On the adaptive wavelet estimation of a multidimensional regression function under α-mixing dependence: Beyond the standard assumptions on the noise

Christophe Chesneau

Abstract. We investigate the estimation of a multidimensional regression function f from n observations of an α-mixing process (Y, X), where $Y = f(X) + \xi$, X represents the design and ξ the noise. We concentrate on wavelet methods. In most papers considering this problem, either the proposed wavelet estimator is not adaptive (i.e., it depends on the knowledge of the smoothness of f in its construction) or it is supposed that ξ is bounded or/and has a known distribution. In this paper, we go far beyond this classical framework. Under no boundedness assumption on ξ and no a priori knowledge on its distribution, we construct adaptive term-by-term thresholding wavelet estimators attaining "sharp" rates of convergence under the mean integrated squared error over a wide class of functions f.

Keywords: nonparametric regression; α-mixing dependence; adaptive estimation; wavelet methods; rates of convergence

Classification: 62G08, 62G20

1. Introduction

We consider the nonparametric multidimensional regression model with uniform design described as follows. Let $(Y_t, X_t)_{t\in\mathbb{Z}}$ be a strictly stationary random process defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$, where

$$Y_t = f(X_t) + \xi_t, \tag{1.1}$$

$f : [0,1]^d \to \mathbb{R}$ is the unknown d-dimensional regression function, d is a positive integer, X_1 follows the uniform distribution on $[0,1]^d$ and $(\xi_t)_{t\in\mathbb{Z}}$ is a strictly stationary centered random process independent of $(X_t)_{t\in\mathbb{Z}}$ (the uniform distribution of X_1 will be discussed in Remark 4.6 below). Given n observations $(Y_1, X_1), \ldots, (Y_n, X_n)$ drawn from $(Y_t, X_t)_{t\in\mathbb{Z}}$, we aim to estimate f globally on $[0,1]^d$. Applications of this nonparametric estimation problem can be found in numerous areas as economics, finance and signal processing. See, e.g., [58], [37] and [38].

The performance of an estimator $\hat{f}$ of f can be evaluated by different measures as the Mean Integrated Squared Error (MISE) defined by

$$\mathbf{R}(\hat{f}, f) = \mathbf{E}\left(\int_{[0,1]^d} (\hat{f}(x) - f(x))^2 \, dx\right),$$

where $\mathbf{E}$ denotes the expectation. The smaller $\mathbf{R}(\hat{f}, f)$ is for a large class of f, the better $\hat{f}$ is. Several nonparametric methods for $\hat{f}$ are candidates to achieve this goal. Most of them are presented in [56]. In this paper, we focus our attention on the wavelet methods because of their spatial adaptivity, computational efficiency and asymptotic optimality properties under the MISE. For exhaustive discussions of wavelets and their applications in nonparametric statistics, see, e.g., [1], [57] and [38].

The feature of this study is to consider (1.1) under the following general setting:

- **(i)** $(Y_t, X_t)_{t\in\mathbb{Z}}$ is a dependent process following an α-mixing structure,
- **(ii)** ξ_1 is not necessarily bounded and its distribution is not necessarily known,

(the precise definitions are given in Section 2).

In order to clarify the interest of **(i)** and **(ii)**, let us now present a brief review on the wavelet estimation of f. In the common case where $(Y_1, X_1), \ldots, (Y_n, X_n)$ are *i.i.d.*, various wavelet methods have been developed. The most famous of them can be found in, e.g., [29], [30], [31], [27], [36], [2], [3], [4], [22], [61], [10], [11], [12], [13], [40], [17], [53] and [8]. In view of the structure of the data of many applications, the issue of the relaxation of the independence assumption naturally arises. Among the answers, there are the considerations of various kinds of mixing dependences as the β-mixing dependence (see, e.g., [5]) and the α-mixing dependence mentioned in **(i)**, and several kinds of correlated errors as α-mixing errors (see, e.g., [51], [59], [44], [43] and [42]), long-range dependent errors (see, e.g., [45], [54], [41] and [7]) and martingale difference errors (see, e.g., [62]). Even if some connections exist, these dependent conditions are of different natures.

The interest of **(i)** is justified by its numerous applications in dynamic economic systems and its relative weakness (see, e.g., [58] and [37]). In such an α-mixing context, recent wavelet regression methods and their properties can be found in, e.g., [49], [52], [32], [34], [18], [6] (exploring the nonparametric regression model for censored data), [15] and [16] (both considering the nonparametric regression model for biased data). However, in most of these works, either the proposed wavelet estimator is not adaptive, i.e., its construction depends on the knowledge of the smoothness of f, or it is supposed that ξ_1 (or Y_1) is bounded or has a known distribution. In fact, to the best of our knowledge, [18] is the only work which deals with such an adaptive wavelet regression function estimation problem under **(i)** and **(ii)** (with $d = 1$). However, the construction of the proposed wavelet estimator deeply depends on a parameter θ related to the α-mixing dependence. Since θ is a priori unknown, this estimator can be considered as non adaptive.

The aim of this paper is to provide a theoretical contribution to the full adaptive wavelet estimation of f under **(i)** and **(ii)**. We develop two adaptive wavelet estimators $\hat{f}_\delta$ and $\hat{f}^*_\delta$, both using a term-by-term thresholding rule δ as the hard thresholding rule or the soft thresholding rule (see, e.g., [29], [30] and [31]). We evaluate their performances under the MISE over a wide class of functions f: the Besov balls. In a first part, under mild assumptions on (1.1), we show that the rate of convergence achieved by $\hat{f}_\delta$ is exactly the one of the standard term-by-term wavelet thresholding estimator for f in the classical *i.i.d.* framework. It corresponds to the optimal one in the minimax sense within a logarithmic term. In a second part, with less restrictive assumptions on ξ_1 (only moments of order 2 is required), we show that $\hat{f}^*_\delta$ achieves the same rate of convergence to $\hat{f}_\delta$ up to a logarithmic term. Thus $\hat{f}^*_\delta$ is somewhat less efficient than $\hat{f}_\delta$ in terms of asymptotic MISE but can be used under very mild assumptions on (1.1). To prove our main theorems, we establish a general result on the performance of wavelet term-by-term thresholding estimators which may be of independent interest.

Our contribution can also be viewed as an extension of well-known adaptive wavelet estimation results in the standard *i.i.d.*; for example, Gaussian case to a more general setting allowing weak dependence on the observations and a wide variety of distributions for ξ_1. This complements recent studies investigating other sophisticated dependent contexts as, e.g., [54], [41] and [7] (but with independent $(X_t)_{t\in\mathbb{Z}}$, Gaussian distribution on ξ_1 and $d = 1$).

The rest of this paper is organized as follows. Section 2 clarifies the assumptions on the model and introduces some notations. Section 3 describes the considered wavelet basis on $[0,1]^d$ and the Besov balls. Section 4 is devoted to our adaptive wavelet estimators and their MISE properties over Besov balls. The technical proofs are postponed to Section 5.

2. Assumptions

We make the following assumptions on the model (1.1).

Assumptions on the noise. Let us recall that $(\xi_t)_{t\in\mathbb{Z}}$ is a strictly stationary random process independent of $(X_t)_{t\in\mathbb{Z}}$ such that $\mathbf{E}(\xi_1) = 0$.

H1. We suppose that there exist two constants $\sigma > 0$ and $\omega > 0$ such that, for any $t \in \mathbb{R}$,

$$\mathbf{E}(e^{t\xi_1}) \leq \omega e^{t^2\sigma^2/2}.$$

H2. We suppose that $\mathbf{E}(\xi_1^2) < \infty$.

Remark 2.1. Note that **H1** and **H2** are satisfied for a wide variety of ξ_1, including Gaussian distributions and the bounded distributions. Obviously **H1** implies **H2**.

Remark 2.2. It follows from **H1** that

- for any $p \geq 1$, we have $\mathbf{E}(|\xi_1|^p) < \infty$,

- for any $\lambda > 0$, we have

$$\mathbf{P}(|\xi_1| \geq \lambda) \leq 2\omega e^{-\lambda^2/(2\sigma^2)}. \tag{2.1}$$

α-mixing assumption. For any $m \in \mathbb{Z}$, we define the m-th strongly mixing coefficient of $(Y_t, X_t)_{t\in\mathbb{Z}}$ by

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

where $\mathcal{F}_{-\infty,0}^{(Y,X)}$ is the σ-algebra generated by $\ldots, (Y_{-1}, X_{-1}), (Y_0, X_0)$ and $\mathcal{F}_{m,\infty}^{(Y,X)}$ is the σ-algebra generated by $(Y_m, X_m), (Y_{m+1}, X_{m+1}), \ldots$.

The assumption **H3** below measuring the α-mixing dependence between $(Y_t, X_t)_{t\in\mathbb{Z}}$ will be at the heart of our study.

H3. We suppose that there exist two constants $\gamma > 0$ and $\beta > 0$ such that

$$\sup_{m\geq 1}\left(\alpha_m e^{\beta m}\right) \leq \gamma.$$

Further details on the α-mixing dependence can be found in, e.g., [35], [14] and [9]. Applications and advantages of assuming α-mixing condition on (1.1) can be found in, e.g., [58], [55], [37] and [47].

Remark 2.3. The particular case where $(X_t)_{t\in\mathbb{Z}}$ are independent and $(\xi_t)_{t\in\mathbb{Z}}$ is an α-mixing process with an exponential decay rate is covered by **H3**. Various kinds of correlated errors are permitted including certain short-range dependent errors as strictly stationary AR(1) processes (see, e.g., [35]). However, for instance, the long-range dependence on $(\xi_t)_{t\in\mathbb{Z}}$ as described in [41, Section 1] is not covered.

Boundedness assumptions.

H4. We suppose that there exists a constant $K > 0$ such that

$$\sup_{x\in[0,1]^d} |f(x)| \leq K.$$

H5. For any $m \in \mathbb{Z}$, let $g_{(X_0,X_m)}$ be the density of (X_0, X_m). We suppose that there exists a constant $L > 0$ such that

$$\sup_{m\geq 1}\sup_{(x,x_*)\in[0,1]^{2d}} g_{(X_0,X_m)}(x, x_*) \leq L. \tag{2.2}$$

These boundedness assumptions are standard for (1.1) under α-mixing dependence. See, e.g., [49] and [52].

3. Preliminaries on wavelets

This section contains some facts about the wavelet tensor-product basis on $[0,1]^d$ and the considered function space in terms of wavelet coefficients that will be used in the sequel.

3.1 Wavelet tensor-product basis on $[0,1]^d$. For any $p \geq 1$, set

$$\mathbb{L}_p([0,1]^d) = \left\{ h : [0,1]^d \to \mathbb{R};\ ||h||_p = \left(\int_{[0,1]^d} |h(x)|^p \, dx \right)^{1/p} < \infty \right\}.$$

For the purpose of this paper, we use a compactly supported wavelet-tensor product basis on $[0,1]^d$ based on the Daubechies wavelets. Let N be a positive integer, ϕ be "father" Daubechies-type wavelet and ψ be a "mother" Daubechies-type wavelet of the family $db2N$. In particular, mention that ϕ and ψ have compact supports (see [24] and [48]).

Then, for any $x = (x_1, \ldots, x_d) \in [0,1]^d$, we construct 2^d functions as follows:

- a scale function

$$\Phi(x) = \prod_{u=1}^{d} \phi(x_u)$$

- $2^d - 1$ wavelet functions

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

  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.

We set

$$D_j = \{0, \ldots, 2^j - 1\}^d,$$

for any $j \geq 0$ and $k = (k_1, \ldots, k_d) \in D_j$,

$$\Phi_{j,k}(x) = 2^{jd/2} \Phi(2^j x_1 - k_1, \ldots, 2^j x_d - k_d)$$

and, for any $u \in \{1, \ldots, 2^d - 1\}$,

$$\Psi_{j,k,u}(x) = 2^{jd/2} \Psi_u(2^j x_1 - k_1, \ldots, 2^j x_d - k_d).$$

Then there exists an integer τ such that, for any $j_* \geq \tau$, the collection

$$\mathcal{B} = \{\Phi_{j_*,k}, k \in D_{j_*};\ (\Psi_{j,k,u})_{u \in \{1, \ldots, 2^d-1\}}, \quad j \in \mathbb{N} - \{0, \ldots, j_* - 1\},\ k \in D_j\}$$

(with appropriated treatments at the boundaries) forms an orthonormal basis of $\mathbb{L}_2([0,1]^d)$.

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

$$h(x) = \sum_{k \in D_{j_*}} c_{j_*,k} \Phi_{j_*,k}(x) + \sum_{u=1}^{2^d-1} \sum_{j=j_*}^{\infty} \sum_{k \in D_j} d_{j,k,u} \Psi_{j,k,u}(x),$$

where

$$c_{j,k} = \int_{[0,1]^d} h(x)\Phi_{j,k}(x)\,dx, \qquad d_{j,k,u} = \int_{[0,1]^d} h(x)\Psi_{j,k,u}(x)\,dx. \tag{3.1}$$

The idea behind this wavelet representation is to decompose h into a set of wavelet approximation coefficients, i.e., $\{c_{j_*,k};\ k \in D_{j_*}\}$, and wavelet detail coefficients, i.e., $\{d_{j,k,u};\ j \geq j_*,\ k \in D_j,\ u \in \{1, \dots, 2^d - 1\}\}$. For further results and details about this wavelet basis, we refer the reader to [50], [24], [23] and [48].

3.2 Besov balls. Let $M > 0$, $s \in (0, N)$, $p \geq 1$ and $r \geq 1$. A function $h \in \mathbb{L}_2([0,1]^d)$ belongs to the Besov balls $B^s_{p,r}(M)$ if and only if there exists a constant $M^* > 0$ such that the associated wavelet coefficients (3.1) satisfy

$$\left(\sum_{k \in D_\tau} |c_{\tau,k}|^p \right)^{1/p} + \left(\sum_{j=\tau}^{\infty} \left(2^{j(s+d(1/2-1/p))} \left(\sum_{u=1}^{2^d-1} \sum_{k \in D_j} |d_{j,k,u}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*$$

and with the usual modifications for $p = \infty$ or $r = \infty$.

For a particular choice of parameters s, p and r, these sets contain Sobolev and Hölder balls as well as function classes of significant spatial inhomogeneity (such as the Bump Algebra and Bounded Variations balls). Details about Besov balls can be found in, e.g., [28], [50] and [38].

4. Wavelet estimators and results

4.1 Introduction. We consider the model (1.1) with $f \in \mathbb{L}_2([0,1]^d)$ and we adopt the notations introduced in Sections 2 and 3. The first step to the wavelet estimation of f is its expansion into $\mathcal{B}$ as

$$f(x) = \sum_{k \in D_{j_0}} c_{j_0,k} \Phi_{j_0,k}(x) + \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{\infty} \sum_{k \in D_j} d_{j,k,u} \Psi_{j,k,u}(x), \tag{4.1}$$

where $j_0 \geq \tau$, $c_{j,k} = \int_{[0,1]^d} f(x)\Phi_{j,k}(x)\,dx$ and $d_{j,k,u} = \int_{[0,1]^d} f(x)\Psi_{j,k,u}(x)\,dx$. In the next section, we construct two different adaptive wavelet estimators for f according to the two following lists of assumptions:

- List 1: **H1**, **H3**, **H4** and **H5**,
- List 2: **H2**, **H3** and **H4**,

both used a term-by-term thresholding of suitable wavelet estimators for $c_{j,k}$ and $d_{j,k}$.

4.2 Wavelet estimator I and result. Suppose that **H1**, **H3**, **H4** and **H5** hold. We define the term-by-term thresholding estimator $\hat{f}_\delta$ by

$$\hat{f}_\delta(x) = \sum_{k \in D_{j_0}} \hat{c}_{j_0,k}\Phi_{j_0,k}(x) + \sum_{u=1}^{2^d-1}\sum_{j=j_0}^{j_1}\sum_{k \in D_j} \delta(\hat{d}_{j,k,u}, \kappa\lambda_n)\Psi_{j,k,u}(x), \tag{4.2}$$

where $\hat{c}_{j,k}$ and $\hat{d}_{j,k,u}$ are the empirical wavelet coefficients estimators of $c_{j,k}$ and $d_{j,k,u}$, i.e.,

$$\hat{c}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \Phi_{j,k}(X_i), \qquad \hat{d}_{j,k,u} = \frac{1}{n}\sum_{i=1}^{n} Y_i \Psi_{j,k,u}(X_i), \tag{4.3}$$

$\delta : \mathbb{R} \times (0, \infty) \to \mathbb{R}$ is a term-by-term thresholding rule satisfying that there exists a constant $C > 0$ such that, for any $(u, v, \lambda) \in \mathbb{R}^2 \times (0, \infty)$,

$$|\delta(v, \lambda) - u| \leq C\left(\min(|u|, \lambda) + |v - u|\mathbf{1}_{\{|v-u|>\lambda/2\}}\right). \tag{4.4}$$

Furthermore, κ is a large enough constant,

$$\lambda_n = \sqrt{\frac{\ln n}{n}}, \tag{4.5}$$

j_0 and j_1 are integers satisfying

$$\frac{1}{2}(\ln n)^2 < 2^{j_0 d} \leq (\ln n)^2, \qquad \frac{1}{2}\frac{n}{(\ln n)^4} < 2^{j_1 d} \leq \frac{n}{(\ln n)^4}.$$

Remark 4.1. The estimators $\hat{c}_{j,k}$ and $\hat{d}_{j,k,u}$ (4.3) are unbiased. Indeed the independence of X_1 and ξ_1, and $\mathbf{E}(\xi_1) = 0$ imply that

$$\mathbf{E}(\hat{c}_{j,k}) = \mathbf{E}(Y_1\Phi_{j,k}(X_1)) = \mathbf{E}(f(X_1)\Phi_{j,k}(X_1)) = \int_{[0,1]^d} f(x)\Phi_{j,k}(x)\,dx = c_{j,k}.$$

Similarly we prove that $\mathbf{E}(\hat{d}_{j,k,u}) = d_{j,k,u}$.

Remark 4.2. Among the thresholding rules δ satisfying (4.4), there are

- the hard thresholding rule defined by $\delta(v, \lambda) = v\mathbf{1}_{\{|v|\geq\lambda\}}$, where $\mathbf{1}$ denotes the indicator function,
- the soft thresholding rule defined by $\delta(v, \lambda) = \mathrm{sign}(v)\max(|v| - \lambda, 0)$, where sign denotes the sign function.

The technical details can be found in [27, Lemma 1].

The idea behind the term-by-term thresholding rule δ in $\hat{f}_\delta$ is to only estimate the "large" wavelet coefficients of f (and to remove the others). The reason is that wavelet coefficients having small absolute value are considered to encode mostly noise whereas the important information of f is encoded by the coefficients having large absolute value. This term-by-term selection gives to $\hat{f}_\delta$ an extraordinary local adaptability in handling discontinuities. For further details on such estimators in various statistical framework, we refer the reader to, e.g., [29], [30], [31], [27], [2] and [38]. For the constructions of such estimators under **H3** in a regression context, we refer to [52], [18], [6] and [15].

The considered threshold λ_n (4.5) corresponds to the universal one determined in the standard Gaussian *i.i.d.* case (see [29], [30]).

Remark 4.3. It is important to underline that $\hat{f}_\delta$ is adaptive; its construction does not depend on the smoothness of f.

Theorem 4.1 below explores the performance of $\hat{f}_\delta$ under the MISE over Besov balls.

Theorem 4.1. *Let us consider the model* (1.1) *under* **H1**, **H3**, **H4** *and* **H5**. *Let* $\hat{f}_\delta$ *be* (4.2). *Suppose that* $f \in B^s_{p,r}(M)$ *with* $r \geq 1$, $\{p \geq 2$ *and* $s \in (0, N)\}$ *or* $\{p \in [1, 2)$ *and* $s \in (d/p, N)\}$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbf{R}(\hat{f}_\delta, f) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+d)},$$

for n *large enough.*

The proof of Theorem 4.1 is based on a general result on the performance of the wavelet term-by-term thresholding estimators (see Theorem 5.1 below) and some statistical properties on (4.3) (see Proposition 5.1 below).

The rate of convergence $((\ln n)/n)^{2s/(2s+d)}$ is the near optimal one in the minimax sense for the standard Gaussian *i.i.d.* case (see, e.g., [38] and [56]). "Near" is due to the extra logarithmic term $(\ln n)^{2s/(2s+d)}$. Also, following the terminology of [38], note that this rate of convergence is attained over both the homogeneous zone of the Besov balls corresponding to $p \geq 2$ and the inhomogeneous zone corresponding to $p \in [1, 2)$. This shows that the performance of $\hat{f}_\delta$ is unaffected by the presence of discontinuities in f.

In view of Theorem 4.1, it is natural to address the following question: is it possible to construct an adaptive wavelet estimator reaching the two following objectives:

- relax some assumptions on the model,
- attain a suitable rate of convergence, i.e., as close as possible to the optimal one $n^{-2s/(2s+d)}$.

An answer is provided in the next section.

4.3 Wavelet estimator II and result. Suppose that **H2**, **H3** and **H4** hold (only moments of order 2 are required on ξ_1 and we have no a priori assumption on $g_{(X_0,X_m)}$ as in (2.2)). We define the term-by-term thresholding estimator $\hat{f}^*_\delta$ by

$$\hat{f}^*_\delta(x) = \sum_{k \in D_{j_0}} \hat{c}^*_{j_0,k} \Phi_{j_0,k}(x) + \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{k \in D_j} \delta(\hat{d}^*_{j,k,u}, \kappa\lambda_n) \Psi_{j,k,u}(x), \tag{4.6}$$

where $\hat{c}^*_{j,k}$ and $\hat{d}^*_{j,k,u}$ are the wavelet coefficients estimators of $c_{j,k}$ and $d_{j,k,u}$ defined by

$$\hat{c}^*_{j,k} = \frac{1}{n} \sum_{i=1}^n A_{i,j,k}, \qquad \hat{d}^*_{j,k,u} = \frac{1}{n} \sum_{i=1}^n B_{i,j,k,u}, \tag{4.7}$$

$$A_{i,j,k} = Y_i \Phi_{j,k}(X_i) \mathbf{1}_{\left\{|Y_i \Phi_{j,k}(X_i)| \le \frac{\sqrt{n}}{\ln n}\right\}},$$

$$B_{i,j,k,u} = Y_i \Psi_{j,k,u}(X_i) \mathbf{1}_{\left\{|Y_i \Psi_{j,k,u}(X_i)| \le \frac{\sqrt{n}}{\ln n}\right\}},$$

$\delta : \mathbb{R} \times (0,\infty) \to \mathbb{R}$ is a term-by-term thresholding rule satisfying (4.4), κ is a large enough constant,

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

and j_0 and j_1 are integers such that

$$j_0 = \tau, \qquad \frac{1}{2} \frac{n}{(\ln n)^2} < 2^{j_1 d} \le \frac{n}{(\ln n)^2}.$$

The role of the thresholding selection in (4.7) is to remove the large $|Y_i|$. This allows us to replace **H1** by the less restrictive assumption **H2**. Such an observations thresholding technique has already been used in various contexts of wavelet regression function estimation in [27], [19], [18] and [20].

Remark 4.4. It is important to underline that $\hat{f}^*_\delta$ is adaptive.

Theorem 4.2 below investigates the performance of $\hat{f}^*_\delta$ under the MISE over Besov balls.

Theorem 4.2. *Let us consider the regression model* (1.1) *under* **H2**, **H3** *and* **H4**. *Let $\hat{f}^*_\delta$ be* (4.6). *Suppose that $f \in B^s_{p,r}(M)$ with $r \ge 1$, $\{p \ge 2$ and $s \in (0,N)\}$ or $\{p \in [1,2)$ and $s \in (d/p,N)\}$. Then there exists a constant $C > 0$ such that*

$$\mathbf{R}(\hat{f}^*_\delta, f) \le C \left(\frac{(\ln n)^2}{n} \right)^{2s/(2s+d)},$$

for n large enough.

The proof of Theorem 4.2 is based on a general result on the performance of the wavelet term-by-term thresholding estimators (see Theorem 5.1 below) and some statistical properties on (4.7) (see Proposition 5.2 below).

Theorem 4.2 significantly improves [18, Theorem 1] in terms of rates of convergence and provides an extension to the multidimensional setting.

Remark 4.5. In the case where ξ_1 is bounded, the only interest of Theorem 4.2, and a fortiori $\hat{f}^*_\delta$, is to relax **H5**.

Remark 4.6. Our work can be extended to any compactly supported regression function f and any random design X_1 having a known density g bounded from below over the support of f (including $X_1(\Omega) = \mathbb{R}^d$). In this case, it suffices to adapt the considered wavelet basis to the support of f and to replace Y_i by $Y_i/g(X_i)$ in the definitions of $\hat{f}_\delta$ and $\hat{f}^*_\delta$ to be able to prove Theorems 4.1 and 4.2. Some technical ingredients can be found in [21, Proof of Proposition 2].

When g is unknown, a possible approach following the idea of [52] is to consider $\widehat{fg} = \hat{f}_\delta$ (or $\hat{f}^*_\delta$) to estimate fg, then estimate the unknown density g by a term-by-term wavelet thresholding estimator $\hat{g}$ (as the one in [38]) and finally consider $\hat{f}^\dagger = \widehat{fg}/\hat{g}$. This estimator is particularly useful if we work with (1.1) in an autoregressive framework (see, e.g., [26] and [33]). However, we do not claim it to be near optimal in the minimax sense.

Remark 4.7. Theorems 4.1 and 4.2 are established without necessary knowledge of the distribution of ξ_1. This flexibility seems difficult to reach for other dependent contexts as the long-range dependence on the errors. See, e.g., [45], [54], [41] and [7], where the Gaussian distribution of ξ_1 is supposed and extensively used in the proofs.

Conclusion and discussion. This paper provides some theoretical contributions to the adaptive wavelet estimation of a multidimensional regression function from the α-mixing sequence $(Y_t, X_t)_{t\in\mathbb{Z}}$ defined by (1.1). Two different wavelet term-by-term thresholding estimators $\hat{f}_\delta$ and $\hat{f}^*_\delta$ are constructed. Under very mild assumptions on (1.1) (including unbounded ξ_1 and no a priori knowledge on the distribution of ξ_1), we determine their rates of convergence under the MISE over Besov balls $B^s_{p,r}(M)$. To be more specific, for any $r \geq 1$, $\{p \geq 2$ and $s \in (0, N)\}$ or $\{p \in [1, 2)$ and $s \in (d/p, N)\}$, we prove that

Results	Assumptions	Estimators	Rates of convergence
Theorem 4.1	**H1**, **H3**, **H4**, **H5**	$\hat{f}_\delta$ (4.2)	$((\ln n)/n)^{2s/(2s+d)}$
Theorem 4.2	**H2**, **H3**, **H4**	$\hat{f}^*_\delta$ (4.6)	$((\ln n)^2/n)^{2s/(2s+d)}$

Since $n^{-2s/(2s+d)}$ is the optimal rate of convergence in the minimax sense for the standard *i.i.d.* framework, these results show the good performances of $\hat{f}_\delta$ and $\hat{f}^*_\delta$.

Let us now discuss several aspects of our study.

- Some useful assumptions in Theorem 4.1 are relaxed in Theorem 4.2 and the rate of convergence attained by $\hat{f}_\delta^*$ is close to the one of $\hat{f}_\delta$ (up to the logarithmic term $(\ln n)^{2s/(2s+d)}$).
- Stricto sensu, $\hat{f}_\delta$ is more efficient to $\hat{f}_\delta^*$. Moreover the construction of $\hat{f}_\delta^*$ is more complicated to the one of $\hat{f}_\delta$ due to the presence of the thresholding in (4.7). This could be an obstacle from a practical point of view.

Possible perspectives of this work are to

- determine the optimal lower bound for (1.1) under the α-mixing dependence,
- consider a random design X_1 with unknown or/and unbounded density,
- relax the exponential decay assumption of α_m in **H3**,
- improve the rates of convergence by perhaps using a group thresholding rule (see, e.g., [10], [11]),
- consider another type of dependence on $(X_t)_{t\in\mathbb{Z}}$ and/or $(Y_t)_{t\in\mathbb{Z}}$ as long-range dependence.

All these aspects need further investigations that we leave for a future work.

5. Proofs

In the following, the quantity C denotes a generic constant that does not depend on j, k and n. Its value may change from one term to another.

5.1 A general result. Theorem 5.1 below is derived from [39, Theorem 3.1] and [27, Theorem 1]. The main contributions of this result are to clarify

- the minimal assumptions on the wavelet coefficients estimators,
- the possible choices of the levels j_0 and j_1 (which will be crucial in our dependent framework),

to ensure a "suitable" rate of convergence for the corresponding wavelet term-by-term thresholding estimator. This result may be of independent interest.

Theorem 5.1. *We consider a general nonparametric model where an unknown function $f \in \mathbb{L}_2([0,1]^d)$ needs to be estimated from n observations of a random process defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Using the wavelet series expansion (4.1) of f, we define the term-by-term thresholding estimator $\hat{f}_\delta^\diamond$ by*

$$\hat{f}_\delta^\diamond(x) = \sum_{k\in D_{j_0}} \hat{c}_{j_0,k}^\diamond \Phi_{j_0,k}(x) + \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{k\in D_j} \delta(\hat{d}_{j,k,u}^\diamond, \kappa\lambda_n)\Psi_{j,k,u}(x),$$

where $\hat{c}_{j_0,k}^\diamond$ and $\hat{d}_{j,k,u}^\diamond$ are wavelet coefficients estimators of $c_{j_0,k}$ and $d_{j,k,u}$ respectively, $\delta : \mathbb{R}\times(0,\infty) \to \mathbb{R}$ is a term-by-term thresholding satisfying (4.4), κ is a large enough constant, λ_n is a threshold depending on n, and j_0 and j_1 are *integers such that*

$$\frac{1}{2}2^{\tau d}(\ln n)^{\nu} < 2^{j_0 d} \leq 2^{\tau d}(\ln n)^{\nu}, \qquad \frac{1}{2}\frac{1}{\lambda_n^2(\ln n)^{\varrho}} \leq 2^{j_1 d} \leq \frac{1}{\lambda_n^2(\ln n)^{\varrho}},$$

with $\nu \geq 0$ *and* $\varrho \geq 0$.

We suppose that

- $\hat{c}^{\diamond}_{j,k}$, $\hat{d}^{\diamond}_{j,k,u}$, κ, λ_n, ν *and* ϱ *satisfy the following properties:*
  - **(a)** *there exists a constant* $C > 0$ *such that, for any* $k \in D_j$,
  $$\mathbf{E}((\hat{c}^{\diamond}_{j_0,k} - c_{j_0,k})^2) \leq C\lambda_n^2,$$
  - **(b)** *there exist a constant* $C > 0$ *and* ϖ_n *such that, for any* $j \in \{j_0, \ldots, j_1\}$, $k \in D_j$ *and* $u \in \{1, \ldots, 2^d - 1\}$,
  $$\mathbf{P}\left(|\hat{d}^{\diamond}_{j,k,u} - d_{j,k,u}| \geq \frac{\kappa}{2}\lambda_n\right) \leq C\frac{\lambda_n^8}{\varpi_n},$$
  where ϖ_n *satisfies*
  $$\mathbf{E}((\hat{d}^{\diamond}_{j,k,u} - d_{j,k,u})^4) \leq \varpi_n,$$
  - **(c)** $\lim_{n\to\infty}(\ln n)^{\max(\nu,\varrho)}\lambda_n^{2(1-\upsilon)} = 0$ *for any* $\upsilon \in [0, 1)$,
- $f \in B^s_{p,r}(M)$ *with* $r \geq 1$, $\{p \geq 2$ *and* $s \in (0, N)\}$ *or* $\{p \in [1, 2)$ *and* $s \in (d/p, N)\}$.

Then there exists a constant $C > 0$ *such that*

$$\mathbf{R}(\hat{f}^{\diamond}_{\delta}, f) \leq C\left(\lambda_n^2\right)^{2s/(2s+d)},$$

for n *large enough.*

Proof of Theorem 5.1: The orthonormality of the considered wavelet basis yields

$$\mathbf{R}(\hat{f}^{\diamond}_{\delta}, f) = R_1 + R_2 + R_3, \tag{5.1}$$

where

$$R_1 = \sum_{k\in D_{j_0}} \mathbf{E}\left((\hat{c}^{\diamond}_{j_0,k} - c_{j_0,k})^2\right), \quad R_2 = \sum_{u=1}^{2^d-1}\sum_{j=j_0}^{j_1}\sum_{k\in D_j} \mathbf{E}\left((\delta(\hat{d}^{\diamond}_{j,k,u}, \kappa\lambda_n) - d_{j,k,u})^2\right)$$

and

$$R_3 = \sum_{u=1}^{2^d-1}\sum_{j=j_1+1}^{\infty}\sum_{k\in D_j} d^2_{j,k,u}.$$

Bound for R_1: By **(a)** and **(c)** we have

$$R_1 \leq C 2^{j_0 d} \lambda_n^2 \leq C (\ln n)^{\nu} \lambda_n^2 \leq C \left(\lambda_n^2\right)^{2s/(2s+d)}. \tag{5.2}$$

Bound for R_2: The feature of the term-by-term thresholding δ (i.e., (4.4)) yields

$$R_2 \leq C(R_{2,1} + R_{2,2}), \tag{5.3}$$

where

$$R_{2,1} = \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{k \in D_j} \left(\min(|d_{j,k,u}|, \kappa\lambda_n)\right)^2$$

and

$$R_{2,2} = \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{k \in D_j} \mathbf{E}\left(|\hat{d}^{\diamond}_{j,k,u} - d_{j,k,u}|^2 \mathbf{1}_{\left\{|\hat{d}^{\diamond}_{j,k,u} - d_{j,k,u}| \geq \kappa\lambda_n/2\right\}}\right).$$

Bound for $R_{2,1}$: Let j_2 be an integer satisfying

$$\frac{1}{2}\left(\frac{1}{\lambda_n^2}\right)^{1/(2s+d)} < 2^{j_2} \leq \left(\frac{1}{\lambda_n^2}\right)^{1/(2s+d)}.$$

Note that, by **(c)**, $j_2 \in \{j_0 + 1, \ldots, j_1 - 1\}$.

First of all, let us consider the case $p \geq 2$. Since $f \in B^s_{p,r}(M) \subseteq B^s_{2,\infty}(M)$, we have

$$\begin{aligned}
R_{2,1} &= \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_2} \sum_{k \in D_j} \left(\min(|d_{j,k,u}|, \kappa\lambda_n)\right)^2 + \sum_{u=1}^{2^d-1} \sum_{j=j_2+1}^{j_1} \sum_{k \in D_j} \left(\min(|d_{j,k,u}|, \kappa\lambda_n)\right)^2 \\
&\leq \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_2} \sum_{k \in D_j} \kappa^2 \lambda_n^2 + \sum_{u=1}^{2^d-1} \sum_{j=j_2+1}^{j_1} \sum_{k \in D_j} d^2_{j,k,u} \\
&\leq C \left(\lambda_n^2 \sum_{j=\tau}^{j_2} 2^{jd} + \sum_{j=j_2+1}^{\infty} 2^{-2js}\right) \leq C\left(\lambda_n^2 2^{j_2 d} + 2^{-2j_2 s}\right) \leq C \left(\lambda_n^2\right)^{2s/(2s+d)}.
\end{aligned}$$

Let us now explore the case $p \in [1, 2)$. The facts that $f \in B^s_{p,r}(M)$ with $s > d/p$ and $(2s + d)(2 - p)/2 + (s + d(1/2 - 1/p))p = 2s$ lead to

$$\begin{aligned}
R_{2,1} &= \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_2} \sum_{k \in D_j} (\min(|d_{j,k,u}|, \kappa\lambda_n))^2 \\
&\quad + \sum_{u=1}^{2^d-1} \sum_{j=j_2+1}^{j_1} \sum_{k \in D_j} (\min(|d_{j,k,u}|, \kappa\lambda_n))^{2-p+p} \\
&\le \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_2} \sum_{k \in D_j} \kappa^2 \lambda_n^2 + \sum_{u=1}^{2^d-1} \sum_{j=j_2+1}^{j_1} \sum_{k \in D_j} |d_{j,k,u}|^p (\kappa\lambda_n)^{2-p} \\
&\le C \left(\lambda_n^2 \sum_{j=\tau}^{j_2} 2^{jd} + (\lambda_n^2)^{(2-p)/2} \sum_{j=j_2+1}^{\infty} 2^{-j(s+d(1/2-1/p))p} \right) \\
&\le C \left(\lambda_n^2 2^{j_2 d} + (\lambda_n^2)^{(2-p)/2} 2^{-j_2(s+d(1/2-1/p))p} \right) \le C \left(\lambda_n^2\right)^{2s/(2s+d)}.
\end{aligned}$$

Therefore, for any $r \ge 1$, $\{p \ge 2$ and $s \in (0, N)\}$ or $\{p \in [1, 2)$ and $s \in (d/p, N)\}$, we have

$$R_{2,1} \le C \left(\lambda_n^2\right)^{2s/(2s+d)}. \tag{5.4}$$

Bound for $R_{2,2}$: It follows from the Cauchy-Schwarz inequality, **(b)** and **(c)** that

$$\begin{aligned}
R_{2,2} &\le C \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{k \in D_j} \sqrt{\mathbf{E}\left((\hat{d}^{\diamond}_{j,k,u} - d_{j,k,u})^4\right) \mathbf{P}\left(|\hat{d}^{\diamond}_{j,k,u} - d_{j,k,u}| > \kappa\lambda_n/2\right)} \\
&\le C\lambda_n^4 \sum_{j=\tau}^{j_1} 2^{jd} \le C\lambda_n^4 2^{j_1 d} \le C\lambda_n^4 \frac{1}{\lambda_n^2 (\ln n)^{\varrho}} \le C\lambda_n^2 \le C \left(\lambda_n^2\right)^{2s/(2s+d)}.
\end{aligned} \tag{5.5}$$

Putting (5.3), (5.4) and (5.5) together, for any $r \ge 1$, $\{p \ge 2$ and $s \in (0, N)\}$ or $\{p \in [1, 2)$ and $s \in (d/p, N)\}$, we obtain

$$R_2 \le C \left(\lambda_n^2\right)^{2s/(2s+d)}. \tag{5.6}$$

Bound for R_3: In the case $p \ge 2$, we have $f \in B^s_{p,r}(M) \subseteq B^s_{2,\infty}(M)$. This with **(c)** imply that

$$R_3 \le C \sum_{j=j_1+1}^{\infty} 2^{-2js} \le C 2^{-2j_1 s} \le C \left(\lambda_n^2 (\ln n)^{\varrho}\right)^{2s/d} \le C \left(\lambda_n^2\right)^{2s/(2s+d)}.$$

On the other hand, when $p \in [1,2)$, we have $f \in B^s_{p,r}(M) \subseteq B^{s+d(1/2-1/p)}_{2,\infty}(M)$. Observing that $s > d/p$ leads to $(s+d(1/2-1/p))/d > s/(2s+d)$ and using **(c)**, we have

$$R_3 \le C \sum_{j=j_1+1}^{\infty} 2^{-2j(s+d(1/2-1/p))} \le C 2^{-2j_1(s+d(1/2-1/p))}$$
$$\le C \left(\lambda_n^2 (\ln n)^{\varrho}\right)^{2(s+d(1/2-1/p))/d} \le C \left(\lambda_n^2\right)^{2s/(2s+d)}.$$

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

$$R_3 \le C \left(\lambda_n^2\right)^{2s/(2s+d)}. \tag{5.7}$$

Combining (5.1), (5.2), (5.6) and (5.7), we arrive at, for $r \ge 1$, $\{p \ge 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > d/p\}$,

$$\mathbf{R}(\hat{f}^{\diamond}_{\delta}, f) \le C \left(\lambda_n^2\right)^{2s/(2s+d)}.$$

The proof of Theorem 5.1 is completed. □

5.2 Proof of Theorem 4.1. The proof of Theorem 4.1 is a consequence of Theorem 5.1 above and Proposition 5.1 below. To be more specific, Proposition 5.1 shows that (**a**), (**b**) and (**c**) of Theorem 5.1 are satisfied under the following configuration: $\hat{c}^{\diamond}_{j_0,k} = \hat{c}_{j_0,k}$ and $\hat{d}^{\diamond}_{j,k,u} = \hat{d}_{j,k,u}$ from (4.3), $\lambda_n = \sqrt{(\ln n)/n}$, κ is a large enough constant, $\nu = 2$ and $\varrho = 3$.

Proposition 5.1. *Suppose that* **H1**, **H3**, **H4** *and* **H5** *hold. Let $\hat{c}_{j,k}$ and $\hat{d}_{j,k,u}$ be defined by* (4.3), *and*

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

Then

- **(i)** *there exists a constant $C > 0$ such that, for any j satisfying $(\ln n)^2 \le 2^{jd} \le n$ and $k \in D_j$,*
$$\mathbf{E}((\hat{c}_{j,k} - c_{j,k})^2) \le C\frac{1}{n} \qquad \left(\le C\lambda_n^2\right),$$
- **(ii)** *there exists a constant $C > 0$ such that, for any j satisfying $2^{jd} \le n$, $k \in D_j$ and $u \in \{1, \ldots, 2^d - 1\}$,*
$$\mathbf{E}((\hat{d}_{j,k,u} - d_{j,k,u})^4) \le Cn \qquad (= \varpi_n),$$
- **(iii)** *for $\kappa > 0$ large enough, there exists a constant $C > 0$ such that, for any j satisfying $(\ln n)^2 \le 2^{jd} \le n/(\ln n)^4$, $k \in D_j$ and $u \in \{1, \ldots, 2^d - 1\}$,*
$$\mathbf{P}\left(|\hat{d}_{j,k,u} - d_{j,k,u}| \ge \frac{\kappa}{2}\lambda_n\right) \le C\frac{1}{n^5} \qquad \left(\le C\lambda_n^8/\varpi_n\right).$$

Proof of Proposition 5.1: The technical ingredients in our proof are suitable covariance decompositions, a covariance inequality for α-mixing processes (see Lemma 5.3 in Appendix) and a Bernstein-type exponential inequality for α-mixing processes (see Lemma 5.4 in Appendix).

(i) Since $\mathbf{E}(Y_1\Phi_{j,k}(X_1)) = c_{j,k}$, we have

$$\hat{c}_{j,k} - c_{j,k} = \frac{1}{n}\sum_{i=1}^{n} U_{i,j,k},$$

where

$$U_{i,j,k} = Y_i\Phi_{j,k,u}(X_i) - \mathbf{E}(Y_1\Phi_{j,k}(X_1)).$$

Considering the event $\mathcal{A}_i = \left\{|Y_i| \geq \kappa_*\sqrt{\ln n}\right\}$, where κ_* denotes a constant which will be chosen later, we can split $U_{i,j,k}$ as

$$U_{i,j,k} = V_{i,j,k} + W_{i,j,k},$$

where

$$V_{i,j,k} = Y_i\Phi_{j,k}(X_i)\mathbf{1}_{\mathcal{A}_i} - \mathbf{E}\left(Y_1\Phi_{j,k}(X_1)\mathbf{1}_{\mathcal{A}_i}\right)$$

and

$$W_{i,j,k} = Y_i\Phi_{j,k}(X_i)\mathbf{1}_{\mathcal{A}_i^c} - \mathbf{E}\left(Y_1\Phi_{j,k}(X_1)\mathbf{1}_{\mathcal{A}_i^c}\right).$$

It follows from these decompositions and the inequality $(x+y)^2 \leq 2(x^2+y^2)$, $(x,y)\in\mathbb{R}^2$, that

$$\begin{aligned}
\mathbf{E}((\hat{c}_{j,k} - c_{j,k})^2) &= \frac{1}{n^2}\mathbf{E}\left(\left(\sum_{i=1}^{n} U_{i,j,k}\right)^2\right)\\
&= \frac{1}{n^2}\mathbf{E}\left(\left(\sum_{i=1}^{n} V_{i,j,k} + \sum_{i=1}^{n} W_{i,j,k}\right)^2\right) \qquad (5.8)\\
&\leq \frac{2}{n^2}\left(\mathbf{E}\left(\left(\sum_{i=1}^{n} V_{i,j,k}\right)^2\right) + \mathbf{E}\left(\left(\sum_{i=1}^{n} W_{i,j,k}\right)^2\right)\right) = \frac{2}{n^2}(S+T),
\end{aligned}$$

where

$$S = \mathbf{V}\left(\sum_{i=1}^{n} Y_i\Phi_{j,k}(X_i)\mathbf{1}_{\mathcal{A}_i}\right), \qquad T = \mathbf{V}\left(\sum_{i=1}^{n} Y_i\Phi_{j,k}(X_i)\mathbf{1}_{\mathcal{A}_i^c}\right),$$

and $\mathbf{V}$ denotes the variance.

Bound for S: Let us now introduce a result which will be useful in the rest of study.

Lemma 5.1. *Let $p \geq 1$. Consider (1.1). Suppose that $\mathbf{E}(|\xi_1|^p) < \infty$ and* **H4** *holds. Then*

- *there exists a constant $C > 0$ such that, for any $j \geq \tau$ and $k \in D_j$,*

$$\mathbf{E}\left(|Y_1\Phi_{j,k}(X_1)|^p\right) \leq C2^{jd(p/2-1)};$$

- *there exists a constant $C > 0$ such that, for any $j \geq \tau$, $k \in D_j$ and $u \in \{1, \ldots, 2^d - 1\}$,*

$$\mathbf{E}\left(|Y_1\Psi_{j,k,u}(X_1)|^p\right) \leq C2^{jd(p/2-1)}.$$

Using the inequality $\left(\sum_{i=1}^m a_i\right)^2 \leq m\sum_{i=1}^m a_i^2$, $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$, Lemma 5.1 with $p = 4$ (thanks to **H1** implying $\mathbf{E}(|\xi_1|^p) < \infty$ for $p \geq 1$) and $2^{jd} \leq n$, we arrive at

$$\begin{aligned} S &\leq \mathbf{E}\left(\left(\sum_{i=1}^n Y_i\Phi_{j,k}(X_i)\mathbf{1}_{\mathcal{A}_i}\right)^2\right) \leq n^2\mathbf{E}\left((Y_1\Phi_{j,k}(X_1))^2\mathbf{1}_{\mathcal{A}_1}\right) \\ &\leq n^2\sqrt{\mathbf{E}\left((Y_1\Phi_{j,k}(X_1))^4\right)\mathbf{P}(\mathcal{A}_1)} \leq Cn^2 2^{jd/2}\sqrt{\mathbf{P}(\mathcal{A}_1)} \\ &\leq Cn^{5/2}\sqrt{\mathbf{P}(\mathcal{A}_1)}. \end{aligned}$$

Now, using **H4**, **H1** (implying (2.1)) and taking κ_* large enough, we obtain

$$\begin{aligned} \mathbf{P}(\mathcal{A}_1) &\leq \mathbf{P}(|\xi_1| \geq \kappa_*\sqrt{\ln n} - K) \leq \mathbf{P}\left(|\xi_1| \geq \frac{\kappa_*}{2}\sqrt{\ln n}\right) \\ &\leq 2\omega e^{-\kappa_*^2\ln n/(8\sigma^2)} = 2\omega n^{-\kappa_*^2/(8\sigma^2)} \leq C\frac{1}{n^3}. \end{aligned}$$

Hence

$$S \leq Cn^{5/2}\frac{1}{n^{3/2}} = Cn. \tag{5.9}$$

Bound for T: Observe that

$$T \leq C(T_1 + T_2), \tag{5.10}$$

where

$$\begin{aligned} T_1 &= n\mathbf{V}\left(Y_1\Phi_{j,k}(X_1)\mathbf{1}_{\mathcal{A}_1^c}\right), \\ T_2 &= \left|\sum_{v=2}^n\sum_{\ell=1}^{v-1}\mathbf{C}_{ov}\left(Y_v\Phi_{j,k}(X_v)\mathbf{1}_{\mathcal{A}_v^c}, Y_\ell\Phi_{j,k}(X_\ell)\mathbf{1}_{\mathcal{A}_\ell^c}\right)\right| \end{aligned}$$

and $\mathbf{C}_{ov}$ denotes the covariance.
Bound for T_1: Lemma 5.1 with $p = 2$ yields

$$T_1 \leq n\mathbf{E}\left((Y_1\Phi_{j,k}(X_1))^2\mathbf{1}_{\mathcal{A}_1^c}\right) \leq n\mathbf{E}\left((Y_1\Phi_{j,k}(X_1))^2\right) \leq Cn. \tag{5.11}$$

Bound for T_2: The stationarity of $(Y_t, X_t)_{t\in\mathbb{Z}}$ and $2^{jd} \le n$ imply that

$$\begin{aligned}
T_2 &= \left| \sum_{m=1}^{n} (n-m) \mathbf{C}_{ov} \left(Y_0 \Phi_{j,k}(X_0) \mathbf{1}_{\mathcal{A}_0^c}, Y_m \Phi_{j,k}(X_m) \mathbf{1}_{\mathcal{A}_m^c} \right) \right| \\
&\le n \sum_{m=1}^{n} \left| \mathbf{C}_{ov} \left(Y_0 \Phi_{j,k}(X_0) \mathbf{1}_{\mathcal{A}_0^c}, Y_m \Phi_{j,k}(X_m) \mathbf{1}_{\mathcal{A}_m^c} \right) \right| = n(T_{2,1} + T_{2,2}),
\end{aligned} \tag{5.12}$$

where

$$\begin{aligned}
T_{2,1} &= \sum_{m=1}^{[(\ln n)/\beta]-1} \left| \mathbf{C}_{ov} \left(Y_0 \Phi_{j,k}(X_0) \mathbf{1}_{\mathcal{A}_0^c}, Y_m \Phi_{j,k}(X_m) \mathbf{1}_{\mathcal{A}_m^c} \right) \right|, \\
T_{2,2} &= \sum_{m=[(\ln n)/\beta]}^{n} \left| \mathbf{C}_{ov} \left(Y_0 \Phi_{j,k}(X_0) \mathbf{1}_{\mathcal{A}_0^c}, Y_m \Phi_{j,k}(X_m) \mathbf{1}_{\mathcal{A}_m^c} \right) \right|
\end{aligned}$$

and $[(\ln n)/\beta]$ is the integer part of $(\ln n)/\beta$ (where β is the one in **H3**).

Bound for $T_{2,1}$: First of all, for any $m \in \{1, \ldots, n\}$, let $h_{(Y_0,X_0,Y_m,X_m)}$ be the density of (Y_0, X_0, Y_m, X_m) and $h_{(Y_0,X_0)}$ the density of (Y_0, X_0). We set

$$\begin{aligned}
\theta_m(y, x, y_*, x_*) &= h_{(Y_0,X_0,Y_m,X_m)}(y, x, y_*, x_*) \\
&\quad - h_{(Y_0,X_0)}(y, x) h_{(Y_0,X_0)}(y_*, x_*), \\
&\quad (y, x, y_*, x_*) \in \mathbb{R} \times [0,1]^d \times \mathbb{R} \times [0,1]^d.
\end{aligned} \tag{5.13}$$

For any $(x, x_*) \in [0,1]^{2d}$, since the density of X_0 is 1 over $[0,1]^d$ and using **H5**, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\theta_m(y, x, y_*, x_*)| \, dy \, dy_* \\
&\le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_{(Y_0,X_0,Y_m,X_m)}(y, x, y_*, x_*) \, dy \, dy_* \\
&\quad + \left(\int_{-\infty}^{\infty} h_{(Y_0,X_0)}(y, x) \, dy \right)^2 \\
&= g_{(X_0,X_m)}(x, x_*) + 1 \le L + 1.
\end{aligned} \tag{5.14}$$

By a standard covariance equality, the definition of (5.13), (5.14) and Lemma 5.1 with $p = 1$, we obtain

$$\begin{aligned}
&\left|\mathbf{C}_{ov}\left(Y_0\Phi_{j,k}(X_0)\mathbf{1}_{\mathcal{A}_0^c}, Y_m\Phi_{j,k}(X_m)\mathbf{1}_{\mathcal{A}_m^c}\right)\right| \\
&= \Bigg| \int_{-\kappa_*\sqrt{\ln n}}^{\kappa_*\sqrt{\ln n}} \int_{[0,1]^d} \int_{-\kappa_*\sqrt{\ln n}}^{\kappa_*\sqrt{\ln n}} \int_{[0,1]^d} \theta_m(y,x,y_*,x_*) \\
&\quad \times (y\Phi_{j,k}(x)y_*\Phi_{j,k}(x_*))\, dy\, dx\, dy_* dx_* \Bigg| \\
&\le \int_{[0,1]^d}\int_{[0,1]^d} \left(\int_{-\kappa_*\sqrt{\ln n}}^{\kappa_*\sqrt{\ln n}} \int_{-\kappa_*\sqrt{\ln n}}^{\kappa_*\sqrt{\ln n}} |y||y_*||\theta_m(y,x,y_*,x_*)|\, dy\, dy_* \right) \\
&\quad \times |\Phi_{j,k}(x)||\Phi_{j,k}(x_*)|\, dx\, dx_* \\
&\le \kappa_*^2 \ln n \int_{[0,1]^d}\int_{[0,1]^d} \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\theta_m(y,x,y_*,x_*)|\, dy\, dy_* \right) \\
&\quad \times |\Phi_{j,k}(x)||\Phi_{j,k}(x_*)|\, dx\, dx_* \\
&\le C \ln n \left(\int_{[0,1]^d} |\Phi_{j,k}(x)|\, dx \right)^2 \le C \ln n\ 2^{-jd}.
\end{aligned}$$

Therefore, since $2^{jd} \ge (\ln n)^2$,

$$T_{2,1} \le C(\ln n)^2 2^{-jd} \le C. \tag{5.15}$$

Bound for $T_{2,2}$: By the Davydov inequality (see Lemma 5.3 in Appendix with $p = q = 4$), Lemma 5.1 with $p = 4$, $2^{jd} \le n$ and **H3**, we have

$$\begin{aligned}
\left|\mathbf{C}_{ov}\left(Y_0\Phi_{j,k}(X_0)\mathbf{1}_{\mathcal{A}_0^c}, Y_m\Phi_{j,k}(X_m)\mathbf{1}_{\mathcal{A}_m^c}\right)\right| &\le C\sqrt{\alpha_m}\sqrt{\mathbf{E}\left((Y_0\Phi_{j,k}(X_0))^4 \mathbf{1}_{\mathcal{A}_0^c} \right)} \\
&\le C\sqrt{\alpha_m}\sqrt{\mathbf{E}\left((Y_0\Phi_{j,k}(X_0))^4 \right)} \\
&\le C\sqrt{\alpha_m} 2^{jd/2} \le C e^{-\beta m/2}\sqrt{n}.
\end{aligned}$$

The previous inequality implies that

$$T_{2,2} \le C\sqrt{n} \sum_{m=[(\ln n)/\beta]}^{n} e^{-\beta m/2} \le C\sqrt{n} e^{-(\ln n)/2} \le C. \tag{5.16}$$

Combining (5.12), (5.15) and (5.16), we arrive at

$$T_2 \le Cn(T_{2,1} + T_{2,2}) \le Cn. \tag{5.17}$$

Putting (5.10), (5.11) and (5.17) together, we have

$$T \le T_1 + T_2 \le Cn. \tag{5.18}$$

Finally, (5.8), (5.9) and (5.18) lead to

$$\mathbf{E}((\hat{c}_{j,k} - c_{j,k})^2) \le \frac{2}{n^2}(S+T) \le C\frac{1}{n^2}n \le C\frac{1}{n}\,.$$

This ends the proof of **(i)**.

(ii) Using $\mathbf{E}(Y_1\Psi_{j,k,u}(X_1)) = d_{j,k,u}$ the inequality $(\sum_{i=1}^m a_i)^4 \le m^3\sum_{i=1}^m a_i^4$, $a=(a_1,\ldots,a_m)\in\mathbb{R}^m$, the Hölder inequality, Lemma 5.1 with $p=4$ and $2^{jd}\le n$, we obtain

$$\begin{aligned}\mathbf{E}((\hat{d}_{j,k,u} - d_{j,k,u})^4) &= \frac{1}{n^4}\mathbf{E}\left(\left(\sum_{i=1}^n (Y_i\Psi_{j,k,u}(X_i) - \mathbf{E}(Y_1\Psi_{j,k,u}(X_1)))\right)^4\right)\\ &\le C\frac{1}{n^4}n^4\mathbf{E}\left((Y_1\Psi_{j,k,u}(X_1))^4\right) \le C2^{jd} \le Cn.\end{aligned}$$

The proof of **(ii)** is completed.

Remark 5.1. This bound can be improved using more sophisticated moment inequalities for α-mixing processes (as [60, Theorem 2.2]). However, the obtained bound in **(ii)** is enough for the rest of our study.

(iii) Since $\mathbf{E}(Y_1\Psi_{j,k,u}(X_1)) = d_{j,k,u}$, we have

$$\hat{d}_{j,k,u} - d_{j,k,u} = \frac{1}{n}\sum_{i=1}^n P_{i,j,k,u},$$

where

$$P_{i,j,k,u} = Y_i\Psi_{j,k,u}(X_i) - \mathbf{E}(Y_1\Psi_{j,k,u}(X_1)).$$

Considering again the event $\mathcal{A}_i = \{|Y_i| \ge \kappa_*\sqrt{\ln n}\}$, where κ_* denotes a constant which will be chosen later, we can split $P_{i,j,k,u}$ as

$$P_{i,j,k,u} = Q_{i,j,k,u} + R_{i,j,k,u},$$

where

$$Q_{i,j,k,u} = Y_i\Psi_{j,k,u}(X_i)\mathbf{1}_{\mathcal{A}_i} - \mathbf{E}\left(Y_1\Psi_{j,k,u}(X_1)\mathbf{1}_{\mathcal{A}_i}\right)$$

and

$$R_{i,j,k,u} = Y_i\Psi_{j,k,u}(X_i)\mathbf{1}_{\mathcal{A}_i^c} - \mathbf{E}\left(Y_1\Psi_{j,k,u}(X_1)\mathbf{1}_{\mathcal{A}_i^c}\right).$$

Therefore

$$\mathbf{P}\left(|\hat{d}_{j,k,u} - d_{j,k,u}| \ge \frac{\kappa}{2}\lambda_n\right) \le I_1 + I_2, \tag{5.19}$$

where

$$I_1 = \mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^n Q_{i,j,k,u}\right| \ge \frac{\kappa}{4}\lambda_n\right), \qquad I_2 = \mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^n R_{i,j,k,u}\right| \ge \frac{\kappa}{4}\lambda_n\right).$$

Bound for I_1: The Markov inequality, the Cauchy-Schwarz inequality and Lemma 5.1 with $p = 2$ yield

$$I_1 \leq \frac{4}{\kappa n \lambda_n} \mathbf{E}\left(\left| \sum_{i=1}^{n} Q_{i,j,k,u} \right| \right) \leq C\sqrt{n}\mathbf{E}(|Q_{1,j,k,u}|) \leq C\sqrt{n}\mathbf{E}\left(|Y_1 \Psi_{j,k,u}(X_1)| \mathbf{1}_{\mathcal{A}_1}\right)$$
$$\leq C\sqrt{n}\sqrt{\mathbf{E}\left((Y_1 \Psi_{j,k,u}(X_1))^2 \right) \mathbf{P}(\mathcal{A}_1)} \leq C\sqrt{n}\sqrt{\mathbf{P}(\mathcal{A}_1)}.$$

Now, using **H4**, **H1** (implying (2.1)) and taking κ_* large enough, we obtain

$$\mathbf{P}(\mathcal{A}_1) \leq \mathbf{P}(|\xi_1| \geq \kappa_* \sqrt{\ln n} - K) \leq \mathbf{P}\left(|\xi_1| \geq \frac{\kappa_*}{2}\sqrt{\ln n} \right)$$
$$\leq 2\omega e^{-\kappa_*^2 \ln n/(8\sigma^2)} = 2\omega n^{-\kappa_*^2/(8\sigma^2)} \leq C\frac{1}{n^{11}}.$$

Hence

$$I_1 \leq C\sqrt{n}\frac{1}{n^{11/2}} \leq C\frac{1}{n^5}. \tag{5.20}$$

Bound for I_2: We will bound I_2 via the Bernstein inequality for α-mixing process described in Lemma 5.4 (see Appendix).

We have $\mathbf{E}(R_{1,j,k,u}) = 0$ and, since $|Y_1|\mathbf{1}_{\mathcal{A}_1^c} \leq \kappa_* \sqrt{\ln n}$ and $|\Psi_{j,k,u}(x)| \leq C 2^{jd/2} \leq C\sqrt{n}/(\ln n)^2$,

$$|R_{i,j,k,u}| \leq C\sqrt{\ln n} \sup_{x \in [0,1]^d} |\Psi_{j,k,u}(x)| \leq C\sqrt{\ln n}\frac{\sqrt{n}}{(\ln n)^2} = C\sqrt{\frac{n}{(\ln n)^3}}.$$

Using arguments similar to the proofs of the bounds for T_1 and $T_{2,1}$ in **(i)**, for any $l \leq C \ln n$, since $2^{jd} \geq (\ln n)^2$, we have

$$\mathbf{V}\left(\sum_{i=1}^{l} R_{i,j,k,u} \right) = \mathbf{V}\left(\sum_{i=1}^{l} Y_i \Psi_{j,k,u}(X_i) \mathbf{1}_{\mathcal{A}_i^c} \right) \leq C(l + l^2 \ln n 2^{-jd}) \leq Cl.$$

Hence

$$D_m = \max_{l \in \{1,\dots,2m\}} \mathbf{V}\left(\sum_{i=1}^{l} R_{i,j,k,u} \right) \leq Cm.$$

Lemma 5.4 applied with $R_{1,j,k,u}, \ldots, R_{n,j,k,u}$, $\lambda = \kappa\lambda_n/4$, $\lambda_n = \sqrt{(\ln n)/n}$, $m = [u \ln n]$ with $u > 0$ chosen later, $M = C\sqrt{n/(\ln n)^3}$ and **H3** gives

$$\begin{aligned}
I_2 &\le C\left(\exp\left(-C\frac{\kappa^2\lambda_n^2 n}{D_m/m + \kappa\lambda_n m M}\right) + \frac{M}{\lambda_n} n e^{-\beta m}\right)\\
&\le C\left(\exp\left(-C\frac{\kappa^2 \ln n}{1+\kappa\sqrt{(\ln n)/n}\,u\ln n\sqrt{n/(\ln n)^3}}\right) + \sqrt{\frac{n/(\ln n)^3}{(\ln n)/n}}\, n e^{-\beta u \ln n}\right)\\
&\le C\left(n^{-C\kappa^2/(1+\kappa u)} + n^{2-\beta u}\right).
\end{aligned}$$

Therefore, taking $u = \sqrt{\kappa}$ (for instance) and κ large enough, we have

$$I_2 \le C\frac{1}{n^5}\,. \tag{5.21}$$

It follows from (5.19), (5.20) and (5.21) that

$$\mathbf{P}\left(|\hat{d}_{j,k,u} - d_{j,k,u}| \ge \frac{\kappa}{2}\lambda_n\right) \le I_1 + I_2 \le C\frac{1}{n^5}.$$

This completes the proof of **(iii)**. This ends the proof of Proposition 5.1. □

5.3 Proof of Theorem 4.2. The proof of Theorem 4.2 is a consequence of Theorem 5.1 above and Proposition 5.2 below. To be more specific, Proposition 5.2 shows that (**a**), (**b**) and (**c**) of Theorem 5.1 can be applied under the following configuration: $\hat{c}^{\diamond}_{j_0,k} = \hat{c}^{*}_{j_0,k}$ and $\hat{d}^{\diamond}_{j,k,u} = \hat{d}^{*}_{j,k,u}$ from (4.7), $\lambda_n = \ln n/\sqrt{n}$, κ is a large enough constant, $\nu = 0$ and $\varrho = 0$.

Proposition 5.2. *Suppose that* **H2**, **H3** *and* **H4** *hold. Let* $\hat{c}^{*}_{j,k}$ *and* $\hat{d}^{*}_{j,k,u}$ *be defined by* (4.7), *and*

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

Then

- **(i)** *there exists a constant* $C > 0$ *such that, for any* j *satisfying* $2^{jd} \le n$ *and* $k \in D_j$,

$$\mathbf{E}((\hat{c}^{*}_{j,k} - c_{j,k})^2) \le C\frac{(\ln n)^2}{n} \qquad \left(\le C\lambda_n^2\right),$$

- **(ii)** *there exists a constant* $C > 0$ *such that, for any* j *such that* $2^{jd} \le n$, $k \in D_j$ *and* $u \in \{1, \ldots, 2^d - 1\}$,

$$\mathbf{E}((\hat{d}^{*}_{j,k,u} - d_{j,k,u})^4) \le C \qquad (= \varpi_n),$$

(iii) *for $\kappa > 0$ large enough, there exists a constant $C > 0$ such that, for any j satisfying $2^{jd} \leq n$, $k \in D_j$ and $u \in \{1, \ldots, 2^d - 1\}$,*

$$\mathbf{P}\left(|\hat{d}^*_{j,k,u} - d_{j,k,u}| \geq \frac{\kappa}{2}\lambda_n\right) \leq C\frac{1}{n^4} \qquad \left(\leq C\lambda_n^8/\varpi_n\right).$$

Proof of Proposition 5.2. Again the technical tools in this proof are suitable covariance decompositions, a covariance inequality for α-mixing processes (see Lemma 5.3 in Appendix) and a Bernstein-type exponential inequality for α-mixing processes (see Lemma 5.4 in Appendix).

The following result will be useful in the sequel.

Lemma 5.2. *Let $\hat{c}^*_{j,k}$ and $\hat{d}^*_{j,k,u}$ be defined by (4.7). Suppose that* **H2** *and* **H4** *hold. Then*

- *there exists a constant $C > 0$ such that, for any $j \geq \tau$ and $k \in D_j$,*

$$|\hat{c}^*_{j,k} - c_{j,k}| \leq \frac{1}{n}\left|\sum_{i=1}^{n}(A_{i,j,k} - \mathbf{E}(A_{1,j,k}))\right| + C\frac{\ln n}{\sqrt{n}},$$

- *there exists a constant $C > 0$ such that, for any $j \geq \tau$, $k \in D_j$ and $u \in \{1, \ldots, 2^d - 1\}$,*

$$|\hat{d}^*_{j,k,u} - d_{j,k,u}| \leq \frac{1}{n}\left|\sum_{i=1}^{n}(B_{i,j,k,u} - \mathbf{E}(B_{1,j,k,u}))\right| + C\frac{\ln n}{\sqrt{n}}.$$

(i) Lemma 5.2 and the inequality $(x^2 + y^2) \leq 2(x^2 + y^2)$, $(x, y) \in \mathbb{R}^2$, yield

$$\begin{aligned}\mathbf{E}((\hat{c}^*_{j,k} - c_{j,k})^2) &\leq C\left(\frac{1}{n^2}\mathbf{V}\left(\sum_{i=1}^{n} A_{i,j,k}\right) + \frac{(\ln n)^2}{n}\right)\\ &\leq C\frac{1}{n^2}\left(S + T + n(\ln n)^2\right),\end{aligned} \tag{5.22}$$

where

$$S = n\mathbf{V}\left(A_{1,j,k}\right), \qquad T = \left|\sum_{v=2}^{n}\sum_{\ell=1}^{v-1}\mathbf{C}_{ov}\left(A_{v,j,k}, A_{\ell,j,k}\right)\right|.$$

Bound for S: It follows from Lemma 5.1 with $p = 2$ that

$$S \leq n\mathbf{E}\left((A_{1,j,k})^2\right) \leq n\mathbf{E}\left((Y_1\Phi_{j,k}(X_1))^2\right) \leq Cn. \tag{5.23}$$

Bound for T: The stationarity of $(Y_t, X_t)_{t\in\mathbb{Z}}$ implies that

$$\begin{aligned}T &= \left|\sum_{m=1}^{n}(n - m)\mathbf{C}_{ov}\left(A_{0,j,k}, A_{m,j,k}\right)\right| \leq n\sum_{m=1}^{n}\left|\mathbf{C}_{ov}\left(A_{0,j,k}, A_{m,j,k}\right)\right|\\ &= n(T_1 + T_2),\end{aligned} \tag{5.24}$$

where

$$T_1 = \sum_{m=1}^{[(\ln n)/\beta]-1} \left|\mathbf{C}_{ov}\left(A_{0,j,k}, A_{m,j,k}\right)\right|, \qquad T_2 = \sum_{m=[(\ln n)/\beta]}^{n} \left|\mathbf{C}_{ov}\left(A_{0,j,k}, A_{m,j,k}\right)\right|$$

and $[(\ln n)/\beta]$ is the integer part of $(\ln n)/\beta$ (where β is the one in **H3**).

Bound for T_1: The covariance inequality: $\mathbf{C}_{ov}(U,V) \le \mathbf{E}(U^2)$, where U and V are identically distributed random variables admitting moments of order 2, and Lemma 5.1 with $p=2$ lead to

$$T_1 \le C \sum_{m=1}^{[(\ln n)/\beta]-1} \mathbf{E}\left((A_{0,j,k})^2\right) \le C \ln n \mathbf{E}\left((Y_0\Phi_{j,k}(X_0))^2\right) \le C \ln n. \tag{5.25}$$

Bound for T_2: By the Davydov inequality (see Lemma 5.3 in Appendix with $p=q=4$), the Hölder inequality, $(A_{0,j,k})^4 \le n(Y_0\Phi_{j,k}(X_0))^2$, Lemma 5.1 with $p=2$ and **H3**, we have

$$\begin{aligned} \left|\mathbf{C}_{ov}\left(A_{0,j,k}, A_{m,j,k}\right)\right| &\le C\sqrt{\alpha_m}\sqrt{\mathbf{E}\left((A_{0,j,k})^4\right)} \\ &\le C\sqrt{\alpha_m}\sqrt{n}\sqrt{\mathbf{E}\left((Y_0\Phi_{j,k}(X_0))^2\right)} \le Ce^{-\beta m/2}\sqrt{n}. \end{aligned}$$

Owing to the previous inequality, we arrive at

$$T_2 \le C\sqrt{n} \sum_{m=[(\ln n)/\beta]}^{n} e^{-\beta m/2} \le C\sqrt{n}e^{-(\ln n)/2} = C. \tag{5.26}$$

Combining (5.24), (5.25) and (5.26), we obtain

$$T \le n(T_1+T_2) \le Cn\ln n. \tag{5.27}$$

Finally, putting (5.22), (5.23) and (5.27) together, we have

$$\mathbf{E}((\hat{c}^*_{j,k} - c_{j,k})^2) \le C\frac{1}{n^2}\left(S + T + n(\ln n)^2\right) \le C\frac{(\ln n)^2}{n}.$$

This ends the proof of **(i)**.

(ii) Using $|d_{j,k,u}| \le C$ (since $f \in \mathbb{L}_2([0,1]^d)$) and $|B_{i,j,k,u}| \le \sqrt{n}/\ln n$, we have

$$|\hat{d}^*_{j,k,u} - d_{j,k,u}| \le |\hat{d}^*_{j,k,u}| + |d_{j,k,u}| \le \frac{1}{n}\sum_{i=1}^{n}|B_{i,j,k,u}| + C \le \frac{\sqrt{n}}{\ln n} + C \le C\frac{\sqrt{n}}{\ln n}.$$

Moreover, proceeding as in the proof of **(i)** but with $\Psi_{j,k,u}$ instead of $\Phi_{j,k}$, we obtain

$$\mathbf{E}((\hat{d}^*_{j,k,u} - d_{j,k,u})^2) \le C\frac{(\ln n)^2}{n}.$$

Therefore, we have

$$\mathbf{E}((\hat{d}^*_{j,k,u} - d_{j,k,u})^4) \le C\frac{n}{(\ln n)^2}\mathbf{E}((\hat{d}^*_{j,k,u} - d_{j,k,u})^2) \le C\frac{n}{(\ln n)^2}\frac{(\ln n)^2}{n} \le C.$$

This finishes the proof of **(ii)**.

(iii) For any $j \ge \tau$, $k \in D_j$ and $u \in \{1, \dots, 2^d - 1\}$, set

$$W_{i,j,k,u} = B_{i,j,k,u} - \mathbf{E}(B_{1,j,k,u}).$$

Lemma 5.2 and $\lambda_n = (\ln n)/\sqrt{n}$ imply that, for κ large enough,

$$\begin{aligned}\mathbf{P}\left(|\hat{d}^*_{j,k,u} - d_{j,k,u}| \ge \frac{\kappa}{2}\lambda_n\right) &\le \mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^n W_{i,j,k,u}\right| \ge \frac{\kappa}{2}\lambda_n - C\frac{\ln n}{\sqrt{n}}\right)\\ &\le \mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^n W_{i,j,k,u}\right| \ge \frac{\kappa}{4}\lambda_n\right).\end{aligned} \tag{5.28}$$

We will bound this probability term via the Bernstein inequality for α-mixing process (see Lemma 5.4 in Appendix).

We have $\mathbf{E}(W_{1,j,k,u}) = 0$ and, since $|B_{i,j,k,u}| \le \sqrt{n}/\ln n$, we get

$$|W_{i,j,k,u}| \le 2\frac{\sqrt{n}}{\ln n}.$$

Arguments similar to the proofs of the bounds of S and T in **(i)** with $1 \le l \le C\ln n$ lead to

$$\mathbf{V}\left(\sum_{i=1}^l W_{i,j,k,u}\right) = \mathbf{V}\left(\sum_{i=1}^l B_{i,j,k,u}\right) \le C(l + l^2) \le l^2.$$

Hence

$$D_m = \max_{l\in\{1,\dots,2m\}} \mathbf{V}\left(\sum_{i=1}^l W_{i,j,k,u}\right) \le Cm^2.$$

Lemma 5.4 applied with $W_{1,j,k,u}, \dots, W_{n,j,k,u}$, $\lambda = \kappa\lambda_n/4$, $\lambda_n = (\ln n)/\sqrt{n}$, $m = [u\ln n]$ with $u > 0$ chosen later, $M = C\sqrt{n}/\ln n$ and **H3** gives

$$\begin{aligned}&\mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^n W_{i,j,k,u}\right| \ge \frac{\kappa}{4}\lambda_n\right)\\ &\le C\left(\exp\left(-C\frac{\kappa^2\lambda_n^2 n}{D_m/m + \kappa\lambda_n mM}\right) + \frac{M}{\lambda_n}ne^{-\beta m}\right)\end{aligned}$$

$$\leq C\left(\exp\left(-C\frac{\kappa^2(\ln n)^2}{u\ln n+\kappa((\ln n)/\sqrt{n})u\ln n(\sqrt{n}/\ln n)}\right)+\frac{\sqrt{n}/\ln n}{(\ln n)/\sqrt{n}}ne^{-\beta u\ln n}\right)$$
$$\leq C\left(\exp\left(-C\frac{\kappa^2(\ln n)^2}{u\ln n(1+\kappa)}\right)+n^2e^{-\beta u\ln n}\right)$$
$$\leq C\left(n^{-C\kappa^2/u(1+\kappa)}+n^{2-\beta u}\right).$$

Therefore, taking $u=\sqrt{\kappa}$ (for instance) and κ large enough, we have

$$\mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}W_{i,j,k,u}\right|\geq\frac{\kappa}{4}\lambda_n\right)\leq C\frac{1}{n^4}. \tag{5.29}$$

It follows from (5.28) and (5.29) that

$$\mathbf{P}\left(|\hat{d}_{j,k,u}-d_{j,k,u}|\geq\frac{\kappa}{2}\lambda_n\right)\leq C\frac{1}{n^4}.$$

This completes the proof of **(iii)**. This ends the proof of Proposition 5.2. □

5.4 Proof of the auxiliary results.

Proof of Lemma 5.1: Owing to $\mathbf{E}(|\xi_1|^p)<\infty$, **H4**, the inequality $|x+y|^p\leq 2^{p-1}(|x|^p+|y|^p)$, $(x,y)\in\mathbb{R}^2$, $p\geq 1$, the independence between X_1 and ξ_1 and the change of variables $y=2^jx-k$, we obtain

$$\begin{aligned}\mathbf{E}\left(|Y_1\Phi_{j,k}(X_1)|^p\right)&\leq C\mathbf{E}\left((K^p+|\xi_1|^p)|\Phi_{j,k}(X_1)|^p\right)\\&=C(K^p+\mathbf{E}(|\xi_1|^p))\mathbf{E}\left(|\Phi_{j,k}(X_1)|^p\right)\leq C\int_{[0,1]^d}|\Phi_{j,k}(x)|^p\,dx\\&=C2^{jdp/2}\left(\int_{[0,1]}|\phi(2^jx-k)|^p\,dx\right)^d\leq C2^{jd(p/2-1)}.\end{aligned}$$

The proof of the other point is similar; it is enough to replace $\Phi_{j,k}$ by $\Psi_{j,k,u}$. This ends the proof of Lemma 5.1. □

Proof of Lemma 5.2: Since $\mathbf{E}(Y_1\Phi_{j,k}(X_1))=c_{j,k}$, we have

$$c_{j,k}=\mathbf{E}(A_{1,j,k})+\mathbf{E}\left(Y_1\Phi_{j,k}(X_1)\mathbf{1}_{\left\{|Y_1\Phi_{j,k}(X_1)|>\frac{\sqrt{n}}{\ln n}\right\}}\right).$$

Therefore

$$|c^*_{j,k}-c_{j,k}|\leq\frac{1}{n}\left|\sum_{i=1}^{n}(A_{i,j,k}-\mathbf{E}(A_{1,j,k}))\right|+\mathbf{E}\left(|Y_1\Phi_{j,k}(X_1)|\mathbf{1}_{\left\{|Y_1\Phi_{j,k}(X_1)|>\frac{\sqrt{n}}{\ln n}\right\}}\right).$$

Let us now bound the last term. The Markov inequality and Lemma 5.1 with $p = 2$ yield

$$\mathbf{E}\left(|Y_1\Phi_{j,k}(X_1)|\mathbf{1}_{\left\{|Y_1\Phi_{j,k}(X_1)|>\frac{\sqrt{n}}{\ln n}\right\}}\right) \leq \frac{\ln n}{\sqrt{n}}\mathbf{E}\left((Y_1\Phi_{j,k}(X_1))^2\right) \leq C\frac{\ln n}{\sqrt{n}},$$

that ends the proof of the first point. The proof of the second point is identical; it is enough to replace $\Phi_{j,k}$ by $\Psi_{j,k,u}$. Lemma 5.2 is proved. □

Appendix. In this section we give some preliminary lemmas which have been used in the proofs of our main results.

Lemma 5.3 below presents a covariance inequality for α-mixing processes.

Lemma 5.3 ([25]). *Let $(A_t)_{t\in\mathbb{Z}}$ be a strictly stationary α-mixing process with mixing coefficient α_m, $m \geq 0$, and h and k be two measurable functions. Let $p > 0$ and $q > 0$ satisfying $1/p + 1/q < 1$, such that $\mathbf{E}(|h(A_0)|^p)$ and $\mathbf{E}(|k(A_0)|^q)$ exist. Then there exists a constant $C > 0$ such that*

$$|\mathbf{C}_{ov}(h(A_0), k(A_m))| \leq C\alpha_m^{1-1/p-1/q}\left(\mathbf{E}(|h(A_0)|^p)\right)^{1/p}\left(\mathbf{E}(|k(A_0)|^q)\right)^{1/q}.$$

Lemma 5.4 below describes a concentration inequality for α-mixing processes.

Lemma 5.4 ([46]). *Let $(A_t)_{t\in\mathbb{Z}}$ be a strictly stationary process with the m-th strongly mixing coefficient α_m, $m \geq 0$, n be a positive integer, $h : \mathbb{R} \to \mathbb{C}$ be a measurable function and, for any $t \in \mathbb{Z}$, $U_t = h(A_t)$. We assume that $\mathbf{E}(U_1) = 0$ and there exists a constant $M > 0$ satisfying $|U_1| \leq M$. Then, for any $m \in \{1, \dots, [n/2]\}$ and $\lambda > 0$, we have*

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

where

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

Acknowledgments. We thank the referee for the constructive comments which have lead to significant improvement of the paper.

References

[1] Antoniadis A., *Wavelets in statistics: a review (with discussion)*, J. Italian Statistical Society Series B **6** (1997), 97–144.

[2] Antoniadis A., Grégoire G., Vial P., *Random design wavelet curve smoothing*, Statist. Probab. Lett. **35** (1997), 225–232.

[3] Antoniadis A., Bigot J., Sapatinas T., *Wavelet estimators in nonparametric regression: a comparative simulation study*, J. Statist. Software **6** (2001), no. 6.

[4] Antoniadis A., Leporini D.. Pesquet J.-C., *Wavelet thresholding for some classes of non-Gaussian noise*, Statist. Neerlandica **56** (2002), no. 4, 434–453.

[5] Baraud Y., Comte F., Viennet G., *Adaptive estimation in autoregression or β-mixing regression via model selection*, Ann. Statist. **29** (2001), no. 3, 839–875.
[6] Benatia F., Yahia D., *Nonlinear wavelet regression function estimator for censored dependent data*, J. Afr. Stat. **7** (2012), 391–411.
[7] Beran J., Shumeyko Y., *On asymptotically optimal wavelet estimation of trend functions under long-range dependence*, Bernoulli **18** (2012), no. 1, 137–176.
[8] Bochkina N., Sapatinas T., *Minimax rates of convergence and optimality of Bayes factor wavelet regression estimators under pointwise risks*, Statist. Sinica **19** (2009), 1389–1406.
[9] Bradley R.C., *Introduction to Strong Mixing Conditions*, Vol. 1, 2, 3, Kendrick Press, Heber City, UT, 2007.
[10] Cai T., *Adaptive wavelet estimation: a block thresholding and oracle inequality approach*, Ann. Statist. **27** (1999), 898–924.
[11] Cai, T., *On block thresholding in wavelet regression: adaptivity, block size and threshold level*, Statist. Sinica **12** (2002), 1241–1273.
[12] Cai T., Brown L.D., *Wavelet shrinkage for nonequispaced samples*, Ann. Statist. **26** (1998), 1783–1799.
[13] Cai T., Brown L.D., *Wavelet estimation for samples with random uniform design*, Statist. Probab. Lett. **42** (1999), 313–321.
[14] Carrasco M., Chen X., *Mixing and moment properties of various GARCH and stochastic volatility models*, Econometric Theory **18** (2002), 17–39.
[15] Chaubey Y.P., Shirazi E., *On MISE of a nonlinear wavelet estimator of the regression function based on biased data under strong mixing*, Comm. Statist. Theory Methods, to appear.
[16] Chaubey Y.P., Chesneau C., Shirazi E., *Wavelet-based estimation of regression function for dependent biased data under a given random design*, J. Nonparametr. Stat. **25** (2013), no. 1, 53–71.
[17] Chesneau C., *Regression with random design: a minimax study*, Statist. Probab. Lett. **77** (2007), no. 1, 40–53.
[18] Chesneau C., *Adaptive wavelet regression in random design and general errors with weakly dependent data*, Acta Univ. Apulensis Math. Inform **29** (2012), 65–84.
[19] Chesneau C., Fadili J., *Adaptive wavelet estimation of a function in an indirect regression model*, AStA Adv. Stat. Anal. **96** (2012), no. 1, 25–46.
[20] Chesneau C., Shirazi E., *Nonparametric wavelet regression based on biased data*, Comm. Statist. Theory Methods, to appear.
[21] Chesneau C., Kachour M., Navarro F., *A note on the adaptive estimation of a quadratic functional from dependent observations*, IStatistik: Journal of the Turkish Statistical Association **6** (2013), no. 1, 10–26.
[22] Clyde M.A., George E.I., *Empirical Bayes estimation in wavelet nonparametric regression*, in Bayesian Inference in Wavelet-based Models, Lecture Notes in Statistics, 141, Springer, New York, 1999, pp. 309-322.
[23] Cohen A., Daubechies I., Jawerth B., Vial P., *Wavelets on the interval and fast wavelet transforms*, Appl. Comput. Harmon. Anal. **24** (1993), no. 1, 54–81.
[24] Daubechies I., *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.
[25] Davydov Y., *The invariance principle for stationary processes*, Theor. Probab. Appl. **15** (1970), no. 3, 498–509.
[26] Delouille V., Franke J., von Sachs R., *Nonparametric stochastic regression with design-adapted wavelets*, Sankhya Ser. A **63** (2001), 328–366.
[27] Delyon B., Juditsky A., *On minimax wavelet estimators*, Appl. Comput. Harmon. Anal. **3** (1996), 215–228.
[28] DeVore R., Popov V., *Interpolation of Besov spaces*, Trans. Amer. Math. Soc. **305** (1998), 397–414.

[29] Donoho D.L., Johnstone I.M., *Ideal spatial adaptation by wavelet shrinkage*, Biometrika **81** (1994), 425–455.

[30] Donoho D.L., Johnstone I.M., *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc. **90** (1995), no. 432, 1200–1224.

[31] Donoho D.L., Johnstone I.M., Kerkyacharian G., Picard D., *Wavelet shrinkage: asymptopia? (with discussion)*, J. Royal Statist. Soc. Ser. B **57**, 301–369.

[32] Doosti H., Afshari M., Niroumand H.A., *Wavelets for nonparametric stochastic regression with mixing stochastic process*, Comm. Statist. Theory Methods **37** (2008), no. 3, 373–385.

[33] Doosti H., Islam M.S., Chaubey Y.P., Gora P., *Two dimensional wavelets for nonlinear autoregressive models with an application in dynamical system*, Ital. J. Pure Appl. Math. **27** (2010), 39–62.

[34] Doosti H., Niroumand H.A., *Multivariate stochastic regression estimation by wavelets for stationary time series*, Pakistan J. Statist. **25** (2009), no. 1, 37–46.

[35] Doukhan P., *Mixing. Properties and examples*, Lecture Notes in Statistics, 85, Springer, New York, 1994.

[36] Hall P., Turlach B.A., *Interpolation methods for nonlinear wavelet regression with irregularly spaced design*, Ann. Statist. **25** (1997), 1912–1925.

[37] Härdle W., *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990.

[38] Härdle W., Kerkyacharian G., Picard D., Tsybakov A., *Wavelet, approximation and statistical applications*, Lectures Notes in Statistics, 129, Springer, New York, 1998.

[39] Kerkyacharian G., Picard D., *Thresholding algorithms, maxisets and well-concentrated bases*, Test **9** (2000), no. 2, 283–344.

[40] Kerkyacharian G., Picard D., *Regression in random design and warped wavelets*, Bernoulli **10** (2004), no. 6, 1053–1105.

[41] Kulik R., Raimondo M., *Wavelet regression in random design with heteroscedastic dependent errors*, Ann. Statist. **37** (2009), 3396–3430.

[42] Liang H., *Asymptotic normality of wavelet estimator in heteroscedastic model with α-mixing errors*, J. Syst. Sci. Complex. **24** (2011), no. 4, 725–737.

[43] Li Y.M., Guo J.H., *Asymptotic normality of wavelet estimator for strong mixing errors*, J. Korean Statist. Soc. **38** (2009), no. 4, 383–390.

[44] Li Y.M., Yin C.D., Wei G.D., *On the asymptotic normality for mixing dependent errors of wavelet regression function estimator*, Acta Mathematicae Applicatae Sinica **31** (2008), 1046–1055.

[45] Li L., Xiao Y., *Mean integrated squared error of nonlinear wavelet-based estimators with long memory data*, Ann. Inst. Statist. Math. **59** (2007), 299–324.

[46] Liebscher E., *Estimation of the density and the regression function under mixing conditions*, Statist. Decisions **19** (2001), no. 1, 9–26.

[47] Lütkepohl H., *Multiple Time Series Analysis*, Springer, Heidelberg, 1992.

[48] Mallat S., *A Wavelet Tour of Signal Processing. The Sparse Way*, third edition, with contributions from Gabriel Peyré, Elsevier/Academic Press, Amsterdam, 2009.

[49] Masry E., *Wavelet-based estimation of multivariate regression functions in Besov spaces*, J. Nonparametr. Statist. **12** (2000), no. 2, 283–308.

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

[51] Neumann M.H., von Sachs R., *Wavelet thresholding: beyond the Gaussian i.i.d. situation*, in: Antoniadis A. and Oppenheim G. (eds.), Wavelets and Statistics (Villard de Lans, 1994), Lecture Notes in Statistics, 103, Springer, New York, 1995, pp. 301–330.

[52] Patil P.N., Truong Y.K., *Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series*, Ann. Inst. Statist. Math. **53** (2001), no. 1, 159–178.

[53] Pensky M., Sapatinas T., *Frequentist optimality of Bayes factor estimators in wavelet regression models*, Statist. Sinica **17** (2007), 599–633.

[54] Porto R.F., Morettin P.A., Aubin E.C.Q., *Wavelet regression with correlated errors on a piecewise Hölder class*, Statist. Probab. Lett. **78** (2008), 2739–2743.

[55] Roussas G.G., *Nonparametric regression estimation under mixing conditions*, Stochastic Process. Appl. **36** (1990), no. 1, 107–116.

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

[57] Vidakovic B., *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 1999.

[58] White H., Domowitz, I., *Nonlinear regression with dependent observations*, Econometrica **52** (1984), 143–162.

[59] Xue L.G., *Uniform convergence rates of the wavelet estimator of regression function under mixing error*, Acta Math. Ser. A Chi. Ed. **22** (2002), 528–535.

[60] Yang S.C., *Maximal moment inequality for partial sums of strong mixing sequences and application*, Acta. Math. Sin. (Engl. Ser.) **23** (2007), 1013–1024.

[61] Zhang S., Zheng Z., *Nonlinear wavelet estimation of regression function with random design*, Sci.China Ser. A **42** (1999), no. 8, 825–833.

[62] Zhou X.C., Lin J.G., *A wavelet estimator in a nonparametric regression model with repeated measurements under martingale difference errors structure*, Statist. Probab. Lett. **82** (2012), no. 11, 1914–1922.

LMNO, Université de Caen Basse-Normandie, 14032, France

E-mail: christophe.chesneau@unicaen.fr

(*Received* December 16, 2012*, revised* May 29, 2013)