Adaptive estimation of an additive regression function from weakly dependent data

Christophe Chesneau[*], Jalal Fadili[†] and Bertrand Maillot[*]

Abstract

A d-dimensional nonparametric additive regression model with dependent observations is considered. Using the marginal integration technique and wavelets methodology, we develop a new adaptive estimator for a component of the additive regression function. Its asymptotic properties are investigated via the minimax approach under the $\mathbb{L}_2$ risk over Besov balls. We prove that it attains a sharp rate of convergence which turns to be the one obtained in the *i.i.d.* case for the standard univariate regression estimation problem.

Keywords and phrases: Additive regression, Dependent data, Adaptivity, Wavelets, Hard thresholding.

AMS 2000 Subject Classifications: 62G07, 62G20.

1 Introduction

1.1 Problem statement

Let d be a positive integer, $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ be a $\mathbb{R} \times [0,1]^d$-valued strictly stationary process on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and ρ be a given real measurable function. The unknown

[*]LMNO, CNRS-Université de Caen, Campus II, Science 3, 14032, Caen, France.

[†]GREYC, CNRS-ENSICAEN-Universié de Caen, 6, Bd du Maréchal Juin, 14050 Caen Cedex, France.

regression function associated to $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ and ρ is defined by

$$g(\mathbf{x}) = \mathbb{E}(\rho(Y)|\mathbf{X} = \mathbf{x}), \qquad \mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d.$$

In the additive regression model, the function g is considered to have an additive structure, i.e. there exist d unknown real measurable functions $g_1, \ldots, g_d$ and an unknown real number μ such that

$$g(\mathbf{x}) = \mu + \sum_{\ell=1}^{d} g_\ell(x_\ell). \tag{1.1}$$

For any $\ell \in \{1, \ldots, d\}$, our goal is to estimate g_ℓ from n observations $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$ of $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$.

1.2 Overview of previous work

When $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ is a *i.i.d.* process, this additive regression model becomes the standard one. In such a case, Stone in a series of papers [34, 35, 36] proved that g can be estimated with the same rate of estimation error as in the one-dimensional case. The estimation of the component g_ℓ has been investigated in several papers via various methods (kernel, splines, wavelets, etc.). See, e.g., [4], [21], [23], [29, 30], [1], [2], [33], [40], [32] and [17].

In some applications, as dynamic economic systems and financial times series, the *i.i.d.* assumption on the observations is too stringent (see, e.g., [19] and [38]). For this reason, some authors have explored the estimation of g_ℓ in the dependent case. When $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ is a strongly mixing process, this problem has been addressed by [5], [11], and results for continuous time processes under a strong mixing condition have been obtained by [12, 13]. In particular, they have developed non-adaptive kernel estimators for g_ℓ and studied its asymptotic properties.

1.3 Contributions

To the best of our knowledge, adaptive estimation of g_ℓ for dependent processes has been addressed only by [18]. The lack of results for adaptive estimation in this context motivates this work. To reach our goal, as in [40], we combine the marginal integration technique introduced by [28] with wavelet methods. We capitalize on wavelets to construct an adaptive thresholding estimator and show that it attains sharp rates of convergence under mild assumptions on the smoothness of the unknown function. By adaptive, it is meant that the parameters of the estimator do not depend on the parameter(s) of the dependent process nor on those of the smoothness class of the function. In particular, this leads to a simple estimator.

More precisely, our wavelet estimator is based on term-by-term hard thresholding. The idea of this estimator is simple: (i) we estimate the unknown wavelet coefficients of g_ℓ based on the observations; (ii) then we select the greatest ones and ignore the others; (iii) and finally we reconstruct the function estimate from the chosen wavelet coefficients on the considered wavelet basis. Adopting the minimax point of view under the $\mathbb{L}_2$ risk, we prove that our adaptive estimator attains a sharp rate of convergence over Besov balls which capture a variety of smoothness features in a function including spatially inhomogeneous behavior. The attained rate corresponds to the optimal one in the *i.i.d.* case for the univariate regression estimation problem (up to an extra logarithmic term).

1.4 Paper organization

The rest of the paper is organized as follows. Section 2 presents our assumptions on the model. In Section 3, we describe wavelet bases on $[0,1]$, Besov balls and tensor product wavelet bases on $[0,1]^d$. Our wavelet hard thresholding estimator is detailed in Section 4. Its rate of convergence under the $\mathbb{L}_2$ risk over Besov balls is established in Section 5. A comprehensive simulation study is reported and discussed in Section 6. The proofs are detailed in Section 7.

2 Notations and assumptions

In this work, we assume the following on our model:

Assumptions on the variables.

- For any $i \in \{1, \ldots, n\}$, we set $\mathbf{X}_i = (X_{1,i}, \ldots, X_{d,i})$. We suppose that
  - for any $i \in \{1, \ldots, n\}$, $X_{1,i}, \ldots, X_{d,i}$ are identically distributed with the common distribution $\mathcal{U}([0,1])$,
  - $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are identically distributed with the common known density f.

- We suppose that the following identifiability condition is satisfied: for any $\ell \in \{1, \ldots, d\}$ and $i \in \{1, \ldots, n\}$, we have

$$\mathbb{E}(g_\ell(X_{\ell,i})) = 0. \tag{2.1}$$

Strongly mixing assumption. Throughout this work, we use the strong mixing dependence structure on $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$. For any $m \in \mathbb{Z}$, we define the m-th strongly mixing coefficient of $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ by

$$\alpha_m = \sup_{(A,B)\in\mathcal{F}^{(Y,\mathbf{X})}_{-\infty,0}\times\mathcal{F}^{(Y,\mathbf{X})}_{m,\infty}} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|, \tag{2.2}$$

where $\mathcal{F}^{(Y,\mathbf{X})}_{-\infty,0}$ is the σ-algebra generated by $\ldots, (Y_{-1}, \mathbf{X}_{-1}), (Y_0, \mathbf{X}_0)$ and $\mathcal{F}^{(Y,\mathbf{X})}_{m,\infty}$ is the σ-algebra generated by $(Y_m, \mathbf{X}_m), (Y_{m+1}, \mathbf{X}_{m+1}), \ldots$.

We suppose that there exist two constants $\gamma > 0$ and $\upsilon > 0$ such that, for any integer $m \geq 1$,

$$\alpha_m \leq \gamma \exp(-\upsilon m). \tag{2.3}$$

This kind of dependence is reasonably weak. Further details on strongly mixing dependence can be found in [3], [39], [16], [27] and [6].

Boundedness assumptions.

- We suppose that $\rho \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_\infty(\mathbb{R})$, i.e. there exist constants $C_1 > 0$ and $C_2 > 0$ (supposed known) such that

$$\int_{-\infty}^{\infty} |\rho(y)|dy \leq C_1, \tag{2.4}$$

$$\text{and} \quad \sup_{y\in\mathbb{R}} |\rho(y)| \leq C_2. \tag{2.5}$$

- We suppose that there exists a known constant $c_1 > 0$ such that

$$\inf_{\mathbf{x}\in[0,1]^d} f(\mathbf{x}) \geq c_1. \tag{2.6}$$

- For any $m \in \{1, \ldots, n\}$, let $f_{(Y_0, \mathbf{X}_0, Y_m, \mathbf{X}_m)}$ be the density of $(Y_0, \mathbf{X}_0, Y_m, \mathbf{X}_m)$, $f_{(Y_0, \mathbf{X}_0)}$ the density of $(Y_0, \mathbf{X}_0)$ and, for any $(y, \mathbf{x}, y_*, \mathbf{x}_*) \in \mathbb{R} \times [0,1]^d \times \mathbb{R} \times [0,1]^d$,

$$h_m(y, \mathbf{x}, y_*, \mathbf{x}_*) = f_{(Y_0, \mathbf{X}_0, Y_m, \mathbf{X}_m)}(y, \mathbf{x}, y_*, \mathbf{x}_*) - f_{(Y_0, \mathbf{X}_0)}(y, \mathbf{x}) f_{(Y_0, \mathbf{X}_0)}(y_*, \mathbf{x}_*). \tag{2.7}$$

We suppose that there exists a known constant $C_3 > 0$ such that

$$\sup_{m \in \{1, \ldots, n\}} \sup_{(y, \mathbf{x}, y_*, \mathbf{x}_*) \in \mathbb{R} \times [0,1]^d \times \mathbb{R} \times [0,1]^d} |h_m(y, \mathbf{x}, y_*, \mathbf{x}_*)| \leq C_3. \tag{2.8}$$

Such boundedness assumptions are standard for the estimation of g_ℓ from a strongly mixing process. The most common example where this assumption holds is when $\rho(y) = y \mathbf{1}_{\{|y| \leq M\}}$, where M denotes a positive constant. This corresponds to the nonparametric regression model $Y = g(\mathbf{X}) + \varepsilon$ with $\mathbb{E}(\varepsilon) = 0$, provided that ε and g are bounded from above. This is exactly the setting considered in the simulations of Section 6. See, e.g., [12, 13] or, for $\ell = d = 1$, [25] and [31].

3 Wavelets and Besov balls

This section presents basics on wavelets and the sequential definitions of the Besov balls.

3.1 Wavelet bases on $[0, 1]$

Let R be a positive integer. We consider an orthonormal wavelet basis generated by dilations and translations of the scaling and wavelet functions ϕ and ψ from the Daubechies family db_{2R}. In particular, ϕ and ψ have compact supports and unit $\mathbb{L}_2$-norm, and ψ has R vanishing moments, i.e. for any $r \in \{0, \ldots, R-1\}$, $\int x^r \psi(x) dx = 0$.

Define the scaled and translated version of ϕ and ψ

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \qquad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k).$$

Then, with an appropriate treatment at the boundaries, there exists an integer τ satisfying $2^\tau \geq 2R$ such that, for any integer $j_* \geq \tau$, the collection

$$\{\phi_{j_*,k}(.),\ k \in \{0, \ldots, 2^{j_*} - 1\};\ \psi_{j,k}(.);\ j \in \mathbb{N} - \{0, \ldots, j_* - 1\},\ k \in \{0, \ldots, 2^j - 1\}\},$$

is an orthonormal basis of $\mathbb{L}_2([0,1]) = \{h : [0,1] \to \mathbb{R}; \ \int_0^1 h^2(x)dx < \infty\}$. See [9, 24].

Consequently, for any integer $j_* \geq \tau$, any $h \in \mathbb{L}_2([0,1])$ can be expanded into a wavelet series as

$$h(x) = \sum_{k=0}^{2^{j_*}-1} \alpha_{j_*,k}\phi_{j_*,k}(x) + \sum_{j=j_*}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x), \qquad x \in [0,1],$$

where

$$\alpha_{j,k} = \int_0^1 h(x)\phi_{j,k}(x)dx, \qquad \beta_{j,k} = \int_0^1 h(x)\psi_{j,k}(x)dx. \tag{3.1}$$

3.2 Besov balls

As is traditional in the wavelet estimation literature, we will investigate the performance of our estimator by assuming that the unknown function to be estimated belongs to a Besov ball. The Besov norm for a function can be related to a sequence space norm on its wavelet coefficients. More precisely, let $M > 0$, $s \in (0, R)$, $p \geq 1$ and $q \geq 1$. A function h in $\mathbb{L}_2([0,1])$ belongs to the Besov ball $\mathbf{B}^s_{p,q}(M)$ of radius M if, and only if, there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (3.1) satisfy

$$\left(\sum_{j=\tau}^{\infty}\left(2^{j(s+1/2-1/p)}\left(\sum_{k=0}^{2^j-1}|\beta_{j,k}|^p\right)^{1/p}\right)^q\right)^{1/q} \leq M^*.$$

In this expression, s is a smoothness parameter and p and q are norm parameters. Besov spaces include many traditional smoothness spaces. For particular choices of s, p and q, Besov balls contain the standard Hölder and Sobolev balls. See [26].

3.3 Wavelet tensor product bases on $[0,1]^d$

For the purpose of this paper, we will use compactly supported tensor product wavelet bases on $[0,1]^d$ based on the Daubechies family. Let us briefly recall their construction. For any $\mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d$, we construct a scaling function

$$\Phi(\mathbf{x}) = \prod_{v=1}^{d} \phi(x_v) \ ,$$

and $2^d - 1$ wavelet functions

$$\Psi_u(\mathbf{x}) = \begin{cases} \psi(x_u) \displaystyle\prod_{\substack{v=1 \\ v \neq u}}^{d} \phi(x_v) & \text{when } u \in \{1, \ldots, d\}, \\ \displaystyle\prod_{v \in A_u} \psi(x_v) \prod_{v \notin A_u} \phi(x_v) & \text{when } u \in \{d+1, \ldots, 2^d - 1\}, \end{cases}$$

where $(A_u)_{u \in \{d+1,\ldots,2^d-1\}}$ forms the set of all non void subsets of $\{1, \ldots, d\}$ of cardinality greater or equal to 2.

For any integer j and any $\mathbf{k} = (k_1, \ldots, k_d)$, define the translated and dilated versions of Φ and Ψ_u as

$$\Phi_{j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2}\Phi(2^j x_1 - k_1, \ldots, 2^j x_d - k_d),$$
$$\Psi_{j,\mathbf{k},u}(\mathbf{x}) = 2^{jd/2}\Psi_u(2^j x_1 - k_1, \ldots, 2^j x_d - k_d), \text{ for any } u \in \{1, \ldots, 2^d - 1\}.$$

Let $D_j = \{0, \ldots, 2^j - 1\}^d$. Then, with an appropriate treatment at the boundaries, there exists an integer τ such that the collection

$$\{\Phi_{\tau,\mathbf{k}}, \mathbf{k} \in D_\tau;\ (\Psi_{j,\mathbf{k},u})_{u \in \{1,\ldots,2^d-1\}}, \quad j \in \mathbb{N} - \{0, \ldots, \tau - 1\},\ \mathbf{k} \in D_j\}$$

forms an orthonormal basis of $\mathbb{L}_2([0,1]^d) = \{h : [0,1]^d \to \mathbb{R};\ \int_{[0,1]^d} h^2(\mathbf{x})d\mathbf{x} < \infty\}$.

For any integer j_* such that $j_* \geq \tau$, a function $h \in \mathbb{L}_2([0,1]^d)$ can be expanded into a wavelet series as

$$h(\mathbf{x}) = \sum_{\mathbf{k} \in D_{j_*}} \alpha_{j_*,\mathbf{k}} \Phi_{j_*,\mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^d-1} \sum_{j=j_*}^{\infty} \sum_{\mathbf{k} \in D_j} \beta_{j,\mathbf{k},u} \Psi_{j,\mathbf{k},u}(\mathbf{x}), \qquad \mathbf{x} \in [0,1]^d,$$

where

$$\alpha_{j_*,\mathbf{k}} = \int_{[0,1]^d} h(\mathbf{x})\Phi_{j_*,\mathbf{k}}(\mathbf{x})d\mathbf{x}, \qquad \beta_{j,\mathbf{k},u} = \int_{[0,1]^d} h(\mathbf{x})\Psi_{j,\mathbf{k},u}(\mathbf{x})d\mathbf{x}. \tag{3.2}$$

4 The estimator

4.1 Wavelet coefficients estimator

The following proposition provides a wavelet decomposition of g_ℓ based on the "marginal integration" method (introduced by [28]) and a tensor product wavelet basis on $[0,1]^d$.

Proposition 4.1 *Suppose that* (2.1) *holds. Then, for any* $j_* \geq \tau$ *and* $\ell \in \{1, \ldots, d\}$*, we can write*

$$g_\ell(x) = \sum_{k=1}^{2^{j_*}-1} a_{j_*,k,\ell}\phi_{j_*,k}(x) + \sum_{j=j_*}^{\infty}\sum_{k=1}^{2^j-1} b_{j,k,\ell}\psi_{j,k}(x) - \mu, \qquad x \in [0,1],$$

where

$$a_{j_*,k,\ell} = a_{j_*,k_\ell,\ell} = 2^{-j_*(d-1)/2}\int_{[0,1]^d} g(\mathbf{x}) \sum_{\mathbf{k}_{-\ell}\in D^*_{j_*}} \Phi_{j_*,\mathbf{k}}(\mathbf{x})d\mathbf{x}, \tag{4.1}$$

$$b_{j,k,\ell} = b_{j,k_\ell,\ell} = 2^{-j(d-1)/2}\int_{[0,1]^d} g(\mathbf{x}) \sum_{\mathbf{k}_{-\ell}\in D^*_{j}} \Psi_{j,\mathbf{k},\ell}(\mathbf{x})d\mathbf{x}, \tag{4.2}$$

and $\mathbf{k}_{-\ell} = (k_1, \ldots, k_{\ell-1}, k_{\ell+1}, \ldots, k_d)$ *and* $D^*_j = \{0, \ldots, 2^j - 1\}^{d-1}$.

Remark 4.1 *Due to the definitions of* g *and properties of* $\Psi_{j,\mathbf{k},\ell}$*,* $b_{j,k,\ell}$ *is nothing but the wavelet coefficient of* g_ℓ*, i.e.*

$$b_{j,k,\ell} = \int_0^1 g_\ell(x)\psi_{j,k}(x)dx = \beta_{j,k}. \tag{4.3}$$

Proposition 4.1 suggests that a first step to estimate g_ℓ should consist in estimating the unknown coefficients $a_{j,k,\ell}$ (4.1) and $b_{j,k,\ell}$ (4.2). To this end, we propose the following estimators of the coefficients

$$\widehat{a}_{j,k,\ell} = \widehat{a}_{j,k_\ell,\ell} = 2^{-j(d-1)/2}\frac{1}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\sum_{\mathbf{k}_{-\ell}\in D^*_j}\Phi_{j,\mathbf{k}}(\mathbf{X}_i) \tag{4.4}$$

and

$$\widehat{b}_{j,k,\ell} = \widehat{b}_{j,k_\ell,\ell} = 2^{-j(d-1)/2}\frac{1}{n}\sum_{i=1}^{n}\frac{\rho(Y_i)}{f(\mathbf{X}_i)}\sum_{\mathbf{k}_{-\ell}\in D^*_j}\Psi_{j,\mathbf{k},\ell}(\mathbf{X}_i). \tag{4.5}$$

These estimators enjoy powerful statistical properties. Some of them are collected in the following propositions.

Proposition 4.2 (Unbiasedness) *Suppose that* (2.1) *holds. For any* $j \geq \tau$, $\ell \in \{1, \ldots, d\}$ *and* $k \in \{0, \ldots, 2^j - 1\}$, $\widehat{a}_{j,k,\ell}$ *and* $\widehat{b}_{j,k,\ell}$ *in* (4.4) *and* (4.5) *are unbiased estimators of* $a_{j,k,\ell}$ *and* $b_{j,k,\ell}$ *respectively.*

The key ingredient for the proof of Proposition 4.2 is Proposition 4.1.

Proposition 4.3 (Moment inequality I) *Suppose that the assumptions of Section 2 hold. Let $j \geq \tau$ such that $2^j \leq n$, $k \in \{0, \ldots, 2^j - 1\}$, $\ell \in \{1, \ldots, d\}$. Then there exists a constant $C_4 > 0$ such that*

$$\mathbb{E}\left((\widehat{a}_{j,k,\ell} - a_{j,k,\ell})^2\right) \leq C_4 \frac{1}{n}, \qquad \mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2\right) \leq C_4 \frac{1}{n}.$$

The proof of Proposition 4.3 is based on several covariance inequalities and the Davydov inequality for strongly mixing processes (see [10]).

Remark 4.2 *In the proof of Proposition 4.3, for the condition on α_m, we only need to have the existence of two constants $C_5 > 0$ and $q \in (0,1)$ such that $\sum_{m=1}^{n} m^q \alpha_m^q \leq C_5 < \infty$. This latter inequality is obviously satisfied by* (2.3).

Proposition 4.4 (Moment inequality II) *Under the same assumptions of Proposition 4.3, there exists a constant $C_6 > 0$ such that*

$$\mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^4\right) \leq C_6 \frac{2^j}{n}.$$

Proposition 4.5 (Concentration inequality) *Suppose that the assumptions of Section 2 hold. Let $j \geq \tau$ such that $\ln n \leq 2^j \leq n/(\ln n)^3$, $k \in \{0, \ldots, 2^j - 1\}$, $\ell \in \{1, \ldots, d\}$ and $\lambda_n = (\ln n/n)^{1/2}$. Then there exist two constants $C_7 > 0$ and $\kappa > 0$ such that*

$$\mathbb{P}\left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \geq \kappa \lambda_n / 2\right) \leq C_7 \frac{1}{n^4}.$$

The proof of Proposition 4.3 is based on a Bernstein like inequality for strongly mixing processes (see [22]).

4.2 Hard thresholding estimator

We now turn to the estimator of g_ℓ from $\widehat{a}_{j,k,\ell}$ and $\widehat{b}_{j,k,\ell}$ as introduced in (4.4) and (4.5). Towards this goal, we will only keep the significant wavelet coefficients that are above a certain threshold according to the hard thresholding rule, and then reconstruct from these coefficients. In a compact form, this reads

$$\widehat{g}_\ell(x) = \sum_{k=0}^{2^{j_0}-1} \widehat{a}_{j_0,k,\ell} \phi_{j_0,k}(x) + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \widehat{b}_{j,k,\ell} \mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell}| \geq \kappa \lambda_n\right\}} \psi_{j,k}(x) - \widehat{\mu}, \tag{4.6}$$

where j_0 is the resolution level satisfying $2^{j_0} = [\ln n]$,

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \rho(Y_i). \tag{4.7}$$

j_1 is the resolution level satisfying $2^{j_1} = [n/(\ln n)^3]$, $\mathbf{1}$ is the indicator function, κ is a large enough constant (the one in Proposition 4.5) and

$$\lambda_n = \sqrt{\frac{\ln n}{n}}.$$

The definitions of the parameters in $\widehat{g}_\ell$ are based on theoretical considerations (see the proof of Theorem 5.1). Let us mention that the threshold λ_n corresponds to the well-known universal one presented in [15] for the density estimation problem in a i.i.d. setting. Note that, due to the assumptions on the model, our wavelet hard thresholding estimator (4.6) is simpler than the one of [40]. Wavelet hard thresolding estimators for g (1.1) defined with only one component, i.e., $\ell = d = 1$ in a α-mixing dependence setting can be found in, e.g., [31] and [7, 8].

5 Minimax upper-bound result

Theorem 5.1 below investigates the minimax rates of convergence attained by $\widehat{g}_\ell$ over Besov balls under the $\mathbb{L}_2$ risk.

Theorem 5.1 *Let $\ell \in \{1, \ldots, d\}$. Suppose that the assumptions of Section 2 hold. Let $\widehat{g}_\ell$ be the estimator given in* (4.6). *Suppose that $g_\ell \in \mathbf{B}^s_{p,q}(M)$ with $q \geq 1$, $\{p \geq 2$ and $s \in (0, R)\}$ or $\{p \in [1, 2)$ and $s \in (1/p, R)\}$. Then there exists a constant $C_8 > 0$ such that*

$$\mathbb{E}\left(\int_0^1 (\widehat{g}_\ell(x) - g_\ell(x))^2 dx\right) \leq C_8 \left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

The proof of Theorem 5.1 is based on a suitable decomposition of the $\mathbb{L}_2$ risk and the statistical properties of (4.4) and (4.5) summarized in Propositions 4.2, 4.3, 4.4 and 4.5 above.

The rate $(\ln n/n)^{2s/(2s+1)}$ is, up to an extra logarithmic term, known to be the optimal one for the standard one-dimensional regression model with uniform random design in the i.i.d. case. See, e.g., [20] and [37]. In this setting, it is also the rate of convergence attained by the one-dimensional wavelet hard thresholding estimator. See, e.g., [14] and [20].

Theorem 5.1 provides an "adaptive contribution" to the results of [5], [11] and [12, 13]. Furthermore, if we confine ourselves to the *i.i.d.* case, we recover a similar result to [40, Theorem 3] but without the condition $s > \max(d/2, d/p)$. The price to pay is more restrictive assumptions on the model (ρ is bounded from above, the density of $\mathbf{X}$ is known, etc.). Additionally, our estimator has a more straightforward and friendly implementation than the one in [40].

6 Simulation results

In this section, a simulation study is conducted to illustrate the numerical performances of the above estimation procedure. Six test functions ("HeaviSine", "Parabolas', "Blocks", "Bumps", "Wave" and "Doppler") representing different degrees of smoothness were considered. These functions are displayed in Figure 1.

In the following, we will take $d = 2$. To generate n observations of the process $(Y_i, \mathbf{X}_i)$, we first consider the first-order autoregressive AR(1) model

$$\begin{aligned} Z_{\ell,i} &= \rho_\ell Z_{\ell,i-1} + e_{\ell,i}, \quad e_{\ell,i} \sim_{iid} \mathcal{N}(0,1), \quad \rho_\ell \in (0,1), \quad i = 2,\dots,n, \quad \ell \in \{1,2\} \\ V_i &= \mu V_{i-1} + \varepsilon_i, \quad \varepsilon_i \sim_{iid} \mathcal{N}(0,1), \quad \mu \in (0,1), \quad i = 2,\dots,n \ , \end{aligned}$$

with $Z_{\ell,1} \sim \mathcal{N}(0,1)$ (resp. $V_1 \sim \mathcal{N}(0,1)$) and independent of $e_{\ell,i}$ (resp. of ε_i) for $i \geq 2$, and $Z_{\ell,i}$ and V_i are mutually independent. Then, for any $i \geq 1$ and $\ell \in \{1,2\}$ we take

$$\begin{aligned} X_{\ell,i} &= F_{\mathcal{N}}\left(Z_{\ell,i}; 0, \tfrac{1}{1-\rho_\ell^2}\right) \\ W_i &= 2\sigma_W \left(F_{\mathcal{N}}\left(V_i; 0, \tfrac{1}{1-\mu^2}\right) - \tau X_{1,i} - (1+\tau)/2\right), \ \sigma_W > 0, \tau \geq 0 \\ Y_i &= g_1(X_{1,i}) + g_2(X_{2,i}) + W_i \ , \end{aligned}$$

where $F_{\mathcal{N}}(\cdot; 0, \sigma^2)$ is the zero-mean normal distribution with variance σ^2. The functions g_1 and g_2 are chosen among the test functions of Figure 1. Observe that when $\tau > 0$, the processes Y_i and $\mathbf{X}_i$ are not mutually independent.

First, the process $(V_i, \mathbf{Z}_i)$ is strictly stationary and strongly mixing. It follows from continuity of the distribution function that the stationarity and the mixing assumptions are met on the process $(Y_i, \mathbf{X}_i)$; see e.g. [16]. Second, it is immediate that for any i, $X_{1,i} \sim \mathcal{U}([0,1])$ and so is $X_{2,i}$ as required. In turn, this entails that the boundedness assumptions (2.6)-(2.8) hold. Third, as W_i is zero-mean, we obviously have $\mathbb{E}(Y_i|\mathbf{X}_i = \mathbf{x}_i) =$

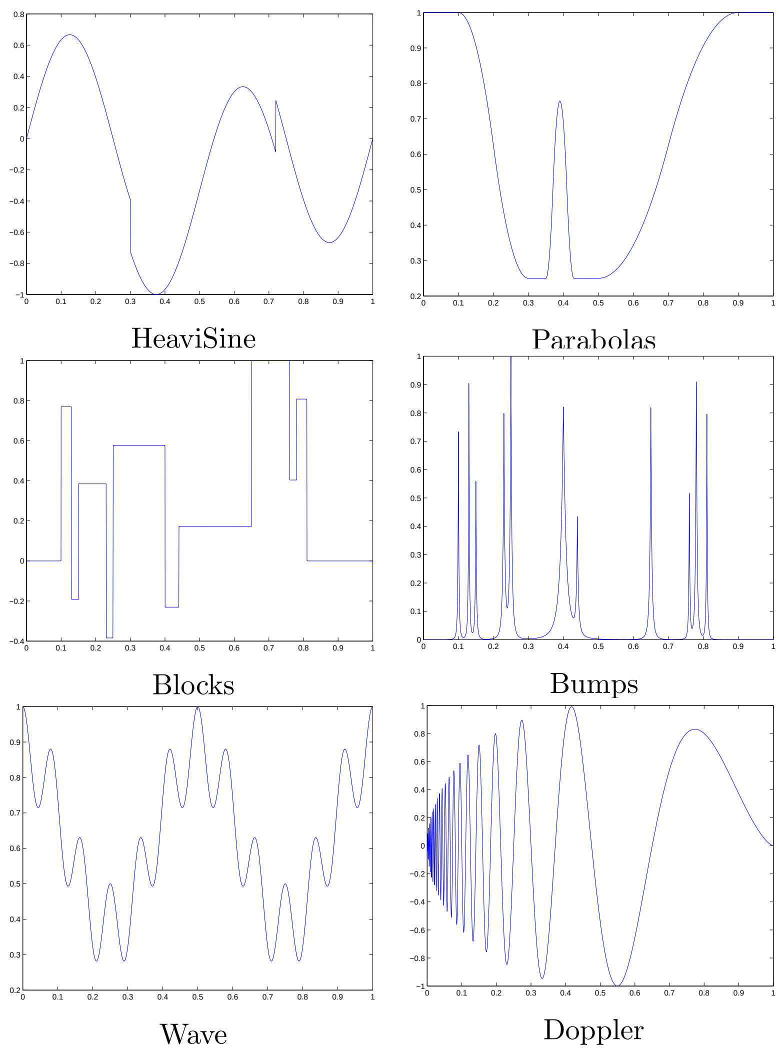

Figure 1: Original test functions.

$g_1(x_{1,i}) + g_2(x_{2,i})$, i.e. $\rho(Y) = Y$. In addition, since the noise process W_i, as well as the test functions considered here are all bounded from above[1], the boundedness assumptions (2.4)-(2.5) are in force.

In the following simulations, we set $\rho_1 = 0.5$, $\rho_2 = 0.9$, $\mu = 0.8$, $\tau = 1/10$, and the scale parameter $\sigma_W = 0.28$, i.e. the signal-to-noise ratio is 5. The Daubechies wavelet db_4 (i.e. $R = 2$) was used. The constant κ in the hard thresholding estimator was set to $\gamma\sigma_W$, where γ was chosen in $[0.25, 2.5]$ where it was observed empirically to lead to the best performance. The numerical performance of the estimator was measured using the Mean Squared Error (MSE), i.e.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(g(x_i) - \widehat{g}(x_i))^2 ,$$

where g is either of g_1 or g_2, $\widehat{g}$ its estimate, and $x_1, \ldots, x_n \in [0,1]$ are the corresponding observed sampling points.

Figure 2 displays the results of the estimator for different pairs of tested functions with two numbers of samples n. Visual inspection shows the good performance of our estimator which is able to adaptively recover a large class of functions spanning a wide range of spatial inhomogeneities. As expected, the estimation quality increases with growing n. This visual impression is confirmed quantitatively by Figure 3 and Figure 4. In these figures, the above simulation was repeated 100 times and the obtained MSE was averaged across these replications. Figure 3 depicts the boxplots of the MSE versus the function. Each plot corresponds to a fixed number of samples increasing from top to bottom. For a given number of samples, the average MSE and its variability is comparable for all functions, though they are slightly higher for the functions HeaviSine and Wave. As observed visually in Figure 2, the average MSE decreases with n as reported in Figure 4. Moreover, the average MSE shows a linear decreasing behaviour in log-log scale, which is clearly consistent with our theoretical convergence result.

[1]For $\tau = 0$, $W_1, \ldots, W_n$ are identically distributed with the common distribution $\mathcal{U}([-\sigma_W, \sigma_W])$. For $\tau > 0$, they are identically distributed according to the triangular distribution with support in $[-\sigma_W(1+\tau), \sigma_W(1+\tau)]$.

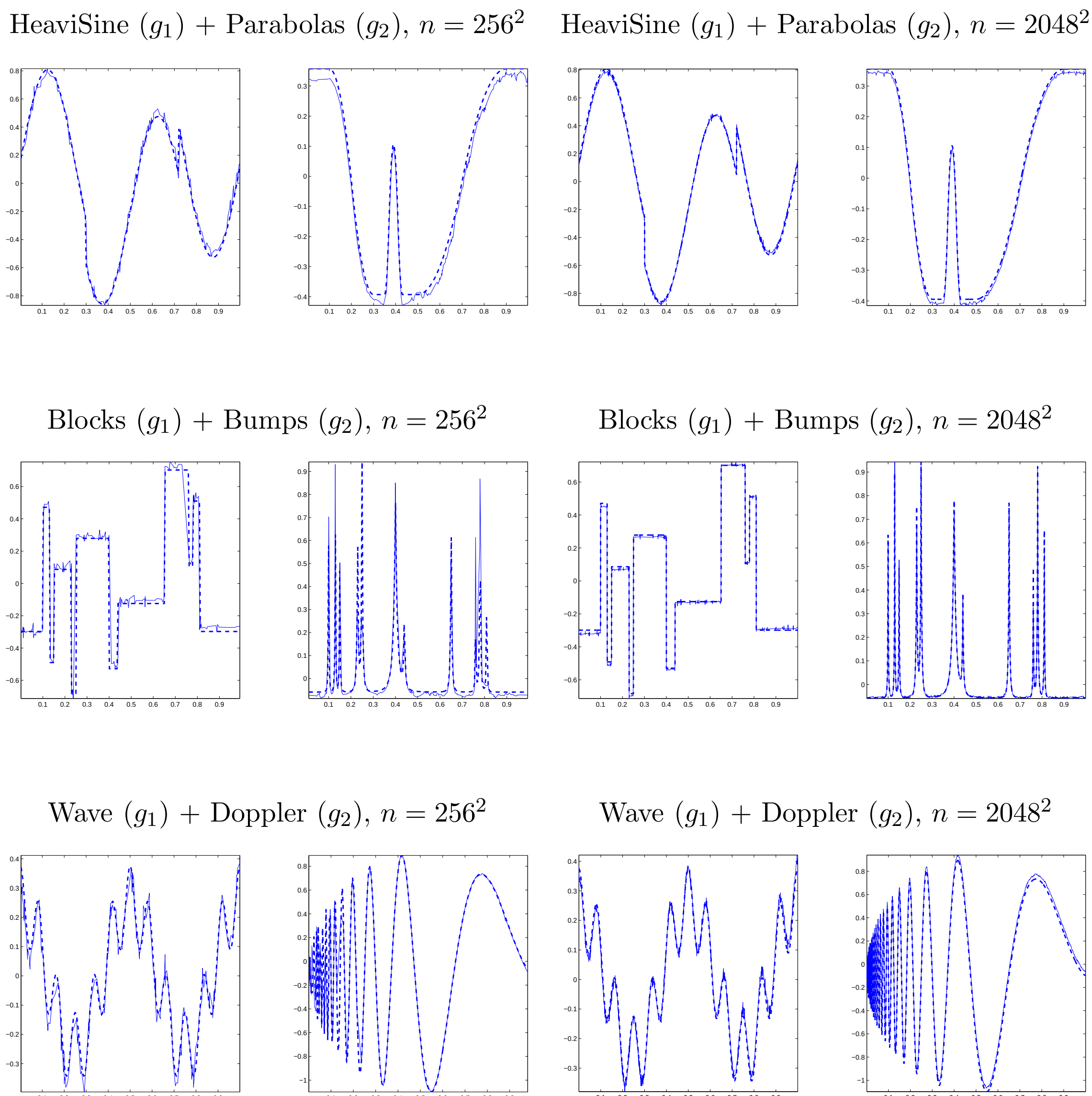

Figure 2: Original functions g_ℓ (dashed) and their estimates $\widehat{g}_\ell$ (solid), with $\ell = 1$ (left) and $\ell = 2$ (right) for different pairs of test functions, and two samples sizes ($n = 256^2$ and $n = 2048^2$).

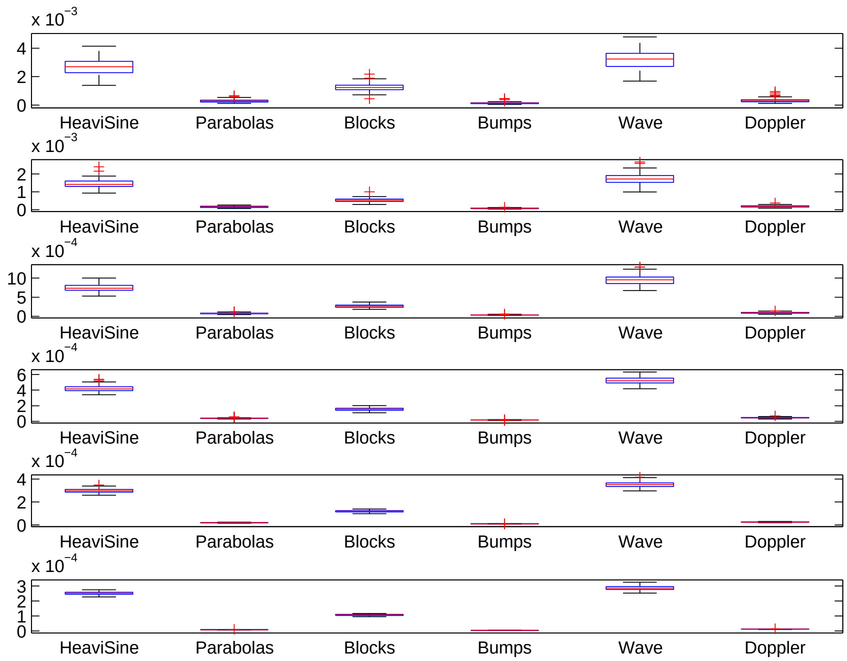

Figure 3: Boxplots of the average MSE for each tested function with the same pairs as those of Figure 2. Each plot corresponds to a number of samples $n \in \{2^{12}, 2^{14}, \ldots, 2^{22}\}$ increasingly from top to bottom.

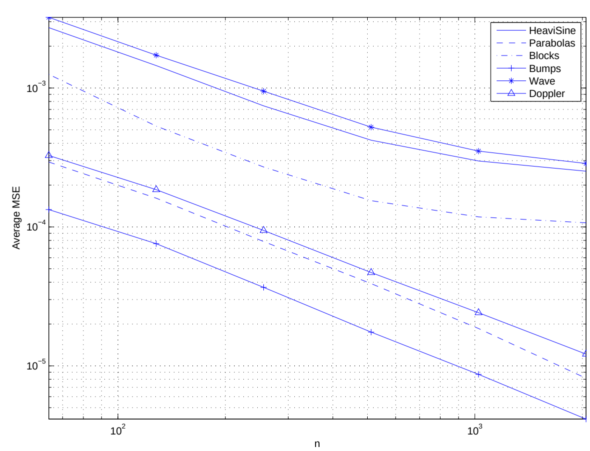

Figure 4: Average MSE over 100 replications as a function of the number of samples.

7 Proofs

In this section, the quantity 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 ϕ or ψ.

7.1 Technical results on wavelets

Proof of Proposition 4.1. Because of (2.5), we have $g \in \mathbb{L}_2([0,1]^d)$. For any $j_* \geq \tau$, we can expand g on our wavelet-tensor product basis as

$$g(\mathbf{x}) = \sum_{\mathbf{k} \in D_{j_*}} \alpha_{j_*,\mathbf{k}} \Phi_{j_*,\mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^d-1} \sum_{j=j_*}^{\infty} \sum_{\mathbf{k} \in D_j} \beta_{j,\mathbf{k},u} \Psi_{j,\mathbf{k},u}(\mathbf{x}), \qquad \mathbf{x} \in [0,1]^d \tag{7.1}$$

where

$$\alpha_{j_*,\mathbf{k}} = \int_{[0,1]^d} g(\mathbf{x}) \Phi_{j_*,\mathbf{k}}(\mathbf{x}) d\mathbf{x}, \qquad \beta_{j,\mathbf{k},u} = \int_{[0,1]^d} g(\mathbf{x}) \Psi_{j,\mathbf{k},u}(\mathbf{x}) d\mathbf{x}.$$

Moreover, using the "marginal integration" method based on (2.1), we can write

$$g_\ell(x_\ell) = \int_{[0,1]^{d-1}} g(\mathbf{x}) \prod_{\substack{v=1 \\ v \neq \ell}}^{d} dx_v - \mu, \qquad x_\ell \in [0,1]. \tag{7.2}$$

Since $\int_0^1 \phi_{j,k}(x)dx = 2^{-j/2}$ and $\int_0^1 \psi_{j,k}(x)dx = 0$, observe that

$$\int_{[0,1]^{d-1}} \Phi_{j_*,\mathbf{k}}(\mathbf{x}) \prod_{\substack{v=1 \\ v \neq \ell}}^{d} dx_v = 2^{-j_*(d-1)/2} \phi_{j_*,k_\ell}(x_\ell)$$

and

$$\int_{[0,1]^{d-1}} \Psi_{j,\mathbf{k},u}(\mathbf{x}) \prod_{\substack{v=1 \\ v \neq \ell}}^{d} dx_v = \begin{cases} 2^{-j(d-1)/2} \psi_{j,k_\ell}(x_\ell) & \text{if } u = \ell, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, putting (7.1) in (7.2) and writing $x = x_\ell$, we obtain

$$g_\ell(x) = \sum_{\mathbf{k} \in D_{j_*}} 2^{-j_*(d-1)/2} \alpha_{j_*,\mathbf{k}} \phi_{j_*,k_\ell}(x) + \sum_{j=j_*}^{\infty} \sum_{\mathbf{k} \in D_j} 2^{-j(d-1)/2} \beta_{j,\mathbf{k},\ell} \psi_{j,k_\ell}(x) - \mu.$$

Or, equivalently,

$$g_\ell(x) = \sum_{k=1}^{2^{j_*}-1} a_{j_*,k,\ell}\phi_{j_*,k}(x) + \sum_{j=j_*}^{\infty}\sum_{k=1}^{2^j-1} b_{j,k,\ell}\psi_{j,k}(x) - \mu,$$

where

$$a_{j,k,\ell} = a_{j,k_\ell,\ell} = 2^{-j(d-1)/2}\int_{[0,1]^d} g(\mathbf{x}) \sum_{\mathbf{k}_{-\ell}\in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x})d\mathbf{x}$$

and

$$b_{j,k,\ell} = b_{j,k_\ell,\ell} = 2^{-j(d-1)/2}\int_{[0,1]^d} g(\mathbf{x}) \sum_{\mathbf{k}_{-\ell}\in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x})d\mathbf{x}.$$

Proposition 4.1 is proved.

■

Proposition 7.1 *For any $\ell \in \{1,\dots,d\}$, $j \geq \tau$ and $k = k_\ell \in \{0,\dots,2^j-1\}$, set*

$$h_{j,k}^{(1)}(\mathbf{x}) = \sum_{\mathbf{k}_{-\ell}\in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}), \qquad h_{j,k}^{(2)}(\mathbf{x}) = \sum_{\mathbf{k}_{-\ell}\in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}), \qquad \mathbf{x}\in[0,1]^d.$$

Then there exists a constant $C>0$ such that, for any $a\in\{1,2\}$,

$$\sup_{\mathbf{x}\in[0,1]^d} |h_{j,k}^{(a)}(\mathbf{x})| \leq C2^{jd/2}, \qquad \int_{[0,1]^d} |h_{j,k}^{(a)}(\mathbf{x})|d\mathbf{x} \leq C2^{-j/2}2^{j(d-1)/2}$$

and

$$\int_{[0,1]^d} (h_{j,k}^{(a)}(\mathbf{x}))^2 d\mathbf{x} = 2^{j(d-1)}.$$

Proof:

- Since $\sup_{x\in[0,1]} |\phi_{j,k}(x)| \leq C2^{j/2}$ and $\sup_{x\in[0,1]} \sum_{k=0}^{2^j-1} |\phi_{j,k}(x)| \leq C2^{j/2}$, we obtain

$$\sup_{\mathbf{x}\in[0,1]^d} |h_{j,k}^{(1)}(\mathbf{x})| = (\sup_{x\in[0,1]} |\phi_{j,k}(x)|)\left(\sup_{x\in[0,1]}\sum_{k=0}^{2^j-1} |\phi_{j,k}(x)|\right)^{d-1} \leq C2^{jd/2}.$$

- Using $\int_0^1 |\phi_{j,k}(x)|dx = C2^{-j/2}$, we obtain

$$\begin{aligned}\int_{[0,1]^d} |h^{(1)}_{j,k}(\mathbf{x})|d\mathbf{x} &\leq \left(\int_0^1 |\phi_{j,k}(x)|dx\right)\left(\sum_{k=0}^{2^j-1}\int_0^1 |\phi_{j,k}(x)|dx\right)^{d-1} \\ &= C2^{-j/2}2^{j(d-1)/2}.\end{aligned}$$

- Since, for any $(u_{\mathbf{k}})_{\mathbf{k}\in D_j}$, $\int_{[0,1]^d}\left(\sum_{\mathbf{k}\in D_j} u_{\mathbf{k}}\Phi_{j,\mathbf{k}}(\mathbf{x})\right)^2 d\mathbf{x} = \sum_{\mathbf{k}\in D_j} u_{\mathbf{k}}^2$, we obtain

$$\int_{[0,1]^d} (h^{(1)}_{j,k}(\mathbf{x}))^2 d\mathbf{x} = \int_{[0,1]^d}\left(\sum_{\mathbf{k}_{-\ell}\in D_j^*}\Phi_{j,\mathbf{k}}(\mathbf{x})\right)^2 d\mathbf{x} = 2^{j(d-1)}.$$

  Proceeding in a similar fashion, using $\sup_{x\in[0,1]} |\psi_{j,k}(x)| \leq C2^{j/2}$, $\int_0^1 |\psi_{j,k}(x)|dx = C2^{-j/2}$ and, for any $(u_{\mathbf{k}})_{\mathbf{k}\in D_j}$, $\int_{[0,1]^d}\left(\sum_{\mathbf{k}\in D_j} u_{\mathbf{k}}\Psi_{j,\mathbf{k},\ell}(\mathbf{x})\right)^2 d\mathbf{x} = \sum_{\mathbf{k}\in D_j} u_{\mathbf{k}}^2$, we obtain the same results for $h^{(2)}_{j,k}$.

This ends the proof of Proposition 7.1.

■

7.2 Statistical properties of the coefficients estimators

Proof of Proposition 4.2. We have

$$\begin{aligned}\mathbb{E}(\widehat{a}_{j,k,\ell}) &= 2^{-j(d-1)/2}\mathbb{E}\left(\frac{\rho(Y_1)}{f(\mathbf{X}_1)}\sum_{\mathbf{k}_{-\ell}\in D_j^*}\Phi_{j,\mathbf{k}}(\mathbf{X}_1)\right) \\ &= 2^{-j(d-1)/2}\mathbb{E}\left(\mathbb{E}(\rho(Y_1)|\mathbf{X}_1)\frac{1}{f(\mathbf{X}_1)}\sum_{\mathbf{k}_{-\ell}\in D_j^*}\Phi_{j,\mathbf{k}}(\mathbf{X}_1)\right) \\ &= 2^{-j(d-1)/2}\mathbb{E}\left(\frac{g(\mathbf{X}_1)}{f(\mathbf{X}_1)}\sum_{\mathbf{k}_{-\ell}\in D_j^*}\Phi_{j,\mathbf{k}}(\mathbf{X}_1)\right) \\ &= 2^{-j(d-1)/2}\int_{[0,1]^d}\frac{g(\mathbf{x})}{f(\mathbf{x})}\sum_{\mathbf{k}_{-\ell}\in D_j^*}\Phi_{j,\mathbf{k}}(\mathbf{x})f(\mathbf{x})d\mathbf{x} \\ &= 2^{-j(d-1)/2}\int_{[0,1]^d} g(\mathbf{x})\sum_{\mathbf{k}_{-\ell}\in D_j^*}\Phi_{j,\mathbf{k}}(\mathbf{x})d\mathbf{x} = a_{j,k,\ell}.\end{aligned}$$

Proceeding in a similar fashion, we prove that $\mathbb{E}(\widehat{b}_{j,k,\ell}) = b_{j,k,\ell}$.

■

Proof of Proposition 4.3. For the sake of simplicity, for any $i \in \{1, \ldots, n\}$, set

$$Z_i = \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{X}_i).$$

Thanks to Proposition 4.2, we have

$$\mathbb{E}\left((\widehat{a}_{j,k,\ell} - a_{j,k,\ell})^2\right) = \mathbb{V}(\widehat{a}_{j,k,\ell}) = 2^{-j(d-1)} \frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^{n} Z_i\right). \tag{7.3}$$

An elementary covariance decomposition gives

$$\begin{aligned} \mathbb{V}\left(\sum_{i=1}^{n} Z_i\right) &= n\mathbb{V}(Z_1) + 2\sum_{v=2}^{n}\sum_{u=1}^{v-1} \mathrm{Cov}(Z_v, Z_u) \\ &\leq n\mathbb{V}(Z_1) + 2\left|\sum_{v=2}^{n}\sum_{u=1}^{v-1} \mathrm{Cov}(Z_v, Z_u)\right|. \end{aligned} \tag{7.4}$$

Using (2.5), (2.6) and Proposition 7.1, we have

$$\begin{aligned} \mathbb{V}(Z_1) &\leq \mathbb{E}(Z_1^2) \leq \frac{\sup_{y\in\mathbb{R}} \rho^2(y)}{\inf_{\mathbf{x}\in[0,1]^d} f(\mathbf{x})} \mathbb{E}\left(\frac{1}{f(\mathbf{X}_1)} \left(\sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{X}_1)\right)^2\right) \\ &\leq C \int_{[0,1]^d} \frac{1}{f(\mathbf{x})} \left(\sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x})\right)^2 f(\mathbf{x}) d\mathbf{x} \\ &= C \int_{[0,1]^d} \left(\sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x})\right)^2 d\mathbf{x} = C 2^{j(d-1)}. \end{aligned} \tag{7.5}$$

It follows from the stationarity of $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ and $2^j \leq n$ that

$$\left|\sum_{v=2}^{n}\sum_{u=1}^{v-1} \mathrm{Cov}(Z_v, Z_u)\right| = \left|\sum_{m=1}^{n} (n-m)\mathrm{Cov}(Z_0, Z_m)\right| \leq R_1 + R_2, \tag{7.6}$$

where

$$R_1 = n \sum_{m=1}^{2^j - 1} \left|\mathrm{Cov}(Z_0, Z_m)\right|, \qquad R_2 = n \sum_{m=2^j}^{n} \left|\mathrm{Cov}(Z_0, Z_m)\right|.$$

It remains to bound R_1 and R_2.

(i) **Bound for R_1.** Let, for any $(y, \mathbf{x}, y_*, \mathbf{x}_*) \in \mathbb{R} \times [0,1]^d \times \mathbb{R} \times [0,1]^d$, $h_m(y, \mathbf{x}, y_*, \mathbf{x}_*)$ be (2.7). Using (2.8), (2.4) and Proposition 7.1, we obtain

$$\begin{aligned}
&|\mathrm{Cov}\,(Z_0, Z_m)| \\
&= \Bigg| \int_{-\infty}^{\infty} \int_{[0,1]^d} \int_{-\infty}^{\infty} \int_{[0,1]^d} h_m(y, \mathbf{x}, y_*, \mathbf{x}_*) \times \\
&\quad \left(\frac{\rho(y)}{f(\mathbf{x})} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) \frac{\rho(y_*)}{f(\mathbf{x}_*)} \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}_*) \right) dy d\mathbf{x} dy_* d\mathbf{x}_* \Bigg| \\
&\leq \int_{-\infty}^{\infty} \int_{[0,1]^d} \int_{-\infty}^{\infty} \int_{[0,1]^d} |h_m(y, \mathbf{x}, y_*, \mathbf{x}_*)| \times \\
&\quad \left| \frac{\rho(y)}{f(\mathbf{x})} \right| \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) \right| \left| \frac{\rho(y_*)}{f(\mathbf{x}_*)} \right| \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}_*) \right| dy d\mathbf{x} dy_* d\mathbf{x}_* \\
&\leq C \left(\int_{-\infty}^{\infty} |\rho(y)| dy \right)^2 \left(\int_{[0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) \right| d\mathbf{x} \right)^2 \leq C 2^{-j} 2^{j(d-1)}.
\end{aligned}$$

Therefore

$$R_1 \leq C n 2^{-j} 2^{j(d-1)} 2^j = C n 2^{j(d-1)}. \tag{7.7}$$

(ii) **Bound for R_2.** By the Davydov inequality for strongly mixing processes (see [10]), for any $q \in (0,1)$, we have

$$\begin{aligned}
&|\mathrm{Cov}\,(Z_0, Z_m)| \leq 10 \alpha_m^q \left(\mathbb{E} \left(|Z_0|^{2/(1-q)} \right) \right)^{1-q} \\
&\leq 10 \alpha_m^q \left(\frac{\sup_{y \in \mathbb{R}} |\rho(y)|}{\inf_{\mathbf{x} \in [0,1]^d} f(\mathbf{x})} \sup_{\mathbf{x} \in [0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) \right| \right)^{2q} \left(\mathbb{E}(Z_0^2) \right)^{1-q}.
\end{aligned}$$

By (2.5), (2.6) and Proposition 7.1, we have

$$\begin{aligned}
\frac{\sup_{y \in \mathbb{R}} |\rho(y)|}{\inf_{\mathbf{x} \in [0,1]^d} f(\mathbf{x})} \sup_{\mathbf{x} \in [0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) \right| &\leq C \sup_{\mathbf{x} \in [0,1]^d} \left| \sum_{\mathbf{k}_{-\ell} \in D_j^*} \Phi_{j,\mathbf{k}}(\mathbf{x}) \right| \\
&\leq C 2^{jd/2}.
\end{aligned}$$

By (7.5), we have

$$\mathbb{E}\left(Z_0^2\right) \leq C 2^{j(d-1)}.$$

Therefore

$$|\mathrm{Cov}\,(Z_0, Z_m)| \leq C2^{qj}2^{j(d-1)}\alpha_m^q.$$

Observe that $\sum_{m=1}^{\infty} m^q \alpha_m^q = \gamma^q \sum_{m=1}^{\infty} m^q \exp(-cqm) < \infty$. Hence

$$R_2 \leq Cn2^{qj}2^{j(d-1)} \sum_{m=2^j}^{n} \alpha_m^q \leq Cn2^{j(d-1)} \sum_{m=2^j}^{n} m^q \alpha_m^q \leq Cn2^{j(d-1)}. \tag{7.8}$$

Putting (7.6), (7.7) and (7.8) together, we have

$$\left| \sum_{v=2}^{n} \sum_{u=1}^{v-1} \mathrm{Cov}\,(Z_v, Z_u) \right| \leq Cn2^{j(d-1)}. \tag{7.9}$$

Combining (7.3), (7.4), (7.5) and (7.9), we obtain

$$\mathbb{E}\left((\widehat{a}_{j,k,\ell} - a_{j,k,\ell})^2\right) \leq C2^{-j(d-1)}\frac{1}{n^2}n2^{j(d-1)} = C\frac{1}{n}.$$

Proceeding in a similar fashion, we prove that

$$\mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2\right) \leq C\frac{1}{n}.$$

This ends the proof of Proposition 4.3.

■

Proof of Proposition 4.4. It follows from (2.5), (2.6) and Proposition 7.1 that

$$\begin{aligned}
|\widehat{b}_{j,k,\ell}| &\leq 2^{-j(d-1)/2}\frac{1}{n}\sum_{i=1}^{n}\frac{|\rho(Y_i)|}{|f(\mathbf{X}_i)|}\left|\sum_{\mathbf{k}_{-\ell}\in D_j^*}\Psi_{j,\mathbf{k},\ell}(\mathbf{X}_i)\right| \\
&\leq 2^{-j(d-1)/2}\frac{\sup_{y\in\mathbb{R}}|\rho(y)|}{\inf_{\mathbf{x}\in[0,1]^d} f(\mathbf{x})}\sup_{\mathbf{x}\in[0,1]^d}\left|\sum_{\mathbf{k}_{-\ell}\in D_j^*}\Psi_{j,\mathbf{k},\ell}(\mathbf{x})\right| \\
&\leq C2^{-j(d-1)/2}2^{jd/2} = C2^{j/2}.
\end{aligned}$$

Because of (2.5), we have $\sup_{\mathbf{x}\in[0,1]^d} |g(\mathbf{x})| \leq C$. It follows from Proposition 7.1 that

$$\begin{aligned} |b_{j,k,\ell}| &\leq 2^{-j(d-1)/2} \int_{[0,1]^d} |g(\mathbf{x})| \left| \sum_{\mathbf{k}_{-\ell}\in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}) \right| d\mathbf{x} \\ &\leq C 2^{-j(d-1)/2} \int_{[0,1]^d} \left| \sum_{\mathbf{k}_{-\ell}\in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{x}) \right| d\mathbf{x} \\ &\leq C 2^{-j(d-1)/2} 2^{-j} 2^{jd/2} = C 2^{-j/2}. \end{aligned} \tag{7.10}$$

Hence

$$|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \leq |\widehat{b}_{j,k,\ell}| + |b_{j,k,\ell}| \leq C 2^{j/2}. \tag{7.11}$$

It follows from (7.11) and Proposition 4.3 that

$$\mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^4\right) \leq C 2^j \mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2\right) \leq C\frac{2^j}{n}.$$

The proof of Proposition 4.4 is complete.

■

Proof of Proposition 4.5. Let us first state a Bernstein inequality for exponentially strongly mixing process.

Lemma 7.1 ([22]) *Let $(Y_i)_{i\in\mathbb{Z}}$ be a strictly stationary process with the m-th strongly mixing coefficient α_m (2.2). Let n be a positive integer, $h : \mathbb{R} \to \mathbb{C}$ be a measurable function and, for any $i \in \mathbb{Z}$, $U_i = h(Y_i)$. We assume that $\mathbb{E}(U_1) = 0$ and there exists a constant $M > 0$ satisfying $|U_1| \leq M$. Then, for any $m \in \{1, \ldots, [n/2]\}$ and $\lambda > 0$, we have*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n U_i\right| \geq \lambda\right) \leq 4\exp\left(-\frac{\lambda^2 n}{16(D_m/m + \lambda M m/3)}\right) + 32\frac{M}{\lambda} n\alpha_m,$$

where $D_m = \max_{l\in\{1,\ldots,2m\}} \mathbb{V}\left(\sum_{i=1}^l U_i\right)$.

We now apply this lemma by setting for any $i \in \{1, \ldots, n\}$,

$$U_i = 2^{-j(d-1)/2} \frac{\rho(Y_i)}{f(\mathbf{X}_i)} \sum_{\mathbf{k}_{-\ell}\in D_j^*} \Psi_{j,\mathbf{k},\ell}(\mathbf{X}_i) - b_{j,k,\ell}.$$

Then we can write

$$\widehat{b}_{j,k,\ell} - b_{j,k,\ell} = \frac{1}{n}\sum_{i=1}^{n} U_i.$$

So

$$\mathbb{P}\left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \geq \kappa\lambda_n/2\right) = \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_i\right| \geq \kappa\lambda_n/2\right),$$

where $U_1, \ldots, U_n$ are identically distributed, depend on $(Y_i, \mathbf{X}_i)_{i\in\mathbb{Z}}$ satisfying (2.3),

- by Proposition 4.2, we have $\mathbb{E}(U_1) = 0$,
- using arguments similar to the bound of R_1 in the proof of Proposition 4.3 with l instead of n satisfying $l \leq C\ln n$ and $2^{-j} \leq 1/\ln n$, we prove that

$$\mathbb{V}\left(\sum_{i=1}^{l} U_i\right) \leq C\left(l + l^2 2^{-j}\right) \leq C\left(l + \frac{l^2}{\ln n}\right) \leq Cl.$$

  Hence

$$D_m = \max_{l\in\{1,\ldots,2m\}} \mathbb{V}\left(\sum_{i=1}^{l} U_i\right) \leq Cm.$$

- proceeding in a similar fashion to (7.11), we obtain $|U_1| \leq C2^{j/2}$.

Lemma 7.1 applied with the random variables $U_1, \ldots, U_n$, $\lambda = \kappa\lambda_n/2$, $\lambda_n = (\ln n/n)^{1/2}$, $m = [u\ln n]$ with $u > 0$ (chosen later), $M = C2^{j/2}$, $2^j \leq n/(\ln n)^3$ and (2.3) gives

$$\begin{aligned}
&\mathbb{P}\left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \geq \kappa\lambda_n/2\right)\\
&\leq C\left(\exp\left(-C\frac{\kappa^2\lambda_n^2 n}{D_m/m + \kappa\lambda_n mM}\right) + \frac{M}{\lambda_n} n\exp(-\upsilon m)\right)\\
&\leq C\left(\exp\left(-C\frac{\kappa^2\ln n}{1 + \kappa u 2^{j/2}\ln n(\ln n/n)^{1/2}}\right) + \frac{2^{j/2}}{(\ln n/n)^{1/2}} n\exp(-\upsilon u\ln n)\right)\\
&\leq C\left(n^{-C\kappa^2/(1+\kappa u)} + n^{2-\upsilon u}\right).
\end{aligned}$$

Therefore, for large enough κ and u, we have

$$\mathbb{P}\left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| \geq \kappa\lambda_n/2\right) \leq C\frac{1}{n^4}.$$

This ends the proof of Proposition 4.5.

■

7.3 Proof of Theorem 5.1

Using Proposition 4.1, we have

$$\begin{aligned}
&\widehat{g}_\ell(x) - g_\ell(x) \\
&= \sum_{k=0}^{2^{j_0}-1} (\widehat{\alpha}_{j_0,k,\ell} - \alpha_{j_0,k,\ell})\phi_{j_0,k}(x) + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} (\widehat{b}_{j,k,\ell}\mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell}|\geq \kappa\lambda_n\right\}} - b_{j,k,\ell})\psi_{j,k}(x) \\
&- \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j-1} b_{j,k,\ell}\psi_{j,k}(x) - (\widehat{\mu} - \mu).
\end{aligned}$$

Using the elementary inequality: $(x+y)^2 \leq 2(x^2+y^2)$, $(x,y) \in \mathbb{R}^2$, and the orthonormal property of the wavelet basis, we have

$$\mathbb{E}\left(\int_0^1 (\widehat{g}_\ell(x) - g_\ell(x))^2 dx\right) \leq 2(T + U + V + W), \tag{7.12}$$

where

$$T = \mathbb{E}((\widehat{\mu} - \mu)^2), \qquad U = \sum_{k=0}^{2^{j_0}-1} \mathbb{E}\left((\widehat{\alpha}_{j_0,k,\ell} - \alpha_{j_0,k,\ell})^2\right),$$

$$V = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left((\widehat{b}_{j,k,\ell}\mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell}|\geq \kappa\lambda_n\right\}} - b_{j,k,\ell})^2\right), \qquad W = \sum_{j=j_1+1}^{\infty} \sum_{k=0}^{2^j-1} b_{j,k,\ell}^2.$$

(i) **Bound for T.** We proceed as in the proof of Proposition 4.3. By (2.1), we have $\mathbb{E}(\rho(Y_1)) = \mu$. Thanks to the stationarity of $(Y_i)_{i\in\mathbb{Z}}$, we have

$$T = \mathbb{V}(\widehat{\mu}) \leq \frac{1}{n}\mathbb{V}(\rho(Y_1)) + 2\frac{1}{n}\sum_{m=1}^{n} \left|\mathrm{Cov}\left(\rho(Y_0), \rho(Y_m)\right)\right|.$$

Using (2.5), the Davydov inequality (see [10]) and (2.3), we obtain

$$T \leq C\frac{1}{n}\left(1 + \sum_{m=1}^{n} \alpha_m^q\right) \leq C\frac{1}{n} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{7.13}$$

(ii) **Bound for U.** Using Proposition 4.3, we obtain

$$U \leq C2^{j_0}\frac{1}{n} \leq C\frac{\ln n}{n} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{7.14}$$

(iii) **Bound for W.** For $q \geq 1$ and $p \geq 2$, we have $g_\ell \in \mathbf{B}^s_{p,q}(M) \subseteq \mathbf{B}^s_{2,\infty}(M)$. Hence, by (4.3),

$$W \leq C \sum_{j=j_1+1}^{\infty} 2^{-2js} \leq C 2^{-2j_1 s} \leq C \left(\frac{(\ln n)^3}{n} \right)^{2s} \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+1)}.$$

For $q \geq 1$ and $p \in [1,2)$, we have $g_\ell \in \mathbf{B}^s_{p,q}(M) \subseteq \mathbf{B}^{s+1/2-1/p}_{2,\infty}(M)$. Since $s > 1/p$, we have $s + 1/2 - 1/p > s/(2s+1)$. So, by (4.3),

$$\begin{aligned} W &\leq C \sum_{j=j_1+1}^{\infty} 2^{-2j(s+1/2-1/p)} \leq C 2^{-2j_1(s+1/2-1/p)} \\ &\leq C \left(\frac{(\ln n)^3}{n} \right)^{2(s+1/2-1/p)} \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

Hence, for $q \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 1/p\}$, we have

$$W \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+1)}. \tag{7.15}$$

(iv) **Bound for V.** We will use arguments similar to [20, Proposition 10.3]. Observe that

$$V = V_1 + V_2 + V_3 + V_4, \tag{7.16}$$

where

$$V_1 = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell}| \geq \kappa\lambda_n,\ |b_{j,k,\ell}| < \kappa\lambda_n/2\right\}} \right),$$

$$V_2 = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell}| \geq \kappa\lambda_n,\ |b_{j,k,\ell}| \geq \kappa\lambda_n/2\right\}} \right),$$

$$V_3 = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left(b^2_{j,k,\ell} \mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell}| < \kappa\lambda_n,\ |b_{j,k,\ell}| \geq 2\kappa\lambda_n\right\}} \right)$$

and

$$V_4 = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left(b^2_{j,k,\ell} \mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell}| < \kappa\lambda_n,\ |b_{j,k,\ell}| < 2\kappa\lambda_n\right\}} \right).$$

- **Bounds for V_1 and V_3.** The following inclusions hold:
$\left\{|\widehat{b}_{j,k,\ell}| < \kappa\lambda_n,\ |b_{j,k,\ell}| \geq 2\kappa\lambda_n\right\} \subseteq \left\{|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa\lambda_n/2\right\}$,
$\left\{|\widehat{b}_{j,k,\ell}| \geq \kappa\lambda_n,\ |b_{j,k,\ell}| < \kappa\lambda_n/2\right\} \subseteq \left\{|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa\lambda_n/2\right\}$
and $\left\{|\widehat{b}_{j,k,\ell}| < \kappa\lambda_n,\ |b_{j,k,\ell}| \geq 2\kappa\lambda_n\right\} \subseteq \left\{|b_{j,k,\ell}| \leq 2|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}|\right\}$.
So
$$\max(V_1, V_3) \leq C \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa\lambda_n/2\right\}}\right).$$
Applying the Cauchy-Schwarz inequality and using Propositions 4.4, 4.5 and $2^j \leq n$, we have
$$\begin{aligned}
&\mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2 \mathbf{1}_{\left\{|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa\lambda_n/2\right\}}\right) \\
&\leq \left(\mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^4\right)\right)^{1/2} \left(\mathbb{P}\left(|\widehat{b}_{j,k,\ell} - b_{j,k,\ell}| > \kappa\lambda_n/2\right)\right)^{1/2} \\
&\leq C\left(\frac{2^j}{n}\right)^{1/2} \left(\frac{1}{n^4}\right)^{1/2} \leq C\frac{1}{n^2}.
\end{aligned}$$
Therefore
$$\max(V_1, V_3) \leq C\frac{1}{n^2} \sum_{j=j_0}^{j_1} 2^j \leq C\frac{1}{n^2} 2^{j_1} \leq C\frac{1}{n} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \quad (7.17)$$

- **Bound for V_2.** Using Proposition 4.3, we obtain
$$\mathbb{E}\left((\widehat{b}_{j,k,\ell} - b_{j,k,\ell})^2\right) \leq C\frac{1}{n} \leq C\frac{\ln n}{n}.$$
Hence
$$V_2 \leq C\frac{\ln n}{n} \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbf{1}_{\left\{|b_{j,k,\ell}| > \kappa\lambda_n/2\right\}}.$$
Let j_2 be the integer defined by
$$2^{j_2} = \left[\left(\frac{n}{\ln n}\right)^{1/(2s+1)}\right]. \quad (7.18)$$
We have
$$V_2 \leq V_{2,1} + V_{2,2},$$

where

$$V_{2,1} = C\frac{\ln n}{n}\sum_{j=j_0}^{j_2}\sum_{k=0}^{2^j-1}\mathbf{1}_{\left\{|b_{j,k,\ell}|>\kappa\lambda_n/2\right\}}$$

and

$$V_{2,2} = C\frac{\ln n}{n}\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}\mathbf{1}_{\left\{|b_{j,k,\ell}|>\kappa\lambda_n/2\right\}}.$$

We have

$$V_{2,1} \le C\frac{\ln n}{n}\sum_{j=j_0}^{j_2}2^j \le C\frac{\ln n}{n}2^{j_2} \le C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

For $q \ge 1$ and $p \ge 2$, we have $g_\ell \in \mathbf{B}^s_{p,q}(M) \subseteq \mathbf{B}^s_{2,\infty}(M)$. So, by (4.3),

$$\begin{aligned} V_{2,2} &\le C\frac{\ln n}{n\lambda_n^2}\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}b^2_{j,k,\ell} \le C\sum_{j=j_2+1}^{\infty}\sum_{k=0}^{2^j-1}\beta^2_{j,k} \le C2^{-2j_2 s} \\ &\le C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

For $q \ge 1$, $p \in [1,2)$ and $s > 1/p$, using (4.3), $\mathbf{1}_{\left\{|b_{j,k,\ell}|>\kappa\lambda_n/2\right\}} \le C|b_{j,k,\ell}|^p/\lambda_n^p = C|\beta_{j,k}|^p/\lambda_n^p$, $g_\ell \in \mathbf{B}^s_{p,q}(M)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} V_{2,2} &\le C\frac{\ln n}{n\lambda_n^p}\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^p \le C\left(\frac{\ln n}{n}\right)^{(2-p)/2}\sum_{j=j_2+1}^{\infty}2^{-j(s+1/2-1/p)p} \\ &\le C\left(\frac{\ln n}{n}\right)^{(2-p)/2}2^{-j_2(s+1/2-1/p)p} \le C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

So, for $r \ge 1$, $\{p \ge 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 1/p\}$, we have

$$V_2 \le C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{7.19}$$

- **Bound for V_4.** We have

$$V_4 \le \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1}b^2_{j,k,\ell}\mathbf{1}_{\left\{|b_{j,k,\ell}|<2\kappa\lambda_n\right\}}.$$

Let j_2 be the integer (7.18). Then

$$V_4 \leq V_{4,1} + V_{4,2},$$

where

$$V_{4,1} = \sum_{j=j_0}^{j_2} \sum_{k=0}^{2^j-1} b_{j,k,\ell}^2 \mathbf{1}_{\left\{|b_{j,k,\ell}|<2\kappa\lambda_n\right\}}, \qquad V_{4,2} = \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} b_{j,k,\ell}^2 \mathbf{1}_{\left\{|b_{j,k,\ell}|<2\kappa\lambda_n\right\}}.$$

We have

$$V_{4,1} \leq C \sum_{j=j_0}^{j_2} 2^j \lambda_n^2 = C\frac{\ln n}{n} \sum_{j=j_0}^{j_2} 2^j \leq C\frac{\ln n}{n} 2^{j_2} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

For $q \geq 1$ and $p \geq 2$, we have $g_\ell \in \mathbf{B}_{p,q}^s(M) \subseteq \mathbf{B}_{2,\infty}^s(M)$. Hence, by (4.3),

$$V_{4,2} \leq \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \leq C2^{-2j_2 s} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

For $q \geq 1$, $p \in [1,2)$ and $s > 1/p$, using (4.3), $b_{j,k,\ell}^2 \mathbf{1}_{\left\{|b_{j,k,\ell}|<2\kappa\lambda_n\right\}} \leq C\lambda_n^{2-p}|b_{j,k,\ell}|^p = C\lambda_n^{2-p}|\beta_{j,k}|^p$, $g_\ell \in \mathbf{B}_{p,q}^s(M)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} V_{4,2} &\leq C\lambda_n^{2-p} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p = C\left(\frac{\ln n}{n}\right)^{(2-p)/2} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \\ &\leq C\left(\frac{\ln n}{n}\right)^{(2-p)/2} \sum_{j=j_2+1}^{\infty} 2^{-j(s+1/2-1/p)p} \\ &\leq C\left(\frac{\ln n}{n}\right)^{(2-p)/2} 2^{-j_2(s+1/2-1/p)p} \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

Thus, for $q \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 1/p\}$, we have

$$V_4 \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{7.20}$$

It follows from (7.16), (7.17), (7.19) and (7.20) that

$$V \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}. \tag{7.21}$$

Combining (7.12), (7.13), (7.14), (7.15) and (7.21), we have, for $q \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 1/p\}$,

$$\mathbb{E}\left(\int_0^1 (\widehat{g}_\ell(x) - g_\ell(x))^2 dx\right) \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+1)}.$$

The proof of Theorem 5.1 is complete.

■

Acknowledgement. JF would like to thank Institut Universitaire de France for partially supporting this work. The authors would like to thank the referees for their valuable comments that helped to improve the quality of this paper.

References

[1] Amato, U. and Antoniadis, A. (2001). Adaptive wavelet series estimation in separable nonparametric regression models. *Statistics and Computing*, 11, 373-394.

[2] Amato, U., Antoniadis, A. and De Feis, I. (2002). Fourier series approximation of separable models. *Journal of Computational and Applied Mathematics*, 146, 459-479.

[3] Bradley, R.C. (2007). *Introduction to strong mixing conditions.* Vol. 1,2,3. Kendrick Press.

[4] Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics*, 17, 453-555.

[5] Camlong-Viot, C., Rodrìguez-Pòo, J. M. and Vieu, P. (2006). Nonparametric and semiparametric estimation of additive models with both discrete and continuous variables under dependence. In *The art of semiparametrics*, Contrib. Statist., pages 155-178. Physica-Verlag/Springer, Heidelberg.

[6] Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39.

[7] Chesneau, C. (2013). On the adaptive wavelet estimation of a multidimensional regression function under α-mixing dependence: Beyond the standard assumptions on the noise, *Commentationes Mathematicae Universitatis Carolinae*, 4, 527-556.

[8] Chesneau, C. (2014). A general result on the mean integrated squared error of the wavelet hard thresholding estimator under α-mixing dependence, *Journal of Probability and Statistics*, (to appear).

[9] Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24, 1, 54–81.

[10] Davydov, Y. (1970). The invariance principle for stationary processes. *Theor. Probab. Appl.*, 15, 3, 498-509.

[11] Debbarh, M. (2006). Asymptotic normality for the wavelets estimator of the additive regression components. *Comptes Rendus Mathematique*, 343, 9, 1, 601-606.

[12] Debbarh, M. and Maillot, B. (2008a). Additive regression model for continuous time processes, *Comm. Statist. Theory Methods*, 37,13-15, 2416-2432.

[13] Debbarh, M. and Maillot, B. (2008b). Asymptotic normality of the additive regression components for continuous time processes, *C. R. Math. Acad. Sci. Paris* 346 , 15-16, 901-906.

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

[15] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics*, 24, 508-539.

[16] Doukhan, P. (1994). Mixing. Properties and Examples. Lecture Notes in Statistics 85. *Springer Verlag*, New York.

[17] Fan, J. and Jiang, J. (2005). Nonparametric inferences for additive models. *Journal of the Americal Statistical Association*, 100, 890-907.

[18] Gao, J., Tong, H. and Wolff, R. (2002). Adaptive orthogonal series estimation in additive stochastic regression models. *Statist. Sinica*, 12 (2), 409-428.

[19] Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press.

[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] Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized additive models*. London: Chapman and Hall.

[22] Liebscher, E. (2001). Estimation of the density and the regression function under mixing conditions. *Statist. Decisions*, 19, (1), 9-26.

[23] Linton, O.B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika*, 84, 469-473.

[24] Mallat, S. (2009). A wavelet tour of signal processing. Elsevier/ Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.

[25] Masry, E. (1993). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stoch. Processes Appl.*, 47, 53-74.

[26] Meyer, Y., *Wavelets and Operators*. Cambridge University Press, Cambridge, 1992.

[27] Modha, D. and Masry, E. (1996). Minimum complexity regression estimation with weakly dependent observations. *IEEE Trans. Inform. Theory*, 42, 2133-2145.

[28] Newey, W.K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10 (2), 233-253.

[29] Opsomer, J.D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, 25, 186-211.

[30] Opsomer, J.D. and Ruppert, D. (1998). A fully automated bandwidth selection method for fitting additive models. *Journal of the American Statistical Association*, 93, 605-619.

[31] Patil, P.N. and Truong, Y.K. (2001). Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series. *Annals of the Institute of Statistical Mathematics*, 53, 1, 159-178.

[32] Sardy, S. and Tseng, P. (2004). AMlet, RAMlet, and GAMlet: automatic nonlinear fitting of additive models, robust and generalized, with wavelets. *Journal of Computational and Graphical Statistics*, 13, 283-309.

[33] Sperlich, S., Tjostheim, D. and Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory*, 18, 197-251.

[34] Stone, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, 13, 689-705.

[35] Stone, C.J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.*, 14, 590-606.

[36] Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.*, 22, 118-184.

[37] Tsybakov, A.B., *Introduction à l'estimation non paramétrique*, Springer, 2004.

[38] White, H. and Domowitz, I. (1984). Nonlinear Regression with Dependent Observations. *Econometrica*, 52, 143-162.

[39] Withers, C.S. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57, 477-480.

[40] Zhang, S. and Wong, M.-Y. (2003). Wavelet threshold estimation for additive regression models. *Annals of Statistics*, 31, 152-173.