



**HAL**  
open science

# A general result on the performance of the wavelet hard thresholding estimator under $\alpha$ -mixing dependence

Christophe Chesneau

► **To cite this version:**

Christophe Chesneau. A general result on the performance of the wavelet hard thresholding estimator under  $\alpha$ -mixing dependence. 2013. hal-00809862v1

**HAL Id: hal-00809862**

**<https://hal.science/hal-00809862v1>**

Preprint submitted on 10 Apr 2013 (v1), last revised 5 Jan 2014 (v3)

**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.

# A general result on the performance of the wavelet hard thresholding estimator under $\alpha$ -mixing dependence

Christophe Chesneau

*Laboratoire de Mathématiques Nicolas Oresme, Université de Caen BP 5186, F 14032 Caen Cedex, FRANCE. e-mail: chesneau@math.unicaen.fr*

**Abstract:** In this note we consider the estimation of an unknown function  $f$  for weakly dependent data ( $\alpha$ -mixing) in a general setting. Our contribution is theoretical: we prove that a wavelet hard thresholding estimator attains a sharp rate of convergence under the mean integrated squared error (MISE) over Besov balls without imposing too restrictive assumptions on the model. Applications are given for two types of inverse problems: the deconvolution density estimation and the density estimation in a GARCH-type model, both improve existing results in this dependent context.

**AMS 2000 subject classifications:** Primary 62G07, 62G20; secondary 60K35.

**Keywords and phrases:** Nonparametric statistics,  $\alpha$ -mixing dependence, Wavelets, Besov balls, Hard thresholding.

## 1. Introduction

A general nonparametric problem is adopted: we aim to estimate an unknown function  $f$  via  $n$  random variables  $V_1, \dots, V_n$  from a strictly stationary stochastic process  $(V_t)_{t \in \mathbb{Z}}$ . We suppose that  $(V_t)_{t \in \mathbb{Z}}$  has a weak dependence structure; the  $\alpha$ -mixing case is considered. This kind of dependence naturally appears in numerous models as Markov chains, GARCH-type models and discretely observed diffusions. See, e.g., Doukhan (1994), Carrasco and Chen (2002) and Bradley (2007). The problems where  $f$  is the density of  $V_1$  or a regression function have received a lot of attention. A partial list of related works includes Robinson (1983), Roussas (1987, 1990), Truong and Stone (1992), Tran (1993), Masry (1996a,b), Masry and Fan (1997), Bosq (1998) and Liebscher (2001).

For an efficient estimation of  $f$ , many methods can be considered. The most popular of them are based on kernels, splines and wavelets. In this note we deal with the wavelet methods. They has been introduced in a i.i.d. setting by Donoho and Johnstone (1994, 1995) and Donoho et al. (1995, 1996). Among their interests, they enjoy remarkable local adaptivity against discontinuities and spatially varying degree of oscillations. Complete reviews and discussions on wavelets in statistics can be found in, e.g., Antoniadis (1997) and Härdle et al. (1998). In the context of  $\alpha$ -mixing dependence, various wavelet methods have been elaborated for a wide variety of nonparametric problems. Recent developments can be found in, e.g., Leblanc (1996), Tribouley and Viennet (1998), Masry (2000), Patil and

Truong (2001), Doosti et al. (2008), Doosti and Niroumand (2009), Doosti et al. (2010), Cai and Liang (2011), Niu and Liang (2011), Chesneau (2012, 2013a,b), Chaubey and Shirazi (2013) and Abbaszadeh and Emadi (2013).

In the general dependent setting described above, we provide a theoretical contribution to the performance of a wavelet estimator based on a hard thresholding. This nonlinear wavelet procedure has the features to be fully adaptive and efficient over a large class of functions  $f$ . See, e.g., Donoho and Johnstone (1994, 1995), Donoho et al. (1995, 1996) and Delyon and Juditsky (1996). Following the spirit of Kerkycharian and Picard (2000), we determine necessary assumptions on  $(V_t)_{t \in \mathbb{Z}}$  and the wavelet basis to ensure that the considered estimator attains a fast rate of convergence under the MISE over Besov balls. The obtained rate of convergence often corresponds to the near optimal one in the minimax sense for the standard i.i.d. case. The originality of our result is to be general and sharp; it can be applied for nonparametric models of different natures and improves some existing results. This fact is illustrated by the consideration of the density deconvolution estimation problem and the density estimation problem in a GARCH-type model, improving (Chesneau, 2012, Proposition 5.1) and (Chesneau, 2013a, Theorem 2) respectively.

The organization of this note is as follows. In the next section we describe the considered wavelet setting. The wavelet hard thresholding estimator and its theoretical performances are presented in Section 3. Applications of our general result are given in Section 4. The proofs are carried out in Section 5.

## 2. Wavelets and Besov balls

In this section we introduce some notations corresponding to wavelets and Besov balls.

### 2.1. Wavelet basis

We consider the wavelet basis on  $[0, 1]$  constructs from the Daubechies wavelets  $db2N$  with  $N \geq 1$  (see, e.g., Daubechies (1992)). A brief description of this basis is given below. Let  $\phi$  and  $\psi$  be the initial wavelet functions of the family  $db2N$ . These functions have the particularity to be compactly supported and to belong to the class  $\mathcal{C}^a$  for  $N > 10(a + 1)$ . For any  $j \geq 0$ , we set  $\Lambda_j = \{0, \dots, 2^j - 1\}$  and, for  $k \in \Lambda_j$ ,

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

With appropriated treatments at the boundaries, there exists an integer  $\tau$  such that, for any integer  $\ell \geq \tau$ ,  $\mathcal{B} = \{\phi_{\ell,k}, k \in \Lambda_\ell; \psi_{j,k}; j \in \mathbb{N} - \{0, \dots, \ell - 1\}, k \in \Lambda_j\}$  is an orthonormal basis of  $\mathbb{L}^2([0, 1]) = \left\{ f : [0, 1] \rightarrow \mathbb{R}; \|f\|_2 = \left( \int_0^1 |f(x)|^2 dx \right)^{1/2} < \infty \right\}$ . For any integer  $\ell \geq \tau$  and  $f \in \mathbb{L}^2([0, 1])$ , we have the following wavelet expansion

sion:

$$f(x) = \sum_{k \in \Lambda_\ell} c_{\ell,k} \phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k \in \Lambda_j} d_{j,k} \psi_{j,k}(x), \quad x \in [0, 1],$$

where  $c_{j,k}$  and  $d_{j,k}$  denotes the wavelet coefficients of  $f$  defined by

$$c_{j,k} = \int_0^1 f(x) \phi_{j,k}(x) dx, \quad d_{j,k} = \int_0^1 f(x) \psi_{j,k}(x) dx. \quad (2.1)$$

Technical details can be found in, e.g., Cohen et al. (1993) and Mallat (2009).

In the main result of this study, we will investigate the performances of the proposed estimator by assuming that the unknown function of interest  $f$  belongs to a wide class of functions: the Besov class. Its definition in terms of wavelet coefficients is presented below.

## 2.2. Besov balls

We say that  $f \in B_{p,r}^s(M)$  with  $s > 0$ ,  $p, r \geq 1$  and  $M > 0$  if and only if there exists a constant  $C > 0$  such that the wavelet coefficients of  $f$  given by (2.1) satisfy

$$2^{\tau(1/2-1/p)} \left( \sum_{k \in \Lambda_\tau} |c_{\tau,k}|^p \right)^{1/p} + \left( \sum_{j=\tau}^{\infty} \left( 2^{j(s+1/2-1/p)} \left( \sum_{k \in \Lambda_j} |d_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq C,$$

with the usual modifications if  $p = \infty$  or  $r = \infty$ . Note that, for particular choices of  $s$ ,  $p$  and  $r$ ,  $B_{p,r}^s(M)$  contains the classical Hölder and Sobolev balls. See, e.g., Meyer (1990) and Härdle et al. (1998).

**Remark 2.1.** *We have chosen a wavelet basis on  $[0,1]$  to fix the notations; wavelet basis on another interval can be considered in the rest of the study.*

## 3. Statistical framework, estimator and result

### 3.1. Statistical framework

As mentioned in Section 1, a nonparametric estimation setting as general as possible is adopted: we aim to estimate an unknown function  $f \in \mathbb{L}^2([0, 1])$  via  $n$  random variables (or vectors)  $V_1, \dots, V_n$  from a strictly stationary stochastic process  $(V_t)_{t \in \mathbb{Z}}$  defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . We suppose that  $(V_t)_{t \in \mathbb{Z}}$  has a  $\alpha$ -mixing dependence structure with exponential decay rate, i.e., there exist two constants  $\gamma > 0$  and  $\theta > 0$  such that

$$\sup_{m \geq 1} (e^{\theta m} \alpha_m) \leq \gamma,$$

where  $\alpha_m = \sup_{(A,B) \in \mathcal{F}_{-\infty,0}^V \times \mathcal{F}_{m,\infty}^V} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ ,  $\mathcal{F}_{-\infty,0}^V$  is the  $\sigma$ -algebra generated by the random variables (or vectors)  $\dots, V_{-1}, V_0$  and  $\mathcal{F}_{m,\infty}^V$  is the  $\sigma$ -algebra generated by the random variables (or vectors)  $V_m, V_{m+1}, \dots$

The  $\alpha$ -mixing dependence is reasonably weak; it is satisfied by a wide variety of models including Markov chains, GARCH-type models and discretely observed diffusions. See, for instance, Bradley (2007), Carrasco and Chen (2002), Doukhan (1994) and Genon-Catalot et al. (2000).

The considered estimator for  $f$  is presented below.

### 3.2. Estimator

We define the hard thresholding wavelet estimator  $\hat{f}$  by

$$\hat{f}(x) = \sum_{k \in \Lambda_{j_0}} \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \hat{d}_{j,k} \mathbf{1}_{\{|\hat{d}_{j,k}| \geq \kappa \lambda_j\}} \psi_{j,k}(x), \quad (3.1)$$

where

$$\hat{c}_{j,k} = \frac{1}{n} \sum_{i=1}^n q(\phi_{j,k}, V_i), \quad \hat{d}_{j,k} = \frac{1}{n} \sum_{i=1}^n q(\psi_{j,k}, V_i), \quad (3.2)$$

$\mathbf{1}$  is the indicator function,  $\kappa > 0$  is a large enough constant,  $j_0$  is the integer satisfying

$$2^{j_0} = \lceil \tau (\ln n)^\nu \rceil, \quad (3.3)$$

where  $[a]$  denotes the integer part of  $a$ ,  $j_1$  is the integer satisfying

$$2^{j_1} = \left\lceil \left[ \left( \frac{n}{(\ln n)^3} \right)^{1/(2\rho+1)} \right] \right\rceil \quad (3.4)$$

and

$$\lambda_j = 2^{\rho j} \sqrt{\frac{\ln n}{n}}. \quad (3.5)$$

Here it is supposed that there exists a function  $q : \mathbb{L}^2([0, 1]) \times V_1(\Omega) \rightarrow \mathbb{R}$  such that

**(H1)** for  $\gamma \in \{\phi, \psi\}$ , any integer  $j \geq j_0$  and  $k \in \Lambda_j$ ,

$$\mathbb{E}(q(\gamma_{j,k}, V_1)) = \int_0^1 f(x) \gamma_{j,k}(x) dx,$$

where  $\mathbb{E}$  denotes the expectation,

**(H2)** there exist three constants,  $C > 0$ ,  $\rho \geq 0$  and  $\nu \geq 0$ , satisfying, for  $\gamma \in \{\phi, \psi\}$ , for any integer  $j \geq j_0$  and  $k \in \Lambda_j$ ,

- (i)  $\sup_{x \in V_1(\Omega)} |q(\gamma_{j,k}, x)| \leq C 2^{\rho j} 2^{j/2}$ ,
- (ii)  $\mathbb{E} \left( (q(\gamma_{j,k}, V_1))^2 \right) \leq C 2^{2\rho j}$ ,
- (iii) for any  $m \geq 1$ ,

$$\mathbb{C}_{ov} (q(\gamma_{j,k}, V_{m+1}), q(\gamma_{j,k}, V_1)) \leq C 2^{2\rho j} (\ln n)^{\nu-1} 2^{-j},$$

where  $\mathbb{C}_{ov}$  denotes the covariance.

For well-known nonparametric models in the i.i.d. setting, wavelet hard thresholding estimators and important results can be found in, e.g., Donoho and Johnstone (1994, 1995), Donoho et al. (1995, 1996), Delyon and Juditsky (1996), Kerkycharian and Picard (2000) and Fan and Koo (2002).

The assumption **(H1)** ensures that (3.2) are unbiased estimators for  $c_{j,k}$  and  $d_{j,k}$  given by (2.1), whereas **(H2)** is related to their good performance. See Proposition 5.1 below. These assumptions are not too restrictive. For instance, if we consider the standard density estimation problem where  $(V_t)_{t \in \mathbb{Z}}$  are i.i.d. random variables with bounded density  $f$ , the function  $q(\gamma, x) = \gamma(x)$  satisfies **(H1)** and **(H2)** with  $\rho = 0$  and  $\nu = 0$  (note that, thanks to the independence of  $(V_t)_{t \in \mathbb{Z}}$ , the covariance term in **(H2)**-**(iii)** is zero). The technical details are given in Donoho et al. (1996).

In the  $\alpha$ -mixing context,  $\hat{f}$  defined by (3.1) is a general and improved version of the estimator considered in Chesneau (2012, 2013a). The main differences are the presence of the tuning parameters  $\nu$  and  $\rho$ , and the global definition of the function  $q$  offering numerous possibilities of applications. Two of them are explored in Section 4.

### 3.3. Result

Theorem 3.1 below determines the rate of convergence attained by  $\hat{f}$  under the MISE over Besov balls.

**Theorem 3.1.** *We consider the general statistical setting described in Subsection 3.1. Let  $\hat{f}$  be (3.1) under **(H1)** and **(H2)**. Suppose that  $f \in B_{p,r}^s(M)$  with  $r \geq 1$ ,  $\{p \geq 2 \text{ and } s \in (0, N)\}$  or  $\{p \in [1, 2) \text{ and } s \in ((2\rho + 1)/p, N)\}$ . Then there exists a constant  $C > 0$  such that*

$$\mathbb{E} \left( \|\hat{f} - f\|_2^2 \right) \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}.$$

The rate of convergence “ $((\ln n)/n)^{2s/(2s+2\rho+1)}$ ” is often the near optimal one in the minimax sense for numerous statistical problems in a i.i.d. setting. See, e.g., Härdle et al. (1998) and Tsybakov (2004). Moreover, note that Theorem 3.1 is flexible; the assumptions on  $(V_t)_{t \in \mathbb{Z}}$ , related to the definition of  $q$  in **(H1)** and **(H2)**, are mild. In the next section, this flexibility is illustrated for two sophisticated nonparametric estimation problems: the density deconvolution estimation problem and the density estimation problem in a GARCH-type model.

## 4. Applications

### 4.1. Density deconvolution

Let  $(V_t)_{t \in \mathbb{Z}}$  be a strictly stationary stochastic process such that

$$V_t = X_t + \epsilon_t, \quad t \in \mathbb{Z}, \quad (4.1)$$

where  $(X_t)_{t \in \mathbb{Z}}$  is a strictly stationary stochastic process with unknown density  $f$  and  $(\epsilon_t)_{t \in \mathbb{Z}}$  is a strictly stationary stochastic process with known density  $g$ . It is supposed that  $\epsilon_t$  and  $X_t$  are independent for any  $t \in \mathbb{Z}$  and  $(V_t)_{t \in \mathbb{Z}}$  is a  $\alpha$ -mixing process with exponential decay rate (see Subsection 3.1 for a precise definition). Our aim is to estimate  $f$  via  $V_1, \dots, V_n$  from  $(V_t)_{t \in \mathbb{Z}}$ . Some related works are Masry (1993), Kulik (2008), Comte et al. (2008) and van Zanten and Zareba (2008).

We formulate the following assumptions:

- (G1) The support of  $f$  is  $[0, 1]$ .  
 (G2) There exists a constant  $C > 0$  such that

$$\sup_{x \in \mathbb{R}} f(x) \leq C < \infty.$$

- (G3) Let  $u$  be the density of  $V_1$ . We suppose that there exists a constant  $C > 0$  such that

$$\sup_{x \in \mathbb{R}} u(x) \leq C.$$

- (G4) For any  $m \in \mathbb{Z}$ , let  $u_{(V_1, V_{m+1})}$  be the density of  $(V_1, V_{m+1})$ . We suppose that there exists a constant  $C > 0$  such that

$$\sup_{m \in \mathbb{Z}} \sup_{(x, y) \in \mathbb{R}^2} u_{(V_1, V_{m+1})}(x, y) \leq C.$$

- (G5) For any integrable function  $\gamma$ , we define its Fourier transform by

$$\mathcal{F}(\gamma)(x) = \int_{-\infty}^{\infty} \gamma(y) e^{-ixy} dy, \quad x \in \mathbb{R}.$$

We suppose that there exist three known constants  $C > 0$ ,  $c > 0$  and  $\delta > 1$  such that, for any  $x \in \mathbb{R}$ ,

- the Fourier transform of  $g$  satisfies

$$|\mathcal{F}(g)(x)| \geq \frac{c}{(1+x^2)^{\delta/2}},$$

- for any  $\ell \in \{0, 1, 2\}$ , the  $\ell$ -th derivative of the Fourier transform of  $g$  satisfies

$$|(\mathcal{F}(g)(x))^{(\ell)}| \leq \frac{C}{(1+|x|)^{\delta+\ell}}.$$

We are now in the position to present the result.

**Theorem 4.1.** *We consider the model (4.1). Suppose that **(G1)**-**(G5)** are satisfied. Let  $\hat{f}$  be defined as in (3.1) with the function  $q : \mathbb{L}^2([0, 1]) \times V_1(\Omega) \rightarrow \mathbb{R}$  defined by*

$$q(\gamma, x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\gamma)}(y)}{\mathcal{F}(g)(y)} e^{-iyx} dy, \quad (4.2)$$

where  $\overline{\mathcal{F}(\gamma)}(y)$  denotes the complex conjugate of  $\mathcal{F}(\gamma)(y)$ ,  $\nu = 1$  and  $\rho = \delta$  (appearing in **(G5)**).

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

$$\mathbb{E} \left( \|\hat{f} - f\|_2^2 \right) \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\delta+1)}.$$

Theorem 4.1 improves (Chesneau, 2012, Proposition 5.1) in terms of rate of convergence; we gain a logarithmic term.

Moreover, it is established that, in the i.i.d. setting, “ $((\ln n)/n)^{2s/(2s+2\delta+1)}$ ,” is

- exactly the rate of convergence attained by the wavelet hard thresholding estimator,
- the near optimal rate of convergence in the minimax sense.

The details can be found in Fan and Koo (2002). Thus Theorem 4.1 can be viewed as an extension of this existing result to the weak dependent case.

#### 4.2. GARCH-type model

We consider the strictly stationary stochastic process  $(V_t)_{t \in \mathbb{Z}}$  where, for any  $t \in \mathbb{Z}$ ,

$$V_t = \sigma_t^2 Z_t, \quad (4.3)$$

$(\sigma_t^2)_{t \in \mathbb{Z}}$  is a strictly stationary stochastic process with unknown density  $f$  and  $(Z_t)_{t \in \mathbb{Z}}$  is a strictly stationary stochastic process with known density  $g$ . It is supposed that  $\sigma_t^2$  and  $Z_t$  are independent for any  $t \in \mathbb{Z}$  and  $(V_t)_{t \in \mathbb{Z}}$  is a  $\alpha$ -mixing process with exponential decay rate (see Subsection 3.1 for a precise definition). Our aim is to estimate  $f$  via  $V_1, \dots, V_n$  from  $(V_t)_{t \in \mathbb{Z}}$ . Some related works are Comte et al. (2008) and Chesneau (2013a).

We formulate the following assumptions:

**(J1)** There exists a positive integer  $\delta$  such that

$$g(x) = \frac{1}{(\delta - 1)!} (-\ln x)^{\delta-1}, \quad x \in [0, 1].$$

Let us remark that  $g$  is the density of  $\prod_{i=1}^{\delta} U_i$ , where  $U_1, \dots, U_{\delta}$  are  $\delta$  i.i.d. random variables having the common distribution  $\mathcal{U}([0, 1])$ .



- (J2) The support of  $f$  is  $[0, 1]$  and  $f \in \mathbb{L}^2([0, 1])$ .  
(J3) Let  $u$  be the density of  $V_1$ . We suppose that there exists a constant  $C > 0$  such that

$$\sup_{x \in \mathbb{R}} u(x) \leq C.$$

- (J4) For any  $m \in \mathbb{Z}$ , let  $u_{(V_1, V_{m+1})}$  be the density of  $(V_1, V_{m+1})$ . We suppose that there exists a constant  $C > 0$  such that

$$\sup_{m \in \mathbb{Z}} \sup_{(x, y) \in \mathbb{R}^2} u_{(V_1, V_{m+1})}(x, y) \leq C.$$

We are now in the position to present the result.

**Theorem 4.2.** *We consider the model (4.3). Suppose that (J1)-(J4) are satisfied. Let  $\hat{f}$  be defined as in (3.1) with*

$$q(\gamma, x) = T_\delta(\gamma)(x), \quad (4.4)$$

where, for any positive integer  $\ell$ ,  $T(\gamma)(x) = (x\gamma(x))'$  and  $T_\ell(\gamma)(x) = T(T_{\ell-1}(\gamma))(x)$ ,  $\nu = 1$  and  $\rho = \delta$  (appearing in (J1)).

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

$$\mathbb{E} \left( \|\hat{f} - f\|_2^2 \right) \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\delta+1)}.$$

Theorem 4.2 significantly improves (Chesneau, 2013a, Theorem 2) in terms of rate of convergence; we gain an exponent  $1/2$ .

## Conclusion

Considering the weak dependent case on the observations, we prove a general result on the rate of convergence attained by a hard wavelet thresholding estimator under the MISE over Besov balls. This result is flexible; it can be applied for a wide class of statistical models. Moreover, the obtained rate of convergence is sharp; it can correspond to the near optimal one in the minimax sense for the standard i.i.d. case. Some recent results on sophisticated statistical problems are improved. Thanks to its flexibility, the perspectives of applications of our theoretical result in other contexts are numerous.

## 5. Proofs

### 5.1. Intermediary result

Proposition 5.1 below proves probability and moments inequalities satisfied by the estimators (3.2).

**Proposition 5.1.** Let  $\hat{\alpha}_{j,k}$  and  $\hat{\beta}_{j,k}$  be defined as in (3.2) under **(H1)** and **(H2)**,  $j_0$  be (3.3) and  $j_1$  be (3.4).

(a) There exists a constant  $C > 0$  such that, for any  $j \in \{j_0, \dots, j_1\}$  and  $k \in \Lambda_j$ ,

$$\mathbb{E}((\hat{c}_{j,k} - c_{j,k})^2) \leq C 2^{2\rho j} \frac{1}{n}$$

and

$$\mathbb{E}\left(\left(\hat{d}_{j,k} - d_{j,k}\right)^2\right) \leq C 2^{2\rho j} \frac{1}{n}.$$

(b) There exists a constant  $C > 0$  such that, for any  $j \in \{j_0, \dots, j_1\}$  and  $k \in \Lambda_j$ ,

$$\mathbb{E}\left(\left(\hat{d}_{j,k} - d_{j,k}\right)^4\right) \leq C 2^{4\rho j}.$$

(c) Let  $\lambda_j$  be defined as in (3.5). There exists a constant  $C > 0$  such that, for any  $\kappa$  large enough,  $j \in \{j_0, \dots, j_1\}$  and  $k \in \Lambda_j$ , we have

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

**Proof of Proposition 5.1.**

(a) Using **(H1)** and the stationarity of  $(V_t)_{t \in \mathbb{Z}}$ , we obtain

$$\begin{aligned} \mathbb{E}((\hat{c}_{j,k} - c_{j,k})^2) &= \mathbb{V}(\hat{c}_{j,k}) \\ &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(q(\phi_{j,k}, V_i)) + \frac{2}{n^2} \sum_{v=2}^n \sum_{\ell=1}^{v-1} \mathbb{C}_{ov}(q(\phi_{j,k}, V_v), q(\phi_{j,k}, V_\ell)) \\ &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(q(\phi_{j,k}, V_i)) + \frac{2}{n^2} \sum_{m=1}^{n-1} (n-m) \mathbb{C}_{ov}(q(\phi_{j,k}, V_{m+1}), q(\phi_{j,k}, V_1)) \\ &\leq \frac{1}{n} \left( \mathbb{E}((q(\phi_{j,k}, V_1))^2) + 2 \sum_{m=1}^{n-1} |\mathbb{C}_{ov}(q(\phi_{j,k}, V_{m+1}), q(\phi_{j,k}, V_1))| \right). \end{aligned} \tag{5.1}$$

By **(H2)-(ii)** we get

$$\mathbb{E}((q(\phi_{j,k}, V_1))^2) \leq C 2^{2\rho j}. \tag{5.2}$$

For the covariance term, note that

$$\sum_{m=1}^{n-1} |\mathbb{C}_{ov}(q(\phi_{j,k}, V_{m+1}), q(\phi_{j,k}, V_1))| = A + B,$$

where

$$A = \sum_{m=1}^{[(\ln n)/\theta]-1} |\mathbb{C}_{ov}(q(\phi_{j,k}, V_{m+1}), q(\phi_{j,k}, V_1))|$$

and

$$B = \sum_{m=[(\ln n)/\theta]}^{n-1} |\mathbb{C}_{ov}(q(\phi_{j,k}, V_{m+1}), q(\phi_{j,k}, V_1))|.$$

It follows from **(H2)**-(iii) and  $2^{-j} \leq 2^{-j_0} < 2(\ln n)^{-\nu}$  that

$$A \leq C2^{2\rho j} (\ln n)^{\nu-1} 2^{-j} [(\ln n)/\theta] \leq C2^{2\rho j}.$$

The Davydov inequality described in Lemma 5.1 with  $p = q = 4$ , **(H2)**-(i)-(ii) and  $2^j \leq 2^{j_1} \leq n$  give

$$\begin{aligned} B &\leq C \sqrt{\mathbb{E}((q(\phi_{j,k}, V_1))^4)} \sum_{m=[(\ln n)/\theta]}^{n-1} \sqrt{\alpha_m} \\ &\leq C2^{2\rho j} 2^{j/2} \sqrt{\mathbb{E}((q(\phi_{j,k}, V_1))^2)} \sum_{m=[(\ln n)/\theta]}^{\infty} e^{-\theta m/2} \\ &= C2^{2\rho j} \sqrt{n} e^{-(\ln n)/2} \leq C2^{2\rho j}. \end{aligned}$$

Thus

$$\sum_{m=1}^{n-1} |\mathbb{C}_{ov}(q(\phi_{j,k}, V_{m+1}), q(\phi_{j,k}, V_1))| \leq C2^{2\rho j}. \quad (5.3)$$

Putting (5.1), (5.2) and (5.3) together, the first point in **(a)** is proved. The proof of the second point is identical with  $\psi$  instead of  $\phi$ .

- (b) Thanks to **(H2)**-(i), we have  $|\hat{d}_{j,k}| \leq \sup_{x \in V_1(\Omega)} |q(\psi_{j,k}, x)| \leq C2^{2\rho j} 2^{j/2}$ . It follows from the triangular inequality and  $|d_{j,k}| \leq \|f\|_2 \leq C$  that

$$|\hat{d}_{j,k} - d_{j,k}| \leq |\hat{d}_{j,k}| + |d_{j,k}| \leq C2^{2\rho j} 2^{j/2}.$$

This inequality and the second result of **(a)** yield

$$\mathbb{E}((\hat{d}_{j,k} - d_{j,k})^4) \leq C2^{2\rho j} 2^j \mathbb{E}((\hat{d}_{j,k} - d_{j,k})^2) \leq C2^{4\rho j} 2^j \frac{1}{n}.$$

Using  $2^j \leq 2^{j_1} \leq n$ , the proof of **(b)** is completed.

- (c) We will use the Liebscher inequality described in Lemma 5.2. Let us set

$$U_i = q(\psi_{j,k}, V_i) - \mathbb{E}(q(\psi_{j,k}, V_1)).$$

We have  $\mathbb{E}(U_1) = 0$  and, by **(H2)**-(i) and  $2^j \leq 2^{j_1} \leq n/(\ln n)^3$ ,

$$|U_i| \leq 2 \sup_{x \in V_1(\Omega)} |q(\psi_{j,k}, x)| \leq C2^{2\rho j} 2^{j/2} \leq C2^{2\rho j} \sqrt{\frac{n}{(\ln n)^3}},$$

(so  $M = C2^{2\rho j} \sqrt{n/(\ln n)^3}$ ).

Proceeding as for the proofs of the bounds in **(a)**, for any integer  $l \leq C \ln n$ , since  $2^{-j} \leq 2^{-j_0} \leq 2(\ln n)^{-\nu}$ , we show that

$$\mathbb{V} \left( \sum_{i=1}^l U_i \right) = \mathbb{V} \left( \sum_{i=1}^l q(\psi_{j,k}, V_i) \right) \leq C2^{2\rho j} (l + l^2 (\ln n)^{\nu-1} 2^{-j}) \leq C2^{2\rho j} l.$$

Therefore

$$D_m = \max_{l \in \{1, \dots, 2m\}} \mathbb{V} \left( \sum_{i=1}^l U_i \right) \leq C2^{2\rho j} m. \quad (5.4)$$

Owing to Lemma 5.2 applied with  $U_1, \dots, U_n$ ,  $\lambda = \kappa\lambda_j/2$ ,  $m = \lceil \sqrt{\kappa} \ln n \rceil$ ,  $M = C2^{\rho j} \sqrt{n}/(\ln n)^3$  and the bound (5.4), we obtain

$$\begin{aligned} & \mathbb{P}\left(|\hat{d}_{j,k} - d_{j,k}| \geq \kappa\lambda_j/2\right) \\ & \leq C \left( \exp\left(-C \frac{\kappa^2 \lambda_j^2 n}{D_m/m + \kappa\lambda_j m M}\right) + \frac{M}{\lambda_j} n e^{-\theta m} \right) \\ & \leq C \left( \exp\left(-C \frac{\kappa^2 2^{2\rho j} \ln n}{2^{2\rho j} + \kappa 2^{\rho j} \sqrt{(\ln n)/n} \lceil \sqrt{\kappa} \ln n \rceil 2^{\rho j} \sqrt{n}/(\ln n)^3}\right) \right) \\ & \quad + \sqrt{\frac{n/(\ln n)^3}{(\ln n)/n}} n e^{-\theta \lceil \sqrt{\kappa} \ln n \rceil} \\ & \leq C \left( n^{-C\kappa^2/(1+\kappa^{3/2})} + n^{2-\theta\sqrt{\kappa}} \right). \end{aligned}$$

Taking  $\kappa$  large enough, the last term is bounded by  $C/n^4$ . This completes the proof of (c).

This completes the proof of Proposition 5.1.  $\square$

### 5.2. Proof of Theorem 3.1

Theorem 3.1 can be proved by combining arguments of (Kerkycharian and Picard, 2000, Theorem 5.1) and (Chesneau, 2008, Theorem 4.2). It is closed to (Chesneau, 2012, Proof of Theorem 2) by taking  $\theta \rightarrow \infty$ . The interested reader can find the details below.

We consider the following wavelet decomposition for  $f$ :

$$f(x) = \sum_{k \in \Lambda_{j_0}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_{k \in \Lambda_j} d_{j,k} \psi_{j,k}(x),$$

where  $c_{j_0,k} = \int_0^1 f(x) \phi_{j_0,k}(x) dx$  and  $d_{j,k} = \int_0^1 f(x) \psi_{j,k}(x) dx$ .

Using the orthonormality of the wavelet basis  $\mathcal{B}$ , the MISE of  $\hat{f}$  can be express as

$$\mathbb{E}\left(\|\hat{f} - f\|_2^2\right) = P + Q + R, \tag{5.5}$$

where

$$P = \sum_{k \in \Lambda_{j_0}} \mathbb{E}\left((\hat{c}_{j_0,k} - c_{j_0,k})^2\right), \quad Q = \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \mathbb{E}\left(\left(\hat{d}_{j,k} \mathbf{1}_{\{|\hat{d}_{j,k}| \geq \kappa\lambda_j\}} - d_{j,k}\right)^2\right)$$

and

$$R = \sum_{j=j_1+1}^{\infty} \sum_{k \in \Lambda_j} d_{j,k}^2.$$

Let us now investigate sharp upper bounds for  $P$ ,  $R$  and  $Q$  successively.

*Upper bound for  $P$ :*

The point (a) of Proposition 5.1 and  $2s/(2s + 2\rho + 1) < 1$  yield

$$P \leq C \frac{2^{j_0}}{n} \leq C \frac{(\ln n)^\nu}{n} \leq C \left(\frac{\ln n}{n}\right)^{2s/(2s+2\rho+1)}. \tag{5.6}$$

*Upper bound for  $R$ :*

- For  $r \geq 1$  and  $p \geq 2$ , we have  $f \in B_{p,r}^s(M) \subseteq B_{2,\infty}^s(M)$ . Using  $2s/(2s+2\rho+1) < 2s/(2\rho+1)$ , we obtain

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

- For  $r \geq 1$  and  $p \in [1, 2)$ , we have  $f \in B_{p,r}^s(M) \subseteq B_{2,\infty}^{s+1/2-1/p}(M)$ . The condition  $s > (2\rho+1)/p$  implies that  $(s+1/2-1/p)/(2\rho+1) > s/(2s+2\rho+1)$ . Thus

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

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

$$R \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}. \quad (5.7)$$

Upper bound for  $Q$ :

Adopting the notation  $\hat{D}_{j,k} = \hat{d}_{j,k} - d_{j,k}$ ,  $Q$  can be written as

$$Q = \sum_{i=1}^4 Q_i, \quad (5.8)$$

where

$$\begin{aligned} Q_1 &= \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \mathbb{E} \left( \hat{D}_{j,k}^2 \mathbf{1}_{\{|\hat{d}_{j,k}| \geq \kappa \lambda_j, |d_{j,k}| < \kappa \lambda_j / 2\}} \right), \\ Q_2 &= \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \mathbb{E} \left( \hat{D}_{j,k}^2 \mathbf{1}_{\{|\hat{d}_{j,k}| \geq \kappa \lambda_j, |d_{j,k}| \geq \kappa \lambda_j / 2\}} \right), \\ Q_3 &= \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \mathbb{E} \left( d_{j,k}^2 \mathbf{1}_{\{|\hat{d}_{j,k}| < \kappa \lambda_j, |d_{j,k}| \geq 2\kappa \lambda_j\}} \right) \end{aligned}$$

and

$$Q_4 = \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \mathbb{E} \left( d_{j,k}^2 \mathbf{1}_{\{|\hat{d}_{j,k}| < \kappa \lambda_j, |d_{j,k}| < 2\kappa \lambda_j\}} \right).$$

Upper bound for  $Q_1 + Q_3$  :

Owing to the inequalities  $\mathbf{1}_{\{|\hat{d}_{j,k}| < \kappa \lambda_j, |d_{j,k}| \geq 2\kappa \lambda_j\}} \leq \mathbf{1}_{\{|\hat{D}_{j,k}| > \kappa \lambda_j / 2\}}$ ,  $\mathbf{1}_{\{|\hat{d}_{j,k}| \geq \kappa \lambda_j, |d_{j,k}| < \kappa \lambda_j / 2\}} \leq \mathbf{1}_{\{|\hat{D}_{j,k}| > \kappa \lambda_j / 2\}}$  and  $\mathbf{1}_{\{|\hat{d}_{j,k}| < \kappa \lambda_j, |d_{j,k}| \geq 2\kappa \lambda_j\}} \leq \mathbf{1}_{\{|d_{j,k}| \leq 2|\hat{D}_{j,k}|\}}$ , the Cauchy-Schwarz inequality, the points (b) and (c) of Proposition 5.1, we have

$$\begin{aligned} Q_1 + Q_3 &\leq C \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \mathbb{E} \left( \hat{D}_{j,k}^2 \mathbf{1}_{\{|\hat{D}_{j,k}| > \kappa \lambda_j / 2\}} \right) \\ &\leq C \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \left( \mathbb{E} \left( \hat{D}_{j,k}^4 \right) \right)^{1/2} \left( \mathbb{P} \left( |\hat{D}_{j,k}| > \kappa \lambda_j / 2 \right) \right)^{1/2} \\ &\leq C \frac{1}{n^2} \sum_{j=j_0}^{j_1} 2^{j(1+2\rho)} \leq C \frac{1}{n} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}. \end{aligned} \quad (5.9)$$

Upper bound for  $Q_2$  :

It follows from the point (a) of Proposition 5.1 that

$$\begin{aligned} Q_2 &\leq \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} \mathbb{E} \left( \hat{D}_{j,k}^2 \right) \mathbf{1}_{\{|d_{j,k}| \geq \kappa \lambda_j / 2\}} \\ &\leq C \frac{1}{n} \sum_{j=j_0}^{j_1} 2^{2\rho j} \sum_{k \in \Lambda_j} \mathbf{1}_{\{|d_{j,k}| > \kappa \lambda_j / 2\}}. \end{aligned}$$

Let us now introduce the integer  $j_*$  defined by

$$2^{j_*} = \left[ \left( \frac{n}{\ln n} \right)^{1/(2s+2\rho+1)} \right]. \quad (5.10)$$

Note that  $j_* \in \{j_0, \dots, j_1\}$  for  $n$  large enough.

Then  $Q_2$  can be bounded as

$$Q_2 \leq Q_{2,1} + Q_{2,2},$$

where

$$Q_{2,1} = C \frac{1}{n} \sum_{j=j_0}^{j_*} 2^{2\rho j} \sum_{k \in \Lambda_j} \mathbf{1}_{\{|d_{j,k}| > \kappa \lambda_j / 2\}}$$

and

$$Q_{2,2} = C \frac{1}{n} \sum_{j=j_*+1}^{j_1} 2^{2\rho j} \sum_{k \in \Lambda_j} \mathbf{1}_{\{|d_{j,k}| > \kappa \lambda_j / 2\}}.$$

On the one hand we have

$$Q_{2,1} \leq C \frac{\ln n}{n} \sum_{j=j_0}^{j_*} 2^{j(1+2\rho)} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}.$$

On the other hand,

- for  $r \geq 1$  and  $p \geq 2$ , the Markov inequality and  $f \in B_{p,r}^s(M) \subseteq B_{2,\infty}^s(M)$  yield

$$\begin{aligned} Q_{2,2} &\leq C \frac{\ln n}{n} \sum_{j=j_*+1}^{j_1} 2^{2\rho j} \frac{1}{\lambda_j^2} \sum_{k \in \Lambda_j} d_{j,k}^2 \leq C \sum_{j=j_*+1}^{\infty} \sum_{k \in \Lambda_j} d_{j,k}^2 \\ &\leq C \sum_{j=j_*+1}^{\infty} 2^{-2js} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}. \end{aligned}$$

- for  $r \geq 1$ ,  $p \in [1, 2)$  and  $s > (2\rho + 1)/p$ , the Markov inequality,  $f \in B_{p,r}^s(M)$  and  $(2s + 2\rho + 1)(2 - p)/2 + (s + 1/2 - 1/p + \rho - 2\rho/p)p = 2s$  imply that

$$\begin{aligned} Q_{2,2} &\leq C \frac{\ln n}{n} \sum_{j=j_*+1}^{j_1} 2^{2\rho j} \frac{1}{\lambda_j^p} \sum_{k \in \Lambda_j} |d_{j,k}|^p \\ &\leq C \left( \frac{\ln n}{n} \right)^{(2-p)/2} \sum_{j=j_*+1}^{\infty} 2^{j\rho(2-p)} 2^{-j(s+1/2-1/p)p} \\ &\leq C \left( \frac{\ln n}{n} \right)^{(2-p)/2} 2^{-j_*(s+1/2-1/p+\rho-2\rho/p)p} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}. \end{aligned}$$

Therefore, for  $r \geq 1$ ,  $\{p \geq 2$  and  $s > 0\}$  or  $\{p \in [1, 2)$  and  $s > (2\rho + 1)/p\}$ , we have

$$Q_2 \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}. \quad (5.11)$$

Upper bound for  $Q_4$  :  
We have

$$Q_4 \leq \sum_{j=j_0}^{j_1} \sum_{k \in \Lambda_j} d_{j,k}^2 \mathbf{1}_{\{|d_{j,k}| < 2\kappa \lambda_j\}}.$$

Let  $j_*$  be the integer (5.10). Then  $Q_4$  can be bound as

$$Q_4 \leq Q_{4,1} + Q_{4,2},$$

where

$$Q_{4,1} = \sum_{j=j_0}^{j_*} \sum_{k \in \Lambda_j} d_{j,k}^2 \mathbf{1}_{\{|d_{j,k}| < 2\kappa \lambda_j\}}, \quad Q_{4,2} = \sum_{j=j_*+1}^{j_1} \sum_{k \in \Lambda_j} d_{j,k}^2 \mathbf{1}_{\{|d_{j,k}| < 2\kappa \lambda_j\}}.$$

On the one hand, we have

$$Q_{4,1} \leq C \sum_{j=j_0}^{j_*} 2^j \lambda_j^2 = C \frac{\ln n}{n} \sum_{j=j_0}^{j_*} 2^{j(1+2\rho)} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}.$$

On the other hand,

- for  $r \geq 1$  and  $p \geq 2$ , since  $f \in B_{p,r}^s(M) \subseteq B_{2,\infty}^s(M)$ , we have

$$Q_{4,2} \leq \sum_{j=j_*+1}^{\infty} \sum_{k \in \Lambda_j} d_{j,k}^2 \leq C \sum_{j=j_*+1}^{\infty} 2^{-2js} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}.$$

- for  $r \geq 1$ ,  $p \in [1, 2)$  and  $s > (2\rho + 1)/p$ , owing to the Markov inequality,  $f \in B_{p,r}^s(M)$  and  $(2s + 2\rho + 1)(2 - p)/2 + (s + 1/2 - 1/p + \rho - 2\rho/p)p = 2s$ , we get

$$\begin{aligned} Q_{4,2} &\leq C \sum_{j=j_*+1}^{j_1} \lambda_j^{2-p} \sum_{k \in \Lambda_j} |d_{j,k}|^p \\ &= C \left( \frac{\ln n}{n} \right)^{(2-p)/2} \sum_{j=j_*+1}^{j_1} 2^{j\rho(2-p)} \sum_{k \in \Lambda_j} |d_{j,k}|^p \\ &\leq C \left( \frac{\ln n}{n} \right)^{(2-p)/2} \sum_{j=j_*+1}^{\infty} 2^{j\rho(2-p)} 2^{-j(s+1/2-1/p)p} \\ &\leq C \left( \frac{\ln n}{n} \right)^{(2-p)/2} 2^{-j_*(s+1/2-1/p+\rho-2\rho/p)p} \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}. \end{aligned}$$

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

$$Q_4 \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}. \quad (5.12)$$

Putting (5.8), (5.9), (5.11) and (5.12) together, for  $r \geq 1$ ,  $\{p \geq 2$  and  $s > 0\}$  or  $\{p \in [1, 2)$  and  $s > (2\rho + 1)/p\}$ , we obtain

$$Q \leq C \left( \frac{\ln n}{n} \right)^{2s/(2s+2\rho+1)}. \quad (5.13)$$

Combining (5.5), (5.6), (5.7) and (5.13), we complete the proof of Theorem 3.1.  $\square$

### 5.3. Proof of Theorem 4.1

The proof of Theorem 4.1 is a direct application of Theorem 3.1: under **(G1)**-**(G5)**, the function  $q$  defined by (4.2) satisfies **(H1)**, see (Fan and Koo, 2002, equation (2)) and **(H2)**: **(i)** see (Fan and Koo, 2002, Lemma 6), **(ii)** see (Fan and Koo, 2002, equation (11)) and **(iii)** see (Chesneau, 2012, Proof of Proposition 6.1).  $\square$

### 5.4. Proof of Theorem 4.2

The proof of Theorem 4.2 is a consequence of Theorem 3.1: under **(J1)**-**(J4)**, the function  $q$  defined by (4.4) satisfies **(H1)** and **(H2)**: **(i)**- **(ii)** see (Chesneau, 2013a, Proposition 1) and **(iii)** see (Chesneau and Doosti, 2012, equation (26)).  $\square$

### Key lemmas

In this section we present two lemmas which have been used in the proofs.

Lemma 5.1 below shows a sharp covariance inequality under the  $\alpha$ -mixing condition.

**Lemma 5.1** (Davydov (1970)). *Let  $(W_t)_{t \in \mathbb{Z}}$  be a strictly 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  $\mathbb{E}(|h(W_1)|^p)$  and  $\mathbb{E}(|k(W_1)|^q)$  exist. Then there exists a constant  $C > 0$  such that*

$$|\mathbb{C}_{ov}(h(W_1), k(W_{m+1}))| \leq C \alpha_m^{1-1/p-1/q} (\mathbb{E}(|h(W_1)|^p))^{1/p} (\mathbb{E}(|k(W_1)|^q))^{1/q}.$$

Lemma 5.2 below presents a concentration inequality for  $\alpha$ -mixing processes.

**Lemma 5.2** (Liebscher (2001)). *Let  $(W_t)_{t \in \mathbb{Z}}$  be a strictly stationary process with the  $m$ -th strongly mixing coefficient  $\alpha_m$ ,  $m \geq 0$ ,  $n$  be a positive integer,  $h : \mathbb{R} \rightarrow \mathbb{C}_{ov}$  be a measurable function and, for any  $t \in \mathbb{Z}$ ,  $U_t = h(W_t)$ . We assume that  $\mathbb{E}(U_1) = 0$  and there exists a constant  $M > 0$  satisfying  $|U_1| \leq M$ . Then, for any  $m \in \{1, \dots, [n/2]\}$  and  $\lambda > 0$ , we have*

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

where

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



## References

- Abbaszadeh, M. and Emadi, M. (2013). Wavelet Density Estimation and Statistical Evidences Role for a GARCH Model in the Weighted Distribution, *Applied Mathematics*, 4, 2, 10-416.
- Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *Journal of the Italian Statistical Society*, 6, 97-144.
- Bosq, D. (1998). *Nonparametric statistics for stochastic processes*. Estimation and Prediction. Lecture Notes in Statistics, 110, Springer Verlag, New York.
- Bradley, R.C. (2007). *Introduction to strong mixing conditions*, Vol. 1,2,3. Kendrick Press.
- Cai, J.J. and Liang, H.Y. (2011). Nonlinear Wavelet Density Estimation for Truncated and Dependent Observations. *International Journal of Wavelets, Multiresolution and Information Processing*, 9, 4, 587-609.
- Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39.
- Chaubey, Y.P. and Shirazi, E. (2013). On MISE of a Nonlinear Wavelet Estimator of the Regression Function Based on Biased Data under Strong Mixing. *Communications in Statistics: Theory and Methods*, (to appear).
- Chesneau, C. (2008). Wavelet estimation via block thresholding: A minimax study under  $L^p$ -risk. *Statist. Sinica*, 18, 1007-1024.
- Chesneau, C. and Doosti, H. (2012). Wavelet linear density estimation for a GARCH model under various dependence structures. *Journal of the Iranian Statistical Society*, 11, 1, 1-21.
- Chesneau, C. (2012). On the adaptive wavelet deconvolution of a density for strong mixing sequences. *Journal of the Korean Statistical Society*, 41, 4, 423-436.
- Chesneau, C. (2013a). Wavelet Estimation of a Density in a GARCH-type Model. *Communications in Statistics: Theory and Methods*, 42, 1, 98-117.
- Chesneau, C. (2013b). On the adaptive wavelet estimation of a multidimensional regression function under  $\alpha$ -mixing dependence: Beyond the boundedness assumption on the noise. *Preprint LMNO*, <http://hal.archives-ouvertes.fr/hal-00727150/>.
- Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24, 54-81.
- Comte, F., Dedecker, J. and Taupin, M.-L. (2008). Adaptive density deconvolution for dependent inputs with measurement errors. *Mathematical Methods of Statistics*, 17, 2, 87-112.
- Daubechies, I. (1992). *Ten lectures on wavelets*, SIAM.
- Davydov, Y. (1970). The invariance principle for stationary processes. *Theor. Probab. Appl.*, 15, 3, 498-509.
- Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators. *Applied Computational Harmonic Analysis*, 3, 215-228.
- 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, 1200-1224.
- Donoho, D., Johnstone, I., Kerkycharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia?. *Journal of the Royal Statistical Society, Ser. B*, 57, 301-369.
- Donoho, D.L., Johnstone, I.M., Kerkycharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics*, 24, 508-539.
- 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.
- Fan, J. and Koo, J.Y. (2002). Wavelet deconvolution. *IEEE transactions on information theory*, 48, 734-747.
- Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, 6(6), 1051-1079.
- Härdle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*. Lectures Notes in Statistics, 129, Springer Verlag, New York.
- Kerkycharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2), 283-344.
- Kulik, R. (2008). Nonparametric deconvolution problem for dependent sequences. *Electronic J. of Statist.*, 2, 722-740.
- Leblanc, F. (1996). Wavelet linear density estimator for a discrete time stochastic process:  $L_p$ -losses. *Statistics and Probability Letters*, 27, 71-84.
- Liebscher, E. (2001). Estimation of the density and the regression function under mixing conditions. *Statist. Decisions*, 19, (1), 9-26.
- 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. (1993). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stoch. Processes Appl.*, 47, 53-74.
- Masry, E. (1996a). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Ser. Anal.*, 17, 6, 571-599.
- Masry, E. (1996b). Multivariate regression estimation-local polynomial fitting for time series. *Stochastic Process. Appl.*, 65, 1, 81-101.
- Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.*, 24, 2, 165-179.

- Masry, E. (2000). Wavelet-Based estimation of multivariate regression functions in besov spaces. *Journal of Nonparametric Statistics*, 12, 2, 283-308.
- Meyer, Y. (1990). *Ondelettes et Opérateurs*. Hermann, Paris.
- Niu, S.L.. and Liang, H.Y. (2011). Nonlinear Wavelet Estimation of Conditional Density under Left-Truncated and  $\alpha$ -Mixing Assumptions. *International Journal of Wavelets, Multiresolution and Information Processing*, 9, 6, 989-1023.
- 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.
- Robinson, P.M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4, 185-207.
- Roussas, G.G. (1987). Nonparametric estimation in mixing sequences of random variables. *J. Statist. Plann. Inference*, 18, 135-149
- Roussas, G.G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Process. Appl.*, 36 (1), 107-116.
- Tran L.H. (1993). Nonparametric function estimation for time series by local average estimators. *The Annals of Statistics*, 21, 1040-1057.
- Tribouley, K. and Viennet, G. (1998).  $L_p$  adaptive density estimation in a  $\beta$ -mixing framework. *Ann. Inst. H. Poincar Probab. Statist.*, 34, 179-208.
- Truong, Y.K. and Stone, C.J. (1992). Nonparametric function estimation involving time series. *The Annals of Statistics*, 20, 77-97.
- Tsybakov, A.B. (2004). *Introduction à l'estimation non-paramétrique*, Springer, New York.
- van Zanten, H. and Zareba, P. (2008). A note on wavelet density deconvolution for weakly dependent data. *Stat. Inference Stoch. Process.*, 11, 207-219.