DATA DRIVEN DENSITY ESTIMATION IN PRESENCE OF ADDITIVE NOISE WITH UNKNOWN DISTRIBUTION.

F. COMTE*,(1) AND C. LACOUR(2)

Abstract. We study the following model: $Y = X + \varepsilon$. We assume that we have at our disposal i.i.d. observations $Y_1, \dots, Y_n$ and $\varepsilon_{-1}, \dots, \varepsilon_{-M}$. The $(X_j)_{1\le j\le n}$ are i.i.d. with density f, independent of the $(\varepsilon_j)_{1\le j\le n}$, i.i.d. with density f_ε. The aim of the paper is to estimate f without knowing f_ε. We first define an estimator, for which we provide bounds for the integrated L^2-risk. We consider ordinary smooth and supersmooth noise ε with regard to ordinary smooth and supersmooth densities f. Then we present an adaptive estimator of the density of f. This estimator is obtained by penalization of a projection contrast, and yields to model selection. Lastly, we present simulation experiments to illustrate the good performances of our estimator and study from the empirical point of view the importance of theoretical constraints.

Keywords. Adaptive estimation. Deconvolution. Density estimation. Mean square risk. Minimax rates. Nonparametric methods.

*,(1): F. Comte, corresponding author.
MAP5, Université Paris Descartes,
45 rue des Saints-Pères, 75006 PARIS, FRANCE,
email: fabienne.comte@parisdescartes.fr

(2): C. Lacour.
Laboratoire de Mathématiques, Université Paris-Sud 11,
91405 Orsay cedex, FRANCE,
email: claire.lacour@math.u-psud.fr

1. Introduction

Let us consider the following model:

$$Y_j = X_j + \varepsilon_j \qquad j = 1, \dots, n \tag{1}$$

where $(X_j)_{1\le j\le n}$ and $(\varepsilon_j)_{1\le j\le n}$ are independent sequences of i.i.d. variables. We denote by f the density of X_j and by f_ε the density of ε_j. The aim is to estimate f when only $Y_1, \dots, Y_n$ are observed. In the classical convolution model, f_ε is assumed to be known, and this is often considered as an important drawback of this simple model. Indeed, in most practical applications, the distribution of the errors cannot be perfectly known. Sometimes, this problem can be circumvented by repeated observations of the same variable of interest, each time with an independent error. This is the model of panel data, see for example Li and Vuong (1998), Delaigle et al. (2008), or Neumann (2007) and references therein.

Date: January 26, 2011.

However, there are also many application fields where it is not possible to do repeated measurements of the *same* variable. In that case, information about the error distribution can be drawn from an additional experiment: a training set is used by experimenters to estimate the noise distribution. Think of ε as a measurement error due to the measuring device; then preliminary calibration measures can be obtained in the absence of any signal X (this is often called the instrument line shape of the measuring device). Mathematically, this means that the knowledge of f_ε can be replaced by observations $\varepsilon_{-1}, \dots, \varepsilon_{-M}$, a noise sample with distribution f_ε, independent of $(Y_1, \dots, Y_n)$. It has the advantage that only one measuring device is needed, instead of two or more for repeated measurement strategies. Note that the availability of two distinct samples makes the problem identifiable.

Actually, this is a natural method used by practitioners or theoreticians. See for example Kerkyacharian et al. (2010) who study spherical deconvolution and replace the characteristic function of the noise by its empirical version, since the noise is unknown in the context of astrophysics. One of the most typical domains where a preliminary estimation of the measurement error is done is spectrometry, or spectro-fluorimetry, but let us detail an example in microscopy. Odiachi and Prieve (2004) study the effect of additive noise in Total Internal Reflection Microscopy (TIRM) experiments. This is an optical technique for monitoring Brownian fluctuations in separation between a single microscopic sphere and a flat plate in aqueous medium. It is used to detect extremely weak interactions that cannot be measured by mechanical techniques. The elevation of the sphere can be detected by measuring the light scattered by the sphere when illuminated by an evanescent wave. The data of scattering intensity I_s are corrupted by a background noise I_b so that only a noisy signal $I_b + I_s$ is observed. Thus a deconvolution is needed to retrieve the distribution of actual scattering intensity, which is the variable of interest. The authors explain that the histogram of the noise is obtained from the measurements when there is no sphere in the observation window (whereas the noisy data are the observed intensities with the sphere in the window).

In the present paper , we study Model (1), completed with a noise sample $\varepsilon_{-1}, \dots, \varepsilon_{-M}$ with distribution f_ε, independent of $(Y_1, \dots, Y_n)$. On the one hand, there exists a huge literature concerning the estimation of f when f_ε is known: see Carroll and Hall (1988), Devroye (1989), Fan (1991), Liu and Taylor (1989), Masry (1991), Stefanski and Carroll (1990), Zhang (1990), Hesse (1999), Cator (2001), Delaigle and Gijbels (2004) for mainly kernel methods, Koo (1999) for a spline method, Pensky and Vidakovic (1999) and Fan and Koo (2002) for wavelet strategies, Fan (1991), Butucea (2004) and Butucea and Tsybakov (2007) for studies of optimality in the minimax sense of the rates, Comte et al. (2006) for adaptive projection strategies. See also the specific studies of Efromovich (1997) (particular strategy in supersmooth case) and Meister (2004) (on the effect of noise misspecification). On the other hand, a few authors have studied the exact problem which is considered in this paper, but only for particular type of smoothness for f_ε or f or other type of risks. We provide a general study of the mean integrated squared error (MISE) which substantially generalizes existing results. Then, as the estimators depend on a bandwidth-type parameter, we propose a model selection strategy: we consider – from both theoretical and practical point of view – the difficult problem of automatic selection of this quantity. In other words, we explain how to select a relevant estimator in the collection. The study of a completely data-driven procedure in this context is new and contains true interesting ideas:

here, the penalty is not only data driven, but the collection of models is also randomly selected on the basis of the observations.

Let us describe what has been done on the subject. In the deconvolution context, four cases must be considered, depending on the smoothness of f and f_ε: roughly speaking, the functions are called ordinary smooth when the rate of decay of their Fourier Transform near infinity is polynomial, and super smooth if it is exponential. The first work of the subject is to be found in Diggle and Hall (1993), who study the case of ordinary smooth noise density and distribution function f, under $M \geq n$, and present interesting heuristics. Next, our work is related to the paper of Neumann (1997), since our estimator is similar to his and we borrow a useful Lemma from his study. Neumann (1997) also considers the case of both ordinary smooth noise and distribution function. He does not perform any bandwidth selection, but he proves the minimax optimality of the bound he obtains in the case he considers, and we shall refer to this lower bound. Lastly, Johannes (2009) recently studied the density deconvolution with unknown (but observed) noise and is interested in the relation between M and n. His estimator and his approach are very interesting and rather different from ours, his estimator depends on two bandwidth-type parameters, which, if relevantly chosen, lead to rates that are the same as in our work. But the data-driven selection of these bandwidths is not done.

It is worth mentioning that the problem of adaptation which is studied here is of non linear type and thus difficult to solve, in spite of the apparent simplicity of the estimator of the Fourier Transform of f_ε. See the study of similar questions in the context of inverse problem with error in the operator in Hoffmann and Reiss (2008) or Cavalier and Raimondo (2007).

Here is the plan of the present paper. In Section 2, we give the notations and define the estimator, first directly, and then as a projection-type estimator. We study in Section 3 the integrated mean square risk (MISE) of one estimator, which allows us to build general tables for the rates. Then, we study the link between M and n if one wants to preserve the rate found in the case where f_ε^* is known. Such a complete panorama is new in this setting. Next, we define and study in Section 4 a model selection estimator by proposing a penalization device. A general integrated risk bound for the resulting estimator is given. The estimator is studied through simulation experiments in Section 5, and its performances are compared with Neumann (1997)'s and Johannes (2009)'s ones. The influence of the size M of the noise sample is studied as well as the importance of some other theoretical constraints on the size of the collection of models. We can check there that the estimator is easy to implement and works very well. Most proofs are gathered in Section 6.

2. Estimation procedure

2.1. **Notations.** For two real numbers a and b, we denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For z a complex number, $\bar{z}$ denotes its conjugate and $|z|$ its modulus. For functions $s, t : \mathbb{R} \mapsto \mathbb{R}$ belonging to $\mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R})$, we denote by $\|t\|$ the $\mathbb{L}^2$ norm of t, that is $\|t\|^2 = \int_{\mathbb{R}} |t(x)|^2 dx$, and by $\langle s, t\rangle$ the scalar product: $\langle s, t\rangle = \int_{\mathbb{R}} s(x)\overline{t(x)}dx$. The Fourier transform t^* of t is defined by

$$t^*(u) = \int e^{-ixu} t(x) dx.$$

Note that, if t^* also belongs to $\mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R})$, then the function t is the inverse Fourier transform of t^* and can be written $t(x) = 1/(2\pi)\int e^{ixu}t^*(u)du$. Finally, the convolution product is defined by $(t * s)(x) = \int t(x-y)s(y)dy$.

2.2. **Basic definition of the adaptive estimator.** It easily follows from Model (1) and independence assumptions that, if f_Y denotes the common density of the Y_j's, then $f_Y = f * f_\varepsilon$ and thus $f_Y^* = f^* f_\varepsilon^*$. Note that this basic equality can be obtained for a noise with discrete distribution, and the whole method can be generalized to that case.

Therefore, under the classical assumption:

(A1) $\forall x \in \mathbb{R},\ f_\varepsilon^*(x) \neq 0$,

the equality $f^* = f_Y^*/f_\varepsilon^*$ yields an estimator of f^* by considering the following estimate of f_Y^*:

$$\hat{f_Y^*}(u) = \frac{1}{n}\sum_{j=1}^n e^{-iuY_j}.$$

Indeed, if f_ε^* is known, we can use the following estimate of f^*: $\hat{f_Y^*}/f_\varepsilon^*$. Then, we should use inverse Fourier transform to get an estimate of f. As $1/f_\varepsilon^*$ is in general not integrable (think of a Gaussian density for instance), this inverse Fourier transform does not exist, and a cutoff is used. The final estimator for known f_ε can thus be written: $(2\pi)^{-1}\int_{|u|\leq \pi m} e^{iux}\hat{f_Y^*}(u)/f_\varepsilon^*(u)du$. This estimator is classical in the sense that it corresponds both to a kernel estimator built with the sinc kernel (see Butucea (2004)) or to a projection type estimator as in Comte et al. (2006), as will be showed below.

Now, f_ε^* is unknown and we have to estimate it. Therefore, we use the preliminary noise sample and we define the natural estimator of f_ε^*

$$\hat{f_\varepsilon^*}(x) = \frac{1}{M}\sum_{j=1}^M e^{-ix\varepsilon_{-j}}.$$

Next, we introduce as in Neumann (1997) the truncated estimator:

$$\frac{1}{\tilde{f_\varepsilon^*}(x)} = \frac{\mathbb{1}_{\{|\hat{f_\varepsilon^*}(x)|\geq M^{-1/2}\}}}{\hat{f_\varepsilon^*}(x)} = \begin{cases} \dfrac{1}{\hat{f_\varepsilon^*}(x)} & \text{if } |\hat{f_\varepsilon^*}(x)| \geq M^{-1/2} \\ 0 & \text{otherwise.}\end{cases} \tag{2}$$

Then we can consider

$$\hat{f}_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\frac{\hat{f_Y^*}(u)}{\tilde{f_\varepsilon^*}(u)}du. \tag{3}$$

Note that this estimator is such that $(\hat{f}_m)^* = (\hat{f_Y^*}/\tilde{f_\varepsilon^*})\mathbb{1}_{[-\pi m,\pi m]}$.

Lastly, we need a strategy to select the parameter m. Indeed, m plays a bandwidth-type role, and has to be relevantly selected to lead to an adequate bias variance compromise. Let

$$f_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} f^*(u)e^{iux}du.$$

Then we have $f_m^*(u) = f^*(u)\mathbb{1}_{[-\pi m,\pi m]}(u)$, and

$$\begin{aligned}\|f-\hat f_m\|^2 &= \frac{1}{2\pi}\|f^*-\hat f_m^*\|^2 = \frac{1}{2\pi}(\|f^*-f_m^*\|^2+\|f_m^*-\hat f_m^*\|^2)\\ &= \frac{1}{2\pi}(\|f^*\|^2-\|f_m^*\|^2+\|f_m^*-\hat f_m^*\|^2)\\ &= \|f\|^2-\|f_m\|^2+\|f_m-\hat f_m\|^2. \qquad (4)\end{aligned}$$

The bias term is $\|f\|^2-\|f_m\|^2$ and can thus be estimated by $-\|\hat f_m\|^2$, because the constant $\|f\|^2$ plays no role in the compromise. The variance term has the order of $\mathbb{E}(\|f_m-\hat f_m\|^2)$. Therefore, the estimation procedure is completed as follows. We choose the best estimator among the collection $(\hat f_m)_{m\in\mathcal{M}_n}$ where $\mathcal{M}_n\subset\{1,\dots,n\}$ is the set of all considered indices by setting:

$$\hat m = \arg\min_{m\in\mathcal{M}_n}\{-\|\hat f_m\|^2+\mathrm{pen}(m)\} \qquad (5)$$

where pen is a penalty term to be specified later, which has the variance order.

2.3. **Equivalent definition of the adaptive estimator.** The following view of the estimator (5) is useful in the proofs and for practical implementation.

2.3.1. *Projection spaces.* Let us consider the function

$$\varphi(x) = \sin(\pi x)/(\pi x)$$

and, for m in $\mathbb{N}^*$, j in $\mathbb{Z}$, $\varphi_{m,j}(x) = \sqrt{m}\varphi(mx-j)$. As $\varphi^*(x) = \mathbb{1}_{[-\pi,\pi]}(x)$, we have, as a key property of the functions $\varphi_{m,j}$, that $\varphi_{m,j}^*(x) = e^{-ixj/m}\mathbb{1}_{[-\pi m,\pi m]}(x)/\sqrt{m}$. It is proven that $\{\varphi_{m,j}\}_{j\in\mathbb{Z}}$ is an orthonormal basis of the space of integrable functions having a Fourier transform with compact support included into $[-\pi m,\pi m]$. Note that m can be chosen in other sets than $\mathbb{N}^*$, and thinner grids may be useful in practice.

In the sequel, we use the following notation:

$$S_m = \mathrm{Span}\{\varphi_{m,j}\}_{j\in\mathbb{Z}}.$$

We know (see Comte et al. (2006)) that the orthogonal projection of a function g in $(\mathbb{L}^1\cap\mathbb{L}^2)(\mathbb{R})$ on S_m, denoted by g_m, is such that $g_m^* = g^*\mathbb{1}_{[-\pi m,\pi m]}$. With Fourier inverse formula, we get:

$$g_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}g^*(u)du. \qquad (6)$$

2.3.2. *Adaptive estimation of f.* Let us define the following contrast

$$\gamma_n(t) = \frac{1}{n}\sum_{j=1}^n[\|t\|^2-2\tilde v_t(Y_j)] \quad \text{with} \quad \tilde v_t^*(u) = \frac{t^*(u)}{\tilde f_\varepsilon^*(-u)}. \qquad (7)$$

Clearly, $\tilde v_t$ is an approximation of

$$v_t^*(u) = \frac{t^*(u)}{f_\varepsilon^*(-u)}. \qquad (8)$$

It is easy to see that, for $t\in S_m$, $\gamma_n(t) = \|t\|^2-2\langle t,\hat f_m\rangle$ so that

$$\hat f_m = \arg\min_{t\in S_m}\gamma_n(t). \qquad (9)$$

But the definition using (7) has the advantage to be written without specifying m.

Moreover, this gives another formula for $\hat{f}_m$:

$$\hat{f}_m = \sum_{l\in\mathbb{Z}} \hat{a}_{m,l}\varphi_{m,l} \quad \text{with} \quad \hat{a}_{m,l} = \frac{1}{n}\sum_{j=1}^{n} \tilde{v}_{\varphi_{m,l}}(Y_j). \tag{10}$$

Actually, we use the strategy given by (10) in practice, because it allows us to use fast algorithms as Inverse Fast Fourier Transform (IFFT). Thus, we use in fact $\hat{f}_m = \sum_{|l|\le K_n} \hat{a}_{ml}\varphi_{m,l}$ because we can estimate only a finite number of coefficients. If K_n is large enough, it does not change the rate of convergence since the additional terms can be made negligible. For the sake of simplicity, we let the sum over $\mathbb{Z}$. For an example of detailed study of theoretical truncation see Comte et al. (2006).

Finally, as obviously $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2$, (5) is equivalent to select the model which minimizes the following penalized criterion:

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n}\{\gamma_n(\hat{f}_m) + \text{pen}(m)\} \tag{11}$$

where pen is the same penalty term as in (5) and will be specified later. Our aim is to study $\hat{f}_{\hat{m}}$ as final estimator of the density.

3. Bound on the L^2 risk

3.1. **Notations.** Let us recall first the following key lemma, proved in Neumann (1997) for $p = 1$:

Lemma 1. *Let $p \ge 1$ be an integer and*

$$R(x) = \left(\frac{1}{\tilde{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)}\right).$$

Then there exists a positive constant C_p such that

$$\mathbb{E}|[R(x)|^{2p}] \le C_p \left(\frac{1}{|f_\varepsilon^*(x)|^{2p}} \wedge \frac{M^{-p}}{|f_\varepsilon^*(x)|^{4p}}\right).$$

The extension from $p = 1$ to any integer p is straightforward and therefore the proof is omitted.

We introduce the notations

$$\Delta(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} du \quad \text{and} \quad \Delta_f(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|f_\varepsilon^*(u)|^2} du. \tag{12}$$

As we shall see, these quantities are involved in the bounds on the variance of our estimators.

3.2. **Bound on the MISE.** Let us study the integrated mean square risk. By (4), we have

$$\|f - \hat{f}_m\|^2 = \|f - f_m\|^2 + \|f_m - \hat{f}_m\|^2. \tag{13}$$

Moreover, writing $\hat{f}_m - f_m$ according to (6) and (3) and applying the Parseval formula, we obtain

$$\|f_m - \hat{f}_m\|^2 = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \left| \frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^2 du.$$

It follows that

$$\|f_m - \hat{f}_m\|^2 \le \frac{1}{\pi}\int_{-\pi m}^{\pi m} |\hat{f}_Y^*(u)|^2 |R(u)|^2 du + \frac{1}{\pi}\int_{-\pi m}^{\pi m} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^2}{|f_\varepsilon^*(u)|^2} du. \tag{14}$$

The last term of the right-hand-side of (14) is the usual term that is found when f_ε^* is known, and the first one is specific to the present framework. Using this decomposition, we can prove the following result:

Proposition 1. *Consider model (1) under* **(A1)***, then $\hat{f}_m$ defined by (9) satisfies:*

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \le \|f_m - f\|^2 + 4C_1\frac{\Delta(m)}{n} + (4C_1 + 2)\frac{\Delta_f(m)}{M} \tag{15}$$

where C_1 is the numerical constant defined in Lemma 1.

The first two terms in the right-hand-side of (15) are the usual terms when f_ε^* is known (see Comte et al. (2006)) and correspond to the bias and the variance term. The last term $\Delta_f(m)/M$ is due to the estimation of f_ε^*.

Remark. As $|f^*| \le 1$, we have $\Delta_f(m) \le \Delta(m)$. It follows that for any $M \ge n$, then $\mathbb{E}\|\hat{f}_m - f\|^2 \le \|f_m - f\|^2 + C\Delta(m)/n$ and we recover the usual risk bound for deconvolution estimation when f_ε^* is known. Therefore, in all cases, the condition $M \ge n$ ensures that the rate of the estimator is the same as when f_ε^* was known.

3.3. **Discussion about the resulting rates.** Assumption **(A1)** is generally strengthened by a parametric description of the rate of decrease of f_ε^* written as follows:

(A2) There exist $s \ge 0, b > 0$, $\gamma \in \mathbb{R}$ ($\gamma > 0$ if $s = 0$) and $k_0, k_1 > 0$ such that

$$k_0(x^2+1)^{-\gamma/2}\exp(-b|x|^s) \le |f_\varepsilon^*(x)| \le k_1(x^2+1)^{-\gamma/2}\exp(-b|x|^s)$$

Moreover, the desnity f to estimate generally belongs to the following type of smoothness spaces:

$$\mathcal{A}_{\delta,r,a}(l) = \{f \text{ density on } \mathbb{R} \text{ and } \int |f^*(x)|^2(x^2+1)^\delta \exp(2a|x|^r)dx \le l\} \tag{16}$$
$$\text{with } r \ge 0, a > 0,\ \ \delta \in \mathbb{R} \ \text{ and } \delta > 1/2 \text{ if } r = 0,\ l > 0.$$

When $r > 0$, the function f is called supersmooth, and ordinary smooth otherwise. In the same way, the noise distribution is called ordinary smooth if $s = 0$ and supersmooth otherwise. The spaces of ordinary smooth functions correspond to classic Sobolev classes, while supersmooth functions are infinitely differentiable. It includes for example normal ($r = 2$) and Cauchy ($r = 1$) densities. We take the convention $(a, r) = (0, 0)$ if $a = 0$ or $r = 0$ and $(b, s) = (0, 0)$ if $b = 0$ or $s = 0$.

In this section, we assume that f_ε^* satisfies Assumption **(A2)**, with parameters γ, b, s and that the unknown function f belongs to a smoothness class $\mathcal{A}_{\delta,r,a}(l)$ given by (16). It is then possible to evaluate orders for the different terms involved in the bound (15).

Since $f_m^* = f^* \mathbb{1}_{[-\pi m, \pi m]}$, the bias term can be bounded in the following way

$$\|f - f_m\|^2 = \frac{1}{2\pi} \int_{([-\pi m, \pi m])^c} |f^*(u)|^2 du \le \frac{l}{2\pi} ((\pi m)^2 + 1)^{-\delta} e^{-2a(\pi m)^r}$$

The other terms are evaluated in the following lemma proved in Section 6. We use the notation $g_1(m) \lesssim g_2(m)$ if there exists a constant $0 < C < +\infty$ such that $\forall m, \;\; g_1(m) \le C g_2(m)$, and the notation $g_1(m) \asymp g_2(m)$ if $g_1(m) \lesssim g_2(m)$ and $g_2(m) \lesssim g_1(m)$.

Lemma 2. *If f_ε^* satisfies Assumption* **(A2)** *then*
(1) $\Delta(m) \asymp (\pi m)^{2\gamma+1-s} e^{2b(\pi m)^s}$,
(2) $\Delta_f(m) \lesssim (\pi m)^{(1+2\gamma-s)\wedge 2(\gamma-\delta)_+} e^{2b(\pi m)^s} \mathbb{1}_{\{s>r\}} + (\pi m)^{2(\gamma-\delta)_+} e^{2(b-a)(\pi m)^s} \mathbb{1}_{\{r=s, b\ge a\}}$
$+\mathbb{1}_{\{r>s\}\cup\{r=s,b<a\}}$.

Now distinguishing the different cases, we can state the following propositions.

Proposition 2. *Assume that* **(A2)** *holds and that $f \in \mathcal{A}_{\delta,r,a}(l)$ given by (16). If $s = 0$ (ordinary smooth noise) or $s > 0$ (supersmooth noise), and for $r = 0$ (ordinary smooth function f), then*

$$\mathbb{E}\|\hat{f}_m - f\|^2 \le C_0 m^{-2\delta} + C \frac{m^{2\gamma+1-s} e^{2b(\pi m)^s}}{n} + C' \frac{m^{(1+2\gamma-s)\wedge 2(\gamma-\delta)_+} e^{2b(\pi m)^s}}{M},$$

where C_0, C and C' are constants which do not depend on M nor n.

Case $s = 0$. It is known from Fan (1991), that the optimal minimax rate when f_ε^* is known is $n^{\frac{-2\delta}{2\gamma+2\delta+1}}$. It is preserved with unknown f_ε^* as soon as $M \ge n^{\frac{2(\gamma\vee\delta)}{2\gamma+2\delta+1}}$. This bound is tighter than $M \ge n$.
Now, choose $m_0 = Int[n^{\frac{1}{2\gamma+2\delta+1}} \wedge M^{\frac{1}{2(\gamma\vee\delta)}}]$ where $Int[.]$ denotes the integer part. We obtain

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 = O\left(n^{-\frac{2\delta}{2\gamma+2\delta+1}} + M^{-(1\wedge(\delta/\gamma))}\right).$$

This is the lower bound proved by Neumann (1997), and thus the rate of our estimator is the optimal rate.
Case $s > 0$. For known f_ε^*, Fan (1991) proves that the optimal rate is of order $(\log n)^{-\frac{2\delta}{s}}$. It is preserved here with unknown f_ε^* as soon as $M \ge n(\log n)^{-\frac{s+[2(\delta\wedge\gamma)+1-s]_+}{s}}$.

Choose $m_0 = Int[(1/\pi)\left(\frac{1}{2b}\log[n(\log n)^{-\frac{2\delta+2\gamma+1}{s}} \wedge M(\log M)^{-\frac{2\delta+s+(1+2\gamma-s)\wedge 2(\gamma-\delta)_+}{s}}]\right)^{1/s}]$.
This yields

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 = O\left((\log n)^{\frac{-2\delta}{s}} + (\log M)^{\frac{-2\delta}{s}}\right).$$

Proposition 3. *Assume that* **(A2)** *holds and that $f \in \mathcal{A}_{\delta,r,a}(l)$ given by (16). If $s = 0$ (ordinary smooth noise) and $r > 0$ (supersmooth function f), then*

$$\mathbb{E}\|\hat{f}_m - f\|^2 \le C_0 m^{-2\delta} e^{-2a(\pi m)^r} + C \frac{m^{2\gamma+1}}{n} + \frac{C'}{M},$$

where C_0, C and C' are constants which do not depend on M nor n.

The optimal rate in this case is studied by Butucea (2004) when f_ε^* is known and is of order $(\log n)^{\frac{2\gamma+1}{r}}/n$. It is preserved even when f_ε^* is estimated, if the sample size for estimating it, M, is such that $M \geq n(\log n)^{-\frac{2\gamma+1}{r}}$.

Let us choose now $m_0 = Int[(1/\pi)\left(\frac{1}{2a}\log[n(\log n)^{\frac{r-2\delta-2\gamma-1}{r}} \wedge M(\log M)^{\frac{-2\delta}{r}}]\right)^{1/r}]$. We get

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 = O\left(\frac{(\log n)^{\frac{2\gamma+1}{r}}}{n} + \frac{1}{M}\right).$$

These three cases are summarized in Table 1.

	$s = 0$	$s > 0$
$r = 0$	$n^{-\frac{2\delta}{2\delta+2\gamma+1}} + M^{-[1\wedge(\frac{\delta}{\gamma})]}$	$(\log n)^{-\frac{2\delta}{s}} + (\log M)^{\frac{-2\delta}{s}}$
$r > 0$	$\frac{(\log n)^{\frac{2\gamma+1}{r}}}{n} + \frac{1}{M}$	see the discussion below.

TABLE 1. Rates of convergence for the MISE.

The last case, when both functions are supersmooth, is much more tedious, in particular if one wants to evaluate the rates. In the case of known error distribution, these are implicitly given in Butucea and Tsybakov (2007), who also study optimality; explicit formulae are available in Lacour (2006), see Theorem 3.1 therein.

Proposition 4. *Assume that* **(A2)** *holds and that* $f \in \mathcal{A}_{\delta,r,a}(l)$ *given by (16). If* $s > 0$ *(supersmooth noise) and* $r > 0$ *(supersmooth function* f*), then*

$$\mathbb{E}\|\hat{f}_m - f\|^2 \leq C_0 m^{-2\delta} e^{-2a(\pi m)^r} + C\frac{m^{2\gamma+1-s}e^{2b(\pi m)^s}}{n} + C'\frac{\Delta_f(m)}{M},$$

where C_0*,* C *and* C' *are constants which do not depend on* M *nor* n*.*

The three following cases can be deduced from Theorem 3.1 in Lacour (2006).

Case $r = s$. We define $\xi = [2b\delta - a(2\gamma+1-s)]/[(a+b)s]$ and $\omega = [2(b-a)\delta - 2a(\gamma-\delta)_+]/[bs]$ if $b \geq a$, $\omega = 0$ if $b < a$. We obtain that

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 = O\left(n^{-\frac{a}{a+b}}(\log n)^{-\xi} + M^{-\frac{a}{a\vee b}}(\log M)^{-\omega}\right), \tag{17}$$

for $\pi m_0 = Int[\left(\frac{\log(n)-(\alpha/s)\log\log(n)}{2a+2b}\right)^{1/s} \wedge \left(\frac{\log(M)-(\beta/s)\log\log(M)}{2(a\vee b)}\right)^{1/s}]$ where $\alpha = 2\delta + 2\gamma + 1 - s$ and $\beta = 2\delta + 2(\gamma-\delta)_+\mathbb{1}_{b\geq a}$.

Case $r < s$. We define $k = \lceil (s/r-1)^{-1}\rceil - 1$, where $\lceil . \rceil$ is the ceiling function (i.e. $\lceil x\rceil$ is the smallest integer larger than or equal to x). There exist coefficients b_i recursively defined (see Lacour (2006)) and a choice m_0 such that

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 = O\Big((\log n)^{-2\delta/s}\exp[\sum_{i=0}^{k} b_i(\log n)^{(i+1)r/s-i}]$$

$$+(\log M)^{-2\delta/s}\exp[\sum_{i=0}^{k} b_i(\log M)^{(i+1)r/s-i}]\Big) \tag{18}$$

Case $r > s$. We define $k = \lceil (r/s-1)^{-1}\rceil - 1$. There exist coefficients d_i recursively defined and a choice m_0 such that

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 = O\Big(\frac{(\log n)^{(1+2\gamma-s)/r}}{n}\exp[-\sum_{i=0}^{k} d_i(\log n)^{(i+1)s/r-i}] + \frac{1}{M}\Big) \tag{19}$$

3.4. **Lower bounds for the additional problem of estimating f_ε.** As mentioned above, Neumann (1997) only studied one particular case from the lower bound point of view. But his proof (for the additional problem of estimating f_ε) can be checked to be suitable in other cases. The following proposition establishes the optimality of our estimator in the case where f is smoother than f_ε and $r \leq 1$.

Proposition 5. *Let*

$$\mathcal{F}_{\gamma,b,s} = \{f_\varepsilon \textit{ density such that there exist } k_0, k_1 > 0 \textit{ such that}$$
$$\forall x \in \mathbb{R} \quad k_0 \leq |f_\varepsilon^*(x)|(x^2+1)^{\gamma/2}\exp(b|x|^s) \leq k_1\}$$

If $r = s = 0$ *and* $\gamma < \delta - 1/2$, *or if* $0 \leq s < r \leq 1$ *then*

$$\inf_{\hat{f}} \sup_{f\in\mathcal{A}_{\delta,a,r}(l), f_\varepsilon\in\mathcal{F}_{\gamma,b,s}} \mathbb{E}\|\hat{f} - f\|_2^2 \geq CM^{-1}$$

where the infimum is taken over all estimators $\hat{f}$ of f based on the observations $Y_1, \ldots, Y_n$.

Proof of Proposition 5. The proof of the lower bound (for the additional problem of estimating f_ε) given by Neumann (1997) can be used here. Thus, if $r = s = 0$ and $\gamma < \delta - 1/2$, we obtain a lower bound CM^{-1} which proves the optimality of our estimator in this case.

In addition, this result can be generalized to a supersmooth noise distribution if $s \leq 1$. Indeed, such densities verify the property (3.1) used by Neumann (1997). In the same way, an extension to supersmooth functions f can be done provided that $r \leq 1$. Thus, if $0 \leq s < r \leq 1$, the rate of convergence M^{-1} of our estimator is optimal for the integrated risk. □

4. Model selection

The above study shows that the choice of m is both crucial and difficult. Thus, we provide a data driven strategy to perform automatically this choice. We assume that we are able to manage with M larger than n: this means that we need a careful calibration step, but this step is done once for all. More precisely, we assume in the following that there exists $\epsilon > 0$ such that $M \geq n^{1+\epsilon}$. This preliminary ε-sample will enable us to provide a density estimator for any new n-sample of the Y_i's.

Consequently, our aim is to preserve here the rate corresponding to the case of an n-sample of observations Y_i when f_ε^* is known. We consider the estimator $\hat{f}_{\hat{m}}$ defined by (11) where we have to define the penalty pen(.).

4.1. **A theoretical estimator.** We will work under Assumption **(A2)** and, concerning the collection of models $\mathcal{M}_n = \{1, 2, \dots, m_n\}$, we define m_n by

$$m_n = \arg\max\{m \in \{1, \dots, n\}, 1 \le \Delta(m)/n \le 2\}.$$

If it exists, we simply take m_n such that $\Delta(m_n) = 2n$.

We can prove the preliminary result given in Theorem 1.

Theorem 1. *Assume that assumption* **(A2)** *is fulfilled, take $M \ge n^{1+\epsilon}$ for $\epsilon > 0$ and consider the estimator $\hat{f}_{\hat{m}}$ defined by (9) and (11) with*

$$\text{pen}(m) = K_0 \left(\frac{\log(\Delta(m))}{\log(m+1)} \right)^2 \frac{\Delta(m)}{n}, \tag{20}$$

where K_0 is a numerical constant.

Then there exists $C > 0$ such that

$$\mathbb{E}\|\hat{f}_{\hat{m}} - f\|^2 \le 4 \inf_{m \in \mathcal{M}_n} \{\|f_m - f\|^2 + \text{pen}(m)\} + \frac{C}{n} \tag{21}$$

where f_m is the orthogonal projection of f on S_m.

This result requires several comments.

(1) First, let us examine the penalty and its order, which contains two terms: a main part $\Delta(m)/n$ and an auxiliary one, $(\log(\Delta(m))/\log(m+1))^2$. The variance has order $\Delta(m)/n$. If the errors are ordinary smooth, the term $(\log(\Delta(m))/\log(m+1))^2$ is a constant and the penalty has exactly the order of the variance (see the results in Lemma 2). In the super-smooth case, this term adds a multiplicative factor $m^{2s}/\log^2(m+1)$.

(2) The result of Theorem 1 is an oracle inequality which states that the estimator $\hat{f}_{\hat{m}}$ makes the compromise between the squared bias term $\|f - f_m\|^2$ and the penalty. If the penalty has exactly the order of the variance $\Delta(m)/m$, then the optimal rate is reached. This occurs at least $-$ but not only $-$ in the ordinary smooth case. Otherwise, a loss may occur. The discussion below shows that, even if the penalty is larger than the variance, the rate can be preserved. Moreover, it follows from results in Butucea and Tsybakov (2007) that a loss may be unavoidable in the adaptive procedure; in that case, the rate is called adaptive optimal.

(3) We can note that the penalty (20) generalizes the one obtained in Comte et al. (2006) with known f_ε^*. In Comte et al. (2006), f_ε is known and fulfills **(A2)** and the oracle inequality is obtained for the estimator with a penalty proportional to

$$m^\omega \frac{\Delta(m)}{n}, \quad \text{with } \omega = \begin{cases} 0 & \text{if} \quad 0 \le s \le 1/3 \\ \frac{3s-1}{2} & \text{if} \quad 1/3 < s \le 1 \\ s & \text{if} \quad s > 1. \end{cases} \tag{22}$$

Clearly, the penalty (20) has the same order for $s = 0$, i.e. in the ordinary smooth case. Otherwise, for $s > 0$, the proposal (20) overestimates the order of (22) for known f_ε. The reason for this sacrifice is the fact that, as we are in a context of unknown f_ε, the term we propose is easier to estimate than the terms in (22).

(4) Let us now discuss the loss which may occur in the rate for $s > 0$. In the case $r = 0, s > 0$, the rate of convergence of the estimate is not affected, because it is clear from Table 1 that the rate does not depend on the powers in the variance term (i.e. it does not depend on γ): the bias term is the one which governs the rate in this case, and a small loss in the variance does not change the rate. In the case $r > 0, s > 0$, if moreover $r < s$, it follows from (18) that the optimal rate is also preserved, for the same reason (the bias term mainly determines the rate). If $r \geq s$, it follows from (17) and (19) that the loss in the rate concerns only the logarithmic terms, which are negligible w.r.t. the rate. Therefore, if a loss in the rate occurs, as price of the adaptive property of the procedure, we know that it is negligible with respect to the rate of convergence of the estimator.

Let us emphasize again here that the interest of the penalty (20) is that the terms required in the supersmooth case are added without requiring the information: are the errors ordinary smooth or supersmooth, and what is the value of s?

4.2. **An effective model selection estimator.** The previous estimator is unrealistic for two reasons:

– The penalty can not be computed because $\Delta(m)$ also depends on f_ε^*.
– The choice of m_n depends on $\Delta(m)$

For the first problem, we define

$$\widetilde{\mathrm{pen}}(m) = K_1 \left(\frac{\log(\hat{\Delta}(m))}{\log(m+1)} \right)^2 \frac{\hat{\Delta}(m)}{n}, \quad \hat{\Delta}(m) = \int_{-\pi m}^{\pi m} |\tilde{f}_\varepsilon^*(x)|^{-2} dx. \tag{23}$$

For the last one, we take

$$\hat{m}_n = \arg\max \left\{ m \in \{1, \ldots, n\}, 1/4 \leq \hat{\Delta}(m)/n \leq 1/2 \right\}.$$

It is useful to note here that, in theory as in practice, the m's need not be integers, and can be chosen in a discrete set with thinner or larger step than 1.

Then the following theorem shows that we have completely solved our problem with a data-driven procedure.

Theorem 2. *Assume that assumption* **(A2)** *is fulfilled and* $M \geq n^{1+\epsilon}$ *for* $\epsilon > 0$*. Consider the estimator* $\tilde{f} = \hat{f}_{\hat{\hat{m}}}$ *defined by (9) and*

$$\hat{\hat{m}} = \underset{m \in \{1, \ldots, \hat{m}_n\}}{\arg\min} \{ \gamma_n(\hat{f}_m) + \widetilde{\mathrm{pen}}(m) \} \tag{24}$$

with $\widetilde{\mathrm{pen}}(m)$ *defined by (23),* K_1 *being a pure numerical constant (*$K_1 = 128$ *would work). Then, for* n *large enough, we have*

$$\mathbb{E}\|\tilde{f} - f\|^2 \leq C_1 \inf_{m \in \mathcal{M}_n} \{ \|f_m - f\|^2 + \mathrm{pen}(m) \} + \frac{C_2}{n}, \tag{25}$$

where C_1 *is a pure numerical constant, and* C_2 *is a constant depending on* f *and* f_ε*.*

It appears that the right-hand-side of Inequality (25) is the same as in Inequality (21) and the comments about (21) following Theorem 1 are therefore valid: the estimator makes an automatic bias-penalty compromise, which leads either to the optimal rate, or to the optimal rate up to a negligible loss. But the left-hand-side term of Inequality (25) involves the new estimator $\hat{f}_{\hat{m}}$ which is now completely data driven, up to the "universal" value of K_1 which is proposed in the next section.

It is worth mentioning that, if we know that the noise is ordinary smooth, then we must take $\widetilde{\text{pen}}(m) = \hat{\Delta}(m)/n$, and the adaptive procedure automatically reaches the optimal rate.

5. Simulations

5.1. **Practical estimation procedure.** Let us describe the estimation procedure as it is implemented. As noticed in (10), for each m, the estimator $\hat{f}_m$ of f can be written

$$\hat{f}_m = \sum_{|l|\le K_n} \hat{a}_{m,l}\varphi_{m,l} \quad \text{with} \quad \hat{a}_{m,l} = \frac{1}{n}\sum_{j=1}^{n} \tilde{v}_{\varphi_{m,l}}(Y_j).$$

To compute the coefficients $\hat{a}_{m,l}$, we use the Inverse Fast Fourier Transform. Indeed, using $\varphi^*_{m,l}(u) = e^{-ilu/m}\mathbb{1}_{[-\pi m,\pi m]}/\sqrt{m}$,

$$\hat{a}_{m,l} \quad = \quad \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{1}{\sqrt{m}} e^{-ilu/m}\frac{\hat{f}^*_Y(-u)}{\tilde{f}^*_\varepsilon(-u)}du = \frac{\sqrt{m}}{2}(-1)^l \int_0^2 e^{il\pi x}\frac{\hat{f}^*_Y}{\tilde{f}^*_\varepsilon}(\pi m(x-1))dx$$

Then, for $l = 0, ..., N-1$, denoting $h_m(x) = (\hat{f}^*_Y/\tilde{f}^*_\varepsilon)(\pi m(x-1))$, $\hat{a}_{m,l}$ can be approximated by

$$\sqrt{m}(-1)^l \frac{1}{N}\sum_{k=0}^{N-1} e^{il\pi\frac{2k}{N}} h_m(\frac{2k}{N}) = \sqrt{m}(-1)^l(\text{IFFT}(H))_l$$

where H is the vector $(h_m(0), h_m(2/N), \ldots, h_m(2(N-1)/N)$. For $l < 0$, it is sufficient to replace $h_m(x)$ by $h_m(-x) = \overline{h_m(x)}$, i.e. H by $\overline{H}$. Following Comte et al. (2006), we choose $K_n = N - 1 = 2^8 - 1$: indeed, a larger K_n does not significantly improve the results.

Thus, to compute $\tilde{f}$, we use the following steps:

- For each m and for each l, compute $\hat{a}_{m,l}$ using function $\hat{f}^*_Y/\tilde{f}^*_\varepsilon$ and IFFT as described above.
- For each m compute $\hat{\Delta}(m)$ and $\gamma_n(\hat{f}_m) + \widetilde{\text{pen}}(m) = -\sum_l |\hat{a}_{m,l}|^2 + \widetilde{\text{pen}}(m)$.
- Compute $\hat{m}_n$.
- Select the $\hat{m}$ which minimizes $\gamma_n(\hat{f}_m) + \widetilde{\text{pen}}(m)$.
- Compute $\tilde{f} = \sum_{|l|\le K_n} \hat{a}_{\hat{m},l}\varphi_{\hat{m},l}$.

Clearly, the use of FFT makes the procedure fast. As in Comte et al. (2007), we consider that m can be fractional. More precisely, we take here

$$\mathcal{M}_n = \left\{ m = \frac{k}{4\pi}, k \in \mathbb{N}^*, k \le \hat{k}_n \right\}$$

where $\hat{k}_n = \arg\max \left\{ k \in \{1, \ldots, n\}, \hat{\Delta}(k/(4\pi)) \le n/2 \right\}$. The penalty is chosen according to Theorem 2. More precisely we take

$$\widetilde{\text{pen}}(m) = 2\left(\frac{\log(\hat{\Delta}(m))}{\log(4\pi m + 1)}\right)^2 \frac{\hat{\Delta}(m)}{n}.$$

It is worth mentioning that, empirically, the procedure is rather robust with respect to the choice of the constant before the penalty and the maximal model $\hat{m}_n (= \hat{k}_n/(4\pi))$. It seems due to the truncation of the error density described in (2).

5.2. **Comparison with existing results.** Let us first compare our estimator to the one of Neumann (1997). He denotes by $f_0(x) = e^{-|x|}/2$ and he considers two examples :
- example 1: $f = f_0 * f_0 * f_0 * f_0$ and $f_\varepsilon = f_0 * f_0$
- example 2: $f = f_0 * f_0$ and $f_\varepsilon = f_0 * f_0 * f_0 * f_0$

We set, as in Neumann (1997), $n = 200$ and $M = 10$ and the L^2 risk is computed with 100 random samples. We also compute the estimator with known noise, replacing $\tilde{f}_\varepsilon$ by f_ε in the procedure. In this case, we choose $\mathcal{M}_n = \{m = k/(4\pi), k \in \mathbb{N}^*, k \leq n^{1/4}\}$. Moreover, we take here $\text{pen}(m) = 4\Delta(m)/n = (2/n\pi) \int_{-\pi m}^{\pi m} (1+x^2)^{2d} dx$ with $d = 2$ in example 1 and $d = 4$ in example 2. The integrated L^2 risks for 100 replications are given in Table 2 and show our improvement of the results of Neumann (1997).

	ex 1	ex 2
f_ε known	0.00257	0.01904
f_ε unknown	0.00828	0.06592

	ex 1	ex 2
f_ε known	0.00243	0.01791
f_ε unknown	0.00612	0.03427

TABLE 2. MISE for the estimators of Neumann (1997) (left) and for the penalized estimator (right).

In these examples, the signal and the noise are ordinary smooth ($r = s = 0$): this induces the rates of convergence $n^{-\frac{15}{24}} + M^{-1}$ and $n^{-\frac{7}{24}} + M^{-\frac{7}{16}}$ for examples 1 and 2 respectively.

An example of estimation for supersmooth functions is given in Johannes (2009). In his example 5.1, he considers a standard Gaussian noise and $X \sim \mathcal{N}(5, 9)$. For a known noise, we use the penalty $\text{pen}(m) = (\pi m)^3 \int_0^1 \exp\{(\pi m x)^2\} dx/(2n)$ and the collection of models $\mathcal{M}_n = \{m = k/(4\pi), k \in \mathbb{N}^*, k \leq \sqrt{n}\}$. As Johannes (2009) presents only boxplots and for the sake of comparison, we give the third quartile for the L^2 risk in Table 3. In this case $r = 2$, $\delta = 1/2$ and $s = 2$, $\gamma = 0$ and the rate of convergence is $n^{-\frac{9}{10}} (\log n)^{-1/2} + M^{-1}$. The improvement brought by our method is striking.

	$n = 100$	$n = 250$	$n = 500$
f_ε known	2.0	0.9	0.6
$M = 100$	2.0	1.0	0.7
$M = 250$	1.9	1.0	0.6
$M = 500$	1.9	0.9	0.6

	$n = 100$	$n = 250$	$n = 500$
f_ε known	0.71	0.23	0.12
$M = 100$	0.24	0.11	0.07
$M = 250$	0.21	0.12	0.07
$M = 500$	0.21	0.12	0.07

TABLE 3. Third quartile of the MISE $\times 100$ for the estimators of Johannes (2009) (left) and for the penalized estimator (right).

5.3. **Other examples and influence of M.** Now we compute estimators for different signal densities and different noises. Following Comte et al. (2006) we study the following densities on the interval I:

(i) Laplace distribution: $f(x) = e^{-\sqrt{2}|x|}/\sqrt{2}$, $I = [-5, 5]$ (regularities $\delta = 2, r = 0$),

(ii) Mixed Gamma distribution: $X = W/\sqrt{5.48}$ with $W \sim 0.4\Gamma(5,1) + 0.6\Gamma(13,1)$, $I = [-1.5, 26]$ (regularities $\delta = 5, r = 0$),
(iii) Cauchy distribution: $f(x) = (\pi(1+x^2))^{-1}$, $I = [-10, 10]$ (regularities $\delta = 0, r = 1$),
(iv) Standard Gaussian distribution, $I = [-4, 4]$ (regularities $\delta = 1/2, r = 2$).

		$n = 100$		$n = 250$		$n = 500$	
f \ f_ε		Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.
Laplace	f_ε known	2.426	3.052	1.196	1.792	0.835	1.180
	$M = \lfloor\sqrt{n}\rfloor$	2.772	2.826	1.539	1.651	1.059	1.232
	$M = n^2$	2.500	2.667	1.311	1.423	0.815	0.942
Mixed Gamma	f_ε known	1.019	0.907	0.637	0.520	0.312	0.313
	$M = \lfloor\sqrt{n}\rfloor$	1.155	1.088	0.609	0.652	0.366	0.380
	$M = n^2$	1.167	1.145	0.607	0.594	0.382	0.360
Cauchy	f_ε known	1.144	0.945	0.486	0.436	0.341	0.255
	$M = \lfloor\sqrt{n}\rfloor$	1.138	1.072	0.533	0.475	0.323	0.302
	$M = n^2$	1.044	1.030	0.438	0.443	0.275	0.273
Gaussian	f_ε known	0.858	0.466	0.676	0.301	0.483	0.220
	$M = \lfloor\sqrt{n}\rfloor$	0.921	0.914	0.630	0.537	0.407	0.405
	$M = n^2$	0.715	0.593	0.529	0.476	0.383	0.259

TABLE 4. MISE $\mathbb{E}(\|f - \tilde{f}\|^2) \times 100$ averaged over 100 samples

We consider two different noises with same variance 1/10:

Laplace noise: In this case, the density of ε_i is given by

$$f_\varepsilon(x) = \frac{\lambda}{2} e^{-\lambda|x|}; \quad f_\varepsilon^*(x) = \frac{\lambda^2}{\lambda^2 + x^2}; \quad \lambda = 2\sqrt{5}.$$

The smoothness parameters are $\gamma = 2$ and $b = s = 0$. In the case when f_ε is known, we use $\text{pen}(m) = 4(\pi m + (2/(3\lambda^2))(\pi m)^3 + (1/(5\lambda^4))(\pi m)^5)/n$.

Gaussian noise: In this case, the density of ε_i is given by

$$f_\varepsilon(x) = \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{x^2}{2\lambda^2}}; \quad f_\varepsilon^*(x) = e^{-\frac{\lambda^2 x^2}{2}}; \quad \lambda = \frac{1}{\sqrt{10}}.$$

So $\gamma = 0$, $b = \lambda^2/2$ and $s = 2$. In the case when f_ε is known, we use $\text{pen}(m) = 0.5(\pi m)^3 \int_0^1 e^{(\lambda\pi m x)^2} dx/n$.

For a known noise, we take here $\mathcal{M}_n = \{m = k/(4\pi), k \in \mathbb{N}^*, k \leq k_n\}$ where

$$k_n = \arg\max \{k \in \{1, \dots, n\}, \Delta(k/(4\pi)) \leq n\}.$$

The results are given in Table 4 and are very comparable to those of Comte et al. (2006). We notice that the estimation of the characteristic function of the noise does not spoil so much the procedure. It even happens that the estimation with unknown noise works better, which is likely due to the truncation (2). We can also observe that, as expected, the risk decreases when M increases. The cases where the risk is larger for $M = n^2$ correspond to a stabilization of the decrease and are due to the variance of the results.

Figure 1 illustrates these results for two cases: a mixed Gamma density estimated through Laplace noise and a Laplace density estimated through Gaussian noise. The curves for $M = 5$ show that our method is still very satisfactory for small values of M.

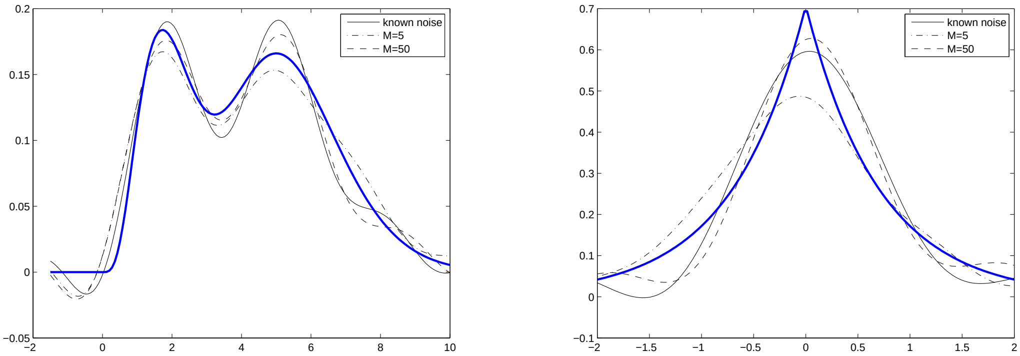

FIGURE 1. True function f (bold line) and estimators for $n = 500$. Left: mixed Gamma density with Laplace noise. Right : Laplace density with Gaussian noise

6. Proofs

For two sequences $u_{n,M,m}$ and $v_{n,M,m}$, we denote $u_{n,M,m} \lesssim v_{n,M,m}$ if there exists a positive constant C such that $u_{n,M,m} \leq C v_{n,M,m}$ for all n, M, m.

6.1. **Proof of Proposition 1.** We start from (14) and take the expectation:

$$\begin{aligned}\mathbb{E}(\|f_m - \hat{f}_m\|^2) &\leq \frac{2}{\pi}\int_{-\pi m}^{\pi m} \mathbb{E}(|\hat{f}_Y^*(u) - f_Y^*(u)|^2 |R(u)|^2)du \\ &+\frac{2}{\pi}\int_{-\pi m}^{\pi m} |f_Y^*(u)|^2 \mathbb{E}(|R(u)|^2)du + \frac{1}{\pi}\int_{-\pi m}^{\pi m} \frac{n^{-1}}{|f_\varepsilon^*(u)|^2}du.\end{aligned}$$

Applying Lemma 1 yields:

$$\begin{aligned}\mathbb{E}(\|f_m - \hat{f}_m\|^2) &\leq \frac{2}{\pi}\int_{-\pi m}^{\pi m} \mathbb{E}(|\hat{f}_Y^*(u) - f_Y^*(u)|^2)\mathbb{E}(|R(u)|^2)du \\ &+\frac{2}{\pi}\int_{-\pi m}^{\pi m} |f^*(u)|^2|f_\varepsilon^*(u)|^2 \mathbb{E}(|R(u)|^2)du + 2\frac{\Delta(m)}{n} \\ &\leq \frac{2C_1}{\pi}\int_{-\pi m}^{\pi m} n^{-1}|f_\varepsilon^*(u)|^{-2}du \\ &+\frac{2C_1}{\pi}\int_{-\pi m}^{\pi m} |f^*(u)|^2|f_\varepsilon^*(u)|^2 \frac{M^{-1}}{|f_\varepsilon^*(u)|^4}du + 2\frac{\Delta(m)}{n} \\ &\leq \frac{2C_1}{\pi M}\int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|f_\varepsilon^*(u)|^2}du + (4C_1+2)\frac{\Delta(m)}{n} \qquad (26)\end{aligned}$$

By gathering (13) and (26), we obtain the result. □

6.2. **Proof of Lemma 2.** The proof of the first result is omitted. It is obtained by distinguishing the cases $s > 2\gamma+1$ and $s \leq 2\gamma+1$ and with standard evaluations of integrals. For the second point, we first remark that $\Delta_f(m) \leq \Delta(m)$. Next, using Assumption **(A2)**,

$$\begin{aligned}\Delta_f(m) &\leq \frac{k_0^{-2}}{2\pi}\int_{-\pi m}^{\pi m}(x^2+1)^\gamma e^{2b|x|^s}|f^*(x)|^2dx \\ &\leq \frac{k_0^{-2}}{2\pi} l \sup_{x\in[-\pi m,\pi m]}((x^2+1)^{\gamma-\delta}e^{2(b|x|^s-a|x|^r)})\end{aligned}$$

Then, if $s > r$,

$$\Delta_f(m) \leq \frac{k_0^{-2}}{2\pi} l((\pi m)^2+1)^{(\gamma-\delta)_+}e^{2b(\pi m)^s}$$

If $r = s$ and $b \geq a$,

$$\Delta_f(m) \leq \frac{k_0^{-2}}{2\pi} l((\pi m)^2+1)^{(\gamma-\delta)_+}e^{2(b-a)(\pi m)^s}$$

If $r > s$ or $r = s$ and $a > b$, $\Delta_f(m)$ is bounded by a constant. □

6.3. **Proof of Theorem 1.** We observe that for all t, t'

$$\gamma_n(t) - \gamma_n(t') = \|t-f\|^2 - \|t'-f\|^2 - 2\nu_n(t-t')$$

where

$$\nu_n(t) = n^{-1}\sum_j\left\{\tilde{v}_t(Y_j) - \int t(x)f(x)dx\right\}.$$

Let us fix $m \in \mathcal{M}_n$ and recall that f_m is the orthogonal projection of f on S_m. Since $\gamma_n(\tilde{f}) + \text{pen}(\hat{m}) \leq \gamma_n(f_m) + \text{pen}(m)$, we have

$$\begin{aligned}\|\hat{f}_{\hat{m}} - f\|^2 &\leq \|f_m - f\|^2 + 2\nu_n(\hat{f}_{\hat{m}} - f_m) + \text{pen}(m) - \text{pen}(\hat{m}) \\ &\leq \|f_m - f\|^2 + 2\|\hat{f}_{\hat{m}} - f_m\|\sup_{t\in B(m,\hat{m})}\nu_n(t) + \text{pen}(m) - \text{pen}(\hat{m})\end{aligned}$$

where, for all m, m', $B(m, m') = \{t \in S_m + S_{m'},\quad \|t\| = 1\}$. Then, using inequality $2xy \leq x^2/4 + 4y^2$,

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq \|f_m - f\|^2 + \frac{1}{4}\|\hat{f}_{\hat{m}} - f_m\|^2 + 4\sup_{t\in B(m,\hat{m})}\nu_n^2(t) + \text{pen}(m) - \text{pen}(\hat{m}). \tag{27}$$

But $\|\hat{f}_{\hat{m}} - f_m\|^2 \leq 2\|\hat{f}_{\hat{m}} - f\|^2 + 2\|f - f_m\|^2$ so that, introducing a function $p(.,.)$

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq 3\|f_m - f\|^2 + 8[\sup_{t\in B(m,\hat{m})}\nu_n^2(t) - p(m,\hat{m})] + 8p(m,\hat{m}) + 2\text{pen}(m) - 2\text{pen}(\hat{m}).$$

If p is such that for all m, m',

$$4p(m,m') \leq \text{pen}(m) + \text{pen}(m') \tag{28}$$

then

$$\mathbb{E}\|\hat{f}_{\hat{m}} - f\|^2 \leq 3\|f_m - f\|^2 + 8\mathbb{E}[\sup_{t\in B(m,\hat{m})}\nu_n^2(t) - p(m,\hat{m})] + 4\text{pen}(m). \tag{29}$$

Using $\mathbb{E}(v_t(Y_j)|X_j) = t(X_j)$ (recall that v_t is defined in (8)), we split $\nu_n(t)$ into two terms: $\nu_n(t) = \nu_{n,1}(t) + S_n(t)$ with

$$\left\{ \begin{array}{l} \nu_{n,1}(t) = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} \{v_t(Y_j) - \mathbb{E}[v_t(Y_j)]\} \\ S_n(t) = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} (\tilde{v}_t - v_t)(Y_j), , \end{array} \right. \tag{30}$$

For S_n we need additional decompositions. We write

$$\begin{aligned} S_n(t) &= \frac{1}{n}\sum_{j=1}^{n} (\tilde{v}_t - v_t)(Y_j) = \frac{1}{2\pi}\int (\frac{1}{n}\sum_{j=1}^{n} e^{iuY_j})t^*(u)R(-u)du \\ &= \frac{1}{2\pi}\int \hat{f}_Y^*(u)t^*(-u)R(u)du \\ &= \frac{1}{2\pi}\int (\hat{f}_Y^*(u) - f_Y^*(u))t^*(-u)R(u)du + \frac{1}{2\pi}\int f_Y^*(u)t^*(-u)R(u)du \end{aligned}$$

Now, let $E(x) = \{|\hat{f}_\varepsilon^*(x)| \geq 1/\sqrt{M}\}$ and write

$$\begin{aligned} R(x) &= \frac{\mathbb{1}_{E(x)}}{\hat{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)} \\ &= \mathbb{1}_{E(x)}\left(\frac{1}{\hat{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)}\right) - \frac{\mathbb{1}_{E(x)^c}}{f_\varepsilon^*(x)} \\ &= \frac{(f_\varepsilon^*(x) - \hat{f}_\varepsilon^*(x))}{f_\varepsilon^*(x)}\frac{\mathbb{1}_{E(x)}}{\hat{f}_\varepsilon^*(x)} - \frac{\mathbb{1}_{E(x)^c}}{f_\varepsilon^*(x)} \\ &= \frac{(f_\varepsilon^*(x) - \hat{f}_\varepsilon^*(x))}{f_\varepsilon^*(x)}R(x) + \frac{(f_\varepsilon^*(x) - \hat{f}_\varepsilon^*(x))}{(f_\varepsilon^*(x))^2} - \frac{\mathbb{1}_{E(x)^c}}{f_\varepsilon^*(x)}. \end{aligned}$$

Thus we have

$$S_n(t) = R_{n,1}(t) + R_{n,2}(t) - R_{n,3}(t) - R_{n,4}(t)$$

where

$$\left\{ \begin{array}{l} R_{n,1}(t) = \dfrac{1}{2\pi}\displaystyle\int (\hat{f}_Y^*(u) - f_Y^*(u))t^*(-u)R(u)du, \\ R_{n,2}(t) = \dfrac{1}{2\pi}\displaystyle\int f^*(u)t^*(-u)(f_\varepsilon^*(u) - \hat{f}_\varepsilon^*(u)))R(u)du, \\ R_{n,3}(t) = \dfrac{1}{2\pi}\displaystyle\int f^*(u)t^*(-u)\frac{\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)}{f_\varepsilon^*(u)}du, \\ R_{n,4}(t) = \dfrac{1}{2\pi}\displaystyle\int f^*(u)t^*(-u)\mathbb{1}_{E(x)^c}du. \end{array} \right.$$

Now, let us define for all m and m' the function

$$p_0(m, m') = 2(\log(\Delta(m^\diamond)/\log(m^\diamond + 1))^2\Delta(m^\diamond)/n \tag{31}$$

where $\Delta(m)$ is defined in (12) and $m^\diamond = \max(m, m')$.

For $\nu_{n,1}$, we use the following proposition, mainly proved in Comte et al. (2006):

Proposition 6. *Under assumptions of Theorem 1, there exists a positive constant C such that*

$$\mathbb{E}_0 := \mathbb{E}\left(\left[\sup_{t\in B(m,\hat{m})} \nu_{n,1}^2(t) - p_0(m,\hat{m})\right]_+\right) \leq \frac{C}{n}. \tag{32}$$

Note that Theorem 1 in Comte et al. (2006) is proved with a smaller penalty, but it is easy to see that, under (**A2**), the above proposal would *a fortiori* work. Therefore, we omit the proof of Proposition 6.

For the other terms, which, for three of them, do not behave as usual residual terms, we prove:

Proposition 7. *Under assumptions of Theorem 1, there exist positive constants $C_1, \dots, C_4$, such that for $i = 1, 2, 3$,*

$$\mathbb{E}_i := \mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,i}(t)|^2 - p_0(m,m')\right) \leq \frac{C_i}{n}, \tag{33}$$

where $p_0(m,m')$ is given by (31) and

$$\mathbb{E}_4 := \mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,4}(t)|^2\right) \leq \frac{C_4}{n}. \tag{34}$$

It follows from Proposition 6 and Proposition 7 and from

$$(\nu_{n,1}(t) + \sum_{i=1}^{4} R_{n,i}(t))^2 \leq 5(\nu_{n,1}^2(t) + \sum_{i=1}^{4} R_{n,i}^2(t)),$$

that there exists a constant C such that

$$\mathbb{E}[\sup_{t\in B(m,\hat{m})} \nu_n^2(t) - p(m,\hat{m})]_+ \quad \leq \quad \frac{C}{n}. \tag{35}$$

for $p(m,m') = 20p_0(m,m')$. Moreover, (28) holds if

$$20 \times 4 \times 2 \left(\frac{\log(\Delta(m^\diamond))}{\log(m^\diamond + 1)}\right)^2 \frac{\Delta(m^\diamond)}{n} \leq \mathrm{pen}(m) + \mathrm{pen}(m')$$

for all m, m' in $\mathcal{M}_n$ (recall that $m^\diamond = \max(m, m')$). This is ensured by the choice

$$\mathrm{pen}(m) = K_0 \left(\frac{\log(\Delta(m))}{\log(m+1)}\right)^2 \Delta(m)/n,$$

with $K_0 = 160$ here. This is the choice of pen(.) given in Theorem 1.
Now, gathering (32), (33), (34) for $i = 2, \dots, 4$ and (35) yields that, $\forall m \in \mathcal{M}_n$,

$$\mathbb{E}\|\hat{f}_{\hat{m}} - f\|^2 \leq 3\|f_m - f\|^2 + 4\mathrm{pen}(m) + \frac{C}{n}.$$

This ends the proof of Theorem 1. □

6.4. **Proof of Proposition 7.** We shall use in the sequel the following Lemma:

Lemma 3. *Under Assumption* **(A2)**, *for all* $m \in \mathcal{M}_n$ *and all* $p \geq 1/2$, *we have*

$$\int_{-\pi m}^{\pi m} |f_\varepsilon^*(x)|^{-4p} dx \lesssim \Delta(m)^{2p} (\log(n))^{(2p-1)\mathbb{1}_{s>1}}.$$

Proof of Lemma 3. Using assumption **(A2)**,

$$\int_{-\pi m}^{\pi m} |f_\varepsilon^*|^{-4p} \leq k_0^{-4p} \int_{-\pi m}^{\pi m} (x^2+1)^{2\gamma p} \exp(4pb|x|^s) dx \lesssim (\pi m)^{4\gamma p + (1-s)} e^{4pb(\pi m)^s}$$

so that, using Lemma 2

$$\begin{aligned} \int_{-\pi m}^{\pi m} |f_\varepsilon^*|^{-4p} &\lesssim [(\pi m)^{2\gamma + (1-s)} e^{2b(\pi m)^s}]^{2p} m^{1-s-2p+2sp} \\ &\lesssim \Delta(m)^{2p} m^{-(2p-1)(1-s)}. \end{aligned} \tag{36}$$

Then $m^{-(2p-1)(1-s)} \leq 1$ if $s \leq 1$ and if $s > 1$, as $m \in \mathcal{M}_n$, we have $m^{s-1} \lesssim \log(n)$, which explains the result. □

In the sequel, we denote by m^* the maximum $\max(m, \hat{m})$.

6.4.1. *Study of* $R_{n,1}(t)$. We define $\Omega(x)$ the set $\Omega(x) = \Omega_1(x) \cap \Omega_2(x)$ where

$$\Omega_1(x) = \{|\hat{f}_Y^*(x) - f_Y^*(x)| \leq 4n^{-1/2}\sqrt{\log(n)}\}$$

and

$$\Omega_2(x) = \{|R(x)| \leq (M^{-1/2}/|f_\varepsilon^*(x)|^2)(n^{\epsilon/2}/\log(n))\}.$$

For t in $S_m + S_{\hat{m}} = S_{m^*}$, we can bound the term $|R_{n,1}(t)|^2$ in the following way

$$|R_{n,1}(t)|^2 \leq \frac{1}{4\pi^2} \int_{-\pi m^*}^{\pi m^*} |\hat{f}_Y^* - f_Y^*|^2 |R|^2 \int |t^*|^2$$

But, on $B(m, \hat{m})$, $\int |t^*|^2 = 2\pi \|t\|^2 \leq 2\pi$ so that

$$\sup_{t \in B(m,\hat{m})} |R_{n,1}(t)|^2 \leq \frac{1}{2\pi} \int_{-\pi m^*}^{\pi m^*} |\hat{f}_Y^* - f_Y^*|^2 |R|^2 \mathbb{1}_\Omega + \frac{1}{2\pi} \int_{-\pi m^*}^{\pi m^*} |\hat{f}_Y^* - f_Y^*|^2 |R|^2 \mathbb{1}_{\Omega^c}$$

On the one hand

$$\int_{-\pi m^*}^{\pi m^*} |\hat{f}_Y^* - f_Y^*|^2 |R|^2 \mathbb{1}_\Omega \leq 16 \frac{\log(n)}{n} \frac{n^\epsilon}{M \log^2(n)} \int_{-\pi m^*}^{\pi m^*} |f_\varepsilon^*|^{-4} \lesssim \frac{\Delta(m^*)}{n} \tag{37}$$

by using the definition of Ω, Lemma 3 and $M \geq n^{1+\epsilon}$. On the other hand

$$\begin{aligned} \mathbb{E}(\int_{-\pi m^*}^{\pi m^*} |\hat{f}_Y^* - f_Y^*|^2 |R|^2 \mathbb{1}_{\Omega^c}) &\leq \int_{-\pi m_n}^{\pi m_n} \mathbb{E}^{1/2}(|\hat{f}_Y^* - f_Y^*|^4) \mathbb{E}^{1/2}(|R|^4) \mathbb{P}^{1/2}(\Omega^c) \\ &\lesssim \int_{-\pi m_n}^{\pi m_n} n^{-1} M^{-1} |f_\varepsilon^*(x)|^{-4} \mathbb{P}^{1/2}(\Omega(x)^c) dx \\ &\lesssim \frac{\Delta(m_n)^2 \log(n)}{Mn} \|\mathbb{P}^{1/2}(\Omega^c)\|_\infty \end{aligned}$$

with Lemma 3. Now, using Bernstein inequality, it is easy to see that for all x,

$$\mathbb{P}(\Omega_1(x)^c) \leq \frac{2}{n^2}.$$

On the other hand, using the Markov inequality, for any $p \geq 1$,

$$\mathbb{P}(\Omega_2(x)^c) \leq \mathbb{E}\left(|R(x)|^{2p} M^p |f_\varepsilon^*(x)|^{4p}\right) \frac{(\log n)^{2p}}{n^{p\epsilon}} \lesssim \frac{(\log n)^{2p}}{n^{p\epsilon}}$$

by using Lemma 1. Now taking $p = Int[2/\epsilon] + 1$ yields $\mathbb{P}(\Omega_2(x)^c) \leq C_\epsilon/n^2$. As a consequence

$$\mathbb{E}(\int_{-\pi m^*}^{\pi m^*} |\hat{f}_Y^* - f_Y^*|^2 |R|^2 \mathbb{1}_{\Omega^c}) \lesssim c_\epsilon/n. \tag{38}$$

Gathering (37) and (38) gives (33) for $i = 1$. □

6.4.2. *Study of* $R_{n,2}(t)$. The following result obviously holds $\mathbb{E}[|f_\varepsilon^* - \hat{f}_\varepsilon^*|^p] \lesssim M^{-p/2}$. Moreover let

$$\Xi(x) = \{|f_\varepsilon^*(x) - \hat{f}_\varepsilon^*(x)| \leq 4M^{-1/2}\sqrt{\log(n)}\} \cap \Omega_2(x).$$

As previously, we have $\mathbb{P}(\Xi(x)^c) \leq c'_\epsilon/n^2$.

We can bound the term $|R_{n,2}(t)|^2$ in the following way

$$\sup_{t \in B(m,\hat{m})} |R_{n,2}(t)|^2 \leq \frac{1}{2\pi}\int_{-\pi m^*}^{\pi m^*} |f^*|^2 |f_\varepsilon^* - \hat{f}_\varepsilon^*|^2 |R|^2 \mathbb{1}_\Xi + \frac{1}{2\pi}\int_{-\pi m^*}^{\pi m^*} |f^*|^2 |f_\varepsilon^* - \hat{f}_\varepsilon^*|^2 |R|^2 \mathbb{1}_{\Xi^c}$$

On the one hand

$$\begin{aligned}\int_{-\pi m^*}^{\pi m^*} |f^*|^2 |f_\varepsilon^* - \hat{f}_\varepsilon^*|^2 |R|^2 \mathbb{1}_\Xi &\leq 64 \int_{-\pi m^*}^{\pi m^*} |f^*|^2 \frac{\log(n) n^\epsilon}{\log^2(n) M^2 |f_\varepsilon^*|^4} \\ &\lesssim \frac{n^\epsilon [\Delta(m^*)]^2}{M^2} \lesssim \frac{\Delta(m^*)}{M} \leq p(m, \hat{m})\end{aligned}$$

where we used the definition of $\Xi(x)$ and Lemma 3 again with $p = 1$.

On the other hand

$$\begin{aligned}\mathbb{E}(\int_{-\pi m^*}^{\pi m^*} |f^*|^2 |f_\varepsilon^* - \hat{f}_\varepsilon^*|^2 |R|^2 \mathbb{1}_{\Xi^c}) &\leq \int_{-\pi m_n}^{\pi m_n} |f^*|^2 \mathbb{E}^{1/4}(|f_\varepsilon^* - \hat{f}_\varepsilon^*|^8) \mathbb{E}^{1/4}(|R|^8) \mathbb{P}^{1/2}(\Xi^c) \\ &\lesssim \int_{-\pi m_n}^{\pi m_n} M^{-2} |f^*(x)|^2 |f_\varepsilon^*(x)|^{-4} \mathbb{P}^{1/2}(\Xi(x)^c) dx \\ &\lesssim [\Delta(m_n)]^2 \log(n) M^{-2} \|\mathbb{P}^{1/2}(\Xi^c)\|_\infty \lesssim c/n.\end{aligned}$$

6.4.3. *Study of* $R_{n,3}(t)$. We can write

$$R_{n,3}(t) = \frac{1}{M}\sum_{k=1}^{M} [F_t(\varepsilon_{-k}) - \mathbb{E}(F_t(\varepsilon_{-k}))]$$

with

$$F_t(u) = \frac{1}{2\pi}\int \frac{f^*(x)}{f_\varepsilon^*(x)} t^*(-x) e^{-ixu} dx.$$

Moreover,

$$\mathbb{E}\left[\sup_{t \in B(m,\hat{m})} |R_{n,3}(t)|^2 - p(m,\hat{m})\right]_+ \leq \sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\sup_{t \in B(m,m')} |R_{n,3}(t)|^2 - p(m,m')\right]_+$$

which replaces the supremum on a random unit ball ($\hat{m}$ is random) by suprema on deterministic unit balls. Then we use the following Lemma

Lemma 4. *Let $T_1, \dots, T_M$ be independent random variables and $\nu_M(r) = (1/M)\sum_{j=1}^M [r(T_j) - \mathbb{E}(r(T_j)]$, for r belonging to a countable class $\mathcal{R}$ of measurable functions. Then, for $\epsilon > 0$,*

$$\mathbb{E}[\sup_{r\in\mathcal{R}} |\nu_M(r)|^2 - (1+2\epsilon)H^2]_+ \leq C\left(\frac{v}{M}e^{-K_1\epsilon\frac{MH^2}{v}} + \frac{B^2}{M^2C^2(\epsilon)}e^{-K_2C(\epsilon)\sqrt{\epsilon}\frac{MH}{B}}\right) \tag{39}$$

with $K_1 = 1/6$, $K_2 = 1/(21\sqrt{2})$, $C(\epsilon) = \sqrt{1+\epsilon} - 1$ and C a universal constant and where

$$\sup_{r\in\mathcal{R}} \|r\|_\infty \leq B, \quad \mathbb{E}\left(\sup_{r\in\mathcal{R}} |\nu_M(r)|\right) \leq H, \quad \sup_{r\in\mathcal{R}} \frac{1}{M}\sum_{j=1}^M \mathrm{Var}(r(T_j) \leq v.$$

Inequality (39) is a straightforward consequence of the Talagrand (1996) inequality given Birgé and Massart (1997). Moreover, standard density arguments allow to apply it to the unit ball $B(m, m')$.
We can determine B, H and v is our problem, and we get $B = \sqrt{\Delta(m^\diamond)}$, $H^2 = \sqrt{p_0(m,m')/n}$ (which is an over-estimation) and $v = C\min(\Delta(m^\diamond), \|f_\varepsilon\|\sqrt{\Delta_2(m^\diamond)})$ with $\Delta_2(m) = \int |f_\varepsilon^*|^{-4}$. We do not give the details since the study is the same as for Proposition 6, which is proved in Comte et al. (2006). □

6.4.4. *Study of $R_{n,4}(t)$.* It is easy to see that

$$\sup_{t\in B(m,\hat{m})} |R_{n,4}(t)|^2 \leq \frac{1}{2\pi}\int_{-\pi m^*}^{\pi m^*} |f^*(u)|^2 \mathbb{1}_{E^c}(u)du,$$

and thus

$$\mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,4}(t)|^2\right) \leq \frac{1}{2\pi}\int_{-\pi m_n}^{\pi m_n} |f^*(u)|^2 \mathbb{P}(E^c(u))du.$$

Now, $\mathbb{P}(E^c(x)) = \mathbb{P}(|\hat{f}_\varepsilon^*(x)| < 1/\sqrt{M})$. We use that, as $\Delta(m_n) \leq 2n$, it holds that for $x \in [-\pi m_n, \pi m_n]$, $|f_\varepsilon^*(x)|^{-2} \leq \Delta(m_n) \leq 2n \leq 2M^{1/(1+\epsilon)}$. Thus $|f_\varepsilon^*(x)| \geq 2/\sqrt{M}$, and proceeding as in Neumann (1997), we apply Bernstein Inequality and we get

$$\begin{aligned}
\mathbb{P}(|\hat{f}_\varepsilon^*(x)| < M^{-1/2}) &\leq \mathbb{P}(|\hat{f}_\varepsilon^*(x) - f_\varepsilon^*(x)| > |f_\varepsilon^*(x)| - M^{-1/2}) \\
&\leq \mathbb{P}(|\hat{f}_\varepsilon^*(x) - f_\varepsilon^*(x)| > M^{-1/2}) \\
&\leq \kappa\exp(-\kappa M|f_\varepsilon^*(x)|^2) = O(M^{-p}|f_\varepsilon^*(x)|^{-2p}),
\end{aligned} \tag{40}$$

for all $p \geq 1$. Then,

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,4}(t)|^2\right) &\lesssim \int_{-\pi m_n}^{\pi m_n} |f^*(u)|^2 M^{-p}|f_\varepsilon^*(u)|^{-2p}du \\
&\lesssim \frac{\Delta(m_n)^p\log(n)^{p-1/2}}{M^p} \lesssim n^{-p\epsilon}\log(n)^{p-1/2} \lesssim n^{-1},
\end{aligned}$$

for $p \geq Int[1/\epsilon] + 1$. This gives the result for $R_{n,4}$. □

6.5. **Proof of Theorem 2.**

6.5.1. *Proof of Theorem 2.* Let us define $m_{opt}(n)$ as the solution of the bias-variance compromise, that is the index m such that

$$m_{opt}(n) = \inf_{m\in\{1,\dots,m_n\}} (\|f-f_m\|^2 + \text{pen}(m)).$$

It follows that $\text{pen}(m_{opt}(n))$ has the order of a rate of convergence; therefore, it tends to 0 when n goes to infinity, and so does $\Delta(m_{opt}(n))/n$. We assume that n is large enough so that $\Delta(m_{opt}(n))/n$ is less than 1/8.

Next, we use the following sets:

$$\Omega = \{m_{opt}(n) \le \hat{m}_n \le m_n\},$$

and

$$\Lambda = \{\forall m \in \mathcal{M}_n, \quad \frac{1}{2\pi}\int_{-\pi m}^{\pi m} |R|^2 \le \frac{\Delta(m)}{4}\}.$$

Let

$$\hat{\Delta}(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} dx/|\tilde{f}_\varepsilon(x)|^2.$$

Since $\Delta(m) \le \frac{2}{2\pi}\int_{-\pi m}^{\pi m}|R|^2 + 2\hat{\Delta}(m)$, we can write on Λ, $\Delta(m) \le \Delta(m)/2 + 2\hat{\Delta}(m)$ and then

$$\Delta(m)\mathbb{1}_\Lambda \le 4\hat{\Delta}(m)\mathbb{1}_\Lambda.$$

Analogously, $\hat{\Delta}(m) \le \frac{2}{2\pi}\int_{-\pi m}^{\pi m}|R|^2 + 2\Delta(m)$, and we can write on Λ, $\hat{\Delta}(m) \le \Delta(m)/2 + 2\Delta(m)$ and then

$$\hat{\Delta}(m)\mathbb{1}_\Lambda \le (5/2)\Delta(m)\mathbb{1}_\Lambda.$$

Note that this implies, since $\log((5/2)x) \le 2\log(x), \forall x \ge 2.5$,

$$\widetilde{\text{pen}}(m)\mathbb{1}_\Lambda \le 10\frac{K_1}{K_0}\text{pen}(m)\mathbb{1}_\Lambda. \tag{41}$$

Reasoning as in the proof of Theorem 1, if p is such that for all m, m',

$$4p(m,m')\mathbb{1}_{\Lambda\cap\Omega} \le \widetilde{\text{pen}}(m)\mathbb{1}_{\Lambda\cap\Omega} + \widetilde{\text{pen}}(m')\mathbb{1}_{\Lambda\cap\Omega}$$

then

$$\begin{aligned}\|\tilde{f}-f\|^2\mathbb{1}_{\Lambda\cap\Omega} &\le 3\|f_{m_{opt}(n)}-f\|^2 + 8[\sup_{t\in B(m_{opt}(n),\hat{m})} \nu_n^2(t) - p(m_{opt}(n),\hat{m})]\mathbb{1}_{\Lambda\cap\Omega} \\ &\quad +4\widetilde{\text{pen}}(m_{opt}(n))\mathbb{1}_{\Lambda\cap\Omega}.\end{aligned}$$

It follows from the proof of Theorem 1 that

$$8\mathbb{E}[\sup_{t\in B(m_{opt}(n),\hat{\hat{m}})} \nu_n^2(t) - p(m_{opt}(n),\hat{\hat{m}})]_+ \le C/n$$

with $p(m,m') = 2(\log(\Delta(m^\diamond))/\log(m^\diamond+1))^2\Delta(m^\diamond)/n$ and $m^\diamond = m\vee m'$.
Thus, choosing $\widetilde{\text{pen}}(m) = 128((\log(\hat{\Delta}(m))/\log(m+1))^2\hat{\Delta}(m)/n$, on $\Lambda\cap\Omega$,

$$\begin{aligned}4p(m,m') &= 8(\log(\Delta(m^\diamond))/\log(m^\diamond+1))^2\Delta(m^\diamond)/n \\ &\le 8(\log(\Delta(m))/\log(m+1))^2\Delta(m)/n + 8(\log(\Delta(m'))/\log(m'+1))^2\Delta(m')/n \\ &\le 128(\log(\hat{\Delta}(m))/\log(m+1))^2\hat{\Delta}(m)/n + 128(\log(\hat{\Delta}(m'))/\log(m'+1))^2\hat{\Delta}(m')/n \\ &\le \widetilde{\text{pen}}(m) + \widetilde{\text{pen}}(m'),\end{aligned}$$

where we use that $\log(4\hat{\Delta}(m)) \leq 2\log(\hat{\Delta}(m))$ which holds as $m \geq m_{opt(n)}$ and n large enough. Then, using (41) and the definition of $m_{opt(n)}$,

$$\begin{aligned}\mathbb{E}(\|\tilde{f}-f\|^2 \mathbb{1}_{\Lambda\cap\Omega}) &\leq 3\|f_{m_{opt(n)}}-f\|^2 + 4\mathbb{E}(\widetilde{\text{pen}}(m_{opt(n)})\mathbb{1}_{\Lambda\cap\Omega}) + \frac{C}{n} \\ &\leq 3\|f_{m_{opt(n)}}-f\|^2 + 40\frac{K_0}{K_1}\text{pen}(m_{opt(n)}) + \frac{C}{n} \\ &\leq (3 \vee 40\frac{K_0}{K_1}) \inf_{m\in\{1,\dots,m_n\}} (\|f_m - f\|^2 + \text{pen}(m)) + \frac{C}{n}. \qquad (42)\end{aligned}$$

We prove below that

$$\mathbb{E}(\|\tilde{f}-f\|^2 \mathbb{1}_{\Lambda^c\cap\Omega}) \leq \frac{C}{n} \qquad (43)$$

and

$$\mathbb{E}(\|\tilde{f}-f\|^2 \mathbb{1}_{\Omega^c}) \leq \frac{C}{n}. \qquad (44)$$

Gathering the results in (42), (43) and (44) gives the result of Theorem 2. □

6.5.2. *Proof of (43) and (44). Proof of (43).* First we compute, using formula (3) ,

$$\|\tilde{f}\|^2 = \frac{1}{2\pi}\int |\tilde{f}^*_{\hat{\hat{m}}}|^2 = \frac{1}{2\pi}\int_{-\pi\hat{\hat{m}}}^{\pi\hat{\hat{m}}} \frac{|\hat{f}^*_Y|^2}{|\tilde{f}^*_\varepsilon|^2} \leq \frac{1}{2\pi}\int_{-\pi\hat{\hat{m}}}^{\pi\hat{\hat{m}}} |\tilde{f}^*_\varepsilon|^{-2}$$

But $|\tilde{f}^*_\varepsilon(x)|^{-2} = |\hat{f}^*_\varepsilon(x)|^{-2}\mathbb{1}_{\{|\hat{f}^*_\varepsilon(x)|\geq M^{-1/2}\}} \leq M$. Then

$$\|\tilde{f}\|^2 \leq M\hat{\hat{m}} \leq Mn \leq n^{2+\epsilon} \leq n^3 \qquad (45)$$

and thus $\mathbb{E}(\|\tilde{f}-f\|^2\mathbb{1}_{\Lambda^c\cap\Omega}) \lesssim n^3 P(\Lambda^c\cap\Omega)$. Now, using Markov and Jensen inequalities, $\forall p \geq 1$

$$\begin{aligned}\mathbb{P}(\Lambda^c\cap\Omega) &\leq \sum_{m\in\mathcal{M}_n} \mathbb{P}(\frac{1}{2\pi}\int_{-\pi m}^{\pi m}|R|^2 > \frac{\Delta(m)}{4}) \leq \sum_{m\in\mathcal{M}_n}\left(\frac{4}{\Delta(m)}\right)^p \mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m}|R|^2\right)^p\right] \\ &\leq \left(\frac{4}{2\pi}\right)^p \sum_{m\in\mathcal{M}_n}\Delta(m)^{-p}\mathbb{E}\left[(2\pi m)^{p-1}\int_{-\pi m}^{\pi m}|R|^{2p}\right] \\ &\leq \frac{4^p}{2\pi}\sum_{m\in\mathcal{M}_n}\Delta(m)^{-p}m^{p-1}\int_{-\pi m}^{\pi m}\mathbb{E}|R|^{2p}\end{aligned}$$

Since $\mathbb{E}|R|^{2p} \lesssim M^{-p}|f^*_\varepsilon|^{-4p}$ (Lemma 1),

$$\mathbb{P}(\Lambda^c\cap\Omega) \lesssim \frac{4^p}{2\pi}M^{-p}\sum_{m\in\mathcal{M}_n}\Delta(m)^{-p}m^{p-1}\int_{-\pi m}^{\pi m}|f^*_\varepsilon|^{-4p}$$

Now, using the proof of Lemma 3, we have

$$\int_{-\pi m}^{\pi m}|f^*_\varepsilon|^{-4p} \lesssim \Delta(m)^{2p}m^{-(2p-1)(1-s)}$$

Hence

$$\mathbb{P}(\Lambda^c\cap\Omega) \lesssim M^{-p}\sum_{m\in\mathcal{M}_n}\Delta(m)^p m^{p-1-(2p-1)(1-s)} \lesssim M^{-p}\Delta(m_n)^p\sum_{m\in\mathcal{M}_n}m^{p-1-(2p-1)(1-s)}.$$

If $s=0$, $p-1-(2p-1)(1-s)=-p$ and $\sum_{m\in\mathcal{M}_n} m^{-p}=O(1)$ if $p>1$; if $s>0$, as m_n has logarithmic order and $\sum_{m\in\mathcal{M}_n} m^{p-1-(2p-1)(1-s)}$ has also logarithmic order. Therefore

$$\mathbb{P}(\Lambda^c\cap\Omega)\lesssim n^{-p\epsilon}\sum_{m\in\mathcal{M}_n} m^{p-1-(2p-1)(1-s)}\le c/n^4$$

for $p\ge Int[4/\epsilon]+1$.

Finally $\mathbb{E}(\|\tilde f-f\|^2\mathbb{1}_{\Lambda^c\cap\Omega})\lesssim n^3P(\Lambda^c\cap\Omega)\lesssim c/n$ and this ends the proof of (43). □

Proof of (44). As (45) is still valid, we have to prove that $\mathbb{P}(\Omega^c)\le c/n^4$. For the study of Ω, we write

$$\mathbb{P}(\Omega^c)\le\mathbb{P}(\hat m_n<m_{opt}(n))+\mathbb{P}(m_n<\hat m_n).$$

We study separately the two terms. It follows from the definition of $\hat m_n$ and m_n, that, if $m_n<\hat m_n$, then $\hat\Delta(\hat m_n)\ge\hat\Delta(m_n)$. Moreover $\hat\Delta(\hat m_n)/n<1/2$ and $\Delta(m_n)/n\ge 1$. Therefore, if $m_n<\hat m_n$, then $1/2>\hat\Delta(\hat m_n)/n\ge\hat\Delta(m_n)/n$ while $\Delta(m_n)/n>1$; therefore the distance between $\Delta(m_n)/n$ and $\hat\Delta(m_n)/n$ is more than $1/2$. This yields

$$\mathbb{P}(m_n<\hat m_n)\le\mathbb{P}(\Delta(m_n)/n-\hat\Delta(m_n)/n\ge 1/2)\le\mathbb{P}(|\Delta(m_n)-\hat\Delta(m_n)|/n\ge 1/2).$$

Similarly, if $\hat m_n<m_{opt}(n)$, then $\hat\Delta(m_n)\le\hat\Delta(m_{opt}(n))$, so that $1/4\le\hat\Delta(m_n)/n\le\hat\Delta(m_{opt}(n))/n$ while $\Delta(m_{opt}(n))/n\le 1/8$; therefore the distance between $\Delta(m_{opt}(n))/n$ and $\hat\Delta(m_{opt}(n))/n$ is larger than $1/8$. Thus

$$\mathbb{P}(\hat m_n<m_{opt}(n))\quad\le\quad\mathbb{P}(|\hat\Delta(m_{opt}(n))-\Delta(m_{opt}(n))|/n\ge 1/8).$$

As we prove below that

$$\mathbb{P}(|\Delta(m_n)-\hat\Delta(m_n)|/n\ge 1/2)\le\frac{C}{n^4},\tag{46}$$

and

$$\mathbb{P}(|\hat\Delta(m_{opt}(n))-\Delta(m_{opt}(n))|/n\ge 1/8)\le\frac{C'}{n^4}\tag{47}$$

we obtain (44). □

6.5.3. *Proof of (46) and (47). Proof of (46).* With Markov Inequality,

$$\begin{aligned}\mathbb{P}(|\Delta(m_n)-\hat\Delta(m_n)|/n\ge 1/2)&\le\frac{2^p}{n^p}\mathbb{E}[|\Delta(m_n)-\hat\Delta(m_n)|^p]\\&\le\frac{2^p(2\pi m_n)^{p-1}}{n^p}\mathbb{E}\left(\int_{-\pi m_n}^{\pi m_n}\left|\frac{1}{|f_\varepsilon^*(x)|^2}-\frac{1}{|\tilde f_\varepsilon^*(x)|^2}\right|^p dx\right)\\&\le T_1+T_2\end{aligned}$$

where

$$T_1=\frac{C_pm_n^{p-1}}{n^p}\int_{-\pi m_n}^{\pi m_n}\frac{\mathbb{E}(\mathbb{1}_{|\hat f_\varepsilon^*|<1/\sqrt{M}})}{|f_\varepsilon^*(x)|^{2p}}dx$$

and

$$T_2=\frac{C_pm_n^{p-1}}{n^p}\mathbb{E}\left(\int_{-\pi m_n}^{\pi m_n}\left|\frac{|f_\varepsilon^*(x)|^2-|\hat f_\varepsilon^*(x)|^2}{|f_\varepsilon^*(x)|^2|\tilde f_\varepsilon^*(x)|^2}\right|^p dx\right).$$

It follows from (40) that

$$T_1 \lesssim m_n^{p-1} n^{-p} M^{-p} \int_{-\pi m_n}^{\pi m_n} \frac{dx}{|f_\varepsilon^*(x)|^{4p}} \lesssim \log(n)^{a_p} n^{-p\epsilon},$$

where $a_p = 0$ in the ordinary smooth case. Therefore, choosing $p > Int[4/\epsilon] + 1$ implies $T_1 \lesssim C/n^4$.

The study of T_2 relies on the Lemma

Lemma 5. $\forall r \geq 1$, $\mathbb{E}(||\hat{f}_\varepsilon^*(x)|^2 - |f_\varepsilon^*(x)|^2|^r) \leq C_r(M^{-r} + |f_\varepsilon^*(x)|^r M^{-r/2})$.

Proof of Lemma 5. The result follows from the equality $|\hat{f}_\varepsilon^*(x)|^2 - |f_\varepsilon^*(x)|^2 = |\hat{f}_\varepsilon^*(x) - f_\varepsilon^*(x)|^2 + 2Re(\bar{f}_\varepsilon^*(x)(\hat{f}_\varepsilon^*(x) - f_\varepsilon^*(x))$, and $\forall r \geq 1$, with Rosenthal Inequality, $\mathbb{E}(|\hat{f}_\varepsilon^*(x) - f_\varepsilon^*(x)|^r) \leq C_r' M^{-r/2}$. □

Lemma 5 implies, as $1/|\tilde{f}_\varepsilon^*|^2 \leq 2|R|^2(x) + 2/|f_\varepsilon^*|^2$,

$$\begin{aligned}
T_2 &\leq \frac{C_p m_n^{p-1}}{n^p} \int_{-\pi m_n}^{\pi m_n} \frac{1}{|f_\varepsilon^*(x)|^{2p}} \mathbb{E}^{1/2}(\frac{1}{|\tilde{f}_\varepsilon^*(x)|^{4p}}) \mathbb{E}^{1/2}(\left||f_\varepsilon^*(x)|^2 - |\tilde{f}_\varepsilon^*(x)|^2\right|^{2p})dx \\
&\leq \frac{C_p' m_n^{p-1}}{n^p} \int_{-\pi m_n}^{\pi m_n} \frac{M^{-p} + |f_\varepsilon^*(x)|^p M^{-p/2}}{|f_\varepsilon^*(x)|^{2p}} (\frac{M^{-p}}{|f_\varepsilon^*(x)|^{4p}} + \frac{1}{|f_\varepsilon^*(x)|^{2p}})dx \\
&\leq \frac{C_p' m_n^{p-1}}{n^p} \int_{-\pi m_n}^{\pi m_n} \left(\frac{M^{-2p}}{|f_\varepsilon^*(x)|^{6p}} + \frac{M^{-p}}{|f_\varepsilon^*(x)|^{4p}} + \frac{M^{-3p/2}}{|f_\varepsilon^*(x)|^{5p}} + \frac{M^{-p/2}}{|f_\varepsilon^*(x)|^{3p}} \right) dx.
\end{aligned}$$

As $\Delta(m_n) \leq 2n$, $\forall q \geq 1/4$,

$$\int_{-\pi m_n}^{\pi m_n} \frac{dx}{|f_\varepsilon^*(x)|^q} dx \lesssim (\pi m_n)^{(1-q/2)(1-s)} \Delta(m_n)^{q/2} \lesssim (\pi m_n)^{(1-q/2)(1-s)} n^{q/2},$$

we find

$$\begin{aligned}
T_2 &\leq \frac{C_p m_n^{p-1}}{n^p} \Big(M^{-2p} (\pi m_n)^{(1-3p)(1-s)} n^{3p} + M^{-p} (\pi m_n)^{(1-2p)(1-s)} n^{2p} \\
&\qquad + M^{-3p/2} (\pi m_n)^{(1-s)(1-5p/2)} n^{5p/2} + M^{-p/2} (\pi m_n)^{(1-3p/2)(1-s)} n^{3p/2} \Big) \\
&\leq C_p m_n^{p-1} ((\pi m_n)^{(1-3p)(1-s)} n^{-2p\epsilon} + (\pi m_n)^{(1-2p)(1-s)} n^{-p\epsilon} \\
&\qquad + (\pi m_n)^{(1-s)(1-5p/2)} n^{-3p\epsilon/2} + n^{p/2} (\pi m_n)^{(1-3p/2)(1-s)} n^{-p\epsilon/2}) \\
&= O(m_n^{p-1-(1-s)(3p/2-1)} n^{-p\epsilon/2})
\end{aligned}$$

where $p - 1 - (1-s)(3p/2 - 1) = -p/2 \leq 0$ if $s = 0$, and if $s > 0$, the power does not matter as m_n has logarithmic order. It follows that $T_2 = O(n^{-4})$ for $p/2 > [4/\epsilon] + 1$.

The proof of (47) follows the same line and, with now $\Delta(m_{opt}(n))/n \leq 1/8$ and $m_{opt}(n)$ much smaller than m_n. This leads clearly to the same result. □.

References

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York.

Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.*, 32(2):181–192.

Butucea, C. and Tsybakov, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.*, 52(1):111–128.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83(404):1184–1186.

Cator, E. A. (2001). Deconvolution with arbitrarily smooth kernels. *Statist. Probab. Lett.*, 54(2):205–214.

Cavalier, L. and Raimondo, M. (2007). Wavelet deconvolution with noisy eigenvalues. *IEEE Trans. Signal Process.*, 55(6, part 1):2414–2424.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.*, 34(3):431–452.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2007). Finite sample penalization in adaptive density deconvolution. *J. Stat. Comput. Simul.*, 77(11):977–1000.

Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.*, 56(1):19–47.

Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, 36(2):665–685.

Devroye, L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.*, 17(2):235–239.

Diggle, P. J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B*, 55(2):523–531.

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.*, 92(438):526–535.

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

Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory*, 48(3):734–747.

Hesse, C. H. (1999). Data-driven deconvolution. *J. Nonparametr. Statist.*, 10(4):343–373.

Hoffmann, M. and Reiss, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.*, 36(1):310–336.

Johannes, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.*, 37(5A):2301–2323.

Kerkyacharian, G., Pham Ngoc, T. M., and Picard, D. (2010). Localized spherical deconvolution. *arXiv:0908.1952*.

Koo, J.-Y. (1999). Logspline deconvolution in Besov space. *Scand. J. Statist.*, 26(1):73–86.

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

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.*, 65(2):139–165.

Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.*, 17(4):427–438.

Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory*, 37(4):1105–1115.

Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canad. J. Statist.*, 32(4):439–449.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.*, 7(4):307–330.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.*, 98(10):1955–1968.

Odiachi, P. and Prieve, D. (2004). Removing the effects of additive noise in tirm measurements. *J. Colloid Interface Sci.*, 270:113–122.

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

Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, 21(2):169–184.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, 18(2):806–831.