REVSTAT – Statistical Journal
Volume 11, Number 2, June 2013, 211–230

# NONPARAMETRIC ESTIMATION FOR FUNCTIONAL DATA BY WAVELET THRESHOLDING

Authors: Christophe Chesneau
– Université de Caen, LMNO, Campus II, Science 3,
14032, Caen, France
christophe.chesneau@unicaen.fr

Maher Kachour
– École supérieure de commerce IDRAC,
47, rue Sergent Michel Berthet CP 607, 69258 Lyon Cedex 09, France
maher.kachour@idraclyon.com

Bertrand Maillot
– Université de Caen, LMNO, Campus II, Science 3,
14032, Caen, France
Bertrand.Maillot@unicaen.fr

Received: February 2012 Revised: October 2012 Accepted: December 2012

Abstract:

- This paper deals with density and regression estimation problems for functional data. Using wavelet bases for Hilbert spaces of functions, we develop a new adaptive procedure based on a term-by-term selection of the wavelet coefficients estimators. We study its asymptotic performances by considering the mean integrated squared error over adapted decomposition spaces.

Key-Words:

- *functional data; density estimation; nonparametric regression; wavelets; hard thresholding.*

AMS Subject Classification:

- 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., [26], [27], [14], [7], [16], [17] and [18] for monographs and collective books on this specific subject). However, whereas there has been substantial work on the nonparametric estimation of the probability density function for univariate and multivariate random variables since the papers of [22] and [28], much less attention has been paid to 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 $\mu$ 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 [33], [34], [8] and [31].

The first consistency result for a kernel estimator of the density function for infinite dimensional random variables has been obtained in [4] 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 (see [5]), delta sequences (see [25]) or wavelets (see [24]), 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 [6]).

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 [12] and then using only probabilities of small balls in [13]. However, since these pioneering works, no adaptive estimator has been proposed.

Considering the density estimation problem from functional data, [24] has recently developed a new procedure based on the multiresolution approach on a separable Hilbert space introduced by [19]. This procedure belongs to the family of the linear wavelet estimators. As proved in [24, Theorem 3.1], it enjoys powerful asymptotic properties. However, such a linear wavelet estimator has two drawbacks: it is not adaptive (i.e., its performances 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, [24, Page 2 lines 14-16] states *"it would be interesting to investigate the advantage of these wavelet estimators for functional data by using wavelet thresholding suggested by [11]"*. This perspective motivates our study.

Adopting the multiresolution approach on a separable Hilbert space $H$ of [19], we construct an adaptive wavelet procedure extending the hard thresholding rule introduced by [11] to a general nonparametric estimation context for functional data. In order to study its asymptotic properties, we introduce two different kinds of decomposition spaces expressed in terms of wavelet coefficients via the new basis (see Section 2). They are related to the maxiset approach introduced by [3] and of interest as they contain a wide variety of unknown functions, complex or not. Exploring the density model and the regression model for functional data, we determine the rates of convergence attained by our estimator under the mean integrated squared error on $H$ and over the intersection of the two considering decomposition spaces. To the best of our knowledge, this study is the first one developing a wavelet-based adaptive estimator in the context of functional data (and studying it theoretically). Let us mention that the new findings includes several obtained results for $H = \mathbb{L}_p([a, b])$.

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 [19]. Let $H$ be a separable Hilbert space of real- or complex-valued functions defined on a complete separable metric space or a normed vector space $S$. Since $H$ is separable, it has an orthonormal basis $\mathcal{E} = \{e_j;\ j \in \Lambda\}$ for some countable index set $\Lambda$. As usual, we denote by $\langle ., . \rangle$ and $||.||$ 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 $\Lambda$ 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

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

satisfy one of the two following conditions:

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

(**A2**) $A_k A_k^* = \mathrm{diag}(d_{k,j})_{j\in\mathcal{I}_k}$ and $B_k B_k^* = \mathrm{diag}(t_{k,j})_{j\in\mathcal{J}_k}$, 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

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

where

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

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 [19, Theorem 2 (a)]).

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

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

where

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

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

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

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

  This assumption is obviously satisfied under (**A1**) with $C_1 = 1$. Remark also that the second example in [19, Section 4] satisfies both **(A2)** and (2.2).

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

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

  This assumption is satisfied by the three examples in [19] (we have $\sup_{x \in S} \sup_{j \in \mathcal{J}_k} |e_j(x)| \leq 1$). Remark that it contains [24, (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 $\mathcal{B}^s_\infty(H)$ by

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

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

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

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., [10] and [30]. From a statistical point of view, they are connected to the maxiset approach. See, e.g., [3], [21] and [1].

## 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 complete separable metric space or a Hilbert space $S$ associated with the corresponding Borel $\sigma$-algebra $\mathcal{B}$. Let $P_X$ be the probability measure induced by $X_1$ on $(S, \mathcal{B})$. Suppose that there exists a $\sigma$-finite measure $\nu$ on the measurable space $(S, \mathcal{B})$ such that $P_X$ is dominated by $\nu$. The Radon-Nikodym theorem ensures the existence of a nonnegative measurable function $f$ such that

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

In this context, we aim to estimate $f$ based on $n$ observed functional data $X_1, ..., X_n$.

We suppose that $f \in H$, where $H$ is a separable Hilbert space of real-valued functions defined on $S$ and square integrable with respect to the $\sigma$-finite measure $\nu$.

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

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

The estimation of the density function for functional data has been first addressed in [4], and the consistency in $L^2$-norm has been established in [5] for a projection estimator. More recently, [24] established the convergence in mean square -with rate- of a non adaptive wavelets based estimator. We refer to these papers for details and applications of the model.

### 3.2. Estimator

Following the procedure of [11] and adopting the notation of Section 2, we define the wavelet hard thresholding estimator $\hat{f}$ by

$$\hat{f}(x) = \sum_{\ell \in \mathcal{I}_0} \hat{\alpha}_{0,\ell} \phi_0(x; \zeta_{0,\ell}) + \sum_{k=0}^{m_n} \sum_{\ell \in \mathcal{J}_k} \hat{\beta}_{k,\ell} \mathbb{1}_{\left\{|\hat{\beta}_{k,\ell}| \geq \kappa \sqrt{\frac{\ln n}{n}}\right\}} \psi_k(x; \eta_{k,\ell}), \tag{3.2}$$

$x \in S$, where

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

$\kappa$ 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 and thirdly we reconstruct the selected elements of the initial wavelet basis. The choices of the threshold $\kappa(\ln n/n)^{1/2}$ (corresponding to the "universal threshold") and the definition of $m_n$ are based on theoretical considerations (see Theorem 3.1 below).

Note that $\hat{f}$ is adaptive, i.e., it does not depend on the knowledge of the smoothness of $f$. It can be viewed as an adaptive and thresholded version of the linear wavelet estimator proposed by [24]

Details on the wavelet hard thresholding estimator for $H = \mathbb{L}_p([a, b])$ and the standard nonparametric models can be found in [11], [9], [20] and [32].

### 3.3. Results

Theorem 3.1 below evaluates the performance of $\hat{f}$ assuming that $f$ belongs to the decomposition spaces described in Subsection 2.2.

**Theorem 3.1.** *Consider the density estimation problem described in Subsection 3.1. Suppose that $\mathcal{E}$ satisfies (2.2) and (2.3). Let $\hat{f}$ be given by (3.2). Suppose that $f$ satisfies (3.1) and, for any $\theta \in (0, 1)$, $f \in \mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$, where $\mathcal{B}_\infty^{\theta/2}(H)$ is (2.4) with $s = \theta/2$ and $\mathcal{W}^{2(1-\theta)}(H)$ (2.5) with $r = 2(1 - \theta)$. Then there exists a constant $C > 0$ such that*

$$E(||\hat{f} - f||^2) \leq C \left( \frac{\ln n}{n} \right)^\theta$$

*for $n$ large enough.*

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., [20]).

Moreover, applying [21, Theorem 3.2], one can prove that $\mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$ is the "maxiset" associated to $\hat{f}$ at the rate of convergence $(\ln n/n)^\theta$, i.e.,

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

These new theoretical results complete the work of [24] in the sense that our wavelet-based procedure is adaptive thanks to its term-by-term selection of the $\hat{\beta}_{k,\ell}$ and we prove that it achieves a suitable rate of convergence over a wide class of functions well adapted to our setting.

The next section considers another statistical problem: the the regression estimation for functional data. We show how adapt our wavelet methodology to this problem.

## 4. REGRESSION ESTIMATION FOR FUNCTIONAL DATA

### 4.1. Problem statement

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 complete separable metric space or a Hilbert space $S$ associated with the corresponding Borel $\sigma$-algebra $\mathcal{B}$ such that

$$Y = f(X) + \epsilon, \tag{4.1}$$

$f$ denotes an unknown regresion function and $\epsilon$ is a random variable independent of $X$ with $\epsilon \sim \mathcal{N}(0, 1)$. We suppose that $f \in H$ where $H$ is a separable Hilbert space of real-valued functions defined on $S$. Let $P_X$ be the probability measure induced by $X_1$ on $(S, \mathcal{B})$. Suppose that there exists a $\sigma$-finite measure $\nu$ on the measurable space $(S, \mathcal{B})$ such that $P_X$ is dominated by $\nu$. As a consequence of the Radon-Nikodym theorem, there exists a nonnegative measurable function $g$ such that

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

We suppose that $g$ is known.

In this context, we want to estimate $f$ from $(X_1, Y_1), ..., (X_n, Y_n)$.

The kernel estimator of the regression function for functional data has been proposed by [13] and the convergence in mean square of that estimator has been established by [15] with the rate $\mathcal{O}(h^2 + (nP(X \in \mathcal{B}(x, h))^{-1})$ where $h$ is the bandwidth. Note that the optimal choice of $h$ depends on the underlying unknown distribution.

We shall suppose that there exist two known constants $C_f > 0$ and $c_g > 0$ such that

$$\sup_{x \in S} f(x) \leq C_f, \qquad \inf_{x \in S} g(x) \geq c_g. \tag{4.2}$$

### 4.2. Results

Theorem 4.1 below explores the performance of $\hat{f}$ assuming that $f$ belongs to the decomposition spaces described in Subsection 2.2.

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

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

*$\kappa$ is a large enough constant and $m_n$ is the integer satisfying*

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

*Suppose that $f$ and $g$ satisfy* (4.2) *and, for any $\theta \in (0,1)$, $f \in \mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$, where $\mathcal{B}_\infty^{\theta/2}(H)$ is* (2.4) *with $s = \theta/2$ and $\mathcal{W}^{2(1-\theta)}(H)$* (2.5) *with $r = 2(1-\theta)$. Then there exists a constant $C > 0$ such that*

$$E(||\hat{f} - f||^2) \leq C\left(\frac{(\ln n)^2}{n}\right)^{\theta}$$

*for $n$ large enough.*

Again, note that, 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

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

This rate of convergence corresponds to the near optimal one in the "standard" minimax setting (see, e.g., [20]) up to an extra logarithmic term. To the best of our knowledge, Theorem 4.1 is first one studying an adaptive wavelet-based estimator for functional data in the nonparametric regression context.

## CONCLUSION AND PERSPECTIVES

We construct an efficient and new adaptive estimator for an unknown function $f$ belonging to a separable Hilbert space $H$. To reach this goal, we combine several existing techniques: the wavelet basis on $H$ developed by [19], the hard thresholding rule introduced by [11] and some elements related to the maxiset approach proposed by [3]. Rates of convergence are determined under the mean integrated squared error on $H$ over $\mathcal{B}_\infty^{\theta/2}(H) \cap \mathcal{W}^{2(1-\theta)}(H)$. Perspectives of this work are

- To determine the optimal lower bounds over the considered spaces,
- To remove the logarithmic term by perhaps considering other thresholding techniques. Thanks to its performances in numerous *i.i.d.* nonparametric models, the block thresholding introduced by [2] is a good candidate.

- Consider the regression model (4.1) with an unknown $g$.

These aspects require further investigations that we leave for a future work.

## 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 depends on $\phi$ or $\psi$.

**Proof of Theorem 3.1:** The proof of Theorem 3.1 is a consequence of [21, Theorem 3.1] with $c(n) = (\ln n/n)^{1/2}$, $\sigma_i = 1$, $r = 2$ and the following proposition.

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

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

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

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

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

**(iii)** *For $\kappa > 0$ large enough, there exists a constant $C > 0$ such that*

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

Let us now prove **(i)**, **(ii)** and **(iii)** of Proposition 5.1 (which corresponds to [21, (3.1) and (3.2) of Theorem 3.1]).

**(i)** We have

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

So

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

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

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

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)** Proceeding as in (5.1), we show that $E\left(\psi_k(X_i;\eta_{k,\ell})\right) = \beta_{k,\ell}$. Hence

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

where

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

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

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

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

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

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

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

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

$$E(|U_{1,k,\ell}|^4) \leq CE\left(|\psi_k(X_1;\eta_{k,\ell})|^4\right) \leq CnE\left(|\psi_k(X_1;\eta_{k,\ell})|^2\right) \leq Cn. \tag{5.6}$$

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

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

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 notation as in **(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),

$$|U_{1,k,\ell}| \leq C\sup_{x\in S}|\psi_k(x;\eta_{k,\ell})| \leq C\sqrt{\frac{n}{\ln n}}.$$

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

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

where

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

Since $\lim_{\kappa\to\infty} w(\kappa) = \infty$, combining (5.17) and (5.19), and taking $\kappa$ such that $w(\kappa) = 2$, we have

$$P\left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \leq C\frac{1}{n^2} \leq C\left(\frac{\ln n}{n}\right)^2.$$

The points **(i)**, **(ii)** and **(iii)** of Proposition 5.1 are proved. The proof of Theorem 3.1 is complete. □

**Proof of Theorem 4.1:** As in the proof of Theorem 3.1, we only need to prove **(i)**, **(ii)** and **(iii)** of Proposition 5.1.

**(i)** Since $X_1$ and $\epsilon_1$ are independent and $E(\epsilon_1) = 0$, we have

$$\begin{aligned} E(\hat{\alpha}_{k,\ell}) &= E\left(\frac{Y_1}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right) = E\left(\frac{f(X_1)}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right) \\ &= \int_S \frac{f(x)}{g(x)}\phi_k(x;\zeta_{k,\ell})g(x)\nu(dx) = \alpha_{k,\ell}. \end{aligned} \tag{5.10}$$

So

$$\begin{aligned} E(|\hat{\alpha}_{k,\ell} - \alpha_{k,\ell}|^2) &= V\left(\hat{\alpha}_{k,\ell}\right) = \frac{1}{n}V\left(\frac{Y_1}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right) \\ &\leq \frac{1}{n}E\left(\left|\frac{Y_1}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right|^2\right). \end{aligned}$$

It follows from (4.2), $|Y_1| \leq C_f + |\epsilon_1|$, $g(X_1) \geq c_g$, the independence between $X_1$ and $\epsilon_1$, $E(\epsilon_1^2) = 1$, the fact that $\mathcal{E}$ is an orthonormal basis of $H$ and (2.2) that

$$\begin{aligned} E\left(\left|\frac{Y_1}{g(X_1)}\phi_k(X_1;\zeta_{k,\ell})\right|^2\right) &\leq (C_f^2+1)\frac{1}{c_g}E\left(|\phi_k(X_1;\zeta_{k,\ell})|^2\,\frac{1}{g(X_1)}\right) \\ &= (C_f^2+1)\frac{1}{c_g}\int_S |\phi_k(x;\zeta_{k,\ell})|^2\frac{1}{g(x)}g(x)\nu(dx) \\ &= C\int_S |\phi_k(x;\zeta_{k,\ell})|^2\nu(dx) \\ &= C\int_S \left|\sum_{j\in\mathcal{I}_k}\frac{1}{\sqrt{g_{j,k,\ell}}}e_j(\zeta_{k,\ell})e_j(x)\right|^2\nu(dx) \\ &= C\sum_{j\in\mathcal{I}_k}\frac{1}{g_{j,k,\ell}}|e_j(\zeta_{k,\ell})|^2 \leq C. \end{aligned} \tag{5.11}$$

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)** Proceeding as in (5.10), we show that $E\left(Y_i\psi_k(X_i;\eta_{k,\ell})/g(X_i)\right) = \beta_{k,\ell}$. Set

$$U_{i,k,\ell} = \frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell}) - E\left(\frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\right).$$

and observe that

$$E(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}|^4) = \frac{1}{n^4}E\left(\left|\sum_{i=1}^n U_{i,k,\ell}\right|^4\right). \tag{5.12}$$

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

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

By the Hölder inequality and (5.11) with $\psi_k(X_1;\eta_{k,\ell})$ instead of $\phi_k(X_1;\zeta_{k,\ell})$, we obtain

$$E(|U_{1,k,\ell}|^2) \leq CE\left(\left|\frac{Y_1}{g(X_1)}\psi_k(X_1;\eta_{k,\ell})\right|^2\right) \leq C. \tag{5.13}$$

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

$$\begin{aligned}
\sup_{x\in S}|\psi_k(x;\eta_{k,\ell})| &\leq \sup_{x\in S}\sum_{j\in\mathcal{J}_k}\frac{1}{\sqrt{h_{j,k,\ell}}}|e_j(\eta_{k,\ell})||e_j(x)| \\
&\leq \left(\sum_{j\in\mathcal{J}_k}\frac{1}{h_{j,k,\ell}}|e_j(\eta_{k,\ell})|^2\right)^{1/2}\left(\sup_{x\in S}\sum_{j\in\mathcal{J}_k}|e_j(x)|^2\right)^{1/2} \\
&\leq C_1^{1/2}C_2^{1/2}\sqrt{|\mathcal{J}_k|} \leq C\sqrt{|\mathcal{J}_{m_n}|} \leq C\sqrt{\frac{n}{(\ln n)^2}}. \qquad (5.14)
\end{aligned}$$

The Hölder inequality, (5.14) and (5.13) yield

$$\begin{aligned}
E(|U_{1,k,\ell}|^4) &\leq CE\left(|\psi_k(X_1;\eta_{k,\ell})|^4\right) \leq CnE\left(|\psi_k(X_1;\eta_{k,\ell})|^2\right) \\
&\leq Cn. \qquad (5.15)
\end{aligned}$$

It follows from the Rosenthal inequality, (5.13) and (5.15) that

$$\begin{aligned}
\frac{1}{n^4}E\left(\left|\sum_{i=1}^n U_{i,k,\ell}\right|^4\right) &\leq C\frac{1}{n^4}\max\left(nE\left(|U_{1,k,\ell}|^4\right), \left(nE\left(|U_{1,k,\ell}|^2\right)\right)^2\right) \\
&\leq C\frac{1}{n^2} \leq C\left(\frac{\ln n}{n}\right)^2. \qquad (5.16)
\end{aligned}$$

By (5.12) and (5.16), 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 notation as in **(ii)**. Since $E\left(Y_i\psi_k(X_i;\eta_{k,\ell})/g(X_i)\right)$ $= \beta_{k,\ell}$, we can write

$$U_{i,k,\ell} = V_{i,k,\ell} + W_{i,k,\ell},$$

where

$$V_{i,k,\ell} = \frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\mathbb{I}_{\mathcal{A}_i} - E\left(\frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\mathbb{I}_{\mathcal{A}_i}\right),$$

$$W_{i,k,\ell} = \frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\mathbb{I}_{\mathcal{A}_i^c} - E\left(\frac{Y_i}{g(X_i)}\psi_k(X_i;\eta_{k,\ell})\mathbb{I}_{\mathcal{A}_i^c}\right),$$

$$\mathcal{A}_i = \left\{|\epsilon_i| \geq c_*\sqrt{\ln n}\right\}$$

and $c_*$ denotes a constant which will be chosen later.

We have

$$
\begin{aligned}
P\left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) &= P\left(\left|\sum_{i=1}^{n} U_{i,k,\ell}\right| \geq n\frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \\
&\leq I_1 + I_2, \qquad (5.17)
\end{aligned}
$$

where

$$
I_1 = P\left(\left|\sum_{i=1}^{n} V_{i,k,\ell}\right| \geq \frac{\kappa}{4}\sqrt{n \ln n}\right)
$$

and

$$
I_2 = P\left(\left|\sum_{i=1}^{n} W_{i,k,\ell}\right| \geq \frac{\kappa}{4}\sqrt{n \ln n}\right).
$$

Let us now bound $I_1$ and $I_2$.

*Upper bound for $I_1$.* The Markov inequality and the Cauchy-Schwarz inequality yield

$$
\begin{aligned}
I_1 &\leq \frac{4}{\kappa\sqrt{n \ln n}} E\left(\left|\sum_{i=1}^{n} V_{i,k,\ell}\right|\right) \leq C\sqrt{\frac{n}{\ln n}} E(|V_{1,k,\ell}|) \\
&\leq C\sqrt{\frac{n}{\ln n}} E\left(\left|\frac{Y_1}{g(X_1)}\psi_k(X_1;\eta_{k,\ell})\right| \mathbb{1}_{\mathcal{A}_1}\right) \\
&\leq C\sqrt{\frac{n}{\ln n}} \left(E\left(\left|\frac{Y_1}{g(X_1)}\psi_k(X_1;\eta_{k,\ell})\right|^2\right)\right)^{1/2} (P(\mathcal{A}_1))^{1/2}.
\end{aligned}
$$

Using (5.13), an elementary Gaussian inequality and taking $c_*$ large enough, we obtain

$$
I_1 \leq C\sqrt{\frac{n}{\ln n}} e^{-c_*^2 \ln n/4} \leq C\frac{1}{n^2}. \qquad (5.18)
$$

*Upper bound for $I_2$.* We will bound this probability via the Bernstein inequality (recalled in the Appendix).

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

Using (4.2) which implies $|Y_1 \mathbb{1}_{\mathcal{A}_1^c}| \leq C_f + c_*\sqrt{\ln n} \leq C\sqrt{\ln n}$ and $g(X_1) \geq c_g$, and (5.14), we obtain

$$
|W_{i,k,\ell}| \leq C\sqrt{\ln n} \sup_{x \in S} |\psi_k(x;\eta_{k,\ell})| \leq C\sqrt{\ln n}\sqrt{\frac{n}{(\ln n)^2}} = C\sqrt{\frac{n}{\ln n}}.
$$

Applying (5.11) with $\psi_k(X_1;\eta_{k,\ell})$ instead of $\phi_k(X_1;\zeta_{k,\ell})$, we obtain $E(|W_{1,k,\ell}|^2) \leq C$.

It follows from the Bernstein inequality that

$$I_2 \leq 2\exp\left(-\frac{Cn^2\kappa^2\frac{\ln n}{n}}{n + n\kappa\sqrt{\frac{\ln n}{n}}\sqrt{\frac{n}{\ln n}}}\right) \leq 2n^{-w(\kappa)}, \tag{5.19}$$

where

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

Since $\lim_{\kappa\to\infty} w(\kappa) = \infty$, taking $\kappa$ such that $w(\kappa) = 2$, we have

$$I_2 \leq 2\frac{1}{n^2}.$$

It follows from (5.17), (5.18) and (5.19) that

$$P\left(|\hat{\beta}_{k,\ell} - \beta_{k,\ell}| \geq \frac{\kappa}{2}\sqrt{\frac{\ln n}{n}}\right) \leq C\frac{1}{n^2} \leq C\left(\frac{\ln n}{n}\right)^2.$$

Hence the points **(i)**, **(ii)** and **(iii)** of Proposition 5.1 are satisfied by our estimators. The proof of Theorem 4.1 is complete. □

## APPENDIX

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

**Lemma A.1** ([29]). *Let $n$ be a positive integer, $p \geq 2$ and $V_1, ..., V_n$ 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*

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

**Lemma A.2** ([23]). *Let $n$ be a positive integer and $V_1, ..., V_n$ be $n$ i.i.d. zero mean random variables such that there exists a constant $M > 0$ satisfying $|V_1| \leq M < \infty$. Then, for any $\upsilon > 0$,*

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

## ACKNOWLEDGMENTS

The authors thank two anonymous referees for their thorough and useful comments.

## REFERENCES

[1] Autin, F. (2004). *Point de vue Maxiset en estimation non paramétrique*, PhD thesis, Université Paris 7. (tel-00008542).

[2] Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, **27**, 898–924.

[3] 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.

[4] Dabo-Niang, S. (2002). Estimation de la densité dans un espace de dimension infinie: application aux diffusions, *C. R. Math. Acad. Sci. Paris*, **334**(3), 213–216.

[5] 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**(1–2), 171–186.

[6] Dabo-Niang, S. (2006). Mode estimation for functional random variable and its application for curves classification, *Far East J. Theor. Stat.*, **18**(1), 93–119.

[7] Dabo-Niang, S. and Ferraty, F. (Eds) (2008). *Functional and Operatorial Statistics*, Contributions to Statistics, Physica-Verlag, Heidelberg.

[8] Dalecky, Yu.L. and Fomin, S.V. (1991). *Measures and Differential Equations in Infinite-Dimensional Space*, Dordrecht, Kluwer Academic Publishers Group.

[9] Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis*, **3**, 215–228.

[10] DeVore, R. (1998). Non linear approximation, *Acta Numerica*, **7**, 51–150.

[11] Donoho, D.; Johnstone, I.; Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Annals of Statistics*, **24**(2), 508–539.

[12] Ferraty, F. and Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés, *C. R. Acad. Sci. Paris Sér. I Math.*, **330**(2), 139–142.

[13] 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.

[14] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis. Theory and Practice*, Springer-Verlag, New York.

[15] Ferraty, F.; Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects, *Aust. N. Z. J. Stat.*, **49**, 267–286.

[16] 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.*, **101**(2).

[17] Ferraty, F. and Romain, Y. (2010). *The Oxford Handbook of Functional Data Analysis*, Oxford University Press.

[18] Ferraty, F. (Ed) (2011). *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, Physica-Verlag, Heidelberg.

[19] Goh, S.S. (2007). Wavelet bases for Hilbert spaces of functions, *Complex Variables and Elliptic Equations*, **52**, 245–260.

[20] Härdle, W.; Kerkyacharian, G.; Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics, 129, Springer-Verlag, New York.

[21] Kerkyacharian, 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.

[22] Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065Ű-1076.

[23] Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford: Clarendon Press.

[24] Prakasa Rao, B.L.S. (2010a). Nonparametric density estimation for functional data via wavelets, *Communications in Statistics — Theory and Methods*, **39**(8–9), 1608–1618.

[25] Prakasa Rao, B.L.S. (2010b). Nonparametric density estimation for functional data by delta sequences, *Brazilian Journal of Probability and Statistics*, **24**(3), 468–478.

[26] Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*, New York, Springer.

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

[28] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.*, **27**, 832–837.

[29] 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.

[30] Temlyakov, V.N. (1998). The best m-term approximation and greedy algorithms, *Adv. Comput. Math.*, **8**, 249–265.

[31] Uglanov, A.V. (2000). *Integration on Infinite-Dimensional Surfaces and its Applications*, Dordrecht, Kluwer Academic Publishers.

[32] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.

[33] Xia, D.X. (1972). *Measure and Integration Theory on Infinite-Dimensional Spaces. Abstract Harmonic Analysis*, New York, Academic Press.

[34] Yamasaki, Y. (1985). *Measures on Infinite-Dimensional Spaces*, Series in Pure Mathematics 5, World Scientific, Singapore.