PENALIZED CONTRAST ESTIMATOR FOR DENSITY DECONVOLUTION WITH MIXING VARIABLES

F. COMTE*[,1] AND M.-L. TAUPIN[2]

Abstract

We consider the problem of estimating the density g of identically distributed variables X_i, from a sample $Z_1, \dots, Z_n$ where $Z_i = X_i + \varepsilon_i$, $i = 1, \dots, n$ and ε is a noise independent of X. We present a model selection procedure allowing to find non-asymptotic risk bounds for the adaptive estimators. These bounds perform an automatic trade-off between the squared bias term and a penalty term which has the same order as the variance, up to possibly some (negligible) logarithmic factors. Consequently, in all cases (ordinary or super smooth errors, ordinary or super smooth densities, independent or absolutely regular random variables), our estimator is proved to reach automatically the optimal rates, up to negligible logarithmic terms in some case.

Résumé

Considérons le problème de déconvolution c'est-à-dire de l'estimation de la densité de variables aléatoires identiquement distribuées X_i, $i = 1, \cdots, n$ à partir de l'observation de $Z_1, \cdots, Z_n$ où $Z_i = X_i + \varepsilon_i$, pour $i = 1, \dots, n$ et où les erreurs ε_i sont des variables aléatoires indépendantes des X_i, de densité connue. Par une procédure de sélection de modèles qui permet d'obtenir des bornes de risque non asymptotiques, nous construisons un estimateur adaptatif de la densité des X_i. Ces bornes de risque fournissent un compromis automatique entre un terme de biais et un terme de pénalité qui a pour ordre de grandeur l'ordre de grandeur de la variance, éventuellement à un facteur logarithmique négligeable près. Par conséquent notre estimateur atteint de façon automatique la vitesse minimax, ou presque dans certains cas, que les erreurs ou la densité à estimer soient peu ou très régulières. Ces résultats sont valables aussi bien dans le cas où les variables (X_i, ε_i) sont indépendantes que dans le cas où elles sont β-mélangeantes.

Keywords and Phrases. Absolutely regular random variables. Adaptive estimation. Berbee's coupling Lemma. Density deconvolution. Model selection. Penalized contrast estimator. Projection Estimator. Talagrand's Inequality.

---

[1] Université Paris V, MAP5, FRE CNRS 2428.

[2] IUT de Paris V et Université d'Orsay, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8628.

1. Introduction

In this paper, we consider the problem of the nonparametric estimation of the density g, of identically distributed variables X_i, from a sample $Z_1, \dots, Z_n$ where $Z_i = X_i + \varepsilon_i$, $i = 1, \dots, n$, X and ε are independent sequences and ε is a strong white noise with known density f_ε. Due to the independence between the X_i's and the ε_i's, the problem is to estimate g using the observations $Z_1, \cdots, Z_n$ with density

$$h(z) = \int g(u-z) f_\varepsilon(u) du.$$

The function f_ε is called the convolution kernel.

This deconvolution problem has been widely studied using kernel estimators and Fourier Transform properties. See for instance, Carroll and Hall (1988), Devroye (1989), Fan (1991a, b), Liu and Taylor (1989), Masry (1991, 1993a, b), Stefansky (1990), Stefansky and Carroll (1990), Taylor and Zhang (1990), Zhang (1990). Those estimators were studied from many points of view: pointwise and global asymptotic optimality, asymptotic normality, case of dependent ε_i's... Recently, Cator (2001) studies the rate of a pointwise kernel estimator under slightly more general conditions on the kernel and proves its optimality in most cases.

In the special case of circular deconvolution, Efromovich (1997) proposes an estimator built as a truncated development of g in an orthogonal basis, where the theoretical coefficients are replaced by estimators. He gives lower and upper bounds when the errors are super smooth. "Super smooth" for the errors density or for the convolution kernel, means that the Fourier transform has an exponential rate of decay, as opposed to "ordinary smooth", which means that the Fourier transform has a polynomial rate of decay. Rigorous definitions are given in Section 2.1.

In a slightly different context, Goldenshluger (1999) proposes an adaptive estimator of Sobolev-type density. Its pointwise risk converges with with the minimax rate when the Fourier transform $f_\varepsilon^*(t)$ of the errors density has exponential decay but with a rate that is logarithmically slower than the minimax rate under the assumption that $f_\varepsilon^*(t)$ decreases as a power of t.

Following ideas of the same type and more specifically the development of wavelet methods (see Donoho and Johnstone (1995), Donoho, Johnstone, Kerkyacharian and Picard (1996)), Pensky and Vidakovic (1999) study a wavelet thresholding method to build an adaptive estimator of the density in the convolution model. They construct linear and nonlinear wavelet estimators based on Meyer-type wavelets. Their estimators are asymptotically optimal and adaptive if g belongs to some Sobolev space and a priori adjusted when g is supersmooth. The same procedure is used by Pensky (2002), but using wavelets with bounded supports in order to better perform the estimation of very irregular functions g. Lastly, Fan and Koo (2002), use both wavelet with bounded supports or Meyer-type wavelets, to estimate densities belonging to Besov spaces $B_{\sigma,p,q}$ with $p < 2$. They establish lower bounds for the density deconvolution in both cases, ordinary and super-smooth errors. In particular, they show that no linear deconvolution estimator can achieve the optimal rates of convergence but that, when the errors are ordinary smooth, a non linear thresholding estimator is asymptotically minimax within logarithmic term, which is generally unavoidable in case of $B_{\sigma,p,q}$ with $p < 2$. Furthermore, when the errors are ordinary smooth, they construct an adaptive estimator which is minimax within a logarithmic factor.

Let us mention that very recently, Butucea (2003) and Butucea and Tsybakov (2003) study the estimation of the density by a kernel estimator and establish lower bounds for pointwise and $\mathbb{L}_2$-risk. In particular, Butucea and Tsybakov (2003) show that the rates computed by Pensky and Vidakovic (1999) when dealing with the so-called super-smooth functions in presence of super-smooth errors, are sub-optimal.

The aim of our paper is to construct an adaptive and optimal density estimator in the deconvolution problem, by using model selection. It is now well known (see Birgé and Massart (1997), Barron et al. (1999)) that thresholding methods can be viewed as model selection procedures based on relevant penalized contrast estimators. Our penalized estimator is constructed by using the orthonormal basis $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$, of the space of square integrable functions having a Fourier transform with compact support included into $[-\pi D_m, \pi D_m]$, that is $\varphi_{m,j}(x) = \sqrt{D_m}\varphi(D_m x - j)$ with $\varphi(x) = \sin(\pi x)/\pi x$, and $D_m = 2^m$. By using a basis different from Pensky and Vidakovic (1999)'s one, we are getting outside the wavelet framework and this allows to recover optimal rates. The projection estimator on this basis provides bias and variance of the same order as those obtained by Butucea (2003) and Butucea and Tsybakov (2003). Since the solution of the explicit trade-off between those squared bias term and variance term provides the optimal rate, it follows that the projection estimator based on $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ is asymptotically optimal for a well chosen m. Nevertheless, this optimal m depends on the smoothness of the density g. Therefore, we consider a penalized contrast estimator built in this basis, for which we provide non asymptotic $\mathbb{L}_2$-risk bounds that ensure an implicit trade-off between the squared bias term and a penalty term; the resulting penalty term has either the same order as the variance term, or the same order as the variance term up to some logarithmic factors in some case. It follows that our estimator is adaptive and asymptotically optimal, or asymptotically optimal up to some logarithmic factors. Consequently, our adaptive procedure improves the rates given by Pensky and Vidakovic (1999), when both the density and the errors are super smooth. It is noteworthy that these logarithmic factors only appear in case where the optimal rate is of order power of n, that is, only when such factors are negligible. By using this methodology we show how useful an adaptive procedure is. In particular, since in the case where both the density f_ε of the noise and the density g of the X_i's are super smooth, these optimal rates are not explicitly computable, it is of high interest that the risk bounds allow for an implicit trade-off: unknown optimal rates are reached automatically (up to possible negligible logarithmic factors). When f_ε is super smooth and g is ordinary smooth, then the optimal rates of convergence of the estimators are desperately slow (powers of $\ln(n)$). The particular interest of the case where g is super smooth is pointed out by Pensky and Vidakovic (1999), Butucea (2003) and Butucea and Tsybakov (2003) as a particularly interesting case since it allows to recover "fast" rates (power of n or faster than any power of $\ln(n)$). Nevertheless, the study of an optimal adaptive procedure when both g and f_ε are super smooth is not complete in Pensky and Vidakovic (1999).

Our procedure has some other advantages: it gives another light on the solution of the problem, and presents a very simple scheme of proof, corresponding to the standard scheme going with the methodology of model selection. Thanks to this simplicity, it allows straightforward extensions to the non independent case (β-mixing variables) using also standard related techniques. Moreover, it leads to some improvements, among which non asymptotic risk bounds, given in a very general way, and using linear (but penalized) estimators.

From a technical point of view, there are three key points in the methodology: the choice of the basis; the use of Talagrand's Inequality (1996) that allows to deal with the supremum of the empirical centered process in an independent framework (see Birgé and Massart (1997)), and Berbee's (1979) coupling Lemma, Delyon's (1990) covariance Inequality and Bryc's (1982) construction of approximating variables that allow to deal with the β-dependence. The paper is organized as follows. Section 2 gives a description of the problem with first the model and assumptions, and second, the estimators and the aim of the study. Section 3 contains the results in the independent framework and then in the β-mixing framework. The proofs are gathered in Section 4, and Section 5 collects some technical lemmas.

2. Description of the problem

2.1. Model and Assumptions.

Consider the model

$$Z_i = X_i + \varepsilon_i, \quad i = 1, \dots, n.$$

The following assumptions hold in all the sequel.

A1 The X_i's and the ε_i's are identically distributed random variables.

A2 The sequences $(X_i)_{i\in\mathbb{N}}$ and $(\varepsilon_i)_{i\in\mathbb{N}}$ are independent.

A3 The ε_i's and the X_i's are both independent random variables.

A3' The ε_i's and the X_i's are both absolutely regular (or β-mixing, see Doukhan (1994), pp.4-5).

Under **A2** and **A3**, the ε_i's and the X_i's are both independent and identically distributed random variables and therefore so is the sequence of the Z_i's.

Whereas, under **A2** and **A3'**, the sequence of the (X_i, ε_i)'s is also absolutely regular and therefore so is the sequence of the Z_i's. We denote by $(\beta_k)_{k\in\mathbb{N}}$ the mixing coefficients of this last sequence. Note that if $\beta_k(X)$ and $\beta_k(\varepsilon)$ denote the β-mixing coefficients of X and ε, then $\beta_k \le \beta_k(X) + \beta_k(\varepsilon)$.

We denote by g the unknown density of the X_i's and by f_ε the **known** density of the ε_i's. The following assumptions will be required for g.

A4 The function g is assumed to be square-integrable on $\mathbb{R}$.
A5 $\int x^2 g^2(x)dx \leq A_g < \infty$.

The problem we are interested in is to estimate g from a sample $(Z_1, \ldots, Z_n)$.
For u and v two square integrable functions, we denote by u^* the Fourier transform of u, $u^*(x) = \int e^{itx}u(t)dt$ and by $u*v$ the convolution product, $u*v(x) = \int u(t)v(x-t)dt$.
According to Assumption **A2** the (unknown density) h of the Z_i's equals $g*f_\varepsilon$, where f_ε denotes the density of the ε_i's. It follows that $h^* = g^* f_\varepsilon^*$. It is well known that the rate of convergence for estimating g is strongly related to the rate of decrease of the Fourier Transform of the errors density $f_\varepsilon(t)$ as t goes to infinity. More precisely, the smoother f_ε, the slower is the rate of convergence for estimating g. Nevertheless, this rate of convergence can be improved by assuming some additional regularity conditions on g. These regularity conditions are described by considering functions in the space:

$$\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha) = \left\{ g \text{ density } : \int_{-\infty}^{+\infty} |g^*(x)|^2 (x^2+1)^\alpha \exp\{2\rho|x|^\nu\}dx \leq A_\alpha \right\}, \tag{1}$$

where α, ρ, ν are nonnegative real numbers. When $\nu = 0$, this corresponds to Sobolev spaces of order α. When $\nu > 0$, this corresponds to analytic functions, often called "super-smooth" functions. For the sake of simplicity, we shall set the constant $\rho = 0$ when $\nu = 0$ and assume that $\rho > 0$ as soon as $\nu > 0$.
Moreover, we assume at the same time that f_ε is such that

A6 f_ε^* does not vanish and is square-integrable on $\mathbb{R}$.
A7 There exist nonnegative real numbers γ, B and δ such that, for all t in $\mathbb{R}$,

$$|f_\varepsilon^*(t)| \geq A_0(t^2+1)^{-\gamma/2} \exp\{-B|t|^\delta\}.$$

Assumption **A7** when $\delta = 0$ amounts to consider what is usually called "ordinary smooth" errors, and "super smooth" errors when $\delta > 0$. For the sake of simplicity, we set $B = 0$ when $\delta = 0$ and we assume that $B > 0$ when $\delta > 0$. The standard examples are the following : Gaussian or Cauchy distributions are super smooth of order $\gamma = 0, \delta = 2$ and $\gamma = 0, \delta = 1$ respectively, and the double exponential distribution is ordinary smooth ($\delta = 0$) of order $\gamma = 2$.

2.2. **The projection spaces and the estimators.** Let us define

$$\varphi(x) = \frac{\sin(\pi x)}{\pi x},$$

and let $\varphi_{m,j}(x) = \sqrt{D_m}\varphi(D_m x - j)$, where D_m is the analogue of the dimension in usual finite dimensional problems. We will consider the dyadic case $D_m = 2^m$ only, by analogy with wavelets but most results may be extended to the more general case of D_m varying in the set of all integers. It is well known (see for instance Meyer (1990), p.22) that $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ is an orthonormal basis of the space of square integrable functions having a Fourier transform with compact support included into

$[-\pi D_m, \pi D_m]$. We denote by S_m such a space and by $(S_m)_{m\in\mathcal{M}_n}$, with $\mathcal{M}_n = \{1,\dots,m_n\}$, this collection of linear spaces that is

$$\begin{aligned} S_m &= \mathrm{Vect}\{\varphi_{m,j},\ j\in\mathbb{Z}\},\\ &= \{f\in\mathbb{L}_2(\mathbb{R}),\ \text{with } \mathrm{supp}(f^*) \text{ included into } [-D_m\pi, D_m\pi]\}. \end{aligned}$$

Denoting by g_m the orthogonal projection of g on S_m, g_m is given by

$$g_m = \sum_{j\in\mathbb{Z}} a_{m,j}\varphi_{m,j} \text{ with } a_{m,j} = \int_{\mathbb{R}} \varphi_{m,j}(x)g(x)dx.$$

Associate this collection of models to the following contrast function, for t belonging to some S_m of the collection $(S_m)_{m\in\mathcal{M}_n}$

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

Here and in the following, we denote by $\|t\|^2 = \int t^2(x)dx$ and by $\langle s,t\rangle = \int s(x)t(x)dx$. Since

$$\mathbb{E}[u_t^*(Z_i)] = \frac{1}{2\pi}\langle \left(\frac{t^*}{f_\varepsilon^*}(-.)\right)^*, g * f_\varepsilon\rangle = \frac{1}{2\pi}\langle \frac{t^*}{f_\varepsilon^*}, g^* f_\varepsilon^*\rangle = \frac{1}{2\pi}\langle t^*, g^*\rangle = \langle t, g\rangle,$$

we find that $\mathbb{E}(\gamma_n(t)) = \|t-g\|^2 - \|g\|^2$ is all the smaller that t is nearer of g. This illustrates that $\gamma_n(t)$ is the natural choice for the empirical version oh the $\mathbb{L}_2$ distance between t and g.
Associated to the collection of models, the collection of estimators $\hat{g}_m$ of g is defined by

$$\hat{g}_m = \text{ Argmin }_{t\in S_m}\gamma_n(t), \tag{2}$$

where by using that, $t\mapsto u_t$ is linear, and that $(\varphi_{m,j})_{j\in\mathbb{Z}}$ is an orthonormal basis of S_m, we have

$$\hat{g}_m = \sum_{j\in\mathbb{Z}} \hat{a}_{m,j}\varphi_{m,j} \quad \text{where} \quad \hat{a}_{m,j} = \frac{1}{n}\sum_{i=1}^n u^*_{\varphi_{m,j}}(Z_i),$$

with $\mathbb{E}(\hat{a}_{m,j}) =< g, \varphi_{m,j} >= a_{m,j}$.
Note that the previous estimators $\hat{g}_m$ involve infinite sums and therefore require the estimation of an infinite number of coefficients $a_{m,j}$. In order to avoid this problem, we also consider the truncated spaces $S_m^{(n)}$ defined as

$$S_m^{(n)} = \mathrm{Vect}\{\varphi_{m,j}, |j|\le K_n\}$$

where K_n is an integer. Associated to those spaces we consider the collection of estimators $\hat{g}_m^{(n)}$ defined by

$$\hat{g}_m^{(n)} = \text{ Argmin }_{t\in S_m^{(n)}}\gamma_n(t). \tag{3}$$

It is easy to see that, $\{\varphi_{m,j}\}_{|j|\leq K_n}$ is an orthonormal basis of $S_m^{(n)}$, and we have

$$\hat{g}_m^{(n)} = \sum_{|j|\leq K_n} \hat{a}_{m,j}\varphi_{m,j} \quad \text{where} \quad \hat{a}_{m,j} = \frac{1}{n}\sum_{i=1}^{n} u^*_{\varphi_{m,j}}(Z_i),$$

with $\mathbb{E}(\hat{a}_{m,j}) =< g, \varphi_{m,j} >= a_{m,j}$.
Subsequently, $\hat{g}_m$ will be considered as a preliminary estimator and $\hat{g}_m^{(n)}$ as an effective estimator. The (non asymptotic) results will be stated on $\hat{g}_m$ in a first time and extended to $\hat{g}_m^{(n)}$ in a second time.

2.3. **The aim of the study.** In order to motivate our approach let us study the rate of convergence of one estimator $\hat{g}_m^{(n)}$, starting by the asymptotic properties of one preliminary estimator $\hat{g}_m$. According to (2), for any given m belonging to $\mathcal{M}_n$, $\hat{g}_m$ satisfies,

$$\gamma_n(\hat{g}_m) - \gamma_n(g_m) \leq 0.$$

Denoting by $\nu_n(t)$ the centered empirical process

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left[u_t^*(Z_i) - \langle t, g\rangle\right], \tag{4}$$

we have that

$$\gamma_n(t) - \gamma_n(s) = \|t-s\|^2 - 2\nu_n(t-s) - 2\langle t-s, g-s\rangle \tag{5}$$

and therefore, by using that $\langle g_m - \hat{g}_m, g - g_m\rangle = 0$, we deduce that $\| \hat{g}_m - g_m \|^2 \leq 2\nu_n(\hat{g}_m - g_m)$. Apply Pythagoras Theorem to get that $\|g - \hat{g}_m\|^2 = \|g - g_m\|^2 + \|g_m - \hat{g}_m\|^2$, and therefore

$$\|g - \hat{g}_m\|^2 \leq \|g - g_m\|^2 + 2\nu_n(\hat{g}_m - g_m) \tag{6}$$

Since $\hat{a}_{m,j} - a_{m,j} = \nu_n(\varphi_{m,j})$, we get that

$$\nu_n(\hat{g}_m - g_m) = \sum_{j\in\mathbb{Z}}(\hat{a}_{m,j} - a_{m,j})\nu_n(\varphi_{m,j}) = \sum_{j\in\mathbb{Z}}[\nu_n(\varphi_{m,j})]^2, \tag{7}$$

and consequently

$$\mathbb{E}\|g - \hat{g}_m\|^2 \leq \|g - g_m\|^2 + 2\sum_{j\in\mathbb{Z}}\mathrm{Var}[\nu_n(\varphi_{m,j})]. \tag{8}$$

The rate of convergence of $\hat{g}_m$ is obtained by selecting the space S_m that makes $\|g - g_m\|^2 + 2\sum_{j\in\mathbb{Z}}\mathrm{Var}[\nu_n(\varphi_{m,j})]$ as small as possible. Let us study the order of the bias term $\|g - g_m\|^2$ which depends on the smoothness of the function g. By combining the fact that g_m is the orthogonal projection of g on S_m, closed subspace of functions $f \in \mathbb{L}^2(\mathbb{R})$ such that $\mathrm{supp}(f^*) \subset [-D_m\pi, D_m\pi]$, and that $g*\varphi_{m,0}$ belongs to S_m, we get that $\|g-g_m\|^2 = \inf_{f\in S_m}\|g-f\|^2$ is bounded by $\|g-\sqrt{D_m}g*\varphi_{m,0}\|^2$, and therefore

$$\|g-g_m\|^2 \leq \frac{1}{2\pi}\|g^*-\sqrt{D_m}g^*\varphi^*_{m,0}\|^2 \leq \frac{1}{2\pi}\int_{|x|\geq \pi D_m}|g^*|^2(x)dx.$$

Consider now that g belongs $\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)$ defined by (1). The above bound implies immediately that the squared bias term is less than $A_\alpha(D_m^2\pi^2+1)^{-\alpha}\exp\{-2\rho\pi^\nu D_m^\nu\}$. Now assume for a moment that the Z_i's are independent (i.e. that the X_i's and the ε_i's are independent and identically distributed random variables); this implies that

$$\mathrm{Var}[\nu_n(\varphi_{m,j})] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left[u^*_{\varphi_{m,j}}(Z_i)\right] = \frac{1}{n}\mathrm{Var}\left[u^*_{\varphi_{m,j}}(Z_1)\right].$$

Apply Lemma 5.1 to get that

$$\sum_{j\in\mathbb{Z}}\mathrm{Var}[\nu_n(\varphi_{m,j})] \leq D_m\Delta_1(m)/(2\pi n)$$

where

$$\Delta_1(m) = \int |\varphi^*(x)/f^*_\varepsilon(D_m x)|^2 dx, \tag{9}$$

which is bounded under **A7** in the following way,

$$\frac{D_m\Delta_1(m)}{2\pi n} \leq \frac{1}{\pi n A_0^2}(D_m^2\pi^2+1)^\gamma \int_0^{\pi D_m}\exp\{2Bu^\delta\}du.$$

If $\delta=0$, the convention $B=0$ implies that

$$\int_0^{\pi D_m}\exp\{2Bu^\delta\}du \leq \pi D_m.$$

Consider now the case $0<\delta\leq 1$. Easy calculations provide that

$$\begin{aligned}\int_0^{\pi D_m}\exp\{2Bu^\delta\}du &= \int_0^{\pi D_m}\left(2B\delta u^{\delta-1}\exp\{2Bu^\delta\}\right)\frac{du}{2B\delta u^{\delta-1}}\\ &\leq \frac{(\pi D_m)^{1-\delta}}{2B\delta}\left[\exp(2Bu^\delta)\right]_0^{\pi D_m}\end{aligned}$$

and therefore

$$\int_0^{\pi D_m}\exp\{2Bu^\delta\}du \leq \frac{(\pi D_m)^{1-\delta}}{2B\delta}\exp(2B(\pi D_m)^\delta).$$

Now, if $\delta>1$, then by using that $u^\delta = u^{\delta-1}u$ we get that

$$\int_0^{\pi D_m}\exp\{2Bu^\delta\}du \leq \int_0^{\pi D_m}\exp\{2B(\pi D_m)^{\delta-1}u\}du$$

and consequently

$$\int_0^{\pi D_m}\exp\{2Bu^\delta\}du \leq \frac{(\pi D_m)^{1-\delta}}{2B}\exp(2B(\pi D_m)^\delta).$$

If we denote by $R(B,\delta)$ the constant

$$R(B,\delta) = \mathbb{1}_{\{\delta=0\}} + \frac{1}{2B(\delta\wedge 1)}\mathbb{1}_{\{\delta>0\}} = \begin{cases} 1 & \text{if } \delta = 0 \\ 1/2B\delta & \text{if } 0<\delta\leq 1 \\ 1/2B & \text{if } \delta>1 \end{cases},$$

then

$$\begin{aligned} \sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})] &\leq \frac{D_m\Delta_1(m)}{2\pi n} \\ &\leq \frac{1}{\pi A_0^2 R(B,\delta) n}(\pi D_m)^{1-\delta}(D_m^2\pi^2+1)^\gamma \exp\{2B\pi^\delta D_m^\delta\}, \end{aligned} \tag{10}$$

and consequently, according to (8)

$$\mathbb{E}(\|g-\hat{g}_m\|^2) \leq \|g-g_m\|^2 + 2\lambda_1(\gamma, A_0, B,\delta)\frac{D_m^{(2\gamma+1-\delta)}\exp\left\{2B\pi^\delta D_m^\delta\right\}}{n}$$

with $\lambda_1 = \lambda_1(\gamma, A_0, B, \delta)$ defined by

$$\lambda_1 = \frac{(\pi^2+1)^\gamma}{\pi^\delta A_0^2 R(B,\delta)}. \tag{11}$$

Now if we consider that the Z_i's are absolutely regular variables, according to Inequality (42) in Lemma 5.1 (see Appendix), based on the Delyon 's (1990) covariance Inequality, successfully exploited by Viennet (1997), Inequality (10) becomes

$$\begin{aligned} \sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})] &\leq 4(\sum_k \beta_k) D_m\Delta_1(m)/(2\pi n) \\ &\leq 4(\sum_k \beta_k)\lambda_1(\gamma, A_0, B,\delta)\frac{D_m^{2\gamma+1-\delta}\exp\{2B\pi^\delta D_m^\delta\}}{n}. \end{aligned} \tag{12}$$

Therefore,

$$\mathbb{E}(\|g-\hat{g}_m\|^2) \leq \|g-g_m\|^2 + 8(\sum_k \beta_k)\lambda_1(\gamma, A_0, B,\delta)\frac{D_m^{(2\gamma+1-\delta)}\exp\left\{2B\pi^\delta D_m^\delta\right\}}{n}.$$

We now give the analogous bound of the risk for one effective estimator $\hat{g}_m^{(n)}$. By applying Lemma 5.4 (in the Appendix below), and Bounds (10) or (12), it follows that the truncated estimator has its risk bounded in the following way. In the independent case, we have

$$\begin{aligned} \mathbb{E}(\|g-\hat{g}_m^{(n)}\|^2) \leq \|g-g_m\|^2 &+ 2\lambda_1(\gamma, A_0, B,\delta)\frac{D_m^{(2\gamma+1-\delta)}\exp\left\{2B\pi^\delta D_m^\delta\right\}}{n} \\ &+ \frac{D_m^2}{K_n}(\|\varphi\|^2 A_g + 1). \end{aligned} \tag{13}$$

In the β-mixing case, we have

(14)

$$\mathbb{E}(\|g-\hat{g}_m^{(n)}\|^2) \le \|g-g_m\|^2 + 8(\sum_k \beta_k)\lambda_1(\gamma, A_0, B, \delta)\frac{D_m^{(2\gamma+1-\delta)}\exp\left\{2B\pi^\delta D_m^\delta\right\}}{n} + \frac{D_m^2}{K_n}(\|\varphi\|^2 A_g + 1).$$

It follows that we can find a relevant choice of K_n (independent of g) that makes the last supplementary term unconditionally negligible with respect to the others, and consequently, the risk of $\hat{g}_m^{(n)}$ has the same order as the risk of $\hat{g}_m$.
Let us now study this risk. If g belongs to $\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)$ with $\nu = 0$ and f_ε satisfies Assumption **A7** with $\delta = 0$, then $\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})]$ has the order $D_m^{1+2\gamma}/n$. On the other hand, $\|g-g_m\|^2$ has the order $D_m^{-2\alpha}$, which leads to choose the space $S_{\breve{m}}$ with dimension $D_{\breve{m}} = n^{1/(2\alpha+2\gamma+1)}$. Therefore the estimator $g_{\breve{m}}$ attains the rate $n^{-2\alpha/(2\alpha+2\gamma+1)}$, which is known to be the optimal rate (see Fan (1991a)). We can see that this choice of $D_{\breve{m}}$ depends on the unknown smoothness of g, α. This motivates us to complete the procedure by an automatic selection of the space via a penalization of the contrast. Table 1 presents the optimal choices of D_m and the associated rates in function of the smoothness of the errors ε and of the function g, the first one being known and not the second one.

f_ε / g	$\delta = 0$	$\delta > 0$
$\nu = 0$	$\pi D_{\breve{m}} = n^{1/(2\alpha+2\gamma+1)}$ rate $n^{-2\alpha/(2\alpha+2\gamma+1)}$	$\pi D_{\breve{m}} = [\ln(n)/(2B+1)]^{1/\delta}$ rate $(\ln(n))^{-2\alpha/\delta}$
$\nu > 0$	$\pi D_{\breve{m}} = \left[\frac{\ln(n)}{2\rho}\right]^{1/\nu}$ rate $\dfrac{\ln(n)^{(2\gamma+1)/\nu}}{n}$	Non explicit optimal $D_{\breve{m}}$ and consequently non explicit rate.

Table 1. Optimal choice of the dimension ($D_{\breve{m}}$) and obtained (optimal) rates.

The main comment here is that, these rates of convergence being the same as those obtained in Butucea (2003) and Butucea and Tsybakov (2003) by using a kernel estimator associated corresponding lower bounds, they are thus the optimal ones.

Some others comments follow.

Remark 2.1. When $\delta > 0$ and $\nu = 0$, the optimal $D_{\breve{m}}$ does not depend on the density g. It follows that $\hat{g}_m$ (and consequently $\hat{g}_m^{(n)}$) is adaptive, provided that we know that we are in this case. This was already the case for Pensky and Vidakovic (1999) estimator.

The case $\delta > 0, \nu > 0$ requires a specific discussion.

Remark 2.2. Firstly, it is noteworthy that when both densities g and f_ε are super-smooth ($\nu > 0, \delta > 0$), the rate of convergence of the estimator is faster than those given by Pensky and Vidakovic (1999). This comes from the fact that their scaling function φ satisfies $\varphi^*(t) = 1$ on $[-2\pi/3, 2\pi/3]$ with support included into the interval $[-4\pi/3, 4\pi/3]$ and the wavelet ψ has its support included into the set

$$[-8\pi/3, -2\pi/3] \cup [2\pi/3, 8\pi/3].$$

Consequently the bias term of their estimator is of order $D_m^{-2\alpha}\exp\{-2\rho(2\pi/3)^\nu D_m^\nu\}$ with a variance term of order $D_m^{2\gamma+1-\delta}\exp\{2B(4\pi/3)^\delta D_m^\delta\}$. The rate of convergence can therefore be improved by using the function $\sin(\pi x)/(\pi x)$ which has a Fourier transform $\mathbb{1}_{[-\pi,\pi]}$ and gives a better trade-off between the squared bias of order $D_m^{-2\alpha}\exp\{-2\rho\pi^\nu D_m^\nu\}$ and the variance of order $D_m^{2\gamma+1-\delta}\exp\{2B\pi^\delta D_m^\delta\}$.

Secondly, Table 1 seems incomplete in that case because the explicit $D_{\breve{m}}$ is missing. In fact, the optimal dimension $D_{\breve{m}}$ is the solution of the following equation

$$nO(1) = D_{\breve{m}}^{2\alpha+2\gamma+1-\nu}\exp\{2B\pi^\delta D_{\breve{m}}^\delta + 2\rho\pi^\nu D_{\breve{m}}^\nu\}, \tag{15}$$

which has not an explicit form solution for general ν and δ.
For instance, if $\nu < \delta$, we have to distinguish if $\nu/\delta \le 1/2$ or $1/2 < \nu/\delta \le 2/3, \dots$. More precisely, if $\nu/\delta \le 1/2$, the optimal choice $D_{\breve{m}}$ is

$$\pi D_{\breve{m}} = \left[\frac{\ln(n)}{2B} - \frac{2\rho}{2B}\left(\frac{\ln(n)}{2B}\right)^{\nu/\delta} - c\ln\left(\frac{\ln(n)}{2B}\right)\right]^{1/\delta} \quad \text{with } c = \frac{2\gamma-\nu+2\alpha+1}{2B\delta}$$

and the rate is

$$\ln(n)^{-2\alpha/\delta}\exp\left[-2\rho\left(\frac{\ln(n)}{2B}\right)^{\nu/\delta}\right].$$

If $1/2 < \nu/\delta \le 2/3$ the optimal choice of $\pi D_{\breve{m}}$ is

$$\pi D_{\breve{m}} = \left[\frac{\ln(n)}{2B} - \frac{2\rho}{2B}\left(\frac{\ln(n)}{2B}\right)^{\nu/\delta} + \frac{\nu}{\delta}\frac{(2\rho)^2}{2B}\left(\frac{\ln(n)}{2B}\right)^{2\nu/\delta-1} - c\ln\left(\frac{\ln(n)}{2B}\right)\right]^{1/\delta}$$

with the same c as above, which gives the rate

$$\ln(n)^{-2\alpha/\delta}\exp\left[-2\rho\left(\frac{\ln(n)}{2B}\right)^{\nu/\delta} + \frac{(2\rho)^2}{2B}\frac{\nu}{\delta}\left(\frac{\ln(n)}{2B}\right)^{2\nu/\delta-1}\right].$$

If $2/3 < \nu/\delta \le 3/4$, we have another choice of $\pi D_{\breve{m}}$ with an other rate. This process stops when we find some integer k such that ν/δ belongs to the interval

$I_k =]k/(k+1); (k+1)/(k+2))]$.

Note that the particular case $\nu = \delta$ leads to the explicit solution

$$\pi D_{\breve{m}} = \{[\ln(n/\ln(n)^a)/(2B+2\rho)]\}^{1/\nu} \text{ with } a = \frac{2\alpha + 2\gamma - \nu + 1}{\nu}$$

and to the rate

$$[\ln(n)]^b n^{-\rho/(\rho+B)} \text{ with } b = \frac{-2\alpha B + (2\gamma - \nu + 1)\rho}{\nu(B+\rho)}.$$

Thirdly, it must be noted that in any case even if we choose $\pi D_{\breve{m}} = [\ln(n)/(2B+1)]^{1/\delta}$, which is not the solution of (15), we get a (non optimal) rate which is faster than any power of $\ln(n)$. Therefore, the optimal rates obtained in this case are always faster than any power of $\ln(n)$, and can reach powers of n (see the case $\nu = \delta$ right above). This is why some possible loss in such rates of order a power of $\ln(n)$ are negligible (see Remark 3.1 below).

Remark 2.3. When $\delta = 0$, we can see that the rate $n^{-2\alpha/(2\alpha+2\gamma+1)}$ can be improved if the function g is super smooth ($\nu > 0$) and the adaptive procedure has the feature to reach automatically this better rate without requiring any information about the smoothness of g. The problem may seem quite different when we consider super smooth errors. Indeed if we assume that $\delta > 0$, together with ordinary smooth functions ($\nu = 0$), then we need to choose $D_{\breve{m}} = [\ln(n)/(2B+1)]^{1/\delta}$ and we find the rate $(\ln(n))^{-2\alpha/\delta}$, which is also optimal. This second choice can be done directly since it only depends on known constants. This is the reason why Pensky and Vidakovik (1999) considered the adaptive procedure useless in that case. Nevertheless, it is easy to see that if the function g is super smooth ($\nu > 0$), then the choice of D_m depends on the relative position of ν and δ as explained in Remark 2.2. This implies that an adaptive procedure remains a good strategy, even if in that case, it implies a small loss in the rate as a price of ignoring how smooth is the unknown function g.

3. The results

The model selection procedure is required (see Section 2.3) when the optimal choice of the space $S_m^{(n)}$, or equivalently S_m (here the choice of m) depends on the unknown function g. We aim at finding the best model $\hat{m}$ in $\mathcal{M}_n$, based on the data and not on prior information on g, such that the risk of the resulting estimator is almost as good as the risk of the best estimator in the family. The model selection is performed in an automatic way, using the following penalized criteria

$$\tilde{g} = \hat{g}_{\hat{m}} \text{ with } \hat{m} = \operatorname{Argmin}_{m \in \mathcal{M}_n} [\gamma_n(\hat{g}_m) + \operatorname{pen}(m)] \tag{16}$$

for the preliminary estimator and

$$\tilde{g}^{(n)} = \hat{g}_{\hat{m}}^{(n)} \text{ with } \hat{m} = \operatorname{Argmin}_{m \in \mathcal{M}_n} \left[\gamma_n(\hat{g}_m^{(n)}) + \operatorname{pen}(m)\right] \tag{17}$$

for the effective estimator. In both cases, pen is a penalty function, based on the observations, that we would like to find such that

$$\mathbb{E} \parallel \tilde{g}^{(n)} - g \parallel^2 \leq \inf_{m\in\mathcal{M}_n} \left[\parallel g - g_m \parallel^2 + 2\lambda_1 \frac{D_m^{2\gamma+1-\delta}\exp\{2B(\pi D_m)^\delta\}}{n} \right]. \tag{18}$$

in the independent case or such that

(19)
$$\mathbb{E} \parallel \tilde{g}^{(n)} - g \parallel^2 \leq \inf_{m\in\mathcal{M}_n} \left[\parallel g - g_m \parallel^2 + 8(\sum_k \beta_k)\lambda_1 \frac{D_m^{2\gamma+1-\delta}\exp\{2B(\pi D_m)^\delta\}}{n} \right],$$

in the β-mixing case.

3.1. **The independent framework.** The following theorem states the result on the preliminary estimator $\tilde{g}$.

Theorem 3.1. *Consider the model described in section 2.1 under* **A1**-**A4**, *and* **A6**-**A7**, *and the collection of estimators $\hat{g}_m$ defined by (2) for $1 \leq m \leq m_n \leq \log_2(n)$. Let $\lambda_1 = \lambda_1(\gamma, A_0, B, \delta)$ be defined by (11) and $\lambda_2 = \lambda_2(\gamma, A_0, B, \delta)$ be defined by*

$$\lambda_2(\gamma, A_0, B, \delta) = \begin{cases} \lambda_1(\gamma, A_0, B, \delta) & \text{if } \delta > 1, \\ \lambda_1^{1/2}(\gamma, A_0, B, \delta)(1+\pi^2)^{\gamma/2}\|f_\varepsilon\|A_0^{-1}(2\pi)^{-1/2} & \text{if } \delta \leq 1. \end{cases} \tag{20}$$

1) *Let $0 \leq \delta \leq 1/3$. Let $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (16) with*

$$\text{pen}(m) = \kappa[\lambda_1(\gamma, A_0, B, \delta) + B\pi^\delta \lambda_2(\gamma, A_0, B, \delta)] \frac{D_m^{2\gamma+1-\delta}\exp\{2B(\pi D_m)^\delta\}}{n},$$

for some universal numerical constant κ.
2) *Let $\delta > 1/3$. Let $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (16) with*

$$\text{pen}(m) = \kappa[\lambda_1(\gamma, A_0, B, \delta) + B\pi^\delta \lambda_2(\gamma, A_0, B, \delta)] \frac{D_m^{2\gamma+((1/2+\delta/2)\wedge 1)}\exp\{2B(\pi D_m)^\delta\}}{n}$$

where κ is some universal numerical constant.
Then in these two cases, $\tilde{g}$ satisfies

$$\mathbb{E}(\|g - \tilde{g}\|^2) \leq K \inf_{m\in\{1,\dots,m_n\}} [\|g - g_m\|^2 + \text{pen}(m)] + \frac{c}{n}, \tag{21}$$

where K and c are constants depending on f_ε and on the mixing coefficients.

Corollary 3.1. *Consider the model described in section 2.1 under* **A1**-**A7**, *and the collection of estimators $\hat{g}_m^{(n)}$ defined by (3) for $1 \leq m \leq m_n \leq \log_2(n)$.*
Then the result (21) of Theorem 3.1 holds with $\tilde{g}$ replaced by $\tilde{g}^{(n)}$ and the same penalty functions provided that $\lim_{n\to+\infty} n(\ln n)^{2-(2\gamma+1)/\nu}K_n^{-1} = 0$.

This corollary comes by combining Bound (13) and Theorem 3.1.

Remark 3.1. The rates are easy to deduce from (21) as soon as g belongs to some space $\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)$ defined by (1). The procedure will reach the rate without requiring the knowledge of α, ν and ρ. Indeed, if g belongs to $\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)$, we have $\|g-g_m\|^2 \leq CD_m^{-2\alpha}\exp\{-2\rho\pi^\nu D_m^\nu\}$, which, associated to the values of pen(m), automatically gives a rate. In the cases, $\delta = 0$, or $0 < \delta \leq 1/3$, the penalty function pen(m) is, as expected in (18), of order $D_m^{2\gamma+1-\delta}\exp(2B(\pi D_m)^\delta)/n$, that is of the variance order. It follows that in those cases, the resulting rates are optimal. In the other case, namely when $\delta > 1/3$, the penalty function pen(m) has not exactly of the order of the variance, but is of order $D_m^{2\gamma+((1/2+\delta/2)\wedge 1)}\exp(2B(\pi D_m)^\delta)/n$, with a loss of order $D_m^{(3\delta/2-1/2)\wedge\delta}$ that is of order $D_m^{(3\delta-1)/2}$ if $1/3 < \delta \leq 1$ and of order D_m^δ if $\delta > 1$. It follows that the rate remains optimal if the bias $\|g-g_m\|^2$ is the dominating term in the trade-off between $\|g-g_m\|^2$ and pen(m), which happens for instance, when $\nu = 0$. When pen(m) is the dominating term in the trade-off between $\|g-g_m\|^2$ and pen(m), there is a loss of order a power of $D_m^{(3\delta/2-1/2)\wedge\delta}$. Note that it happens in cases, where D_m is of order $(\ln n)^{1/\delta}$ and consequently the loss is of order at most $\ln n$, when the rate is faster than logarithmic: therefore the loss happens only in cases when it can be seen as negligible. Consequently, the adaptive procedure remains a good strategy, even if in that case, it implies a small loss in the rate as a price to pay for ignoring how smooth is the unknown function g. This strategy is all the more relevant that it solves almost optimally the problem in cases where the rate is impossible to compute analytically (see Remark 2.2).

3.2. **The β-mixing framework.** We consider now the β-mixing case.

Theorem 3.2. *Consider the model described in section 2.1 under* **A1**, **A2**, **A3'**, **A4**, **A6** *and* **A7**, *and the collection of estimators $\hat{g}_m$ defined by (2) for $1 \leq m \leq m_n \leq \log_2(n)$ and pen(m_n) bounded by some constant. Assume moreover that the Z_i's are arithmetically β-mixing, that is $\beta_k \leq Ck^{-(1+\theta)}$, for all $k \in \mathbb{N}$, with $\theta > 3$. Let $\lambda_1(\gamma, A_0, B, \delta)$ be defined by (11) and let $\lambda_2^\star = \lambda_2^\star(\gamma, A_0, B, \delta)$ defined by*

(22)
$$\lambda_2^\star(\gamma, A_0, B, \delta) = \begin{cases} \lambda_1(\gamma, A_0, B, \delta)(\sum_{k\in\mathbb{N}}\beta_k) & \text{if } \delta > 1 \\ \lambda_1^{1/2}(\gamma, A_0, B, \delta)(1+\pi^2)^{\gamma/2}\|h\|_\infty^{1/2}(\sum_{k\in\mathbb{N}}(1+k)\beta_k)^{1/2}A_0^{-1} & \text{if } \delta \leq 1 \end{cases}.$$

1) *Let $0 \leq \delta \leq 1/3$. Let $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (16) with*

$$\mathrm{pen}(m) = \kappa[\lambda_1(\gamma, A_0, B, \delta) + B\pi^\delta\lambda_2^\star(\gamma, A_0, B, \delta)](\sum_{k\in\mathbb{N}}\beta_k)\frac{D_m^{2\gamma+1-\delta}\exp\{2B(\pi D_m)^\delta\}}{n},$$

for some universal numerical constant κ.
2) *Let $\delta > 1/3$ and let $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (16) with*

$$\mathrm{pen}(m) = \kappa[\lambda_1(\gamma, A_0, B, \delta)+B\pi^\delta\lambda_2^\star(\gamma, A_0, B, \delta)](\sum_{k\in\mathbb{N}}\beta_k)\frac{D_m^{2\gamma+((1/2+\delta/2)\wedge 1)}\exp\{2B(\pi D_m)^\delta\}}{n}$$

where κ is some universal numerical constant.
Then in these two cases, $\tilde{g}$ satisfies

$$\mathbb{E}(\|g-\tilde{g}\|^2) \leq K \inf_{m\in\{1,\dots,m_n\}} \left[\|g-g_m\|^2 + \mathrm{pen}(m)\right] + \frac{c}{n}, \tag{23}$$

where K and c are constants depending on f_ε and on the mixing coefficients.

Corollary 3.2. *Consider the model described in section 2.1 under* **A1**, **A2**, **A3'** *and* **A4**-***A7***, *and the collection of estimators $\hat{g}_m^{(n)}$ defined by (3) for $1 \leq m \leq m_n \leq \log_2(n)$. Assume moreover that Z_i's are arithmetically β-mixing, that is $\beta_k \leq Ck^{-(1+\theta)}$, for all $k \in \mathbb{N}$, with $\theta > 3$.*
Then the result (23) of Theorem 3.2 holds with $\tilde{g}$ replaced by $\tilde{g}^{(n)}$ and the same penalty functions provided that $\lim_{n\to+\infty} n(\ln n)^{2-(2\gamma+1)/\nu}K_n^{-1} = 0$.

This corollary comes by combining Bound (14) and Theorem 3.2.
Some remarks immediately follow.

Remark 3.2. Obviously the two previous results hold in the geometrical β-mixing case, that is when $\beta_k \leq Ce^{-\theta k}$, for all $k \in \mathbb{N}$ with no condition on the rate θ.

Remark 3.3. The penalty found in the mixing framework contains two unknown coefficients, namely the norm $\|h\|_\infty$ which depends on the function to be estimated and the term $\sum_k \beta_k$, which depends on the mixing coefficients. The first one may be replaced by an estimator, but no estimator has been found yet to replace the second one. From a practical point of view, there exists some methods for determining the constants in the penalties; those practical methods are known to give good results and are often preferred to methods where the unknown term is replaced by an estimator, in the cases where an estimator is available (which is not the case here). We refer for details about it to Birgé and Rozenholc (2001) or Comte and Rozenholc (2001).

Remark 3.4. The previous results are analogous to the results obtained in the independent case, up to the constants. Therefore the comments in remark 3.1 apply here.

4. Proofs

4.1. **Proof of Theorem 3.1 : the i.i.d. case.** The following decompositions illustrate that the deconvolution problem can be treated using a classical model selection scheme (see for instance Birgé and Massart (1997) or Comte and Merlevède (2002)). By definition, $\tilde{g}$ satisfies that for all $m \in \mathcal{M}_n$,

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

Therefore, by applying (5) and that $2ab \leq x^{-1}a^2 + xb^2$ we get

$$\begin{aligned} \| \tilde{g} - g_m \|^2 &\leq 2\nu_n(\tilde{g} - g_m) + 2\langle \tilde{g} - g_m, g - g_m \rangle + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}) \\ &\leq 2\nu_n(\tilde{g} - g_m) + \left[\frac{1}{4} \| \tilde{g} - g_m \|^2 + 4 \| g - g_m \|^2\right] \\ &\quad + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}). \end{aligned} \tag{24}$$

Next, we use that if $t = t_1 + t_2$ with t_1 in S_m and t_2 in $S_{m'}$, then t is such that t^* has its support in $[-\pi D_{m\vee m'}, \pi D_{m\vee m'}]$ and therefore t belongs to $S_{m\vee m'}$. Denote by $B_{m,m'}(0,1)$ the set

$$B_{m,m'}(0,1) = \{t \in S_{m\vee m'} \ / \ \|t\| = 1\}.$$

It follows that

$$|\nu_n(\tilde{g} - g_m)| \leq \|\tilde{g} - g_m\| \sup_{t\in B_{m,\hat{m}}(0,1)} |\nu_n(t)|.$$

Consequently,

$$\|\tilde{g} - g_m\|^2 \leq \frac{1}{2}\|\tilde{g} - g_m\|^2 + 4\|g - g_m\|^2 + \mathrm{pen}(m) + 4 \sup_{t\in B_{m,\hat{m}}(0,1)} \nu_n^2(t) - \mathrm{pen}(\hat{m})$$

and therefore

$$\frac{1}{2}\|\tilde{g} - g_m\|^2 \leq 4\|g - g_m\|^2 + \mathrm{pen}(m) + 4 \sup_{t\in B_{m,\hat{m}}(0,1)} \nu_n^2(t) - \mathrm{pen}(\hat{m}).$$

Use that $\| \tilde{g} - g \|^2 \leq 2 \| \tilde{g} - g_m \|^2 + 2 \| g_m - g \|^2$ and choose some positive function $p(m, m')$ such that $4p(m, m') \leq \mathrm{pen}(m) + \mathrm{pen}(m')$, to get that

$$\begin{aligned} \frac{1}{4} \| \tilde{g} - g \|^2 &\leq \frac{9}{2} \| g - g_m \|^2 + 2\mathrm{pen}(m) + 4 \left(\sup_{t\in B_{m,\hat{m}}(0,1)} \nu_n^2(t) - p(m, \hat{m})\right)_+ \\ &\leq \frac{9}{2} \| g - g_m \|^2 + 2\mathrm{pen}(m) + 4 \sum_{m'\in\mathcal{M}_n} \left(\sup_{t\in B_{m,m'}(0,1)} \nu_n^2(t) - p(m, m')\right)_+ . \end{aligned} \tag{25}$$

The main point of the proof lies in studying $W_n(m')$ defined as

$$W_n(m') := \left[\left(\sup_{t\in B_{m,m'}(0,1)} |\nu_n(t)|\right)^2 - p(m, m')\right]_+ , \tag{26}$$

and more precisely in finding $p(m, m')$ such that

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_n(m')) \leq C/n, \tag{27}$$

where C is a constant. Combining (25) and (27) we infer that, for all m in $\mathcal{M}_n$,

$$\mathbb{E}\|g - \tilde{g}\|^2 \leq 18\|g - g_m\|^2 + \frac{16C}{n} + 8\mathrm{pen}(m),$$

which can also be written

$$\mathbb{E}\|g-\tilde{g}\|^2 \le K \inf_{m\in\mathcal{M}_n}\left[\|g-g_m\|^2+\mathrm{pen}(m)\right]+\frac{16C}{n}, \tag{28}$$

where $K=18$ suits. If $\mathrm{pen}(m)$ has the same order as the variance order $D_m\Delta_1(m)/n$, then equation (28) guarantees an automatic trade-off between the squared bias term $\|g-g_m\|^2$ and the variance term, up to some multiplicative constant and (18) follows. It remains thus to find $p(m,m')$ such that (27) holds. In the i.i.d. case, (27) follows by applying the following version of Talagrand's Inequality (see Talagrand (1996)) given below.

Lemma 4.1. *Let $X_1,\dots,X_n$ be i.i.d. random variables and $\nu_n(f)$ be defined by $\nu_n(f)=(1/n)\sum_{i=1}^n[f(X_i)-\mathbb{E}(f(X_i))]$ for f belonging to a countable class $\mathcal{F}$ of uniformly bounded measurable functions. Then*

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}|\nu_n(f)|^2-2(4+\xi^2)H^2\right]_+ \le \frac{6}{K_1}\left(\frac{v}{n}e^{-K_1\xi^2\frac{nH^2}{v}}+\frac{4M_1^2}{K_1n^2}e^{-\frac{K_1\xi}{\sqrt{2}}\frac{nH}{M_1}}\right), \tag{29}$$

where K_1 is a universal constant,

$$\sup_{f\in\mathcal{F}}\|f\|_\infty\le M_1,\quad \mathbb{E}\left(\sup_{f\in\mathcal{F}}|\nu_n(f)|\right)\le H,\quad \sup_{f\in\mathcal{F}}\mathrm{Var}(f(X_1))\le v.$$

We refer to Comte and Merlevède (2002) for the integration of the standard Talagrand's Inequality which gives straightforwardly (29). 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 $m^*=m\vee m'$. Combining Lemma 5.2 and Inequality (10) we propose to take

$$H^2=H^2(m^*)=\lambda_1(\gamma,A_0,B,\delta)D_{m^*}^{2\gamma+1-\delta}\exp\{2B(\pi D_{m^*})^\delta\}/n \text{ and } \quad M_1=\sqrt{nH^2},$$

where $\lambda_1=\lambda_1(\gamma,A_0,B,\delta)$ is defined by (11). Again, by applying Lemma 5.2, we take $v\ge\Delta_2(m,h)$ with

$$\Delta_2(m,h)=D_m^2\iint\left|\frac{\varphi^*(x)\varphi^*(y)}{f_\varepsilon^*(D_mx)f_\varepsilon^*(D_my)}h^*(D_m(x-y))\right|^2dxdy. \tag{30}$$

For $\delta>1$ we use a rough bound for $\Delta_2(m,h)$ given by $\sqrt{\Delta_2(m^*,h)}\le 2\pi nH^2$. When $\delta\le 1$, write that

$$\begin{aligned}\Delta_2(m,h) &\le \frac{D_m^2}{A_0^2}(1+\pi^2D_m^2)^\gamma\exp\{2B(\pi D_m)^\delta\}\int_{-\pi}^{\pi}\frac{dx}{|f_\varepsilon^*(D_mx)|^2}\int|h^*(D_mu)|^2du\\ &\le \frac{2\pi\lambda_1(1+\pi^2)^\gamma\|h^*\|^2}{A_0^2}D_m^{4\gamma+1-\delta}\exp\{4B(\pi D_m)^\delta\}.\end{aligned}$$

Using that $\|h^*\|^2\le\|f_\varepsilon^*\|^2<\infty$ by Assumption **A6**, we propose to take

$$v=\lambda_2D_{m^*}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)}\exp\{2B(\pi D_{m^*})^\delta\},$$

where $\lambda_2=\lambda_2(\gamma,A_0,B,\delta)$ is defined in Theorem 3.1.

From the definition (26) of $W_n(m')$, by taking $p(m,m') = 2(4+\xi^2)H^2$, we get that

$$\mathbb{E}(W_n(m')) \le \mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} |\nu_n(t)|^2 - 2(4+\xi^2)H^2\right]_+ .$$

Applying (29), we get the global bound

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_n(m')) \le K \sum_{m'\in\mathcal{M}_n} [I(m^*) + II(m^*)],$$

where $I(m^*)$ and $II(m^*)$ are defined by

$$I(m^*) = \frac{\lambda_2 D_{m^*}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)} \exp\{2B(\pi D_{m^*})^\delta\}}{n} e^{-K_1\xi^2(\lambda_1/\lambda_2)D_{m^*}^{(1/2-\delta/2)+}}$$

$$\text{and} \qquad II(m^*) = \frac{\lambda_1 D_{m^*}^{2\gamma+1-\delta} e^{2B(\pi D_{m^*})^\delta}}{n^2} e^{-\frac{K_1\xi}{\sqrt{2}}\sqrt{n}}.$$

1) Case $0 \le \delta \le 1/3$. In that case, $\delta \le (1/2-\delta/2)_+$ and the choice $\xi^2 = (4B\pi^\delta\lambda_2)/(K_1\lambda_1)$ ensures the convergence of $\sum_{m'\in\mathcal{M}_n} I(m^*)$. Indeed, if we denote by $\psi = 2\gamma + (1/2-\delta/2)\wedge(1-\delta)$, $\omega = (1/2-\delta/2)_+$, $K' = K_1\lambda_1/\lambda_2$, then for $a,b \ge 1$, we infer that

$$\begin{aligned}(a\vee b)^\psi e^{2B\pi^\delta(a\vee b)^\delta} e^{-K'\xi^2(a\vee b)^\omega} &\le \left(a^\psi e^{2B\pi^\delta a^\delta} + b^\psi e^{2B\pi^\delta b^\delta}\right) e^{-(K'\xi^2/2)(a^\omega+b^\omega)} \\ (31) \qquad &\le a^\psi e^{2B\pi^\delta a^\delta} e^{-(K'\xi^2/2)a^\omega} e^{-(K'\xi^2/2)b^\omega} + b^\psi e^{2B\pi^\delta b^\delta} e^{-(K'\xi^2/2)b^\omega}.\end{aligned}$$

Since the function $a \mapsto a^\psi e^{2B\pi^\delta a^\delta} e^{-(K'\xi^2/2)a^\omega}$ is bounded on $\mathbb{R}^+$ by a constant, κ, depending on γ, δ and K' only, it follows that

$$\sum_{m'\in\mathcal{M}_n} I(m^*) \le C/n.$$

The same holds for $\sum_{m'\in\mathcal{M}_n} II(m^*)$ and consequently, for some constant C,

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_n(m')) \le C/n.$$

In that case we choose

$$\text{pen}(m) = 8(4+\xi^2)\lambda_1 D_m^{2\gamma+1-\delta} \exp\{2B(\pi D_m)^\delta\}/n.$$

2) Case $\delta > 1/3$.

In that case, $\delta > (1/2-\delta/2)_+$. Bearing in mind Inequality (31) we choose $\xi^2 = \xi^2(m,m')$ such that

$$2B\pi^\delta D_{m^*}^\delta - K'\xi^2 D_{m^*}^\omega = -2B\pi^\delta D_{m^*}^\delta$$

that is

$$\xi^2 = \xi^2(m,m') = (4B\pi^\delta\lambda_2)/(K_1\lambda_1) D_{m^*}^{\delta-\omega}.$$

These choices ensure that $\sum_{m'\in\mathcal{M}_n} I(m^*) \leq C/n$, that $\sum_{m'\in\mathcal{M}_n} II(m^*) \leq C/n$ and consequently $\sum_{m'\in\mathcal{M}_n} \mathbb{E}[W_n(m')] \leq C/n$. The result follows by taking

$$p(m,m') = 2(4+\xi^2(m,m'))\lambda_1 D_{m^*}^{2\gamma+1-\delta}\exp(2B(\pi D_{m^*})^\delta)/n,$$

and

$$\mathrm{pen}(m) = 8(4+\xi^2(m,m))\lambda_1 D_m^{2\gamma+1-\delta}\exp(2B(\pi D_m)^\delta)/n.$$

4.2. **Proof of Theorem 3.2 : the absolutely regular case.** We use Berbee's coupling Lemma extended to sequences (see Bryc's [5] construction), to build approximating variables for the Z_i's. More precisely, we build variables $Z_i^\star$ such that if $n = 2p_nq_n + r_n$, $0 \leq r_n < q_n$, and $\ell = 0, \cdots, p_n - 1$

$$A_\ell^\star = (Z^\star_{2\ell q_n+1}, ..., Z^\star_{(2\ell+1)q_n}),\ B_\ell^\star = (Z^\star_{(2\ell+1)q_n+1}, ..., Z^\star_{(2\ell+2)q_n}),$$

and analogous definition without stars, then

- $A_\ell^\star$ and A_ℓ have the same law,
- $\mathbb{P}(A_\ell \neq A_\ell^\star) \leq \beta_{q_n}$,
- $A_\ell^\star$ and $(A_0, A_1, ..., A_{\ell-1}, A_0^\star, A_1^\star, \cdots, A_{\ell-1}^\star)$ are independent.

The blocks $B_\ell^\star$ are built in the same way. Without loss of generality and for sake of simplicity we assume that $r_n = 0$.
Starting from (24), and denoting by $\nu_n^\star$ the empirical contrast computed on the $Z_i^\star$, we write

$$\begin{aligned}\|\tilde g - g_m\|^2 &\leq 2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)| + 2|\nu_n^\star(\tilde g - g_m)| \\ &\quad + \frac{1}{4}\|\tilde g - g_m\|^2 + 4\|g - g_m\|^2 + \mathrm{pen}(m) - \mathrm{pen}(\hat m) \\ &\leq \frac{1}{2}\|\tilde g - g_m\|^2 + 4\left(\sup_{t\in B_{m,\hat m}(0,1)} (\nu_n^\star(t))^2 - p(m,\hat m)\right) \\ &\quad + 4\|g - g_m\|^2 + 2\mathrm{pen}(m) + 2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)|, \end{aligned} \tag{32}$$

where $4p(m,m') \leq \mathrm{pen}(m) + \mathrm{pen}(m')$ as previously. Note that there exist some coefficients $\tilde b_{m\vee\hat m,j}$ such that $\tilde g - g_m = \sum_{j=1}^{m\vee\hat m} \tilde b_{m\vee\hat m,j}\varphi_{m\vee\hat m,j}$ and $\sum_{j=1}^{m\vee\hat m} b^2_{m\vee\hat m,j} = \|\tilde g - g_m\|^2$. Therefore,

$$\begin{aligned} 2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)| &= \frac{2}{n}\left|\sum_{i=1}^n [u^*_{\tilde g - g_m}(Z_i) - u^*_{\tilde g - g_m}(Z_i^\star)]\right| \\ &\leq \frac{4}{n}\|\tilde g - g_m\| \left\|\sum_{j\in\mathbb{Z}} (u^*_{\varphi_{m\vee\hat m,j}})^2\right\|_\infty^{1/2} \sum_{i=1}^n \mathbb{1}_{Z_i\neq Z_i^\star}. \end{aligned}$$

According to Bound (40) in Lemma 5.1 (in Appendix),

$$\left\|\sum_{j\in\mathbb{Z}} (u^*_{\varphi_{m\vee\hat m,j}})^2\right\|_\infty \leq D_{m_n}\Delta_1(m_n)/(2\pi),$$

which gives that

(33)
$$2|\nu_n(\tilde g - g_m) - \nu_n^\star(\tilde g - g_m)| \le \frac{1}{4}\|\tilde g - g_m\|^2 + \frac{16 D_{m_n}\Delta_1(m_n)}{2\pi}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{Z_i\neq Z_i^\star}\right)^2.$$

Let $W_n^\star(m')$ be defined as in (26) with $Z_i^\star$ replacing Z_i. By gathering (32) and (33) we get that

$$\begin{aligned}\|\tilde g - g_m\|^2 &\le \frac{3}{4}\|\tilde g - g_m\|^2 + 4\sum_{m'\in\mathcal{M}_n} W_n^\star(m') + 2\mathrm{pen}(m) + 4\|g - g_m\|^2 \\ &\quad + \frac{16 D_{m_n}\Delta_1(m_n)}{2\pi}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{Z_i\neq Z_i^\star}\right)^2.\end{aligned}$$

Consequently

$$\begin{aligned}\frac{1}{4}\|\tilde g - g_m\|^2 &\le 4\|g - g_m\|^2 + 2\mathrm{pen}(m) + 4\sum_{m'\in\mathcal{M}_n} W_n^\star(m') \\ &\quad + 8\frac{D_{m_n}\Delta_1(m_n)}{\pi}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{Z_i\neq Z_i^\star}\right)^2.\end{aligned}$$

Therefore by using that

$$\frac{1}{8}\|\tilde g - g\|^2 \le \frac{1}{4}\|\tilde g - g_m\|^2 + \frac{1}{4}\|g - g_m\|^2$$

and that $\mathbb{E}\left(n^{-1}\sum_{i=1}^n \mathbb{1}_{Z_i\neq Z_i^\star}\right)^2 \le \beta_{q_n}$, we infer that, for all $m\in\mathcal{M}_n$,

$$\mathbb{E}\|g - \tilde g\|^2 \le 34\|g - g_m\|^2 + 16\mathrm{pen}(m) + \frac{C_1 + 32C_2}{n},$$

provided that

$$\frac{64 D_{m_n}\Delta_1(m_n)}{\pi}\beta_{q_n} \le \frac{C_1}{n} \tag{34}$$

and

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}(W_n^\star(m')) \le \frac{C_2}{n}. \tag{35}$$

Using that $D_{m_n}\Delta_1(m_n) \le C' D_{m_n}^{2\gamma+1-\delta} e^{2B\pi^\delta D_{m_n}^\delta}$, we conclude that (34) is fulfilled as soon as

$$D_{m_n}^{2\gamma+1-\delta} e^{2B\pi^\delta D_{m_n}^\delta}\beta_{q_n} \le C_1'/n. \tag{36}$$

In order to ensure that our estimators converge, we only consider models such that $D_{m_n}^{2\gamma+1-\delta} e^{2B\pi^\delta D_{m_n}^\delta} \le n$ (bounded penalty), and therefore (36) requires that $\beta_{q_n} \le C_1'/n^2$. When the mixing is geometrical, this holds with no condition on θ by choosing $q_n = \ln(n)/(4\theta)$. When the mixing is arithmetical, we take $q_n = [n^c]$ with

$c \in]0, 1/2[$ and (36) becomes $n^{-c(1+\theta)} \leq \kappa n^{-2}$ which holds as soon as $\theta > 3$. We now come to the proof of (35), and study $W_n^\star(m')$ which must be split into two terms $\frac{1}{2}(W_{,1}^\star(m') + W_{n,2}^\star(m'))$ involving respectively the odd and even blocks and which are of the same type. More precisely $W_{n,1}^\star(m')$ and $W_{n,2}^\star(m')$ are given by

$$\begin{aligned} W_{n,1}^\star(m') &= \left[\left(\sup_{t\in B_{m,m'}(0,1)}\left|\frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\left(u_t^*(Z_{2\ell q_n+i}^\star) - \langle t, g\rangle\right)\right|\right)^2 - p_1(m,m')\right]_+ \\ W_{n,2}^\star(m') &= \left[\left(\sup_{t\in B_{m,m'}(0,1)}\left|\frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\left(u_t^*(Z_{(2\ell+1)q_n+i}^\star) - \langle t, g\rangle\right)\right|\right)^2 - p_1(m,m')\right]_+. \end{aligned}$$

We only study $W_{n,1}^\star(m')$ and conclude for $W_{n,2}^\star(m')$ by using analogous arguments. The study of $W_{n,1}^\star(m')$ consists in applying Inequality (29) to $\nu_{n,1}^\star(t)$ defined by

$$\begin{aligned} \nu_{n,1}^\star(t) &= \frac{1}{p_nq_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n}\left(u_t^*(Z_{2\ell q_n+i}^\star) - \langle t, g\rangle\right) \\ &= \frac{1}{p_n}\sum_{\ell=1}^{p_n}\nu_{q_n,\ell}^\star(t) \end{aligned} \tag{37}$$

considered as the sum of the p_n independent random variables $\nu_{q_n,\ell}^\star(t)$ defined as

$$\nu_{q_n,\ell}^\star(t) = (1/q_n)\sum_{j=1}^{q_n} u_t^*(Z_{2\ell q_n+j}^\star) - \langle t, g\rangle. \tag{38}$$

Denote by $M_1^\star$, $H^\star$ and $v^\star$ the quantities such that

$$\sup_{t\in B_{m,m'}(0,1)} \| \nu_{q_n,\ell}^\star(t) \|_\infty \leq M_1^\star,\ \mathbb{E}(\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}^\star(t)|) \leq H^\star,\ \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu_{q_n,\ell}^\star(t)) \leq v^\star.$$

Clearly, $M_1^\star = 2M_1$. By combining Lemma 5.3 in Appendix to Inequality (12), we propose to take $(H^\star)^2$ given by

$$(H^\star)^2 = (H^\star(m^*))^2 = \frac{4\lambda_1(\sum_k \beta_k)D_{m^*}^{2\gamma+1-\delta}e^{2B(\pi D_{m^*})^\delta}}{n},$$

where λ_1 is still given by (11). By using again Lemma 5.3 and arguing as in the independent case, we get that for $\delta > 1$

$$\Delta_2(m, bh) \leq (2\pi)^2\lambda_1^2 \| (bh)^* \|_\infty^2 D_m^{4\gamma+2-2\delta}\exp(4B\pi^\delta D_m^\delta)$$

and for $\delta \leq 1$, we use that

$$\Delta_2(m, bh) \leq \frac{2\pi\lambda_1(1+\pi^2)^\gamma}{A_0^2} \| (bh)^* \|^2 D_m^{4\gamma+1-\delta}\exp(4B\pi^\delta D_m^\delta).$$

Next, use that $\| (bh)^* \|_\infty \leq \| bh \|_1$ with $\| bh \|_1 = \sum_k \beta_k$ and that, by Parseval's Formula, $\|(bh)^*\|^2 = 2\pi\|bh\|^2 \leq 2\pi\|h\|_\infty \int b^2(x)h(x)dx$ to get, by applying (44), that

$$\begin{aligned} \| (bh)^* \|_\infty &\leq \sum_k \beta_k \\ \text{and } \|(bh)^*\|^2 &\leq 4\pi\|h\|_\infty \sum_{k\in\mathbb{N}} (1+k)\beta_k. \end{aligned}$$

Note that under **A4** and **A6**, $\|(bh)^*\|$ is bounded as soon as $\theta > 1$. Therefore we propose to take $v^\star$ as

$$v^\star = C_2^\star D_{m^*}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)} \exp\{2B(\pi D_{m^*})^\delta\}/q_n,$$

where

$$C_2^\star = C_2^\star(\gamma, A_0, B, \delta) = \begin{cases} 4\sqrt{2}\lambda_2^\star & \text{if } \delta \leq 1 \\ 4\lambda_2^\star & \text{if } \delta > 1. \end{cases}$$

with $\lambda_2^\star = \lambda_2^\star(\gamma, A_0, B, \delta)$ defined by (22). By taking $p_1(m,m') = 2(4+\xi^2)(H^\star)^2$, we get that

$$\mathbb{E}(W_{n,1}^\star(m')) \leq \mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} |\nu_{n,1}^\star(t)|^2 - 2(4+\xi^2)(H^\star)^2\right]_+ .$$

Apply (29) to the sum $\nu_{n,1}^\star = (p_n)^{-1}\sum_{\ell=1}^{p_n} \nu_{q_n,\ell}^\star(t)$, to obtain the global bound

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_{n,1}^\star(m')) \leq K \sum_{m'\in\mathcal{M}_n} [I^\star(m^*) + II^\star(m^*)],$$

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

$$\begin{aligned} I^\star(m^*) &= \frac{\lambda_2^\star D_{m^*}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)} \exp\{2B(\pi D_{m^*})^\delta\}}{n} e^{-2K_1\xi^2(\sum_k \beta_k)(\lambda_1/C_2^\star)D_{m^*}^{(1/2-\delta/2)_+}} \\ \text{and} \qquad II^\star(m^*) &= \frac{\lambda_1^2 q_n^2 D_{m^*}^{4\gamma+2-2\delta} e^{4B(\pi D_{m^*})^\delta}}{n^2} e^{-K_1\xi\sqrt{\sum_k \beta_k}/\sqrt{2}(\sqrt{n}/q_n)}. \end{aligned}$$

Since the term $I^\star(m^*)$ has the same order as the analogous term $I(m^*)$ in the independent case, it is bounded in the same manner. More precisely, when $0 \leq \delta \leq 1/3$, then $\delta \leq (1/2-\delta/2)_+$ and the choice $\xi^2 = (2B\pi^\delta C_2^\star)/[K_1\lambda_1(\sum_k \beta_k)]$ ensures the convergence of $\sum_{m'\in\mathcal{M}_n} I(m^*)$. When $\delta > 1/3$, $\delta > (1/2-\delta/2)_+$ and we choose $\xi^2 = \xi^2(m,m')$ such that

$$2B\pi^\delta D_{m^*}^\delta - K_1\xi^2 2\lambda_1(\sum_k \beta_k) D_{m^*}^\omega / C_2^\star = -2B\pi^\delta D_{m^*}^\delta$$

that is

$$\xi^2 = \xi^2(m,m') = (2B\pi^\delta C_2^\star)/[K_1(\sum_k \beta_k)\lambda_1] D_{m^*}^{\delta-\omega}.$$

These choices ensure that in both cases, $\sum_{m'\in\mathcal{M}_n} I(m^*) \leq C/n$

The main difference lies in the study of the second term $II^\star(m^*)$. When the mixing is geometrical then $q_n = \ln(n)/(4\theta)$ ensures that

$$\sum_{m'\in\mathcal{M}_n} \frac{q_n^2 D_{m^*}^{2\gamma+1-\delta} e^{2B(\pi D_{m^*})^\delta}}{n^2} e^{-K_1\xi \frac{\sqrt{n}\sqrt{(\sum_k \beta_k)}}{\sqrt{2}q_n}} \leq C/n. \tag{39}$$

When the mixing is arithmetical, then $q_n = [n^c]$, with c in $]0, 1/2[$ ensures that (39) holds and consequently

$$\sum_{m'\in\mathcal{M}_n} II^\star(m^*) \leq C/n.$$

Finally $\sum_{m'\in\mathcal{M}_n} \mathbb{E}[W_n^\star(m')] \leq 2\sum_{m'\in\mathcal{M}_n} \mathbb{E}[W_{n,1}^\star(m') + W_{n,2}^\star(m')] \leq C/n$. The result follows by choosing $p_1(m,m') = p_2(m,m') = 8\lambda_1(4+\xi^2(m,m'))(\sum_k \beta_k) D_{m^*}^{2\gamma+1-\delta} e^{2B(\pi D_{m^*})^\delta}/n$ with $p(m,m') = 2p_1(m,m') + 2p_2(m,m')$ and

$$\text{pen}(m) = 128\lambda_1(4+\xi^2(m,m))(\sum_k \beta_k) D_m^{2\gamma+1-\delta} e^{2B(\pi D_m)^\delta}/n.$$

5. Appendix

Lemma 5.1. *Let $\nu_n(t)$ be defined by (4), $\Delta_1(m)$ be defined by (9).*
• *If the random variables $Z_1, \cdots, Z_n$, are i.i.d. random variables, then we have*

$$\sup_{g\in\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)} \| \sum_{j\in\mathbb{Z}} |u^*_{\varphi_{m,j}}|^2 \|_\infty \leq D_m \Delta_1(m)/(2\pi), \tag{40}$$

and

$$\sup_{g\in\mathcal{S}_{\alpha,\nu,\rho}(A_\alpha)} \sum_{j\in\mathbb{Z}} \text{Var}[\nu_n(\varphi_{m,j})] \leq D_m\Delta_1(m)/(2\pi n). \tag{41}$$

• *If the random variables $Z_1, \cdots, Z_n$, are absolutely regular random variables, then we have*

$$\sum_{j\in\mathbb{Z}} Var[\nu_n(\varphi_{m,j})] \leq \frac{4(\sum_k \beta_k) D_m \Delta_1(m)}{2\pi n}. \tag{42}$$

Proof of Lemma 5.1. The proof follows the lines of the proof of Lemma 3 in Pensky and Vidakovic (1999). Use the definition of $u^*_{\varphi_{m,j}}(z)$ to get that

$$\begin{aligned}\sum_{j\in\mathbb{Z}} \left|u^*_{\varphi_{m,j}}(z)\right|^2 &= \sum_{j\in\mathbb{Z}} \left|\int e^{ixz} u_{\varphi_{m,j}}(x)dx\right|^2 \\ &= \frac{D_m}{(2\pi)^2} \sum_{j\in\mathbb{Z}} \left|\int e^{-ixD_m} e^{ijx} \frac{\varphi^*(x)}{f_\varepsilon^*(xD_m)} dx\right|^2.\end{aligned}$$

Note that $\int e^{ixD_m}e^{ijx}\varphi^*(x)/f_\varepsilon^*(xD_m)dx$ are the Fourier coefficients of the function $x \mapsto e^{ixD_m}\varphi^*(x)/f_\varepsilon^*(xD_m)$ and therefore by Parseval's Formula,

$$\sum_{j\in\mathbb{Z}} \left|u^*_{\varphi_{m,j}}(z)\right|^2 = (2\pi)^{-1}D_m \int \left|\frac{\varphi^*(x)}{f_\varepsilon^*(xD_m)}\right|^2 dx = \frac{1}{2\pi}D_m\Delta_1(m), \tag{43}$$

which entails that Bound (40) is proved.
• The i.i.d. case. Bound (41) follows by using that $\mathrm{Var}(Y) \le \mathbb{E}(Y^2)$, which implies that

$$\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})] \;\; \le \;\; n^{-1}\int \sum_{j\in\mathbb{Z}} \left|u^*_{\varphi_{m,j}}(z)\right|^2 h(z)dz.$$

• The dependent case. We use Delyon's (1990) covariance Inequality, successfully exploited by Viennet (1997) for partial sums of strictly stationary processes, that we start by recalling.

Theorem 5.1. *(Delyon (1990), Viennet (1997)) Let P be the distribution of Z_0 on a probability space $\mathcal{X}$, $\int f dP = \mathbb{E}_P(f)$ for any function f P-integrable. For $r \ge 2$, let $\mathcal{L}(r,\beta,P)$ be the set of functions $b : \mathcal{X} \to \mathbb{R}^+$ such that*

$$b = \sum_{l\ge 0}(l+1)^{r-2}b_l \text{ with } 0 \le b_l \le 1 \text{ and } \mathbb{E}_P(b_l) \le \beta_l,$$

We define B_r as $B_r = \sum_{l\ge 0}(l+1)^{r-2}\beta_l$. Then for $1 \le p < \infty$ and any function b in $\mathcal{L}(2,\beta,P)$,

$$\mathbb{E}_P(b^p) \le pB_{p+1}, \tag{44}$$

as soon as $B_{p+1} < \infty$. The following result holds for a strictly stationary absolutely regular sequence, $(Z_i)_{i\in\mathbb{Z}}$, with β-mixing coefficients $(\beta_k)_{k\ge 0}$: if $B_2 < +\infty$, there exists $b \in \mathcal{L}(2,\beta,\infty)$ such that for any positive integer n and any measurable function $f \in \mathbb{L}_2(P)$, we have

$$\mathrm{Var}\left(\sum_{i=1}^n f(Z_i)\right) \le 4n\mathbb{E}_P(bf^2) = 4n\int b(x)f^2(x)dP(x). \tag{45}$$

Apply (45), to infer that

$$\mathrm{Var}(\nu_n(\varphi_{m,j})) \le \frac{4}{n}\int b(z)(u^*_{\varphi_{m,j}})^2(z)h(z)dz,$$

which combined with (44) and Bound (40) entails Bound (42).

Lemma 5.2. *Let $\nu_n(t)$ be defined by (4), $\Delta_1(m)$ be defined by (9). If the random variables $Z_1, \cdots, Z_n$, are i.i.d. random variables, then we have*

$$\begin{aligned}
\sup_{t\in B_{m,m'}(0,1)} \| u_t^* \|_\infty &\leq \sqrt{D_{m^*}\Delta_1(m^*)/(2\pi)} \\
\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)} |\nu_n(t)|] &\leq \sqrt{D_{m^*}\Delta_1(m^*)/(2\pi n)}, \\
\text{and} \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(u_t^*(Z_1)) &\leq \sqrt{\Delta_2(m^*,h)}/(2\pi),
\end{aligned}$$

where $\Delta_2(m,h)$ is defined by (30).

Proof of Lemma 5.2. The proof follows the lines of Lemma 4 in Pensky and Vidakovic (1999). The first term $\sup_{t\in B_{m,m'}(0,1)} \| u_t^* \|_\infty$ is, by applying Cauchy-Schwarz Inequality and (43), bounded by

$$\sqrt{\sum_{j\in\mathbb{Z}} \int \left| \frac{\varphi_{m^*,j}^*}{f_\varepsilon^*}(u) \right|^2 du} = \sqrt{D_{m^*}\Delta_1(m^*)/(2\pi)}.$$

Now,

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

which is bounded, by applying Bound (41) in Lemma 5.1, by $\sqrt{D_{m^*}\Delta_1(m^*)/(2\pi n)}$. Now write that

$$\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(u_t^*(Z_1)) \leq \sup_{t\in B_{m,m'}(0,1)} \mathbb{E}[|u_t^*(Z_1)|^2] \leq \sqrt{\sum_{j,k\in\mathbb{Z}} |Q_{j,k}(m^*)|^2},$$

with $Q_{j,k}(m) = \mathbb{E}[u_{\varphi_{m,j}}^*(Z_1)u_{\varphi_{m,k}}^*(-Z_1)]$ also given by

$$Q_{j,k}(m) = \frac{D_m}{(2\pi)^2} \iint e^{ijx-iky} \frac{\varphi^*(x)\varphi^*(y)}{f_\varepsilon^*(D_m x) f_\varepsilon^*(D_m y)} h^*(D_m(x-y))dxdy.$$

Apply Parseval's Formula to get that

$$\sum_{j,k\in\mathbb{Z}} |Q_{j,k}(m)|^2 = \frac{D_m^2}{(2\pi)^2} \iint \left| \frac{\varphi^*(x)\varphi^*(y)}{f_\varepsilon^*(D_m x) f_\varepsilon^*(D_m y)} h^*(D_m(x-y)) \right|^2 dxdy,$$

which entails the result.

Lemma 5.3. *Let $\nu_{n,1}^\star(t)$ and $\nu_{q_n,\ell}^\star(t)$ be defined by (37) and (38), and $\Delta_1(m)$ be defined by (9). If the random variables $Z_1, \cdots, Z_n$, are β-mixing random variables,* *then we have*

$$\begin{aligned}
\sup_{t\in B_{m,m'}(0,1)} \parallel \nu^\star_{q_n,\ell}(t) \parallel_\infty &\leq \sqrt{D_{m^*}\Delta_1(m^*)/(2\pi)} \\
\mathbb{E}[\sup_{t\in B_{m,m'}(0,1)} |\nu^\star_{n,1}(t)|] &\leq \sqrt{\frac{2(\sum_k \beta_k) D^*_m \Delta_1(m^*)}{\pi n}}, \\
\text{and} \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(\nu^\star_{q_n,\ell}(t)) &\leq \frac{2\sqrt{\Delta_2(m^*, bh)}}{\pi q_n}.
\end{aligned}$$

Proof of Lemma 5.3. Arguing as in Lemma 5.2, the first term $\sup_{t\in B_{m,m'}(0,1)} \parallel \nu^\star_{q_n,\ell}(t) \parallel_\infty$ is, by using Cauchy-Schwarz Inequality and (43), bounded by

$$\sqrt{\sum_{j\in\mathbb{Z}} \int \left|\frac{\varphi^*_{m^*,j}}{f^*_\varepsilon}(u)\right|^2 du} = \sqrt{D_{m^*}\Delta_1(m^*)/(2\pi)}.$$

Next

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} \left|\nu^\star_{n,1}(t)\right|\right] &= \mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} \left|\frac{1}{p_n q_n}\sum_{\ell=1}^{p_n}\sum_{i=1}^{q_n} u^*_t(Z^\star_{2\ell q_n+i}) - \langle t, g\rangle\right|\right] \\
&\leq \sqrt{\sum_{j\in\mathbb{Z}} \mathrm{Var}(\nu^\star_{n,1}(\varphi^*_{m,j}))}
\end{aligned}$$

According to (45),

$$\mathbb{E}\left[\sup_{t\in B_{m,m'}(0,1)} \left|\nu^\star_{n,1}(t)\right|\right] \leq \sqrt{\frac{2(\sum_k \beta_k) D^*_m \Delta_1(m^*)}{\pi n}}.$$

Lastly, by applying again Inequality (45) we get that

$$\sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}\left(\frac{1}{q_n}\sum_{i=1}^{q_n} u^*_t(Z^\star_{2\ell q_n+i})\right) \leq \frac{4}{q_n}\sup_{t\in B_{m,m'}(0,1)} \int |u^*_t(z)|^2 b(z)h(z)dz.$$

Now write that

$$\frac{4}{q_n}\sup_{t\in B_{m,m'}(0,1)} \int |u^*_t(z)|^2 b(z)h(z)dz \leq \frac{4}{q_n}\sqrt{\sum_{j,k\in\mathbb{Z}} |\bar{Q}_{j,k}(m^*)|^2},$$

with $\bar{Q}_{j,k}(m) = \int u^*_{\varphi_{m,j}}(z) u^*_{\varphi_{m,k}}(-z) b(z)h(z)dz$ also given by

$$\bar{Q}_{j,k}(m) = \frac{D_m}{(2\pi)^2}\iint e^{ijx-iky}\frac{\varphi^*(x)\varphi^*(y)}{f^*_\varepsilon(D_m x) f^*_\varepsilon(D_m y)}(bh)^*(D_m(x-y))dxdy.$$

Apply Parseval's Formula to get that

$$\begin{aligned}\sum_{j,k\in\mathbb{Z}} |\bar{Q}_{j,k}(m)|^2 &= \frac{D_m^2}{(2\pi)^2}\iint \left|\frac{\varphi^*(x)\varphi^*(y)}{f_\varepsilon^*(D_m x)f_\varepsilon^*(D_m y)}(bh)^*(D_m(x-y))\right|^2 dxdy \\ &= \Delta_2(m,bh)/(2\pi)^2,\end{aligned}$$

and the result follows.

Lemma 5.4. *Consider the model described in section 2.1 under* **A1**-**A7** *and the collection of estimators* $\hat{g}_m^{(n)}$ *defined by (3) for* $1 \leq m_n \leq \log_2(n)$.
Then

$$\mathbb{E} \parallel g - \hat{g}_m^{(n)} \parallel^2 \leq \parallel g - g_m \parallel^2 + 2\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})] + (\|\varphi\|^2 A_g + 1)\frac{D_m^2}{K_n}.$$

Proof of Lemma 5.4 : The proof follows the lines of the proof of Theorem 3 in Pensky and Vidakovic (1999). Arguing as for Inequality (8), we obtain

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \leq \|g - \mathbb{E}(\hat{g}_m^{(n)})\|^2 + 2\sum_{|j|\leq K_n} \mathrm{Var}[\nu_n(\varphi_{m,j})]. \tag{46}$$

By applying the triangular inequality, we have

$$\mathbb{E} \parallel g - \hat{g}_m^{(n)} \parallel^2 \leq \|g - g_m\|^2 + \|g_m - \mathbb{E}(\hat{g}_m^{(n)})\|^2 + 2\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})].$$

Using that $\mathbb{E}(\hat{g}_m^{(n)})$ is the orthogonal projection of the density g on $S_m^{(n)}$, we get that

$$\mathbb{E}(\hat{g}_m^{(n)}) = \sum_{|j|\leq K_n} a_{m,j}\varphi_{m,j},$$

and consequently

$$\|g_m - \mathbb{E}(\hat{g}_m^{(n)})\|^2 = \sum_{|j|\geq K_n} a_{m,j}^2 \leq (\sup_j j a_{m,j})^2 \sum_{|j|\geq K_n} j^{-2}.$$

Now we write that

$$\begin{aligned} j a_{m,j} &= j\sqrt{D_m}\int \varphi(D_m x - j)g(x)dx \\ &\leq D_m^{3/2}\int |x||\varphi(D_m x - j)|g(x)dx + \sqrt{D_m}\int |D_m x - j||\varphi(D_m x - j)|g(x)dx \\ &\leq D_m^{3/2}\left(\int |\varphi(D_m x - j)|^2 dx\right)^{1/2}\left(\int x^2 g^2(x)dx\right)^{1/2} + \sqrt{D_m}\sup_x |x\varphi(x)|.\end{aligned}$$

This implies finally that

$$j a_{m,j} \leq D_m\|\varphi\|\left(\int x^2 g^2(x)dx\right)^{1/2} + \sqrt{D_m},$$

and the result follows.

References

[1] A.R. Barron, Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields* **113**, 301-413.

[2] Berbee, H.C.P (1979) *Random walks with stationary increments and renewal theory.* Cent. Math. Tracts, Amsterdam.

[3] Birgé, L. and Massart, P. (1997) From model selection to adaptive estimation. In: D. Pollard, E. Torgersen and G. Yang, eds. *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* (Springer-Verlag, New-York) pp. 55-87.

[4] Birgé, L. and Rozenholc, Y. (2002) How many bins must be put in a regular histogram. Preprint du LPMA 721, http://www.proba.jussieu.fr/mathdoc/preprints/index.html.

[5] Bryc, W. (1982) On the approximation theorem of Berkes and Philipp. *Demonstratio Mathematica* **15**, 807-815.

[6] Butucea, C. (2003) Deconvolution of supersmooth densities with smooth noise. Submitted for publication.

[7] Butucea, C. and Tsybakov, A.B. (2003) Fast asymptotics in density deconvolution. Working paper.

[8] Carroll, R.J. and Hall, P. (1988) Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184-1186.

[9] Cator, E. A. (2001) Deconvolution with arbitrarily smooth kernels. *Stat. Probab. Lett.* **54**, 205-214.

[10] Comte, F. and Merlevède, F. (2002) Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM P&S* **6**, 211-238.

[11] Comte, F. and Rozenholc, Y. (2001) Adaptive estimation of mean and volatility functions in (auto-)regressive models. *Stoch. Proc. Appl.* **97**, 111-145.

[12] Delyon, B. (1990) Limit theorem for mixing processes. *Tech. Report IRISA* Rennes **1**, 546.

[13] Devroye, L.(1989) Consistent deconvolution in density estimation. *Canad. J. Statist.* **17**, 235-239.

[14] Donoho, D.L. and Johnstone, I.M. (1998) Minimax estimation with wavelet shrinkage. *Ann. Statist.* **26**, 879-921.

[15] Donoho, D.L. Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996) Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508-539.

[16] Efromovich, S. (1997) Density estimation in the case of supersmooth measurement errors. *J. Amer. Statist. Assoc.* **92**, 526-535.

[17] Ermakov, M. S. (1999) Minimax estimation of the solution of an ill-posed convolution type problem. *Problems of Information Transmission* **25**, 191-200.

[18] Fan, J. (1991a) On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.* **19**, 1257-1272.

[19] Fan, J. (1991b) Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1**, 541-551.

[20] Fan, J. and Koo J.-Y. (2002) Wavelet deconvolution. *IEEE TransacT. on Information Theory* **48**, 734-747.

[21] Goldenshluger, A. (1999) On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5** 907-926.

[22] Liu, M.C. and Taylor, R.L. (1989) A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17**, 427-438.

[23] Masry, E. (1991) Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory* **37**, 1105-1115.

[24] Masry, E. (1993a) Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Process. Appl.* **47**, 53-74.

[25] Masry, E. (1993b) Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *J. Multivariate Anal.* **44**, 47-68.
[26] Meyer, Y. (1990), Ondelettes et opérateurs, Tome I, Hermann.
[27] Pensky, M. (2002) Density deconvolution based on wavelets with bounded supports. *Stat. Probab. Lett.* **56** 261-269.
[28] Pensky, M. and Vidakovic, B. (1999) Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 6, 2033-2053.
[29] Stefansky, L. (1990) Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Letters* **9**, 229-235.
[30] Stefansky, L. and Carroll, R.J. (1990) Deconvolution kernel density estimators. *Statistics* **21**, 169-184.
[31] Talagrand, M. (1996) New concentration inequalities in product spaces. *Invent. Math.* **126**, 505-563.
[32] Taylor, R.L. and Zhang, H.M. (1990) On strongly consistent non-parametric density estimator for deconvolution problem. *Comm. Statist. Theory Methods* **19**, 3325-3342.
[33] Viennet, G.(1997) Inequalities for absolutely regular sequences: application to density estimation. *Probab. Th. Rel. Fields* **107**, 467-492.
[34] Zhang, C.H. (1990) Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806-831.

F. Comte,
MAP5 (Statistique)
Université René Descartes-Paris 5,
45 rue des Saints-Pères,
75 270 PARIS cedex 06, France.
email: comte@biomedicale.univ-paris5.fr.

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