Data driven density estimation in presence of unknown convolution operator

Fabienne Comte, Claire Lacour

DECONVOLUTION WITH ESTIMATED CHARACTERISTIC FUNCTION OF THE ERRORS.

F. COMTE AND C. LACOUR

ABSTRACT. We study the following model of deconvolution $Y = X+\varepsilon$ with 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 ε_j. The aim of the paper is to estimate f without knowing the density f_ε of the ε_j. We first define an estimator, for which we provide bounds for the pointwise and 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.

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. In many contexts however, preliminary calibration measures can be obtained in the absence of any signal X. This can be done each time a physical machine takes measures of a signal; when no signal is in input, only the noise is measured. In that case, the knowledge of f_ε can be replaced by the observations of $\varepsilon_{-1}, \dots, \varepsilon_{-M}$, a noise sample with distribution f_ε, independent of $(Y_1, \dots, Y_n)$. Thus, a study can be conducted, in which we do not assume that f_ε is known. Note that the availability of two distinct samples makes the problem identifiable.

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 kernel methods, Koo (1999) for a spline method, Pensky and Vidakovic (1999) and Fan and Koo (2002) for wavelet strategies, , Butucea (2004) and Butucea and Tsybakov (2007) for studies of minimaxity of the rates, Comte et al. (2006), Comte et al. (2007) for adaptive projection strategies. On the other hand, several 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. In this regard, we provide the first study of pointwise mean square risk (MSE). Moreover, we provide a general study of the mean integrated squared error (MISE) which substantially generalizes existing

Date: November 14, 2008.

results. Then, to propose a model selection strategy, we use a minimum projection contrast expression of our collections of estimators: they depend on a bandwidth-type parameter for which we consider 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 an adaptive procedure in this context is essentially new.

Let us describe what has been done on the subject.

– Diggle and Hall (1993) consider the same model and obtain result of the same type of our first Proposition (see Proposition 2). But then, they study the case of ordinary smooth noise and distribution function under $M \geq n$.

– We may mention the work by Efromovich (1997), but his context is slightly different, since he considers circular data. He assumes that the noise is supersmooth and the distribution function ordinary smooth. In this context, he proposes a data driven choice of M to make his estimator adaptive.

– Our work is more related to the paper of Neumann (1997), since our estimator is rather equivalent to his and we borrow a useful Lemma from his work. He mainly 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. We shall of course refer to this lower bound.

– Meister (2004) takes a rather different point of view, compared to our problem: he studies what happens when the function f_ε^* used for estimation is not the true one. For instance, he shows that it is safer to use an ordinary smooth noise characteristic function, if it is unknown.

– Lastly, Johannes (2007) recently studied the density deconvolution with unknown (but observed) noise and he is interested in the relation between M and n. Note that 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 rate that are the same as in our work. But the data-driven selection of these bandwidths is not done.

Note that a similar question in the context of inverse problem is studied in Cavalier and Raimondo (2009).

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 both the pointwise and the integrated mean square risk (MSE and MISE) of one estimator, which allows to build general tables for the rates in both cases. 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 an adaptive estimator by proposing a penalization device. A general integrated risk bound for the resulting estimator is given. The estimator is studied through simulation experiments, and its performances are compared with Neumann (1997)'s and Johannes (2007)'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. Most proofs are gathered in Section 6.

2. Estimation procedure

2.1. **Notations.** 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 estimator.** It follows easily 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 $f_Y^* = f^* f_\varepsilon^*$. 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: $\int_{|u|\le \pi m} e^{iux} \widehat{f_Y^*}(u)/f_\varepsilon^*(u)du$. 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 sample and we define

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

the natural estimator of f_ε^*. Next, we introduce as in Neumann (1997) the truncated estimator:

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

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 $(\hat{f}_m)^* = (\hat{f}_Y^*/\tilde{f}_\varepsilon^*)\mathbb{1}_{[-\pi m, \pi m]}$. This formula is the only needed for the study of the MSE and the MISE, but it is not convenient to present the strategy which is required 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. And the definition of this procedure must be done in terms that leads to MISE bounds on the final estimator.

2.3. Definion of the adaptive estimator.

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 $\varphi^*_{m,j}(x) = e^{-ixj/m}\mathbb{1}_{[-\pi m,\pi m]}(x)/\sqrt{m}$. Note 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 $g^*_m = g^*\mathbb{1}_{[-\pi m,\pi m]}$, i.e. with Fourier inverse formula:

$$g_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu} g^*(u)du. \tag{4}$$

This explains why the order of bias terms is the same for the two expressions of the estimator.

2.3.2. *Estimation of f for the classical deconvolution problem.* We want to estimate f, the density of the X_j in model (1). When f_ε is known, we can estimate f by minimizing a contrast built as follows. A standard contrast in density estimation is

$$\frac{1}{n}\sum_j [\|t\|^2 - 2t(X_j)].$$

It is not possible to use this contrast in the convolution model because we do not observe $X_1, \ldots, X_n$. Only the noisy data Y_j are available. The solution is given by exploiting the following lemma.

Lemma 1. *For any function t, let v_t be the inverse Fourier transform of $t^*/f^*_\varepsilon(-.)$, i.e.*

$$v_t(x) = \frac{1}{2\pi}\int e^{ixu}\frac{t^*(u)}{f^*_\varepsilon(-u)}du.$$

Then, for all $1 \le j \le n$,

(1) $\mathbb{E}[v_t(Y_j)|X_j] = t(X_j)$
(2) $\mathbb{E}[v_t(Y_j)] = \mathbb{E}[t(X_j)]$

The second assertion in Lemma 1 is an obvious consequence of the first one and leads us to consider the following contrast:

$$\gamma_n^0(t) = \frac{1}{n}\sum_{j=1}^{n}[\|t\|^2 - 2v_t(Y_j)] \quad \text{with} \quad v_t^*(u) = \frac{t^*(u)}{f_\varepsilon^*(-u)}.$$

Indeed, since $t(X_j)$ and $v_t(Y_j)$ have the same expectation, it is natural to replace the unknown quantity $t(X_j)$ in the contrast by $v_t(Y_j)$. We can observe that

$$\begin{aligned}\mathbb{E}\gamma_n^0(t) &= (1/n)\sum_{j=1}^{n}[\|t\|^2 - 2\mathbb{E}[v_t(Y_j)]] = (1/n)\sum_{j=1}^{n}[\|t\|^2 - 2\mathbb{E}[t(X_j)]] \\ &= \|t\|^2 - 2\int tf = \|t-f\|^2 - \|f\|^2.\end{aligned}$$

This contrast is used in Comte et al. (2006) to define a collection of estimators on each space S_m and then a penalty is given to select an adequate space S_m.

2.3.3. *Estimation of f if the distribution of ε is unknown.* Now, f_ε^* is unknown and we replace it by the estimator (2). We shall study in the following the new contrast

$$\gamma_n(t) = \frac{1}{n}\sum_{j=1}^{n}[\|t\|^2 - 2\tilde{v}_t(Y_j)] \quad \text{with} \quad \tilde{v}_t^*(u) = \frac{t^*(u)}{\tilde{f}_\varepsilon^*(-u)}. \tag{5}$$

We define our estimators by minimizing this contrast on the projection spaces S_m:

$$\hat{f}_m = \arg\min_{t\in S_m} \gamma_n(t) \tag{6}$$

or, equivalently,

$$\hat{f}_m = \sum_{l\in\mathbb{Z}} \hat{a}_{m,l}\varphi_{m,l} \quad \text{with} \quad \hat{a}_{m,l} = \frac{1}{n}\sum_{j=1}^{n}\tilde{v}_{\varphi_{m,l}}(Y_j). \tag{7}$$

It is sufficient to differentiate the contrast (5) to obtain formula (7). Actually, we should define $\hat{f}_m = \sum_{|l|\le K_n}\hat{a}_{ml}\varphi_{m,l}$ because we can estimate only a finite number of coefficients. If K_n is suitably chosen, 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 truncation see Comte et al. (2006).

The notation $\hat{f}_m$ is the same because the estimators coincide. Indeed, starting with (7), we have the following equalities:

$$\begin{aligned}\hat{f}_m(x) &= \sum_{l\in\mathbb{Z}}\frac{1}{n}\sum_{j=1}^{n}\tilde{v}_{\varphi_{m,l}}(Y_j)\varphi_{m,l}(x) = \frac{1}{2\pi}\sum_{l\in\mathbb{Z}}\left(\int \hat{f}_Y^*(-u)\frac{\varphi_{m,l}^*(u)}{\tilde{f}_\varepsilon(-u)}du\right)\varphi_{m,l}(x) \\ &= \frac{1}{2\pi}\sum_{l\in\mathbb{Z}}\langle\frac{\hat{f}_Y^*}{\tilde{f}_\varepsilon^*},\varphi_{m,l}^*\rangle\varphi_{m,l}(x) = \frac{1}{2\pi}\sum_{l\in\mathbb{Z}}\langle\left(\frac{\hat{f}_Y^*}{\tilde{f}_\varepsilon^*}\right)^*(-.),\varphi_{m,l}\rangle\varphi_{m,l}(x). \end{aligned}\tag{8}$$

This is the expression of the orthogonal projection on S_m of $(1/2\pi)(\hat{f}_Y^*/\tilde{f}_\varepsilon^*)^*(-.)$. Using (4) and (8) yields (3). In practice, the coincidence is not exact because the sums over $\mathbb{Z}$ are truncated.

To complete the estimation procedure, 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 indexes. To do this, we select the model which minimizes the following penalized criterion:

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n}\{\gamma_n(\hat{f}_m) + \mathrm{pen}(m)\} \tag{9}$$

where pen is a penalty term to be specified later. Finally, we consider $\hat{f}_{\hat{m}}$ as 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 2. *Let $p \geq 1$ 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

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

The extension from $p = 1$ to any integer p is straightforward and therefore the proof is omitted.
Moreover, we introduce the notations

$$\Delta(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2}du \text{ and } \Delta^0(m) = \frac{1}{2\pi}\left(\int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-1}du\right)^2 \tag{10}$$

and

$$\Delta_f(m) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|f_\varepsilon^*(u)|^2}du \text{ and } \Delta_f^0(m) = \frac{1}{2\pi}\left(\int_{-\pi m}^{\pi m} \frac{|f^*(u)|}{|f_\varepsilon^*(u)|}du\right)^2. \tag{11}$$

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

3.2. **Pointwise mean square risk.** First, we study quickly the so-called MSE, the pointwise mean square error of the estimator. Let us denote by f_m the orthogonal projection of f on S_m. Then we have the following decomposition:

$$\begin{aligned}
\mathbb{E}[(\hat{f}_m(x) - f(x))^2] &\leq 2(f_m(x) - f(x))^2 + 2\mathbb{E}[(\hat{f}_m(x) - f_m(x))^2] \\
&\leq 2(f_m(x) - f(x))^2 + 4\mathrm{Var}\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(-u)}du\right) \\
&\quad +4\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\hat{f}_Y^*(u)R(u)du\right)^2\right]
\end{aligned} \tag{12}$$

The first (squared bias term) and second (variance term) terms of the right-hand-side of (12) are the usual ones, and are also found when f_ε^* is known; they are studied in Butucea and Comte (2007). The last one is studied by analogous methods and Lemma 2, and is specific to the present context. We find that the following risk bound holds:

Proposition 1. *Consider model (1) under* **(A1)***, then* $\hat{f}_m$ *defined by (6) satisfies:*

$$\mathbb{E}[(\hat{f}_m(x) - f(x))^2] \leq 2\left(\frac{1}{2\pi}\int_{|t|\geq \pi m} |f^*(t)|dt\right)^2 + \frac{C}{n}\inf(\int |f_Y^*|\Delta(m), \Delta^0(m)) + C'\frac{\Delta_f^0(m)}{M}$$

It follows that, as f belongs to $\mathbb{L}^2(\mathbb{R})$, if $\int |f_Y^*| < +\infty$ and $M \geq n$, then the risk bound obtained in Proposition 1 is the same as the one we get when f_ε^* is known. Indeed,

$$\Delta_f^0(m) \leq \|f\|^2\Delta(m) \text{ and } \Delta_f^0(m) \leq \Delta^0(m).$$

This is summarized by the Corollary:

Corollary 1. *Consider model (1) under* **(A1)***. Assume moreover that* $\int |f_Y^*| < +\infty$ *and* $M \geq n$*. Then* $\hat{f}_m$ *defined by (6) satisfies:*

$$\mathbb{E}[(\hat{f}_m(x) - f(x))^2] \leq 2\left(\frac{1}{2\pi}\int_{|t|\geq \pi m} |f^*(t)|dt\right)^2 + \frac{K}{n}\inf(\Delta(m), \Delta^0(m)), \tag{13}$$

where K *is a constant depending on* $\int |f_Y^*|$ *and* $\|f\|$*.*

3.3. **Pointwise rates under regularity conditions.** Assumption **(A1)** is generally strengthened by a parametric description of the rate of decrease of f_ε^* written as follows:

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

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

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

$$\mathcal{A}_{\delta,r,a}(l) = \{f \text{ density on } \mathbb{R} \text{ and } \int |f^*(x)|^2(x^2+1)^\delta \exp(2a|x|^r)dx \leq l\} \tag{14}$$
$$\text{with } r \geq 0, a > 0, \ \ \delta \in \mathbb{R} \ \text{ and } \delta > 1/2 \text{ if } r = 0, \ l > 0.$$

When $r > 0$, the function f is known as supersmooth, and as 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 $s = 0$.

Remark. If f_ε^* satisfies **(A2)** and f belongs to $\mathcal{A}_{\delta,r,a}(l)$ as defined in (14), then

$$\left(\int |f_Y^*|\right)^2 = \left(\int |f_\varepsilon^* f^*|\right)^2 \leq k_1^2 l \int (x^2+1)^{-(\gamma+\delta)}\exp(-2b|x|^s - 2a|x|^r)dx.$$

Thus definition (14) implies that $\int |f_Y^*| < +\infty$.

The optimality (minimaxity) of the rates resulting from (13) for known f_ε^*, when f_ε^* satisfies **(A2)** and f belongs to $\mathcal{A}_{\delta,r,a}(l)$, has been studied in Fan (1991), Butucea (2004) and Butucea and Tsybakov (2007).

More generally, we can see that, if f_ε^* satisfies **(A2)** and if $f \in \mathcal{A}_{\delta,r,a}(l)$:

$$\Delta_f^0(m) \leq \frac{l}{2\pi k_0^2}\int_{-\pi m}^{\pi m}(x^2+1)^{\gamma-\delta}\exp\left(2b|x|^s - 2a|x|^r\right)dx.$$

Now, we can combine the rates related to M with standard pointwise rates (see Table 1 in Lacour (2006) or Table 2 in Butucea and Comte (2007)) and we obtain the Table 2 here.

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

TABLE 1. Rates of convergence for the MSE if f_ε^* satisfies **(A2)** and $f \in \mathcal{A}_{\delta,r,a}(l)$.

We discuss the case $r>0, s>0$ for the integrated risk only, and thus omit this part of the study here. The principle would be the same, with slightly different orders. See also Lacour (2006).

3.4. **Bound on the MISE.** We shall study in more detail the integrated mean square risk, which is slightly simpler. Indeed now, by Pythagoras theorem, we have

$$\|f-\hat{f}_m\|^2 = \|f-f_m\|^2 + \|f_m-\hat{f}_m\|^2. \tag{15}$$

Moreover, writing $\hat{f}_m - f_m$ according to (4) 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

$$\|f_m-\hat{f}_m\|^2 \leq \frac{1}{\pi}\int_{-\pi m}^{\pi m}|\hat{f}_Y^*(u)|^2|R(u)|^2du + \frac{1}{\pi}\int_{-\pi m}^{\pi m}\frac{|\hat{f}_Y^*(u)-f_Y^*(u)|^2}{|f_\varepsilon^*(u)|^2}du. \tag{16}$$

The last term of the right-hand-side of (16) is the usual term that is found when f_ε^* is known, and the first one is specific to the present framework.

We can prove the following result:

Proposition 2. *Consider model (1) under* **(A1)***, then $\hat{f}_m$ defined by (6) satisfies:*

$$\mathbb{E}(\|\hat{f}_m-f\|^2) \leq \|f_m-f\|^2 + C\frac{\Delta(m)}{n} + C'\frac{\Delta_f(m)}{M} \tag{17}$$

with C and C' numerical constants.

The first two terms in the right-hand-side of (17) 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 $|f^*(x)| \leq 1$, we have $\Delta_f(m) \leq \Delta(m)$. It follows that for any $M \geq n$, then $\mathbb{E}\|\hat{f}_m - f\|^2 \leq \|f_m - f\|^2 + C\Delta(m)/n$ and we recover the usual risk bound for deconvolution estimation when f_ε^* is known. Therefore, in all cases, the condition $M \geq n$ ensures that the rate of the estimator is the same as when f_ε^* was known.

3.5. **Discussion about the resulting rates.** 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 (14). It is then possible to evaluate orders for the different terms involved in the bound (17).

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

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

The other terms are evaluated in the following lemma proved in Section 6.

Lemma 3. *If f_ε^* satisfies Assumption* **(A2)** *then*
(1) $\Delta(m) \lesssim (\pi m)^{2\gamma+1-s} e^{2b(\pi m)^s}$,
(2) $\Delta_f(m) \lesssim (\pi m)^{(1+2\gamma-s)\wedge 2(\gamma-\delta)_+} e^{2b(\pi m)^s}\mathbb{1}_{\{s>r\}} + (\pi m)^{2(\gamma-\delta)_+} e^{2(b-a)(\pi m)^s}\mathbb{1}_{\{r=s,b\geq a\}}$
$+\mathbb{1}_{\{r>s\}\cup\{r=s,b<a\}}$.

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

Proposition 3. *Assume that* **(A2)** *holds and that $f \in \mathcal{A}_{\delta,r,a}(l)$ given by (14). If $s = 0$ (ordinary smooth noise) and $r = 0$ (ordinary smooth function f), then*

$$\mathbb{E}\|\hat{f}_m - f\|^2 \leq C_0 m^{-2\delta} + C\frac{m^{2\gamma+1}}{n} + C'\frac{m^{2(\gamma-\delta)_+}}{M}$$

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

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 $M \geq n^{\frac{2(\gamma\vee\delta)}{2\gamma+2\delta+1}}$. This bound is tighter than $M \geq n$.

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.

Proposition 4. *Assume that* **(A2)** *holds and that $f \in \mathcal{A}_{\delta,r,a}(l)$ given by (14). If $s > 0$ (supersmooth noise) and $r = 0$ (ordinary smooth function f), then*

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

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

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 $M \geq n(\log n)^{-\frac{s+[2(\delta\wedge\gamma)+1-s]_+}{s}}$.

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 5. *Assume that* **(A2)** *holds and that $f \in \mathcal{A}_{\delta,r,a}(l)$ given by (14). If $s = 0$ (ordinary smooth noise) and $r > 0$ (supersmooth function f), then*

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

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

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

This is summarized in Table 2.

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

TABLE 2. 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. 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 6. *Assume that* **(A2)** *holds and that $f \in \mathcal{A}_{\delta,r,a}(l)$ given by (14). 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.

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 $b < a$. It follows from Theorem 3.1 in Lacour (2006) that

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

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 Theorem 3.1 in Lacour (2006)) and a choice m_0 such that

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

Case $r > s$. We define $k = \lceil (r/s - 1)^{-1}\rceil - 1$. There exist coefficients d_i recursively defined (see Theorem 3.1 in Lacour (2006)) and a choice m_0 such that

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

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

Proposition 7. *Let*

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

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

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

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

Proof of Proposition 7. The proof of the lower bound (for the additional problem of estimating f_ε) given by Neumann (1997) can be used for the study of the pointwise risk. Indeed it suffices to use the same hypothesis functions $f_{X,N,1}$ and $f_{X,N,2}$ shifted at point x and to compute the distance $|f_{X,N,1}(0) - f_{X,N,2}(0)|^2$. These functions have been adjusted to deal with the integrated risk but in the case where the two risks have the same order, they can suit. 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 both the integrated and the pointwise risk. □

4. Adaptation

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 in case $M \geq n$, so that our aim is to preserve here the rate corresponding to the case where f_ε^* is known. We consider thus the estimator $\hat{f}_{\hat{m}}$ defined by (9) where we have to define the penalty pen(.).
We will work under Assumption **(A2)** and the following one, concerning the collection of models $\mathcal{M}_n = \{1, 2, \ldots, m_n\}$:

$$\textbf{(A3)} \quad \exists \alpha \in]0,1[, \beta \in]0,1/2[, (\pi m_n)^{2\gamma} e^{2b(\pi m_n)^s} \lesssim M^{1-\alpha} \text{ and } n \lesssim M \lesssim \exp(n^{1/2-\beta}).$$

The first Inequality in **(A3)** together with **(A2)** implies that $\forall x \in [-\pi m_n, \pi m_n], |f_\varepsilon^*(m_n)|^{-2} \lesssim M^{1-\alpha}$.

If we choose $M = n$, we can see that Assumption **(A3)** is ensured if pen(.) is bounded over $\mathcal{M}_n$ in the case $b = s = 0$ (that is if $\pi m_n \leq n^{1/(2\gamma+1)}$) and if $\pi m_n \leq (\log(n)/(2b+1))^{1/s}$ for $b > 0, s > 0$.

Clearly, it is difficult to choose $\mathcal{M}_n$ and thus m_n so that **(A3)** is fulfilled. This is a problem in the practical procedure which requires an explicit upper bound m_n of $\mathcal{M}_n$. Diggle and Hall (1993) suggest that regression methods applied to $\log(|\hat{f}_\varepsilon^*|)$ may deliver some estimates of the parameter γ in the ordinary smooth case. An estimator of s is proposed in a semi-parametric framework in a recent work of Butucea et al. (2008): if it is known that f_ε is supersmooth, this strategy may be used to estimate s. Lastly, one can think of taking m_n of order $(|\hat{f}_\varepsilon^*|^2)^{-1}(\sqrt{M})$, where the exponent -1 here denotes the reciprocal function.

We can provide another set of assumptions ensuring **(A3)**. Assume that $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, $s \in [\underline{s}, \overline{s}]$ and $b \in [\underline{b}, \overline{b}]$ with $\underline{s} > 0$, $\underline{b} > 0$ whenever $s > 0$, $b > 0$. And consider also the following assumption: $M \lesssim \exp(n^{1/4})$ and

Case $(b, s) = (0, 0)$**:** (ordinary smooth noise):

$$\textbf{(A4)} \quad \begin{cases} \mathcal{M}_n = \left\{m, (\pi m)^{2\underline{\gamma}+1} \leq n\right\} \\ \\ M \geq n^{\frac{2\overline{\gamma}+1}{2\underline{\gamma}+1}} \end{cases} \quad \text{or} \quad \begin{cases} \mathcal{M}_n = \left\{m, (\pi m) \leq \sqrt{n}\right\} \\ \\ M \geq n^{\overline{\gamma}+1/2}. \end{cases}$$

Case $b > 0, s > 0$**:** (supersmooth noise)

$$\textbf{(A4)} \quad \begin{cases} \mathcal{M}_n = \left\{m, (\pi m)^{2\underline{\gamma}+1-\overline{s}} e^{2\underline{b}(\pi m)^{\underline{s}}} \leq n\right\} \\ \\ M \geq \exp\left\{2 \dfrac{2\overline{b}}{(2\underline{b})^{\underline{s}}} (\log n)^{\overline{s}/\underline{s}}\right\} \end{cases}$$

Then **(A4)** may require greater values of M, but also ensures **(A3)**.

The simulation experiments of Section 5 are very useful here, to study the influence of the value of M and the importance of the set $\mathcal{M}_n$. It illustrates that we obtain very good results in practice, even when arbitrary limitations are set on m_n or M.

We can prove the preliminary result given in Theorem 1. Note that the proof of this theorem is not standard in the present setting for the following reason. Usually, the contrast decomposition gives two types of terms:

- supremum of centered empirical processes (which in the independent bounded case are controlled thanks to Talagrand's type results): these terms impose the form of the penalty function.
- residual terms which are of negligible orders and which are not centered.

Here, the residual terms which appear, even when controlled by using Lemma 2, are not negligible and have a weight in the penalty function. This makes the proof quite difficult.

Theorem 1. *Assume that assumptions* **(A2)** *and* **(A3)** *are fulfilled and consider the estimator* $\hat{f}_{\hat{m}}$ *defined by (6) and (9) with*

$$\mathrm{pen}(m) = K_0(\pi m)^{[s-(1-s)_+/2]_+}\frac{\Delta(m)}{n}.$$

Then there exists $C > 0$ *such that*

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

where f_m *is the orthogonal projection of* f *on* S_m*.*

The presence of $|f_\varepsilon^*|$ and s in the penalty is not admissible as it is an unknown quantity. That is why we state the following theorem.

Theorem 2. *Assume that assumptions* **(A2)** *and* **(A3)** *are fulfilled and that* $s \leq \overline{s}$ *for some given upper value* $\overline{s}$*. Consider the estimator* $\tilde{f} = \hat{f}_{\hat{\hat{m}}}$ *defined by (6) and*

$$\hat{\hat{m}} = \arg\min_{m\in\mathcal{M}_n}\{\gamma_n(\hat{f}_m) + \widetilde{\mathrm{pen}}(m)\} \tag{18}$$

with

$$\widetilde{\mathrm{pen}}(m) = K_1(\pi m)^{[\overline{s}-(1-\overline{s})_+/2]_+}\frac{\int_{-\pi m}^{\pi m}|\tilde{f}_\varepsilon^*|^{-2}}{n}.$$

Then there exists $C > 0$ *such that*

$$\mathbb{E}\|\tilde{f} - f\|^2 \leq 4 \inf_{m\in\mathcal{M}_n}\{\|f_m - f\|^2 + \mathbb{E}\widetilde{\mathrm{pen}}(m)\} + \frac{C}{n}$$

Note that the restriction $s \leq 2$ is very classical, so that $\overline{s} = 2$ is generally suitable.

Concluding Remarks. As we can prove that $\mathbb{E}\widetilde{\mathrm{pen}}(m) \lesssim (\pi m)^{[\overline{s}-(1-\overline{s})_+/2]_+}\Delta(m)/n$, it follows from Theorem 2 that $\tilde{f}$ automatically reaches the same rate as when f_ε is known if $\overline{s} = s$ or if $\overline{s} = 0$ (and thus $s = 0$ and this is known). For a discussion about the optimality of these rates (which holds in most cases), see Comte et al. (2006).

In particular, in the case of ordinary smooth errors, the procedure is data driven and reaches the optimal rate, provided that M is taken large enough (Assumption **(A4)** case 1.

5. Simulations

Let us describe the estimation procedure. As noticed in (7), for each m, the estimator $\hat{f}_m$ of f can be written

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

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

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

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 $m \in \mathcal{M}_n$ 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 $\gamma_n(\hat{f}_m) + \widetilde{\text{pen}}(m) = -\sum_l |\hat{a}_{m,l}|^2 + \widetilde{\text{pen}}(m)$.
- Select the $\hat{m}$ which minimizes $\gamma_n(\hat{f}_m) + \widetilde{\text{pen}}(m)$.
- Compute $\tilde{f} = \sum_{|l|\leq K_n} \hat{a}_{\hat{m},l}\varphi_{\hat{m},l}$.

Clearly, the use of FFT makes the procedure very fast. The penalty is chosen according to Theorem 2 with $\bar{s} = 2$. Indeed densities with $s > 2$ are difficult to express in a closed form whereas usual densities all verify $0 \leq s \leq 2$. The constant K_1 is chosen equal to $1/2$ after intensive simulation experiments. However other values of K_1 can suit and, empirically, the procedure seems rather robust with respect to the choice of this constant. Thus, in all the examples below, we take:

$$\widetilde{\text{pen}}(m) = \frac{1}{2n}(\pi m)^2\int_{-\pi m}^{\pi m} |\tilde{f}^*_\varepsilon|^{-2}.$$

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. As in Comte et al. (2007), we consider that m can be fractional. More precisely, we take here

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

for the estimation with unknown noise. Actually, the bound on m, which is crucial in theory, turns out to be of little importance for practical purposes. The procedure chooses the appropriate model (often very small) even if the maximal model is very large. That is why in all our experiments we choose to keep a maximal m with order $\sqrt{n}$. We also compute the estimator with known noise, replacing $\tilde{f}_\varepsilon$ by f_ε in the procedure. In this case, the choice of the maximal model has more impact, so 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 3 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	0.00225	0.01641
f_ε unknown	0.00619	0.03327

TABLE 3. 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 (2007). In his example 5.1, he considers a standard Gaussian noise and $X \sim \mathcal{N}(5, 9)$. Again we use $\mathcal{M}_n = \{m = k/(4\pi), k \in \mathbb{N}^*, k \leq \sqrt{n}\}$. The penalty for a known noise is $\text{pen}(m) = (\pi m)^3 \int_0^1 \exp\{(\pi m x)^2\}dx/(2n)$. As Johannes (2007) presents only boxplots and for the sake of comparison, we give the third quartile for the L^2 risk in Table 4. 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	0.34	0.20	0.08
$M = 100$	0.29	0.15	0.09
$M = 250$	0.27	0.13	0.08
$M = 500$	0.23	0.11	0.08

TABLE 4. Third quartile of the MISE $\times 100$ for the estimators of Johannes (2007) (left) and for the penalized estimator (right).

Now we compute estimators for different signal densities and different noises. For the sake of simplicity (and since the chosen model are here larger), we take now

$$\mathcal{M}_n = \left\{ m = \frac{k}{2\pi}, k \in \mathbb{N}^*, k \leq \sqrt{n} \right\}$$

(for both known and unknown noise). 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 $\delta = 1/2, r = 2$)

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
f \ f_ε		Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.
Laplace	f_ε known	2.185	2.250	1.261	1.168	0.836	0.924	0.583	0.633
	$M = \lfloor\sqrt{n}\rfloor$	4.868	4.791	2.811	2.845	1.777	1.781	1.153	1.109
	$M = n$	5.107	5.114	2.892	2.876	1.757	1.748	1.090	1.110
Mixed	f_ε known	1.001	0.945	0.603	0.554	0.278	0.274	0.177	0.202
Gamma	$M = \lfloor\sqrt{n}\rfloor$	0.971	1.025	0.751	0.777	0.454	0.472	0.232	0.230
	$M = n$	1.037	1.039	0.745	0.765	0.467	0.500	0.222	0.218
Cauchy	f_ε known	1.072	0.979	0.468	0.475	0.341	0.251	0.243	0.137
	$M = \lfloor\sqrt{n}\rfloor$	1.276	1.343	0.791	0.802	0.400	0.398	0.189	0.194
	$M = n$	1.266	1.362	0.762	0.782	0.363	0.364	0.172	0.172
Gaussian	f_ε known	0.810	0.589	0.771	0.287	0.500	0.191	0.373	0.134
	$M = \lfloor\sqrt{n}\rfloor$	1.045	1.114	0.397	0.346	0.241	0.181	0.139	0.170
	$M = n$	0.904	0.986	0.252	0.256	0.150	0.182	0.100	0.094

TABLE 5. 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 $\mathrm{pen}(m) = 4(\pi m + (2/(3\lambda^2))(\pi m)^3 + (1/(5\lambda^4))(\pi m)^5)/n$.

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 $\mathrm{pen}(m) = 0.5(\pi m)^3 \int_0^1 e^{(\lambda\pi m x)^2} dx/n$.

The results are given in Table 5 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. We can also observe that, as expected, the risk decreases when M increases. The cases where the risk is larger for $M = n$ 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.

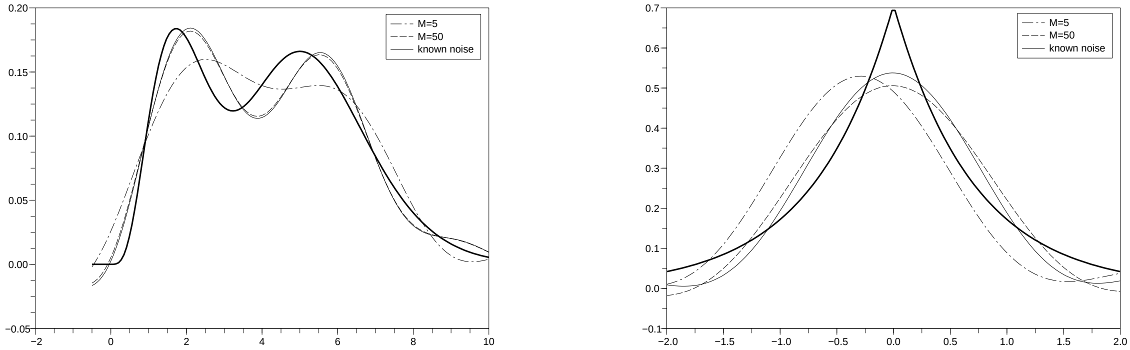

FIGURE 1. True function f (bold line) and estimators for $n = 500$. Left: mixed Gamma density with Laplace noise. Right : Laplace density with Gaussian noise

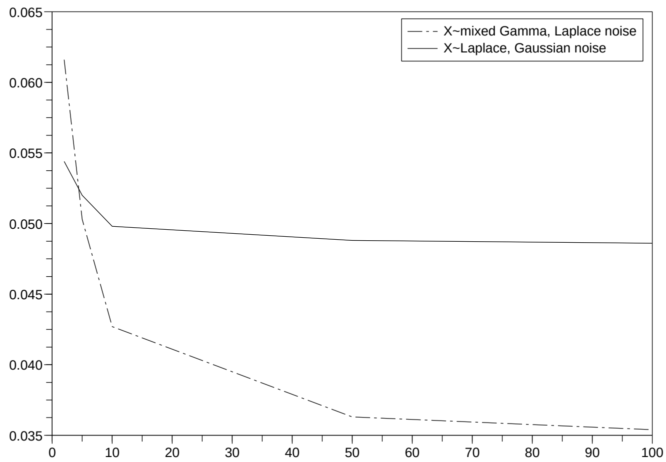

FIGURE 2. MISE against M for $n = 100$ in two cases

Figure 2 shows the decrease of the integrated risk in these two cases. These curves confirm the theoretical result since the rate of convergence for a fixed n is M^{-1} in the first case and $(\log M)^{-2}$ in the second case.

6. PROOFS

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

6.1. **Proof of Lemma 1.** It is sufficient to prove the first assertion. First we write that $v_t(Y_j) = (1/2\pi)\int e^{iY_j u} t^*(u)/f_\varepsilon^*(-u)du$ so that

$$E[v_t(Y_j)|X_j] = \frac{1}{2\pi}\int E[e^{iY_j u}|X_j]\frac{t^*(u)}{f_\varepsilon^*(-u)}du.$$

By using the independence between X_j and ε_j, we compute

$$E[e^{iY_j u}|X_j] = E[e^{iX_j u}e^{i\varepsilon_j u}|X_j] = e^{iX_j u}E[e^{i\varepsilon_j u}] = e^{iX_j u}f_\varepsilon^*(-u).$$

Then

$$E[v_t(Y_j)|X_j] = \frac{1}{2\pi}\int e^{iX_j u}f_\varepsilon^*(-u)\frac{t^*(u)}{f_\varepsilon^*(-u)}du = \frac{1}{2\pi}\int e^{iX_j u}t^*(u)du = t(X_j).$$

$\square$

6.2. **Proof of Proposition 1.** We start from Inequality (12). It follows from Butucea and Comte (2007) that:

$$\text{Var}\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(-u)}du\right) \le \frac{1}{2\pi n}\inf(\int |f_Y^*|\Delta(m), \Delta^0(m)), \tag{19}$$

and (also, to see this, use (4) and $(f - f_m)(x) = (1/2\pi)(f^* - f_m^*)^*(-x)$),

$$(f_m(x) - f(x))^2 \le \left(\frac{1}{2\pi}\int_{|t|\ge \pi m}|f^*(t)|dt\right)^2. \tag{20}$$

For the remaining term in (12), we write first:

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}\hat{f}_Y^*(u)R(u)du\right)^2\right] &\le 2\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}(\hat{f}_Y^*(u) - f_Y^*(u))R(u)du\right)^2\right]\\
&\quad +2\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}f_Y^*(u)R(u)du\right)^2\right]\\
&:= 2T_1 + 2T_2.
\end{aligned}$$

Then we find

$$\begin{aligned}
T_1 &= \frac{1}{4\pi^2}\iint e^{ix(u-v)}\text{cov}(\hat{f}_Y^*(u), \hat{f}_Y^*(v))\mathbb{E}(R(u)\bar{R}(v))dudv\\
&\le \frac{1}{4\pi^2 n}\iint |f_Y^*(u-v)|\sqrt{\mathbb{E}(|R(u)|^2)\mathbb{E}(|R(v)|^2)}dudv\\
&\lesssim \frac{1}{4\pi^2 n}\iint |f_Y^*(u-v)|\frac{1}{|f_\varepsilon(u)f_\varepsilon(v)|}dudv
\end{aligned}$$

by using Lemma 2. This term is clearly bounded by $\Delta^0(m)$. Moreover writing it as

$$\iint \frac{\sqrt{|f_Y^*(u-v)|}}{|f_\varepsilon(u)|}\frac{\sqrt{|f_Y^*(u-v)|}}{|f_\varepsilon(v)|}dudv$$

and using first the Schwarz Inequality, and second the Fubini Theorem yields the bound $\int |f_Y^*|\Delta(m)$. Therefore

$$\mathbb{E}\left[\left(\frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{ixu}(\hat{f}_Y^*(u)-f_Y^*(u))R(u)du\right)^2\right] \lesssim \frac{1}{2\pi n}\inf\left(\int |f_Y^*|\Delta(m), \Delta^0(m)\right), \tag{21}$$

and thus it has the same order as the usual variance term. Lastly,

$$\begin{aligned} T_2 &\le \frac{1}{4\pi^2}\iint_{|u|,|v|\le \pi m} |f_Y^*(u)f_Y^*(v)|\sqrt{\mathbb{E}(|R(u)|^2)\mathbb{E}(|R(v)|^2)}dudv \\ &\le \frac{1}{4\pi^2}\left(\int_{-\pi m}^{\pi m} |f_Y^*(u)|\sqrt{\mathbb{E}(|R(u)|^2)}du\right)^2 \\ &\lesssim \frac{1}{4\pi^2 M}\left(\int_{-\pi m}^{\pi m} |f_Y^*(u)|\frac{1}{|f_\varepsilon^*(u)|^2}du\right)^2 \\ &= \frac{1}{4\pi^2 M}\left(\int_{-\pi m}^{\pi m} \frac{|f^*(u)|}{|f_\varepsilon^*(u)|}du\right)^2 = \frac{\Delta_f^0(m)}{2\pi M}. \end{aligned} \tag{22}$$

Inserting the bounds (19) to (22) in Inequality (12), we obtain the result of Proposition 1. □

6.3. **Proof of Proposition 2.** We start from (16) and take the expectation:

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

Applying Lemma 2 yields:

$$\begin{aligned} \mathbb{E}(\|f_m-\hat{f}_m\|^2) &\le \frac{2}{\pi}\int_{-\pi m}^{\pi m} \mathbb{E}(|\hat{f}_Y^*(u)-f_Y^*(u)|^2)\mathbb{E}(|R(u)|^2)du \\ &\quad +\frac{2}{\pi}\int_{-\pi m}^{\pi m} |f^*(u)|^2|f_\varepsilon^*(u)|^2\mathbb{E}|R(u)|^2du + 2\frac{\Delta(m)}{n} \\ &\lesssim \int_{-\pi m}^{\pi m} n^{-1}|f_\varepsilon^*(u)|^{-2}du \\ &\quad +\int_{-\pi m}^{\pi m} |f^*(u)|^2|f_\varepsilon^*(u)|^2\frac{M^{-1}}{|f_\varepsilon^*(u)|^4}du + \frac{\Delta(m)}{n} \\ &\lesssim \frac{1}{M}\int_{-\pi m}^{\pi m}\frac{|f^*(u)|^2}{|f_\varepsilon^*(u)|^2}du + \frac{\Delta(m)}{n} \end{aligned} \tag{23}$$

By gathering (15) and (23), we obtain the result. □

6.4. **Proof of Lemma 3.** The proof of the first result is omitted. It is obtained by distinguishing the cases $s > 2\gamma+1$ and $s \le 2\gamma+1$ and with standard evaluations of integrals.

For the second point, we first remark that $\Delta_f(m) \le \Delta(m)$. Next, using Assumption **(A2)**,

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

Then, if $s > r$,

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

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

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

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

6.5. **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 $\gamma_n(\tilde f) + \mathrm{pen}(\hat m) \le \gamma_n(f_m) + \mathrm{pen}(m)$, we have

$$\begin{aligned}\|\hat f_{\hat m} - f\|^2 &\le \|f_m - f\|^2 + 2\nu_n(\hat f_{\hat m} - f_m) + \mathrm{pen}(m) - \mathrm{pen}(\hat m)\\ &\le \|f_m - f\|^2 + 2\|\tilde f - f_m\| \sup_{t\in B(m,\hat m)}\nu_n(t) + \mathrm{pen}(m) - \mathrm{pen}(\hat m)\end{aligned}$$

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

$$\|\hat f_{\hat m} - f\|^2 \le \|f_m - f\|^2 + \frac{1}{4}\|\hat f_{\hat m} - f_m\|^2 + 4\sup_{t\in B(m,\hat m)}\nu_n^2(t) + \mathrm{pen}(m) - \mathrm{pen}(\hat m). \tag{24}$$

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

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

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

$$4p(m,m') \le \mathrm{pen}(m) + \mathrm{pen}(m') \tag{25}$$

then

$$\mathbb{E}\|\hat f_{\hat m} - f\|^2 \le 3\|f_m - f\|^2 + 8\mathbb{E}[\sup_{t\in B(m,\hat m)}\nu_n^2(t) - p(m,\hat m)] + 4\mathrm{pen}(m). \tag{26}$$

With Lemma 1 in mind, $\nu_n(t)$ can be split into two terms : $\nu_n(t) = \nu_{n,1}(t) + S_n(t)$ with

$$\left\{ \begin{array}{l} \nu_{n,1}(t) = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} \{v_t(Y_j) - \mathbb{E}[v_t(Y_j)]\} \\ S_n(t) = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} (\tilde{v}_t - v_t)(Y_j), , \end{array} \right. \tag{27}$$

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

Proposition 8. *Let $p_1(m,m') = K(\pi m'')^{[s-(1-s)_+/2]_+}\Delta(m'')/n$ where $\Delta(m)$ is defined in* (10) *and $m'' = \max(m,m')$ and K is a constant. Then, under assumptions of Theorem 1, there exists a positive constant C such that*

$$\mathbb{E}_0 := \mathbb{E}\left(\left[\sup_{t\in B(m,\hat{m})} \nu_{n,1}^2(t) - p_1(m,\hat{m})\right]_+\right) \leq \frac{C}{n}. \tag{28}$$

Note that Theorem 1 in Comte et al. (2006) is proved under the assumption that the penalty is bounded, but it is easy to check that it also holds under **(A3)** (and the assumption $M \lesssim \exp(n^{1/2-\beta})$ is used here).

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

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

Now, we prove in the following subsections the result:

Proposition 9. *We denote here by $m^* = m \vee \hat{m}$. Under assumptions of Theorem 1, there exists a positive constant C such that*

$$\mathbb{E}_1 := \mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,1}(t)|^2 - C_1\frac{\Delta(m^*)}{n}\right) \leq \frac{C}{n}. \tag{29}$$

$$\mathbb{E}_2 := \mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,2}(t)|^2 - C_2\frac{\Delta_f(m^*)}{M}\right) \leq \frac{C}{n}. \tag{30}$$

$$\mathbb{E}_3 := \mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,3}(t)|^2 - p_3(m,\hat{m})\right)_+ \leq \frac{C}{n}. \tag{31}$$

with $p_3(m,m') = K'(\pi m'')^{[s-(1-s)_+/2]_+}\Delta(m'')/M$, $m'' = \max(m,m')$.

$$\mathbb{E}_4 := \mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,4}(t)|^2\right) \leq \frac{C}{n}. \tag{32}$$

It follows that

$$\mathbb{E}[\sup_{t\in B(m,\hat{m})} \nu_n^2(t) - p(m,\hat{m})]_+ \quad \leq \quad 5(\mathbb{E}_1 + \mathbb{E}_2 + \mathbb{E}_3 + \mathbb{E}_4 + \mathbb{E}_0) \tag{33}$$

as soon as

$$5\left(\frac{\Delta(m")}{n} + \frac{\Delta_f(m")}{n} + p_3(m,m') + p_1(m,m')\right) \leq p(m,m')$$

for all m, m' in $\mathcal{M}_n$. Therefore, the choice $p(m,m') = K"(\pi m'')^{[s-(1-s)_+/2]_+}\Delta(m'')/n$ for a numerical constant $K"$ large enough, is suitable. The choice of pen(.) given in Theorem 1 ensures then that (25) holds true.
Now, gathering (28), (29)–(32) and (33) yields

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

which, together with (26), ends the proof of Theorem 1. □

6.6. **Proof of Proposition 9.**

6.6.1. *Study of $R_{n,1}(t)$ and proof of (29).* We recall that m^* is the maximum $\max(m, \hat{m})$ and we define $\Omega(x)$ the set $\Omega(x) = \Omega_1(x) \cap \Omega_2(x)$ where

$$\Omega_1(x) = \{|\hat{f_Y^*}(x) - f_Y^*(x)| \le n^{\alpha/8-1/2} M^{\alpha/4}\} \text{ and } \Omega_2(x) = \{|R(x)| \le M^{\alpha/8-1/2}/|f_\varepsilon^*(x)|^2)\},$$

for $\alpha \in (0,1)$ as defined in **(A3)**.

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

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

so that

$$\sup_{t \in B(m,\hat{m})} |R_{n,1}(t)|^2 \le \frac{1}{4\pi^2} \int_{-\pi m^*}^{\pi m^*} |\hat{f_Y^*} - f_Y^*|^2 |R|^2 \mathbb{1}_\Omega + \frac{1}{4\pi^2} \int_{-\pi m^*}^{\pi m^*} |\hat{f_Y^*} - f_Y^*|^2 |R|^2 \mathbb{1}_{\Omega^c}$$

On the one hand

$$\begin{aligned}
\int_{-\pi m^*}^{\pi m^*} |\hat{f_Y^*} - f_Y^*|^2 |R|^2 \mathbb{1}_\Omega &\le \int_{-\pi m^*}^{\pi m^*} n^{\alpha/4-1} M^{\alpha/2} M^{\alpha/4-1} |f_\varepsilon^*|^{-4} \\
&\le n^{\alpha/4-1} M^{3\alpha/4-1} \int_{-\pi m^*}^{\pi m^*} M^{1-\alpha} |f_\varepsilon^*(x)|^{-2} dx \\
&\lesssim n^{\alpha/4-1} M^{-\alpha/4} \Delta(m^*) \lesssim \frac{\Delta(m^*)}{n} \left(\frac{n}{M}\right)^{\alpha/4} \lesssim \frac{\Delta(m^*)}{n}
\end{aligned}$$

On the other hand

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

But, using the Markov inequality,

$$\begin{aligned}
\mathbb{P}(\Omega(x)^c) &\le n^{-p(\alpha/8-1/2)} M^{-p\alpha/4} \mathbb{E}|\hat{f_Y^*} - f_Y^*|^p + M^{-2p(\alpha/8-1/2)} |f_\varepsilon^*(x)|^{4p} \mathbb{E}|R|^{2p} \\
&\le n^{-p\alpha/8} M^{-p\alpha/4} + M^{-p\alpha/4} \lesssim M^{-p\alpha/4},
\end{aligned}$$

then

$$\begin{aligned}
\mathbb{E}(\int_{-\pi m^*}^{\pi m^*} |\hat{f_Y^*} - f_Y^*|^2 |R|^2 \mathbb{1}_{\Omega^c}) &\lesssim n^{-1} M^{-\alpha} \Delta(m_n) M^{-p\alpha/8}) \\
&\lesssim \frac{m_n M^{1-\alpha}}{Mn} M^{1-\alpha-p\alpha/8} \le \frac{M^{1+1/(2\tilde{\gamma})-2\alpha-\alpha p/8}}{n}
\end{aligned}$$

where $1/(2\tilde{\gamma}) = 1/(2\gamma)$ if $s = 0, \gamma > 0$ and $1/(2\tilde{\gamma}) = 1$ if $\gamma = 0, s > 0$.
We choose p large enough ($p \ge 8(1/(2\tilde{\gamma}) + 1 - 2\alpha)/\alpha$) so that $M^{1+1/(2\tilde{\gamma})-2\alpha-p\alpha/8} = O(1)$.
We obtain (29). □

6.6.2. *Study of $R_{n,2}(t)$ and proof of (30).* The following result obviously holds $\mathbb{E}[|f_\varepsilon^* - \hat{f}_\varepsilon^*|^p] \lesssim M^{-p/2}$. Moreover, let

$$\Xi(x) = \{|f_\varepsilon^*(x) - \hat{f}_\varepsilon^*(x)| \le M^{-\omega} \text{ and } |R(x)| \le M^{-\omega}/|f_\varepsilon^*(x)|^2\},$$

where $0 < \omega < 1/2$. We can bound the term $|R_{n,2}(t)|^2$ in the following way

$$\sup_{t\in S_m+S_{\hat{m}}} |R_{n,2}(t)|^2 \le \frac{1}{4\pi^2}\int_{-\pi m^*}^{\pi m^*} |f^*|^2|f_\varepsilon^* - \hat{f}_\varepsilon^*|^2|R|^2\mathbb{1}_\Xi + \frac{1}{4\pi^2}\int_{-\pi m^*}^{\pi m^*} |f^*|^2|f_\varepsilon^* - \hat{f}_\varepsilon^*|^2|R|^2\mathbb{1}_{\Xi^c}$$

On the one hand

$$\begin{aligned}\int_{-\pi m^*}^{\pi m^*} |f^*|^2|f_\varepsilon^* - \hat{f}_\varepsilon^*|^2|R|^2\mathbb{1}_\Xi &\le \int_{-\pi m^*}^{\pi m^*} |f^*|^2 M^{-4\omega}|f_\varepsilon^*|^{-4} \\ &\le \int_{-\pi m^*}^{\pi m^*} |f^*|^2 M^{-4\omega}|f_\varepsilon^*|^{-2}M^{1-\alpha} \\ &\lesssim \frac{\Delta_f(m^*)}{M}(M^{2-4\omega-\alpha}) \lesssim \frac{\Delta_f(m^*)}{M},\end{aligned}$$

as soon as ω satisfies $(2-\alpha)/4 \le \omega < 1/2$, e.g. we can take $0 < \omega = (2-\alpha)/4 < 1/2$. On the other hand

$$\begin{aligned}\mathbb{E}(\int_{-\pi m^*}^{\pi m^*} |f^*|^2|f_\varepsilon^* - \hat{f}_\varepsilon^*|^2|R|^2\mathbb{1}_{\Xi^c}) &\le \int_{-\pi m_n}^{\pi m_n} |f^*|^2\mathbb{E}^{1/4}(|f_\varepsilon^* - \hat{f}_\varepsilon^*|^8)\mathbb{E}^{1/4}(|R|^8)\mathbb{P}^{1/2}(\Xi^c) \\ &\lesssim \int_{-\pi m_n}^{\pi m_n} M^{-2}|f^*(x)|^2|f_\varepsilon^*(x)|^{-4}\mathbb{P}^{1/2}(\Xi(x)^c)dx \\ &\lesssim \Delta_f(m_n)M^{-1-\alpha}\|\mathbb{P}^{1/2}(\Xi^c)\|_\infty \lesssim m_n M^{-\alpha}\|\mathbb{P}^{1/2}(\Xi^c)\|_\infty \\ &\lesssim M^{1/(2\tilde{\gamma})-\alpha}\|\mathbb{P}^{1/2}(\Xi^c)\|_\infty.\end{aligned}$$

Then, using the Markov inequality,

$$\begin{aligned}\mathbb{P}(\Xi(x)^c) &\le M^{2p\omega}\mathbb{E}|f_\varepsilon^* - \hat{f}_\varepsilon^*|^{2p} + M^{2\omega p}|f_\varepsilon^*(x)|^{4p}\mathbb{E}|R|^{2p} \\ &\lesssim M^{p(2\omega-1)}.\end{aligned}$$

Thus

$$\mathbb{E}(\int_{-\pi m^*}^{\pi m^*} |\hat{f_Y^*} - f_Y^*|^2|R|^2\mathbb{1}_{\Xi^c}) \lesssim \frac{M^{1/(2\tilde{\gamma})-\alpha-p(1-2\omega)}}{n} \lesssim \frac{1}{M}$$

for $p \ge (1/(2\tilde{\gamma}) - \alpha)/(1-2\omega) = 2(1/(2\tilde{\gamma}) - \alpha)/\alpha$. This yields (30). □

6.6.3. *Study of $R_{n,3}(t)$ and proof of (31).* 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,

$$\mathbb{E}\left[\sup_{t\in B(m,\hat{m})} |R_{n,3}(t)|^2 - p_3(m,\hat{m})\right]_+ \le \sum_{m'\in\mathcal{M}_n} \mathbb{E}\left[\sup_{t\in B(m,m')} |R_{n,3}(t)|^2 - p_3(m,m')\right]_+$$

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 $\epsilon > 0$,*

$$\mathbb{E}[\sup_{r\in\mathcal{R}} |\nu_M(r)|^2 - (1+2\epsilon)H^2]_+ \le C\left(\frac{v}{M}e^{-K_1\epsilon\frac{MH^2}{v}} + \frac{B^2}{M^2C^2(\epsilon)}e^{-K_2 C(\epsilon)\sqrt{\epsilon}\frac{MH}{B}}\right) \tag{34}$$

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

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

Inequality (34) 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 of a finite dimensional linear space.

Let us determine B, H and v is our problem.
For $t \in S_m + S_{m'} = S_{m''}$,

$$\|F_t\|_\infty^2 \le \frac{1}{4\pi^2}\int_{-\pi m''}^{\pi m''} \frac{|f^*(x)|^2}{|f_\varepsilon^*(x)|^2}dx \int_{-\pi m''}^{\pi m''} |t^*(-x)|^2 dx.$$

Then $\sup_{t\in B(m,m')} \|F_t\|_\infty^2 \le \Delta_f(m'')$ and we set

$$B = \sqrt{\Delta(m'')}.$$

If t belongs to $B(m,m')$, it can be written $t = \sum_{l\in\mathbb{Z}} a_{m'',l}\varphi_{m'',l}$ and

$$|R_{n,3}(t)|^2 \le \sum_{l\in\mathbb{Z}} a_{m'',l}^2 \sum_{l\in\mathbb{Z}} |\frac{1}{M}\sum_{j=1}^M [F_{\varphi_{m'',l}}(\varepsilon_{-j}) - \mathbb{E}(F_{\varphi_{m'',l}}(\varepsilon_{-j}))]|^2$$

As the ε_{-j} are i.i.d.,

$$\mathbb{E}(\sup_{t\in B(m,m')} |R_{n,3}(t)|^2) \le \sum_{l\in\mathbb{Z}} \mathrm{Var}\left(\frac{1}{M}\sum_{j=1}^M F_{\varphi_{m'',l}}(\varepsilon_{-j})\right) \le \frac{1}{M}\sum_{l\in\mathbb{Z}} \mathrm{Var}\left(F_{\varphi_{m'',l}}(\varepsilon_1)\right)$$

Now, using the Parseval formula,

$$\sum_{l\in\mathbb{Z}} \left|\frac{1}{2\pi}\int \frac{f^*(x)}{f_\varepsilon^*(x)}\varphi_{m'',l}^*(-x)e^{-ixu}dx\right|^2 = \frac{1}{2\pi}\int_{-\pi m''}^{\pi m''} \left|\frac{f^*(x)}{f_\varepsilon^*(x)}\right|^2 dx.$$

Then

$$\mathbb{E}(\sup_{t\in B(m,m')} |R_{n,3}(t)|^2) \le \frac{1}{M}\mathbb{E}\left(\sum_{l\in\mathbb{Z}} |F_{\varphi_{m'',l}}(\varepsilon_1)|^2\right) \le \frac{1}{2\pi M}\int_{-\pi m''}^{\pi m''} \left|\frac{f^*(x)}{f_\varepsilon^*(x)}\right|^2 dx \le \frac{\Delta_f(m'')}{M}$$

and we set $H = \sqrt{\Delta(m'')/M}$ as $\|f\|_1 = 1$.
Lastly, standard methods give $v = C\min(\Delta(m''), \|f_\varepsilon\|\sqrt{\Delta_2(m'')})$ with

$$\Delta_2(m) = \int \left|\frac{1}{f_\varepsilon^*}\right|^4$$

Then we insert these quantities in the Inequality given in Lemma 4. For the first term of the right-hand-side bound, we obtain

$$\sum_{m'\in\mathcal{M}_n} \frac{v}{M}\exp(-K_1\epsilon\frac{MH^2}{v}) \leq \frac{C}{M}$$

with adequate choices of ϵ. The study of this term is the same as in the proof of Theorem 1 in Comte et al. (2006) and is omitted here. The second term of the right-hand-side bound is less than (up to multiplicative constants):

$$\sum_{m'\in\mathcal{M}_n} \frac{\Delta(m'')}{M^2}\exp(-C\sqrt{M}) \leq M^{1/\tilde{\gamma}-1-\alpha}\exp(-C\sqrt{M}) \lesssim \frac{1}{M}.$$

Inserting the value of ϵ and applying Lemma 4 leads to

$$\sum_{m'\in\mathcal{M}_n} \mathbb{E}\left[\sup_{t\in B(m,m')} |R_{n,3}(t)|^2 - p_3(m,m')\right]_+ \leq \frac{C}{M}$$

which implies (31). □

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

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

and thus

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

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

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

for all $p \geq 1$. Then

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

Then as $\Delta_f(m_n) \leq \Delta(m_n) \leq m_n M \leq M^{1+1/(2\tilde{\gamma})}$, it is sufficient to take $p \geq 2+1/(2\alpha\tilde{\gamma}) > 0$ to obtain

$$\mathbb{E}\left(\sup_{t\in B(m,\hat{m})} |R_{n,4}(t)|^2\right) \lesssim \frac{1}{M} \lesssim \frac{1}{n}.$$

Therefore, (32) holds. □

6.7. **Proof of Theorem 2.** We use the following set

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

Let

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

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

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

Reasoning as in the proof of Theorem 1, if p is such that for all m, m', $4p(m,m')\mathbb{1}_\Lambda \le \widetilde{\text{pen}}(m)\mathbb{1}_\Lambda + \widetilde{\text{pen}}(m')\mathbb{1}_\Lambda$ then

$$\|\tilde{f} - f\|^2\mathbb{1}_\Lambda \le 3\|f_m - f\|^2 + 8[\sup_{t\in B(m,\hat{m})} \nu_n^2(t) - p(m,\hat{m})]\mathbb{1}_\Lambda + 4\widetilde{\text{pen}}(m)\mathbb{1}_\Lambda.$$

It follows from the proof of Theorem 1 that

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

with $p(m,m') = K(\pi m'')^{[s-(1-s)_+/2]_+}\ \Delta(m'')/n$.
Thus, choosing $\widetilde{\text{pen}}(m) = 16K(\pi m)^{[\bar{s}-(1-\bar{s})_+/2]_+}\widetilde{\Delta}(m)/n$, on Λ,

$$\begin{aligned}
4p(m,m') &= 4K(\pi m'')^{[s-(1-s)_+/2]_+}\Delta(m'')/n \\
&\le 4K(\pi m)^{[\bar{s}-(1-\bar{s})_+/2]_+}\Delta(m)/n + 4K(\pi m')^{[\bar{s}-(1-\bar{s})_+/2]_+}\Delta(m')/n \\
&\le 16K(\pi m)^{[\bar{s}-(1-\bar{s})_+/2]_+}\widetilde{\Delta}(m)/n + 16K(\pi m')^{[\bar{s}-(1-\bar{s})_+/2]_+}\widetilde{\Delta}(m')/n \\
&\le \widetilde{\text{pen}}(m) + \widetilde{\text{pen}}(m').
\end{aligned}$$

Then

$$\mathbb{E}(\|\tilde{f} - f\|^2\mathbb{1}_\Lambda) \le 4\inf_{m\in\mathcal{M}_n}\{\|f_m - f\|^2 + \mathbb{E}\widetilde{\text{pen}}(m)\} + \frac{C}{n}$$

We still have to prove that

$$\mathbb{E}(\|\tilde{f} - f\|^2\mathbb{1}_{\Lambda^c}) \le \frac{C}{n}$$

First we compute, using formula (3) ,

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

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

$$\|\tilde{f}\|^2 \le M\hat{m} \le Mm_n \le M^{1/(2\tilde{\gamma})+1}$$

and thus $\mathbb{E}(\|\tilde{f}-f\|^2\mathbb{1}_{\Lambda^c}) \lesssim M^{1/(2\tilde{\gamma})+1}P(\Lambda^c)$. Now, using Markov and Jensen inequalities,

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

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

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

Now, using 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

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

Hence

$$\mathbb{P}(\Lambda^c) \lesssim M^{-ap}\sum_{m\in\mathcal{M}_n}m^{-s+sp} \lesssim M^{-\alpha p}(m_n)^{1+s(p-1)}.$$

Finally $\mathbb{E}(\|\tilde{f}-f\|^2\mathbb{1}_{\Lambda^c}) \lesssim M^{1+1/(2\tilde{\gamma})}P(\Lambda^c) \lesssim M^{-\alpha p+1+1/(2\tilde{\gamma})}m_n^{1+s(p-1)}$. If $s=0$, then $\gamma>0$ and the bound become $M^{-\alpha p+1+1/\gamma}$, so that $p \geq (2+1/\gamma)/\alpha$ implies $\mathbb{E}(\|\tilde{f}-f\|^2\mathbb{1}_{\Lambda^c}) \lesssim M^{-1}$. If $s>0$, then under **(A3)**, $m_n \lesssim (\log(M))^{1/s}$, so that the previous inequality holds if $p > 2/\alpha$. □

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 Comte, F. (2007). Adaptive estimation of linear functionals in the convolution model and applications. *To appear in Bernoulli, Preprint MAP5*, 2007-4.

Butucea, C., Matias, C., and Pouet, C. (2008). Adaptivity in convolution models with partially known noise distribution. *Electron. J. Stat.*, 2:897–915.

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. (2009). Wavelet deconvolution with noisy eigen-values. *To appear in IEEE Trans. Signal Process.*

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.

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.

Johannes, J. (2007). Deconvolution with unknown error distribution. *ArXiv.*

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.

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.

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.