



**HAL**  
open science

## Adaptive density estimation for stationary processes

Matthieu Lerasle

► **To cite this version:**

Matthieu Lerasle. Adaptive density estimation for stationary processes. *Mathematical Methods of Statistics*, 2009, 18 (1), pp.59–83. 10.3103/S1066530709010049 . hal-00413692

**HAL Id: hal-00413692**

**<https://hal.science/hal-00413692v1>**

Submitted on 4 Sep 2009

**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.

# Adaptive density estimation of stationary $\beta$ -mixing and $\tau$ -mixing processes.

Matthieu Lerasle\*

Abstract:

We propose an algorithm to estimate the common density  $s$  of a stationary process  $X_1, \dots, X_n$ . We suppose that the process is either  $\beta$  or  $\tau$ -mixing. We provide a model selection procedure based on a generalization of Mallows'  $C_p$  and we prove oracle inequalities for the selected estimator under a few prior assumptions on the collection of models and on the mixing coefficients. We prove that our estimator is adaptive over a class of Besov spaces, namely, we prove that it achieves the same rates of convergence as in the i.i.d framework.

**Key words:** Density estimation, weak dependence, model selection.

**2000 Mathematics Subject Classification:** 62G07, 62M99.

## 1 Introduction

We consider the problem of estimating the unknown density  $s$  of  $P$ , the law of a random variable  $X$ , based on the observation of  $n$  (possibly) dependent data  $X_1, \dots, X_n$  with common law  $P$ . We assume that  $X$  is real valued, that  $s$  belongs to  $L^2(\mu)$  where  $\mu$  denotes the Lebesgue measure on  $\mathbb{R}$  and that  $s$  is compactly supported, say in  $[0, 1]$ . Throughout the chapter, we consider least-squares estimators  $\hat{s}_m$  of  $s$  on a collection  $(S_m)_{m \in \mathcal{M}_n}$  of linear subspaces of  $L^2(\mu)$ . Our final estimator is chosen through a model selection algorithm.

Model selection has received much interest in the last decades. When its final goal is prediction, it can be seen more generally as the question of choosing between the outcomes of several prediction algorithms. With such a general formulation, a very natural answer is the following. First, estimate the prediction error for each model, that is  $\|s - \hat{s}_m\|_2^2$ . Then, select the model which minimizes this estimate.

It is natural to think of the empirical risk as an estimator of the prediction error. This can fail dramatically, because it uses the same data for building predictors and for comparing them, making these estimates strongly biased for models involving a number of parameters growing with the sample size.

In order to correct this drawback, penalization's methods state that a good choice can be made by minimizing the sum of the empirical risk (how do algorithms fit the data) and some complexity measure of the algorithms (called the penalty). This method was first developed in the work of Akaike [2] and [1] and Mallows [19].

In the context of density estimation, with independent data, Birgé & Massart [8] used penalties of order  $L_n D_m/n$ , where  $D_m$  denotes the dimension of  $S_m$  and  $L_n$  is a constant depending on the complexity of the collection  $\mathcal{M}_n$ . They used Talagrand's inequality (see for example Talagrand [24] for an overview) to prove that this penalization procedure is

---

\*Institut de Mathématiques (UMR 5219), INSA de Toulouse, Université de Toulouse, France

efficient *i.e.* the integrated quadratic risk of the selected estimator is asymptotically equivalent to the risk of the oracle (see Section 2 for a precise definition). They also proved that the selected estimator achieves adaptive rates of convergence over a large class of Besov spaces. Moreover, they showed that some methods of adaptive density estimation like the unbiased cross validation (Rudemo [23]) or the hard thresholded estimator of Donoho *et al.* [16] can be viewed as special instances of penalized projection estimators. More recently, Arlot [5] introduced new measures of the quality of penalized least-squares estimators (PLSE). He proved pathwise oracle inequalities, that is deviation bounds for the PLSE that are harder to prove but more informative from a practical point of view (see also Section 2 for details).

When the process  $(X_i)_{i=1,\dots,n}$  is  $\beta$ -mixing (Rozanov & Volkonskii [26] and Section 2), Talagrand's inequality can not be used directly. Baraud *et al.* [6] used Berbee's coupling lemma (see Berbee ([7]) and Viennet's covariance inequality (Viennet [25]) to overcome this problem and build model selection procedure in the regression problem. Then Comte & Merlevède [13] used this algorithm to investigate the problem of density estimation for a  $\beta$ -mixing process. They proved that under reasonable assumptions on the collection  $\mathcal{M}_n$  and on the coefficients  $\beta$ , one can recover the results of Birgé & Massart [8] in the i.i.d. framework.

The main drawback of those results is that many processes, even simple Markov chains are not  $\beta$ -mixing. For instance, if  $(\epsilon_i)_{i \geq 1}$  is iid with marginal  $\mathcal{B}(1/2)$ , then the stationary solution  $(X_i)_{i \geq 0}$  of the equation

$$X_n = \frac{1}{2}(X_{n-1} + \epsilon_n), \quad X_0 \text{ independent of } (\epsilon_i)_{i \geq 1} \quad (1)$$

is not  $\beta$ -mixing (Andrews [3]). More recently, Dedecker & Prieur [15] introduced new mixing-coefficients, in particular the coefficients  $\tau$ ,  $\tilde{\phi}$  and  $\tilde{\beta}$  and proved that many processes like (1) happen to be  $\tau$ ,  $\tilde{\phi}$  and  $\tilde{\beta}$ -mixing. They proved a coupling lemma for the coefficient  $\tau$  and covariance inequalities for  $\tilde{\phi}$  and  $\tilde{\beta}$ . Gannaz & Wintenberger [18] used the covariance inequality to extend the result of Donoho *et al.* [16] for the wavelet thresholded estimator to the case of  $\tilde{\phi}$ -mixing processes. They recovered (up to a  $\log(n)$  factor) the adaptive rates of convergence over Besov spaces.

In this article, we first investigate the case of  $\beta$ -mixing processes. We prove a pathwise oracle inequality for the PLSE. We extend the result of Comte & Merlevède [13] under weaker assumptions on the mixing coefficients. Then, we consider  $\tau$ -mixing processes. The problem is that the coupling result is weaker for the coefficient  $\tau$  than for  $\beta$ . Moreover, in order to control the empirical process we use a covariance inequality that is harder to handle. Hence, the generalization of the procedure of Baraud *et al.* [6] to the framework of  $\tau$ -mixing processes is not straightforward. We recover the optimal adaptive rates of convergence over Besov spaces (that is the same as in the independent framework) for  $\tau$ -mixing processes, which is new as far as we know.

The chapter is organized as follows. In Section 2, we give the basic material that we will use throughout the chapter. We recall the definition of some mixing coefficients and we state their properties. We define the penalized least-squares estimator (PLSE). Sections 3 and 4 are devoted to the statement of the main results, respectively in the  $\beta$ -mixing case and in the  $\tau$ -mixing case. In Section 5, we derive the adaptive properties of the PLSE. Finally, Section 6 is devoted to the proofs. Some additional material has been reported in the Appendix in Section 7.

## 2 Preliminaries

### 2.1 Notation.

Let  $(\Omega, \mathcal{A}, \mathbb{P})$  be a probability space. Let  $\mu$  be the Lebesgue measure on  $\mathbb{R}$ , let  $\|\cdot\|_p$  be the usual norm on  $L^p(\mu)$  for  $1 \leq p \leq \infty$ . For all  $y \in \mathbb{R}^l$ , let  $|y|_l = \sum_{i=1}^l |y_i|$ . Denote by  $\lambda_\kappa$  the set of  $\kappa$ -Lipschitz functions, *i.e.* the functions  $t$  from  $(\mathbb{R}^l, |\cdot|_l)$  to  $\mathbb{R}$  such that  $\text{Lip}(t) \leq \kappa$  where

$$\text{Lip}(t) = \sup \left\{ \frac{|t(x) - t(y)|}{|x - y|_l}, x, y \in \mathbb{R}^l, x \neq y \right\} \leq \kappa.$$

Let  $BV$  and  $BV_1$  be the set of functions  $t$  supported on  $\mathbb{R}$  satisfying respectively  $\|t\|_{BV} < \infty$  and  $\|t\|_{BV} \leq 1$  where

$$\|t\|_{BV} = \sup_{n \in \mathbb{N}^*} \sup_{-\infty < a_1 < \dots < a_n < \infty} |t(a_{i+1}) - t(a_i)|.$$

### 2.2 Some measures of dependence.

#### 2.2.1 Definitions and assumptions

Let  $Y = (Y_1, \dots, Y_l)$  be a random variable defined on  $(\Omega, \mathcal{A}, \mathbb{P})$  with values in  $(\mathbb{R}^l, |\cdot|_l)$ . Let  $\mathcal{M}$  be a  $\sigma$ -algebra of  $\mathcal{A}$ . Let  $\mathbb{P}_{Y|\mathcal{M}}, \mathbb{P}_{Y_1|\mathcal{M}}$  be conditional distributions of  $Y$  and  $Y_1$  given  $\mathcal{M}$ , let  $\mathbb{P}_Y, \mathbb{P}_{Y_1}$  be the distribution of  $Y$  and  $Y_1$  and let  $F_{Y_1|\mathcal{M}}, F_{Y_1}$  be distribution functions of  $\mathbb{P}_{Y_1|\mathcal{M}}$  and  $\mathbb{P}_{Y_1}$ . Let  $\mathcal{B}$  be the Borel  $\sigma$ -algebra on  $(\mathbb{R}^l, |\cdot|_l)$ . Define now

$$\begin{aligned} \beta(\mathcal{M}, \sigma(Y)) &= \mathbb{E} \left( \sup_{A \in \mathcal{B}} |\mathbb{P}_{Y|\mathcal{M}}(A) - \mathbb{P}_Y(A)| \right), \\ \tilde{\beta}(\mathcal{M}, Y_1) &= \mathbb{E} \left( \sup_{x \in \mathbb{R}} |F_{Y_1|\mathcal{M}}(x) - F_{Y_1}(x)| \right), \\ \text{and if } \mathbb{E}(|Y|) < \infty, \tau(\mathcal{M}, Y) &= \mathbb{E} \left( \sup_{t \in \lambda_1} |\mathbb{P}_{Y|\mathcal{M}}(t) - \mathbb{P}_Y(t)| \right). \end{aligned}$$

The coefficient  $\beta(\mathcal{M}, \sigma(Y))$  is the mixing coefficient introduced by Rozanov & Volkonskii [26]. The coefficients  $\tilde{\beta}(\mathcal{M}, Y_1)$  and  $\tau(\mathcal{M}, Y)$  have been introduced by Dedecker & Prieur [15].

Let  $(X_k)_{k \in \mathbb{Z}}$  be a stationary sequence of real valued random variables defined on  $(\Omega, \mathcal{A}, \mathbb{P})$ . For all  $k \in \mathbb{N}^*$ , the coefficients  $\beta_k, \tilde{\beta}_k$  and  $\tau_k$  are defined by

$$\beta_k = \beta(\sigma(X_i, i \leq 0), \sigma(X_i, i \geq k)), \quad \tilde{\beta}_k = \sup_{j \geq k} \{\tilde{\beta}(\sigma(X_p, p \leq 0), X_j)\}.$$

If  $\mathbb{E}(|X_1|) < \infty$ , for all  $k \in \mathbb{N}^*$  and all  $r \in \mathbb{N}^*$ , let

$$\tau_{k,r} = \max_{1 \leq l \leq r} \frac{1}{l} \sup_{k \leq i_1 < \dots < i_l} \{\tau(\sigma(X_p, p \leq 0), (X_{i_1}, \dots, X_{i_l}))\}, \quad \tau_k = \sup_{r \in \mathbb{N}^*} \tau_{k,r}.$$

Moreover, we set  $\beta_0 = 1$ . In the sequel, the processes of interest are either  $\beta$ -mixing or  $\tau$ -mixing, meaning that, for  $\gamma = \beta$  or  $\tau$ , the  $\gamma$ -mixing coefficients  $\gamma_k \rightarrow 0$  as  $k \rightarrow +\infty$ . For  $p \in \{1, 2\}$ , we define  $\kappa_p$  as:

$$\kappa_p = p \sum_{l=0}^{\infty} l^{p-1} \beta_l, \tag{2}$$

where  $0^0 = 1$ , when the series are convergent. Besides, we consider two kinds of rates of convergence to 0 of the mixing coefficients, that is for  $\gamma = \beta$  or  $\tau$ ,

**[AR]** arithmetical  $\gamma$ -mixing with rate  $\theta$  if there exists some  $\theta > 0$  such that  $\gamma_k \leq (1+k)^{-(1+\theta)}$  for all  $k$  in  $\mathbb{N}$ ,

**[GEO]** geometrical  $\gamma$ -mixing with rate  $\theta$  if there exists some  $\theta > 0$  such that  $\gamma_k \leq e^{-\theta k}$  for all  $k$  in  $\mathbb{N}$ .

## 2.2.2 Properties

### Coupling

Let  $X$  be an  $\mathbb{R}^l$ -valued random variable defined on  $(\Omega, \mathcal{A}, \mathbb{P})$  and let  $\mathcal{M}$  be a  $\sigma$ -algebra. Assume that there exists a random variable  $U$  uniformly distributed on  $[0, 1]$  and independent of  $\mathcal{M} \vee \sigma(X)$ . There exist two  $\mathcal{M} \vee \sigma(X) \vee \sigma(U)$ -measurable random variables  $X_1^*$  and  $X_2^*$  distributed as  $X$  and independent of  $\mathcal{M}$  such that

$$\beta(\mathcal{M}, \sigma(X)) = \mathbb{P}(X \neq X_1^*) \text{ and} \quad (3)$$

$$\tau(\mathcal{M}, X) = \mathbb{E}(|X - X_2^*|_l). \quad (4)$$

Equality (3) has been established by Berbee [7], Equality (4) has been established in Dedecker & Prieur [15], Section 7.1.

### Covariance inequalities

Let  $X, Y$  be two real valued random variables and let  $f, h$  be two measurable functions from  $\mathbb{R}$  to  $\mathbb{C}$ . Then, there exist two measurable functions  $b_1 : \mathbb{R} \rightarrow \mathbb{R}$  and  $b_2 : \mathbb{R} \rightarrow \mathbb{R}$  with  $\mathbb{E}(b_1(X)) = \mathbb{E}(b_2(Y)) = \beta(\sigma(X), \sigma(Y))$  such that, for any conjugate  $p, q \geq 1$  (see Viennet [25] Lemma 4.1)

$$|\text{Cov}(f(X), h(Y))| \leq 2\mathbb{E}^{1/p}(|f(X)|^p b_1(X)) \mathbb{E}^{1/q}(|h(Y)|^q b_2(Y)).$$

There exists a random variable  $b(\sigma(X), Y)$  such that  $\mathbb{E}(b(\sigma(X), Y)) = \tilde{\beta}(\sigma(X), Y)$  and such that, for all Lipschitz functions  $f$  and all  $h$  in  $BV$  (Dedecker & Prieur [15] Proposition 1)

$$|\text{Cov}(f(X), h(Y))| \leq \|h\|_{BV} \mathbb{E}(|f(X)| b(\sigma(X), Y)) \leq \|h\|_{BV} \|f\|_{\infty} \tilde{\beta}(\sigma(X), Y). \quad (5)$$

### Comparison results

Let  $(X_k)_{k \in \mathbb{Z}}$  be a sequence of identically distributed real random variables. If the marginal distribution satisfies a concentration's condition  $|F_X(x) - F_X(y)| \leq K|x - y|^a$  with  $a \leq 1$ ,  $K > 0$ , then (Dedecker *et al.* [14] Remark 5.1 p 104)

$$\tilde{\beta}_k \leq 2K^{1/(1+a)} \tau_{k,1}^{a/(a+1)} \leq 2K^{1/(1+a)} \tau_k^{a/(a+1)}.$$

In particular, if  $\mathbb{P}_X$  has a density  $s$  with respect to the Lebesgue measure  $\mu$  and if  $s \in L^2(\mu)$ , we have from Cauchy-Schwarz inequality

$$|F_X(x) - F_X(y)| = \left| \int \mathbf{1}_{[x,y]} s d\mu \right| \leq \|s\|_2 \left( \int \mathbf{1}_{[x,y]} d\mu \right)^{1/2} = \|s\|_2 |x - y|^{1/2},$$

thus

$$\tilde{\beta}_k \leq 2 \|s\|_2^{2/3} \tau_k^{1/3}.$$

In particular, for any arithmetically **[AR]**  $\tau$ -mixing process with rate  $\theta > 2$ , we have

$$\tilde{\beta}_k \leq 2 \|s\|_2^{2/3} (1+k)^{-(1+\theta)/3}. \quad (6)$$

### 2.2.3 Examples

Examples of  $\beta$ -mixing and  $\tau$ -mixing sequences are well known, we refer to the books of Doukhan [17] and Bradley [11] for examples of  $\beta$ -mixing processes and to the book of Dedecker *et. al* [14] or the articles of Dedecker & Prieur [15], Prieur [21], and Comte *et. al* [12] for examples of  $\tau$ -mixing sequences. One of the most important example is the following: a stationary, irreducible, aperiodic and positively recurrent Markov chain  $(X_i)_{i \geq 1}$  is  $\beta$ -mixing. However, many simple Markov chains are not  $\beta$ -mixing but are  $\tau$ -mixing. For instance, it is known for a long time that if  $(\epsilon_i)_{i \geq 1}$  are i.i.d Bernoulli  $\mathcal{B}(1/2)$ , then a stationary solution  $(X_i)_{i \geq 0}$  of the equation

$$X_n = \frac{1}{2}(X_{n-1} + \epsilon_n), \quad X_0 \text{ independent of } (\epsilon_i)_{i \geq 1}$$

is not  $\beta$ -mixing since  $\beta_k = 1$  for any  $k \geq 1$  whereas  $\tau_k \leq 2^{-k}$  (see Dedecker & Prieur [15] Section 4.1). Another advantage of the coefficient  $\tau$  is that it is easy to compute in many situations (see Dedecker & Prieur [15] Section 4).

### 2.3 Collections of models

We observe  $n$  identically distributed real valued random variables  $X_1, \dots, X_n$  with common density  $s$  with respect to the Lebesgue measure  $\mu$ . We assume that  $s$  belongs to the Hilbert space  $L^2(\mu)$  endowed with norm  $\|\cdot\|_2$ . We consider an orthonormal system  $\{\psi_{j,k}\}_{(j,k) \in \Lambda}$  of  $L_2(\mu)$  and a collection of models  $(S_m)_{m \in \mathcal{M}_n}$  indexed by subsets  $m \subset \Lambda$  for which we assume that the following assumptions are fulfilled:

[M<sub>1</sub>] for all  $m \in \mathcal{M}_n$ ,  $S_m$  is the linear span of  $\{\psi_{j,k}\}_{(j,k) \in m}$  with finite dimension  $D_m = |m| \geq 2$  and  $N_n = \max_{m \in \mathcal{M}_n} D_m$  satisfies  $N_n \leq n$ ;

[M<sub>2</sub>] there exists a constant  $\Phi$  such that

$$\forall m, m' \in \mathcal{M}_n, \forall t \in S_m, \forall t' \in S_{m'}, \|t + t'\|_\infty \leq \Phi \sqrt{\dim(S_m + S_{m'})} \|t + t'\|_2;$$

[M<sub>3</sub>]  $D_m \leq D_{m'}$  implies that  $m \subset m'$  and so  $S_m \subset S_{m'}$ .

As a consequence of Cauchy-Schwarz inequality, we have

$$\left\| \sum_{(j,k) \in m \cup m'} \psi_{j,k}^2 \right\|_\infty = \sup_{t \in S_m + S_{m'}, t \neq 0} \frac{\|t\|_\infty^2}{\|t\|_2^2} \quad (7)$$

see Birgé & Massart [8] p 58. Three examples are usually developed as fulfilling this set of assumptions:

[T] trigonometric spaces:  $\psi_{0,0}(x) = 1$  and for all  $j \in \mathbb{N}^*$ ,  $\psi_{j,1}(x) = \cos(2\pi jx)$ ,  $\psi_{j,2}(x) = \sin(2\pi jx)$ .  $m = \{(0,0), (j,1), (j',2), 1 \leq j, j' \leq J_m\}$  and  $D_m = 2J_m + 1$ ;

[P] regular piecewise polynomial spaces:  $S_m$  is generated by  $r$  polynomials  $\psi_{j,k}$  of degree  $k = 0, \dots, r-1$  on each subinterval  $[(j-1)/J_m, j/J_m]$  for  $j = 1, \dots, J_m$ ,  $D_m = rJ_m$ ,  $\mathcal{M}_n = \{m = \{(j,k), j = 1, \dots, J_m, k = 0, \dots, r-1\}, 1 \leq J_m \leq [n/r]\}$ ;

[W] spaces generated by dyadic wavelet with regularity  $r$  as described in Section 4.

For a precise description of those spaces and their properties, we refer to Birgé & Massart [8].

### 2.4 The estimator

Let  $(X_n)_{n \in \mathbb{Z}}$  be a real valued stationary process and let  $P$  denote the law of  $X_0$ . Assume that  $P$  has a density  $s$  with respect to the Lebesgue measure  $\mu$  and that  $s \in L_2(\mu)$ .

Let  $(S_m)_{m \in \mathcal{M}_n}$  be a collection of models satisfying assumptions  $[M_1]$ - $[M_3]$ . We define  $S_n = \cup_{m \in \mathcal{M}_n} S_m$ ,  $s_m$  and  $s_n$  the orthogonal projections of  $s$  onto  $S_m$  and  $S_n$  respectively, let  $\mathbb{P}$  be the joint distribution of the observations  $(X_n)_{n \in \mathbb{Z}}$  and let  $\mathbb{E}$  be the corresponding expectation. We define the operators  $P_n$ ,  $P$  and  $\nu_n$  on  $L^2(\mu)$  by

$$P_n t = \frac{1}{n} \sum_{i=1}^n t(X_i), \quad P t = \int t(x) s(x) d\mu(x), \quad \nu_n(t) = (P_n - P)t.$$

All the real numbers that we shall introduce and which are not indexed by  $m$  or  $n$  are fixed constants. In order to define the penalized least-squares estimator, let us consider on  $\mathbb{R} \times S_n$  the contrast function  $\gamma(x, t) = -2t(x) + \|t\|_2^2$  and its empirical version  $\gamma_n(t) = P_n \gamma(\cdot, t)$ . Minimizing  $\gamma_n(t)$  over  $S_m$  leads to the classical projection estimator  $\hat{s}_m$  on  $S_m$ . Let  $\hat{s}_n$  be the projection estimator on  $S_n$ . Since  $\{\psi_{j,k}\}_{(j,k) \in m}$  is an orthonormal basis of  $S_m$  one gets

$$\hat{s}_m = \sum_{(j,k) \in m} (P_n \psi_{j,k}) \psi_{j,k} \text{ and } \gamma_n(\hat{s}_m) = - \sum_{(j,k) \in m} (P_n \psi_{j,k})^2.$$

Now, given a penalty function  $\text{pen} : \mathcal{M}_n \rightarrow \mathbb{R}^+$ , we define a selected model  $\hat{m}$  as any element

$$\hat{m} \in \arg \min_{m \in \mathcal{M}_n} (\gamma_n(\hat{s}_m) + \text{pen}(m)) \quad (8)$$

and a PLSE is defined as any  $\tilde{s} \in S_{\hat{m}} \subset S_n$  such that

$$\gamma_n(\tilde{s}) + \text{pen}(\hat{m}) = \inf_{m \in \mathcal{M}_n} (\gamma_n(\hat{s}_m) + \text{pen}(m)). \quad (9)$$

## 2.5 Oracle inequalities

An ideal procedure for estimation chooses an oracle

$$m_o \in \text{Arg} \min_{m \in \mathcal{M}_n} \{\|s - \hat{s}_m\|_2\}.$$

An oracle depends on the unknown  $s$  and on the data so that it is unknown in practice. In order to validate our procedure, we try to prove: -non asymptotic oracle inequalities for the PLSE:

$$\mathbb{E} \left( \|s - \tilde{s}\|_2^2 \right) \leq L \inf_{m \in \mathcal{M}_n} \{ \mathbb{E} (\|s - \hat{s}_m\|_2^2 + R(m, n)) \}, \quad (10)$$

for some constant  $L \geq 1$  (as close to 1 as possible) and a remainder term  $R(m, n) \geq 0$  possibly random, and small compared to  $\mathbb{E} \left( \|s - \tilde{s}\|_2^2 \right)$  if possible. This inequality compares the risk of the PLSE with the best deterministic choice of  $m$ . Since  $\hat{m}$  is random, we prefer to prove a stronger form of oracle inequality :

$$\mathbb{E} \left( \|s - \tilde{s}\|_2^2 \right) \leq L \mathbb{E} \left( \inf_{m \in \mathcal{M}_n} \{ \|s - \hat{s}_m\|_2^2 + R(m, n) \} \right), \quad (11)$$

or, when it is possible, deviation bounds for the PLSE:

$$\mathbb{P} \left( \|s - \tilde{s}\|_2^2 > L \inf_{m \in \mathcal{M}_n} \left( \|s - \hat{s}_m\|_2^2 + R(m, n) \right) \right) \leq c_n, \quad (12)$$

where typically  $c_n \leq C/n^{1+\gamma}$  for some  $\gamma > 0$ . Inequality (12) proves that, asymptotically, the risk  $\|s - \tilde{s}\|_2^2$  is almost surely the one of the oracle. Let

$$\Omega = \left\{ \|s - \tilde{s}\|_2^2 > L \inf_{m \in \mathcal{M}_n} \left( \|s - \hat{s}_m\|_2^2 + R(m, n) \right) \right\}.$$

We have

$$\mathbb{E} \left( \|s - \tilde{s}\|_2^2 \right) = \mathbb{E} \left( \|s - \tilde{s}\|_2^2 1_\Omega \right) + \mathbb{E} \left( \|s - \tilde{s}\|_2^2 1_{\Omega^c} \right).$$

It is clear that  $\mathbb{E} \left( \|s - \tilde{s}\|_2^2 1_{\Omega^c} \right) \leq L \mathbb{E} \left( \inf_{m \in \mathcal{M}_n} \left\{ \|s - \hat{s}_m\|_2^2 + R(m, n) \right\} \right)$ . Moreover, we have  $\|s - \tilde{s}\|^2 = \|s - s_{\hat{m}}\|^2 + \|s_{\hat{m}} - \tilde{s}\|^2 \leq \|s\|^2 + \Phi^2 D_{\hat{m}} \leq \|s\|^2 + \Phi^2 n$ , thus, when (12) holds, we have

$$\mathbb{E} \left( \|s - \tilde{s}\|_2^2 1_{\Omega^c} \right) \leq (\|s\|^2 + \Phi^2 n) c_n \leq \frac{C}{n^\gamma}.$$

Therefore, inequality (12) implies

$$\mathbb{E} \left( \|s - \tilde{s}\|_2^2 \right) \leq \mathbb{E} \left( \inf_{m \in \mathcal{M}_n} \left\{ \|s - \hat{s}_m\|_2^2 + R(m, n) \right\} \right) + \frac{C}{n^\gamma}.$$

We can derive from these inequalities adaptive rates of convergence of the PLSE on Besov spaces (see Birgé & Massart [8] for example). In order to achieve this goal, we only have to prove a weaker form of oracle inequality where the remainder term  $R(m, n) \leq LD_m/n$  for some constant  $L$ , for all the models  $m$  with sufficiently large dimension. This will be detailed in Section 5.

### 3 Results for $\beta$ -mixing processes

From now on, the letters  $\kappa$ ,  $L$  and  $K$ , with various sub- or superscripts, will denote some constants which may vary from line to line. One shall use  $L$  to indicate more precisely the dependence on various quantities, especially those which are related to the unknown  $s$ .

In this section, we give the following theorem for  $\beta$ -mixing sequences. It can be seen as a pathwise version of Theorem 3.1 in Comte & Merlevède [13].

**Theorem 3.1** *Consider a collection of models satisfying  $[M_1]$ ,  $[M_2]$  and  $[M_3]$ . Assume that the process  $(X_n)_{n \in \mathbb{Z}}$  is strictly stationary and arithmetically **[AR]**  $\beta$ -mixing with mixing rate  $\theta > 2$  and that its marginal distribution admits a density  $s$  with respect to the Lebesgue measure  $\mu$ , with  $s \in L_2(\mu)$ .*

*Let  $\kappa_1$  be the constant defined in (2) and let  $\tilde{s}$  be the PLSE defined by (9) with*

$$\text{pen}(m) = \frac{K\Phi^2\kappa_1 D_m}{n}, \text{ where } K > 4.$$

*Then, for all  $\kappa > 2$  there exist  $c_0 > 0, L_s > 0, \gamma_1 > 0$  and a sequence  $\epsilon_n \rightarrow 0$ , such that*

$$\mathbb{P} \left( \|\tilde{s} - s\|_2^2 > (1 + \epsilon_n) \inf_{m \in \mathcal{M}_n, D_m \geq c_0(\log n)^{\gamma_1}} \left( \|s - s_m\|_2^2 + \text{pen}(m) \right) \right) \leq L_s \frac{(\log n)^{(\theta+2)\kappa}}{n^{\theta/2}}. \quad (13)$$

**Remark:** The term  $K\Phi^2\kappa_1$  is the same as in Theorem 3.1 of Comte & Merlevède [13] but with a constant  $K > 4$  instead of 320. The main drawback of this result is that the penalty term involves the constant  $\kappa_1$  which is unknown in practice. However, Theorem 3.1 ensures



that penalties proportional to the linear dimension of  $S_m$  lead to efficient model selection procedures. Thus we can use this information to apply the slope heuristic algorithm introduced by Birgé & Massart [9] in a Gaussian regression context and generalized by Arlot & Massart [4] to more general M-estimation frameworks. This algorithm calibrates the constant in front of the penalty term when the shape of an ideal penalty is available. The result of Arlot & Massart is proven for independent sequences, in a regression framework, but it can be generalized to the density estimation framework, for independent as well as for  $\beta$  or  $\tau$  dependent data. This result is beyond the scope of this chapter and will be proved in chapter 4.

We have to consider the infimum in equation (13) over the models with sufficiently large dimensions. However, as noted by Arlot [5] (Remark 9 p 43), we can take the infimum over all the models in (13) if we add an extra term in (13). More precisely, we can prove that, with probability larger than  $1 - L_s(\log n)^{(\theta+2)\kappa}/n^{\theta/2}$

$$\|\tilde{s} - s\|_2^2 \leq (1 + \epsilon_n) \inf_{m \in \mathcal{M}_n} \left( \|s - \hat{s}_m\|_2^2 + \text{pen}(m) \right) + L \frac{(\log n)^{\gamma_2}}{n}, \quad (14)$$

where  $L > 0$  and  $\gamma_2 > 0$ .

**Remark :** The main improvement of Theorem 3.1 is that it gives an oracle inequality in probability, with a deviation bound of order  $o(1/n)$  as soon as  $\theta > 2$  instead of  $\theta > 3$  in Comte & Merlevède [13]. Moreover, we do not require  $s$  to be bounded to prove our result.

**Remark:** When the data are independent, the proof of Theorem 3.1 can be used to obtain that the estimator  $\tilde{s}$  chosen with a penalty term of order  $K\Phi D_m/n$  satisfy an oracle inequality as (13). The main difference would be that  $\kappa_1 = 1$ , thus it can be used without a slope heuristic (even if this algorithm can be used also in this context to optimize the constant  $K$ ) and the control of the probability would be  $L_s e^{-\ln(n)^2/C_s}$  for some constants  $L_s, C_s$  instead of  $L_s(\log n)^{(\theta+2)\kappa} n^{-\theta/2}$  in our theorem.

## 4 Results for $\tau$ -mixing sequences

In order to deal with  $\tau$ -mixing sequences, we need to specify the basis  $(\psi_{j,k})_{(j,k) \in \Lambda}$ .

### 4.1 Wavelet basis

Throughout this section,  $r$  is a real number,  $r \geq 1$  and we work with an  $r$ -regular orthonormal multiresolution analysis of  $L_2(\mu)$ , associated with a compactly supported scaling function  $\phi$  and a compactly supported mother wavelet  $\psi$ . Without loss of generality, we suppose that the support of the functions  $\phi$  and  $\psi$  is an interval  $[A_1, A_2]$  where  $A_1$  and  $A_2$  are integers such that  $A_2 - A_1 = A \geq 1$ . Let us recall that  $\phi$  and  $\psi$  generate an orthonormal basis by dilatations and translations.

For all  $k \in \mathbb{Z}$  and  $j \in \mathbb{N}^*$ , let  $\psi_{0,k} : x \rightarrow \sqrt{2}\phi(2x - k)$  and  $\psi_{j,k} : x \rightarrow 2^{j/2}\psi(2^j x - k)$ . The family  $\{(\psi_{j,k})_{j \geq 0, k \in \mathbb{Z}}\}$  is an orthonormal basis of  $L_2(\mu)$ . Let us recall the following inequalities: for all  $p \geq 1$ , let  $K_p = (\sqrt{2}\|\phi\|_p) \vee \|\psi\|_p$ ,  $K_L = (2\sqrt{2}\text{Lip}(\phi)) \vee \text{Lip}(\psi)$ ,  $K_{BV} = AK_L$ . Then for all  $j \geq 0$ , we have  $\|\psi_{j,k}\|_\infty \leq K_\infty 2^{j/2}$ ,

$$\left\| \sum_{k \in \mathbb{Z}} |\psi_{j,k}| \right\|_\infty \leq AK_\infty 2^{j/2} \quad (15)$$

$$\text{Lip}(\psi_{j,k}) \leq K_L 2^{3j/2}, \quad (16)$$

$$\|\psi_{j,k}\|_{BV} \leq K_{BV} 2^{j/2}. \quad (17)$$

We assume that our collection  $(S_m)_{m \in \mathcal{M}_n}$  satisfies the following assumption:

**[W]** dyadic wavelet generated spaces: let  $J_n = \lceil \log(n/2(A+1))/\log(2) \rceil$  and for all  $J_m = 1, \dots, J_n$ , let

$$m = \{(0, k), -A_2 < k < 2 - A_1\} \cup \{(j, k), 1 \leq j \leq J_m, -A_2 < k < -A_1 + 2^j\}$$

and  $S_m$  the linear span of  $\{\psi_{j,k}\}_{(j,k) \in m}$ . In particular, we have  $D_m = (A-1)(J_m+1) + 2^{J_m+1}$  and thus  $2^{J_m+1} \leq D_m \leq (A-1)(J_m+1) + 2^{J_m+1} \leq A2^{J_m+1}$ .

## 4.2 The $\tau$ -mixing case

The following result proves that we keep the same rate of convergence for the PLSE based on  $\tau$ -mixing processes.

**Theorem 4.1** *Consider the collection of models [W]. Assume that  $(X_n)_{n \in \mathbb{Z}}$  is strictly stationary and arithmetically [AR]  $\tau$ -mixing with mixing rate  $\theta > 5$  and that its marginal distribution admits a density  $s$  with respect to the Lebesgue measure  $\mu$ . Let  $\tilde{s}$  be the PLSE defined by (9) with*

$$\text{pen}(m) = KAK_\infty K_{BV} \left( \sum_{l=0}^{\infty} \tilde{\beta}_l \right) \frac{D_m}{n}, \text{ where } K \geq 8.$$

Then there exist constants  $c_0 > 0, \gamma_1 > 0$  and a sequence  $\epsilon_n \rightarrow 0$  such that

$$\mathbb{E} (\|\tilde{s} - s\|_2^2) \leq (1 + \epsilon_n) \left( \inf_{m \in \mathcal{M}_n, D_m \geq c_0(\log n)^{\gamma_1}} \|s - s_m\|_2^2 + \text{pen}(m) \right). \quad (18)$$

**Remark :** As in Theorem 3.1, the penalty term involves an unknown constant and we have a condition on the dimension of the models in (18). However, the slope heuristic can also be used in this context to calibrate the constant and a careful look at the proof shows that we can take the infimum over all models  $m \in \mathcal{M}_n$  provided that we increase the constant  $K$  in front of the penalty term. Our result allows to derive rates of convergence in Besov spaces for the PLSE that correspond to the rates in the i.i.d. framework (see Proposition 5.2).

**Remark :** Theorem 4.1 gives an oracle inequality for the PLSE built on  $\tau$ -mixing sequences. This inequality is not pathwise and the constants involved in the penalty term are not optimal. This is due to technical reasons, mainly because we use the coupling result (4) instead of (3). However, we recover the same kind of oracle inequality as in the i.i.d. framework (Birgé and Massart [8]) under weak assumptions on the mixing coefficients since we only require arithmetical [AR]  $\tau$ -mixing assumptions on the process  $(X_n)_{n \in \mathbb{Z}}$ . This is the first result for these processes up to our knowledge.

Let us mention here Theorem 4.1 in Comte & Merlevède [13]. They consider  $\alpha$ -mixing processes (for a definition of the coefficient  $\alpha$  and its properties, we refer to Rio [22]). They make geometrical [GEO]  $\alpha$ -mixing assumptions on the processes and consider penalties of order  $L \log(n) D_m / n$  to get an oracle inequality. This leads to a logarithmic loss in the rates of convergence. They get the optimal rate under an extra assumption (namely Assumption [Lip] in Section 3.2). There exist random processes that are  $\tau$ -mixing and not  $\alpha$ -mixing (see Dedecker & Prieur [15]), however, the comparison of these coefficients is difficult in general and our method can not be applied in this context.

The constants  $c_0, \gamma_1, n_o$  are given in the end of the proof.

**Remark :** Inequality (2.6) can be improved under stronger assumptions on  $s$ . For example, when  $s$  is bounded, we have  $\hat{\beta}_k \leq C\sqrt{\tau_k}$ . Under this assumption and  $\theta > 3$ , we can prove that the estimator  $\tilde{s}$  satisfies the inequality

$$\mathbb{E} (\|\tilde{s} - s\|_2^2) \leq (1 + \epsilon_n) \left( \inf_{m \in \mathcal{M}_n, D_m \geq c_0(\log n)^{\gamma_1}} \|s - s_m\|_2^2 + \text{pen}(m) \right) + \frac{(\log n)^{\kappa(\theta+1)}}{n^{(\theta-3)/2}}.$$

When  $\theta < 5$ , the extra term  $(\log n)^{\kappa(\theta+1)}/n^{(\theta-3)/2}$  may be larger than the main term  $\inf_{m \in \mathcal{M}_n, D_m \geq c_0(\log n)^{\gamma_1}} \|s - s_m\|_2^2 + \text{pen}(m)$ . In this case, we don't know if our control remains optimal. On the other hand, Proposition 5.2 ensures that  $\tilde{s}$  is adaptive over the class of Besov balls when  $\theta \geq 5$ .

## 5 Minimax results

### 5.1 Approximation results on Besov spaces

#### Besov balls.

Throughout this section,  $\Lambda = \{(j, k), j \in \mathbb{N}, k \in \mathbb{Z}\}$  and  $\{\psi_{j,k}, (j, k) \in \Lambda\}$  denotes an  $r$ -regular wavelet basis as introduced in Section 4.1. Let  $\alpha, p$  be two positive numbers such that  $\alpha + 1/2 - 1/p > 0$ . For all functions  $t \in L_2(\mu)$ ,  $t = \sum_{(j,k) \in \Lambda} t_{j,k} \psi_{j,k}$ , we say that  $t$  belongs to the Besov ball  $B_{\alpha,p,\infty}(M_1)$  on the real line if  $\|t\|_{\alpha,p,\infty} \leq M_1$  where

$$\|t\|_{\alpha,p,\infty} = \sup_{j \in \mathbb{N}} 2^{j(\alpha+1/2-1/p)} \left( \sum_{k \in \mathbb{Z}} |t_{j,k}|^p \right)^{1/p}.$$

It is easy to check that if  $p \geq 2$   $B_{\alpha,p,\infty}(M_1) \subset B_{\alpha,2,\infty}(M_1)$  so that upper bounds on  $B_{\alpha,2,\infty}(M_1)$  yield upper bounds on  $B_{\alpha,p,\infty}(M_1)$ .

#### Approximation results on Besov spaces.

We have the following result (Birgé & Massart [8] Section 4.7.1). Suppose that the support of  $s$  equals  $[0, 1]$  and that  $s$  belongs to the Besov ball  $B_{\alpha,2,\infty}(1)$ , then whenever  $r > \alpha - 1$ ,

$$\|s - s_m\|_2^2 \leq \frac{\|s\|_{\alpha,2,\infty}^2}{4(4^\alpha - 1)} 2^{-2J_m \alpha} \leq \frac{(2A)^{2\alpha} \|s\|_{\alpha,2,\infty}^2}{4(4^\alpha - 1)} D_m^{-2\alpha} \quad (19)$$

### 5.2 Minimax rates of convergence for the PLSE

We can derive from Theorems 3.1 and 4.1 adaptation results to unknown smoothness over Besov Balls.

**Proposition 5.1** *Assume that the process  $(X_n)_{n \in \mathbb{Z}}$  is strictly stationary and arithmetically [AR]  $\beta$ -mixing with mixing rate  $\theta > 2$  and that its marginal distribution admits a density  $s$  with respect to the Lebesgue measure  $\mu$ , that  $s$  is supported in  $[0, 1]$  and that  $s \in L^2(\mu)$ . For all  $\alpha, M_1 > 0$ , the PLSE  $\tilde{s}$  defined in Theorem 3.1 for the collection of models [W] satisfies*

$$\forall \kappa > 2, \quad \sup_{s \in B_{\alpha,2,\infty}(M_1)} \mathbb{P} \left( \|\tilde{s} - s\|_2^2 > L_{M_1,\alpha,\theta} n^{-2\alpha/(2\alpha+1)} \right) \leq \frac{L_{M_1} (\log n)^{(\theta+2)\kappa}}{n^{\theta/2}}.$$

**Proposition 5.2** *Assume that the process  $(X_n)_{n \in \mathbb{Z}}$  is strictly stationary and arithmetically [AR]  $\tau$ -mixing with mixing rate  $\theta > 5$  and that its marginal distribution admits a density*

$s$  with respect to the Lebesgue measure  $\mu$ , that  $s$  is supported in  $[0, 1]$  and that  $s \in L^2(\mu)$ . For all  $\alpha, M_1 > 0$ , the PLSE  $\tilde{s}$  defined in Theorem 4.1 satisfies

$$\sup_{s \in B_{\alpha, 2, \infty}(M_1)} \mathbb{E} \left( \|\tilde{s} - s\|_2^2 \right) \leq L_{M_1, \alpha, \theta} n^{-2\alpha/(2\alpha+1)}.$$

**Remark:** Proposition 5.2 can be compared to Theorem 3.1 in Gannaz & Wintenberger [18]. They prove near minimax results for the thresholded wavelet estimator introduced by Donoho *et al.* [16] in a  $\tilde{\phi}$ -dependent setting (for a definition of the coefficient  $\tilde{\phi}$ , we refer to Dedecker & Prieur [15]). Basically, with our notations, their result can be stated as follows: if  $(X_n)_{n \in \mathbb{Z}}$  is  $\tilde{\phi}$ -mixing with  $\tilde{\phi}_1(r) \leq Ce^{-ar^b}$  for some constants  $C, a, b$ , then the thresholded wavelet estimator  $\hat{s}$  of  $s$  satisfies

$$\forall \alpha > 0, \forall p > 1, \sup_{s \in B_{\alpha, p, \infty}(M_1) \cap L^\infty(M)} \mathbb{E} \left( \|\hat{s} - s\|_2^2 \right) \leq L_{M, M_1, \alpha, p} \left( \frac{\log n}{n} \right)^{2\alpha/(2\alpha+1)}.$$

The main advantage of their result is that they can deal with Besov balls with regularity  $1 < p < 2$ . However, in the regular case, when  $p \geq 2$ , we have been able to remove the extra  $\log n$  factor. Moreover, our result only requires arithmetical [AR] rates of convergence for the mixing coefficients and we do not have to suppose that  $s$  is bounded.

## 6 Proofs.

### 6.1 Proofs of the minimax results.

**Proof of Proposition 5.1:**

Let  $\alpha > 0$  and  $M_1 > 0$  and assume that  $s \in B_{\alpha, 2, \infty}(M_1)$ . Let  $\tilde{\mathcal{M}}_n = \{m \in \mathcal{M}_n, D_m > c_0(\log n)^{\gamma_1}\}$ . By Theorem 3.1, there exists a constant  $L_\theta > 0$  such that

$$\mathbb{P} \left( \|\tilde{s} - s\|_2^2 > L_\theta \inf_{m \in \tilde{\mathcal{M}}_n} \left\{ \|s - s_m\|_2^2 + \frac{D_m}{n} \right\} \right) \leq \frac{L_s (\log n)^{(\theta+2)\kappa}}{n^{\theta/2}}. \quad (20)$$

It appears from the proof of Theorem 3.1 that the constant  $L_s$  depends only on  $\|s\|_2$  and that it is a nondecreasing function of  $\|s\|_2$  so that  $L_s$  can be uniformly bounded over  $B_{\alpha, 2, \infty}(M_1)$  by a constant  $L_{M_1}$  so that, by (20)

$$\mathbb{P} \left( \|\tilde{s} - s\|_2^2 > L_\theta \inf_{m \in \tilde{\mathcal{M}}_n} \left\{ \|s - s_m\|_2^2 + \frac{D_m}{n} \right\} \right) \leq \frac{L_{M_1} (\log n)^{(\theta+2)\kappa}}{n^{\theta/2}}.$$

In particular, for a model  $m$  in  $\mathcal{M}_n$  with dimension  $D_m$  such that

$$c_0(\log n)^{\gamma_1} \leq L_1 n^{1/(2\alpha+1)} \leq D_m \leq L_2 n^{1/(2\alpha+1)},$$

we have

$$\mathbb{P} \left( \|\tilde{s} - s\|_2^2 > L_\theta \left( \|s - s_m\|_2^2 + \frac{D_m}{n} \right) \right) \leq \frac{L_{M_1} (\log n)^{(\theta+2)\kappa}}{n^{\theta/2}}.$$

Since  $s$  belongs to  $B_{\alpha, 2, \infty}(M_1)$ , we can use Inequality (19) to get

$$\|s - s_m\|_2^2 \leq L_{\alpha, M_1} D_m^{-2\alpha}.$$

Thus we obtain

$$\mathbb{P} \left( \|\tilde{s} - s\|_2^2 > L_{M_1, \alpha, \theta} n^{-2\alpha/(2\alpha+1)} \right) \leq \frac{L_{M_1} (\log n)^{(\theta+2)\kappa}}{n^{\theta/2}}. \quad \square$$

**Proof of Proposition 5.2:**

Let  $\alpha > 0$  and  $M_1 > 0$  and assume that  $s \in B_{\alpha,2,\infty}(M_1)$ . By Theorem 4.1, we have

$$\mathbb{E} \left( \|\tilde{s} - s\|_2^2 \right) \leq L_\theta \left( \inf_{m \in \mathcal{M}_n} \left\{ \|s - s_m\|_2^2 + \frac{D_m}{n} \right\} \right).$$

Inequality (19) leads to  $\|s - s_m\|_2^2 \leq L_{\alpha, M_1} D_m^{-2\alpha}$ , so that for a model  $m$  in  $\tilde{\mathcal{M}}_n$  with dimension  $D_m$  such that

$$c_0(\log n)^{\gamma_1} \leq L_1 n^{1/(2\alpha+1)} \leq D_m \leq L_2 n^{1/(2\alpha+1)},$$

we find

$$\mathbb{E} \left( \|\tilde{s} - s\|_2^2 \right) \leq L_{\theta, \alpha, M_1} n^{-2\alpha/(2\alpha+1)}. \square$$

## 6.2 Proof of Theorem 3.1:

For all  $m_o$  in  $\mathcal{M}_n$ , we have, by definition of  $\hat{m}$

$$\begin{aligned} \gamma_n(\tilde{s}) + \text{pen}(\hat{m}) &\leq \gamma_n(\hat{s}_{m_o}) + \text{pen}(m_o) \\ P\gamma(\tilde{s}) + \nu_n \gamma(\tilde{s}) + \text{pen}(\hat{m}) &\leq P\gamma(\hat{s}_{m_o}) + \nu_n \gamma(\hat{s}_{m_o}) + \text{pen}(m_o) \\ P\gamma(\tilde{s}) - P\gamma(s) - 2\nu_n \tilde{s} + \text{pen}(\hat{m}) &\leq P\gamma(\hat{s}_{m_o}) - P\gamma(s) - 2\nu_n \hat{s}_{m_o} + \text{pen}(m_o) \end{aligned}$$

Since for all  $t \in L_2(\mu)$ ,  $P\gamma(t) - P\gamma(s) = \|t - s\|_2^2$ , we have

$$\|s - \tilde{s}\|_2^2 \leq \|s - \hat{s}_{m_o}\|_2^2 + \text{pen}(m_o) - V(m_o) - (\text{pen}(\hat{m}) - V(\hat{m})) - 2\nu_n(s_{m_o} - \hat{s}_{m_o}), \quad (21)$$

where, for all  $m \in \mathcal{M}_n$

$$V(m) = 2\nu_n(\hat{s}_m - s_m) = 2 \sum_{(j,k) \in m} \nu_n^2(\psi_{j,k}).$$

This decomposition is different from the one used in Birgé & Massart [8] and in Comte & Merlevède [13]. It allows to improve the constant in the oracle inequality in the  $\beta$ -mixing case. Moreover, we choose to prove an oracle inequality of the form (12) for  $\beta$ -mixing sequences, which allows to assume only  $\theta > 2$  instead of  $\theta > 3$ . Let us now give a sketch of the proof:

1. we build an event  $\Omega_C$  with  $\mathbb{P}(\Omega_C^c) \leq p\beta_q$  such that, on  $\Omega_C$ ,  $\nu_n = \nu_n^*$ , where  $\nu_n^*$  is built with independent data. A suitable choice of the integers  $p$  and  $q$  leads to  $p\beta_q \leq C(\ln n)^r n^{-\theta/2}$ .
2. We use the concentration's inequality (7.4) of Birgé & Massart [8] for  $\chi^2$ -type statistics, derived from Talagrand's inequality. This allows us to find  $p_1(m)$  such that on an event  $\Omega_1$  with  $\mathbb{P}(\Omega_1^c \cap \Omega_C) \leq L_{1,s} c_n$

$$\sup_{m \in \mathcal{M}_n} \{V(m) - p_1(m)\} \leq 0.$$

$c_n < C(\ln n)^r n^{-\theta/2}$  and  $L_{1,s}$  is some constant depending on  $s$ .

3. From Bernstein's inequality, we prove that, for all  $m, m' \in \mathcal{M}_n$ , there exists  $p_2(m, m')$  such that, for all  $\eta > 0$ , on an event  $\Omega_2$  with  $\mathbb{P}(\Omega_2^c \cap \Omega_C) \leq L_{2,s} c_n$ ,

$$\sup_{m, m' \in \mathcal{M}_n} \left\{ \nu_n(s_m - s_{m'}) - \frac{\eta}{2} p_2(m, m') - \frac{\|s_m - s_{m'}\|_2^2}{2\eta} \right\} \leq 0.$$

Moreover, for all  $m, m' \in \mathcal{M}_n$ ,  $p_2(m, m') \leq p_2(m, m) + p_2(m', m')$ .

4. We have  $\|s_{\hat{m}} - s_{m_o}\|_2^2 \leq \|s_{\hat{m}} - s\|_2^2 + \|s - s_{m_o}\|_2^2$  because  $s_{\hat{m}} - s_{m_o}$  is either the projection of  $s_{\hat{m}} - s$  onto  $S_{m_o}$  or the projection of  $s - s_{m_o}$  onto  $S_{\hat{m}}$ . Take  $\text{pen}(m) \geq p_1(m) + \eta p_2(m, m)$ , we have, on  $\Omega_1 \cap \Omega_2 \cap \Omega_C$

$$\|s - \tilde{s}\|_2^2 \leq \|s - \hat{s}_{m_o}\|_2^2 - \frac{V_{m_o}}{2} + \text{pen}(m_o) - \frac{V_{m_o}}{2} \quad (22)$$

$$\begin{aligned} & -(\text{pen}(\hat{m}) - p_1(\hat{m})) - (p_1(\hat{m}) - V(\hat{m})) - 2\nu_n(s_{m_o} - s_{\hat{m}}) \\ & \leq \|s - s_{m_o}\|_2^2 + \text{pen}(m_o) - \frac{V(m_o)}{2} - \eta p_2(\hat{m}, \hat{m}) \\ & \quad + \eta p_2(\hat{m}, m_o) + \frac{\|s_{m_o} - s_{\hat{m}}\|_2^2}{\eta} \end{aligned} \quad (23)$$

$$\left(1 - \frac{1}{\eta}\right) \|s - \tilde{s}\|_2^2 \leq \left(1 + \frac{1}{\eta}\right) \|s - s_{m_o}\|_2^2 + \text{pen}(m_o) + \eta p_2(m_o, m_o). \quad (24)$$

In (23), we used that  $V(m_o) = 2\|s_{m_o} - \hat{s}_{m_o}\|_2^2 \geq 0$ . In (24), we used that  $V_{m_o} \geq 0$ . Pythagoras Theorem gives

$$\|s - \hat{s}_{m_o}\|_2^2 - \frac{V(m_o)}{2} = \|s - s_{m_o}\|_2^2 \text{ and } \|s - s_{\hat{m}}\|_2^2 \leq \|s - \tilde{s}\|_2^2.$$

Finally, we prove that we can choose  $\eta = (\log n)^\gamma$ , with  $\gamma > 0$  such that  $\eta p_2(m_o, m_o) = o(\text{pen}(m_o))$  and we conclude the proof of (3.1) from the previous inequalities.

We decompose the proof in several claims corresponding to the previous steps.

**Claim 1 :** For all  $l = 0, \dots, p-1$ , let us define  $A_l = (X_{2lq+1}, \dots, X_{(2l+1)q})$  and  $B_l = (X_{(2l+1)q+1}, \dots, X_{(2l+2)q})$ . There exist random vectors  $A_l^* = (X_{2lq+1}^*, \dots, X_{(2l+1)q}^*)$  and  $B_l^* = (X_{(2l+1)q+1}^*, \dots, X_{(2l+2)q}^*)$  such that for all  $l = 0, \dots, p-1$  :

1.  $A_l^*$  and  $A_l$  have the same law,
2.  $A_l^*$  is independent of  $A_0, \dots, A_{l-1}, A_0^*, \dots, A_{l-1}^*$
3.  $\mathbb{P}(A_l \neq A_l^*) \leq \beta_q$

the same being true for the variables  $B_l$ .

**Proof of Claim 1 :**

The proof is derived from Berbee's lemma, we refer to Proposition 5.1 in Viennet [25] for further details about this construction.  $\square$

Hereafter, we assume that, for some  $\kappa > 2$ ,  $\sqrt{n}(\log n)^\kappa/2 \leq p \leq \sqrt{n}(\log n)^\kappa$  and for the sake of simplicity that  $pq = n/2$ , the modifications needed to handle the extra term when  $q = \lceil n/(2p) \rceil$  being straightforward. Let  $\Omega_C = \{\forall l = 0, \dots, p-1 A_l = A_l^*, B_l = B_l^*\}$ . We have

$$\mathbb{P}(\Omega_C^c) \leq 2p\beta_q \leq 2^{2+\theta} \frac{(\log n)^{(\theta+2)\kappa}}{n^{\theta/2}}.$$

Let us first deal with the quadratic term  $V(m)$ .

**Claim 2 :** Under the assumptions of Theorem 3.1, let  $\epsilon > 0$ ,  $1 < \gamma < \kappa/2$ . We define  $L_1^2 = 2\Phi^2\kappa_1$ ,  $L_2^2 = 8\Phi^{3/2}\sqrt{\kappa_2}$ ,  $L_3 = 2\Phi\kappa(\epsilon)$  and

$$L_{1,m} = 4 \left( (1 + \epsilon)L_1 + L_2 \sqrt{\frac{(\log n)^\gamma}{D_m^{1/4}}} + \frac{L_3}{(\log n)^{\kappa-\gamma}} \right)^2. \quad (25)$$

Then, we have

$$\mathbb{P} \left( \sup_{m \in \mathcal{M}_n} \left\{ V(m) - \frac{L_{1,m} D_m}{n} \right\} \geq 0 \cap \Omega_C \right) \leq L_{s,\gamma} \exp \left( -\frac{(\log n)^\gamma}{\sqrt{\|s\|_2}} \right).$$

where  $L_{s,\gamma} = 2 \sum_{D=1}^{\infty} \exp(-(\log D)^\gamma / \|s\|_2^{1/2})$ . In particular, for all  $r > 0$ , there exists a constant  $L'_{s,r}$  depending on  $\|s\|_2$ , such that

$$\mathbb{P} \left( \sup_{m \in \mathcal{M}_n} \left\{ V(m) - \frac{L_{1,m} D_m}{n} \right\} \geq 0 \cap \Omega_C \right) \leq \frac{L'_{s,r}}{n^r}.$$

**Remark :** When  $(L_2/L_1)^8 (\log n)^{4(2\kappa-\gamma)} \leq D_m \leq n$ , we have

$$L_{1,m} \leq \left[ 1 + \epsilon + \left( 1 + \frac{\sqrt{2}\kappa(\epsilon)}{\sqrt{\kappa_1}} \right) (\log n)^{-(\kappa-\gamma)} \right]^2 4L_1^2.$$

**Proof of Claim 2 :**

Let  $P_n^*(t) = \sum_{i=1}^n t(X_i^*)/n$  and  $\nu_n^*(t) = (P_n^* - P)t$ , we have

$$V(m)\mathbf{1}_{\Omega_C} = 2 \sum_{(j,k) \in m} (\nu_n^*)^2(\psi_{j,k})\mathbf{1}_{\Omega_C}.$$

Let  $B_1(S_m) = \{t \in S_m; \|t\|_2 \leq 1\}$ .  $\forall t \in B_1(S_m)$ , let  $\bar{t}(x_1, \dots, x_q) = \sum_{i=1}^q t(x_i)/2q$  and for all functions  $g : \mathbb{R}^q \rightarrow \mathbb{R}$  let

$$P_{A,p}^* g = \frac{1}{p} \sum_{j=0}^{p-1} g(A_j^*), \quad P_{B,p}^* g = \frac{1}{p} \sum_{j=0}^{p-1} g(B_j^*), \quad \bar{P}g = \int g \mathbb{P}_A(d\mu),$$

$$\text{and } \bar{\nu}_{A,p} g = (P_{A,p}^* - \bar{P})g, \quad \bar{\nu}_{B,p} g = (P_{B,p}^* - \bar{P})g.$$

Now we have

$$\sum_{(j,k) \in m} (\nu_n^*)^2(\psi_{j,k}) \leq 2 \sum_{(j,k) \in m} \bar{\nu}_{A,p}^2 \bar{\psi}_{j,k} + 2 \sum_{(j,k) \in m} \bar{\nu}_{B,p}^2 \bar{\psi}_{j,k}.$$

In order to handle these terms, we use Proposition 7.4 which is stated in Section 7. Taking

$$B_m^2 = \sum_{(j,k) \in m} \text{Var}(\bar{\psi}_{j,k}(A_1)), \quad V_m^2 = \sup_{t \in B_1(S_m)} \text{Var}(\bar{t}(A_1)), \quad \text{and } H_m^2 = \left\| \sum_{(j,k) \in m} (\bar{\psi}_{j,k})^2 \right\|_\infty,$$

we have

$$\forall x > 0, \quad \mathbb{P} \left( \sqrt{\sum_{(j,k) \in m} \bar{\nu}_{A,p}^2 \bar{\psi}_{j,k}} \geq \frac{(1+\epsilon)}{\sqrt{p}} B_m + V_m \sqrt{\frac{2x}{p}} + \kappa(\epsilon) \frac{H_m x}{p} \right) \leq e^{-x}. \quad (26)$$

In order to evaluate  $B_m$ ,  $V_m$  and  $H_m$ , we use Viennet's inequality (54). There exists a function  $b$  such that, for all  $p = 1, 2$ ,  $P|b|^p \leq \kappa_p$  where  $\kappa_p$  is defined in (2) and for all functions  $t \in L_2(\bar{P})$ ,

$$\text{Var}(\bar{t}(A_1)) \leq \frac{1}{q} P b t^2.$$

Thus

$$B_m^2 = \sum_{(j,k) \in m} \text{Var}(\bar{\psi}_{j,k}(A_1)) \leq \frac{1}{q} \sum_{(j,k) \in m} P b \psi_{j,k}^2 \leq \left\| \sum_{(j,k) \in m} \psi_{j,k}^2 \right\|_{\infty} \frac{\kappa_1}{q}.$$

From Assumption  $[M_2]$ ,  $\left\| \sum_{(j,k) \in m} \psi_{j,k}^2 \right\|_{\infty} \leq \Phi^2 D_m$ , thus,

$$B_m^2 \leq \frac{\Phi^2 \kappa_1 D_m}{q}. \quad (27)$$

From Viennet's and Cauchy-Schwarz inequalities

$$V_m^2 = \sup_{t \in B_1(S_m)} \text{Var}(\bar{t}(A_1)) \leq \sup_{t \in B_1(S_m)} \frac{P b t^2}{q} \leq \sup_{t \in B_1(S_m)} \|t\|_{\infty} \frac{(P t^2)^{1/2} (P b^2)^{1/2}}{q}.$$

Since  $t \in B_1(S_m)$ , we have by Cauchy-Schwarz inequality

$$(P t^2)^{1/2} \leq (\|t\|_{\infty} \|t\|_2 \|s\|_2)^{1/2} \leq (\|t\|_{\infty} \|s\|_2)^{1/2}.$$

From Assumption  $[M_2]$ , we have  $\|t\|_{\infty} \leq \Phi \sqrt{D_m}$ , and from Viennet's inequality  $P b^2 \leq \kappa_2 < \infty$ , thus we obtain

$$V_m^2 \leq \Phi^{3/2} (\|s\|_2 \kappa_2)^{1/2} \frac{D_m^{3/4}}{q}. \quad (28)$$

Finally, from Assumption  $[M_2]$ , we have, using Cauchy-Schwarz inequality

$$H_m^2 = \left\| \sum_{(j,k) \in m} \bar{\psi}_{j,k}^2 \right\|_{\infty} \leq \frac{1}{4} \left\| \sum_{(j,k) \in m} \psi_{j,k}^2 \right\|_{\infty} \leq \frac{\Phi^2 D_m}{4}. \quad (29)$$

Let  $y_n > 0$ . We define

$$L_m = \left( (1 + \epsilon) L_1 + L_2 \sqrt{\frac{(\log D_m)^{\gamma} + y_n}{2 D_m^{1/4}}} + L_3 \frac{(\log D_m)^{\gamma} + y_n}{2 (\log n)^{\kappa}} \right)^2.$$

We apply Inequality (26) with  $x = ((\log D_m)^{\gamma} + y_n) / \|s\|_2^{1/2}$  and the evaluations (27), (28) and (29). Recalling that  $1/p \leq 2/(\sqrt{n}(\log n)^{\kappa})$ , this leads to

$$\mathbb{P} \left( \sum_{(j,k) \in m} \bar{v}_{A,p}^2 \bar{\psi}_{j,k} \geq \frac{L_m D_m}{n} \right) \leq \exp \left( -\frac{(\log D_m)^{\gamma}}{\sqrt{\|s\|_2}} \right) \exp \left( -\frac{y_n}{\sqrt{\|s\|_2}} \right).$$

In order to give an upper bound on  $H_m x$ , we used that the support of  $s$  is included in  $[0, 1]$ , thus

$$1 = \|s\|_1 \leq \|s\|_2.$$

The result follows by taking  $y_n = (\log n)^{\gamma} \geq (\log D_m)^{\gamma}$ .  $\square$

**Claim 3.** *We keep the notations  $\kappa/2 > \gamma > 1$ ,  $L_2$  of the proof of Claim 2. For all  $m, m' \in \mathcal{M}_n$  we take*

$$L_{m,m'} = 4 \left( L_2 \sqrt{\frac{(\log n)^{\gamma}}{(D_m \vee D_{m'})^{1/4}}} + \frac{4\Phi}{3(\log n)^{\kappa-\gamma}} \right)^2, \quad (30)$$



we have, for all  $\eta > 0$ ,

$$\mathbb{P} \left( \sup_{m, m' \in \mathcal{M}_n} \nu_n^*(s_m - s_{m'}) - \frac{\|s_m - s_{m'}\|_2^2}{2\eta} - \frac{\eta L_{m, m'}(D_m \vee D_{m'})}{2n} > 0 \right) \leq L_{s, \gamma} e^{-\frac{(\log n)^\gamma}{\|s\|_2^{1/2}}}$$

with  $L_{s, \gamma} = 2 \sum_{m, m' \in \mathcal{M}_n} e^{-\frac{(\log(D_m \vee D_{m'}))^\gamma}{\|s\|_2^{1/2}}}$ .

**Remark :** The constant  $L_{s, \gamma}$  is finite since for all  $x, y > 0$ ,  $(\log(x \vee y))^\gamma \geq ((\log x)^\gamma + (\log y)^\gamma)/2$ .

As in Claim 2, when  $(L_2/L_1)^8(\log n)^{4(2\kappa-\gamma)} \leq D_m \leq n$ , we have

$$L_{m, m'} \leq \left( 1 + \frac{2^{3/2}}{3\sqrt{\kappa_1}} \right)^2 (\log n)^{-2(\kappa-2\gamma)} 4L_1^2.$$

**Proof of Claim 3.**

We keep the notations of the proof of Claim 2 and for  $m, m' \in \mathcal{M}_n$ , let  $t_{m, m'} = (s_m - s_{m'})/\|s_m - s_{m'}\|_2$ . We use the inequality  $2ab \leq a^2\eta^{-1} + b^2\eta$ , which holds for all  $a, b \in \mathbb{R}$ ,  $\eta > 0$ . This leads to

$$\begin{aligned} \nu_n^*(s_m - s_{m'}) &= \|s_m - s_{m'}\|_2 \nu_n^*(t_{m, m'}) \leq \frac{\|s_m - s_{m'}\|_2^2}{2\eta} + \frac{\eta}{2} (\nu_n^*(t_{m, m'}))^2 \\ &= \frac{\|s_m - s_{m'}\|_2^2}{2\eta} + \frac{\eta}{2} (\bar{\nu}_{A, p}(\bar{t}_{m, m'}) + \bar{\nu}_{B, p}(\bar{t}_{m, m'}))^2 \\ &\leq \frac{\|s_m - s_{m'}\|_2^2}{2\eta} + \eta(\bar{\nu}_{A, p}(\bar{t}_{m, m'}))^2 + \eta(\bar{\nu}_{B, p}(\bar{t}_{m, m'}))^2. \end{aligned}$$

Now from Bernstein's inequality (see Section 7), we have

$$\forall x > 0, \mathbb{P} \left( \bar{\nu}_{A, p}(\bar{t}_{m, m'}) > \sqrt{\frac{2\text{Var}(\bar{t}_{m, m'}(A_1))x}{p}} + \frac{\|\bar{t}_{m, m'}\|_\infty x}{3p} \right) \leq e^{-x}. \quad (31)$$

From Viennet's and Cauchy-Schwarz inequalities, we have

$$\text{Var}(\bar{t}_{m, m'}(A_1)) \leq \frac{Pbt_{m, m'}^2}{q} \leq \frac{\|t_{m, m'}\|_\infty \sqrt{Pb^2 Pt_{m, m'}^2}}{q}.$$

Moreover

$$Pb^2 \leq \kappa_2, \quad Pt_{m, m'}^2 \leq \|t_{m, m'}\|_\infty \|t_{m, m'}\|_2 \|s\|_2.$$

Since  $t_{m, m'} \in S_m \cup S_{m'}$  and  $\|t_{m, m'}\|_2 = 1$ , we have, from Assumption  $[M_2]$   $\|t_{m, m'}\|_\infty \leq \Phi\sqrt{D_m \vee D_{m'}}$ . Let  $y_n > 0$ . We apply Inequality (31) with  $x = [(\log(D_m \vee D_{m'}))^\gamma + y_n]/\|s\|_2^{1/2}$ . We define

$$\frac{L'_{m, m'}}{4} = \left( L_2 \sqrt{\frac{(\log(D_m \vee D_{m'}))^\gamma + y_n}{2(D_m \vee D_{m'})^{1/4}}} + \frac{4\Phi [(\log(D_m \vee D_{m'}))^\gamma + y_n]}{6(\log n)^\kappa} \right)^2,$$

we have

$$\mathbb{P} \left( \bar{\nu}_{A, p}(\bar{t}_{m, m'}) > \sqrt{\frac{L'_{m, m'}(D_m \vee D_{m'})}{4n}} \right) \leq \exp \left( -\frac{(\log(D_m \vee D_{m'}))^\gamma}{\|s\|_2^{1/2}} \right) e^{-y_n/\|s\|_2^{1/2}}.$$

The result follows by taking  $y_n = (\log n)^\gamma$  and using  $2 \leq D_m \leq n$ .

**Conclusion of the proof:**

Let  $\eta > 0$  and  $\text{pen}'(m) \geq (L_{1,m} + \eta L_{m,m})D_m/n$  where  $L_{1,m}$  and  $L_{m,m}$  are defined respectively by (25) and (30). From Claims 1, 2 and 3 and (24), we obtain that, for all  $m_o$  and with probability larger than  $L_{s,\theta}(\log n)^{(\theta+2)\kappa}n^{-\theta/2}$

$$\left(1 - \frac{1}{\eta}\right) \|s - \tilde{s}\|_2^2 \leq \left(1 + \frac{1}{\eta}\right) \|s - s_{m_o}\|_2^2 + \text{pen}'(m_o) + \eta L(m_o, m_o) \frac{D_{m_o}}{n}. \quad (32)$$

Assume that  $D_m \geq (L_2/L_1)^8(\log n)^{4(2\kappa-\gamma)}$ , then we have from remarks 6.2 and 6.2

$$\begin{aligned} L_{1,m} &\leq \left[1 + \epsilon + \left(1 + \frac{2\kappa(\epsilon)}{\sqrt{\kappa_1}}\right) (\log n)^{-(\kappa-2\gamma)}\right]^2 4L_1^2 \text{ and} \\ L_{m,m} &\leq \left(1 + \frac{2^{3/2}}{3\sqrt{\kappa_1}}\right)^2 (\log n)^{-2(\kappa-\gamma)} 4L_1^2. \end{aligned}$$

Take  $\eta = (\log n)^{\kappa-\gamma}$ , we have  $(L_{1,m_o} + \eta L_{m_o,m_o})D_{m_o}/n \leq C\text{pen}(m_o)$ . Fix  $\epsilon > 0$  such that  $[1 + \epsilon]^2 < K/4$ . Since  $\kappa > \gamma$ , for  $n \geq n_o$ , we have  $L_{1,m} + \eta L_{m,m} \leq KL_1^2$ , thus, inequality (13) follows from (32) as soon as  $n > n_o$ . We remove the condition  $n > n_o$  by improving the constant  $L_s$  in (13) if necessary.  $\square$

### 6.3 Proof of Theorem 4.1.

The proof follows the previous one, the main difference is that the coupling lemma (Claim 1) as well as the covariance inequalities are much harder to handle in the  $\tau$ -mixing case. This leads to more technical computations to recover the results obtained in the  $\beta$ -mixing case (see Claims 2, 3 and the proof of inequality (45)). We start with the decomposition (21). As in the previous proof, the decomposition of the risk given in Birgé & Massart [8] or in Comte & Merlevède [13] could be used. This leads to a loss in the constant in front of the main term in (18) without avoiding any of the main difficulties. We divide the proof in four claims.

**Claim 1 :** For all  $l = 0, \dots, p-1$ , let us denote by  $A_l = (X_{2lq+1}, \dots, X_{(2l+1)q})$  and  $B_l = (X_{(2l+1)q+1}, \dots, X_{(2l+2)q})$ . There exist random vectors  $A_l^* = (X_{2lq+1}^*, \dots, X_{(2l+1)q}^*)$  and  $B_l^* = (X_{(2l+1)q+1}^*, \dots, X_{(2l+2)q}^*)$  such that for all  $l = 0, \dots, p-1$  :

- $A_l^*$  and  $A_l$  have the same law,
- $A_l^*$  is independent of  $A_0, \dots, A_{l-1}, A_0^*, \dots, A_{l-1}^*$
- $\mathbb{E}(|A_l - A_l^*|_q) \leq q\tau_q$

the same being true for the variables  $B_l$ .

**Proof of Claim 1 :**

We use the same recursive construction as Viennet [25].

Let  $(\delta_j)_{0 \leq j \leq p-1}$  be a sequence of independent random variables uniformly distributed over  $[0, 1]$  and independent of the sequence  $(A_j)_{0 \leq j \leq p-1}$ . Let  $A_0^* = (X_1^*, \dots, X_q^*)$  be the random variable given by equality (4) for  $\mathcal{M} = \sigma(X_i, i \leq -q)$ ,  $A_0$  and  $\delta_0$ .

Now suppose that we have built the variables  $A_l^*$  for  $l < l'$ . From equality (4) applied to the  $\sigma$ -algebra  $\sigma(A_l, A_l^*, l < l')$ ,  $A_{l'}$  and  $\delta_{l'}$ , there exists a random variable  $A_{l'}^*$  satisfying the hypotheses of Claim 1.

We build in the same way the variables  $B_l^*$  for all  $l = 0, \dots, p-1$ .  $\square$

We keep the notations  $\nu_n^*$ ,  $\bar{\nu}_{A,p}$ ,  $\bar{\nu}_{B,p}$ ,  $\bar{t}$  and  $B_1(S_m)$  that we introduced in the proof of Theorem 3.1. As in the proof of Theorem 3.1, we assume that, for some  $\kappa > 2$ ,  $\sqrt{n}(\log n)^\kappa/2 \leq p \leq \sqrt{n}(\log n)^\kappa$  and for the sake of simplicity that  $pq = n/2$ , the modifications needed to handle the extra term when  $q = \lceil n/(2p) \rceil$  being straightforward. We have

$$V(\hat{m}) = \sum_{(j,k) \in \hat{m}} \nu_n^2(\psi_{j,k}) \leq 2 \sum_{(j,k) \in \hat{m}} (P_n - P_n^*)^2(\psi_{j,k}) + 2 \sum_{(j,k) \in \hat{m}} (\nu_n^*)^2(\psi_{j,k}) \quad (33)$$

**Claim 2 :** *There exists a constant  $L = L_{A,K_L,K_\infty,\kappa,\theta}$  such that*

$$\mathbb{E} \left( \sum_{j,k \in \hat{m}} ((P_n - P_n^*)(\psi_{j,k}))^2 \right) \leq L \frac{(\log n)^{\kappa(\theta+1)}}{n^{(\theta-3)/2}}. \quad (34)$$

*Proof of Claim 2 :*

$$\begin{aligned} \mathbb{E} \left( \sum_{(j,k) \in \hat{m}} (P_n - P_n^*)^2(\psi_{j,k}) \right) &\leq \mathbb{E} \left( \sup_{m \in \mathcal{M}_n} \sum_{(j,k) \in m} (P_n - P_n^*)^2(\psi_{j,k}) \right) \\ &\leq \sum_{m \in \mathcal{M}_n} \sum_{(j,k) \in m} \mathbb{E} ((P_n - P_n^*)^2(\psi_{j,k})) \\ &\leq \frac{2}{p^2} \sum_{m \in \mathcal{M}_n} \sum_{l,l'=1}^p (g_{A,m}(j,k,l,l') + g_{B,m}(j,k,l,l')) \end{aligned}$$

with

$$g_{m,A}(j,k,l,l') = \mathbb{E} \left( \sum_{(j,k) \in m} (\bar{\psi}_{j,k}(A_l) - \bar{\psi}_{j,k}(A_l^*)) (\bar{\psi}_{j,k}(A_{l'}) - \bar{\psi}_{j,k}(A_{l'}^*)) \right).$$

We develop this last term and we get, since

$$|\bar{\psi}_{j,k}(x) - \bar{\psi}_{j,k}(y)| \leq \frac{K_L 2^{3j/2} |x - y|_q}{2q}$$

$$\begin{aligned} g_{A,m}(j,k,l,l') &\leq \mathbb{E} \left( \sum_{(j,k) \in m} |\bar{\psi}_{j,k}(A_l) - \bar{\psi}_{j,k}(A_l^*)| |\bar{\psi}_{j,k}(A_{l'}) - \bar{\psi}_{j,k}(A_{l'}^*)| \right) \\ &\leq \mathbb{E} \left( \sum_{(j,k) \in m} |\bar{\psi}_{j,k}(A_l) - \bar{\psi}_{j,k}(A_l^*)| K_L 2^{3j/2} \frac{|A_{l'} - A_{l'}^*|_q}{2q} \right) \\ &\leq \frac{K_L \tau_q}{2} \sup_{x,y \in \mathbb{R}^q} \left\{ \sum_{(j,k) \in m} 2^{3j/2} |\bar{\psi}_{j,k}(x) - \bar{\psi}_{j,k}(y)| \right\} \\ &\leq \frac{K_L \tau_q}{4} \sum_{j=0}^{J_m} 2^{3j/2} \sup_{x,y \in \mathbb{R}} \left\{ \sum_{k \in \mathbb{Z}} |\psi_{j,k}(x) - \psi_{j,k}(y)| \right\} \\ &\leq \frac{2}{3} A K_L K_\infty 2^{2J_m} \tau_q \text{ since } \left\| \sum_{k \in \mathbb{Z}} |\psi_{j,k}| \right\|_\infty \leq A K_\infty 2^{j/2} \end{aligned}$$

We can do the same computations for the term  $g_{B,m}(j, k, l, l')$  and we obtain

$$\mathbb{E} \left( \sum_{j,k \in \hat{m}} ((P_n - P_n^*)(\psi_{j,k}))^2 \right) \leq L\tau_q \sum_{m \in \mathcal{M}_n} 2^{2J_m} \leq L\tau_q 2^{2J_n} \leq L \frac{(\log n)^{\kappa(\theta+1)}}{n^{(\theta-3)/2}}.$$

The last inequality comes from  $q \geq \sqrt{n}/(2(\log n)^\kappa)$  and Assumption **[AR]**, the one before comes from Assumption **[W]**.  $\square$

**Claim 3.** *Let us keep the notations of Theorem 4.1, let  $u = 6/(7 + \theta) < 1/2$  and recall that  $\kappa > 2$ . Let  $\gamma$  be a real number in  $(1, \kappa/2)$ . Let*

$$L_1^2 = AK_\infty K_{BV} \sum_{l=0}^{\infty} \tilde{\beta}_l, \quad L_2^2 = 2\Phi K_{BV}^u \sum_{k=0}^{\infty} \tilde{\beta}_k^u, \quad L_3 = \kappa(\epsilon)\Phi$$

and  $L_{1,m} = 4(1 + \epsilon) \left( (1 + \epsilon)L_1 + L_2 \sqrt{\frac{(\log D_m)^\gamma}{D_m^{1/2-u}}} + L_3 \frac{(\log D_m)^\gamma}{(\log n)^\kappa} \right)^2, \quad (35)$

There exists a constant  $L_s$  such that

$$\mathbb{E} \left( \sup_{m \in \mathcal{M}_n} \left\{ \sum_{(j,k) \in m} (\nu_n^*)^2(\psi_{j,k}) - \frac{L_{1,m} D_m}{n} \right\} \right) \leq \frac{L_s}{n}.$$

**Remark :** The series  $\sum_{l=0}^{\infty} \tilde{\beta}_l$  and  $\sum_{k=0}^{\infty} \tilde{\beta}_k^u$  are convergent under our hypotheses on the coefficients  $\tau$ . Since  $s \in L^2([0, 1])$ , we have from Inequality (6),  $\tilde{\beta}_l \leq 2\|s\|_2^{2/3} \tau_l^{1/3}$  and thus  $\tilde{\beta}_l \leq 2\|s\|_2^{2/3} (1+l)^{-(1+\theta)/3}$ . The series  $\sum_{k=0}^{\infty} \tilde{\beta}_k^u$  converge since  $\theta > 5$  and

$$\frac{u(1+\theta)}{3} = \frac{2(1+\theta)}{7+\theta} = 1 + \frac{\theta-5}{\theta+7} > 1.$$

We use here  $\tilde{\beta}$  instead of  $\tau$  which allows to take  $L_1$  not depending on  $\|s\|_2$ .

**Proof of Claim 3 :**

As in the previous section we use the following decomposition

$$\begin{aligned} \sum_{(j,k) \in m} (\nu_n^*)^2(\psi_{j,k}) &= \sum_{(j,k) \in m} (\bar{\nu}_{A,p}(\bar{\psi}_{j,k}) + \bar{\nu}_{B,p}(\bar{\psi}_{j,k}))^2 \\ &\leq 2 \sum_{(j,k) \in m} (\bar{\nu}_{A,p}(\bar{\psi}_{j,k}))^2 + 2 \sum_{(j,k) \in m} (\bar{\nu}_{B,p}(\bar{\psi}_{j,k}))^2 \end{aligned}$$

We treat both terms with Proposition 7.4 applied to the random variables  $(A_l^*)_{0=1, \dots, p-1}$  and  $(B_l^*)_{l=0, \dots, p-1}$  and to the class of functions  $\{(\bar{\psi}_{j,k})_{(j,k) \in m}\}$ . Let

$$B_m^2 = \sum_{(j,k) \in m} \text{Var}(\bar{\psi}_{j,k}(A_1)), \quad V_m^2 = \sup_{t \in B_1(S_m)} \text{Var}(\bar{t}(A_1)), \quad H_m^2 = \left\| \sum_{(j,k) \in m} \bar{\psi}_{j,k}^2 \right\|_\infty.$$

We have, from Proposition 7.4

$$\forall x > 0, \quad \mathbb{P} \left[ \sqrt{\sum_{(j,k) \in m} (\bar{\nu}_{A,p})^2 \bar{\psi}_{j,k}} \geq \frac{(1+\epsilon)}{\sqrt{p}} B_m + V_m \sqrt{\frac{2x}{p}} + \kappa(\epsilon) \frac{H_m x}{p} \right] \leq e^{-x}. \quad (36)$$

Let us now evaluate  $B_m$ ,  $V_m$  and  $H_m$ , we have

$$B_m^2 = \frac{1}{(2q)^2} \sum_{(j,k) \in m} \text{Var} \left( \sum_{i=1}^q \psi_{j,k}(X_i) \right).$$

From (17) and (15) we have  $\forall j, k \|\psi_{j,k}\|_{BV} \leq K_{BV} 2^{j/2}$  and  $\forall j \|\sum_{k \in \mathbb{Z}} |\psi_{j,k}|\|_{\infty} \leq AK_{\infty} 2^{j/2}$ . Thus, from Inequality (5)

$$\begin{aligned} \sum_{(j,k) \in m} \text{Var} \left( \sum_{i=1}^q \psi_{j,k}(X_i) \right) &\leq 2 \sum_{(j,k) \in m} \sum_{l=1}^q (q+1-l) |\text{Cov}(\psi_{j,k}(X_1), \psi_{j,k}(X_l))| \\ &\leq 2q \sum_{j=0}^{J_m} \sum_{k \in \mathbb{Z}} \sum_{l=1}^q \|\psi_{j,k}\|_{BV} \mathbb{E}(|\psi_{j,k}(X_1)| b(\sigma(X_1), X_l)) \\ &\leq 2K_{BV} q \sum_{j=0}^{J_m} 2^{j/2} \left\| \sum_{k \in \mathbb{Z}} |\psi_{j,k}(X_0)| \right\|_{\infty} \sum_{l=1}^q \tilde{\beta}_{l-1} \\ &\leq 2q \left( AK_{\infty} K_{BV} \sum_{l=0}^{\infty} \tilde{\beta}_l \right) D_m. \end{aligned}$$

The last inequality comes from Assumption **[W]**.

Since  $L_1^2 = AK_{\infty} K_{BV} \sum_{l=0}^{\infty} \tilde{\beta}_l$  we have

$$B_m^2 \leq \frac{L_1^2 D_m}{2q}. \quad (37)$$

Let us deal with the term  $V_m^2$ . We have

$$V_m^2 \leq \sup_{t \in B_1(S_m)} \text{Var}(\bar{t}(A_1)) \leq \frac{2}{(2q)^2} \sum_{k=1}^q (q+1-k) \sup_{t \in B_1(S_m)} |\text{Cov}(t(X_1), t(X_k))| \quad (38)$$

From Inequality (5), we have

$$|\text{Cov}(t(X_1), t(X_k))| \leq \|t\|_{BV} \|t\|_{\infty} \tilde{\beta}_{k-1}.$$

Since  $t$  belongs to  $B_1(S_m)$ , we have  $t = \sum_{(j,k) \in m} a_{j,k} \psi_{j,k}$ , with  $\sum_{(j,k) \in m} a_{j,k}^2 \leq 1$ . Thus, by Cauchy-Schwarz inequality

$$\begin{aligned} \sum_{i=1}^l |t(x_{i+1}) - t(x_i)| &\leq \sum_{(j,k) \in m} |a_{j,k}| \sum_{i=1}^l |\psi_{j,k}(x_{i+1}) - \psi_{j,k}(x_i)| \\ &\leq \left( \sum_{(j,k) \in m} a_{j,k}^2 \right)^{1/2} \left( \sum_{(j,k) \in m} \left( \sum_i |\psi_{j,k}(x_{i+1}) - \psi_{j,k}(x_i)| \right)^2 \right)^{1/2} \\ &\leq \left( \sum_{(j,k) \in m} \|\psi_{j,k}\|_{BV}^2 \right)^{1/2} \leq K_{BV} D_m. \end{aligned}$$

Thus  $\|t\|_{BV} \leq D_m K_{BV}$ . From Assumption **[M<sub>2</sub>]**, we have  $\|t\|_{\infty} \leq \Phi \sqrt{D_m}$ . Thus

$$|\text{Cov}(t(X_1), t(X_k))| \leq \Phi K_{BV} \tilde{\beta}_{k-1} D_m^{3/2}. \quad (39)$$

Moreover, we have by Cauchy-Schwarz inequality and  $[M_2]$

$$|\text{Cov}(t(X_1), t(X_k))| \leq \|t\|_\infty \|t\|_2 \|s\|_2 \leq \Phi \|s\|_2 \sqrt{D_m}. \quad (40)$$

We use the inequality  $a \wedge b \leq a^u b^{1-u}$  with

$$a = \Phi K_{BV} \tilde{\beta}_{k-1} D_m^{3/2}, \quad b = \Phi \|s\|_2 \sqrt{D_m}, \quad u = \frac{6}{7+\theta} < \frac{1}{2}.$$

From (39) and (40), we derive that

$$|\text{Cov}(t(X_1), t(X_k))| \leq L'_k D_m^{1/2+u} \text{ where } L'_k = \Phi \left( K_{BV} \tilde{\beta}_{k-1} \right)^u \|s\|_2^{1-u}.$$

Plugging this inequality in (38), we obtain

$$V_m^2 \leq \frac{L_2^2 \|s\|_2^{1-u} D_m^{1/2+u}}{4q} \text{ since } L_2^2 = 2\Phi K_{BV}^u \sum_{k=0}^{\infty} \tilde{\beta}_k^u. \quad (41)$$

Finally, we have from hypothesis  $[M_2]$

$$H_m^2 \leq \frac{1}{4} \left\| \sum_{(j,k) \in m} \psi_{j,k}^2 \right\|_\infty \leq \frac{\Phi^2 D_m}{4}. \quad (42)$$

Let  $y > 0$  and let us apply Inequality (36) with  $x = ((\log D_m)^\gamma / \|s\|_2^{1-u}) + (y/D_m^{1/2+u})$ . We have, from (37), (41) and (42)

$$\begin{aligned} \mathbb{P} \left[ \sum_{(j,k) \in m} (\bar{\nu}_{A,p})^2 (\bar{\psi}_{j,k}) > \left( (1+\epsilon) \sqrt{\frac{L_1^2 D_m}{2pq}} + \frac{L_3 \sqrt{D_m}}{2p} \left( \frac{(\log D_m)^\gamma}{\|s\|_2^{1-u}} + \frac{y}{D_m^{1/2+u}} \right) \right) \right. \\ \left. + \sqrt{\frac{L_2^2 \|s\|_2^{1-u} D_m^{1/2+u}}{2pq} \left( \frac{(\log D_m)^\gamma}{\|s\|_2^{1-u}} + \frac{y}{D_m^{1/2+u}} \right)} \right] \leq e^{-\frac{(\log D_m)^\gamma}{\|s\|_2^{1-u}}} e^{-D_m^{-(1/2+u)} y}. \end{aligned}$$

Then, we use the inequality  $\sqrt{\alpha + \beta} \leq \sqrt{\alpha} + \sqrt{\beta}$  with

$$\alpha = \frac{(\log D_m)^\gamma}{\|s\|_2^{1-u}} \text{ and } \beta = \frac{y}{D_m^{1/2+u}}$$

and the inequality  $(a+b)^2 \leq (1+\epsilon)a^2 + (1+\epsilon^{-1})b^2$  with

$$a = \left( (1+\epsilon)L_1 + L_2 \sqrt{\frac{(\log D_m)^\gamma}{D_m^{1/2+u}}} + \frac{L_3 (\log D_m)^\gamma}{\|s\|_2^{1-u} (\log n)^\kappa} \right) \sqrt{\frac{D_m}{n}}$$

$$\text{and } b = \frac{1}{\sqrt{n}} \left( L_2 \sqrt{\|s\|_2^{1-u} y} + \frac{L_3 y}{(\log n)^\kappa D_m^u} \right).$$

Setting  $L_m = (1+\epsilon)a^2 n / D_m$ , we obtain

$$\begin{aligned} \mathbb{P} \left( \sum_{(j,k) \in m} (\bar{\nu}_{A,p})^2 (\bar{\psi}_{j,k}) - \frac{L_m D_m}{n} > \frac{(1+\epsilon^{-1})}{n} \left( L_2 \sqrt{\|s\|_2^{1-u} y} + \frac{L_3 y}{(\log n)^\kappa D_m^u} \right)^2 \right) \\ \leq e^{-\frac{(\log D_m)^\gamma}{\|s\|_2^{1-u}}} e^{-D_m^{-(1/2+u)} y}. \end{aligned}$$

Thus, for all  $y > 0$ ,

$$\mathbb{P} \left( \sup_{m \in \mathcal{M}_n} \left\{ \sum_{(j,k) \in m} (\bar{\nu}_{A,p})^2 (\bar{\psi}_{j,k}) - \frac{L_m D_m}{n} \right\} > \frac{L_s}{n} (y + y^2) \right) \leq \sum_{m \in \mathcal{M}_n} e^{-\frac{(\log D_m)^\gamma}{\|s\|_2^{1-u}} - D_m^{-(1/2+u)} y}$$

where  $L_s = 2(1 + \epsilon^{-1}) \left[ (L_2 \sqrt{\|s\|_2^{1-u}}) \vee L_3 / ((\log 2)^\kappa 2^u) \right]^2$ . We can integrate this last inequality to prove Claim 3.  $\square$

**Claim 4 :** We keep the notations of the previous Claims. Let

$$L_2(m, m') = 4 \left( L_2 \sqrt{\frac{(\log(D_m \vee D_{m'}))^\gamma}{(D_m \vee D_{m'})^{1/2-u}}} + \frac{\Phi}{3(\log n)^{\kappa-\gamma}} \right)^2. \quad (43)$$

Then there exists a constant  $L_{s,\theta}$  depending on  $\|s\|_2$  and  $\theta$  such that, for all  $\eta > 0$

$$\mathbb{E} \left( \sup_{m, m' \in \mathcal{M}_n} \left\{ \nu_n(s_m - s_{m'}) - \frac{\|s_m - s_{m'}\|_2^2}{2\eta} - \eta \frac{L_2(m, m')(D_m \vee D_{m'})}{n} \right\} \right) \leq \frac{\eta L_{s,\theta}}{n}.$$

**Proof of Claim 4 :**

$$\begin{aligned} & \mathbb{E} \left( \sup_{m, m' \in \mathcal{M}_n} \left\{ \nu_n(s_m - s_{m'}) - \frac{\|s_m - s_{m'}\|_2^2}{2\eta} - \eta \frac{L_2(m, m')(D_m \vee D_{m'})}{n} \right\} \right) \\ & \leq \mathbb{E} \left( \sup_{m, m'} (P_n - P_n^*)(s_m - s_{m'}) \right) \\ & + \mathbb{E} \left( \sup_{m, m'} \left\{ \nu_n^*(s_m - s_{m'}) - \frac{\|s_m - s_{m'}\|_2^2}{2\eta} - \eta \frac{L_2(m, m')(D_m \vee D_{m'})}{n} \right\} \right). \quad (44) \end{aligned}$$

Since  $\forall l = 0, \dots, p-1$ ,  $\mathbb{E}(|A_l - A_l^*|_q) \leq q\tau_q$ , we have

$$\begin{aligned} \mathbb{E} \left( \sup_{m, m'} (P_n - P_n^*)(s_m - s_{m'}) \right) & \leq 2 \sum_{m, m'} \mathbb{E} (|(\bar{s}_m - \bar{s}_{m'})(A_1) - (\bar{s}_m - \bar{s}_{m'})(A_1^*)|) \\ & \leq \tau_q \sum_{m, m'} \text{Lip}(s_m - s_{m'}). \end{aligned}$$

When  $m \subset m'$ , we have, for all  $x, y \in \mathbb{R}$ , using Assumption **[W]**,

$$\frac{|(s_m - s_{m'})(x - y)|}{|x - y|} \leq \sum_{j=J_m+1}^{J_{m'}} \sum_{k=-A_2}^{2^j - A_1} |P\psi_{j,k}| \frac{|\psi_{j,k}(x) - \psi_{j,k}(y)|}{|x - y|}$$

Let us fix  $j \in [J_m + 1, J_{m'}]$ , from Assumption **[W]**, there is less than  $A$  indexes  $k \in \mathbb{Z}$  such that  $\psi_{j,k}(x) \neq 0$ , thus there is less than  $2A$  indexes such that  $|\psi_{j,k}(x) - \psi_{j,k}(y)| \neq 0$ . Hence

$$\begin{aligned} \sum_{k \in \mathbb{Z}} |P\psi_{j,k}| \frac{|\psi_{j,k}(x) - \psi_{j,k}(y)|}{|x - y|} & \leq 2A \sup_{k \in \mathbb{Z}} |P\psi_{j,k}| \text{Lip}(\psi_{j,k}) \\ & \leq 2A \|s\|_2 K_L 2^{3j/2}. \end{aligned}$$

Thus,  $\text{Lip}(s_m - s_{m'}) \leq A \|s\|_2 K_L \sqrt{8} 2^{3J_{m'}/2} / (\sqrt{8} - 1)$  and by Assumptions **[W]**, **[AR]** and the value of  $q$ ,

$$\mathbb{E} \left( \sup_{m,m'} (P_n - P_n^*)(s_m - s_{m'}) \right) \leq L_s n^{3/2} (\log n) \tau_q \leq L_s \frac{(\log n)^{\kappa(\theta+1)+1}}{n^{(\theta-2)/2}}. \quad (45)$$

Let us deal with the other term in (44). We have,  $\forall \eta > 0$

$$\begin{aligned} \nu_n^*(s_m - s_{m'}) &\leq \frac{\|s_m - s_{m'}\|_2^2}{2\eta} + \frac{\eta}{2} (\bar{\nu}_{A,p}(\bar{t}_{m,m'}) + \bar{\nu}_{B,p}(\bar{t}_{m,m'}))^2 \\ &\leq \frac{\|s_m - s_{m'}\|_2^2}{2\eta} + \eta (\bar{\nu}_{A,p}(\bar{t}_{m,m'}))^2 + \eta (\bar{\nu}_{B,p}(\bar{t}_{m,m'}))^2 \end{aligned} \quad (46)$$

where, as in the proof of Theorem 3.1,  $t_{m,m'} = (s_m - s_{m'}) / \|s_m - s_{m'}\|_2$ . We apply Bernstein's inequality to the function  $\bar{t}_{m,m'}$  and the variables  $A_i^*$ , we have

$$\forall x > 0, \mathbb{P} \left( \bar{\nu}_{A,p}(\bar{t}_{m,m'}) > \sqrt{\frac{2\text{Var}(\bar{t}_{m,m'}(A_0))x}{p}} + \frac{\|\bar{t}_{m,m'}\|_\infty x}{3p} \right) \leq e^{-x}. \quad (47)$$

We proceed as in the proof of Claim 3 to control this variance. We have, by stationarity of the process  $(X_n)_{n \in \mathbb{Z}}$ ,

$$\text{Var}(\bar{t}_{m,m'}(A_0)) = \frac{1}{2q^2} \sum_{k=0}^{q-1} (q-k) \text{Cov}(t_{m,m'}(X_1), t_{m,m'}(X_{k+1})).$$

From Inequality (5), we have

$$|\text{Cov}(t_{m,m'}(X_1), t_{m,m'}(X_{k+1}))| \leq \|t_{m,m'}\|_{BV} \|t_{m,m'}\|_\infty \tilde{\beta}_k.$$

Let  $m \triangle m'$  be the set of indexes that belong to  $m \cup m'$  but do not belong to  $m \cap m'$ . We use the same computations as in the proof of Claim 3 to get

$$\|t_{m,m'}\|_{BV} \leq \frac{\left\| \sum_{(j,k) \in m' \triangle m} (P\psi_{j,k}) \psi_{j,k} \right\|_{BV}}{\|s_m - s_{m'}\|_2} \leq \sqrt{\sum_{(j,k) \in m' \triangle m} \|\psi_{j,k}\|_{BV}^2} \leq K_{BV} (D_m \vee D_{m'}).$$

Since  $\|t_{m,m'}\|_\infty = \Phi \sqrt{D_m \vee D_{m'}}$ , we have

$$|\text{Cov}(t_{m,m'}(X_1), t_{m,m'}(X_{k+1}))| \leq \Phi K_{BV} \tilde{\beta}_k (D_m \vee D_{m'})^{3/2}. \quad (48)$$

Moreover, we have

$$\text{Cov}(t_{m,m'}(X_1), t_{m,m'}(X_{k+1})) \leq \|t_{m,m'}\|_\infty \|t_{m,m'}\|_2 \|s\|_2 \leq \Phi \|s\|_2 \sqrt{(D_m \vee D_{m'})}. \quad (49)$$

Thus, using  $a \wedge b \leq a^u b^{1-u}$  with

$$a = \Phi K_{BV} \tilde{\beta}_k (D_m \vee D_{m'})^{3/2}, \quad b = \Phi \|s\|_2 \sqrt{(D_m \vee D_{m'})}, \quad \text{and } u = \frac{6}{7+\theta} < \frac{1}{2},$$

we have

$$|\text{Cov}(t_{m,m'}(X_1), t_{m,m'}(X_{k+1}))| \leq \Phi K_{BV}^u \tilde{\beta}_k^u \|s\|_2^{1-u} (D_m \vee D_{m'})^{1/2+u}.$$



Thus

$$\text{Var}(\bar{t}_{m,m'}(A_0)) \leq \Phi K_{BV}^u \left( \sum_{k=0}^{\infty} \tilde{\beta}_k^u \right) \|s\|_2^{1-u} \frac{(D_m \vee D_{m'})^{1/2+u}}{2q}. \quad (50)$$

Moreover

$$\|\bar{t}_{m,m'}\|_{\infty} \leq \frac{1}{2} \|t_{m,m'}\|_{\infty} \leq \frac{1}{2} \Phi \sqrt{D_m \vee D_{m'}}. \quad (51)$$

Now, we use (47) with  $x = (\log(D_m \vee D_{m'}))^{\gamma} / \|s\|_2^{1-u} + y / (D_m \vee D_{m'})^{1/2+u}$ . From (50) and (51), we have for all  $y > 0$ ,

$$\begin{aligned} & \mathbb{P} \left( \bar{\nu}_{A,p}(\bar{t}_{m,m'}) > L_2 \sqrt{\frac{(D_m \vee D_{m'})^{1/2+u}}{2pq} \left( (\log(D_m \vee D_{m'}))^{\gamma} + \frac{\|s\|_2^{1-u} y}{(D_m \vee D_{m'})^{1/2+u}} \right)} \right. \\ & \quad \left. + \frac{\Phi \sqrt{D_m \vee D_{m'}}}{6p} \left( \frac{(\log(D_m \vee D_{m'}))^{\gamma}}{\|s\|_2^{1-u}} + \frac{y}{(D_m \vee D_{m'})^{1/2+u}} \right) \right) \\ & \leq e^{-\frac{(\log(D_m \vee D_{m'}))^{\gamma}}{\|s\|_2^{1-u}}} e^{-\frac{y}{(D_m \vee D_{m'})^{1/2+u}}}. \end{aligned}$$

Now we use the inequality  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$  with

$$a = (\log(D_m \vee D_{m'}))^{\gamma} \text{ and } b = \frac{\|s\|_2^{1-u} y}{(D_m \vee D_{m'})^{1/2+u}}$$

and we obtain, using Assumption  $[M_1]$

$$\begin{aligned} & \mathbb{P} \left( \bar{\nu}_{A,p} \bar{t}_{m,m'} - \sqrt{\frac{L_2(m, m')(D_m \vee D_{m'})}{n}} > \frac{L_s}{\sqrt{n}} (\sqrt{y} + y) \right) \\ & \leq e^{-\frac{(\log(D_m \vee D_{m'}))^{\gamma}}{\|s\|_2^{1-u}}} e^{-(D_m \vee D_{m'})^{-(1/2+u)} y}, \end{aligned}$$

with

$$\begin{aligned} L_2(m, m') &= \left( L_2 \sqrt{\frac{(\log(D_m \vee D_{m'}))^{\gamma}}{(D_m \vee D_{m'})^{1/2-u}} + \frac{\Phi (\log(D_m \vee D_{m'}))^{\gamma}}{3(\log n)^{\kappa}}} \right)^2, \\ \text{and } L_s &= L_2 \sqrt{\|s\|_2^{1-u}} \vee \frac{\Phi}{3(\log 2)^{\kappa} 2^u}. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} & \mathbb{P} \left( (\bar{\nu}_{A,p} \bar{t}_{m,m'})^2 > 2 \frac{L_2(m, m')(D_m \vee D_{m'})}{n} + 4 \frac{L_s^2}{n} (y + y^2) \right) \\ & \leq e^{-\frac{(\log(D_m \vee D_{m'}))^{\gamma}}{\|s\|_2^{1-u}}} e^{-\frac{y}{(D_m \vee D_{m'})^{1/2+u}}}. \end{aligned}$$

The same result holds for  $\bar{\nu}_{B,p} \bar{t}_{m,m'}$ . Thus we obtain from (46)

$$\begin{aligned} & \mathbb{P} \left( \nu_n^*(s_m - s_{m'}) \geq \frac{\|s_m - s_{m'}\|_2^2}{2\eta} + 4\eta \frac{L_2(m, m')(D_m \vee D_{m'})}{n} + 8\eta \frac{L_s^2}{n} (y + y^2) \right) \\ & \leq 2e^{-\frac{(\log(D_m \vee D_{m'}))^{\gamma}}{\|s\|_2^{1-u}}} e^{-\frac{y}{(D_m \vee D_{m'})^{1/2+u}}}. \end{aligned}$$

We deduce that

$$\begin{aligned} & \mathbb{P} \left( \exists m, m' \in \mathcal{M}_n, \nu_n^*(s_m - s_{m'}) - \frac{\|s_m - s_{m'}\|_2^2}{2\eta} - 4\eta \frac{L_2(m, m')(D_m \vee D_{m'})}{n} \right. \\ & \left. \geq 8\eta \frac{L_s^2}{n} (y + y^2) \right) \leq 2 \sum_{m, m' \in \mathcal{M}_n} \left( e^{-\frac{(\log(D_m \vee D_{m'}))^\gamma}{\|s\|_2^{1-u}}} \right) e^{-\frac{y}{(D_m \vee D_{m'})^{1/2+u}}}. \end{aligned}$$

We integrate this last inequality to get Claim 4.  $\square$

### Conclusion of the proof:

Take

$$\text{pen}'(m) \geq (2L_{1,m} + \eta L_2(m, m)) \frac{D_m}{n},$$

where  $L_{1,m}$  and  $L_2(m, m)$  are defined by (35) and (43) respectively. From Claims 2, 3 and 4, if we take the expectation in (21), we have, for some constant  $L_s$ ,

$$\mathbb{E} \left( \|s - \hat{s}\|_2^2 \right) \leq \mathbb{E} \left( \|s - \hat{s}_{m_o}\|_2^2 + \text{pen}'(m_o) - V(m_o) + 2\eta L_2(m_o, m_o) \frac{D_{m_o}}{n} \right) + \frac{\eta L_s}{n}. \quad (52)$$

Moreover, if  $D_m \geq ((L_2/L_1)(\log n)^{\kappa-\gamma/2})^{2(7+\theta)/(\theta-5)}$ , we have

$$\begin{aligned} \frac{L_{1,m}}{4L_1^2} & \leq (1 + \epsilon) \left( (1 + \epsilon) + \left( 1 + \frac{L_3}{2L_1} \right) (\log n)^{-(\kappa-\gamma)} \right)^2 \\ & \leq (1 + \epsilon)^3 + (1 + \epsilon^{-1})(1 + \epsilon) \left( 1 + \frac{L_3}{2L_1} \right)^2 (\log n)^{-2(\kappa-\gamma)}. \end{aligned} \quad (53)$$

We use the inequality  $(a + b)^2 \leq (1 + \epsilon)a^2 + (1 + \epsilon^{-1})b^2$  to obtain (53). Moreover, we have

$$L_2(m, m) \leq 4L_1^2 \left( \left( 1 + \frac{\Phi}{6L_1} \right) (\log n)^{-(\kappa-\gamma)} \right)^2.$$

As in the proof of Theorem 3.1, we take  $\eta = (\log n)^{\kappa-\gamma}$  and we fix  $\epsilon$  sufficiently small. For  $n \geq n_o$ , we have  $2L_{1,m} + \eta L_2(m, m) < KL_1^2$ . Thus inequality (18) follows from (52).  $\square$

## 7 Appendix

This section is devoted to technical lemmas that are needed in the proofs.

### 7.1 Covariance inequality

**Lemma 7.1** *Viennet's inequality* Let  $(X_n)_{n \in \mathbb{Z}}$  be a stationary and  $\beta$ -mixing process. There exists a positive function  $b$  such that  $P(b) \leq \sum_{l=0}^{\infty} \beta_l$ ,  $P(b^p) \leq p \sum_{l=1}^{\infty} l^{p-1} \beta_l$ , and for all function  $h \in L_2(P)$

$$\text{Var} \left( \sum_{l=1}^q h(X_l) \right) \leq 4qP(bh^2). \quad (54)$$

## 7.2 Concentration inequalities

We sum up in this section the concentration inequalities we used in the proofs. We begin with Bernstein's inequality

### Proposition 7.2 *Bernstein's inequality*

Let  $X_1, \dots, X_n$  be iid random variables valued in a measurable space  $(X, \mathcal{X})$  and let  $t$  be a measurable real valued function. Let  $v = \text{Var}(t(X_1))$  and  $b = \|t\|_\infty$ , then, for all  $x > 0$ , we have

$$\mathbb{P} \left( (P_n - P)t > v\sqrt{\frac{2x}{n}} + \frac{bx}{3n} \right) \leq e^{-x}.$$

Now we give the most important tool of our proof, it is a concentration's inequality for the supremum of the empirical process over a class of function. We give here the version of Bousquet [10].

### Theorem 7.3 *Talagrand's Theorem*

Let  $X_1, \dots, X_n$  be i.i.d random variables valued in some measurable space  $[X, \mathcal{X}]$ . Let  $\mathcal{F}$  be a separable class of bounded functions from  $X$  to  $\mathbb{R}$  and assume that all functions  $t$  in  $\mathcal{F}$  are  $P$ -measurable, and satisfy  $\text{Var}(t(X_1)) \leq \sigma^2$ ,  $\|t\|_\infty \leq b$ . Then

$$\mathbb{P} \left( \sup_{t \in \mathcal{F}} \nu_n(t) > \mathbb{E} \left( \sup_{t \in \mathcal{F}} \nu_n(t) \right) + \sqrt{\frac{2x(\sigma^2 + 2b\mathbb{E}(\sup_{t \in \mathcal{F}} \nu_n(t)))}{n}} + \frac{bx}{3n} \right) \leq e^{-x}.$$

In particular, for all  $\epsilon > 0$ , if  $\kappa(\epsilon) = 1/3 + \epsilon^{-1}$ , we have

$$\mathbb{P} \left( \sup_{t \in \mathcal{F}} \nu_n(t) > (1 + \epsilon)\mathbb{E} \left( \sup_{t \in \mathcal{F}} \nu_n(t) \right) + \sigma\sqrt{\frac{2x}{n}} + \kappa(\epsilon)\frac{bx}{n} \right) \leq e^{-x}.$$

We can deduce from this Theorem a concentration's inequality for  $\chi$ -square type statistics. This is Proposition (7.3) of Massart [20].

**Proposition 7.4** Let  $X_1, \dots, X_n$  be independent and identically distributed random variables valued in some measurable space  $(X, \mathcal{X})$ . Let  $P$  denote their common distribution. Let  $\phi_\lambda$  be a finite family of measurable and bounded functions on  $(X, \mathcal{X})$ . Let

$$H_\Lambda^2 = \left\| \sum_{\lambda \in \Lambda} \phi_\lambda^2 \right\|_\infty \text{ and } B_\Lambda^2 = \sum_{\lambda \in \Lambda} \text{Var}(\phi_\lambda(X_1)).$$

Moreover, let  $\mathcal{S}_\Lambda = \{a \in \mathbb{R}^\Lambda : \sum_{\lambda \in \Lambda} a_\lambda^2 = 1\}$  and

$$V_\Lambda^2 = \sup_{a \in \mathcal{S}_\Lambda} \left\{ \text{Var} \left( \sum_{\lambda \in \Lambda} a_\lambda \phi_\lambda(X_1) \right) \right\}.$$

Then the following inequality holds, for all positive  $x$  and  $\epsilon$

$$\mathbb{P} \left[ \left( \sum_{\lambda \in \Lambda} (P_n - P)^2 \phi_\lambda \right)^{1/2} \geq \frac{1 + \epsilon}{\sqrt{n}} B_\Lambda + V_\Lambda \sqrt{\frac{2x}{n}} + \kappa(\epsilon) \frac{H_\Lambda x}{n} \right] \leq e^{-x}, \quad (55)$$

where  $\kappa(\epsilon) = \epsilon^{-1} + 1/3$ .

**Proof :**

Following Massart [20] Proposition 7.3, we remark that, by Cauchy-Schwarz's inequality

$$\left( \sum_{\lambda \in \Lambda} \nu_n^2 \phi_\lambda \right)^{1/2} = \sup_{a \in \mathcal{S}_\Lambda} \sum_{\lambda \in \Lambda} a_\lambda \nu_n \phi_\lambda = \sup_{a \in \mathcal{S}_\Lambda} \nu_n \left( \sum_{\lambda \in \Lambda} a_\lambda \phi_\lambda \right).$$

Thus the result follows by applying Talagrand's Theorem to the class of functions

$$\mathcal{F} = \left\{ t = \sum_{\lambda \in \Lambda} a_\lambda \phi_\lambda; a \in \mathcal{S}_\Lambda \right\}.$$

## References

- [1] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281. Akadémiai Kiadó, Budapest, 1973.
- [2] Hirotugu Akaike. Statistical predictor identification. *Ann. Inst. Statist. Math.*, 22:203–217, 1970.
- [3] Donald W. K. Andrews. Nonstrong mixing autoregressive processes. *J. Appl. Probab.*, 21(4):930–934, 1984.
- [4] S. Arlot and P. Massart. Data-driven calibration of penalties for least squares regression. *Submitted to Journal of Machine learning research*, 2008.
- [5] Sylvain Arlot. Model selection by resampling penalization. *hal-00262478*, 2008.
- [6] Y. Baraud, F. Comte, and G. Viennet. Adaptive estimation in autoregression or  $\beta$ -mixing regression via model selection. *Ann. Statist.*, 29(3):839–875, 2001.
- [7] Henry C. P. Berbee. *Random walks with stationary increments and renewal theory*, volume 112 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam, 1979.
- [8] Lucien Birgé and Pascal Massart. From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York, 1997.
- [9] Lucien Birgé and Pascal Massart. Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields*, 138(1-2):33–73, 2007.
- [10] Olivier Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris*, 334(6):495–500, 2002.
- [11] Richard C. Bradley. *Introduction to strong mixing conditions. Vol. 1*. Kendrick Press, Heber City, UT, 2007.
- [12] F. Comte, J. Dedecker, and M. L. Taupin. Adaptive density deconvolution with dependent inputs. *Math. Methods Statist.*, 17(2):87–112, 2008.
- [13] Fabienne Comte and Florence Merlevède. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, 6:211–238 (electronic), 2002. New directions in time series analysis (Luminy, 2001).
- [14] Jérôme Dedecker, Paul Doukhan, Gabriel Lang, José Rafael León R., Sana Louhichi, and Clémentine Prieur. *Weak dependence: with examples and applications*, volume 190 of *Lecture Notes in Statistics*. Springer, New York, 2007.
- [15] Jérôme Dedecker and Clémentine Prieur. New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields*, 132(2):203–236, 2005.
- [16] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539, 1996.
- [17] Paul Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples.

- [18] Irène Gannaz and Olivier Wintenberger. Adaptive density estimation under dependence. *forthcoming in ESAIM, Probab. and Statist.*, 2008.
- [19] C.L. Mallows. Some comments on  $c_p$ . *Technometrics*, 15:661–675, 1973.
- [20] Pascal Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.
- [21] C. Prieur. Change point estimation by local linear smoothing under a weak dependence condition. *Math. Methods Statist.*, 16(1):25–41, 2007.
- [22] Emmanuel Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2000.
- [23] Mats Rudemo. Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, 9(2):65–78, 1982.
- [24] Michel Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.
- [25] Gabrielle Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 107(4):467–492, 1997.
- [26] V. A. Volkonskiĭ and Yu. A. Rozanov. Some limit theorems for random functions. I. *Teor. Veroyatnost. i Primenen.*, 4:186–207, 1959.