ADAPTIVE DENSITY DECONVOLUTION WITH DEPENDENT INPUTS

F. COMTE[*,1], J. DEDECKER[2], AND M. L. TAUPIN [3]

ABSTRACT. In the convolution model $Z_i = X_i + \varepsilon_i$, we give a model selection procedure to estimate the density of the unobserved variables $(X_i)_{1\le i\le n}$, when the sequence $(X_i)_{i\ge 1}$ is strictly stationary but not necessarily independent. This procedure depends on wether the density of ε_i is super smooth or ordinary smooth. The rates of convergence of the penalized contrast estimators are the same as in the independent framework, and are minimax over most classes of regularity on $\mathbb{R}$. Our results apply to mixing sequences, but also to many other dependent sequences. When the errors are super smooth, the condition on the dependence coefficients is the minimal condition of that type ensuring that the sequence $(X_i)_{i\ge 1}$ is not a long-memory process.

May 30, 2006

MSC 2000 Subject Classifications. 62G07-62G20

Keywords and phrases. Adaptive estimation. Deconvolution. Dependence. Mixing. Penalized Contrast. Hidden Markov models

1. Introduction

The problem of estimating the density of identically distributed but not independent random variables $X_1, \dots, X_n$ when they are observed with an additive and independent noise is encountered in numerous contexts. This problem is described by the model

$$Z_i = X_i + \varepsilon_i, \quad \text{for } i = 1, \dots, n, \tag{1.1}$$

where one observes $Z_1, \dots, Z_n$, and where $(\varepsilon_i)_{1\le i\le n}$ are independent and identically distributed (i.i.d.), and independent of $(X_i)_{1\le i\le n}$. When $(X_i)_{1\le i\le n}$ is a Markov chain, the model (1.1) is a particular case of hidden Markov models, with an additive structure.

Our aim is the adaptive estimation of g, the common distribution of the unobserved variables $(X_i)_{1\le i\le n}$, when the density f_ε of ε_i is known. More precisely we shall build an estimator of g without any prior knowledge on its smoothness, using the observations $(Z_i)_{1\le i\le n}$ and the knowledge of the convolution kernel f_ε. We shall assume that the known density f_ε belongs to various collections of densities, and that the dependence properties of the sequence $(X_i)_{i\ge 1}$ are described by appropriate dependence coefficients. More precisely, we consider two types of dependent sequences. We assume either that the sequence $(X_i)_{i\ge 1}$ is absolutely regular in the sense of Rozanov and Volkonskii (1960), or that it is τ-dependent in the sense of Dedecker and Prieur (2005). These dependence conditions are presented in Section 2 and motivated through various examples.

In density deconvolution, two factors determine the estimation accuracy. First, the smoothness of the density g to be estimated, and second the smoothness of the error density, the worst rates of

[1] Université Paris V, MAP5, UMR CNRS 8145.
[2] Université Paris 6, Laboratoire de Statistique Théorique et Appliquée.
[3] IUT de Paris V et Université Paris Sud, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8628.

convergence being obtained for the smoothest errors densities. We shall consider two classes of densities for f_ε: first the so called super smooth densities with exponential decay of their Fourier transform, and next the class of ordinary smooth densities with Fourier transform having a polynomial decay.

Let us briefly recall the previous results in the independent framework. To our knowledge, the first adaptive estimator has been proposed by Pensky and Vidakovic (1999). It is a wavelet estimator constructed *via* a thresholding procedure. This estimator achieves the minimax rates when g belongs to a Sobolev class, but it fails to reach the minimax rates when both the errors density and g are supersmooth. More recently, Comte *et al.* (2006) have proposed an adaptive estimator of g constructed by minimizing an appropriate penalized contrast function only depending on the observations and on f_ε. This estimator is minimax (sometimes within a negligible logarithmic factor) in all cases where lower bounds are previously known (i.e. in most cases). More precisely, the authors obtain non-asymptotic upper bounds for the Mean Integrated Squared Error (MISE), which ensure an automatic trade-off between a bias term and the penalty term. Hence, the estimator automatically achieves the best rate obtained by the collection of non-penalized estimators when the (unknown) optimal space is selected (sometimes up to a negligible logarithmic factor). When both the density and the errors are super smooth, this adaptive estimator significantly improves on the rates given by the adaptive estimator built in Pensky and Vidakovic (1999), whereas both adaptive estimators have the same rate in the other cases. This improvement partly comes from the choice of the Shannon basis (see Section 3.2) instead of the wavelet basis considered in Pensky and Vidakovic.

In the dependent context, we follow the approach proposed in Comte *et al.* (2006). We give adaptive estimators of g, constructed by minimizing an appropriate penalized contrast function. The penalty function depends on the known density f_ε, but it does not depend on the dependence coefficients of the sequence $(X_i)_{i\geq 1}$. The adaptive estimators have the same rates as in the independent case, under mild conditions on the dependence coefficients of $(X_i)_{i\geq 1}$. The important point here is that the penalty functions are the same (or almost the same) as in the independent framework. This is a bit surprising: indeed, when the $(X_i)_{1\leq i\leq n}$ are observed (i.e. $\varepsilon_i = 0$), the threshold level proposed in Tribouley and Viennet (1998) as well as the penalty function given in Comte and Merlevède (2002) (see also our Corollary 5.2) depend on the mixing coefficients of the sequence $(X_i)_{i\geq 1}$.

In Section 4 we deal with non adaptive estimators. As usual, we show that the MISE of the minimum contrast estimator is bounded by a squared bias plus a variance term. The variance term can be split into two terms. The first and dominating term of the variance is exactly the variance of a density deconvolution estimator in the independent context. It is as usual related to $\int_{|x|\leq C_n} |f_\varepsilon^*(x)|^{-2}dx$, $C_n \to \infty$. The second and negligible term in the variance is the term involving the dependence structure of the sequence $(X_i)_{i\geq 1}$. The main consequence of this first result is that this non adaptive estimator reaches the (minimax) rates of the i.i.d. case (as given in Fan (1991), Butucea (2004), and Butucea and Tsybakov (2005)), as soon as the dependence coefficients are summable. Moreover, even if the coefficients are not summable, there is no loss in the rate provided that the partial sums of the coefficients do not grow too fast with respect to $\int_{|x|\leq C_n} |f_\varepsilon^*(x)|^{-2}dx$. These results have to be compared with previously known results for non adaptive density deconvolution in dependent contexts. For strongly mixing sequences in the sense of Rosenblatt (1956), Masry (1993) propose a kernel-type estimator for the joint density g_p of $(X_1, \ldots, X_p)$ when it exists. For the (pointwise) Mean Square Error, he obtains the same rates as in the i.i.d. case provided that $\alpha(n) = O(n^{-2-\delta})$ for ordinary smooth f_ε, and provided that $\alpha(n) = O(n^{-1-\delta})$ for super smooth f_ε. When $p = 1$, our assumption on the mixing coefficients is weaker, since we only need $\sum_{n>0} \alpha(n) < \infty$ in both cases (see our Remark 4.1).

In the main part (Section 5), we study the adaptive estimators. We show that the squared bias term and the variance term obtained in the upper bound of the MISE of the adaptive estimator are the same as in the independent case. The model selection procedure depends on wether the density f_ε is super smooth or ordinary smooth.

When f_ε is super smooth, the adaptive estimator, is constructed with the exact penalty of the independent context. Its rate of convergence is exactly the same as in the independent case, provided that the dependence coefficients of $(X_i)_{i\geq 1}$ are summable. The main tools in this case are covariance inequalities for dependent variables, and concentration inequalities. The case of super smooth errors is particularly important, since it contains the case of Gaussian errors. It also contains the stochastic volatility model, in which $\varepsilon_i \sim \ln(\mathcal{N}(0,1)^2)$ (see Van Es *et al.* (2003, 2005), Comte and Genon-Catalot (2006)).

When f_ε is ordinary smooth, the adaptive estimator, is constructed with a penalty of the same order as in the independent context. Its rate of convergence is exactly the same as in the independent case. For ordinary smooth errors, the main tools are the coupling properties of the dependence coefficients (see Section 2.1). To use these properties, we need to consider a more restrictive type of dependence than for super smooth errors, and we need to impose a polynomial decrease of the coefficients.

In both cases, super and ordinary smooth, the results hold for β-mixing and τ-dependent random variables $(X_i)_{i\geq 1}$. To our knowledge, this is the first time that adaptive density deconvolution in a dependent context is considered. The robustness of this estimation procedure to dependency relies on the independence between $(X_i)_{1\leq i\leq n}$ and $(\varepsilon_i)_{i\leq 1\leq n}$, and the fact that the errors are i.i.d. random variables. We refer to Comte *et al.* (2005, 2006) for practical implementation of the estimators, and for the calibration of the constants in the penalty functions. In Comte *et al.* (2005), the robustness of the procedure to various dependency has been experimented in practice (see Tables 4 and 5 therein).

2. Some measures of dependence

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let Y be a random variable with values in a Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$, and let $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$. Let $\mathbb{P}_{Y|\mathcal{M}}$ be a conditional distribution of Y given $\mathcal{M}$, and let P_Y be the distribution of Y. Let $\mathcal{B}(\mathbb{B})$ be the borel σ-algebra on $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$, and let $\Lambda_1(\mathbb{B})$ be the set of 1-Lipschitz functions from $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ to $\mathbb{R}$. Define now

$$\begin{aligned} \beta(\mathcal{M}, \sigma(Y)) &= \mathbb{E}\Big(\sup_{A\in\mathcal{B}(\mathcal{X})} |\mathbb{P}_{Y|\mathcal{M}}(A) - \mathbb{P}_Y(A)| \Big), \\ \text{and if } \mathbb{E}(\|Y\|) < \infty, \quad \tau(\mathcal{M}, Y) &= \mathbb{E}\Big(\sup_{f\in\Lambda_1(\mathbb{B})} |\mathbb{P}_{Y|\mathcal{M}}(f) - \mathbb{P}_Y(f)| \Big). \end{aligned}$$

The coefficient $\beta(\mathcal{M}, \sigma(Y))$ is the usual mixing coefficient, introduced by Rozanov and Volkonskii (1960). The coefficient $\tau(\mathcal{M}, Y)$ has been introduced by Dedecker and Prieur (2005).

Let $\mathbf{X} = (X_i)_{i\geq 1}$ be a strictly stationary sequence of real-valued random variables. For any $k \geq 0$, the coefficients $\beta_{\mathbf{X},1}(k)$ and $\tau_{\mathbf{X},1}(k)$ are defined by

$$\begin{aligned} \beta_{\mathbf{X},1}(k) &= \beta(\sigma(X_1), \sigma(X_{1+k})), \\ \text{and if } \mathbb{E}(|X_1|) < \infty, \quad \tau_{\mathbf{X},1}(k) &= \tau(\sigma(X_1), X_{1+k}). \end{aligned}$$

On $\mathbb{R}^l$, we put the norm $\|x-y\|_{\mathbb{R}^l} = l^{-1}(|x_1-y_1|+\cdots+|x_l-y_l|)$. Let $\mathcal{M}_i = \sigma(X_k, 1 \le k \le i)$. The coefficients $\beta_{\mathbf{X},\infty}(k)$ and $\tau_{\mathbf{X},\infty}(k)$ are defined by

$$\begin{aligned} \beta_{\mathbf{X},\infty}(k) &= \sup_{i\ge 1, l\ge 1} \sup\left\{\beta(\mathcal{M}_i, \sigma(X_{i_1},\ldots,X_{i_l})), i+k \le i_1 < \cdots < i_l\right\}, \\ \text{and if } \mathbb{E}(|X_1|)<\infty, \quad \tau_{\mathbf{X},\infty}(k) &= \sup_{i\ge 1, l\ge 1} \sup\left\{\tau(\mathcal{M}_i, (X_{i_1},\ldots,X_{i_l})), i+k \le i_1 < \cdots < i_l\right\}. \end{aligned}$$

2.1. **Coupling.** We recall the coupling properties of these coefficients. Assume that Ω is rich enough, which means that there exists U uniformly distributed over $[0,1]$ and independent of $\mathcal{M}\vee\sigma(X)$. There exist two $\mathcal{M}\vee\sigma(U)\vee\sigma(X)$-measurable random variables X_1^* and X_2^* distributed as X and independent of $\mathcal{M}$ such that

$$\beta(\mathcal{M},\sigma(X)) = \mathbb{P}(X \ne X_1^*) \quad \text{and} \quad \tau(\mathcal{M}, X) = \mathbb{E}(\|X - X_2^*\|_{\mathbb{B}}). \tag{2.1}$$

The first equality in (2.1) is due to Berbee (1979), and the second one has been established in Dedecker and Prieur (2005), Section 7.1.

2.2. **Covariance inequalities.** Denote by $\|\cdot\|_{\infty,\mathbb{P}}$ the $\mathbb{L}^\infty(\Omega,\mathbb{P})$-norm. Let X, Y be two real-valued random variables, and let f, h be two measurable functions from $\mathbb{R}$ to $\mathbb{C}$. Then

$$|\mathrm{Cov}(f(Y), h(X))| \le 2\|f(Y)\|_{\infty,\mathbb{P}}\|h(X)\|_{\infty,\mathbb{P}}\,\beta(\sigma(X),\sigma(Y)), \tag{2.2}$$

and if $\mathrm{Lip}(h)$ is the Lipschitz coefficient of h,

$$|\mathrm{Cov}\ (f(Y), h(X))| \le \|f(Y)\|_{\infty,\mathbb{P}}\mathrm{Lip}(h)\,\tau(\sigma(Y), X). \tag{2.3}$$

Inequalities (2.2) and (2.3) follow from the coupling properties (2.1) by noting that if X^* is distributed as X and independent of Y,

$$\mathrm{Cov}\ (f(Y), h(X)) = \mathbb{E}(\overline{f(Y)}(h(X) - h(X^*))).$$

2.3. **Examples.** Examples of β-mixing sequences are well known (we refer to the books by Doukhan (1994) and Bradley (2002)). One of the most important examples is the following: a stationary, irreducible, aperiodic and positively recurrent Markov chain $(X_i)_{i\ge1}$ is β-mixing, which means that $\beta_{\mathbf{X},\infty}(k)$ tends to zero as k tends to infinity.

Unfortunately, many simple Markov chains are not β-mixing (and not even strongly mixing in the sense of Rosenblatt (1956)). For instance, if $(\epsilon_i)_{i\ge1}$ is i.i.d. with marginal $\mathcal{B}(1/2)$, then the stationary solution $(X_i)_{i\ge0}$ of the equation

$$X_n = \frac{1}{2}(X_{n-1}+\epsilon_n), \quad X_0 \text{ independent of } (\epsilon_i)_{i\ge1} \tag{2.4}$$

is not β-mixing (and not even strongly mixing) since $\beta_{\mathbf{X},1}(k) = 1$ for any $k \ge 0$. By contrast, for this particular example, one has $\tau_{\mathbf{X},\infty}(k) \le 2^{-k}$. More generally, the coefficient $\tau_{\mathbf{X},\infty}(k)$ is easy to compute in many situations (see Dedecker and Prieur (2005)). Let us recall some important examples:

Linear processes. Assume that $X_i = \sum_{j\ge0} a_j \xi_{n-j}$, where $(\xi_i)_{i\in\mathbb{Z}}$ is i.i.d. One has the bounds

$$\tau_{\mathbf{X},\infty}(k) \le 2\mathbb{E}(|\xi_0|)\sum_{j\ge k}|a_j| \quad \text{and} \quad \tau_{\mathbf{X},\infty}(k) \le \sqrt{2\mathrm{Var}(\xi_0)\sum_{j\ge k} a_j^2}.$$

Markov chains. Let $(X_n)_{n\geq 0}$ be a stationary Markov chain such that $X_n = F(X_{n-1}, \xi_n)$ for some measurable function F and some i.i.d. sequence $(\xi_i)_{i\geq 1}$ independent of X_0. Assume that there exists $\kappa < 1$ such that

$$\mathbb{E}(|F(x,\xi_0) - F(y,\xi_0)|) \leq a|x-y|\,.$$

Then one has the inequality

$$\tau_{\mathbf{X},\infty}(k) \leq 2\mathbb{E}(|X_0|)a^k\,.$$

An important example is $X_n = f(X_{n-1}) + \xi_n$ for some a-lipschitz function f.

Expanding maps. Let T be a Borel-measurable map from $[0,1]$ to $[0,1]$. If the probability μ is invariant by T, the sequence $(Y_i = T^i)_{i\geq 0}$ of random variables from $([0,1],\mu)$ to $[0,1]$ is strictly stationary. Define the operator K from $\mathbb{L}^1([0,1],\mu)$ to $\mathbb{L}^1([0,1],\mu)$ *via* the equality

$$\int_0^1 (Kh)(x)k(x)\mu(dx) = \int_0^1 h(x)(k\circ T)(x)\mu(dx)$$

where $h \in \mathbb{L}^1([0,1],\mu)$ and $k \in \mathbb{L}^\infty([0,1],\mu)$. It is easy to check that $(Y_1, Y_2, \ldots, Y_n)$ has the same distribution as $(X_n, X_{n-1}, \ldots, X_1)$ where $(X_i)_{i\in\mathbb{Z}}$ is a stationary Markov chain with invariant distribution μ and transition kernel K. If T is uniformly expanding (see for instance the assumptions on page 218 in Dedecker and Prieur (2005)), then there exist $C > 0$ and ρ in $]0,1[$ such that

$$\tau_{\mathbf{X},\infty}(k) \leq C\rho^k$$

(see Dedecker and Prieur page 230). Note that the Markov chain $(X_i)_{i\geq 1}$ is not β-mixing (and not even strongly mixing). Indeed $\beta(\sigma(X_1), \sigma(X_n)) = \beta(\sigma(T^n), \sigma(T))$. Since $\sigma(T^n) \subset \sigma(T)$, it follows that

$$\beta(\sigma(X_1),\sigma(X_n)) \geq \beta(\sigma(T^n),\sigma(T^n)) = \beta(\sigma(T),\sigma(T))$$

and the later is positive as soon as μ is non trivial.

3. Assumptions and estimators

For two complex-valued functions u and v in $\mathbb{L}_2(\mathbb{R}) \cap \mathbb{L}_1(\mathbb{R})$, let

$$u^*(x) = \int e^{itx}u(t)dt, \quad u * v(x) = \int u(y)v(x-y)dy, \text{ and } \quad < u, v >= \int u(x)\overline{v}(x)dx$$

with $\overline{z}$ the conjugate of a complex number z. We also use the notations

$$\|u\|_1 = \int |u(x)|dx, \quad \|u\|^2 = \int |u(x)|^2dx, \quad \text{and} \quad \|u\|_\infty = \sup_{x\in\mathbb{R}} |u(x)|.$$

3.1. **Assumptions for density deconvolution.** The smoothness of f_ε is described by the following assumption.

There exist nonnegative numbers $\kappa_0, \gamma,\ \mu,$ and δ such that f_ε^* satisfies

$(\mathbf{A}_1^\varepsilon)$ $$\kappa_0(x^2+1)^{-\gamma/2}\exp\{-\mu|x|^\delta\} \leq |f_\varepsilon^*(x)| \leq \kappa_0'(x^2+1)^{-\gamma/2}\exp\{-\mu|x|^\delta\}.$$

$(\mathbf{A}_2^\varepsilon)$ The density f_ε belongs to $\mathbb{L}_2(\mathbb{R})$ and for all $x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$.

Since f_ε is known, the constants $\mu, \delta, \kappa_0,$ and γ defined in $(\mathbf{A}_1^\varepsilon)$ are also known.

When $\delta = 0$ in $(\mathbf{A}_1^\varepsilon)$, f_ε is usually called "ordinary smooth". When $\mu > 0$ and $\delta > 0$, f_ε is called "super smooth". Densities satisfying $(\mathbf{A}_1^\varepsilon)$ with $\delta > 0$ and $\mu > 0$ are infinitely differentiable. The standard examples for super smooth densities are the following: Gaussian or Cauchy distributions are super smooth of order $\gamma = 0, \delta = 2$ and $\gamma = 0, \delta = 1$ respectively. When $\varepsilon = \ln(\eta^2)$ with $\eta \sim \mathcal{N}(0,1)$ as in Van Es *et al.* (2003, 2005), then ε is super-smooth with $\delta = 1, \gamma = 0$ and $\mu = \pi/2$. For ordinary smooth densities, one can cite for instance the double exponential (also called Laplace) distribution with $\delta = 0 = \mu$ and $\gamma = 2$. Although densities with $\delta > 2$ exist, they are difficult to express in a closed form. Nevertheless, our results hold for such densities. Furthermore, the square integrability of f_ε in $(\mathbf{A}_2^\varepsilon)$ require that $\gamma > 1/2$ when $\delta = 0$ in $(\mathbf{A}_1^\varepsilon)$.

Classically, the slowest rates of convergence for estimating g are obtained for super smooth error densities. In particular, when ε is Gaussian and g belongs to Sobolev classes, the minimax rates are negative powers of $\ln(n)$ (see Fan (1991)). Nevertheless, the rates are improved if g has stronger smoothness properties, described by the set

$$\mathcal{S}_{s,r,b}(C_1) = \left\{\psi \text{ such that } \int_{-\infty}^{+\infty} |\psi^*(x)|^2 (x^2+1)^s \exp\{2b|x|^r\} dx \le C_1\right\} \tag{3.1}$$

for s, r, b non-negative numbers.

Such smoothness classes are classically considered both in deconvolution and in density estimation without errors. When $r = 0$, (3.1) corresponds to a Sobolev ball. The functions in (3.1) with $r > 0$ and $b > 0$ are infinitely many times differentiable. They admit analytic continuation on a finite width strip when $r = 1$ and on the whole complex plane if $r = 2$.

Subsequently, the density g is supposed to satisfy the following assumption.

$(\mathbf{A}_3^X)$ The density $g \in \mathbb{L}_2(\mathbb{R})$ and there exists $M_2 > 0$, such that $\displaystyle\int x^2 g^2(x)dx < M_2 < \infty.$

Assumption $(\mathbf{A}_3^X)$ which is due to the construction of the estimator, is quite unusual in density estimation. It already appears in density deconvolution in the independent framework in Comte *et al.* (2005, 2006). It also appears in a slightly different way in Pensky and Vidakovic (1999) who assume, instead of $(\mathbf{A}_3^X)$ that $\sup_{x\in\mathbb{R}} |x|g(x) < \infty$. It is important to note that Assumption $(\mathbf{A}_3^X)$ is very unrestrictive.

All densities having tails of order $|x|^{-(s+1)}$ as x tends to infinity satisfy $(\mathbf{A}_3^X)$ only if $s > 1/2$. One can cite for instance the Cauchy distribution or all stable distributions with exponent $r > 1/2$ (see Devroye (1986)). The Lévy distribution, with exponent $r = 1/2$ does not satisfies $(\mathbf{A}_3^X)$.

3.2. **The projection spaces.** Let $\varphi(x) = \sin(\pi x)/(\pi x)$. For $m \in \mathbb{N}$ and $j \in \mathbb{Z}$, set $\varphi_{m,j}(x) = \sqrt{m}\varphi(mx - j)$. The functions $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ constitute an orthonormal system in $\mathbb{L}^2(\mathbb{R})$ (see e.g. Meyer (1990), p.22). For $m = 2^k$, it is known as the Shannon basis. Though we choose here integer values for m, a thinner grid would also be possible. Let us define

$$S_m = \overline{\text{span}}\{\varphi_{m,j},\ j \in \mathbb{Z}\},\ m \in \mathbb{N}.$$

The space S_m is exactly the subspace of $\mathbb{L}_2(\mathbb{R})$ of functions having a Fourier transform with compact support contained in $[-\pi m, \pi m]$.

The orthogonal projections of g on S_m is $g_m = \sum_{j\in\mathbb{Z}} a_{m,j}(g)\varphi_{m,j}$ where $a_{m,j}(g) =< \varphi_{m,j}, g >$. To obtain representations having a finite number of "coordinates", we introduce

$$S_m^{(n)} = \overline{\text{span}}\,\{\varphi_{m,j}, |j| \le k_n\}$$

with integers k_n to be specified later. The family $\{\varphi_{m,j}\}_{|j|\le k_n}$ is an orthonormal basis of $S_m^{(n)}$ and the orthogonal projections of g on $S_m^{(n)}$ is given by $g_m^{(n)} = \sum_{|j|\le k_n} a_{m,j}(g)\varphi_{m,j}$.

3.3. **Construction of the minimum contrast estimators.** For an arbitrary fixed integer m, an estimator of g belonging to $S_m^{(n)}$ is defined by

$$\hat{g}_m^{(n)} = \arg\min_{t \in S_m^{(n)}} \gamma_n(t), \tag{3.2}$$

where, for t in $S_m^{(n)}$,

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\|t\|^2 - 2u_t^*(Z_i)\right], \quad \text{with} \quad u_t(x) = \frac{1}{2\pi}\left(\frac{t^*(-x)}{f_\varepsilon^*(x)}\right).$$

By using Parseval and inverse Fourier formulae we obtain that $\mathbb{E}\left[u_t^*(Z_i)\right] = \langle t, g\rangle$, so that $\mathbb{E}(\gamma_n(t)) = \|t-g\|^2 - \|g\|^2$ is minimal when $t = g$. This shows that $\gamma_n(t)$ suits well for the estimation of g. Classical calculations show that

$$\hat{g}_m^{(n)} = \sum_{|j|\le k_n} \hat{a}_{m,j}\varphi_{m,j} \text{ with } \hat{a}_{m,j} = \frac{1}{n}\sum_{i=1}^{n} u^*_{\varphi_{m,j}}(Z_i), \quad \text{and} \quad \mathbb{E}(\hat{a}_{m,j}) = < g, \varphi_{m,j} > = a_{m,j}.$$

3.4. **Minimum penalized contrast estimator.** As in the independent framework, the minimum penalized estimator of g is defined as $\tilde{g} = \hat{g}_{\hat{m}_g}$ where $\hat{m}_g$ is chosen in a purely data-driven way. The main point of the estimation procedure lies in the choice of $m = \hat{m}_g$ for the estimators $\hat{g}_m$ from Section 3.3 in order to mimic the oracle parameter

$$\breve{m}_g = \arg\min_m \mathbb{E} \parallel \hat{g}_m - g \parallel_2^2 . \tag{3.3}$$

The model selection is performed in an automatic way, using the following penalized criteria

$$\tilde{g} = \hat{g}_{\hat{m}}^{(n)} \text{ with } \hat{m} = \arg\min_{m\in\{1,\cdots,m_n\}} \left[\gamma_n(\hat{g}_m^{(n)}) + \text{ pen}(m)\right], \tag{3.4}$$

where $\text{pen}(m)$ is a penalty function, precised in the Theorems, that depends on f_ε^* through $\Delta(m)$ defined by

$$\Delta(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(x)|^2} dx. \tag{3.5}$$

The key point in the dependent context is to find a penalty function not depending on the mixing coefficients such that

$$\mathbb{E} \parallel \tilde{g} - g \parallel^2 \le C \inf_{m\in\{1,\cdots,m_n\}} \mathbb{E} \parallel \hat{g}_m - g \parallel^2 .$$

4. Risk bounds for the minimum contrast estimators $\hat{g}_m^{(n)}$

We focus here on non adaptive estimation, starting with the presentation of general upper bounds for MISEs of the minimum contrast estimators $\hat{g}_m^{(n)}$.

Proposition 4.1. *If* $(\mathbf{A}_2^\varepsilon)$ *and* $(\mathbf{A}_3^X)$ *hold, then*

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \le \|g - g_m\|^2 + \frac{m^2(M_2+1)}{k_n} + \frac{2\Delta(m)}{n} + \frac{2R_m}{n},$$

where

$$R_m = \frac{1}{\pi} \sum_{k=2}^{n} \int_{-\pi m}^{\pi m} \left| \mathrm{Cov}\left(e^{ixX_1}, e^{ixX_k}\right) \right| dx. \tag{4.1}$$

Moreover, $R_m \le \min(R_{m,\beta}, R_{m,\tau})$, *where*

$$R_{m,\beta} = 4m \sum_{k=1}^{n-1} \beta_{\mathbf{X},1}(k) \quad \text{and} \quad R_{m,\tau} = \pi m^2 \sum_{k=1}^{n-1} \tau_{\mathbf{X},1}(k)\,.$$

Remark 4.1. The term R_m can be easily bounded for many other dependent sequences. For instance, if $\alpha_{\mathbf{X},1} = \alpha(\sigma(X_1), \sigma(X_{1+k}))$ is the usual strong mixing coefficient of Rosenblatt (1956), one has the upper bound $R_m \le 16m \sum_{k=1}^{n-1} \alpha_{\mathbf{X},1}(k)$. If $\mathbf{X}$ is a stationary sequence of associated random variables (see Esary *et al.* (1967) for the definition), then $|\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})| \le 4x^2 \mathrm{Cov}(X_1, X_k)$, so that $R_m \le (8\pi^2/3) m^3 \sum_{k=2}^{n} \mathrm{Cov}(X_1, X_k)$. For general treatment in this case, see Marsy (2003).

We now comment the rates resulting from Proposition 4.1. As usual, the variance term $n^{-1}\Delta(m)$ depends on the rate of decay of the Fourier transform of f_ε. According to Lemma 7.2 and according to Butucea and Tsybakov (2005), under $(\mathbf{A}_1^\varepsilon)$-$(\mathbf{A}_2^\varepsilon)$, we have

$$\begin{aligned} &\lambda_1(f_\varepsilon, \kappa_0')\Gamma(m)(1+o(1)) \le \Delta(m) \le \lambda_1(f_\varepsilon, \kappa_0)\Gamma(m)(1+o(1)) \quad \text{as } m \to \infty \\ &\text{where} \quad \Gamma(m) = (1+(\pi m)^2)^\gamma (\pi m)^{1-\delta} \exp\left\{2\mu(\pi m)^\delta\right\}, \end{aligned} \tag{4.2}$$

$$\lambda_1(f_\varepsilon, \kappa_0) = \frac{1}{\kappa_0^2 \pi R(\mu, \delta)}, \text{ and } R(\mu,\delta) = \mathbb{1}_{\{\delta=0\}} + 2\mu\delta \mathbb{1}_{\{\delta>0\}}. \tag{4.3}$$

If $(\mathbf{A}_1^\varepsilon)$-$(\mathbf{A}_2^\varepsilon)$ and $(\mathbf{A}_3^X)$ hold, and if $k_n \ge n$, we have the upper bound

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \le \|g - g_m\|^2 + \frac{m^2(M_2+1)}{n} + \frac{2\lambda_1(f_\varepsilon, \kappa_0)\Gamma(m)}{n} + \frac{2R_m}{n}. \tag{4.4}$$

Finally, since g_m is the orthogonal projection of g on S_m, we get that $g_m^* = g^* \mathbb{1}_{[-m\pi, m\pi]}$ and therefore

$$\|g - g_m\|^2 = \frac{1}{2\pi}\|g^* - g_m^*\|^2 = \frac{1}{2\pi} \int_{|x| \ge \pi m} |g^*|^2(x) dx.$$

If g belongs to the class $\mathcal{S}_{s,r,b}(C_1)$ defined in (3.1), then

$$\|g - g_m\|^2 \le \frac{C_1}{2\pi}(m^2\pi^2+1)^{-s} \exp\{-2b\pi^r m^r\}.$$

Hence, according to (4.4), if $(\mathbf{A}_3^X)$ holds and $k_n \ge n$, the risk of $\hat{g}_m^{(n)}$ is bounded by

$$\frac{C_1}{2\pi}(m^2\pi^2+1)^{-s} \exp\{-2b\pi^r m^r\} + \frac{2\lambda_1(f_\varepsilon, \kappa_0)(1+(\pi m)^2))^\gamma (\pi m)^{1-\delta} \exp\left\{2\mu\pi^\delta m^\delta\right\}}{n} + \frac{m^2(M_2+1)}{n} + \frac{2R_m}{n}.$$

Assume now that either $\sum_{k>0} \beta_{\mathbf{X},1}(k) < \infty$ or $\sum_{k>0} \tau_{\mathbf{X},1}(k) < \infty$, so that the residual terms $n^{-1}R_m + n^{-1}m^2(M_2+1)$ are of order $n^{-1}m^2$. As in the independent case, we choose $\breve{m}$ as the minimizer of

$$(m^2\pi^2+1)^{-s}\exp\{-2b\pi^r m^r\}+\frac{(\pi m)^{2\gamma+1-\delta}\exp\left\{2\mu\pi^\delta m^\delta\right\}}{n}.$$

The behavior of $\breve{m}$ is recalled in Table 1. We see that in all cases, the residual terms $n^{-1}R_{\breve{m}}+n^{-1}\breve{m}^2(M_2+1)$ of order $n^{-1}\breve{m}^2$ are negligible with respect to the main terms since $n^{-1}\Delta(m)$ grows faster than $n^{-1}m^2$ (recall that if $\delta=0$, we have the restriction $\gamma>1/2$ (cf. Section 3.1)). Hence the rate of convergence of $\hat{g}^{(n)}_{\breve{m}}$ is the same as in the i.i.d. case (see Table 1 below).

Table 1. Choice of $\breve{m}$ and corresponding rates under Assumptions $(\mathbf{A}_1^\varepsilon)$-$(\mathbf{A}_2^\varepsilon)$ and (3.1).

		f_ε: $\delta=0$ ordinary smooth	f_ε: $\delta>0$ supersmooth
g	$r=0$ Sobolev(s)	$\pi\breve{m}=O(n^{1/(2s+2\gamma+1)})$ rate $=O(n^{-2s/(2s+2\gamma+1)})$ *minimax rate*	$\pi\breve{m}=[\ln(n)/(2\mu+1)]^{1/\delta}$ rate $=O((\ln(n))^{-2s/\delta})$ *minimax rate*
	$r>0$ $\mathcal{C}^\infty$	$\pi\breve{m}=[\ln(n)/2b]^{1/r}$ rate $=O\left(\frac{\ln(n)^{(2\gamma+1)/r}}{n}\right)$ *minimax rate*	$\breve{m}$ solution of $\breve{m}^{2s+2\gamma+1-r}\exp\{2\mu(\pi\breve{m})^\delta+2b\pi^r\breve{m}^r\}$ $=O(n)$ *minimax rate if* $r<\delta$ *and* $s=0$

When $r>0,\delta>0$ the value of $\breve{m}$ is not explicitly given. It is obtained as the solution of the equation

$$\breve{m}^{2s+2\gamma+1-r}\exp\{2\mu(\pi\breve{m})^\delta+2b\pi^r\breve{m}^r\}=O(n).$$

Consequently, the rate of $\hat{g}^{(n)}_{\breve{m}}$ is not explicit and depends on the ratio r/δ. If r/δ or δ/r belongs to $]k/(k+1);(k+1)/(k+2)]$ with k integer, the rate of convergence can be expressed as a function of k. We refer to Comte *et al.* (2006) for further discussions about those rates. We refer to Lacour (2006) for explicit formulae for the rates in the special case $r>0,\delta>0$.

5. Risk bounds for adaptive estimators

In the previous section, the construction of the estimators require the knowledge of the smoothness of g. We now come to adaptive estimation, without such prior knowledge.

5.1. **A first bound in adaptive density deconvolution.** Theorem 5.1 gives a general bound which holds under mild dependence conditions, for f_ε being either ordinary or super smooth. For $a>1$, let $\mathrm{pen}(m)$ be defined by

$$\mathrm{pen}(m)=\begin{cases}24a\dfrac{\Delta(m)}{n} & \text{if } 0\le\delta<1/3,\\[2mm] 8a\left(1+\dfrac{48\mu\pi^\delta\lambda_2(f_\varepsilon,\kappa_0)}{\lambda_1(f_\varepsilon,\kappa_0')}\right)\dfrac{\Delta(m)\,m^{\min((3\delta/2-1/2)_+,\delta))}}{n} & \text{if } \delta\ge1/3.\end{cases}\tag{5.5}$$

The constant $\lambda_1(f_\varepsilon,\kappa_0)$ is defined in (4.3) and $\lambda_2(f_\varepsilon,\kappa_0)$ is given by

$$\lambda_2(f_\varepsilon,\kappa_0)=\| f_\varepsilon \| \kappa_0^{-1}\sqrt{2\lambda_1(f_\varepsilon,\kappa_0)}\mathbb{I}_{0\le\delta\le 1}+2\lambda_1(f_\varepsilon,\kappa_0)\mathbb{I}_{\delta>1}. \tag{5.6}$$

In order to bound up $\mathrm{pen}(m)$, we impose that

$$\pi m_n \le \begin{cases} n^{1/(2\gamma+1)} & \text{if } \delta=0 \\ \left[\dfrac{\ln(n)}{2\mu}+\dfrac{2\gamma+1-\delta}{2\delta\mu}\ln\left(\dfrac{\ln(n)}{2\mu}\right)\right]^{1/\delta} & \text{if } \delta>0. \end{cases} \tag{5.7}$$

Subsequently we set

$$\kappa_a=(a+1)/(a-1), \text{ and } C_a=\max(\kappa_a^2, 2\kappa_a). \tag{5.8}$$

Theorem 5.1. *Assume that f_ε satisfies ($\mathbf{A}_1^\varepsilon$)-($\mathbf{A}_2^\varepsilon$), that g satisfies ($\mathbf{A}_3^X$), and that m_n satisfies (5.7). Consider the collection of estimators $\hat g_m^{(n)}$ defined by (3.2) with $k_n\ge n$ and $1\le m\le m_n$. Let* $\mathrm{pen}(m)$ *be defined by (5.5). The estimator $\tilde g=\hat g_{\hat m}^{(n)}$ defined by (3.4) satisfies*

$$\mathbb{E}(\|g-\tilde g\|^2)\le C_a \inf_{m\in\{1,\cdots,m_n\}}\left[\|g-g_m\|^2+\mathrm{pen}(m)+\frac{m^2(M_2+1)}{n}\right]+\frac{\overline{C}(R_{m_n}+m_n)}{n},$$

where R_m is defined in (4.1), C_a is defined in (5.8), and $\overline{C}$ is a constant depending on f_ε and a.

Let us compare the rate of $\tilde g$ with the rate obtained in the independent framework. The term $\inf_{m\in\{1,\cdots,m_n\}}[\|g-g_m\|^2+\mathrm{pen}(m)+m^2(M_2+1)/n]$ corresponds to the rate of $\tilde g$ when all variables are i.i.d. The dependent context induces the additional term $n^{-1}(R_{m_n}+m_n)$. If the dependence coefficients are summable and the errors are super smooth, then $n^{-1}(R_{m_n}+m_n)$ is negligible and $\tilde g$ achieves the rate of the independent framework. If ε is ordinary smooth, the term $n^{-1}(R_{m_n}+m_n)$ may not be negligible and Theorem 5.1 is not precise enough.

5.2. **Adaptive density deconvolution for super smooth f_ε.** If ($\mathbf{A}_1^\varepsilon$)-($\mathbf{A}_2^\varepsilon$) hold for some $\delta>0$, we have the following corollary.

Corollary 5.1. *Assume that f_ε satisfies ($\mathbf{A}_1^\varepsilon$)-($\mathbf{A}_2^\varepsilon$) with $\delta>0$, that g satisfies ($\mathbf{A}_3^X$), and that m_n satisfies (5.7). Let* $\mathrm{pen}(m)$ *be defined by (5.5). Consider the collection of estimators $\hat g_m^{(n)}$ defined by (3.2) with $k_n\ge n$ and $1\le m\le m_n$.*

(1) *If $\sum_{k>0}\beta_{\mathbf{X},1}(k)<\infty$, the estimator $\tilde g=\hat g_{\hat m}^{(n)}$ defined by (3.4) satisfies*

$$\mathbb{E}(\|g-\tilde g\|^2) \le C_a \inf_{m\in\{1,\cdots,m_n\}}\left[\|g-g_m\|^2+\mathrm{pen}(m)+\frac{m^2(M_2+1)}{n}\right]+\frac{\overline{C}(\ln(n))^{1/\delta}}{n},$$

where C_a is defined in (5.8) and $\overline{C}$ is a constant depending on f_ε, a and $\sum_{k>0}\beta_{\mathbf{X},1}(k)$.

(2) *If $\sum_{k>0}\tau_{\mathbf{X},1}(k)<\infty$, the estimator $\tilde g=\hat g_{\hat m}^{(n)}$ defined by (3.4) satisfies*

$$\mathbb{E}(\|g-\tilde g\|^2) \le C_a \inf_{m\in\{1,\cdots,m_n\}}\left[\|g-g_m\|^2+\mathrm{pen}(m)+\frac{m^2(M_2+1)}{n}\right]+\frac{\overline{C}(\ln(n))^{2/\delta}}{n},$$

where C_a is defined in (5.8) and $\overline{C}$ is a constant depending on f_ε, a and $\sum_{k>0}\tau_{\mathbf{X},1}(k)$.

Corollary 5.1 requires important comments. The terms involving power of $\ln(n)$ are negligible with respect to $\inf_{m\in\{1,\cdots,m_n\}}[\|g-g_m\|^2+\mathrm{pen}(m)+m^2(M_2+1)/n]$. The risk of $\tilde g$ is of order $\inf_{m\in\{1,\cdots,m_n\}}[\|g-g_m\|^2+\mathrm{pen}(m)]$, that is of the best order, as in the independent framework. The penalty does not depend on the dependence coefficients and is the same as in the independent framework.

As a conclusion, we see that the adaptive estimator $\tilde g$ built with the same penalty as in the independent framework, still achieves the best rates under mild conditions on the dependence coefficients.

5.3. **Adaptive density deconvolution for ordinary smooth f_ε.** For $a>1$, define $\mathrm{pen}(m)$ by

$$\mathrm{pen}(m)=\frac{25a\Delta(m)}{n}\,. \tag{5.9}$$

Theorem 5.2. *Assume that f_ε satisfies $(\mathbf{A}_1^\varepsilon)$-$(\mathbf{A}_2^\varepsilon)$ with $\delta=0$, that g satisfies $(\mathbf{A}_3^X)$, and that m_n satisfies (5.7). Let $\mathrm{pen}(m)$ be defined by (5.9). Consider the collection of estimators $\hat g_m^{(n)}$ defined by (3.2) with $k_n\ge n$ and $1\le m\le m_n$.*

1. *If $\beta_{\mathbf{X},\infty}(k)=O(k^{-(1+\theta)})$ for some $\theta>(2\gamma+3)/(2\gamma+1)$, then the estimator $\tilde g=\hat g_{\hat m}^{(n)}$ defined by (3.4) satisfies*

$$\mathbb{E}(\|g-\tilde g\|^2)\le C_a\inf_{m\in\{1,\cdots,m_n\}}\Big[\|g-g_m\|^2+\mathrm{pen}(m)+\frac{m^2(M_2+1)}{n}\Big]+\frac{\overline{C}}{n}, \tag{5.10}$$

where C_a is defined in (5.8) and $\overline{C}$ is a constant depending on f_ε, a, and $\sum_{k>0}\beta_{\mathbf{X},\infty}(k)$.

2. *If $\tau_{\mathbf{X},\infty}(k)=O(k^{-(1+\theta)})$ for some $\theta>(2\gamma+5)/(2\gamma+1)$, then the estimator $\tilde g=\hat g_{\hat m}^{(n)}$ defined by (3.4) satisfies (5.10), where $\overline{C}$ is a constant depending on f_ε, a and $\sum_{k>0}\tau_{\mathbf{X},\infty}(k)$.*

Remark 5.1. Note that the condition for $\beta_{\mathbf{X},\infty}(k)$ is realized for any $\gamma>1/2$ provided $\theta>2$. In the same way, the condition for $\tau_{\mathbf{X},\infty}(k)$ is realized for any $\gamma>1/2$ provided $\theta>3$. In both cases, the condition on θ is weaker as γ increases. In other words, the smoother is f_ε, the weaker is the condition on the dependence coefficients.

Remark 5.2. For m large enough, the penalty function given for ordinary smooth errors in Theorem 5.2 is an upper bound of more precise penalty functions which depend on the dependence coefficients. Under the assumptions of (1) in Theorem 5.2, let $\mathrm{pen}(m)$ be defined by

$$\mathrm{pen}(m)=\frac{24a\Delta(m)+128a\Big(1+4\sum_{k=1}^n\beta_{\mathbf{X},1}(k)\Big)m}{n}. \tag{5.11}$$

Under the assumptions of (2) in Theorem 5.2 let $\mathrm{pen}(m)$ be defined by

$$\mathrm{pen}(m)=\frac{24a\Delta(m)}{n}+\frac{64a\left[1+38\ln(m)\right]\left(m+\pi\sum_{k=1}^n\tau_{\mathbf{X},1}(k)m^2\right)}{n} \tag{5.12}$$

In both cases, the estimator $\tilde g=\hat g_{\hat m}^{(n)}$ defined by (3.4) satisfies (5.10). Remark 5.2 follows from the proof of Theorem 5.2.

5.4. **Case without noise.** One can deduce from Proposition 4.1, Theorem 5.2, its proof and Remark 5.2, a result for density estimation without errors, on the whole real line, that is when $\mathbf{X}$ is observed. If $\varepsilon = 0$, then we can consider that $Z = X$ and replace f_ε^* by 1. It follows that $u_t^*(Z_i) = t(X_i)$ and the contrast γ_n simply becomes

$$\gamma_{n,X}(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n t(X_i). \tag{5.13}$$

Let $k_n \geq n^2$, and consider as previously

$$\hat{g}_m^{(n)} = \arg\min_{t \in S_m^{(n)}} \gamma_{n,X}(t), \;\; \mathrm{pen}(m) = 128a\Big(1 + 4\sum_{k=1}^n \beta_{\mathbf{X},1}(k)\Big)\frac{m}{n}\,, \tag{5.14}$$

and

$$\hat{m} = \arg\min_{m \in \{1,\ldots,n\}} [\gamma_{n,X}(g_m^{(n)}) + \mathrm{pen}(m)]. \tag{5.15}$$

The following results follow straightforwardly.

Corollary 5.2. *Assume that $\varepsilon = 0$. Let $k_n \geq n^2$. Then*
1)

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \leq \|g - g_m\|^2 + \frac{m(M_2+3)}{n} + \frac{2R_m}{n}.$$

2) *If $\beta_{\mathbf{X},\infty} = O(k^{-(1+\theta)})$ for some $\theta > 3$, then the estimator $\tilde{g} = \hat{g}_{\hat{m}}$ defined by (5.14) and (5.15) satisfies*

$$\mathbb{E}(\|g - \tilde{g}\|^2) \leq C_a \inf_{m \in \{1,\cdots,n\}} \Big[\|g - g_m\|^2 + \mathrm{pen}(m) + \frac{m(M_2+1)}{n}\Big] + \frac{\overline{C}}{n},$$

where C_a is defined in (5.8) and $\overline{C}$ is a constant depending on a and $\sum_{k>0}\beta_{\mathbf{X},\infty}(k)$.

The result **1)** shows that if $\sum_{k>0}\beta_{\mathbf{X},1}(k) < \infty$, one obtains the same bounds (and the same rates) as in the i.i.d. case. However, if $\sum_{k>0}\tau_{\mathbf{X},1}(k) < \infty$ the term $n^{-1}R_m$ is of order $n^{-1}m^2$ and the rates for $\hat{g}_m^{(n)}$ are less good than in the i.i.d. case.

This result **2)** shows that this estimation procedure also works in density estimation without errors. It allows to estimate a density on the whole real line and to reach the usual rates of convergence, by using a penalty of the classical order m/n. This remark is valid in the β-mixing framework and in the case of independent X_i's. We refer to Pensky (1999) and Rigollet (2006) for recent results in adaptive density estimation on the whole real line in the i.i.d. case.

6. Proofs

6.1. **Proof of Proposition 4.1.** The proof of the proposition 4.1 follows the same lines as in the independent framework (see Comte *et al.* (2006)). The main difference lies in the control of the variance term. We keep the same notations as in Section 3.3. According to (3.2), for any given m belonging to $\{1, \cdots, m_n\}$, $\hat{g}_m^{(n)}$ satisfies, $\gamma_n(\hat{g}_m^{(n)}) - \gamma_n(g_m^{(n)}) \leq 0$. For a random variable Y with density f_Y, and any function ψ such that $\psi(Y)$ is integrable, let

$$\nu_{n,Y}(\psi) = \frac{1}{n}\sum_{i=1}^n [\psi(Y_i) - \langle \psi, f_Y\rangle], \text{ so that } \quad \nu_{n,Z}(u_t^*) = \frac{1}{n}\sum_{i=1}^n [u_t^*(Z_i) - \langle t, g\rangle]\,. \tag{6.1}$$

Since

$$\gamma_n(t)-\gamma_n(s)=\|t-g\|^2-\|s-g\|^2-2\nu_{n,Z}(u^*_{t-s}), \tag{6.2}$$

we infer that

$$\|g-\hat g_m^{(n)}\|^2\le \|g-g_m^{(n)}\|^2+2\nu_{n,Z}\left(u^*_{\hat g_m^{(n)}-g_m^{(n)}}\right). \tag{6.3}$$

Writing that $\hat a_{m,j}-a_{m,j}=\nu_{n,Z}(u^*_{\varphi_{m,j}})$, we obtain

$$\nu_{n,Z}\left(u^*_{\hat g_m^{(n)}-g_m^{(n)}}\right)=\sum_{|j|\le k_n}(\hat a_{m,j}-a_{m,j})\nu_{n,Z}(u^*_{\varphi_{m,j}})=\sum_{|j|\le k_n}[\nu_{n,Z}(u^*_{\varphi_{m,j}})]^2.$$

Consequently, $\mathbb{E}\|g-\hat g_m^{(n)}\|^2\le\|g-g_m^{(n)}\|^2+2\sum_{j\in\mathbb{Z}}\mathbb{E}[(\nu_{n,Z}(u^*_{\varphi_{m,j}}))^2]$. According to Comte *et al.* (2006),

$$\|g-g_m^{(n)}\|^2=\| g-g_m \|^2+\|g_m-g_m^{(n)}\|^2\le\| g-g_m \|^2+\frac{(\pi m)^2(M_2+1)}{k_n}. \tag{6.4}$$

The variance term is studied by using that for $f\in\mathbb{L}_1(\mathbb{R})$,

$$\nu_{n,Z}(f^*)=\int\nu_{n,Z}(e^{ix\cdot})f(x)dx. \tag{6.5}$$

Now, we use (6.5) and apply Parseval's formula to obtain

$$\mathbb{E}\left(\sum_{j\in\mathbb{Z}}(\nu_{n,Z}(u^*_{\varphi_{m,j}}))^2\right)=\frac{1}{4\pi^2}\sum_{j\in\mathbb{Z}}\mathbb{E}\Big(\int\frac{\varphi^*_{m,j}(-x)}{f^*_\varepsilon(x)}\nu_{n,Z}(e^{ix\cdot})dx\Big)^2=\frac{1}{2\pi}\int_{-\pi m}^{\pi m}\frac{\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2}{|f^*_\varepsilon(x)|^2}dx. \tag{6.6}$$

Since $\nu_{n,Z}$ involves centered and stationary variables,

$$\begin{aligned}\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2 &= \mathrm{Var}|\nu_{n,Z}(e^{ix\cdot})|=\frac{1}{n^2}\left(\sum_{k=1}^n\mathrm{Var}(e^{ixZ_k})+\sum_{1\le k\ne l\le n}\mathrm{Cov}(e^{ixZ_k},e^{ixZ_l})\right)\\ &= \frac1n\mathrm{Var}(e^{ixZ_1})+\frac{1}{n^2}\sum_{1\le k\ne l\le n}\mathrm{Cov}(e^{ixZ_k},e^{ixZ_l}). \end{aligned}\tag{6.7}$$

Since $(X_i)_{i\ge1}$ and $(\varepsilon_i)_{i\ge1}$ are independent, we have $\mathbb{E}(e^{ixZ_k})=f^*_\varepsilon(x)g^*(x)$ so that

$$\mathrm{Cov}(e^{ixZ_k},e^{ixZ_l})=\mathbb{E}(e^{ix(Z_l-Z_k)})-|\mathbb{E}(e^{ixZ_k})|^2=\mathbb{E}(e^{ix(Z_l-Z_k)})-|f^*_\varepsilon(x)g^*(x)|^2.$$

Next, by independence of X and ε, we write, for $k\ne l$,

$$\mathbb{E}(e^{ix(Z_l-Z_k)})=\mathbb{E}(e^{ix(X_l-X_k)})\mathbb{E}(e^{ix(\varepsilon_l-\varepsilon_k)})=\mathbb{E}(e^{ix(X_l-X_k)})|f^*_\varepsilon(x)|^2,$$

and consequently

$$\mathrm{Cov}(e^{ixZ_k},e^{ixZ_l})=\mathrm{Cov}(e^{ixX_k},e^{ixX_l})|f^*_\varepsilon(x)|^2. \tag{6.8}$$

From (6.7), (6.8) and the stationarity of $(X_i)_{i\ge1}$, we obtain that

$$\mathbb{E}|\nu_{n,Z}(e^{ix\cdot})|^2\le\frac1n+\frac2n\sum_{k=2}^n\left|\mathrm{Cov}(e^{ixX_1},e^{ixX_k})\right||f^*_\varepsilon(x)|^2. \tag{6.9}$$

The first part of Proposition 4.1 follows from the stationarity of the X_i's, and from (6.3), (6.4), (6.6) and (6.9).

Let us prove that $R_m \leq \min(R_{m,\beta}, R_{m,\tau})$, where $R_{m,\beta}$ and $R_{m,\tau}$ are defined in Proposition 4.1. Using the inequalities (2.2) and (2.3), we obtain the bounds

$$|\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})| \leq 2\beta_{\mathbf{X},1}(k-1) \quad \text{and } |\mathrm{Cov}(e^{ixX_1}, e^{ixX_k})| \quad \leq \quad |x|\tau_{\mathbf{X},1}(k-1)$$

(for the last inequality, note that $t \to e^{ixt}$ is $|x|$-Lipschitz). The result easily follows.

6.2. **Proof of Theorem 5.1.** By definition, $\tilde{g}$ satisfies that for all $m \in \{1, \cdots, m_n\}$,

$$\gamma_n(\tilde{g}) + \mathrm{pen}(\hat{m}) \leq \gamma_n(g_m) + \mathrm{pen}(m).$$

Therefore, by using (6.2) we get that

$$\|\tilde{g} - g\|^2 \leq \|g_m^{(n)} - g\|^2 + 2\nu_{n,Z}(u^*_{\tilde{g}-g_m^{(n)}}) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}).$$

If $t = t_1 + t_2$ with t_1 in $S_m^{(n)}$ and t_2 in $S_{m'}^{(n)}$, t^* has its support in $[-\pi\max(m,m'), \pi\max(m,m')]$ and t belongs to $S^{(n)}_{\max(m,m')}$. Set $B_{m,m'}(0,1) = \{t \in S^{(n)}_{\max(m,m')} \ / \ \|t\| = 1\}$. For $\nu_{n,Z}$ defined in (6.1) we get

$$|\nu_{n,Z}(u^*_{\tilde{g}-g_m^{(n)}})| \leq \|\tilde{g} - g_m^{(n)}\| \sup_{t \in B_{m,\hat{m}}(0,1)} |\nu_{n,Z}(u_t^*)|.$$

Using that $2uv \leq a^{-1}u^2 + av^2$ for any $a > 1$, leads to

$$\|\tilde{g} - g\|^2 \quad \leq \quad \|g_m^{(n)} - g\|^2 + a^{-1}\|\tilde{g} - g_m^{(n)}\|^2 + a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,Z}(u_t^*))^2 + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}).$$

Now, according to Lemma 7.1, write that $\nu_{n,Z}(u_t^*) = \nu_n^{(1)}(t) + \nu_{n,X}(t)$, where

$$\nu_n^{(1)}(t) = n^{-1}\sum_{i=1}^n [u_t^*(Z_i) - \mathbb{E}(u_t^*(Z_i)|\sigma(X_i,\, i \geq 1))] = n^{-1}\sum_{i=1}^n [u_t^*(Z_i) - t(X_i)]. \tag{6.10}$$

Consequently,

$$\begin{aligned}\|\tilde{g} - g\|^2 \quad \leq \quad & \|g_m^{(n)} - g\|^2 + a^{-1}\|\tilde{g} - g_m^{(n)}\|^2 + 2a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_n^{(1)}(t))^2 + 2a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 \\ & + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}).\end{aligned}$$

Hence by writing that $\|\tilde{g} - g_m^{(n)}\|^2 \leq (1 + \kappa_a^{-1})\|\tilde{g} - g\|^2 + (1 + \kappa_a)\|g - g_m^{(n)}\|^2$ with κ_a defined in (5.8), we have

$$\begin{aligned}\|\tilde{g} - g\|^2 \quad \leq \quad & \kappa_a^2\|g_m^{(n)} - g\|^2 + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_n^{(1)}(t))^2 + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 \\ & + \kappa_a(\mathrm{pen}(m) - \mathrm{pen}(\hat{m})).\end{aligned}$$

Choose some positive function $p(m, m')$ such that

$$2ap(m, m') \leq \mathrm{pen}(m) + \mathrm{pen}(m'). \tag{6.11}$$

For this function $p(m, m')$ we have

$$\begin{aligned}\|\tilde{g} - g\|^2 \quad & \leq \quad \kappa_a^2\|g - g_m^{(n)}\|^2 + 2\kappa_a\mathrm{pen}(m) + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 + 2a\kappa_a W_n(m, \hat{m}) \\ & \leq \quad \kappa_a^2\|g - g_m^{(n)}\|^2 + 2\kappa_a\mathrm{pen}(m) + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 + 2a\kappa_a \sum_{m'=1}^{m_n} W_n(m, m'), \qquad (6.12)\end{aligned}$$

where

$$W_n(m,m') := \Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - p(m,m')\Big]_+, \tag{6.13}$$

The main parts of the proof lies in the two following points :

1) Study of $W_n(m,m')$, and more precisely find $p(m,m')$ such that for a constant A_1,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_n(m,m')) \leq \frac{A_1}{n}. \tag{6.14}$$

2) Study of $\sup_{t\in B_{m,\hat{m}}(0,1)}(\nu_{n,X}(t))^2$ and more precisely prove that

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat{m}}(0,1)}(\nu_{n,X}(t))^2\Big] \leq \frac{m_n+R_{m_n}}{n}, \tag{6.15}$$

where R_m is defined in (4.1). Combining (6.12), (6.14) and (6.15), we infer that, for all $1\leq m\leq m_n$

$$\mathbb{E}\|g-\tilde{g}\|^2 \leq \kappa_a^2\|g-g_m^{(n)}\|^2 + 2\kappa_a \mathrm{pen}(m) + \frac{2a\kappa_a(m_n+R_{m_n})}{n} + \frac{2a\kappa_a A_1}{n}.$$

If we denote by $C_a = \max(\kappa_a^2, 2\kappa_a)$, this can also be written

$$\begin{aligned}
\mathbb{E}\|g-\tilde{g}\|^2 &\leq C_a \inf_{m\in\{1,\cdots,m_n\}}\big[\|g-g_m^{(n)}\|^2 + \|g_m^{(n)}-g_m\| + \mathrm{pen}(m)\big] + \frac{2a\kappa_a(L_{m_n}+R_{m_n})}{n} + \frac{2a\kappa_a A_1}{n}\\
&\leq C_a \inf_{m\in\{1,\cdots,m_n\}}\big[\|g-g_m\|^2 + (M_2+1)m^2/k_n + \mathrm{pen}(m)\big] + \frac{2a\kappa_a(L_{m_n}+R_{m_n})}{n} + \frac{2a\kappa_a A_1}{n}.
\end{aligned}$$

Proof of (6.14) We start by writing $\mathbb{E}(W_n(m,m')) = \mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\nu_n^{(1)}(t)|^2 - p(m,m')]_+$ as

$$\mathbb{E}\Big\{\mathbb{E}_{\mathbf{X}}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu_n^{(1)}(t)|^2 - p(m,m')\Big]_+\Big\},$$

where $\mathbb{E}_{\mathbf{X}}(Y)$ denotes the conditional expectation $\mathbb{E}(Y|\sigma(X_i,\, i\geq 0))$. The point is that, conditionally to $\sigma(X_i,\, i\geq 0)$, the random variables $u_t^*(Z_i) - \mathbb{E}(u_t^*(Z_i)|\sigma(X_i,\, i\geq 0))$ are centered, independent but non identically distributed. We proceed as in the independent case (see Comte *et al.* (2006)), by applying the following Lemma to the expectation $\mathbb{E}_{\mathbf{X}}[\sup_{t\in B_{m,m'}(0,1)}|\nu_n^{(1)}(t)|^2 - p(m,m')]_+$.

Lemma 6.1. *Let $Y_1,\ldots,Y_n$ be independent random variables and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\xi^2>0$*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|^2 - 2(1+2\xi^2)H^2\Big]_+ \leq \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\xi^2\frac{nH^2}{v}} + \frac{98M_1^2}{K_1n^2C^2(\xi^2)}e^{-\frac{2K_1C(\xi)\xi}{7\sqrt{2}}\frac{nH}{M_1}}\right),$$

with $C(\xi)=\sqrt{1+\xi^2}-1$, $K_1 = 1/6$, and

$$\sup_{f\in\mathcal{F}}\|f\|_\infty \leq M_1, \quad \mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|\Big]\leq H, \quad \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{k=1}^n \mathrm{Var}(f(Y_k))\leq v.$$

The proof of this inequality can be found in Appendix. It comes from a concentration Inequality in Klein and Rio (2005) and arguments that can be found in Birgé and Massart (1998). Usual density arguments show that this result can be applied to the class of functions $\mathcal{F} = B_{m,m'}(0,1)$. Let us denote by $m^* = \max(m, m')$. Applying Lemma 6.1, one has the bound

$$\mathbb{E}_{\mathbf{X}}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - 2(1+2\xi^2)H^2\Big]_+ \leq \frac{6}{K_1}\left(\frac{v}{n}e^{-K_1\xi^2\frac{nH^2}{v}} + \frac{98M_1^2}{K_1 n^2 C^2(\xi^2)}e^{-\frac{K_1 C(\xi)\xi}{7\sqrt{2}}\frac{nH}{M_1}}\right),$$

where

$$\sup_{t\in B_{m,m'}(0,1)} \|u_t^*(Z_1)\|_\infty \leq M_1, \quad \mathbb{E}_{\mathbf{X}}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|\Big] \leq H, \qquad \sup_{t\in B_{m,m'}} \frac{1}{n}\sum_{k=1}^n \mathrm{Var}_{\mathbf{X}}(u_t^*(Z_k)) \leq v.$$

By applying Lemma 7.3, we propose to take

$$H^2 = H^2(m^*) = \frac{\Delta(m^*)}{n}, \quad M_1 = M_1(m^*) = \sqrt{nH^2} \text{ and } v = v(m^*) = \frac{\sqrt{\Delta_2(m^*,h)}}{2\pi}$$

with, for f_Z denoting the density of Z_1,

$$\Delta_2(m,h) = \int_{-\pi m}^{\pi m}\int_{-\pi m}^{\pi m} \frac{|f_Z^*(x-y)|^2}{|f_\varepsilon^*(x)f_\varepsilon^*(y)|^2}dxdy. \tag{6.16}$$

From the definition (6.13) of $W_n(m,m')$, by taking $p(m,m') = 2(1+2\xi^2)H^2(m^*)$, we get that

$$\mathbb{E}(W_n(m,m')) \leq \mathbb{E}\Big\{\mathbb{E}_{\mathbf{X}}\Big[\sup_{t\in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - 2(1+2\xi^2)H^2(m^*)\Big]_+\Big\}. \tag{6.17}$$

According to the condition (6.11), we thus take $\mathrm{pen}(m) = 4ap(m,m) = 8n^{-1}a(1+2\xi^2)\Delta(m)$ where ξ^2 is suitably chosen in the control of the sum of the right-hand side of (6.17). Set m_0 such that for $m^* \geq m_0$

$$(1/2)\lambda_1(f_\varepsilon, \kappa_0')\Gamma(m^*) \leq \Delta(m^*) \leq 2\lambda_1(f_\varepsilon,\kappa_0)\Gamma(m^*) \tag{6.18}$$

where $\Gamma(m)$ is defined in (4.2) and $\lambda_1(f_\varepsilon,\kappa_0)$ and $\lambda_1(f_\varepsilon,\kappa_0')$ are defined in (4.3). We split the sum over m' in two parts and write

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_n(m,m')) = \sum_{m'|m^*<m_0} \mathbb{E}(W_n(m,m')) + \sum_{m'|m^*\geq m_0} \mathbb{E}(W_n(m,m')). \tag{6.19}$$

By applying Lemma 6.1 and (6.18), we get the global bound $\mathbb{E}_{\mathbf{X}}(W_n(m,m')) \leq K[I(m^*) + II(m^*)]$, where $I(m^*)$ and $II(m^*)$ are defined by

$$I(m^*) = \frac{v(m^*)}{n}\exp\left\{-K_1\xi^2\frac{\Delta(m^*)}{v(m^*)}\right\}$$

$$\text{and} \qquad II(m^*) = \frac{\Delta(m^*)}{n^2}\exp\left\{-\frac{2K_1\xi C(\xi)}{7\sqrt{2}}\sqrt{n}\right\},$$

Since I and II do not depend on the X_i's, we infer that $\mathbb{E}(W_n(m,m')) \leq K[I(m^*) + II(m^*)]$.

When $m^* \leq m_0$, with m_0 finite, we get that for all $m \in \{1,\cdots,m_n\}$,

$$\sum_{m'|m^*\leq m_0} \mathbb{E}(W_n(m,m')) \leq \frac{C(m_0)}{n}.$$

We now come to the sum over m' such that $m^* > m_0$.

When $\delta > 1$ we use a rough bound for $\Delta_2(m,h)$ given by $\sqrt{\Delta_2(m,h)} \leq 2\pi n H^2(m)$.

When $0 \leq \delta \leq 1$, write that

$$\Delta_2(m,h) \leq \| \, |f_\varepsilon^*|^{-2} \mathbb{I}_{[-\pi m, \pi m]} \, \|_\infty \, \Delta(m) \, \| \, h^* \, \|^2 \, (2\pi).$$

Under $(\mathbf{A}_1^\varepsilon)$-$(\mathbf{A}_2^\varepsilon)$ we use that $\|h^*\|^2 \leq \|f_\varepsilon^*\|^2 < \infty$, that $\sqrt{2\pi}\|f_\varepsilon^*\| = \|f_\varepsilon\|$ and apply (6.18) to infer that for $m^* \geq m_0$,

$$v(m^*) = \frac{\sqrt{\Delta_2(m^*,h)}}{2\pi} \leq \lambda_2(f_\varepsilon, \kappa_0)\Gamma_2(m^*), \tag{6.20}$$

where $\lambda_2(f_\varepsilon, \kappa_0)$ is defined in (5.6) and

$$\Gamma_2(m) = (1+(\pi m)^2)^\gamma (\pi m)^{\min((1/2-\delta/2),(1-\delta))} \exp(2\mu(\pi m)^\delta) = (\pi m)^{-(1/2-\delta/2)_+}\Gamma(m). \tag{6.21}$$

Combining (6.18) and (6.20), we get that for $m^* \geq m_0$,

$$I(m^*) \leq \frac{\lambda_2(f_\varepsilon,\kappa_0)\Gamma_2(m^*)}{n} \exp\left\{ -\frac{K_1\xi^2\lambda_1(f_\varepsilon,\kappa_0')}{2\lambda_2(f_\varepsilon,\kappa_0)} (\pi m^*)^{(1/2-\delta/2)_+} \right\}$$

$$\text{and} \qquad II(m^*) \leq \frac{\Delta(m^*)}{n^2} \exp\left\{ -\frac{2K_1\xi C(\xi)\sqrt{n}}{7\sqrt{2}} \right\}.$$

• Study of $\sum_{m'|m^*\geq m_0} II(m^*)$. According to the choices for $v(m^*)$, $H^2(m^*)$ and $M_1(m^*)$, we have

$$\begin{aligned} \sum_{m'|m^*\geq m_0} II(m^*) &\leq \sum_{m'=1}^{m_n} \frac{\Delta(m^*)}{n^2} \exp\left\{ \frac{-2K_1\xi C(\xi)\sqrt{n}}{7\sqrt{2}} \right\} \\ &\leq \frac{\Delta(m_n)m_n}{n^2} \exp\left\{ \frac{-2K_1\xi C(\xi)\sqrt{n}}{7\sqrt{2}} \right\}. \end{aligned}$$

Since under (5.7), $n^{-1}\Delta(m_n)$ is bounded, we deduce that $\sum_{m'|m^*\geq m_0} II(m^*) \leq n^{-1}C$.

• Study of $\sum_{m'|m^*\geq m_0} I(m^*)$. Denote by $\psi = 2\gamma + \min(1/2-\delta/2, 1-\delta)$, $\omega = (1/2-\delta/2)_+$, and $K' = K_1\lambda_1(f_\varepsilon,\kappa_0')/(2\lambda_2(f_\varepsilon,\kappa_0))$. For $a, b \geq 1$, we have that

$$\begin{aligned} \max(a,b)^\psi e^{2\mu\pi^\delta \max(a,b)^\delta} e^{-K'\xi^2 \max(a,b)^\omega} &\leq (a^\psi e^{2\mu\pi^\delta a^\delta} + b^\psi e^{2\mu\pi^\delta b^\delta}) e^{-(K'\xi^2/2)(a^\omega + b^\omega)} \\ &\leq a^\psi e^{2\mu\pi^\delta a^\delta} e^{-(K'\xi^2/2)a^\omega} e^{-(K'\xi^2/2)b^\omega} + b^\psi e^{2\mu\pi^\delta b^\delta} e^{-(K'\xi^2/2)b^\omega}. \end{aligned} \tag{6.22}$$

Consequently,

$$\begin{aligned} \sum_{m'|m^*\geq m_0} I(m^*) &\leq \sum_{m'=1}^{m_n} \frac{\lambda_2(f_\varepsilon,\kappa_0)\Gamma_2(m^*)}{n} \exp\left\{ -\frac{K_1\xi^2(\lambda_1(f_\varepsilon,\kappa_0')}{2\lambda_2(f_\varepsilon,\kappa_0)} (\pi m^*)^{(1/2-\delta/2)_+} \right\} \\ &\leq \frac{2\lambda_2(f_\varepsilon,\kappa_0)\Gamma_2(m)}{n} \exp\left\{ -\frac{K'\xi^2}{2}(\pi m)^{(1/2-\delta/2)_+} \right\} \sum_{m'=1}^{m_n} \exp\left\{ -\frac{K'\xi^2}{2}(\pi m')^{(1/2-\delta/2)_+} \right\} \\ &\quad + \sum_{m'=1}^{m_n} \frac{2\lambda_2(f_\varepsilon,\kappa_0)\Gamma_2(m')}{n} \exp\left\{ -\frac{K'\xi^2}{2}(\pi m')^{(1/2-\delta/2)_+} \right\}. \end{aligned} \tag{6.23}$$

Case $0 \leq \delta < 1/3$. In that case, since $\delta < (1/2 - \delta/2)_+$, the choice $\xi^2 = 1$ ensures that the quantity $\Gamma_2(m)\exp\{-(K'\xi^2/2)(m)^{(1/2-\delta/2)}\}$ is bounded, and thus the first term in (6.23) is bounded by C/n. Since $1 \leq m \leq m_n$ with m_n satisfying (5.7), $n^{-1}\sum_{m'=1}^{m_n}\Gamma_2(m')\exp\{-(K'/2)(m')^{(1/2-\delta/2)}\}$ is bounded by $\tilde{C}/n$, and hence $\sum_{m'|m^*\geq m_0} I(m^*) \leq Dn^{-1}$. According to (6.11), the result follows by choosing $\mathrm{pen}(m) = 4ap(m,m') = 24an^{-1}\Delta(m)$.

Case $\delta = 1/3$. According to (6.23), we choose ξ^2 such that $2\mu\pi^\delta(m)^\delta - (K'\xi^2/2)m^\delta = -2\mu(\pi m)^\delta$ that is $\xi^2 = (8\mu\pi^\delta\lambda_2(f_\varepsilon,\kappa_0))/(K_1\lambda_1(f_\varepsilon,\kappa_0'))$. Arguing as for the case $0 \leq \delta < 1/3$, this choice ensures that $\sum_{m'|m^*\geq m_0} I(m^*) \leq Dn^{-1}$, and consequently (6.14) holds. The result follows by taking $p(m,m') = 2(1+2\xi^2)\Delta(m^*)n^{-1}$, and $\mathrm{pen}(m) = 8a(1+2\xi^2)\Delta(m)n^{-1}$.

Case $\delta > 1/3$. In that case $\delta > (1/2-\delta/2)_+$. Choose $\xi^2(m)$ such that $2\mu\pi^\delta(m)^\delta - (K'\xi^2/2)m^{(1/2-\delta)_+} = -2\mu\pi^\delta(m)^\delta$. Hence $\xi^2(m) = (8\mu(\pi)^\delta\lambda_2(f_\varepsilon,\kappa_0)/(K_1\lambda_1(f_\varepsilon,\kappa_0'))(\pi m)^{\delta-(1/2-\delta/2)_+}$. This choice ensures that $\sum_{m'|m^*\geq m_0} I(m^*) \leq D/n$, so that (6.14) holds. The result follows by choosing $p(m,m') = 2(1+2\xi^2(m^*))\Delta(m^*)/n$, associated to $\mathrm{pen}(m) = 8a(1+2\xi^2(m))\Delta(m)/n$.

Proof of (6.15). Since $\max(m,\hat{m}) \leq m_n$, according to (6.5),

$$\begin{aligned}
\sup_{t\in B_{m,\hat{m}}(0,1)} \mathbb{E}\left(\nu_{n,X}(t)\right)^2 &\leq \sup_{t\in S_{m_n},\|t\|=1} \mathbb{E}\left(\frac{1}{2\pi}\int \nu_{n,X}(e^{ix\cdot})t^*(-x)dx\right)^2\\
&\leq \frac{1}{2\pi}\int_{-\pi m_n}^{\pi m_n} \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n e^{ixX_k}\right)dx\\
&\leq \frac{m_n}{n} + \frac{1}{\pi n}\int_{-\pi m_n}^{\pi m_n}\sum_{k=2}^n \left|\mathrm{Cov}\left(e^{ixX_1}, e^{ixX_k}\right)\right| dx
\end{aligned}$$

and Theorem 5.1 is proved. □

6.3. **Proofs of Theorem 5.2 (1).** We use the coupling argument recalled in Section 2.1 to build approximating variables for the X_i's. For $n = 2p_nq_n + r_n$, $0 \leq r_n < q_n$, and $\ell = 0, \cdots, p_n - 1$, denote by

$$\begin{aligned}
E_\ell &= (X_{2\ell q_n+1}, ..., X_{(2\ell+1)q_n}), & F_\ell &= (X_{(2\ell+1)q_n+1}, ..., X_{(2\ell+2)q_n}),\\
E_\ell^\star &= (X_{2\ell q_n+1}^\star, ..., X_{(2\ell+1)q_n}^\star), & F_\ell^\star &= (X_{(2\ell+1)q_n+1}^\star, ..., X_{(2\ell+2)q_n}^\star).
\end{aligned}$$

The variables $E_\ell^\star$ and $F_\ell^\star$ are such that

- $E_\ell^\star$, E_ℓ, $F_\ell^\star$ and F_ℓ are identically distributed,
- $\mathbb{P}(E_\ell \neq E_\ell^\star) \leq \beta_{\mathbf{X},\infty}(q_n)$ and $\mathbb{P}(F_\ell \neq F_\ell^\star) \leq \beta_{\mathbf{X},\infty}(q_n)$,
- The variables $(E_\ell^\star)_{0\leq\ell\leq p_n-1}$ are i.i.d., and so are the variables $(F_\ell^\star)_{0\leq\ell\leq p_n-1}$.

Without loss of generality and for sake of simplicity we assume that $r_n = 0$. For κ_a defined in (5.8), we start from

$$\begin{aligned}
\|\tilde{g} - g\|^2 &\leq \kappa_a^2 \|g_m^{(n)} - g\|^2 + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_n^{(1)}(t))^2 + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t))^2 \\
&\quad + \kappa_a(\mathrm{pen}(m) - \mathrm{pen}(\hat{m})) \\
&\leq \kappa_a^2 \|g_m^{(n)} - g\|^2 + 2a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_n^{(1)}(t))^2 + 4a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}^\star(t))^2 \\
&\quad + 4a\kappa \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t) - \nu_{n,X}^\star(t))^2 + \kappa_a(\mathrm{pen}(m) - \mathrm{pen}(\hat{m})),
\end{aligned}$$

where $\nu_{n,X}^\star(t)$ is defined as $\nu_{n,X}(t)$ with $X_i^\star$ instead of X_i. Choose $p_1(m, m')$ and $p_2(m, m')$ such that

$$2ap_1(m, m') \leq [\mathrm{pen}_1(m) + \mathrm{pen}_1(m')] \text{ and } 4ap_2(m, m') \leq [\mathrm{pen}_2(m) + \mathrm{pen}_2(m')],$$

for $\mathrm{pen}(m) = \mathrm{pen}_1(m) + \mathrm{pen}_2(m)$. It follows that

$$\begin{aligned}
\|\tilde{g} - g\|^2 &\leq \kappa_a^2 \|g - g_m^{(n)}\|^2 + 2\kappa_a \mathrm{pen}(m) + 4a\kappa_a W_{n,X}^\star(m, \hat{m}) + 4a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t) - \nu_{n,X}^\star(t))^2 \\
&\quad + 2a\kappa_a W_n(m, \hat{m}) \\
&\leq \kappa_a^2 \|g - g_m^{(n)}\|^2 + 2\kappa_a \mathrm{pen}(m) + 4a\kappa_a \sum_{m'=1}^{m_n} W_{n,X}^\star(m, m') + 2a\kappa_a \sum_{m'=1}^{m_n} W_n(m, m') \qquad (6.24) \\
&\quad + 4a\kappa_a \sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t) - \nu_{n,X}^\star(t))^2,
\end{aligned}$$

where

$$W_n(m, m') := \Big[\sup_{t \in B_{m,m'}(0,1)} |\nu_n^{(1)}(t)|^2 - p_1(m, m')\Big]_+, \tag{6.25}$$

$$W_{n,X}^\star(m, m') := \Big[\sup_{t \in B_{m,m'}(0,1)} |\nu_{n,X}^\star(t)|^2 - p_2(m, m')\Big]_+. \tag{6.26}$$

The main parts of the proof lies in the three following points :

1) Study of $W_n(m, m')$. More precisely, we have to find $p_1(m, m')$ such that for a constant A_2,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_n(m, m')) \leq \frac{A_2}{n}. \tag{6.27}$$

2) Study of $W_{n,X}^\star(m, m')$. More precisely, we have to find $p_2(m, m')$ such that for a constant A_3,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_{n,X}^\star(m, m')) \leq \frac{A_3}{n}. \tag{6.28}$$

3) Study of $\sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}(t) - \nu_{n,X}^\star(t))^2$ and more precisely we have to prove that

$$\mathbb{E}\Big[\sup_{t \in B_{m,\hat{m}}(0,1)} (\nu_{n,X}^\star(t) - \nu_{n,X}(t))^2\Big] \leq 4\beta_{\mathbf{X},\infty}(q_n) m_n \leq \frac{A_4}{n}. \tag{6.29}$$

Proof of (6.27) The proof of (6.27) for ordinary smooth errors ($\delta = 0$ in $(\mathbf{A}_1^\varepsilon)$) is the same as the proof of (6.14) by taking $p_1(m, m') = p(m, m')$, with $p(m, m')$ as in the proof of (6.14) and $\xi^2 = 1$.

Hence we choose $\mathrm{pen}_1(m) = 24an^{-1}\Delta(m)$.

Proof of (6.28) We proceed as in the independent case by applying Lemma 6.1. Set $m^* = \max(m, m')$. The process $W^\star_{n,X}(m, m')$ must be split into two terms $(W^\star_{n,1,X}(m, m') + W^\star_{n,2,X}(m, m'))/2$ involving respectively the odd and even blocks, which are of the same type. More precisely $W^\star_{n,k,X}(m, m')$ is defined, for $k = 1, 2$, by

$$W^\star_{n,k,X}(m, m') = \Big[\sup_{t\in B_{m,m'}(0,1)} \Big|\frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\Big(t(X^\star_{(2\ell+k-1)q_n+i}) - \langle t, g\rangle\Big)\Big|^2 - p_{2,k}(m, m')\Big]_+$$

We only study $W^\star_{n,1,X}(m, m')$ and conclude for $W^\star_{n,2,X}(m, m')$ by using analogous arguments. The study of $W^\star_{n,1,X}(m, m')$ consists in applying Lemma 6.1 to $\nu^\star_{n,1,X}(t)$ defined by

$$\nu^\star_{n,1,X}(t) = \frac{1}{p_n}\sum_{\ell=1}^{p_n}\nu^\star_{q_n,\ell,X}(t) \text{ with } \nu^\star_{q_n,\ell,X}(t) = \frac{1}{q_n}\sum_{j=1}^{q_n} t(X^\star_{2\ell q_n+j}) - \langle t, g\rangle,$$

considered as the sum of the p_n independent random variables $\nu^\star_{q_n,\ell,X}(t)$. Denote by $M_1^\star(m^*)$, $H^\star(m^*)$ and $v^\star(m^*)$ quantities such that

$$\sup_{t\in B_{m,m'}(0,1)} \|\, \nu^\star_{q_n,\ell,X}(t)\,\|_\infty \;\le\; M_1^\star(m^*), \quad \mathbb{E}\Big(\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|\Big) \le H^\star(m^*)$$
$$\text{and} \quad \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu^\star_{q_n,\ell,X}(t)) \le v^\star(m^*).$$

Lemma 7.5 leads to the choices $M_1^\star(m^*) = \sqrt{m^*}$,

$$(H^\star(m^*))^2 = \frac{\Big(1 + 4\sum_{k=1}^{n}\beta_{\mathbf{X},1}(k)\Big)m^*}{n}, \quad \text{and } v^\star(m^*) = \frac{8\Big(\sum_{k=0}^{q_n}(k+1)\beta_{\mathbf{X},1}(k)\|g\|_\infty m^*\Big)^{1/2}}{q_n}.$$

Take $\xi^2(m^*) = 1/2$. We use that for $m^* \ge m_0$,

$$2(1 + 2\xi^2(m^*))(H^\star(m^*))^2 = 4(H^\star(m^*))^2 \le \Delta(m^*)/(4n).$$

Then we take $p_{2,1}(m, m') = \Delta(m)/(4n)$, and get that

$$\begin{aligned}
\sum_{m'=1}^{m_n}\mathbb{E}(W^\star_{n,1,X}(m, m')) &= \sum_{m'|m^*\le m_0}\mathbb{E}(W^\star_{n,1,X}(m, m')) + \sum_{m'|m^*>m_0}\mathbb{E}(W^\star_{n,1,X}(m, m')) \\
&\le \sum_{m'|m^*\le m_0}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 4(H^\star(m^*))^2\Big]_+ \\
&\quad + \sum_{m'|m^*\le m_0}|p_{21}(m, m') - 4(H^\star(m^*))^2| \\
&\quad + \sum_{m'|m^*>m_0}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 4(H^\star(m^*))^2\Big]_+.
\end{aligned}$$

It follows that

$$\begin{aligned}\sum_{m'=1}^{m_n}\mathbb{E}(W^\star_{n,1,X}(m,m')) &\leq 2\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2-4(H^\star(m^*))^2\Big]_+\\&\quad+\sum_{m'|m^*\leq m_0}|p_{2,1}(m,m')-4(H^\star(m^*))^2|\\&\leq 2\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2-4(H^\star(m^*))^2\Big]_+ +\frac{C(m_0)}{n}.\end{aligned}$$

We apply Lemma 6.1 to $\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2-4(H^\star(m^*))^2\Big]_+$ and obtain

$$\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2-4(H^\star(m^*))^2\Big]_+\leq K\sum_{m'=1}^{m_n}[I^\star(m^*)+II^\star(m^*)],$$

with $I^\star(m^*)$ and $II^\star(m^*)$ defined by

$$I^\star(m^*)=\frac{m^*}{n}\exp\Big\{-K_2\sqrt{m^*}\Big\}\text{ and }II^\star(m^*)=\frac{q_n^2m^*}{n^2}\exp\Big\{-\frac{\sqrt{2}K_1\xi C(\xi)}{7}\frac{\sqrt{n}}{q_n}\Big\},$$

where $K_2=(K_1/32)(1+4\sum_{k=1}^n\beta_{\mathbf{X},1}(k))/\sqrt{\|g\|_\infty\sum_{k=0}^{q_n}(k+1)\beta_{\mathbf{X},1}(k)}$.
With our choice of $\xi^2(m)$, if we take $q_n=[n^c]$, for c in $]0,1/2[$, then

$$\sum_{m'}I(m^*)\leq\frac{C}{n},\text{ and }\sum_{m'=1}^{m_n}II^\star(m^*)\leq\frac{C}{n}.$$

Finally

$$\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2-4(H^\star(m^*))^2\Big]_+\leq\frac{C}{n}$$

and

$$\sum_{m'=1}^{m_n}\mathbb{E}[W^\star_{n,X}(m,m')]\leq 2\sum_{m'=1}^{m_n}\mathbb{E}[W^\star_{n,1,X}(m,m')+W^\star_{n,2,X}(m,m')]\leq\frac{C}{n}.$$

The result follows for choosing $p_2(m,m')=2p_{2,1}(m,m')+2p_{2,2}(m,m')=\Delta(m)/n$, and $\mathrm{pen}(m)=25a\Delta(m)/n$.

Proof of (6.29). A rough bound is obtained by writing that

$$\begin{aligned}\sup_{t\in B_{m,\hat m}(0,1)}|\nu^\star_{n,X}(t)-\nu_{n,X}(t)|^2 &= \sup_{t\in S^{(n)}_{\max(m,\hat m)},\|t\|\leq 1}|\nu^\star_{n,X}(t)-\nu_{n,X}(t)|^2\\&\leq\sup_{t\in S_{m_n},\|t\|\leq 1}|\nu^\star_{n,X}(t)-\nu_{n,X}(t)|^2.\end{aligned}$$

According to (6.5),

$$\nu^\star_{n,X}(t)-\nu_{n,X}(t)=\frac{1}{2\pi}\int[\nu^\star_{n,X}(e^{ix\cdot})-\nu_{n,X}(e^{ix\cdot})]t^*(-x)dx.$$

Since $|\nu_{n,X}(e^{ix\cdot}) - \nu^\star_{n,X}(e^{ix\cdot})| \le 2$, we have

$$\begin{aligned}\sup_{t\in B_{m,\hat m}(0,1)} |\nu^\star_{n,X}(t) - \nu_{n,X}(t)|^2 &\le \sup_{t\in S_{m_n}, \|t\|\le 1} \frac{1}{4\pi^2}\left|\int [\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})]t^*(-x)dx\right|^2 \\ &\le \frac{1}{2\pi}\int_{-\pi m_n}^{\pi m_n} |\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})|^2 dx \\ &\le \frac{1}{\pi}\int_{-\pi m_n}^{\pi m_n} |\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})| dx.\end{aligned}$$

According to the properties of the coupling,

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat m}(0,1)} |\nu^\star_{n,X}(t) - \nu_{n,X}(t)|^2\Big] \le \frac{1}{\pi}\int_{-\pi m_n}^{\pi m_n} \mathbb{E}|\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})|dx \le 4\beta_{\mathbf{X},\infty}(q_n)m_n.$$

For ordinary smooth errors, according to (5.7), $m_n \le n^{1/(2\gamma+1)}$. It follows that if we choose q_n such that $\beta_{\mathbf{X},\infty}(q_n) = O(n^{-(2\gamma+2)/(2\gamma+1)})$, then $\beta_{\mathbf{X},\infty}(q_n)m_n = O(n^{-1})$. For $q_n = [n^c]$ and $\beta_{\mathbf{X},\infty}(n) = O(n^{-1-\theta})$, we obtain the condition $n^{-c(1+\theta)} = O(n^{-(2\gamma+2)/(2\gamma+1)})$. If $\theta > (2\gamma+3)/(2\gamma+1)$, one can find $c < 1/2$ such that this condition is satisfied.

6.4. **Proofs of Theorem 5.2 (2).** We proceed as in the β-mixing case, by using the coupling argument given in Section 2.1. The variables $E_\ell, E^\star_\ell, F_\ell, F^\star_\ell$ are build as in Section 6.3 and are such that

- $E^\star_\ell$, E_ℓ, $F^\star_\ell$ and F_ℓ are identically distributed,
- $\sum_{i=1}^{q_n} \mathbb{E}(|X_{2\ell q_n+i} - X^\star_{2\ell q_n+i}|) \le q_n\tau_{\mathbf{X},\infty}(q_n)$ and $\sum_{i=1}^{q_n} \mathbb{E}(|X_{(2\ell+1)q_n+i} - X^\star_{(2\ell+1)q_n+i}|) \le q_n\tau_{\mathbf{X},\infty}(q_n)$,
- The variables $(E^\star_\ell)_{0\le\ell\le p_n-1}$ are i.i.d., and so are the variables $(F^\star_\ell)_{0\le\ell\le p_n-1}$.

Without loss of generality and for sake of simplicity we assume that $r_n = 0$. As for the proof of Theorem 5.2 under **2)**, we start from (6.25). Hence we have to :

1) Study of $W_n(m,m')$, and more precisely in finding $p_1(m,m')$ such that for a constant K_2,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W_n(m,m')) \le \frac{K_2}{n}. \tag{6.30}$$

2) Study of $W^\star_{n,X}(m,m')$, and more precisely in finding $p_2(m,m')$ such that for a constant K_3,

$$\sum_{m'=1}^{m_n} \mathbb{E}(W^\star_{n,X}(m,m')) \le \frac{K_3}{n}. \tag{6.31}$$

3) Study of $\sup_{t\in B_{m,\hat m}(0,1)}(\nu_{n,X}(t) - \nu^\star_{n,X}(t))^2$ and more precisely in proving that

$$\mathbb{E}\big[\sup_{t\in B_{m,\hat m}(0,1)} (\nu^\star_{n,X}(t) - \nu_{n,X}(t))^2\big] \le \pi\tau_{\mathbf{X},\infty}(q_n){m_n}^2 \le \frac{K_4}{n}. \tag{6.32}$$

Proof of (6.30) The proof of (6.30) for ordinary smooth errors is the same as the proof of (6.14).

Proof of (6.31) As for the proof (6.28) we apply Lemma 6.1 with

$$(H^\star(m^*))^2 = \frac{\Big(m^* + \pi\sum_{k=1}^{n-1}\tau_{\mathbf{X},1}(k)(m^*)^2\Big)}{n}, \quad M_1^\star(m^*) = m^*,$$

$$\text{and } v^\star(m^*) = \frac{\Big(m^* + \pi\sum_{k=1}^{n-1}\tau_{\mathbf{X},1}(k)(m^*)^2\Big)}{q_n}.$$

We take $\xi^2 = \xi^2(m) = (3/K_1+1)\ln(m)$. In the same way as for the proof Theorem 5.2(1), we use that for $m^* \geq m_0$,

$$2(1+2\xi^2(m^*))(H^\star(m^*))^2 \leq \Delta(m^*)/(4n).$$

Then we take $p_{21}(m,m') = \Delta(m^*)(4n)^{-1}$ and get that

$$\sum_{m'=1}^{m_n}\mathbb{E}(W^\star_{n,1,X}(m,m')) \leq 2\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi^2(m^*))(H^\star(m^*))^2\Big]_+ + \frac{C(m_0)}{n}.$$

We now apply Lemma 6.1 to $\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi^2(m^*))(H^\star(m^*))^2\Big]_+$ and obtain

$$\sum_{m'=1}^{m_n}\mathbb{E}\Big[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|^2 - 2(1+2\xi^2(m^*))(H^\star(m^*))^2\Big]_+ \leq K\sum_{m'=1}^{m_n}[I^\star(m^*) + II^\star(m^*)],$$

with $I^\star(m^*)$ and $II^\star(m^*)$ now defined by

$$I^\star(m^*) = \frac{m^{*2}}{n}\exp\{-K_1\xi^2(m^*)\}$$

$$\text{and} \qquad II^\star(m^*) = \frac{q_n^2 m^{*2}}{n^2}\exp\left\{-\frac{\sqrt{2}K_1\xi C(\xi)\Big(1+\pi\sum_{k=1}^{n}\tau_{\mathbf{X},1}(k)\Big)}{7}\frac{\sqrt{n}}{q_n}\right\}.$$

With this $\xi^2(m)$, if we take $q_n = [n^c]$, with c in $]0,1/2[$ then

$$\sum_{m'} I(m^*) \leq \frac{C}{n} \quad \text{and} \quad \sum_{m'=1}^{m_n} II(m^*) \leq \frac{C}{n}.$$

Finally $\sum_{m'=1}^{m_n}\mathbb{E}[W^\star_n(m,m')] \leq 2\sum_{m'=1}^{m_n}\mathbb{E}[W^\star_{n,1,X}(m,m') + W^\star_{n,2,X}(m,m')] \leq Cn^{-1}$. The result follows by choosing $p_2(m,m') = 2p_{21}(m,m') + 2p_{22}(m,m') = \Delta(m)n^{-1}$, and $\text{pen}(m) = 25a\Delta(m)n^{-1}$.

Proof of (6.32) The proof of (6.32) is similar to the proof of (6.15). Since $|e^{-ixt} - e^{-ixs}| \leq |x||t-s|$, one has

$$\sum_{i=1}^{q_n}\mathbb{E}(|e^{-iX_{2\ell q_n+i}} - e^{-iX^\star_{2\ell q_n+i}}|) \leq q_n|x|\tau_{\mathbf{X},\infty}(q_n)$$

It follows that

$$\mathbb{E}\Big[\sup_{t\in B_{m,\hat m}(0,1)}|\nu^\star_{n,X}(t) - \nu_{n,X}(t)|^2\Big] \quad\leq\quad \frac{1}{\pi}\int_{-\pi m_n}^{\pi m_n}\mathbb{E}|\nu^\star_{n,X}(e^{ix\cdot}) - \nu_{n,X}(e^{ix\cdot})|dx \leq \pi\tau_{\mathbf{X},\infty}(q_n){m_n}^2.$$

For ordinary smooth errors, according to (5.7), ${m_n}^2 \leq n^{2/(2\gamma+1)}$. It follows that if we choose q_n such that $\tau_{\mathbf{X},\infty}(q_n) = O(n^{-(2\gamma+3)/(2\gamma+1)})$, then $\tau_{\mathbf{X},\infty}(q_n)m_n^2 = O(n^{-1})$. For $q_n = [n^c]$ and $\tau_{\mathbf{X},\infty}(n) = O(n^{-1-\theta})$, we obtain the condition $n^{-c(1+\theta)} = O(n^{-(2\gamma+3)/(2\gamma+1)})$. If $\theta > (2\gamma+5)/(2\gamma+1)$, one can find $c < 1/2$ such that this condition is satisfied. □

6.5. **Proof of Corollary 5.2.** The result follows from the proof of Theorem 5.2 (1), where only the process $\nu_{n,X}$ appears.□

7. Technical lemmas

Lemma 7.1. *If we denote by $\nu_{n,X}(t)$ the quantity defined by (6.1), then*

$$n^{-1}\sum_{k=1}^{n}\mathbb{E}(u_t^*(Z_k)|\sigma(X_i,\, i\geq 0))-<t,g>=\nu_{n,X}(t).$$

The proof of Lemma 7.1, rather straightforward, is omitted.

Lemma 7.2. *Let $\nu_{n,Z}(u_t^*)$ be defined by (6.1), $\Delta(m)$ being defined in (3.5). Then*

$$\sum_{j\in\mathbb{Z}}\left|u^*_{\varphi_{m,j}}(z)\right|^2=(2\pi)^{-1}m\int\left|\frac{\varphi^*(x)}{f_\varepsilon^*(xm)}\right|^2dx=\Delta(m).$$

Lemma 7.3. *Let $\nu_{n,Z}(u_t^*)$, $\Delta(m)$ and $\Delta_2(m,h)$ be defined in (6.1), (3.5) and in (6.16). Then*

$$\sup_{t\in B_{m,m'}(0,1)}\parallel u_t^*\parallel_\infty\leq\sqrt{\Delta(m^*)}\qquad \mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\nu_{n,Z}(u_t^*)|]\leq\sqrt{\Delta(m^*)/n},$$
$$\text{and}\sup_{t\in B_{m,m'}(0,1)}\mathrm{Var}(u_t^*(Z_1))\ \leq\ \sqrt{\Delta_2(m^*,h)}/(2\pi).$$

We refer to Comte *et al.* (2006) for the proofs of Lemmas 7.2 and 7.3.

Lemma 7.4. $\parallel\sum_{j\in\mathbb{Z}}|\varphi_{m,j}|^2\parallel_\infty\leq m$.

Proof of Lemma 7.4 Write

$$\sum_{j\in\mathbb{Z}}|\varphi_{m,j}(x)|^2=\frac{1}{(2\pi)^2}\sum_{j\in\mathbb{Z}}\left|\int e^{-iux}\varphi^*_{m,j}(u)du\right|^2=\frac{m}{(2\pi)^2}\sum_{j\in\mathbb{Z}}\left|\int e^{-ixum}e^{iju}\varphi^*(u)du\right|^2.$$

We conclude by applying Parseval's Formula which gives that

$$\sum_{j\in\mathbb{Z}}|\varphi_{m,j}(x)|^2=(2\pi)^{-1}m\int|\varphi^*(u)|^2\,du=m.$$

Lemma 7.5. *For $B_{m,m'}(0,1)=\{t\in S_{m\vee m'}\ /\ \|t\|_2=1\}$, we have, for $m^*=m\vee m'$,*

$$\sup_{t\in B_{m,m'}(0,1)}\parallel t\parallel_\infty\leq\sqrt{m^*},\quad \mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|]\leq\sqrt{\frac{(1+4\sum_{k=1}^n\beta_{\mathbf{X},1}(k))m^*}{n}}$$
$$\textit{and}\sup_{t\in B_{m,m'}(0,1)}\mathrm{Var}(\nu^\star_{q_n,\ell,X}(t))\leq\frac{[2\|g\|_\infty(1+32\sum_{k=1}^n(1+k)\beta_{\mathbf{X},1}(k))]^{1/2}\sqrt{m^*}}{q_n}.$$

Proof of Lemma 7.5 For t in $B_{m,m'}(0,1)$, with $m^*=m\vee m'$, one has $t=\sum_{j\in\mathbb{Z}}b_{m^*,j}\varphi_{m^*,j}$. Applying Cauchy-Schwarz Inequality and Lemma 7.4 we obtain

$$\sup_{t\in B_{m,m'}(0,1)}\parallel t\parallel_\infty\leq\left\|\sum_{j\in\mathbb{Z}}|\varphi_{m^*,j}|^2\right\|_\infty^{1/2}\leq\sqrt{m^*}.$$

Now, using again Cauchy-Schwarz Inequality

$$\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1,X}(t)|\right] \;\leq\; \mathbb{E}\left[\sqrt{\sum_{j\in\mathbb{Z}}(\nu^\star_{n,1X}(\varphi_{m^*,j}))^2}\right] \leq \sqrt{\sum_{j\in\mathbb{Z}}\mathrm{Var}(\nu^\star_{n,1,X}(\varphi_{m^*,j}))}.$$

By analogy with (6.6), we write

$$\mathbb{E}\left(\sum_{j\in\mathbb{Z}}\left(\nu^\star_{n,1,X}(\varphi_{m,j})\right)^2\right) = \frac{1}{4\pi^2}\sum_{j\in\mathbb{Z}}\mathbb{E}\Big(\int \varphi^*_{m,j}(-x)\nu^\star_{n,1,X}(e^{ix\cdot})dx\Big)^2 = \frac{1}{2\pi}\int_{-\pi m}^{\pi m}\mathbb{E}|\nu^\star_{n,1,X}(e^{ix\cdot})|^2dx.$$

This yields

$$\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)}|\nu^\star_{n,1,X}(t)|\right] \leq \frac{(1+4\sum_{k=1}^n \beta_{\mathbf{X},1}(k))\, m^*}{n}.$$

Finally, we apply Viennet's (1997) variance inequality (see Theorem 2.1 p. 472 and Lemma 4.2 p. 481). Hence there exist some measurable functions b_k, such that $0\leq b_k\leq 1$ and $\mathbb{E}\left[(\sum_{k=1}^n b_k(X_1))^2\right] \leq \sum_{k\geq 1}(1+k)\beta_{\mathbf{X},1}(k)$, for which

$$\sup_{t\in B_{m,m'}(0,1)}\mathrm{Var}(\nu_{q_n,\ell,X}(t)) \leq \sup_{t\in B_{m,m'}(0,1)}\frac{1}{q_n}\int\left(1+4\sum_{k=1}^{q_n}b_k\right)t^2(x)g(x)dx\,.$$

Consequently

$$\begin{aligned}\sup_{t\in B_{m,m'}(0,1)}\mathrm{Var}(\nu_{q_n,\ell,X}(t)) \;&\leq\; \sup_{t\in B_{m,m'}(0,1)}\frac{1}{q_n}\|t\|_\infty\|g\|_\infty^{1/2}\left[\int\left(1+4\sum_{k=1}^{q_n}b_k\right)^2 g(x)dx\right]^{1/2}\\ &\leq\; \sqrt{2\|g\|_\infty(1+32\sum_{k=1}^{q_n}(1+k)\beta_{\mathbf{X},1}(k))}\frac{\sqrt{m^*}}{q_n}.\end{aligned}$$

Proof of Lemma 6.1 : Starting from the concentration inequality given in Klein and Rio (2005) and arguing as in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354) we obtain the upper bound

$$\mathbb{P}\left(\sup_{g\in\mathcal{G}}|\nu_n(g)|\geq(1+\eta)H+\lambda\right)\leq 2\exp\left[-K_1 n\left(\frac{\lambda^2}{v}\wedge\frac{2\lambda(\eta\wedge 1)}{7M_1}\right)\right], \tag{7.1}$$

where $K_1 = 1/6$. By taking $\eta = (\sqrt{1+\epsilon} - 1) \wedge 1 = C(\epsilon) \le 1$ we get

$$\begin{aligned}
\mathbb{E}[\sup_{g\in\mathcal{G}} |\nu_n(g)|^2 - 2(1+2\epsilon)H^2]_+ &\le \int_0^{+\infty} \mathbb{P}\left(\sup_{g\in\mathcal{G}} |\nu_n(g)|^2 \ge 2(1+2\epsilon)H^2 + \tau\right) d\tau \\
&\le \int_0^{+\infty} \mathbb{P}\left(\sup_{g\in\mathcal{G}} |\nu_n(g)| \ge \sqrt{2(1+\epsilon)H^2 + 2(\epsilon H^2 + \tau/2)}\right) d\tau \\
&\le 2\int_0^{+\infty} \mathbb{P}\left(\sup_{g\in\mathcal{G}} |\nu_n(g)| \ge \sqrt{(1+\epsilon)}H + \sqrt{\epsilon H^2 + \tau/2}\right) d\tau \\
&\le 4\left(\int_0^{+\infty} e^{-\frac{K_1 n}{v}(\epsilon H^2 + \tau/2)} d\tau + \int_0^{+\infty} e^{-\frac{2K_1 nC(\epsilon)}{7M_1\sqrt{2}}(\sqrt{\epsilon}H + \sqrt{\tau/2})} d\tau\right) \\
&\le 4e^{-K_1\epsilon\frac{nH^2}{v}} \int_0^{+\infty} e^{-\frac{K_1 n}{2v}\tau} d\tau + 4e^{-\frac{\sqrt{2}K_1 C(\epsilon)\sqrt{\epsilon}}{7}\frac{nH}{M_1}} \int_0^{+\infty} e^{-\frac{K_1 C(\epsilon) n\sqrt{\tau}}{7M_1}} d\tau.
\end{aligned}$$

Using that for any positive constant C, $\int_0^{+\infty} e^{-Cx}dx = 1/C$ and $\int_0^{+\infty} e^{-C\sqrt{x}}dx = 2/C^2$, we get that

$$\mathbb{E}[\sup_{g\in\mathcal{G}} |\nu_n(g)|^2 - 2(1+2\epsilon)H^2]_+ \le \frac{8}{K_1}\left(\frac{v}{n}e^{-K_1\epsilon\frac{nH^2}{v}} + \frac{49M_1^2}{K_1^2 n^2 C^2(\epsilon)} e^{-\frac{\sqrt{2}K_1 C(\epsilon)\sqrt{\epsilon}}{7}\frac{nH}{M_1}}\right). \qquad \square$$

References

Berbee, H. (1979). *Random walks with Stationary Increments and Renewal Theory.* Mathematical Centre Tracts 112. Amsterdam: Mathematisch Centrum.

Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli 4*(3), 329–375.

Bradley, R. C. (2002). Intyroduction to strong mixing conditions. Vol. 1 Technical report, Department of Mathematics, I. U. Bloomington.

Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canadian J. Statist. 32*, 181–192.

Butucea, C. and Tsybakov, A. B. (2005). Sharp optimality for density deconvolution with dominating bias. *Theor. Probab. Appl., to appear*.

Comte, F. and Genon-Catalot, V. (2006). Penalized projection estimator for volatility density. *Scand. J. Statist., to appear*.

Comte, F. and Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist. (electronic) 6*, 211–238.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2005). Finite sample penalization in adaptive density deconvolution. Technical report, MAP5 2005-11, http://www.math-info.univ-paris5.fr/map5/publis/titres05.html.

Comte, F., Rozenholc, Y., and Taupin, M. L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canadian J. Statist. 34*(3), to appear.

Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Relat. Fields 132*(2), 203–236.

Devroye, L. (1986). *Non-uniform random variate generation.* New York etc.: Springer-Verlag.

Doukhan, P. (1994). *Mixing: Properties and examples.* Lecture Notes in Statistics (Springer). New York: Springer-Verlag.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist. 38*, 1466–1474.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Stat. 19*(3), 1257–1272.

Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab. 33*(3), 1060–1077.

Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris Sér. I Math. 342*(11), 877–882.

Masry, E. (1993). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Processes Appl. 47*(1), 53–74.

Masry, E. (2003). Deconvolving multivariate kernel density estimates from contaminated associated observations. *IEEE Trans. Inform. Theory 49*(11), 2941–2952.

Meyer, Y. (1990). *Ondelettes et opérateurs I: Ondelettes.* Actualités Mathématiques. Paris: Hermann, éditeurs des Sciences et des Arts.

Pensky, M. (1999). Estimation of a smooth density function using Meyer-type wavelets. *Stat. Decis. 17*(2), 111–123.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Stat. 27*(6), 2033–2053.

Rigollet, P. (2006). Adaptive density estimation using the blockwise stein method. *Bernoulli 12*(2), 351–370.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA 42*, 43–47.

Tribouley, K. and Viennet, G. (1998). $\mathbb{L}_p$ adaptive density estimation in a β mixing framework. *Ann. Inst. Henri Poincaré, Probab. Stat. 34*(2), 179–208.

van Es, B., Spreij, P., and van Zanten, H. (2003). Nonparametric volatility density estimation. *Bernoulli 9*(3), 451–465.

van Es, B., Spreij, P., and van Zanten, H. (2005). Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Stat. 17*(2), 237–251.

Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields 107*(4), 467–492.

Volkonskij, V. and Rozanov, Y. (1960). Some limit theorems for random functions. I. *Theor. Probab. Appl. 4*, 178–197.

F. Comte, MAP5 UMR 8145
Université René Descartes-Paris 5, 45 rue des Saints-Pères 75270 Paris cedex 06, France.
email: fabienne.comte@univ-paris5.fr.

J. Dedecker, Laboratoire de Statistique Théorique et Appliquée
Université paris 6, 175 rue du Chevaleret 75013 Paris, France.
email: dedecker@ccr.jussieu.fr

M.-L. Taupin, Laboratoire de Probabilités, Statistique et Modélisation, UMR C 8628
Université Paris-Sud, Bâtiment 425, 91405 Orsay Cedex, France.
email: Marie-Luce.Taupin@math.u-psud.fr