \\
&\leq C \frac{n}{\ln n}. \tag{5.6}
\end{aligned}$$

It follows from the Rosenthal inequality, (5.4) and (5.6) that

$$\begin{aligned}
\frac{1}{n^4} E \left(\left| \sum_{i=1}^n U_{i,k,\ell} \right|^4 \right) &\leq C \frac{1}{n^4} \max \left(nE (|U_{1,k,\ell}|^4), (nE (|U_{1,k,\ell}|^2))^2 \right) \\
&\leq C \frac{1}{n^4} \max \left(\frac{n^2}{\ln n}, n^2 \right) = C \frac{1}{n^2} \leq C \left(\frac{\ln n}{n} \right)^2. \tag{5.7}
\end{aligned}$$

By (5.3) and (5.7), we prove the existence of a constant $C > 0$ such that

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) \leq C \left(\frac{\ln n}{n} \right)^2.$$

(iii) We adopt the same notations of (ii). Observe that

$$P \left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}} \right) = P \left(\left| \sum_{i=1}^n U_{i,k,\ell} \right| \geq n \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}} \right). \tag{5.8}$$

We will bound this probability via the Bernstein inequality (recalled in the Appendix).

We have $E(U_{1,k,\ell}) = 0$.

By (5.5),

$$|U_{i,k,\ell}| \leq C \sup_{x \in S} |\psi_k(x; \eta_{k,\ell})| \leq C \sqrt{\frac{n}{\ln n}}.$$

Applying (5.2) with $\psi_k(X_1; \eta_{k,\ell})$ instead of $\phi_k(X_1; \zeta_{k,\ell})$, we obtain $E(U_{1,k,\ell}^2) \leq C$.

It follows from the Bernstein inequality that

$$P \left(\left| \sum_{i=1}^n U_{i,k,\ell} \right| \geq n \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}} \right) \leq 2 \exp \left(- \frac{C n^2 \kappa^2 \frac{\ln n}{n}}{n + \frac{n \kappa \sqrt{\frac{\ln n}{n}} \sqrt{\frac{n}{\ln n}}}{3}} \right) \leq 2n^{-w(\kappa)}, \quad (5.9)$$

where

$$w(\kappa) = \frac{C \kappa^2}{1 + \kappa}.$$

Since $\lim_{\kappa \rightarrow \infty} w(\kappa) = \infty$, combining (5.8) and (5.9), and taking κ such that $w(\kappa) = 2$, we have

$$P \left(\left| \hat{\beta}_{k,\ell} - \beta_{k,\ell} \right| \geq \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}} \right) \leq \frac{2}{n^2} \leq 2 \left(\frac{\ln n}{n} \right)^2.$$

The proof of Proposition 5.1 is complete.

Appendix

Here we state the two inequalities that have been used for proving the results in earlier section.

Lemma 5.1 (Rosenthal (1970)) *Let n be a positive integer, $p \geq 2$ and V_1, \dots, V_n be n zero mean i.i.d. random variables such that $E(|V_1|^p) < \infty$. Then there exists a constant $C > 0$ such that*

$$E \left(\left| \sum_{i=1}^n V_i \right|^p \right) \leq C \max \left(n E(|V_1|^p), n^{p/2} (E(V_1^2))^{p/2} \right).$$

Lemma 5.2 (Petrov (1995)) *Let n be a positive integer and V_1, \dots, V_n be n i.i.d. zero mean independent random variables such that there exists a constant $M > 0$ satisfying $|V_1| \leq M < \infty$. Then, for any $v > 0$,*

$$P \left(\left| \sum_{i=1}^n V_i \right| \geq v \right) \leq 2 \exp \left(- \frac{v^2}{2 (n E(V_1^2) + v M / 3)} \right).$$

References

- Autin, F. (2004). Point de vue Maxiset en estimation non paramétrique. PhD thesis, Université Paris 7. (tel-00008542).
- Cohen, A., De Vore, R., Kerkyacharian, G. and Picard, D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11, 167-191.
- Dabo-Niang, S. (2002). Estimation de la densité dans un espace de dimension infinie: application aux diffusions. *C. R. Math. Acad. Sci. Paris*, 334, Vol. 3, 213–216.
- Dabo-Niang, S. (2004). Density estimation by orthogonal series in an infinite dimensional space: application to processes of diffusion type I, *J. Nonparametr. Stat.*, 16, no. 1-2, 171–186.
- Dabo-Niang, S. (2006). Mode estimation for functional random variable and its application for curves classification, *Far East J. Theor. Stat.*, 18, no. 1, 93–119.
- Dabo-Niang, S. and Ferraty, F. (Eds)(2008) *Functional and Operatorial Statistics. Contributions to Statistics*, Physica-Verlag, Heidelberg.
- Dalecky, Yu. L. and Fomin, S.V. (1991). *Measures and differential equations in infinite-dimensional space*. Dordrecht, Kluwer Academic Publishers Group.
- Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis* **3**, 215–228.
- DeVore, R. (1998). *Non linear approximation*. Acta Numerica, 51-150.
- Donoho, D., Johnstone, I., Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *Annals of Statistics*, 24, 2, 508-539.
- Ferraty, F. and Vieu, P. (2000), Dimension fractale et estimation de la rgression dans des espaces vectoriels semi-norms, *C. R. Acad. Sci. Paris Sr. I Math.*, 330, no. 2, 139–142.

- Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination, *J. Nonparametr. Stat.*, 16(1-2), 111–125.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and practice*. Springer-Verlag, New York.
- Ferraty, F. (Ed) (2010). Special Issue : Statistical Methods and Problems in Infinite-dimensional Spaces. 1st International Workshop on Functional and Operatorial Statistics (IWFOS'2008). *J. Multivariate Anal.*, Vol 101, Issue 2.
- Ferraty, F. and Romain, Y. (2010). *The oxford handbook of functional data analysis*, Oxford University Press.
- Ferraty, F. (Ed) (2011). *Recent Advances in Functional Data Analysis and Related Topics*. Contributions to Statistics, Physica-Verlag, Heidelberg.
- Goh, S.S. (2007). Wavelet bases for Hilbert spaces of functions. *Complex Variables and Elliptic Equations*, 52, 245-260.
- Härdle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998). Wavelet, Approximation and Statistical Applications, Lectures Notes in Statistics, New York 129, Springer Verlag.
- Kerkycharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases (with discussion and a rejoinder by the authors), *Test*, 9, 2, 283-345.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, 1065-1076.
- Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford: Clarendon Press.
- Prakasa Rao, B.L.S. (2010). Nonparametric Density Estimation for Functional Data via Wavelets. *Communications in Statistics - Theory and Methods*, 39:8-9, 1608-1618.

- Prakasa Rao, B.L.S. (2010). Nonparametric density estimation for functional data by delta sequences. *Brazilian Journal of Probability and Statistics*, 24, 3, 468-478.
- Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. New York, Springer.
- Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. New York, Springer.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27, 832-837.
- Rosenthal, H.P. (1970). On the subspaces of \mathbb{L}^p ($p \geq 2$) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8, 273-303.
- Temlyakov, V.N. (1998). The best m-term approximation and greedy algorithms. *Adv. Comput. Math.*, 8, 249-265.
- Uglanov, A.V. (2000). *Integration on infinite-dimensional surfaces and its applications*. Dordrecht, Kluwer Academic Publishers.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. John Wiley & Sons, Inc., New York, 384 pp.
- Xia, D.X. (1972). *Measure and integration theory on infinite-dimensional spaces. Abstract harmonic analysis*, New York, Academic Press.
- Yamasaki, Y. (1985). *Measures on infinite-dimensional spaces*, Series in Pure Mathematics 5, World Scientific, Singapore.