



**HAL**  
open science

# On the adaptive wavelet estimation of a multidimensional regression function under $\alpha$ -mixing dependence: Beyond the boundedness assumption on the noise

Christophe Chesneau

► **To cite this version:**

Christophe Chesneau. On the adaptive wavelet estimation of a multidimensional regression function under  $\alpha$ -mixing dependence: Beyond the boundedness assumption on the noise. 2013. hal-00727150v3

**HAL Id: hal-00727150**

**<https://hal.science/hal-00727150v3>**

Preprint submitted on 23 Mar 2013

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

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

# On the adaptive wavelet estimation of a multidimensional regression function under $\alpha$ -mixing dependence: Beyond the boundedness assumption on the noise

CHRISTOPHE CHESNEAU

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

e-mail: christophe.chesneau@unicaen.fr

March 23, 2013

## Abstract

We investigate the estimation of a multidimensional regression function  $f$  from  $n$  observations of a  $\alpha$ -mixing process  $(Y, X)$  with  $Y = f(X) + \xi$ , where  $X$  represents the design and  $\xi$  the noise. We focus our attention 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  $\xi$  is bounded (excluding the Gaussian case). In this study we go far beyond this classical framework. Under no boundedness assumption on  $\xi$ , we construct adaptive term-by-term thresholding wavelet estimators enjoying powerful mean integrated squared error (MISE) properties. More precisely, we prove that they achieve "sharp" rates of convergence under the MISE over a wide class of functions  $f$ .

**Key words and phrases:** Nonparametric regression,  $\alpha$ -mixing dependence, Adaptive estimation, Wavelet methods, Rates of convergence.

*AMS 2010 Subject Classifications:* 62G08, 62G20.

## 1 Introduction

The nonparametric multidimensional regression model with uniform design can be 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

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

$f : [0, 1]^d \rightarrow \mathbb{R}$  is the unknown 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 4.4 below). Given  $n$  observations  $(Y_1, X_1), \dots, (Y_n, X_n)$  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., White and Domowitz (1984), Härdle (1990) and Härdle *et al.* (1998).

The performance of an estimator  $\hat{f}$  of  $f$  can be evaluated by 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  $\hat{f}$  can be considered. In this study, 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., Antoniadis (1997), Vidakovic (1999) and Härdle *et al.* (1998). We consider (1.1) under following general setting:

- (i)  $(Y_t, X_t)_{t \in \mathbb{Z}}$  is a dependent process following a  $\alpha$ -mixing structure,
- (ii)  $Y_1$  is not necessarily bounded (allowing a Gaussian distribution for  $\xi_1$ ),

(the precise definitions and more details are given in Section 2).

In order to clarify the interest of considering (i) and (ii), let us now present a brief review on the wavelet estimation of  $f$  under various dependence structures on the observations. In the common case where  $(Y_1, X_1), \dots, (Y_n, X_n)$  are *i.i.d.*, various wavelet methods have been developed. The most famous of them can be found in, e.g., Donoho and Johnstone (1994, 1995), Donoho *et al.* (1995), Delyon and Jusitky (1996), Hall and Turlach (1997), Antoniadis *et al.* (1997, 2001, 2002), Clyde and George (1999), Zhang and Zheng (1999), Cai (1999, 2002), Cai and Brown (1998, 1999), Kerkycharian and Picard (2004), Chesneau (2007), Pensky and Sapatinas (2007) and Bochkina and Sapatinas (2009). Relaxations of the independence assumption have been considered in several studies. If we focus our attention on the  $\alpha$ -mixing dependence, which is reasonably weak and particularly interesting for (1.1) thanks to its applications in dynamic economic systems (see, e.g., White and Domowitz (1984) and Härdle (1990)), recent wavelet methods and results can be found in, e.g., Masry (2000), Patil and Truong (2001), Doosti *et al.* (2008), Doosti and Niroumand (2009) and Chesneau (2012). However, in most of these studies, 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  $\xi_1$  is bounded. In fact, to the best of our knowledge, Chesneau (2012) 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  $\theta$  related to the  $\alpha$ -mixing dependence. Since  $\theta$  is unknown a priori, this estimator can be considered as non adaptive.

In this paper we go far beyond this classical framework by providing theoretical contributions 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  $\delta$  (as the hard thresholding rule or the soft thresholding rule, see, e.g., Donoho and Johnstone (1994, 1995) and Donoho *et al.* (1995)). 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 assumptions on (1.1) (only moments of order 2 is required on  $\xi_1$ ), 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 MISE but can be used under very mild assumptions on (1.1). For proving our main theorems, we establish a general result on the performance of wavelet term-by-term thresholding estimators which may be of independent interest.

The rest of this work 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 in 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 three constants  $\nu > 0$ ,  $\sigma > 0$  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  $\mathbf{E}(\xi_1^2) < \infty$ .

Obviously **H1** allows a Gaussian distribution on  $\xi_1$  and **H1** implies **H2**.

**Remark 2.1** *It follows from H1 that*

- for any  $p \geq 1$ , we have  $\mathbf{E}(|\xi_1|^p) < \infty$ ,
- for any  $\lambda > 0$ , we have

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

### $\alpha$ -mixing assumption

**H3.** 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  $\sigma$ -algebra generated by  $\dots, (Y_{-1}, X_{-1}), (Y_0, X_0)$  and  $\mathcal{F}_{m,\infty}^{(Y,X)}$  is the  $\sigma$ -algebra generated by  $(Y_m, X_m), (Y_{m+1}, X_{m+1}), \dots$

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

$$\alpha_m \leq \gamma e^{-\beta m}.$$

Further details on the  $\alpha$ -mixing dependence can be found in, e.g., Bradley (2007), Doukhan (1994) and Carrasco and Chen (2002). Applications on (1.1) under  $\alpha$ -mixing dependence can be found in, e.g., White and Domowitz (1984), Härdle (1990) and Lütkepohl (1992).

### 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  $m \in \{1, \dots, n\}$ , 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

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

These boundedness assumptions are standard for (1.1) under  $\alpha$ -mixing dependence. See, e.g., Masry (2000) and Patil and Truong (2001).

### 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

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

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,  $\phi$  be "father" Daubechies-type wavelet and  $\psi$  be a "mother" Daubechies-type wavelet of the family  $db2N$ . In particular, mention that  $\phi$  and  $\psi$  have compact supports (see Mallat (2009)).

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

- a scale function

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

- $2^d - 1$  wavelet functions

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

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

We set

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

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

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

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

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

Then there exists an integer  $\tau$  such that the collection

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

(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

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

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 on  $[0, 1]^d$ , we refer the reader to Meyer (1992), Cohen *et al.* (1993) and Mallat (2009).

### 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  $B_{p,r}^s(M)$  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  $q = \infty$ .

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., Devore and Popov (1988), Meyer (1992) and Härdle *et al.* (1998).

## 4 Wavelet estimators and results

### 4.1 Introduction

We consider the model (1.1) and we adopt the notations introduced in Sections 2 and 3. We expand the unknown regression function  $f$  as

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

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 will construct two different 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.**

### 4.2 Wavelet estimator I and result

Suppose that **H1, H3, H4** and **H5** hold. We define the hard thresholding estimator  $\hat{f}_\delta$  by

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

where  $j_0 \geq \tau$  is an integer satisfying

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

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

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

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

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

$\kappa$  is a large enough constant,

$$\lambda_n = \sqrt{\frac{\ln n}{n}} \quad (4.5)$$

and  $j_1$  is an integer satisfying

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

**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  $\xi_1$ , 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  $\delta$  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  $\delta(v, \lambda) = \text{sign}(v) \max(|v| - \lambda, 0)$ , where  $\text{sign}$  denotes the sign function.*

*The technical details can be found in (Delyon and Jusitky, 1996, Lemma 1).*

The idea behind the term-by-term thresholding rule  $\delta$  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., Donoho and Johnstone (1994, 1995), Donoho *et al.* (1995), Delyon and Jusitky (1996), Antoniadis *et al.* (1997) and Härdle *et al.* (1998).

The considered threshold  $\lambda_n$  (4.5) corresponds to the universal one determined in the standard Gaussian *i.i.d.* case (see Donoho and Johnstone (1994, 1995)).

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  $f \in B_{p,r}^s(M)$  with  $r \geq 1$ ,  $\{p \geq 2 \text{ and } s \in (0, N)\}$  or  $\{p \in [1, 2) \text{ and } s \in (d/p, N)\}$ . Then there exists a constant  $C > 0$  such that*

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

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., Härdle *et al.* (1998) and Tsybakov (2004)). “Near” is due to the extra logarithmic term  $(\ln n)^{2s/(2s+d)}$ . Also, following the terminology of Härdle *et al.* (1998), 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  $\xi_1$  and we have no a priori assumption on (2.2)). We define the hard thresholding estimator  $\hat{f}_\delta^*$  by

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

where  $j_0 = \tau$ ,  $\hat{c}_{j, k}^*$  and  $\hat{d}_{j, k, u}^*$  are the wavelet coefficients estimators of  $c_{j, k}$  and  $d_{j, k, u}$  defined by

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

where

$$A_{i, j, k} = Y_i \Phi_{j, k}(X_i) \mathbf{1}_{\{|Y_i \Phi_{j, k}(X_i)| \leq \frac{\sqrt{n}}{\ln n}\}}, \quad B_{i, j, k, u} = Y_i \Psi_{j, k, u}(X_i) \mathbf{1}_{\{|Y_i \Psi_{j, k, u}(X_i)| \leq \frac{\sqrt{n}}{\ln n}\}},$$

$\kappa$  is a large enough constant,

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

$\delta : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$  is a term-by-term thresholding rule satisfying (4.4) and  $j_1$  is an integer such that

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

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 used in various contexts of wavelet regression function estimation in Delyon and Jusitky (1996), Chesneau and Fadili (2012), Chesneau (2012) and Chesneau and Shirazi (2013).

Again, let us mention 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  $f \in B_{p,r}^s(M)$  with  $r \geq 1$ ,  $\{p \geq 2 \text{ and } s \in (0, N)\}$  or  $\{p \in [1, 2) \text{ and } s \in (d/p, N)\}$ . Then there exists a constant  $C > 0$  such that*

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

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 4.1 improves (Chesneau, 2012, Theorem 1) in terms of rates of convergence and provides an extension to the multidimensional setting.

**Remark 4.3** *Obviously the assumptions **H1** and **H2** includes the bounded noise. If  $\xi$  is bounded, the only interest of Theorem 4.2, and a fortiori  $\hat{f}_\delta^*$ , is to relax **H5**.*

**Remark 4.4** *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  $\hat{f}_\delta^*$  (more precisely, in (4.3) and (4.7)) to be able to prove Theorems 4.1 and 4.2. Some technical ingredients can be found in (Chesneau et al., 2013, Proof of Proposition 2).*

*When  $g$  is unknown, a possible approach following the idea of Patil and Truong (2001) 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 Härdle et al. (1998)) and finally consider  $\hat{f}^\dagger = \widehat{fg}/\hat{g}$ . This estimator is particularly appropriated if we work with (1.1) in an autoregressive framework (see, e.g., Delouille et al. (2001) and Doosti et al. (2010)). However, we do not claim it to be near optimal in the minimax sense.*

## Conclusion and discussion

This paper provides some theoretical contributions to the adaptive wavelet estimation of a multidimensional regression function from the  $\alpha$ -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  $Y_1$  (and  $\xi_1$ ) and no a priori knowledge on the distribution of  $\xi_1$ ), we determine their rates of convergence under the MISE over Besov balls  $B_{p,r}^s(M)$ . 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

Result	Assumptions	Estimator	Rate of convergence
Theorem 4.1	<b>H1, H3, H4, H5</b>	$\hat{f}_\delta$ (4.2)	$(\ln n/n)^{2s/(2s+d)}$
Theorem 4.2	<b>H2, H3, H4</b>	$\hat{f}_\delta^*$ (4.6)	$((\ln n)^2/n)^{2s/(2s+d)}$

Since  $n^{-2s/(2s+d)}$  is the optimal rate of convergence (in the minimax sense) in 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  $\alpha$ -mixing dependence,
- consider a random design  $X_1$  with unknown density or/and unbounded,
- relax the exponential decay assumption of  $\alpha_m$  in **H3**,
- improve the rates of convergence by perhaps using a group thresholding rule (see, e.g., Cai (1999, 2002)).

All these aspects needs 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 (Kerkyacharian and Picard, 2000, Theorem 3.1) and (Delyon and Jusitky, 1996, Theorem 1). The main contributions of this result is to provide more flexibility on the choices of  $j_0$  and  $j_1$  (which will be crucial in our dependent framework) and to clarified the minimal assumptions on the wavelet coefficients estimators to ensure a “suitable” rate of convergence for the corresponding wavelet term-by-term thresholding estimator (see **(a)** and **(b)** in Theorem 5.1). This result may be of independent interest.

**Theorem 5.1** *Let us 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 its wavelet series expansion (4.1), we define the hard 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  $j_0 \geq \tau$  is an integer such that

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

with  $\nu \geq 0$ ,  $\hat{c}_{j_0, k}^\diamond$  and  $\hat{d}_{j, k, u}^\diamond$  are wavelet coefficients estimators of  $c_{j_0, k}$  and  $d_{j, k, u}$  respectively,  $\kappa$  is a large enough constant,  $\lambda_n$  is a threshold,  $\delta : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$  is a term-by-term thresholding satisfying (4.4) and  $j_1$  is an integer such that

$$\frac{1}{2} \frac{1}{\lambda_n^2 (\ln n)^\varrho} \leq 2^{j_1 d} \leq \frac{1}{\lambda_n^2 (\ln n)^\varrho},$$

with  $\varrho \geq 0$ .

It is understood that

$$\lim_{n \rightarrow \infty} (\ln n)^{\max(\nu, \varrho)} \lambda_n^{2(1-\nu)} = 0$$

for any  $\nu \in (0, 1)$ .

We suppose that  $\nu$ ,  $\hat{c}_{j, k}^\diamond$ ,  $\hat{d}_{j, k, u}^\diamond$ ,  $\kappa$ ,  $\lambda_n$  and  $\varrho$  satisfy the following inequalities:

**(a)** there exists a constant  $C > 0$  such that, for any  $k \in D_j$ ,

$$\mathbf{E}((\hat{c}_{j_0, k}^\diamond - c_{j_0, k})^2) \leq C \lambda_n^2.$$

**(b)** there exists a constant  $C > 0$  such that, for any  $j \in \{j_0, \dots, j_1\}$ ,  $k \in D_j$  and  $u \in \{1, \dots, 2^d - 1\}$ ,

$$\mathbf{P} \left( |\hat{d}_{j, k, u}^\diamond - d_{j, k, u}| \geq \frac{\kappa}{2} \lambda_n \right) \leq C \frac{\lambda_n^8}{\varpi_n},$$

where  $\varpi_n$  is such that

$$\mathbf{E}((\hat{d}_{j,k,u}^\circ - d_{j,k,u})^4) \leq \varpi_n.$$

Furthermore we suppose that  $f \in B_{p,r}^s(M)$  with  $r \geq 1$ ,  $\{p \geq 2 \text{ and } s \in (0, N)\}$  or  $\{p \in [1, 2) \text{ and } s \in (d/p, N)\}$ . Then there exists a constant  $C > 0$  such that

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

**Proof of Theorem 5.1.** The orthonormality of the considered wavelet basis yields

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

where

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

and

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

*Bound for  $R_1$ :* By (a) we have

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

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

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

where

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

and

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

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

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

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

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

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

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

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

$$R_{2,1} \leq C (\lambda_n^2)^{2s/(2s+d)}. \quad (5.4)$$

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

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

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

$$R_2 \leq C (\lambda_n^2)^{2s/(2s+d)}. \quad (5.6)$$

Bound for  $R_3$ : In the case  $p \geq 2$ , we have  $f \in B_{p,r}^s(M) \subseteq B_{2,\infty}^s(M)$  which implies that

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

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

$$\begin{aligned} R_3 &\leq C \sum_{j=j_1+1}^{\infty} 2^{-2j(s+d(1/2-1/p))} \leq C 2^{-2j_1(s+d(1/2-1/p))} \\ &\leq C (\lambda_n^2 (\ln n)^{\varrho})^{2(s+d(1/2-1/p))/d} \leq C (\lambda_n^2)^{2s/(2s+d)}. \end{aligned}$$

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

$$R_3 \leq C (\lambda_n^2)^{2s/(2s+d)}. \quad (5.7)$$

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

$$\mathbf{R}(f_{\delta}^{\circ}, f) \leq C (\lambda_n^2)^{2s/(2s+d)}.$$

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) and (b) of Theorem 5.1 are satisfied under the following configuration:

- $\nu = 2$ ,
- $\hat{c}_{j_0,k}^{\circ} = \hat{c}_{j_0,k}$  and  $\hat{d}_{j,k,u}^{\circ} = \hat{d}_{j,k,u}$  from (4.3),
- $\lambda_n = \sqrt{\ln n/n}$ ,
- $\kappa$  is a large enough constant,
- $\varrho = 3$ .

**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  $(1/2)(\ln n)^2 \leq 2^{jd} \leq n$  and  $k \in D_j$ ,*

$$\mathbf{E}((\hat{c}_{j,k} - c_{j,k})^2) \leq C \frac{1}{n} \quad (\leq C\lambda_n^2).$$

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

$$\mathbf{E}((\hat{d}_{j,k,u} - d_{j,k,u})^4) \leq Cn \quad (= \varpi_n).$$

(iii) *for  $\kappa > 0$  large enough, there exists a constant  $C > 0$  such that, for any  $j$  satisfying  $(1/2)(\ln n)^2 \leq 2^{jd} \leq n/(\ln n)^4$ ,  $k \in D_j$  and  $u \in \{1, \dots, 2^d - 1\}$ ,*

$$\mathbf{P}\left(|\hat{d}_{j,k,u} - d_{j,k,u}| \geq \frac{\kappa}{2}\lambda_n\right) \leq C \frac{1}{n^5} \quad (\leq C\lambda_n^8/\varpi_n).$$

**Proof of Proposition 5.1.** The technical ingredients in our proof are suitable covariance decompositions, a covariance inequality for  $\alpha$ -mixing processes (see Lemma 5.3 in Appendix) and a Bernstein-type exponential inequality for  $\alpha$ -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  $\mathcal{A}_i = \{|Y_i| \geq \kappa_* \sqrt{\ln n}\}$ , where  $\kappa_*$  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

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

and

$$W_{i,j,k} = Y_i \Phi_{j,k}(X_i) \mathbf{1}_{\mathcal{A}_i^c} - \mathbf{E}(Y_1 \Phi_{j,k}(X_1) \mathbf{1}_{\mathcal{A}_1^c}).$$

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

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

where

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

*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  $k \in D_j$ ,*

$$\mathbf{E}(|Y_1 \Phi_{j,k}(X_1)|^p) \leq C 2^{jd(p/2-1)}.$$

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

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

Using the inequality  $(\sum_{i=1}^m a_i)^2 \leq m \sum_{i=1}^m a_i^2$ ,  $a = (a_1, \dots, a_m) \in \mathbb{R}^m$ , 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

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

Now, using **H4**, **H1** (implying (2.1)) and taking  $\kappa_*$  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

$$S \leq C n^{5/2} \frac{1}{n^{3/2}} = Cn. \quad (5.9)$$

*Bound for T:* Observe that

$$T \leq T_1 + T_2, \quad (5.10)$$

where

$$T_1 = n \mathbf{V} (Y_1 \Phi_{j,k}(X_1) \mathbf{1}_{\mathcal{A}_1^c})$$

and

$$T_2 = \left| \sum_{v=2}^n \sum_{\ell=1}^{v-1} \mathbf{C}_{ov} (Y_v \Phi_{j,k}(X_v) \mathbf{1}_{\mathcal{A}_v^c}, Y_\ell \Phi_{j,k}(X_\ell) \mathbf{1}_{\mathcal{A}_\ell^c}) \right|.$$

*Bound for T<sub>1</sub>:* Lemma 5.1 with  $p = 2$  yields

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

*Bound for T<sub>2</sub>:* The stationarity of  $(Y_t, X_t)_{t \in \mathbb{Z}}$  and  $2^{jd} \leq n$  imply that

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

where

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

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

*Bound for  $T_{2,1}$ :* First of all, for any  $m \in \{1, \dots, n\}$ , 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

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

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

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

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

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

Therefore, since  $2^{jd} \geq (1/2)(\ln n)^2$ ,

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

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

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

The previous inequality implies that

$$T_{2,2} \leq C\sqrt{n} \sum_{m=\lceil \ln n/\beta \rceil}^n e^{-\beta m/2} \leq C\sqrt{ne^{-\ln n/2}} \leq C. \quad (5.16)$$

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

$$T_2 \leq n(T_{2,1} + T_{2,2}) \leq Cn. \quad (5.17)$$

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

$$T \leq T_1 + T_2 \leq Cn. \quad (5.18)$$

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

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

This ends the proof of (i).

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

$$\begin{aligned} \mathbf{E}((\hat{d}_{j,k,u} - d_{j,k,u})^4) &= \frac{1}{n^4} \mathbf{E} \left( \left( \sum_{i=1}^n (Y_i \Psi_{j,k,u}(X_i)) - \mathbf{E}(Y_1 \Psi_{j,k,u}(X_1)) \right)^4 \right) \\ &\leq C \frac{1}{n^4} n^4 \mathbf{E}((Y_1 \Psi_{j,k,u}(X_1))^4) \leq C 2^{jd} \leq Cn. \end{aligned}$$

The proof of (ii) is completed.

**Remark 5.1** *This bound can be improved using more sophisticated moment inequalities for  $\alpha$ -mixing processes (as (Yang, 2007, 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  $\mathcal{A}_i = \{|Y_i| \geq \kappa_* \sqrt{\ln n}\}$ , where  $\kappa_*$  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

$$Q_{i,j,k,u} = Y_i \Psi_{j,k,u}(X_i) \mathbf{1}_{\mathcal{A}_i} - \mathbf{E}(Y_1 \Psi_{j,k,u}(X_1) \mathbf{1}_{\mathcal{A}_1})$$

and

$$R_{i,j,k,u} = Y_i \Psi_{j,k,u}(X_i) \mathbf{1}_{\mathcal{A}_i^c} - \mathbf{E}(Y_1 \Psi_{j,k,u}(X_1) \mathbf{1}_{\mathcal{A}_1^c}).$$

Therefore

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

where

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

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

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

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

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

Hence

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

*Bound for  $I_2$ :* We will bound  $I_2$  via the Bernstein inequality for  $\alpha$ -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 the  $\alpha$ -mixing random variables  $R_{1,j,k,u}, \dots, R_{n,j,k,u}$ ,  $\lambda = \kappa \lambda_n/4$ ,  $\lambda_n = \sqrt{\ln n/n}$ ,  $m = \lfloor u \ln n \rfloor$  with  $u > 0$  chosen later,  $M = C \sqrt{n}/(\ln n)^3$  and **H3** gives

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

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

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

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

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

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) and (b) of Theorem 5.1 can be applied under the following configuration:

- $\nu = 0$ ,
- $\hat{c}_{j_0,k}^\diamond = \hat{c}_{j_0,k}^*$  and  $\hat{d}_{j,k,u}^\diamond = \hat{d}_{j,k,u}^*$  from (4.7),
- $\lambda_n = \ln n / \sqrt{n}$ ,
- $\kappa$  is a large enough constant,
- $\varrho = 0$ .

**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  $2^{jd} \leq n$  and  $k \in D_j$ ,*

$$\mathbf{E}((\hat{c}_{j,k}^* - c_{j,k})^2) \leq C \frac{(\ln n)^2}{n} \quad (\leq C\lambda_n^2).$$

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

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

(iii) *for  $\kappa > 0$  large enough, there exists a constant  $C > 0$  such that, for any  $j$  satisfying  $2^{jd} \leq n/(\ln n)^4$ ,  $k \in D_j$  and  $u \in \{1, \dots, 2^d - 1\}$ ,*

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

**Proof of Proposition 5.2.** Again the technical tools in our proof are suitable covariance decompositions, a covariance inequality for  $\alpha$ -mixing processes (see Lemma 5.3 in Appendix) and a Bernstein-type exponential inequality for  $\alpha$ -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  $u \in \{1, \dots, 2^d - 1\}$ ,

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

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

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

where

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

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

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

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

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

where

$$T_1 = \sum_{m=1}^{\lfloor \ln n / \beta \rfloor - 1} |\mathbf{C}_{ov}(A_{0,j,k}, A_{m,j,k})|, \quad T_2 = \sum_{m=\lfloor \ln n / \beta \rfloor}^n |\mathbf{C}_{ov}(A_{0,j,k}, A_{m,j,k})|$$

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

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

$$T_1 \leq C \sum_{m=1}^{\lfloor \ln n / \beta \rfloor - 1} \mathbf{E}((A_{0,j,k})^2) \leq C \ln n \mathbf{E}((Y_0 \Phi_{j,k}(X_0))^2) \leq C \ln n. \quad (5.25)$$

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

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

Owing to the previous inequality, we arrive at

$$T_2 \leq C \sqrt{n} \sum_{m=\lfloor \ln n / \beta \rfloor}^n e^{-\beta m / 2} \leq C \sqrt{n} e^{-\ln n / 2} = C. \quad (5.26)$$

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

$$T \leq n(T_1 + T_2) \leq C n \ln n. \quad (5.27)$$

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

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

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

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

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

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

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

Therefore

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

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

(iii) For any  $j \geq \tau$ ,  $k \in D_j$  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  $\kappa$  large enough,

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

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

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

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

Similar arguments to the proofs of the bounds of  $S$  and  $T$  in (i) with  $1 \leq l \leq C \ln n$  lead to

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

Hence

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

Lemma 5.4 applied with the  $\alpha$ -mixing random variables  $W_{1,j,k,u}, \dots, W_{n,j,k,u}$ ,  $\lambda = \kappa \lambda_n / 4$ ,  $\lambda_n = \ln n / \sqrt{n}$ ,  $m = \lceil u \ln n \rceil$  with  $u > 0$  chosen later,  $M = C \sqrt{n} / \ln n$  and **H3** gives

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

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

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

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  $\xi_1$  and the change of variables  $y = 2^j x - k$ , we obtain

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

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

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

Therefore

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

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

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

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 (Davydov (1970))** *Let  $(A_t)_{t \in \mathbb{Z}}$  be a stationary  $\alpha$ -mixing process with mixing coefficient  $\alpha_m$ ,  $m \geq 0$ , 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*

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

**Lemma 5.4 (Liebscher (2001))** *Let  $(A_t)_{t \in \mathbb{Z}}$  be a stationary process with the  $m$ -th strongly mixing coefficient  $\alpha_m$ ,  $m \geq 0$ ,  $n$  be a positive integer,  $h : \mathbb{R} \rightarrow \mathbb{C}$  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*

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

where

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

## References

- Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion), *Journal of the Italian Statistical Society Series B*, 6, 97-144.
- Antoniadis, A., Grégoire, G. and Vial, P. (1997). Random design wavelet curve smoothing, *Statist. Probab. Lett.*, 35, 225-232.

- Antoniadis, A., Bigot, J. and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study, *Journal of Statistical Software*, 6, i06.
- Antoniadis, A., Leporini, D. and Pesquet, J.-C. (2002). Wavelet thresholding for some classes of non-Gaussian noise, *Statistica Neerlandica*, 56, 4, 434-453.
- Bochkina, N. and Sapatinas, T. (2009). Minimax rates of convergence and optimality of Bayes factor wavelet regression estimators under pointwise risks, *Statistica Sinica*, Vol. 19, 1389-1406.
- Bradley, R.C. (2007). *Introduction to strong mixing conditions*, Vol. 1,2,3. Kendrick Press.
- Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, 27, 898-924.
- Cai, T. (2002). On block thresholding in wavelet regression: adaptivity, block size and threshold level, *Statistica Sinica*, 12, 1241-1273.
- Cai, T. and Brown, L.D. (1998). Wavelet shrinkage for nonequispaced samples, *The Annals of Statistics*, 26, 1783-1799.
- Cai, T. and Brown, L.D. (1999). Wavelet estimation for samples with random uniform design, *Statistics and Probability Letters*, 42, 313-321.
- Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory*, 18, 17-39.
- Chesneau, C. (2007). Regression with random design: a minimax study, *Statistics and Probability Letters*, 77, 1, 40-53.
- Chesneau, C. (2012). Adaptive wavelet regression in random design and general errors with weakly dependent data, *Acta universitatis Apulensis*, 29, 65-84.
- Chesneau, C. and Fadili, J. (2012). Adaptive wavelet estimation of a function in an indirect regression model. *Advances in Statistical Analysis*, 96, 1, 25-46.
- Chesneau, C. and Shirazi, E. (2013). Nonparametric Wavelet Regression Based on Biased Data, *Communications in Statistics: Theory and Methods*, (to appear).
- Chesneau, C., Kachour, M. and Navarro, F. (2013). A note on the adaptive estimation of a quadratic functional from dependent observations, *IStatistik: Journal of the Turkish Statistical Association*, 6, 1, 10 - 26.
- Clyde, M.A. and George, E.I. (1999). *Empirical Bayes estimation in wavelet nonparametric regression*. In Bayesian inference in wavelet-based models, volume 141 of Lecture Notes in Statist., pages 309-322. Springer, New York.

- 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.
- Davydov, Y. (1970). The invariance principle for stationary processes, *Theor. Probab. Appl.*, 15, 3, 498-509.
- Delouille, V., Franke, J. and von Sachs, R. (2001). Nonparametric stochastic regression with design-adapted wavelets, *Sankhya, ser. A*, 63, 328-366.
- Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis*. **3**, 215-228.
- DeVore, R. and Popov, V. (1988). Interpolation of Besov spaces, *Trans. Amer. Math. Soc.*, 305, 397-414.
- Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, 81, 425-455.
- Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, 90, 432, 1200-1224.
- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia ? (with discussion), *Journal of the Royal Statistical Society Serie B*, 57, 301-369.
- Doosti, H., Afshari, M. and Niroumand, H.A. (2008). Wavelets for nonparametric stochastic regression with mixing stochastic process, *Communication in Statistics: Theory and Methods*, 37, **3**, 373-385.
- Doosti, H., Islam, M.S., Chaubey, Y.P. and Gora, P. (2010). Two dimensional wavelets for nonlinear autoregressive models with an application in dynamical system, *Italian Journal of Pure and Applied Mathematics*, 27, 39-62.
- Doosti, H. and Niroumand, H.A. (2009). Multivariate Stochastic Regression Estimation by Wavelets for Stationary Time Series, *Pakistan journal of Statistics*, 27, **1**, 37-46.
- Doukhan, P. (1994). *Mixing. Properties and Examples*, Lecture Notes in Statistics 85. Springer Verlag, New York.
- Hall, P. and Turlach, B.A. (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design, *The Annals of Statistics*, 25, 1912-1925.
- Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press.

- Härdle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics, New York 129, Springer Verlag.
- Kerkycharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2), 283-344.
- Kerkycharian, G. and Picard, D. (2004). Regression in random design and warped wavelets, *Bernoulli*, 10, 6, 1053-1105.
- Liebscher, E. (2001). Estimation of the density and the regression function under mixing conditions. *Statist. Decisions*, 19, (1), 9-26.
- Lütkepohl, H. (1992). *Multiple Time Series Analysis*, Springer-Verlag, Heidelberg.
- Mallat, S. (2009). *A wavelet tour of signal processing*, Elsevier/ Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.
- Masry, E. (2000). Wavelet-Based estimation of multivariate regression functions in besov spaces, *Journal of Nonparametric Statistics*, 12, 2, 283-308.
- Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University Press, Cambridge.
- 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.
- Pensky, M. and Sapatinas, T. (2007). Frequentist optimality of Bayes factor estimators in wavelet regression models, *Statistica Sinica*, 17, 599-633.
- Tsybakov, A.B. (2004). *Introduction à l'estimation non-paramétrique*, Springer.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.
- White, H. and Domowitz, I. (1984). Nonlinear Regression with Dependent Observations. *Econometrica*, 52, 143-162.
- Yang S.C. (2007). Maximal moment inequality for partial sums of strong mixing sequences and application. *Acta. Math. Sinica*, English Series, 23, 1013-1024.
- Zhang, S. and Zheng, Z. (1999). Nonlinear wavelet estimation of regression function with random design, *Science in China Series A: Mathematics*, 42, 8, 825-833.