PENALIZED CONTRAST ESTIMATOR FOR DENSITY DECONVOLUTION

Abstract. We consider the problem of estimating the density g of independent and identically distributed variables X_i, from a sample $Z_1, \dots, Z_n$ where $Z_i = X_i + \sigma\varepsilon_i$, $i = 1, \dots, n$, ε is a noise independent of X, with $\sigma\varepsilon$ having known distribution. We present a model selection procedure allowing to construct an adaptive estimator of g and to find non-asymptotic bounds for its $\mathbb{L}_2(\mathbb{R})$-risk. In all cases where lower bounds are available in density deconvolution, our estimator is proved to reach automatically the optimal rates, except in one case where a negligible logarithmic loss occurs, due to the adaptation. Furthermore we show that these results still hold when the X_i's and the ε_i's are both absolutely regular random variables. A short simulation study gives an illustration of the good practical performances of the method.

MSC 2000 Subject Classifications. Primary 62G07. Secondary 62G20.

Résumé

Considérons le problème de déconvolution c'est-à-dire de l'estimation de la densité de variables aléatoires identiquement distribuées X_i, à partir de l'observation de Z_i où $Z_i = X_i + \sigma\varepsilon_i$, pour $i = 1, \dots, n$ et où les erreurs $\sigma\varepsilon_i$ 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 dans la plupart des 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. Une étude par simulation illustre les bonnes performances pratiques de la méthode.

Keywords and phrases. Adaptive estimation. Density deconvolution. Model selection. Penalized contrast. Projection Estimator. Absolutely regular sequence.

1. Introduction

1.1. **The problem.** In this paper, we consider the problem of the nonparametric estimation of the density g, of independent and identically distributed (i.i.d.) variables X_i, from a sample $Z_1, \dots, Z_n$ in the model

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

where the X_i's and ε_i's are independent sequences, the ε_i's are i.i.d. centered random variables with common density f_ε and noise level σ. Due to the independence between the X_i's and the ε_i's, the observations $Z_1, \cdots, Z_n$ have common density $h(z) = \sigma^{-1} g * f_\varepsilon(./\sigma)(z)$, where $*$ denotes the convolution product. The function $\sigma^{-1} f_\varepsilon(./\sigma)$ is often called the convolution kernel and is here completely **known**. We refer to Matias (2002) or Butucea and Matias (2003) for results about density deconvolution when σ is unknown in such a model.

The aim of our paper is to construct an adaptive and optimal density deconvolution estimator in model (1). More precisely we aim at estimating the density g on $\mathbb{L}_2(\mathbb{R})$ without any prior knowledge on it, that is without the knowledge of its smoothness parameters and even without the knowledge of what type is its smoothness. We want our estimator to be adaptive in a minimax sense, simultaneously on various classes of functions.

We focus here on the independent case and we briefly mention that our results extend when the X_i's and the ε_i's are both absolutely regular random variables.

1.2. **Previous known results.**

1.2.1. *Previous upper and lower bounds for density deconvolution.* The problem of density deconvolution has been widely studied, especially using kernel estimators. It is well known that since $g^*(.) = h^*(.)/f_\varepsilon^*(\sigma.)$, ($u^*$ denoting the Fourier transform of u), two factors determine the estimation accuracy in the standard density deconvolution problem : first the smoothness of the density to be estimated, usually described by

$$\int_{-\infty}^{+\infty} |g^*(x)|^2 (x^2+1)^s \exp\{2b|x|^r\} dx \leq C_1, \tag{2}$$

and second the smoothness of the error density which is described by the rate of decay of the Fourier transform of f_ε

$$\kappa_0 (x^2+1)^{-\gamma/2} \exp\{-\mu|x|^\delta\} \leq |f_\varepsilon^*(x)| \leq \kappa_0' (x^2+1)^{-\gamma/2} \exp\{-\mu|x|^\delta\}, \tag{3}$$

with polynomial decay for ordinary smooth error density and exponential decay for super smooth error density.

Most previous results concern cases where the density g to be estimated belongs to smoothness classes, such as Hölder or Sobolev classes ($r = 0$ in (2)) and ordinary or super smooth error density. One can cite among others 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) and Zhang (1990), Koo (1999), Cator (2001). Most of them propose kernel estimators which are studied from many points of view: pointwise and global asymptotic optimality, asymptotic normality, case of dependent ε_i's... One consequence of those results is that the smoother the error density, the slower the optimal rate of convergence, with logarithmic rates of convergence when g belongs to a Sobolev or a Hölder class in presence of super smooth error density ($\delta > 0$ in (3)).

Much faster rates of convergence can be obtained if the density g to be estimated is much smoother, that is if $r > 0$ in (2). To our knowledge, the first paper dealing with the case of super smooth g is the paper by Pensky and Vidakovic (1999) who propose wavelet estimators.

When $r > 0$ and the errors are ordinary smooth ($\delta = 0$), Butucea (2004) proposes a kernel type estimator which is optimal in sharp asymptotical minimax sense under the pointwise and the $\mathbb{L}_2$-risks. In the same context, Butucea and Tsybakov (2004) deal with super smooth errors, and propose a kernel type estimator which is sharp asymptotical minimax, for the pointwise and $\mathbb{L}_2(\mathbb{R})$-risks, when $0 < r < \delta$ and $s = 0$: in this case, the variance of the estimator turns out to be asymptotically negligible with respect to its squared bias. One consequence of Butucea and Tsybakov's (2004) paper is that Pensky and Vidakovic's (1999) estimators are not optimal in the minimax sense, when both r and δ are positive.

In the same context, Meister (2004) studies the effect of missspecifying the error density on the asymptotic behavior of the risk.

In the previously mentioned papers, except in Pensky and Vidakovic (1999) and marginally in Butucea and Tsybakov (2004) who also consider adaptive estimation, the smoothness parameters of the unknown density are supposed to be known, and thus those papers deal with non adaptive estimation.

1.2.2. *Previous known results on adaptive deconvolution.* Let us now give more details on results about adaptive estimation. We do not develop the results concerning slightly different models such as circular deconvolution (Efromovich (1997)) or inverse problems in Gaussian white noise model (Goldenshluger (1999), Johnstone (1999), Cavalier and Tsybakov (2002)). Those results usually deal with estimation on compact sets and not on $\mathbb{R}$.

We focus on the description of results on adaptive density deconvolution on $\mathbb{R}$, that is in the same context as ours.

Following the ideas associated with wavelet methods (see Donoho and Johnstone (1995), Donoho et al. (1996)), Pensky and Vidakovic (1999) study a wavelet thresholding method to build an adaptive density deconvolution estimator, in the sense that its construction does not depend on unknown smoothness parameters. 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 with wavelets having bounded supports in order to better perform the estimation of very irregular functions g.

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 $p < 2$ and 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.

Using an automatic empirical bandwidth selection procedure, Hesse (1999) propose a data-driven deconvolution estimator, when the errors are ordinary smooth ($\gamma = 2$, and $\delta = 0$ in (3)). The resulting estimator is asymptotically optimal for the integrated squared risk.

More recently, in the specific case $0 < r < \delta/2$, and $s = 0$ in (2) and (3), Butucea and Tsybakov (2004) propose a sharp adaptive kernel type estimator of g for pointwise and $\mathbb{L}_2(\mathbb{R})$-risks. Nevertheless, it requires to know that $0 < r < \delta/2$.

1.3. **Estimator and new results.** Our estimator is constructed by model selection, and more precisely it is a penalized contrast estimator (see Birgé and Massart (1997), Barron et al. (1999)). We show that this penalized contrast estimator $\tilde{g}$ is adaptive and optimal or nearly optimal. More precisely, we establish a non-asymptotic bound for its integrated quadratic risk that ensures an automatic trade-off between a bias term and a penalty term, only depending on the observations, which shows that the estimator $\tilde{g}$ is adaptive in a minimax sense for the $\mathbb{L}_2(\mathbb{R})$-risk simultaneously on Sobolev classes ($r = 0$), on classes of analytical densities ($r = 1$), on some classes of super smooth densities (if $s = 0$ and $r > 0$ and probably if $r > 0$, $s \neq 0$) and on classes of entire functions (having Fourier transform compactly supported), when the errors are either ordinary smooth or super smooth, provided that $\delta \leq 1/3$ or that $0 < r < \delta$.

When $r \geq \delta > 0$ and $\delta \leq 1/3$, then $\tilde{g}$ automatically adjusts and achieves the best rate obtained by the collection of non penalized estimators. If $r \geq \delta > 1/3$, then $\tilde{g}$ automatically adjusts and achieves the best rate obtained by the collection of minimum contrast estimators, within a logarithmic factor. Nevertheless, it always significantly improves the rates given by the adaptive estimator built in Pensky and Vidakovic (1999), when both the density and the errors are super smooth.

The paper is organized as follows. In Section 2, we describe the problem, the assumptions, the construction of the minimum contrast estimator $\hat{g}_m^{(n)}$ and of the penalized minimum contrast estimator $\tilde{g}$. In Section 3 we give upper bounds for the $\mathbb{L}_2(\mathbb{R})$-risk of the minimum contrast estimator $\hat{g}_m^{(n)}$, when the smoothness of g is known, and study the optimality of the resulting rates. In Section 4, we give upper bounds of the $\mathbb{L}_2(\mathbb{R})$-risk of the penalized minimum contrast estimator $\tilde{g}$ when no prior knowledge on g is used, starting with the independent framework and then giving an extension to some β-mixing context. We provide in Section 5 a simulation study that illustrates the good practical results that can be obtained with this method and that compare with some other simulation results described in Delaigle and Gijbels (2004). All the proofs are gathered in Section 6.

2. Construction of the estimators

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, $u*v(x) = \int u(t)v(x-t)dt$. Moreover we denote by $\| u \| = \left(\int |u|^2(x)dx\right)^{1/2}$, and by $\langle s,t\rangle = \int s(x)\overline{t(x)}dx$.

2.1. **Model and Assumptions.** Consider Model (1) under the following assumptions.

($\mathbf{A}_1^X$) The X_i's and the ε_i's are identically distributed random variables.

($\mathbf{A}_2^{X,\varepsilon}$) The sequences $(X_i)_{i\in\mathbb{N}}$ and $(\varepsilon_i)_{i\in\mathbb{N}}$ are independent from each other.

($\mathbf{A}_3^{X,\varepsilon}$) The sequences $(\varepsilon_i)_{i\in\mathbb{N}}$ and $(X_i)_{i\in\mathbb{N}}$ are both sequences of independent random variables .

($\mathbf{A}_4^{\varepsilon}$) The density f_ε belongs to $\mathbb{L}_2(\mathbb{R})$ and is such that for all $x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$.

Under ($\mathbf{A}_1^X$) and ($\mathbf{A}_3^{X,\varepsilon}$), 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.

Under Assumption ($\mathbf{A}_2^{X,\varepsilon}$) the (unknown density) h of the Z_i's equals $\sigma^{-1}g * f_\varepsilon(./\sigma)$, and $h^* = g^*(.)f_\varepsilon^*(\sigma.)$, that is $g^*(.) = h^*(.)/f_\varepsilon^*(\sigma.)$. 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 $f_\varepsilon^*(x)$ as x goes to infinity. More precisely, the smoother f_ε, the slower is the rate of convergence for estimating g. Indeed, if f_ε is very smooth then so is h, the density of the observations and thus it is difficult to recover g. Thie smoothness of f_ε is described by the following assumption.

($\mathbf{A}_5^{\varepsilon}$) There exist nonnegative real numbers $\gamma,\ \mu,$ and δ such that

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

Note that only the left-hand side of ($\mathbf{A}_5^\varepsilon$) is required for upper bounds whereas the right-hand side is useful when we consider lower bounds and optimality problems of our estimators.

Assumptions ($\mathbf{A}_4^\varepsilon$) and ($\mathbf{A}_5^\varepsilon$) are usual for the construction of an estimator in density deconvolution. In particular Assumption ($\mathbf{A}_4^\varepsilon$) ensures that g is identifiable. Let us now comment Assumption ($\mathbf{A}_5^\varepsilon$). When $\delta = 0$ in ($\mathbf{A}_5^\varepsilon$), it amounts to consider what is usually called "ordinary smooth" errors, and when $\mu > 0$ and $\delta > 0$, the error density is usually called "super smooth". Indeed densities satisfying ($\mathbf{A}_5^\varepsilon$) 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. 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, when $\delta = 0$, ($\mathbf{A}_4^\varepsilon$) and ($\mathbf{A}_5^\varepsilon$) require that $\gamma > 1/2$.

By convention, we set $\mu = 0$ when $\delta = 0$ and we assume that $\mu > 0$ when $\delta > 0$. In the same way, if $\sigma = 0$, the X_i's are directly observed without noise and we set $\mu = \gamma = \delta = 0$ in this case.

Although, slower rates of convergence for estimating g are obtained for smoother error density, those rates can be improved by some additional regularity conditions on g. These regularity conditions are described as follows.

($\mathbf{R}_1^X$) There exists some positive real numbers s, r, b such that the density

$$g \text{ belongs to } \mathcal{S}_{s,r,b}(C_1) = \left\{ f \text{ density } : \int_{-\infty}^{+\infty} |f^*(x)|^2 (x^2+1)^s \exp\{2b|x|^r\} dx \leq C_1 \right\}.$$

($\mathbf{R}_2^X$) There exists some positive real numbers C_2 and d such that the density

$$g \text{ belongs to } \mathcal{S}_d(C_2) = \left\{ f \text{ density such that for all } x \in \mathbb{R},\ |f^*(x)| \leq C_2 \mathbb{I}_{[-d,d]}(x) \right\}.$$

The smoothness classes described by ($\mathbf{R}_1^X$) are classically considered both in deconvolution and in "direct" density estimation, since they can be roughly viewed as extensions of Sobolev classes. Note that densities satisfying ($\mathbf{R}_1^X$) with $r > 0, b > 0$ are infinitely many times differentiable. Moreover, such densities admit analytic continuation on a finite width strip when $r = 1$ and on the whole complex plane if $r = 2$. The densities satisfying ($\mathbf{R}_2^X$), often called entire functions, admit analytic continuation in the whole complex plane (see Ibragimov and Hasminskii (1983)).

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

($\mathbf{A}_6^X$) The density g belongs to $\mathbb{L}_2(\mathbb{R})$ and there exists some positive real M_2

$$\text{such that } g \text{ belongs to } \left\{ f \text{ density such that } \int x^2 f^2(x) dx \leq M_2 < \infty \right\}.$$

Assumption ($\mathbf{A}_6^X$) which is due to the construction of the estimator, is quite unusual in density estimation. Nevertheless it already appears in density deconvolution in a slightly different way in Pensky and Vidakovic (1999) who assume, instead of ($\mathbf{A}_6^X$) that $\sup_{x\in\mathbb{R}} |x|g(x) < \infty$. It is important to note that Assumption ($\mathbf{A}_6^X$) is very unrestrictive and can be refined. The main drawback of this condition is that it is not stable by translation, but most practical problems may be avoided by empirical centering of the data. Some improvements of Assumption ($\mathbf{A}_6^X$) may be searched but are omitted for the sake of simplicity. All densities having tails of order

$|x|^{-(m+1)}$ as x tends to infinity satisfy $(\mathbf{A}_6^X)$ only if $m > 1/2$. One can cite for instance the Cauchy distribution or all stable distributions with exponent $r > 1/2$ (see Devroye (1986)). In particular, the Levy distribution, with exponent $r = 1/2$ does not satisfies $(\mathbf{A}_6^X)$.

2.2. **The projection spaces and the estimators.** Consider $\varphi(x) = \sin(\pi x)/(\pi x)$, and let $\varphi_{m,j}(x) = \sqrt{L_m}\varphi(L_m x - j)$, where $L_m = m$ and $m \in \mathcal{M}_n = \{1, \cdots, m_n\}$. When $L_m = 2^m$, the basis $(\varphi_{m,j})$ is known as the Shannon basis. 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 $[-\pi L_m, \pi L_m]$. We denote by S_m such a space and by $(S_m)_{m\in\mathcal{M}_n}$, with $\mathcal{M}_n = \{1, \ldots, m_n\}$, this collection of linear spaces that is $S_m = \text{Vect}\{\varphi_{_{m,j}},\ j \in \mathbb{Z}\} = \{f \in \mathbb{L}_2(\mathbb{R}),\ \text{with supp}(f^*) \text{ included into } [-L_m\pi, L_m\pi]\}$. Denoting by g_m the orthogonal projection of g on S_m, g_m is given by

$$g_m = \sum_{j\in\mathbb{Z}} a_{m,j}\varphi_{m,j} \qquad \text{with} \qquad a_{m,j} =< g, \varphi_{m,j} > .$$

Since the projection g_m of g on S_m, involves infinite sums, we may prefer to consider the truncated spaces $S_m^{(n)}$ defined as

$$S_m^{(n)} = \text{Vect}\left\{\varphi_{m,j}, |j| \leq K_n\right\} \text{ where } K_n \text{ is an integer.}$$

It is easy to see that, $\{\varphi_{m,j}\}_{|j|\leq K_n}$ is an orthonormal basis of $S_m^{(n)}$ and the orthogonal projection $g_m^{(n)}$ of g on $S_m^{(n)}$ is given by $g_m^{(n)} = \sum_{|j|\leq K_n} a_{m,j}\varphi_{m,j}$ with $a_{m,j} =< g, \varphi_{m,j} >$.

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

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

By using Parseval and inverse Fourier formulas we get that

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

It follows that $\mathbb{E}(\gamma_n(t)) = \|t - g\|^2 - \|g\|^2$, which shows that $\gamma_n(t)$ suits well for the estimation of g. This quantity u_t^* also appears in a slightly different way in kernel deconvolution. The problem of its practical calculation is usual and can be solved by using algorithms like Fast Fourier Transform.

2.3. **Construction of the minimum contrast estimators.** Associated to the collection of models, the collection of non penalized estimators $\hat{g}_m^{(n)}$ of g is defined by

$$\hat{g}_m^{(n)} = \arg\min_{t\in S_m^{(n)}} \gamma_n(t). \tag{4}$$

By using that, $t \mapsto u_t$ is linear, and that $(\varphi_{m,j})_{|j|\leq K_n}$ is an orthonormal basis of $S_m^{(n)}$, we have $\hat{g}_m^{(n)} = \sum_{|j|\leq K_n} \hat{a}_{m,j}\varphi_{m,j}$ where $\hat{a}_{m,j} = n^{-1}\sum_{i=1}^{n} u^*_{\varphi_{m,j}}(Z_i)$ and $\mathbb{E}(\hat{a}_{m,j}) =< g, \varphi_{m,j} >= a_{m,j}$.

2.4. **Construction of the penalized contrast estimator.** 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

$$\tilde{g} = \hat{g}_{\hat{m}}^{(n)} \text{ with } \hat{m} = \arg\min_{m\in\mathcal{M}_n} \left[\gamma_n(\hat{g}_m^{(n)}) + \operatorname{pen}(m)\right], \tag{5}$$

where the penalty function pen must be chosen by using only the observations and the knowledge of $f_\varepsilon^*(\sigma.)$.

3. Rates of convergence of the minimum contrast estimators $\hat{g}_m^{(n)}$

In order to motivate our approach let us first give the rate of convergence of one estimator $\hat{g}_m^{(n)}$, when the smoothness of g is known.

Proposition 3.1. *Under Assumptions* ($\mathbf{A}_1^X$)-($\mathbf{A}_4^\varepsilon$) *and* ($\mathbf{A}_6^X$)*, denote by* $\Delta_1(m)$ *the quantity*

$$\Delta_1(m) = L_m \int_{-\pi}^{\pi} |f_\varepsilon^*(L_m x\sigma)|^{-2}\, dx/(2\pi). \tag{6}$$

Then

$$\mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 \le \|g - g_m\|^2 + L_m^2(M_2+1)/K_n + 2\Delta_1(m)/n. \tag{7}$$

First the variance term $\Delta_1(m)/n$, defined in (6) depends on the rate of decay of the Fourier transform of f_ε, with larger variance for smoother f_ε. By applying Lemma 6.3 (See Section 6.4), under ($\mathbf{A}_5^\varepsilon$), we get the bound

$$\Delta_1(m) \le 2\lambda_1\Gamma(m) \quad \text{where} \quad \Gamma(m) = L_m^{(2\gamma+1-\delta)} \exp\left\{2\mu\sigma^\delta\pi^\delta L_m^\delta\right\}, \tag{8}$$

where $\lambda_1 = \lambda_1(\gamma, \kappa_0, \mu, \sigma, \delta)$ is given by

$$\lambda_1(\gamma, \kappa_0, \mu, \sigma, \delta) = \frac{(\sigma^2\pi^2+1)^\gamma}{\pi^\delta\kappa_0^2 R(\mu,\delta,\sigma)} \quad \text{with} \quad R(\mu,\delta,\sigma) = \begin{cases} 1 & \text{if } \delta = 0 \\ 2\mu\delta\sigma^\delta & \text{if } 0 < \delta \le 1 \\ 2\mu\sigma^\delta & \text{if } \delta > 1. \end{cases} \tag{9}$$

According to (8) we only consider $L_m = m \le m_n$ such that $\Gamma(m_n)/n$ is bounded. Consequently, $\mathcal{M}_n = \{1, \cdots, m_n\}$ with

$$m_n \le \begin{cases} \pi^{-1} n^{1/(2\gamma+1)} & \text{if } \delta = 0 \\ \pi^{-1}\left[\dfrac{\ln(n)}{2\mu\sigma^\delta} + \dfrac{2\gamma+1-\delta}{2\delta\mu\sigma^\delta}\ln\left(\dfrac{\ln(n)}{2\mu\sigma^\delta}\right)\right]^{1/\delta} & \text{if } \delta > 0. \end{cases} \tag{10}$$

Second, if $K_n \ge n$, then we have

$$\begin{aligned} \mathbb{E}\|g - \hat{g}_m^{(n)}\|^2 &\le \|g - g_m\|^2 + 2\lambda_1\Gamma(m)/n + L_m^2(M_2+1)/n \\ &\le \|g - g_m\|^2 + (2\lambda_1 + M_2)\Gamma(m)/n. \end{aligned} \tag{11}$$

Remark 3.1. We point out that the $\{\varphi_{m,j}\}$ are $\mathbb{R}$-supported (and not compactly supported) so that we obtain an estimation on the whole line and not only on a compact set as usual for projection estimators. This is a great advantage of this basis. Nevertheless it induces the residual term $L_m^2(M_2+1)/K_n$, due to the truncation $|j| \le K_n$. But the most important thing is that the choice of K_n does not influence the other terms. Consequently, it is easy to check that we can find a relevant choice of K_n ($K_n \geq n$ under ($\mathbf{A}_6^X$)), that makes this last supplementary term unconditionally negligible with respect to the others. The choice of large K_n does not change the efficiency of our estimator from a statistical point of view but only changes some practical computations.

Finally, the bias term $\|g - g_m\|^2$ depends on the smoothness of the function g. It has the expected order for classical smoothness classes since it is given by the distance between g and the classes of entire functions having Fourier transform compactly supported on $[-\pi L_m, \pi L_m]$ (see Ibragimov and Hasminskii (1983)). Indeed, by using 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 $\mathrm{supp}(f^*) \subset [-L_m\pi, L_m\pi]$, we get that $g_m^* = g^* \mathbb{I}_{[-L_m\pi, L_m\pi]}$ and therefore $\|g - g_m\|^2 = (2\pi)^{-1}\|g^* - g_m^*\|^2 = (2\pi)^{-1}\int_{|x|\geq \pi L_m} |g^*|^2(x)dx$.

Let us precise the order of this risk when f_ε satisfies Assumption ($\mathbf{A}_5^\varepsilon$) and g satisfies Assumption ($\mathbf{R}_1^X$) or ($\mathbf{R}_2^X$).

3.1. **Order of the risk of $\hat{g}_m^{(n)}$ under ($\mathbf{R}_2^X$).** Consider that g satisfies ($\mathbf{R}_2^X$). Therefore by choosing $\pi L_m = d$, and $K_n \geq n$, the bias term $\| g - g_m \|^2 = 0$, (11) becomes

$$\mathbb{E}(\|g - \hat{g}_m^{(n)}\|^2) \leq 2\lambda_1 d^{(2\gamma+1-\delta)} \exp\left\{2\mu\sigma^\delta\pi^\delta d^\delta\right\}/n + d^2(M_2+1)/(\pi^2 n), \tag{12}$$

and the parametric rate of convergence for estimating g is achieved. We refer to Ibragimov and Hasminskii (1983) for similar result for the "direct" estimation of a density g using the observations $X_1, \cdots, X_n$ of common density g satisfying Assumption ($\mathbf{R}_2^X$).

3.2. **Order of the risk of $\hat{g}_m^{(n)}$ under ($\mathbf{R}_1^X$).** Consider now that g satisfies ($\mathbf{R}_1^X$). This implies immediately that the squared bias term is less than

$$\|g - g_m\|^2 \quad \leq \quad [C_1/(2\pi)](L_m^2\pi^2 + 1)^{-s}\exp\{-2b\pi^r L_m^r\}. \tag{13}$$

Consequently, under ($\mathbf{A}_6^X$) with $K_n \geq n$, and according to (11), the rate of convergence of $\hat{g}_m^{(n)}$ is obtained by selecting the space $S_m^{(n)}$ that minimizes

$$C_1(2\pi)^{-1}(L_m^2\pi^2+1)^{-s}\exp\{-2b\pi^r L_m^r\} + 2\lambda_1 L_m^{(2\gamma+1-\delta)}\exp\left\{2\mu\sigma^\delta\pi^\delta L_m^\delta\right\}/n + L_m^2(M_2+1)/n.$$

Those optimal choices of L_m with the corresponding rates of convergence are given in Table 1, for different types of smoothness of the unknown density g and different types of known density errors f_ε.

3.3. **About the optimality of the rates.**

3.3.1. *The standard cases.*

- The case $\delta = 0, r = 0$.

If g belongs to $\mathcal{S}_{s,r,b}(A)$ with $r = 0$ and f_ε satisfies Assumption ($\mathbf{A}_5^\varepsilon$) with $\delta = 0$, then $\Delta_1(m)$ has the order $L_m^{1+2\gamma}/n$. On the other hand, $\|g - g_m\|^2$ has the order L_m^{-2s}, which leads to choose the space $S_{\breve{m}}$ with $L_{\breve{m}}$ of order $n^{1/(2s+2\gamma+1)}$. In this case the bias and the variance term have the same order, and the estimator $g_{\breve{m}}$ reaches the rate $n^{-2s/(2s+2\gamma+1)}$, which is known to be the optimal rate (see Fan (1991a)). Nevertheless, the $L_{\breve{m}}$ which realizes the best compromise, depends on s the smoothness parameter of the unknown density g.

- The case $\delta > 0, r = 0$.

In this case, the rate of convergence, given by the order of the bias, is very slow, of order

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

TABLE 1. Optimal choice of the length ($L_{\breve{m}}$) and resulting (optimal) rates under Assumptions ($\mathbf{A}_5^\varepsilon$) and ($\mathbf{R}_1^X$).

$(\ln(n))^{-2s/\delta}$, known to be the minimax rate of convergence (see Fan (1991a)). In this case, the optimal $L_{\breve{m}}$ does not depend on the density g. Therefore adaptation with optimal rate of convergence is simple since it can be achieved by a direct tuning of the smoothing parameters πL_m and $\hat{g}_m^{(n)}$ is thus adaptive. It is important to note that this remark is valid provided that we know that r, which is related to the unknown density g, equals 0.

- The case $\delta = 0, r > 0$.

In this case, the rate of order $\ln(n)^{(2\gamma+1)/r}/n$ is given by the variance term, that is mainly explained by behavior of the noise. This case has been intensively studied by Butucea (2004) who gives the optimal rate with exact constant for the pointwise and the $\mathbb{L}_2$-risks, by using deconvolution kernel estimator with the kernel $\sin(x)/x$. It follows that $\hat{g}_m^{(n)}$ achieves the rate proved to be optimal by Butucea (2004). In this case the $\pi L_{\breve{m}}$ that realizes the best compromise clearly depends on the smoothness parameters of g.

3.3.2. *New results in a non standard case.* The case $\delta > 0, r > 0$ requires a specific discussion. To our knowledge, the first paper dealing with such a case is the paper by Pensky and Vidakovic (1999) who propose estimators based on wavelets and study the problem of adaptive estimation. Their estimators achieve optimal rates of convergence in the three previous cases. But when $r > 0, \delta > 0$, the rate of convergence of their estimator is not optimal as it is shown in Butucea and Tsybakov (2004), who provide sharp minimax results in this case. Since $\hat{g}_m^{(n)}$ has the same bias and the same variance as the kernel estimator of Butucea and Tsybakov (2004), we conclude that the rate of convergence of $\hat{g}_m^{(n)}$ is also the minimax rate of convergence in the case $0 < r < \delta$ and $s = 0$.

An important remark is that when $r > 0$ and $\delta > 0$ the optimal parameter $L_{\breve{m}}$ which has not an explicit form for general $r > 0$ and $\delta > 0$, is the solution of the following equation

$$nO(1) = L_{\breve{m}}^{2s+2\gamma+1-r}\exp\{2\mu\sigma^\delta\pi^\delta L_{\breve{m}}^\delta + 2b\pi^r L_{\breve{m}}^r\}. \tag{14}$$

This non explicit form of the optimal smoothing parameter appears in Butucea and Tsybakov (2004) when $0 < r < \delta$ and $s = 0$ who show that if we denote by $\pi L_{\breve{m}}$ the solution of $2\mu\sigma^\delta(\pi L_{\breve{m}})^\delta + 2b(\pi L_{\breve{m}})^r = \ln n - (\ln\ln n)^2$, then the minimax rate for the $\mathbb{L}_2(\mathbb{R})$ is of order $\exp\{-2b(\pi L_{\breve{m}})^r\}$. Nevertheless, the order of this minimax rate can be precised by using some additional information on the ratio r/δ. For instance, if $r < \delta$, we have to distinguish if $r/\delta \leq 1/2$ or $1/2 < r/\delta \leq 2/3, \ldots$. More precisely, if $r/\delta \leq 1/2$, the optimal choice $L_{\breve{m}}$ is

$$\pi L_{\breve{m}} = \left[\frac{\ln(n)}{2\mu\sigma^\delta} - \frac{2b}{2\mu\sigma^\delta}\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)^{r/\delta} - c\ln\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)\right]^{1/\delta} \text{ with } c = \frac{2\gamma - r + 2s + 1}{2\mu\sigma^\delta\delta}$$

and the rate is

$$\ln(n)^{-2s/\delta}\exp\left[-2b\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)^{r/\delta}\right].$$

If $1/2 < r/\delta \leq 2/3$ the optimal choice of $\pi L_{\breve{m}}$ is

$$\pi L_{\breve{m}} = \left[\frac{\ln(n)}{2\mu\sigma^\delta} - \frac{2b}{2\mu\sigma^\delta}\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)^{r/\delta} + \frac{r}{\delta}\frac{(2b)^2}{2\mu\sigma^\delta}\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)^{2r/\delta - 1} - c\ln\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)\right]^{1/\delta}$$

with the same c as above, which gives the rate

$$\ln(n)^{-2s/\delta}\exp\left[-2b\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)^{r/\delta} + \frac{(2b)^2}{2\mu\sigma^\delta}\frac{r}{\delta}\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)^{2r/\delta-1}\right].$$

If $2/3 < r/\delta \leq 3/4$, we have another choice of $\pi L_{\breve{m}}$ with another rate. It follows that the rate depends on the integer k such that r/δ belongs to the interval $I_k =]k/(k+1); (k+1)/(k+2)]$.

Consider finally the specific case $r = \delta$ which leads to the explicit solution

$$\pi L_{\breve{m}} = \left\{[\ln(n/\ln(n)^a)/(2\mu\sigma^\delta + 2b)]\right\}^{1/r} \text{ with } a = \frac{2s + 2\gamma - r + 1}{r} \tag{15}$$

and to the rate $[\ln(n)]^b n^{-b/(b+\mu\sigma^\delta)}$ with $b = (-2s\mu\sigma^\delta + (2\gamma - r + 1)b)/(r(\mu\sigma^\delta + b))$. The case $r = \delta = 1$ has also been studied by Tsybakov (2000) and Cavalier et al. (2003), in the case of inverse problems with random noise. In this case and in both problems (density deconvolution and inverse problem) the best compromise is explicit and so is the rate of convergence, of order $n^{-b/(b+\mu\sigma)}[\ln n]^{(-2s\mu\sigma+2b\gamma)/(\mu\sigma+b)}$. It is noteworthy that $\hat{g}_m^{(n)}$ seems also to achieve the optimal rate of convergence in this case.

As a conclusion, it is important to note that in this case, the solution of the best compromise between the squared bias and the variance depends on the ration r/δ, with r related to the strongly unknown density g.

All those facts give some strong motivation to consider adaptive estimation since the best compromise depends on g or at least on the knowledge of the behavior of g with respect to f_ε. We aim at finding some procedure that provides an estimator, that does not require prior information on g, and whose risk automatically achieves the optimal rate.

4. Adaptive estimation

We would like to find the penalty function pen, based on the observations, such that, for $K_n \geq n$

$$\mathbb{E} \parallel \tilde{g} - g \parallel^2 \leq \inf_{m\in\mathcal{M}_n} \left[\parallel g - g_m \parallel^2 + L_m^2(M_2+1)/n + 2\lambda_1\Gamma(m)/n\right]. \tag{16}$$

4.1. **Main results : the independent case.** First we give the result concerning the field on which the oracle Inequality (16) is reached, up to some multiplicative constants.

Theorem 4.1. *Under Assumptions* $(\mathbf{A}_1^X)$*-*$(\mathbf{A}_6^X)$*, consider the collection of estimators* $\hat{g}_m^{(n)}$ *defined by (4) with* $1 \leq m \leq m_n$ *satisfying (10) and* $K_n \geq n$*. Let* $\Gamma(m) = L_m^{2\gamma+1-\delta}\exp\{2\mu\sigma^\delta(\pi L_m)^\delta\}$ *and* $\lambda_1 = \lambda_1(\gamma, \kappa_0, \mu, \sigma, \delta)$ *be defined by (9) and* $\lambda_2 = \lambda_2(\gamma, \kappa_0, \mu, \sigma, \delta)$ *be defined by*

$$\lambda_2 = \begin{cases} \lambda_1 & \text{if } \delta > 1, \\ \lambda_1^{1/2}(1+\sigma^2\pi^2)^{\gamma/2}\|f_\varepsilon\|\kappa_0^{-1}(2\pi)^{-1/2} & \text{if } \delta \leq 1. \end{cases} \tag{17}$$

1) *If* $\delta = 0$ *or* $0 < \delta < 1/3$*, let* $\tilde{g} = \hat{g}_{\hat{m}}^{(n)}$ *be defined by (5) with*

$$\text{pen}(m) \geq 6x\lambda_1\Gamma(m)/n,$$

for some universal numerical constants $\xi > 0$ *and* $x > 1$*.*

2) *If* $\delta = 1/3$*, let* $\tilde{g} = \hat{g}_{\hat{m}}$ *be defined by (5) with*

$$\text{pen}(m) \geq 2x[\lambda_1 + \mu\sigma^\delta\pi^\delta\lambda_2]\Gamma(m)/n$$

for some universal numerical constants $\xi > 0$ *and* $x > 1$*.*

Then in both cases, $\tilde{g}$ *satisfies*

$$\mathbb{E}(\|g-\tilde{g}\|^2) \leq C_x \inf_{m\in\{1,\ldots,m_n\}}[\|g-g_m\|^2 + \text{pen}(m) + L_m^2(M_2+1)/n] + x\kappa_x C_\xi^2/n, \tag{18}$$

where $C_x = \kappa_x^2 \vee 2\kappa_x$*,* $\kappa_x = (x+1)/(x-1)$ *and* C_ξ^2 *is a constant depending on* f_ε *and on* ξ^2*.*

Remark 4.1. The rates are easy to deduce from (18) as soon as g belongs to some smoothness class, but the procedure will reach the rate without requiring the knowledge of any smoothness parameter. For instance, if g satisfies $(\mathbf{R}_1^X)$, $\|g-g_m\|^2 \leq (C_1/2\pi)L_m^{-2s}\exp\{-2b\pi^r L_m^r\}$, and, associated to the value of pen(m), of order $\Gamma(m)/n$, the estimator $\tilde{g}$ automatically reaches the best rate, without the knowledge of s, r nor b. This best rate is the minimax rate in all cases here, except if $r \geq \delta > 0$ and $\delta \leq 1/3$ which is a case where no lower bounds are available.

If g satisfies $(\mathbf{R}_2^X)$, then according to Section 3.1, $\|g-g_m\|^2 = 0$ as soon as $\pi L_m \geq d$, and therefore the parametric rate of convergence is still automatically achieved without the knowledge of C_2 and d and especially without requiring to know that $(\mathbf{R}_2^X)$ is fulfilled.

Next, we give a result concerning the case where a loss may occur and oracle Inequality (16) is not completely reached. We explain hereafter why this loss is negligible with respect to the rate.

Theorem 4.2. *Under Assumptions* ($\mathbf{A}_1^X$)-($\mathbf{A}_6^X$) *with* $\delta > 1/3$, *consider the collection of estimators* $\hat{g}_m^{(n)}$ *defined by (4) with* $K_n \geq n$ *and* $1 \leq m \leq m_n$ *satisfying*

$$m_n \leq \pi^{-1}\left[\frac{\ln(n)}{2\mu\sigma^\delta} + \frac{2\gamma + (1/2+\delta/2)\wedge 1}{2\delta\mu\sigma^\delta}\ln\left(\frac{\ln(n)}{2\mu\sigma^\delta}\right)\right]^{1/\delta} \tag{19}$$

Let $\Gamma(m) = L_m^{2\gamma+1-\delta}\exp\{2\mu\sigma^\delta(\pi L_m)^\delta\}$, $\lambda_1 = \lambda_1(\gamma, \kappa_0, \mu, \sigma, \delta)$ *be defined by (9) and* $\lambda_2 = \lambda_2(\gamma, \kappa_0, \mu, \sigma, \delta)$ *be defined by (17).*

Let $\tilde{g} = \hat{g}_{\hat{m}}$ *be defined by (5) with*

$$\text{pen}(m) \geq 2x[\lambda_1 + \mu\sigma^\delta\pi^\delta\lambda_2]L_m^{(3\delta/2-1/2)\wedge\delta}\Gamma(m)/n$$

for some universal numerical constants $\xi > 0$ *and* $x > 1$. *Then* $\tilde{g}$ *satisfies (18).*

Remark 4.2. When $\delta > 1/3$, the penalty function pen(m) has not exactly the order of the variance $\Gamma(m)/n$, but a loss of order $L_m^{(3\delta/2-1/2)\wedge\delta}$ occurs, that is of order $L_m^{(3\delta-1)/2}$ if $1/3 < \delta \leq 1$ and of order L_m^δ if $\delta > 1$. Consequently 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). More precisely, when $r = 0$ and $\delta > 0$, the optimal rate of order $(\ln(n))^{-2s/\delta}$ is given by the bias term, and the loss in the penalty function does not change the rate achieved by the adaptive estimator $\tilde{g}$, which remains thus optimal.

When $0 < r < \delta$, the rate is given by the bias term and thus this loss does not affect the rate of convergence of $\tilde{g}$ either. Therefore, according to Butucea and Tsybakov (2004)'s lower bounds, the rate of convergence of $\tilde{g}$ is still the optimal rate of convergence when $s = 0$ and also probably if $s \neq 0$. In the specific case $0 < r < \delta/2$, Butucea and Tsybakov (2004) also propose an adaptive estimator. But this requires to know that $0 < r < \delta/2$.

Let us now focus our discussion on the case where pen(m) can be the dominating term in the trade-off between $\|g - g_m\|^2$ and pen(m), that is when $r \geq \delta > 1/3$. In that case, there is a loss of order $L_m^{(3\delta/2-1/2)\wedge\delta}$ in the penalty function, compared to the variance term. Since it happens in cases where the order of the optimal L_m is less than $(\ln n)^{1/\delta}$, the loss in the rate is at most of order $\ln n$, when the rate is faster than logarithmic and consequently, the loss appears only in cases where it can be seen as negligible.

For $\mathbb{L}_2$ estimation, such an unavoidable logarithmic loss in adaptation, has been pointed out by Tsybakov (2000) and Cavalier et al. (2003) in case of inverse problems with random noise, when $r = \delta = 1$, which shows that, in a slightly different model but with comparable rates of convergence, a loss due to adaptivity of order $\ln(n)^{b/(\mu\sigma+b)}$ is unavoidable, The main point is that, according to (15), our estimator has its quadratic risk with the same logarithmic loss when $r = \delta = 1$. This logarithmic loss due to adaptation seems thus unavoidable at least in one case.

Remark 4.3. Note that, when $\sigma = 0$, then by convention $\delta = \mu = 0$, $\lambda_1 = 1$ and $\text{pen}(m) = 2x(1+2\xi^2)L_m/n$ which is the penalty function used in direct density estimation. More precisely, if σ is very small, then the procedure selects the parameter L_m closed to the parameter selected in usual density estimation.

In conclusion, according to Fan (1991a), Butucea (2004) and Butucea and Tsybakov (2004), in all cases where lower bounds are available ($r = \delta = 0$, $r = 0$ and $\delta > 0$, $r > 0$ and $\delta = 0$, $0 < r < \delta$ and $s = 0$), $\tilde{g}$ achieves automatically the minimax rate of convergence. When $r = \delta = 1$, according to Tsybakov (2000) and Cavalier et al. (2003), $\tilde{g}$ seems to achieve the optimal rate for adaptive estimators. In the last cases, no lower bounds are yet available. That is, when $r \geq \delta > 0$ and $\delta \leq 1/3$, $\tilde{g}$ automatically adjusts and achieves the best rate of non penalized estimators. And when $r \geq \delta > 1/3$, $\tilde{g}$ automatically adjusts and achieves the best rate of non penalized estimators up to a logarithmic loss. Nevertheless, $\tilde{g}$ always improves the rates given by the adaptive estimator built in Pensky and Vidakovic (1999), when both the density and the errors are super smooth.

The adaptive procedure is all the more relevant that it provides an adaptive estimator which achieves the optimal rate of convergence (possibly up to logarithmic factor) in all the cases, without any prior information on the unknown density g, like the knowledge of its smoothness parameters or the comparison of its smoothness with the error density smoothness. In particular it solves almost optimally the problem when the best compromise would not be explicitly computable (see Section 3.3.2).

4.2. **Extension to the mixing case.** We show in this section that all results stated in the independent framework, still hold when Assumption $(\mathbf{A}_3^{X,\varepsilon})$ is replaced by the following assumption

$(\mathbf{A}_7^{\varepsilon,X})$ The ε_i's and the X_i's are both absolutely regular.

Note that, under $(\mathbf{A}_2^{X,\varepsilon})$ and $(\mathbf{A}_7^{\varepsilon,X})$, 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 $\beta_k \leq \beta_k(X) + \beta_k(\varepsilon)$. We refer to Doukhan (1994), pp.4-5 for further references on absolutely regular variables.

For the non penalized estimator $\hat{g}_m^{(n)}$, the bound (7) in Proposition 3.1 becomes the following. Under Assumptions $(\mathbf{A}_1^X)$-$(\mathbf{A}_2^{X,\varepsilon})$, $(\mathbf{A}_4^{\varepsilon})$, $(\mathbf{A}_6^X)$ and $(\mathbf{A}_7^{\varepsilon,X})$, then we have

$$\|g - \hat{g}_m^{(n)}\|^2 \leq \|g - g_m\|^2 + 8(\sum_k \beta_k)\Delta_1(m)/n + L_m^2(M_2+1)/K_n, \tag{20}$$

with $\Delta_1(m)$ and Γ, defined by (6) and (8). As a consequence, if we assume moreover that Assumptions $(\mathbf{A}_5^{\varepsilon})$ and $(\mathbf{R}_1^X)$ hold then, if $K_n \geq n$, Bound (11) becomes

(21)
$$\mathbb{E}(\|g - \hat{g}_m^{(n)}\|^2) \leq A(\sigma^2 L_m^2 \pi^2 + 1)^{-s} \exp\{-2b\pi^r L_m^r\} + 8(\sum_k \beta_k)\lambda_1 \Gamma(m)/n + L_m^2(M_2+1)/n,$$

with λ_1 defined by (9). Consequently Table 1 is still valid and the rates of convergence are the same as the one in the independent framework, as soon as $\sum_k \beta_k < +\infty$. In the same way, under Assumptions $(\mathbf{A}_5^{\varepsilon})$ and $(\mathbf{R}_2^X)$, if $K_n \geq n$, then as soon as $\pi L_m \geq d$, (12) becomes

$$\mathbb{E}(\|g - \hat{g}_m^{(n)}\|^2) \leq 2(\sum_k \beta_k)\lambda_1 d^{(2\gamma+1-\delta)} \exp\left\{2\mu\sigma^\delta \pi^\delta d^\delta\right\}/n + d^2(M_2+1)/(\pi^2 n),$$

and the parametric rate of convergence for estimating g is achieved as soon as $\sum_k \beta_k < +\infty$. The proof of (20) follows the lines of the proof of Proposition 3.1 combined with Viennet's (1997) variance inequality for mixing variables.

For the adaptation, as in the independent case, we would like to find a penalty function, only depending on the observations, such that for $K_n \geq n$,

$$\text{(22)} \qquad \mathbb{E} \parallel \tilde{g} - g \parallel^2 \leq \inf_{m \in \mathcal{M}_n} [\parallel g - g_m \parallel^2 + L_m^2(M_2+1)/n + 8(\sum_k \beta_k)\lambda_1 \Gamma(m)/n].$$

In the β-mixing framework, Theorems 4.1 and 4.2 become the following.

Theorem 4.3. *Under Assumptions $(\mathbf{A}_1^X)$-$(\mathbf{A}_2^{X,\varepsilon})$, $(\mathbf{A}_4^{\varepsilon})$-$(\mathbf{A}_7^{\varepsilon,X})$, consider the collection of estimators $\hat{g}_m^{(n)}$ defined by (4) with $K_n \geq n$. 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 $\Gamma(m) = L_m^{2\gamma+1-\delta} \exp\{2\mu\sigma^\delta(\pi L_m)^\delta\}$, and $\lambda_1(\gamma, \kappa_0, \mu, \sigma, \delta)$ be defined by (9) and let $\lambda_2^\star = \lambda_2^\star(\gamma, \kappa_0, \mu, \sigma, \delta)$ defined by*

$$\text{(23)} \qquad \lambda_2^\star = \begin{cases} \lambda_1(\sum_{k\in\mathbb{N}} \beta_k) & \text{if } \delta > 1 \\ \lambda_1^{1/2}(1+\sigma^2\pi^2)^{\gamma/2}\|h\|_\infty^{1/2}(\sum_{k\in\mathbb{N}}(1+k)\beta_k)^{1/2}\kappa_0^{-1} & \text{if } \delta \leq 1 \end{cases}.$$

1) *If $0 \leq \delta < 1/3$, let $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (5) with m_n satisfying (10) and*

$$\text{pen}(m) = \kappa\lambda_1(\sum_{k\in\mathbb{N}} \beta_k)\Gamma(m)/n,$$

for some universal numerical constant κ.

2) *If $\delta = 1/3$, let $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (5) with m_n satisfying (10) and*

$$\text{pen}(m) = \kappa[\lambda_1 + \mu\sigma^\delta\pi^\delta\lambda_2^\star](\sum_{k\in\mathbb{N}} \beta_k)\Gamma(m)/n$$

where κ is some universal numerical constant.

3) *If $\delta > 1/3$, let $\tilde{g} = \hat{g}_{\hat{m}}$ be defined by (5) with m_n satisfying (19) and*

$$\text{pen}(m) = \kappa[\lambda_1 + \mu\sigma^\delta\pi^\delta\lambda_2^\star](\sum_{k\in\mathbb{N}} \beta_k)L_m^{(3\delta/2-1/2)\wedge\delta}\Gamma(m)/n$$

where κ is some universal numerical constant.

Then in all the cases, $\tilde{g}$ satisfies

$$\mathbb{E}(\|g-\tilde{g}\|^2) \leq K \inf_{m\in\{1,\dots,m_n\}} \left[\|g-g_m\|^2 + \text{pen}(m) + L_m^2(M_2+1)/n\right] + c/n, \tag{24}$$

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

Remark 4.4. Obviously the two previous results hold in the geometrical β-mixing case, that is when $\beta_k \leq C\exp\{-\theta k\}$, for all $k \in \mathbb{N}$ with no condition on the rate $\theta > 0$.

Remark 4.5. This result is mainly a result of robustness which shows that the procedure still works when the variables are not independent but β-mixing. The main drawback of the result in Theorem 4.3 is that the penalty 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 yet found to replace the second one. From a practical point of view, there exists some methods for finding 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 to Birgé and Rozenholc (2002) or Comte and Rozenholc (2001) for further details in other contexts.

Remark 4.6. The previous results are analogous to the results obtained in the independent case, up to the constants. Therefore the comments in Remarks 4.1-4.3 apply here.

5. Simulation study

The implementation is conducted by using Matlab software. The algorithm uses Fast Fourier Transform in order to compute the empirical coefficients

$$\hat{a}_{m,j} = \frac{1}{n}\sum_{k=1}^{n} u^*_{\varphi_{m,j}}(Z_k) = \frac{1}{2\pi n}\sum_{k=1}^{n}\int e^{-ixZ_k}\frac{\varphi^*_{m,j}(x)}{f^*_\varepsilon(\sigma x)}dx$$

rewritten as

$$\hat{a}_{m,j} = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2\pi\sqrt{L_m}}\int_{-\pi L_m}^{\pi L_m}\frac{e^{ix(Z_k-j/L_m)}}{f^*_\varepsilon(\sigma x)}dx = \frac{\sqrt{L_m}}{2\pi}\int_{-\pi}^{\pi} e^{-ijx}\frac{\psi_Z(L_m x)}{f^*_\varepsilon(\sigma L_m x)}dx,$$

where we denote by $\psi_Z(x) = n^{-1}\sum_{k=1}^{n} e^{ixZ_k}$, the empirical Fourier transform of $h(.) = \sigma^{-1}g * f_\varepsilon(./\sigma))$. The algorithm chooses automatically, $\hat{m}$ or $L_{\hat{m}}$ as the minimizer of $\gamma_n(\hat{g}_m^{(n)}) + \text{pen}(L_m)$ with

$$\gamma_n(\hat{g}_m^{(n)}) = -\sum_{|j|\leq K_n}|\hat{a}_{m,j}|^2 = -\|\hat{g}_m^{(n)}\|^2,$$

where K_n is chosen as $K_n = 2^8$.

The integrated squared error (ISE) $\|\hat{g}_m^{(n)} - g\|^2$ is computed via a standard approximation and discretization of the integral on an interval of $\mathbb{R}$ denoted by I and given in each case. Then the MISE, $\mathbb{E}\|\hat{g}_m^{(n)} - g\|^2$ is computed by the empirical mean of the approximated ISE $\|\hat{g}_m^{(n)} - g\|^2$, over 500 simulation samples.

We illustrate our method on some test densities (with different smoothness properties) and in the two contexts of errors, ordinary and super smooth. We start by describing the errors densities and the associated penalties.

5.1. **Two settings for the errors and the associated penalties.** We consider two types of error density f_ε, the first one is ordinary smooth, with polynomial decay of the Fourier Transform, and the second one is supersmooth, with an exponential decay of the Fourier transform f_ε^*.

- **Case 1: Double exponential (or Laplace) ε's.** In this case, the density of ε is given by

$$f_\varepsilon(x) = e^{-\sqrt{2}|x|}/\sqrt{2}, \quad f_\varepsilon^*(x) = (1 + x^2/2)^{-1}. \tag{25}$$

This density corresponds to centered ε's with variance 1, and satisfying $(\mathbf{A}_5^\varepsilon)$ with $\gamma = 2$, $\kappa_0 = 1/2$ and $\mu = \delta = 0$.
According to Theorem 4.1, the penalty function has the variance order and is in fact evaluated as

$$\kappa(L_m/n)\int_{-\pi}^{\pi} |\varphi^*(x)/f_\varepsilon^*(\sigma L_m x)|^2 dx,$$

where, here,

$$\int_{-\pi}^{\pi} |\varphi^*(x)/f_\varepsilon^*(\sigma L_m x)|^2 dx = 2\pi\left(1 + \frac{\pi^2}{3}\sigma^2 {L_m}^2 + \frac{\pi^4}{20}\sigma^4 {L_m}^4\right).$$

Several simulations lead to fix $\kappa = 3$ and to choose the following penalty

$$\text{pen}(L_m) = \frac{6\pi L_m}{n}\left(1 + \frac{(\ln(L_m))^{2.5}}{L_m} + \frac{\pi^2}{3}\sigma^2 {L_m}^2 + \frac{\pi^4}{20}\sigma^4 {L_m}^4\right).$$

The additional term $(\ln(L_m))^{2.5}/L_m$ is motivated by the works of Birgé and Rozenholc (2002) and Comte and Rozenholc (2004). In our case also, this term improves the quality of the results by making the penalty slightly heavier when L_m becomes large.

- **Case 2: Gaussian ε's.** In that case, the errors density f_ε is given by

$$f_\varepsilon(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \quad f_\varepsilon^*(x) = e^{-x^2/2}. \tag{26}$$

This density satisfies $(\mathbf{A}_5^\varepsilon)$ with $\gamma = 0$, $\kappa_0 = 1$, $\delta = 2$ and $\mu = 1/2$.
In this case, according to Theorems 4.1 and 4.2, the penalty is slightly bigger than the variance term, that is of order

$$\kappa {L_m}^{(3\delta/2 - 1/2)\wedge\delta}(L_m/n)\int_{-\pi}^{\pi} |\varphi^*(x)/f_\varepsilon^*(\sigma L_m x)|^2 dx \text{ with } \delta = 2,$$

and

$$\int_{-\pi}^{\pi} |\varphi^*(x)/f_\varepsilon(\sigma L_m x)|^2 dx = \int_{-\pi}^{\pi} \exp(\sigma^2 {L_m}^2 x^2) dx.$$

As in the previous case, several simulations lead to fix $\kappa = 3$ and to choose the following penalty

$$\text{pen}(L_m) = \frac{6\pi L_m}{n}\left(1 + \frac{(\ln(L_m))^{2.5}}{L_m} + \frac{\pi^2\sigma^2 {L_m}^2}{3}\right)\left(\int_0^{\pi} \exp(\sigma^2 {L_m}^2 x^2) dx/\pi\right),$$

where the integral is numerically computed. According to the theory (see Theorem 4.2), the loss due to the adaptation is the term $\pi^2\sigma^2 {L_m}^2/3$. As previously, the additional term $(\ln(L_m))^{2.5}/L_m$ is motivated by simulations and the works of Birgé and Rozenholc (2002) and Comte and Rozenholc (2004).

Remark 5.1. Note that when $\sigma = 0$, both penalties are equal to $(6\pi L_m)(1+(\ln(L_m))^{2.5}/L_m)/n$.

5.2. **Test densities.** First we consider densities having classical smoothness properties like Hölderian smoothness with polynomial decay of their Fourier transform. Second we consider densities having stronger smoothness properties, with exponential decay of the Fourier transform.

Except in the case of the infinite variance density (Cauchy density), we consider density functions g normalized with unit variance so that $1/\sigma^2$ represents the usual signal-to-noise ratio (variance of the signal divided by the variance of the noise) and is denoted in the sequel by $s2n$ defined as $s2n = 1/\sigma^2$.

The functions which are considered are listed below, associated with the interval I used to evaluate the ISE:

(a) Chi2(3)-type distribution, $X = 1/\sqrt{6}U$, $g_X(x) = \sqrt{6}g(\sqrt{6}x)$, $U \sim \chi^2(3)$ where we know that $U \sim \Gamma(\frac{3}{2}, \frac{1}{2})$,

$$g_U(x) = \frac{1}{2^{5/2}\Gamma(3/2)} e^{-|x|/2}|x|^{1/2}, g_U^*(x) = \frac{1}{(1-2ix)^{3/2}},$$

and $I = [-1, 16]$.

(b) Laplace distribution, as given in (25), $I = [-5, 5]$.

(c) Mixed Gamma distribution, $X = 1/\sqrt{5.48}W$ with $W \sim 0.4\Gamma(5,1) + 0.6\Gamma(13,1)$,

$$g_W(x) = [0.4 * \frac{x^4 e^{-x}}{\Gamma(5)} + 0.6\frac{x^{12}e^{-x}}{\Gamma(13)}]\mathbb{1}_{\mathbb{R}^+}(x), g_W^*(x) = \frac{0.4}{(1-ix)^5} + \frac{0.6}{(1-ix)^{13}},$$

and $I = [-1.5, 26]$.

(d) Cauchy distribution, $g(x) = (1/\pi)(1/(1+x^2))$, $g^*(x) = e^{-|x|}$, $I = [-10, 10]$.

(e) Gaussian distribution, $X \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 1$, $I = [-4, 4]$.

(f) Mixed Gaussian distribution: $X \sim \sqrt{2}V$ with $V \sim 0.5\mathcal{N}(-3,1) + 0.5\mathcal{N}(2,1)$

$$g_V(x) = 0.5\frac{1}{\sqrt{2\pi}}(e^{-(x+3)^2/2} + e^{-(x-2)^2/2)}),\ g_V^*(x) = 0.5(e^{-3ix} + e^{2ix})e^{-x^2/2},$$

and $I = [-8, 7]$.

Densities (a), (b), (c) correspond to cases with $r = 0$ (Hölderian smoothness properties) with different values of s, whereas densities (d), (e), (f) correspond to cases with $r > 0$ (infinitely many times differentiable) with different values for the power r.

5.3. **Results.** Figure 1 illustrates the choice performed by the algorithm, when selecting one dimension among several possibilities.

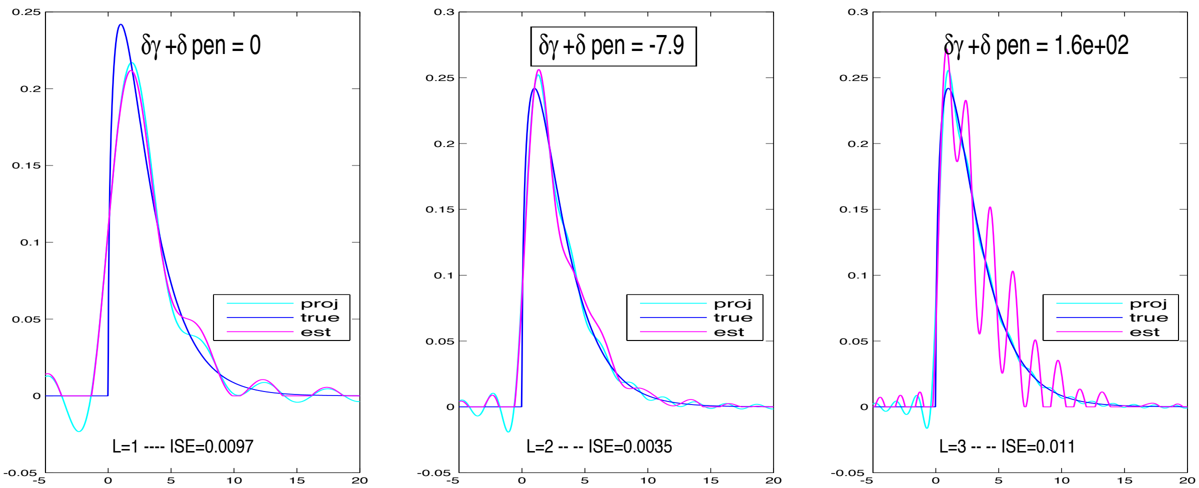

FIGURE 1. Plots of the estimator of the true density and of its projection, when estimating the Chi2 density - Laplace errors - $n = 750$, $s2n$=10, for different values of L. The algorithm chooses $L = 2$, $\delta\gamma + \delta\text{pen}$ is the difference between the penalized contrast for the considered L and for $L = 1$.

Table 2 presents the MISE for the two types of errors, the different tested densities, different $s2n$ and different sample sizes. The greatest values of $s2n$ amount to consider that there is essentially no noise. In those cases, the results would have to be compared with results coming from a direct density estimation method.

The main comment on Table 2 concerns the importance of σ. Clearly the MISE are smaller when there is less noise (σ small, $s2n$ large). Moreover, it appears that the results are globally very good.

We can in particular compare the performances of our adaptive estimator with the performances of the deconvolution kernel as presented in Delaigle and Gijbels (2004). This comparison is done for densities (a), (c), (e) and (f) which correspond to the densities #2, #6, #1 and #3 respectively, in Delaigle and Gijbels (2004). They give median ISE obtained with kernel estimators by using four different methods of bandwidth selection. The comparison is given in Table 3 between the median ISE computed for 500 samples generated with the same interval length and signal to noise ratio as Delaigle and Gijbels (2004). The ISE are computed on the same intervals I as them. We also give our corresponding means since we believe that they are more meaningful than medians since the MISE is $\mathbb{E}\|\hat{g}_m^{(n)} - g\|^2$, but we also give our medians. It is noteworthy that it may happen that medians seems much better because means can become

$\times 10^{-2}$		$n = 100$		$n = 250$		$n = 500$		$n = 1000$		$n = 2500$	
g	$s2n$	Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.
Chi2(3)	2	2.02	4.15	1.39	2.37	1.18	1.72	1.06	1.36	1.03	1.12
	4	1.52	1.79	1.21	1.27	1.07	1.13	1.04	1.04	0.654	0.996
	10	1.31	1.31	1.13	1.11	1.01	1.03	0.505	0.995	0.345	0.974
	10^2	1.22	1.23	0.72	0.884	0.409	0.411	0.327	0.335	0.179	0.232
	10^3	1.22	1.21	0.651	0.638	0.391	0.382	0.293	0.298	0.157	0.157
Laplace	2	3.7	10.6	2.17	5.2	1.61	3.03	1.41	2.07	1.2	1.48
	4	2.5	2.99	1.66	1.93	1.33	1.46	1.26	1.25	0.817	1.12
	10	1.9	1.97	1.43	1.42	1.35	1.22	0.723	1.12	0.441	1.06
	10^2	1.69	1.64	0.883	1.06	0.607	0.538	0.453	0.385	0.343	0.211
	10^3	1.68	1.65	0.814	0.79	0.593	0.561	0.411	0.379	0.284	0.24
Mix.Gamma	2	1.32	3.96	0.547	1.88	0.292	1.01	0.148	0.533	0.06	0.224
	4	0.79	1.05	0.316	0.453	0.151	0.224	0.0815	0.116	0.0361	0.0497
	10	0.495	0.524	0.194	0.215	0.103	0.11	0.0543	0.0565	0.024	0.0246
	10^2	0.369	0.384	0.152	0.149	0.0789	0.0785	0.0409	0.0412	0.0194	0.0186
	10^3	0.364	0.353	0.149	0.15	0.0762	0.0767	0.0404	0.0406	0.0184	0.0185
Cauchy	2	2.72	9.09	1.22	4.26	0.645	2.3	0.353	1.25	0.158	0.513
	4	1.66	2.27	0.716	0.967	0.364	0.514	0.205	0.28	0.138	0.127
	10	1.15	1.13	0.437	0.46	0.249	0.257	0.215	0.142	0.219	0.0764
	10^2	0.815	0.783	0.373	0.351	0.351	0.271	0.206	0.201	0.147	0.0962
	10^3	0.783	0.78	0.366	0.355	0.34	0.331	0.189	0.189	0.121	0.118
Gauss.	2	2.74	9.21	1.1	4.08	0.605	2.14	0.296	1.06	0.143	0.446
	4	1.59	2.23	0.591	0.878	0.362	0.457	0.229	0.227	0.463	0.0894
	10	0.885	1.02	0.397	0.42	0.372	0.21	0.515	0.112	0.229	0.046
	10^2	0.711	0.713	0.565	0.432	0.396	0.394	0.279	0.195	0.171	0.15
	10^3	0.739	0.705	0.606	0.592	0.352	0.355	0.259	0.246	0.167	0.145
Mix. Gauss.	2	2.97	9.98	1.26	4.45	0.693	2.31	0.328	1.26	0.132	0.509
	4	1.73	2.37	0.709	1.02	0.375	0.478	0.185	0.257	0.0751	0.105
	10	1.14	1.21	0.463	0.466	0.237	0.242	0.118	0.122	0.0468	0.0515
	10^2	0.851	0.817	0.359	0.352	0.166	0.167	0.0866	0.0867	0.034	0.0351
	10^3	0.823	0.828	0.344	0.327	0.169	0.163	0.0845	0.0839	0.0334	0.0336

TABLE 2. Mean MISE $\times 100$ obtained with $N = 500$ samples, for different sample size ($n = 100, 250, 500, 1000, 2500$) and different values of $s2n$ (2, 4, 10, 100, 1000), the higher $s2n$ the lower the noise level.

huge simply because a few numbers of bad paths. The cost of such errors seems therefore to have a price given by means and completely hidden by medians.

We can see that our estimation procedure provides better results in all cases except in one case, namely when we aim at estimating a Gaussian density, for both types of error density. This is most probably due to the fact that the bandwidth selection methods are based on computations assuming that the underlying density is Gaussian, so that they perform very well when it is true. For the other cases, even our means are often better than Delaigle and

		$n = 100$		$n = 250$	
density g	method	ε Lap.	ε Gaus.	ε Lap.	ε Gaus.
(a) or #2 $\chi^2(3)$ ($s2n$=4)	DG, lower median	0.015	0.018	—	—
	DG, higher median	0.018	0.022	—	—
	Proj.: median	0.014	0.016	—	—
	Proj.: mean	0.015	0.018	—	—
(c) or #6 Mix.Gamma ($s2n$=10)	DG, lower median	—	—	0.0021	0.0023
	DG, higher median	—	—	0.0024	0.0026
	Proj.: median	—	—	0.0017	0.0020
	Proj., mean	—	—	0.0019	0.0021
(e) or #1 Gauss ($s2n$=4)	DG, lower median	0.0071	0.0080	0.0041	0.0051
	DG, higher median	0.011	0.012	0.0059	0.0072
	Proj.: median	0.012	0.017	0.0049	0.0066
	Proj.: mean	0.016	0.022	0.0059	0.0088
(f) or #3 Mix.Gauss ($s2n$=4)	DG, lower median	0.018	0.027	0.011	0.020
	DG, higher median	0.031	0.034	0.023	0.028
	Proj.: median	0.016	0.022	0.0063	0.0088
	Proj.: mean	0.017	0.024	0.0071	0.010

TABLE 3. Median ISE obtained by Delaigle and Gijbels (2004) with a kernel estimator and four different strategies of bandwidth selection, and with our penalized projection estimator (median and mean).

Gijbels'(2004) medians which shows that our method provides a very good solution to the deconvolution problem. We may also emphasize that our algorithm is a fast algorithm.

6. Proofs

6.1. **Proof of Proposition 3.1.** According to (4), for any given m belonging to $\mathcal{M}_n$, $\hat{g}_m^{(n)}$ satisfies, $\gamma_n(\hat{g}_m^{(n)}) - \gamma_n(g_m^{(n)}) \leq 0$. Denoting by $\nu_n(t)$ the centered empirical process

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left[u_t^*(Z_i) - \langle t, g\rangle\right], \tag{27}$$

we have that

$$\gamma_n(t) - \gamma_n(s) = \|t-g\|^2 - \|s-g\|^2 - 2\nu_n(t-s), \tag{28}$$

and therefore,

$$\|g - \hat{g}_m^{(n)}\|^2 \leq \|g - g_m^{(n)}\|^2 + 2\nu_n(\hat{g}_m^{(n)} - g_m^{(n)}). \tag{29}$$

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

$$\nu_n(\hat{g}_m^{(n)} - g_m^{(n)}) = \sum_{|j|\leq K_n}(\hat{a}_{m,j} - a_{m,j})\nu_n(\varphi_{m,j}) = \sum_{|j|\leq K_n}[\nu_n(\varphi_{m,j})]^2, \tag{30}$$

and consequently

$$\mathbb{E}\|g-\hat{g}_m^{(n)}\|^2 \le \|g-g_m^{(n)}\|^2 + 2\sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})]. \tag{31}$$

Now, since the X_i's and the ε_i's are independent and identically distributed random variables, we get 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 6.2 in Section 6.4 to get that $\sum_{j\in\mathbb{Z}}\mathrm{Var}[\nu_n(\varphi_{m,j})] \le \Delta_1(m)/n$, where $\Delta_1(m)$ is defined by (6). It remains to study $\|g-g_m^{(n)}\|^2$. By applying Pythagoras Theorem, we have $\|g-g_m^{(n)}\|^2 = \| g-g_m \|^2 + \|g_m-g_m^{(n)}\|^2$, where $\|g_m-g_m^{(n)}\|^2 = \sum_{|j|\ge K_n} a_{m,j}^2 \le (\sup_j j a_{m,j})^2 \sum_{|j|\ge K_n} j^{-2}$. Now we write that

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

This implies finally that $j a_{m,j} \le L_m (M_2)^{1/2} + \sqrt{L_m}$, and (7) follows. □

6.2. **Proof of Theorems 4.1 and 4.2 : the i.i.d. case.** By definition, $\tilde{g}$ satisfies that for all $m\in\mathcal{M}_n$, $\gamma_n(\tilde{g}) + \mathrm{pen}(\hat{m}) \le \gamma_n(g_m^{(n)}) + \mathrm{pen}(m)$. Therefore, by applying (28) we get that

$$\| \tilde{g} - g \|^2 \le \| g_m^{(n)} - g \|^2 + 2\nu_n(\tilde{g} - g_m^{(n)}) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}). \tag{32}$$

Next, we use that if $t = t_1 + t_2$ with t_1 in $S_m^{(n)}$ and t_2 in $S_{m'}^{(n)}$, then t is such that t^* has its support in $[-\pi L_{m\vee m'}, \pi L_{m\vee m'}]$ and therefore t belongs to $S_{m\vee m'}^{(n)}$. If we denote by $B_{m,m'}(0,1)$ the set $B_{m,m'}(0,1) = \{t\in S_{m\vee m'}^{(n)} \;/\; \|t\| = 1\}$, then $|\nu_n(\tilde{g}-g_m^{(n)})| \le \|\tilde{g}-g_m^{(n)}\| \sup_{t\in B_{m,\hat{m}}(0,1)} |\nu_n(t)|$. Consequently, by using that $2ab \le x^{-1}a^2 + xb^2$, for $x>1$, we get

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

and therefore, by writing that $\|\tilde{g}-g_m^{(n)}\|^2 \le (1+y^{-1})\|\tilde{g}-g\|^2 + (1+y)\|g-g_m^{(n)}\|^2$, with $y=(x+1)/(x-1)$ for $x>1$, we infer that

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

Choose some positive function $p(m,m')$ such that $xp(m,m') \le \mathrm{pen}(m) + \mathrm{pen}(m')$. Consequently, for $\kappa_x = (x+1)/(x-1)$ we have

$$\begin{aligned} \|\tilde{g}-g\|^2 \le \kappa_x^2\left[\|g-g_m\|^2 + \|g_m - g_m^{(n)})\|^2\right] + x\kappa_x W_n(\hat{m}) \\ + \kappa_x\left(xp(m,\hat{m}) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m})\right) \end{aligned} \tag{33}$$

$$\text{with} \qquad W_n(m') := [\sup_{t\in B_{m,m'}(0,1)} |\nu_n(t)|^2 - p(m,m')]_+, \tag{34}$$

that is, according to the proof of Proposition 3.1,

$$\|\tilde{g}-g\|^2 \le \kappa_x^2\|g-g_m\|^2 + \kappa_x^2(M_2+1)L_m^2/K_n + 2\kappa_x\mathrm{pen}(m) + x\kappa_x \sum_{m'\in\mathcal{M}_n} W_n(m'). \tag{35}$$

The main point of the proof lies in studying $W_n(m')$, and more precisely in finding $p(m,m')$ such that for a constant K,

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}(W_n(m')) \le K/n. \tag{36}$$

In this case, combining (35) and (36) we infer that, for all m in $\mathcal{M}_n$,

$$\mathbb{E}\|g-\tilde{g}\|^2 \le \kappa_x^2\|g-g_m^{(n)}\|^2 + \kappa_x^2(M_2+1)L_m^2/K_n + 2\kappa_x\mathrm{pen}(m) + x\kappa_x K/n,$$

which can also be written

$$\mathbb{E}\|g-\tilde{g}\|^2 \le C_x \inf_{m\in\mathcal{M}_n}\left[\|g-g_m\|^2 + \mathrm{pen}(m) + (M_2+1)L_m^2/K_n\right] + x\kappa_x K/n, \tag{37}$$

where $C_x = \kappa_x^2 \vee 2\kappa_x$ suits. It remains thus to find $p(m,m')$ such that (36) holds. This will be done by applying the following immediate integration of Talagrand's Inequality (see Talagrand (1996)):

Lemma 6.1. *Let $Y_1,\dots,Y_n$ be i.i.d. random variables and $r_n(f) = (1/n)\sum_{i=1}^n [f(X_i) - \mathbb{E}(f(X_i))]$ for f belonging to a countable class $\mathcal{F}$ of uniformly bounded measurable functions. Then for $\xi^2>0$*

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}|r_n(f)|^2 - 2(1+2\xi^2)H^2\right]_+ \le \frac{6}{K_1}\left(\frac{v}{n}\exp\left\{-K_1\xi^2\frac{nH^2}{v}\right\} + \frac{8M_1^2}{K_1n^2C^2(\xi^2)}\exp\left\{-\frac{K_1C(\xi)\xi}{\sqrt{2}}\frac{nH}{M_1}\right\}\right), \tag{38}$$

with $C(\xi)=\sqrt{1+\xi^2}-1$, K_1 is a universal constant, and where

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

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 6.4 and Inequality (8) in Section 6.4, we propose to take

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

where $\lambda_1 = \lambda_1(\gamma,\kappa_0,\mu,\sigma,\delta)$ is defined by (9). Again, by applying Lemma 6.4, we take $v \ge \Delta_2(m^*,h)$ with

$$\Delta_2(m,h) = L_m^2 \iint \left|\frac{\varphi^*(x)\varphi^*(y)}{f_\varepsilon^*(\sigma L_m x) f_\varepsilon^*(\sigma L_m y)} h^*(L_m(x-y))\right|^2 dxdy. \tag{39}$$

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

$$\begin{aligned}\Delta_2(m,h) &\leq \kappa_0^{-2} L_m^2 (1+(\sigma\pi L_m)^2)^\gamma \exp\{2\mu\sigma^\delta(\pi L_m)^\delta\} \int_{-\pi}^{\pi} \frac{dx}{|f_\varepsilon^*(\sigma L_m x)|^2} \int |h^*(L_m u)|^2 du \\ &\leq 2\kappa_0^{-2}\pi\lambda_1(1+\sigma^2\pi^2)^\gamma \|h^*\|^2 L_m^{4\gamma+1-\delta} \exp\{4\mu\sigma^\delta(\pi L_m)^\delta\}.\end{aligned}$$

Using that $\|h^*\|^2 \leq \|f_\varepsilon^*\|^2 < \infty$ under $(\mathbf{A}_4^\varepsilon)$, we take $v = \lambda_2 L_{m^*}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)} \exp\{2\mu\sigma^\delta(\pi L_{m^*})^\delta\}$, where $\lambda_2 = \lambda_2(\gamma,\kappa_0,\mu,\sigma,\delta)$ is defined in Theorem 4.1. From the definition (34) of $W_n(m')$, by taking $p(m,m') = 2(1+2\xi^2)H^2$, we get that

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

By applying (38), we get the global bound $\mathbb{E}(W_n(L_{m'})) \leq K[I(L_{m^*}) + II(m^*)]$, where $I(m^*)$ and $II(m^*)$ are defined by

$$I(m^*) = \frac{\lambda_2 L_{m^*}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)} \exp\{2\mu\sigma^\delta(\pi L_{m^*})^\delta\}}{n} \exp\{-K_1\xi^2(\lambda_1/\lambda_2) L_{m^*}^{(1/2-\delta/2)_+}\}$$

$$\text{and} \qquad II(m^*) = \frac{\lambda_1 L_{m^*}^{2\gamma+1-\delta} e^{2\mu\sigma^\delta(\pi L_{m^*})^\delta}}{n^2} \exp\left\{-K_1\xi C(\xi)\sqrt{n}/\sqrt{2}\right\},$$

with $\lambda_2 = \lambda_2(\gamma,\kappa_0,\mu,\sigma,\delta)$ defined in Theorem 4.1.

• Study of $\sum_{m\in\mathcal{M}_n} II(m^*)$. According to the choices for v, H^2 and M_1 we have

$$\sum_{m\in\mathcal{M}_n} II(m^*) \leq |\mathcal{M}_n| \exp\left\{-K_1\xi C(\xi)\sqrt{n}/\sqrt{2}\right\} 2\lambda_1\Gamma(m_n)/n^2.$$

Consequently, since under (10), $\Gamma(m_n)/n$ is bounded, $\sum_{m\in\mathcal{M}_n} II(m^*) \leq C/n$.

• Study of $\sum_{m\in\mathcal{M}_n} I(m^*)$. 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 $a,b \geq 1$, we infer that

$$\begin{aligned}(a\vee b)^\psi e^{2\mu\sigma^\delta\pi^\delta(a\vee b)^\delta} e^{-K'\xi^2(a\vee b)^\omega} &\leq (a^\psi e^{2\mu\sigma^\delta\pi^\delta a^\delta} + b^\psi e^{2\mu\sigma^\delta\pi^\delta b^\delta}) e^{-(K'\xi^2/2)(a^\omega+b^\omega)} \\ &\leq a^\psi e^{2\mu\sigma^\delta\pi^\delta a^\delta} e^{-(K'\xi^2/2)a^\omega} e^{-(K'\xi^2/2)b^\omega} + b^\psi e^{2\mu\sigma^\delta\pi^\delta b^\delta} e^{-(K'\xi^2/2)b^\omega}.\end{aligned} \tag{40}$$

Consequently, if we denote by $\tilde\Gamma$ the quantity $\tilde\Gamma(m) = L_{m^*}^{2\gamma+(1/2-\delta/2)\wedge(1-\delta)} \exp\{2\mu\sigma^\delta(\pi L_{m^*})^\delta\}$ then

$$\begin{aligned}\sum_{m'\in\mathcal{M}_n} I(m^*) &\leq \frac{2\lambda_2\tilde\Gamma(m)}{n} \exp\{-(K'\xi^2/2)(L_m)^{(1/2-\delta/2)}\} \sum_{m'\in\mathcal{M}_n} \exp\{-(K'\xi^2/2)(L_{m'})^{(1/2-\delta/2)}\} \\ &\quad + \sum_{m'\in\mathcal{M}_n} \frac{2\lambda_2\tilde\Gamma(m')}{n} \exp\{-(K'\xi^2)(L_{m'})^{(1/2-\delta/2)}\}.\end{aligned} \tag{41}$$

1) Case $0 \leq \delta < 1/3$. In that case, since $\delta < (1/2-\delta/2)_+$, the choice $\xi^2 = 1$ ensures that $\tilde\Gamma(m)\exp\{-(K'\xi^2/2)(L_m)^{(1/2-\delta/2)}\}$ is bounded and thus the first term in (41) is bounded by C/n. In the same way, since $1 \leq m \leq m_n$ which satisfies (10),

$$\sum_{m'\in\mathcal{M}_n} \frac{\tilde\Gamma(m')}{n} \exp\{-(K'\xi^2)(L_{m'})^{(1/2-\delta/2)}\} \leq \tilde{\tilde{C}}/n,$$

and hence $\sum_{m'\in\mathcal{M}_n} I(m^*) \le C/n$. Consequently, (36) hold if we choose $\mathrm{pen}(m) = 2x(1+2\xi^2)\lambda_1(L_m)^{2\gamma+1-\delta}\exp\{2\mu\sigma^\delta(\pi L_m)^\delta\}/n$.
2) Case $\delta = 1/3$. In that case, bearing in mind Inequality (40) we choose ξ^2 such that $2\mu\sigma^\delta\pi^\delta(L_{m^*})^\delta - (K'\xi^2/2)L_{m^*}^\delta = -2\mu\sigma^\delta(\pi L_{m^*})^\delta$ that is $\xi^2 = (4\mu\sigma^\delta\pi^\delta\lambda_2)/(K_1\lambda_1)$. By the same arguments as for the case $0\le\delta<1/3$, this choice ensures that $\sum_{m'\in\mathcal{M}_n} I(m^*)\le C/n$, and consequently (36) holds. The result if we take $p(m,m') = 2(1+2\xi^2)\lambda_1 L_{m^*}^{2\gamma+1-\delta}\exp(2\mu\sigma^\delta(\pi L_{m^*})^\delta)/n$, and $\mathrm{pen}(m) = 2x(1+2\xi^2)\lambda_1(L_m)^{2\gamma+1-\delta}\exp(2\mu\sigma^\delta(\pi L_m)^\delta)/n$.
3) Case $\delta > 1/3$. In that case, $\delta > (1/2-\delta/2)_+$. Bearing in mind Inequality (40) we choose $\xi^2 = \xi^2(L_m, L_{m'})$ such that $2\mu\sigma^\delta\pi^\delta(L_{m^*})^\delta - (K'\xi^2/2)L_{m^*}^\omega = -2\mu\sigma^\delta\pi^\delta(L_{m^*})^\delta$ that is $\xi^2 = \xi^2(m,m') = (4\mu\sigma^\delta\pi^\delta\lambda_2)/(K_1\lambda_1)L_{m^*}^{\delta-\omega}$. This choice ensures that $\sum_{m'\in\mathcal{M}_n} I(m^*) \le C/n$, and consequently (36) holds and (18) follows if $p(m,m') = 2(1+2\xi^2(m,m'))\lambda_1 L_{m^*}^{2\gamma+1-\delta}\exp(2\mu\sigma^\delta(\pi L_{m^*})^\delta)/n$, and $\mathrm{pen}(m) = 2x(1+2\xi^2(L_m,m))\lambda_1(L_m)^{2\gamma+1-\delta}\exp(2\mu\sigma^\delta(\pi L_m)^\delta)/n$. □

6.3. **Proof of Theorem 4.3: the absolutely regular case.** The proof of the absolutely regular case is rather similar to the one of the independent case with some additional technicalities due to the approximation of the dependent variables by blockwise independent variables, based on Berbee's coupling Lemma extended to sequences (see Bryc's [6] construction). It uses also Delyon's (1990) covariance inequality, successfully exploited by Viennet (1997) for partial sums of strictly stationary variables. Since the methods and tools are standard, the proof is omitted for the sake of place but is available from the authors upon request.

6.4. **Technical Lemmas.**

Lemma 6.2. *Let $\nu_n(t)$ be defined by (27), $\Delta_1(m)$ be defined by (6). Under Assumptions ($\mathbf{A}_1^X$), ($\mathbf{A}_2^{X,\varepsilon}$), ($\mathbf{A}_3^{X,\varepsilon}$), then we have*

$$\| \sum_{j\in\mathbb{Z}} |u^*_{\varphi_{m,j}}|^2 \|_\infty \le \Delta_1(m), \text{ and } \sup_{g\in\mathcal{S}_{s,r,b}(A)} \sum_{j\in\mathbb{Z}} \mathrm{Var}[\nu_n(\varphi_{m,j})] \le \Delta_1(m)/n. \tag{42}$$

Proof of Lemma 6.2. Use the definition of $u^*_{\varphi_{m,j}}(z)$ to get that

$$\sum_{j\in\mathbb{Z}}\left|u^*_{\varphi_{m,j}}(z)\right|^2 = \sum_{j\in\mathbb{Z}}\left|\int\exp\{ixz\}u_{\varphi_{m,j}}(x)dx\right|^2 = \frac{L_m}{(2\pi)^2}\sum_{j\in\mathbb{Z}}\left|\int\exp\{-ixzL_m\}\exp\{ijx\}\frac{\varphi^*(x)}{f^*_\varepsilon(xL_m\sigma)}dx\right|^2.$$

By Parseval's Formula,

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

which entails that the first part of the bound (42) is proved. The second part follows since

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

Lemma 6.3. *Let $\Delta_1(m)$ and $R(\mu,\delta,\sigma)$ be defined by (6) and (9). Then under Assumption ($\mathbf{A}_5^\varepsilon$), $\Delta_1(m) \le \dfrac{1}{\pi\kappa_0^2R(\mu,\delta,\sigma)}(\pi L_m)^{1-\delta}(\sigma^2L_m^2\pi^2+1)^\gamma\exp\{2\mu\sigma^\delta\pi^\delta L_m^\delta\}$.*

Proof of Lemma 6.3. Under Assumption($\mathbf{A}_5^\varepsilon$), $\Delta_1(m)$ is bounded in the following way,

$$\Delta_1(m) \leq \frac{1}{\pi\kappa_0^2}(\sigma^2 L_m^2\pi^2+1)^\gamma \int_0^{\pi L_m} \exp\{2\mu\sigma^\delta u^\delta\}du.$$

If $\delta = 0$, by convention $\mu = 0$, and hence the integral in the above bound is less than πL_m.

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

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

and therefore

$$\int_0^{\pi L_m} \exp\{2\mu\sigma^\delta u^\delta\}du \leq \frac{(\pi L_m)^{1-\delta}}{2\mu\sigma^\delta\delta}\exp(2\mu\sigma^\delta(\pi L_m\sigma)^\delta).$$

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

$$\int_0^{\pi L_m} \exp\{2\mu\sigma^\delta u^\delta\}du \leq \int_0^{\pi L_m} \exp\{2\mu\sigma^\delta(\pi L_m)^{\delta-1}u\}du \leq \frac{(\pi L_m)^{1-\delta}}{2\mu\sigma^\delta}\exp(2\mu\sigma^\delta(\pi L_m)^\delta),$$

and consequently Lemma 6.3 follows □.

Lemma 6.4. *Let $\nu_n(t)$, $\Delta_1(m)$ and $\Delta_2(m,h)$ be defined by (27), (6) and (39). Then under Assumptions ($\mathbf{A}_1^X$), ($\mathbf{A}_2^{X,\varepsilon}$), ($\mathbf{A}_3^{X,\varepsilon}$) and ($\mathbf{A}_4^\varepsilon$) we have*

$$\sup_{t\in B_{m,m'}(0,1)} \| u_t^* \|_\infty \leq \sqrt{\Delta_1(m^*)} \qquad \mathbb{E}[\sup_{t\in B_{m,m'}(0,1)}|\nu_n(t)|] \leq \sqrt{\Delta_1(m^*)/n},$$

$$\text{and} \sup_{t\in B_{m,m'}(0,1)} \mathrm{Var}(u_t^*(Z_1)) \leq \sqrt{\Delta_2(m^*,h)}/(2\pi).$$

Proof of Lemma 6.4. By combining Cauchy-Schwarz Inequality and (43), the square of the first term $\sup_{t\in B_{m,m'}(0,1)} \| u_t^* \|_\infty^2$ is bounded by

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

Now,

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

which is bounded, by applying the second part of (42) in Lemma 6.2, by $\sqrt{\Delta_1(m^*)/n}$. Now write that

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

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

$$Q_{j,k}(m) = \frac{L_m}{(2\pi)^2}\iint \exp\{ijx - iky\}\frac{\varphi^*(x)\varphi^*(y)}{f^*_\varepsilon(\sigma L_m x)f^*_\varepsilon(\sigma L_m y)}h^*(L_m(x-y))dxdy.$$

Apply Parseval's Formula to get the result since

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

□

References

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

[2] E. Belister & B. Levit (2001). Asymptotically minimax estimation in infinitely smooth density with censored data. *Annals of the Institute of Statistical Mathematics* **53** 289-306.

[3] H.C.P. Berbee (1979). *Random walks with stationary increments and renewal theory.* Central Mathematical Tracts, Amsterdam.

[4] L. Birgé & P. Massart (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) 55-87.

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

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

[7] C. Butucea (2004). Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics* **32**, 181-192.

[8] C. Butucea & A.B. Tsybakov (2004). Sharp optimality and some effects of dominating bias in density deconvolution, Preprint LPMA-898, http://www.proba.jussieu.fr/mathdoc/preprints/index.html#2004.

[9] C. Butucea and C. Matias (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli* **11**(2), to appear.

[10] R.J. Carroll, R.J. & P. Hall (1988). Optimal rates of convergence for deconvolving a density. *Journal of the Amererican Statistical Association* **83**, 1184-1186.

[11] E.A. Cator (2001). Deconvolution with arbitrarily smooth kernels. *Statistics and Probability Letters* **54**, 205-214.

[12] Cavalier L., Golubev Y., Lepsi O. Tsybakov A. B. (2003) Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems.*Theorey Prob. Appl.***48**.

[13] L. Cavalier & A. Tsybakov (2002). Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields* **123**, 323-354.

[14] F. Comte & Y. Rozenholc (2001). Adaptive estimation of mean and volatility functions in (auto-) regressive models. *Stochastic Processes and Their Applications* **97**, 111-145.

[15] Comte, F. and Rozenholc, Y. (2004) A new algorithm for fixed design regression and denoising. *A*nn. Inst. Statist. Math. **56** 449-473.

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

[17] L. Devroye (1986). Non-Uniform Random Variable Generation. Springer-Verlag, New-York.

[18] L. Devroye (1989). Consistent deconvolution in density estimation. *The Canadian Journal of Statistics* **17**, 235-239.

[19] Delaigle, A. and I. Gijbels (2004) Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics and Data Analysis* **45**, 249-267.

[20] D.L. Donoho & I.M. Johnstone (1998). Minimax estimation with wavelet shrinkage. *The Annals of Statistics* **26**, 879-921.

[21] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian & D. Picard (1996). Density estimation by wavelet thresholding. *The Annals of Statistics* **24**, 508-539.

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

[23] S. Efromovich (1997). Density estimation in the case of supersmooth measurement errors. *Journal of the American Statistical Association* **92**, 526-535.

[24] J. Fan (1991a). On the optimal rates of convergence for nonparametric deconvolution problem. *The Annals of Statistics* **19**, 1257-1272.

[25] J. Fan (1991b). Asymptotic normality for deconvolution kernel estimators, *Sankhya Series A* **53**, 97-110.

[26] J. Fan (1991c). Global behavior of deconvolution kernel estimates. *Statistica Sinica* **1**, 541-551.

[27] J. Fan (1993). Adaptively local one-dimensional sub-problems with application to a deconvolution problem, *The Annals of Statistics* **21**, 600-610.

[28] J. Fan & J.-Y. Koo (2002). Wavelet deconvolution. *IEEE Transactions on Information Theory* **48**, 734-747.

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

[30] C. Hesse (1999) Data-driven deconvolution. *Journal of Nonparametric Statistics* **10**, 343-373.

[31] I.A. Ibragimov & R.Z. Hasminskii (1983). Estimation of distribution density. *Journal of Soviet Mathematics* **21**, 40-57.

[32] I.M. Johnstone (1999). Wavelet shrinkage for correlated data and inverse problems : adaptivity results. *Statistica Sinica* **9**, 51-83.

[33] J.-Y Koo (1999). Logspline deconvolution in Besov space. *Scandinavian Journal of Statistics* **26**, 73-86.

[34] M.C. Liu & R.L. Taylor (1989). A consistent nonparametric density estimator for the deconvolution problem. *The Canadian Journal of Statistics* **17**, 427-438.

[35] E. Masry (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory* **37**, 1105-1115.

[36] E. Masry (1993a). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Processes and Their Applications* **47**, 53-74.

[37] E. Masry (1993b). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *Journal of Multivariate Analysis* **44**, 47-68.

[38] C. Matias (2002). Semiparametric deconvolution with unknown noise variance. *ESAIM Probability. & Statistics.* **6** 271–292.

[39] A. Meister (2004). On the effect of missspecifying the error density in a deconvolution problem. The canadian Journal of Statistics **32**(4), 439-449.

[40] Y. Meyer (1990). Ondelettes et opérateurs, Tome I, Hermann.

[41] M. Pensky (2002). Density deconvolution based on wavelets with bounded supports. *Statistics and Probability Letters* **56**, 261-269.

[42] M. Pensky and B. Vidakovic (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics* **27**(6), 2033-2053.

[43] L. Stefansky (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters* **9**, 229-235.

[44] L. Stefansky & R.J. Carroll (1990). Deconvolution kernel density estimators. *Statistics* **21**, 169-184.

[45] M. Talagrand (1996). New concentration inequalities in product spaces. *Inventiones Mathematicae* **126**, 505-563.

[46] R.L. Taylor & H.M. Zhang (1990). On strongly consistent non-parametric density estimator for deconvolution problem. *Communications in Statistics. A. Theory and Methods* **19**, 3325-3342.

[47] A.B. Tsybakov (2000). On the best rate of adaptive estimation in some inverse problems. *Comptes-Rendus de l'Académie des Sciences, Paris, Série I Mathématiques*, **330**, 835-840.

[48] G. Viennet (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probability Theory and Related Fields* **107**, 467-492.

[49] C.H. Zhang (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics* **18**, 806-831.