



HAL
open science

Hermite density deconvolution

Ousmane B Sacko

► **To cite this version:**

Ousmane B Sacko. Hermite density deconvolution. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 2020, 17, pp.419-443. 10.30757/ALEA.v17-17 . hal-01978591

HAL Id: hal-01978591

<https://hal.science/hal-01978591v1>

Submitted on 11 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HERMITE DENSITY DECONVOLUTION

OUSMANE B SACKO

ABSTRACT. We consider the additive model:

$$Z = X + \varepsilon,$$

where X and ε are independent. We construct a new estimator of the density of X from n observations of Z . We propose a projection method which exploits the specific properties of the Hermite basis. We study the quality of the resulting estimator by proving a bound on the integrated quadratic risk. We then propose an adaptive estimation procedure, that is a method of selecting a relevant model. We check that our estimator reaches the classical convergence speeds of deconvolution. Numerical simulations are proposed and a comparison with the results of the method proposed in Comte and Lacour (2011) is performed.

Keywords: Deconvolution, Hermite basis, model selection, projection estimator.

2010 Mathematics Subject Classification: 62G05-62G07.

1 INTRODUCTION

Consider the additive noise model:

$$(1) \quad Z_k = X_k + \varepsilon_k, \quad k = 1, \dots, n$$

where

(H_1) $(X_k)_{k \geq 1}$ are independent and identically distributed (i.i.d.) with unknown density f , with respect to the Lebesgue measure.

(H_2) $(\varepsilon_k)_{k \geq 1}$ are i.i.d. with known common density f_ε , with respect to the Lebesgue measure.

(H_3) $(X_k)_{k \geq 1}$ and $(\varepsilon_k)_{k \geq 1}$ are independent.

We observe n copies Z_1, \dots, Z_n . We want to estimate f , the distribution of X_1 , using Z_1, \dots, Z_n only. Under (H_3), if we denote by f_Z the density of Z_1 , we can write

$$(2) \quad f_Z = f * f_\varepsilon,$$

where $u * v(x) = \int_{\mathbb{R}} u(u)v(x - u)du$ is the convolution product of the functions u and v under adequate assumptions. Formula (2) explains the term of "deconvolution" for density estimation in model (1). Two factors influence the rate of convergence: the regularity of f and the regularity of f_ε , with slower rate of convergence if f_ε is more regular. Two types of errors are considered: errors are called "ordinary smooth" errors, when the Fourier transform of f_ε is polynomially decaying near infinity, and "super smooth", when it is exponentially decaying near infinity. The deconvolution problem has been widely studied in the literature. The first works proposed kernel nonadaptive estimators assuming that f is ordinary smooth and that f_ε is ordinary or super smooth. We can cite Carroll and Hall (1988), Fan (1991), (1993), among others. Adaptive estimation, based on a wavelet method, was first considered by Pensky and Vidakovic (1999). Butucea (2004) establishes the minimax rate in the case where f is super smooth and f_ε is ordinary smooth while Butucea and Tsybakov (2007a, 2007b) study optimality in the very difficult case when both functions are super smooth. Some more recent works were dedicated to this problem: Comte and Lacour (2011) considered the case where the noise density is unknown,

MAP5 UMR 8145, University Paris Descartes, 45 Rue des Saints-Pères, 75006 Paris, France.
Email: ousmane.sacko@parisdescartes.fr.

and propose an adaptive estimator in this setting, later improved by Kappus and Mabon (2014). Mabon (2017) builds a projection estimator in Laguerre basis in the case where the variable of interest is positive. Comte and Genon-Catalot (2017) propose a projection estimator in Hermite basis, but do not provide a complete study of its performance. Our aim here is to improve this estimator and to propose an adaptive model selection procedure. We obtain a simple, fast and powerful procedure, which preserves standard deconvolution rates. Moreover, its numerical performances are very good. The paper is organized as follows: we define our estimator in Section 2.2. We prove a bound on the risk in Section 3, and discuss rates of convergence in Section 3.2. In Section 4, an adaptive estimation procedure is proposed and a risk control of the resulting estimator is provided. We then illustrate the performance and stability of the adaptive estimation procedure in Section 5, and we compare our result with Comte and Lacour (2011). A brief extension to the case of dependent variables is given in Section 6. Proofs of most theoretical results are gathered in Section 7.

2 ESTIMATION PROCEDURE AND HERMITE BASIS

2.1 Useful tools

2.1.1 Notations. For $a, b \in \mathbb{R}$, let $a \vee b = \max(a, b)$, and $a_+ = \max(0, a)$. For f, g in $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$, we denote by $\langle f, g \rangle = \int_{\mathbb{R}} f(u) \overline{g(u)} du$, $\|f\|^2 = \int_{\mathbb{R}} |f(u)|^2 du$, $f^*(x) = \int_{\mathbb{R}} e^{itu} f(u) du$ and $f * g(x) = \int_{\mathbb{R}} f(x-u)g(u) du \forall x \in \mathbb{R}$. Lastly, we recall Plancherel-Parseval formula $\langle f, g \rangle = (2\pi)^{-1} \langle f^*, g^* \rangle$.

Before proposing an estimator, we start by recalling the definition of the Hermite basis.

2.1.2 Hermite basis. The Hermite basis $(\varphi_j)_{j \geq 0}$ is a basis on $\mathbb{L}^2(\mathbb{R})$ defined from Hermite polynomials $(H_j)_{j \geq 0}$: $H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} (e^{-x^2})$. The Hermite polynomials are orthogonal with respect to the weight function e^{-x^2} : $\int_{\mathbb{R}} H_j(x) H_k(x) e^{-x^2} dx = 2^j j! \sqrt{\pi} \delta_{j,k}$ (see Abramowitz and Stegun (1964)). Thus, we deduce that the basis:

$$\varphi_j(x) = c_j H_j(x) e^{-x^2/2}, \quad c_j = (2^j j! \sqrt{\pi})^{-1/2},$$

is orthonormal in $\mathbb{L}^2(\mathbb{R})$. The Hermite basis $(\varphi_j)_{j \geq 0}$ is a bounded basis verifying $\|\varphi_j\|_{\infty} = \sup_{x \in \mathbb{R}} |\varphi_j(x)| \leq \phi_0$, with $\phi_0 \simeq 1,086435/\pi^{1/4} \simeq 0,8160$ (see Abramowitz and Stegun (1964)). The Fourier transform of $(\varphi_j)_{j \geq 0}$ verifies:

$$(3) \quad \varphi_j^* = \sqrt{2\pi} (i)^j \varphi_j.$$

Moreover, according to Askey and Wainger (1965), we have

$$(4) \quad |\varphi_j(x)| < C e^{-\xi x^2}, \quad |x| \geq \sqrt{2j+1}, \quad C > 0$$

where ξ is a positive constant independent of x .

2.1.3 Assumptions on the noise. For the definition of our estimator, we assume the following: (H_4) the noise density f_{ε} is such that $f_{\varepsilon}^* \neq 0$.

We also assume that f_{ε} satisfies:

There exist $c_1 \geq c'_1 > 0$, et $\gamma \geq 0, \mu \geq 0, \delta \geq 0$ (with $\gamma > 0$ si $\delta = 0$) such that

$$(5) \quad c'_1 (1+t^2)^{\gamma} e^{\mu|t|^{\delta}} \leq \frac{1}{|f_{\varepsilon}^*(t)|^2} \leq c_1 (1+t^2)^{\gamma} e^{\mu|t|^{\delta}}.$$

It is standard to assume a condition like (5) in the deconvolution setting. When $\delta = 0$ in (5), the function f_{ε} and the errors are called "ordinary smooth". When $\delta > 0$ (with the convention that $\delta > 0$ if and only if $\mu > 0$), they are called "super smooth".

2.2 Estimation procedure. We denote by $S_m = \text{vect}\{\varphi_0, \dots, \varphi_{m-1}\}$, the space generated by $(\varphi_0, \dots, \varphi_{m-1})$ in $\mathbb{L}^2(\mathbb{R})$. Now, we construct an estimator of f relying on the data Z_1, \dots, Z_n , from model (1). We suppose that f belongs to $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$, thus we can write $f = \sum_{j=0}^{+\infty} a_j \varphi_j$

with $a_j = \langle f, \varphi_j \rangle$ and the orthogonal projection of f on S_m is given by: $f_m = \sum_{j=0}^{m-1} a_j \varphi_j$. In fact, we estimate f_m and therefore, we build m estimators \hat{a}_j of $a_j, j = 0, \dots, m-1$. Under (H_4) and using (2), we have $f^* = \frac{f_Z^*}{f_\varepsilon^*}$. Therefore, using Parseval's Theorem and (3), we have:

$$(6) \quad a_j = \langle f, \varphi_j \rangle = \frac{1}{2\pi} \langle f^*, \varphi_j^* \rangle = \frac{(-i)^j}{\sqrt{2\pi}} \langle f^*, \varphi_j \rangle = \frac{(-i)^j}{\sqrt{2\pi}} \int \frac{f_Z^*(u)}{f_\varepsilon^*(u)} \varphi_j(u) du.$$

Thus, to estimate a_j , we replace f_Z^* by an estimate. As $f_Z^*(t) = \int e^{itu} f_Z(u) du = \mathbb{E}[e^{itZ_1}]$, we set:

$$(7) \quad \hat{f}_Z^*(t) = \frac{1}{n} \sum_{k=1}^n e^{itZ_k}.$$

Plugging (7), into (6), we can propose an unbiased estimator of f_m , provided that $\varphi_j/f_\varepsilon^*$ is integrable on \mathbb{R} , for $j = 0, \dots, m-1$:

$$(8) \quad \hat{f}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j, \quad \hat{a}_j = \frac{(-i)^j}{\sqrt{2\pi}} \int \frac{\hat{f}_Z^*(u)}{f_\varepsilon^*(u)} \varphi_j(u) du.$$

The Hermite basis has the specificity of leading to integrable $\varphi_j/f_\varepsilon^*$ in a large number of cases.

3 RISK STUDY OF THE ESTIMATOR

3.1 Risk of the estimator for fixed m . Under the additional assumption:

(H_5) f_Z is bounded,

we can study the risk of \hat{f}_m and the following proposition states our result.

Proposition 3.1. (i) Under $(H_1), \dots, (H_5)$ and for \hat{f}_m given by (8), we have

$$(9) \quad \mathbb{E}[|\hat{f}_m - f|^2] \leq \|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{2}{n} \|f_Z\|_\infty \sum_{j=0}^{m-1} \int_{|u| > \sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du,$$

where $l > 0$ is a positive constant.

(ii) If in addition $l \geq 2$ and f_ε satisfies (5) with $0 < \delta < 2$ or $(\delta = 2, \text{ with } \mu < \xi)$, where ξ is defined in (4), then

$$(10) \quad \frac{2}{n} \|f_Z\|_\infty \sum_{j=0}^{m-1} \int_{|u| > \sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du = \mathcal{O}\left(\frac{1}{n}\right).$$

The first right-hand side term of (9) is the bias term, it is decreasing with m as $\|f - f_m\|^2 = \sum_{j \geq m} a_j^2$. The second term is the main variance term, it is clearly increasing with m . The last term also comes from the variance computation, but we give in Proposition 3.1, part (ii) conditions ensuring that it is negligible. Thus, choosing m that minimizes the risk requires a bias-variance compromise.

If $\delta = 2$ and $\mu \geq \xi$, we can always transform our variables considering the following model:

$$\widetilde{Z}_k = \widetilde{X}_k + \widetilde{\varepsilon}_k, \quad \widetilde{Z}_k = \sqrt{\frac{\xi}{2\mu}} Z_k, \quad \widetilde{X}_k = \sqrt{\frac{\xi}{2\mu}} X_k \quad \text{and} \quad \widetilde{\varepsilon}_k = \sqrt{\frac{\xi}{2\mu}} \varepsilon_k.$$

As $f = \sqrt{\frac{2\mu}{\xi}} f_1(\sqrt{\frac{2\mu}{\xi}} \cdot)$ where f_1 is the density of \widetilde{X}_k , we can build an estimator of f_1 and f_ε^* satisfies (5), with $\widetilde{\mu} < \xi$.

So under the assumptions of Proposition 3.1, part (ii), (9) becomes:

$$(11) \quad \mathbb{E}[|\widehat{f}_m - f|^2] \leq \|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{c}{n}, \quad c > 0, \quad l \geq 2.$$

Comment about (H_5) : We both have, $\forall x \in \mathbb{R}$, $|f_Z(x)| = |f * f_\varepsilon(x)| \leq \min(\|f\|_\infty, \|f_\varepsilon\|_\infty)$ and $|f_Z(x)| \leq \|f\| \cdot \|f_\varepsilon\|$. Therefore, the density f_Z is bounded if f or f_ε is bounded, or if both functions are square integrable. Condition (H_5) is not very strong.

3.2 Rate of convergence. To obtain rates of convergence, we have to evaluate the order of bias and variance terms. In general each basis is associated with a regularity space: here, we consider Sobolev-Hermite spaces.

3.2.1 Rate on a Sobolev-Hermite space. For $s > 0$, the Sobolev-Hermite space of regularity s (see Bongionni and Torrea (2006)) is given by:

$$W_H^s = \{\theta : \mathbb{R} \rightarrow \mathbb{R}, \theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k \geq 0} k^s a_k^2(\theta) < +\infty\}, \quad a_k(\theta) = \int \theta(u) \varphi_k(u) du$$

and the Sobolev-Hermite ball by:

$$(12) \quad W_H^s(D) = \{\theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k \geq 0} k^s a_k^2(\theta) \leq D\}, \quad D > 0$$

Thus, for f in $W_H^s(D)$, we have $\|f - f_m\|^2 = \sum_{j \geq m} j^s a_j^2 j^{-s} \leq Dm^{-s}$. Under the assumptions of

Proposition 3.1 and for $f \in W_H^s(D)$, we get:

$$(13) \quad \mathbb{E}[|\widehat{f}_m - f|^2] \lesssim Dm^{-s} + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2},$$

where the notation \lesssim means for two functions u, v , we denote $u(x) \lesssim v(x)$ if $u(x) \leq cv(x)$, with c is constant independent of x . This inequality is similar to the one in Comte and Lacour (2011), with m therein replaced now by \sqrt{m} . It is worth underlining that the role of the dimension m in projection methods is played here by \sqrt{m} : this is a specificity of the Hermite basis. The result is the similar in density estimation when X_k are directly observed, (see Comte and Genon-Catalot (2017), Belomestny *et al.* (2017)). Let us denote by m_{opt} the value of m for which the bias-variance compromise is obtained, relying on the same calculations as in Comte and Lacour (2011), the rates and the dimension m_{opt} are given in following table.

	$\delta = 0$	$0 < \delta < 2$ or $\delta = 2, \mu < \xi$
m_{opt}	$\lceil n^{\frac{2}{2s+2\gamma+1}} \rceil$	$\left\lceil \left(\frac{\log n}{2\mu l^\delta} \right)^{\frac{2}{\delta}} \right\rceil$
Rate	$n^{-\frac{2s}{2s+2\gamma+1}}$	$(\log n)^{-\frac{2s}{\delta}}$

TABLE 1. Rate of convergence for the MISE if $f \in W_H^s(D)$

The rates obtained coincide with the ones obtained by Fan (1993), Pensky and Vidakovic (1999). They are known to be optimal: lower bounds corresponding to these rates for f_ε verifying (5) are proved by Fan (1993) when f belongs to a Hölder class, and by Pensky and Vidakovic (1999) for f in a Sobolev class.

3.2.2 Rates of convergence for specific function classes. We can obtain for some specific classes of functions a bias term with much smaller order, for instance the Gaussian density or the mixtures of Gaussian. Indeed, then, we can explicitly compute the coefficients a_j and obtain smaller bias than previously on $W_H^s(D)$. Let

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad g_{p,\sigma}(x) = \frac{x^{2p}}{\sigma^{2p}C_{2p}} f_{0,\sigma}(x), \quad C_{2p} = \mathbb{E}[X^{2p}],$$

for X a standard Gaussian variable. We also define the class of mean mixtures, respectively of variance mixtures of the Gaussian distribution by:

$$\mathcal{F}(C) = \left\{ f : f(x) = \phi \star \Pi(x) = \int \phi(x-u) d\Pi(u), \quad \Pi \in \mathcal{P}(C) \right\},$$

where $\mathcal{P}(C) := \{ \Pi \in \mathcal{P}(\mathbb{R}), \Pi(|u| > t) \leq C \exp(-t^2/C), \quad \forall t \in \mathbb{R}^+ \}$, respectively

$$\mathcal{G}(v) = \left\{ f : f(x) = \int_0^{+\infty} \frac{\phi(x/u)}{u} d\Pi(u), \quad \Pi([1/\sqrt{v}, \sqrt{v}]) = 1 \right\}, \quad v > 1,$$

with ϕ the density of standard Gaussian and $\mathcal{P}(\mathbb{R})$ the set of probability measures on \mathbb{R} . The following results are based on bias evaluation obtained in Belomestny *et al.* (2017). The rate is given by the order of variance term, since in all these cases, the bias term is exponentially small. We can prove the following proposition.

Proposition 3.2. *Assume that f_ε is ordinary smooth. For the choice $m_{opt} = \lceil \log(n)/C_1 \rceil$, with $C_1 = \log(2) + \varepsilon\mu^2$ if $f = f_{\mu,1}$, $C_1 = \log\left(\frac{\sigma^2+1}{\sigma^2-1}\right)^2$ if $f = f_{0,\sigma}$, $C_1 = \frac{1}{(eC + 1/\log(2))}$ if $f \in \mathcal{F}(C)$, $C_1 = \left(\frac{v^2-1}{v^2+1}\right)$ if $f \in \mathcal{G}(v)$, we have*

$$\mathbb{E} \left[\|\widehat{f}_{m_{opt}} - f\|^2 \right] \lesssim \frac{(\log n)^{\gamma+\frac{1}{2}}}{n},$$

where γ is given in (5).

The same result hold for $f = g_{p,\sigma}$. This rate is similar to the one obtained in Butucea (2004) for super-smooth functions f .

However in all previous cases the choice $m = m_{opt}$ depends on the regularity of f and associated parameters, which are unknown. This is why we have to look for another method to make the bias-variance compromise, in a data-driven way (see Section 4).

3.3 Comparison with the classical estimator in deconvolution. The "standard" deconvolution estimator (see Fan (1991), and choose sinus cardinal kernel) is given by:

$$(14) \quad \check{f}_\ell(x) = \frac{1}{2\pi} \int_{-\pi\ell}^{\pi\ell} e^{-ixu} \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(-u)} du, \quad \text{where } \widehat{f}_Z^* \text{ is defined by (7).}$$

We mention that this estimator can be decomposed in an orthonormal basis namely $\psi_{\ell,j}(x) = \sqrt{\ell} \psi(\ell x - j)$, $\psi(x) = \frac{\sin \pi x}{\pi x}$ (see Comte *et al.* (2008), Section 3.2), but the development is infinite:

$$\check{f}_\ell(x) = \sum_{j \in \mathbb{Z}} \widehat{a}_{m,j} \psi_{m,j}, \quad \widehat{a}_{m,j} = \frac{1}{n} \sum_{k=1}^n \frac{1}{2\pi} \int \frac{\psi_{m,j}^*(-u)}{f_\varepsilon^*(u)} e^{iuZ_k} du$$

A finite (computable) development would require an additional approximation (truncation of the sum). From computation point of view, the low complexity of \widehat{f}_m in the Hermite basis is an advantage (see Belomestny *et al.* (2017), Section 4.5). The risk of \check{f} verifies

$$\mathbb{E}[|\check{f}_\ell - f|^2] \leq \frac{1}{2\pi} \int_{|t| > \pi\ell} |f^*(u)|^2 du + \frac{1}{2\pi n} \int_{|u| \leq \pi\ell} \frac{du}{|f_\varepsilon^*(u)|^2}.$$

In this context, the regularity spaces which are considered are Sobolev balls defined by

$$(15) \quad W^s(D') = \left\{ f \in \mathbb{L}^2(\mathbb{R}), \int (1 + u^{2s}) |f^*(u)|^2 du < D' \right\}, \quad D' > 0.$$

Note that it is proved in Comte and Genon-Catalot (2017) that $W_H^s(D) \subset W^s(D)$. For $f \in W^s(D)$ the bias term is such that $\frac{1}{2\pi} \int_{|t| > \pi\ell} |f^*(u)|^2 du \leq \frac{D}{2\pi} (\pi\ell)^{-2s} = C\ell^{-2s}$. Therefore, for $\ell = \sqrt{m}$, the risks of the two estimators have the same order. This implies that they have the same rates of convergence.

4 ADAPTIVE ESTIMATION AND MODEL SELECTION

From now on, l given in Proposition 3.1, part (ii) is assumed to be fixed. In this section we propose an automatic selection of m which performs the bias-variance compromise. The procedure does not depend on the regularity of the density f , but only on data Z_1, \dots, Z_n . Consider the contrast function defined by

$$(16) \quad \gamma_n(t) = \|t\|^2 - \frac{2}{n} \sum_{k=1}^n \phi_t(Z_k), \quad \phi_t(x) = \frac{1}{2\pi} \int \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-ixu} du.$$

It is easy to check that $\widehat{f}_m = \underset{t \in S_m}{\operatorname{argmin}} \gamma_n(t)$. Let

$$\Delta(m) = \frac{1}{\pi} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}.$$

We consider \mathcal{M}_n , the collection of models,

$$\mathcal{M}_n = \{m \in \mathbb{N} \setminus \{0\}, \Delta(m) \leq n\}.$$

This collection is finite and contains models with bounded variance. More precisely, as f_ε is a density, we have $|f_\varepsilon^*(u)| \leq 1$, which implies $\Delta(m) \geq \frac{1}{\pi} \int_{|u| \leq \sqrt{lm}} du = \frac{2\sqrt{lm}}{\pi}$. Therefore, $m \lesssim n^2$.

The cardinal of \mathcal{M}_n is therefore at most of order $\mathcal{O}(n^2)$. Our aim is to find the best model \widehat{m} in \mathcal{M}_n , that is to select \widehat{m} such that, the risk of $\widehat{f}_{\widehat{m}}$ approximately performs the bias-variance trade-off, without any information on f . We set:

$$(17) \quad \widehat{m} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \{ \gamma_n(\widehat{f}_m) + \operatorname{pen}(m) \},$$

where $\operatorname{pen}(m)$ is an increasing function defined by:

$$(18) \quad \operatorname{pen}(m) = \begin{cases} \kappa \frac{\Delta(m)}{n}, & \text{if } f_\varepsilon \text{ is ordinary smooth or super smooth with } \delta < \frac{1}{2}, \\ 2\kappa \left(1 + 24\mu l^{\delta/2} m^{\delta - \frac{1}{2}}\right) \frac{\Delta(m)}{n} & \text{if } f_\varepsilon \text{ is super smooth with } \delta \geq \frac{1}{2}, \end{cases}$$

where $\kappa > 0$ is a numerical constant, μ is the constant given in (5) and $l \geq 2$ given in Proposition 3.1, fixed. As $\gamma_n(\widehat{f}_m) = -\|\widehat{f}_m\|^2 = -\sum_{j=0}^{m-1} \widehat{a}_j^2$, it is worth emphasizing that computing \widehat{m} is

numerically fast. Clearly the choice of m given by (17) is entirely determined by the data. The constant κ is independent of the data. The theoretical results show that $\kappa > 16$ is suitable (see

the proof of Lemma 7.1.) In practice this value is too large and is calibrated by preliminary simulation experiments. They confirm that (see Section 5) smaller practical values must be chosen.

We can prove the following theorem.

Theorem 4.1. *Assume that f_ε is square integrable. Let $\text{pen}(m)$ defined by (18), $\hat{f}_m = \underset{t \in S_m}{\text{argmin}} \gamma_n(t)$ and \hat{m} selected by (17). Then, there exists a constant κ_0 such that, for all $\kappa \geq \kappa_0$, the estimator $\hat{f}_{\hat{m}}$ satisfies*

$$(19) \quad \mathbb{E} \left[\|\hat{f}_{\hat{m}} - f\|^2 \right] \leq C \inf_{m \in \mathcal{M}_n} (\|f - f_m\|^2 + \text{pen}(m)) + \frac{C'}{n},$$

where C is a numerical constants ($C=4$ suits) and C' a constant depending on f_ε .

Remark 4.2. *Assume that the assumptions of Theorem 4.1 are satisfied. Then if $f \in W_H^s(D)$ the estimator $\hat{f}_{\hat{m}}$ converges to f with the rates obtained in Table 1. Indeed, the term C'/n in (19) does not change the order of the rate, and is negligible compared to the term $\|f - f_m\|^2 + \text{pen}(m)$.*

5 SIMULATION AND NUMERICAL RESULTS

5.1 Implementation of the adaptive estimator. In this section, we propose some illustrations of the theoretical results. More precisely, we implement the projection estimator given by (8). To do this, we consider data simulated according to (1). For the density f , we choose the distributions (following Comte and Lacour (2011)):

- (i) Gaussian standard $\mathcal{N}(0, 1)$, $I = [-4, 4]$
- (ii) Cauchy standard: $f(x) = (\pi(1+x^2))^{-1}$, $I = [-10, 10]$
- (iii) Laplace density: $f(x) = e^{-\sqrt{2}|x|}/\sqrt{2}$, $I = [-5, 5]$
- (iv) Gamma density $\Gamma(4, 1/\sqrt{3})/\sqrt{12}$, $I = [0, 6]$
- (v) Mixed Gaussian density $(0.5\mathcal{N}(-2, 1) + 0.5\mathcal{N}(2, 1))/\sqrt{5}$, $I = [-3, 3]$

where I is the interval on which we graphically represent the functions and compute the risks.

• **Case 1 : Laplace noise ("ordinary smooth")**

The density considered of f_ε is :

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

The penalty term is given by:

$$\text{pen}(m) = \frac{\kappa}{n} \Delta(m) = \frac{\kappa}{\pi n} \int_{|u| \leq \sqrt{lm}} \left(1 + \frac{u^2}{\lambda^2}\right)^2 du = \frac{2\kappa}{\pi n} \left(\sqrt{lm} + \frac{2}{3\lambda^2} (\sqrt{lm})^3 + \frac{(\sqrt{lm})^5}{5\lambda^4} \right), \quad l = 6.$$

• **Case 2 : Gaussian noise ("super smooth")**

We have:

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

The penalty proposed is:

$$\text{pen}(m) = 4\kappa \left(1 + 24\sigma_\varepsilon^2 l m^{3/2}\right) \frac{\sqrt{lm}}{\pi n} \left(\int_0^1 e^{u^2 \sigma_\varepsilon^2 l m} du \right),$$

where $l = 4$ here and the integral is computed by a Riemann sum discretized in 300 points. Then, we have to calibrate the penalty constant κ . This constant is fixed through preliminary simulations, by testing set of values on different densities f with a large number of repetitions. The comparison of the risks for these different values of κ makes it possible to make a reasonable choice. Finally, we choose $\kappa = 0.4$ for a Laplace noise, $\kappa = 10^{-3}$ for a Gaussian noise.

The estimation procedure is described as follows:

- For m in \mathcal{M}_n , compute $-\sum_{j=0}^{m-1} \hat{a}_j^2 + \text{pen}(m) = \text{Cr}(m)$, with \hat{a}_j given by (8)
- Select \hat{m} such that $\hat{m} = \underset{m \in \mathcal{M}_n}{\text{argmin}} \text{Cr}(m)$,
- Compute $\hat{f}_{\hat{m}} = \sum_{j=0}^{\hat{m}-1} \hat{a}_j \varphi_j$, and $\int_I (\hat{f}_{\hat{m}}(u) - f(u))^2 du$ by discretization.

5.2 Simulations results. Simulation results are given in Tables 2 and 3. The columns of Table 2 indicate the values of the MISE (Mean Integrated squared Error) multiplied by 100 for a Laplace noise or a Gaussian noise, Table 3 gives the ratio of the risk values obtained in Comte and Lacour (2011) divided by the risk values obtained by our method: the larger it is, the better our method is. The errors obtained by our method are computed by a discretization of the integral as Riemann sums and averaging over 100 independent simulations. We remark that increasing the sample size makes the error smaller and thus improves the estimation. Globally the results of our simulations are satisfactory and our method is often better than Comte and Lacour (2011) for both noise densities. The main exception concerns the Gamma density (iv). Some failures for Cauchy density (ii) and super smooth noise are also observed, especially when n increases.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
f	Noise	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.
	Gaussian		0.44	0.37	0.12	0.06	9.5910^{-2}	4.310^{-2}	7.10^{-2}
Cauchy		0.28	0.89	0.20	0.56	0.14	0.37	0.10	0.29
Laplace		1.65	2.18	1.06	1.34	0.75	1.16	0.57	0.87
Gamma		1.70	1.27	0.98	0.97	0.50	0.90	0.28	0.83
Mixed Gaussian		2.82	1.91	1.09	0.87	0.66	0.69	0.41	0.53

TABLE 2. Empirical integrated mean squared errors computed from $(100 \times \mathbb{E} \|\hat{f}_{\hat{m}} - f\|^2)$ over 100 independent simulations pour $n = 100, 250, 500, 1000$.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
f	Noise	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.
	Gaussian		1.95	1.27	5.67	5.00	5.01	5.11	2.41
Cauchy		4.07	1.07	2.45	0.79	2.43	0.70	1.40	0.52
Laplace		1.47	1.40	1.13	1.34	1.12	1.02	1.04	0.89
Gamma		0.67	0.88	0.66	0.73	0.82	0.49	1	0.37
Mixed Gaussian		1.26	2.17	1.45	2.24	1.17	1.68	0.95	1.15

TABLE 3. Ratio of the risks obtained in Comte and Lacour (2011) divided by those of Table 2.

6 CONCLUSION REMARKS: EXTENSIONS TO THE DEPENDENT CASE

We proposed a projection estimator of the density of X in the convolution model (1), relying on the Hermite basis. We prove a bound on the quadratic risk which shows that the relevant parameter is not the dimension m but its square root \sqrt{m} . A data driven estimator is proposed: the model can be automatically chosen and the resulting estimator reaches optimal rates in most

cases. We also provide numerical simulation results, and the comparison with Comte and Lacour (2011) ensures the good performances of our method.

To conclude we explain that the results may be extended to the context of dependent X_i 's. We first define the mixing coefficients.

Definition 1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and \mathcal{U}, \mathcal{V} two σ -algebras of \mathcal{A} . The β -mixing coefficient is defined by

$$(20) \quad \beta(\mathcal{U}, \mathcal{V}) = \frac{1}{2} \sup \left\{ \sum_{i=1}^{\mathcal{I}} \sum_{j=1}^{\mathcal{J}} |\mathbb{P}(U_i \cap V_j) - \mathbb{P}(U_i)\mathbb{P}(V_j)| \right\},$$

where the supremum is taken over all pairs finite partitions $\{U_1, \dots, U_{\mathcal{I}}\}$ and $\{V_1, \dots, V_{\mathcal{J}}\}$ of Ω , such that $U_i \in \mathcal{U}$ and $V_j \in \mathcal{V}$.

Let $(X_k)_{k \in \mathbb{Z}}$ a strictly stationary process. Let $\mathcal{F}_0 = \sigma(X_i, i \leq 0)$ and $\mathcal{F}_k = \sigma(X_i, i \leq k)$ for all $k \in \mathbb{Z}$, where \mathcal{F}_0 is the σ -algebra generated by the X_i for $i \leq 0$ and \mathcal{F}_k generated by X_i for $i \leq k$. The mixing coefficient β_k is defined by $\beta_k = \beta(\mathcal{F}_0, \mathcal{F}_k)$, where β is defined by (20).

The process $(X_k)_{k \in \mathbb{Z}}$ is β -mixing if the sequence β_k tends to zero at infinity.

In this section, we still consider model (1), but we replace (H_1) by (H'_1) : $(X_k)_{k \geq 1}$ is strictly stationary and β -mixing. The estimator is the same as in the independent case.

We can prove a bound on the risk.

Proposition 6.1. Let $1 \leq p < +\infty$ and q two numbers such that $1/p + 1/q = 1$ and assume that

$$(21) \quad \mathbb{E}[|X_1|^{2q/3}] < +\infty. \text{ If the mixing coefficient are such that } \sum_{k=0}^{+\infty} (k+1)^{p-1} \beta_k < +\infty, \text{ then}$$

$$\mathbb{E}[|\hat{f}_m - f|^2] \leq \|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{2}{n} \|f_Z\|_\infty \sum_{j=0}^{m-1} \int_{|u| > \sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du + c' \frac{\sqrt{m}}{n},$$

where $l \geq 2$ is a positive constant, and $c' = c' \left(\mathbb{E}[|X_1|^{2q/3}], \sum_{k=0}^{+\infty} (k+1)^{p-1} \beta_k \right)$.

Now we comment this bound of risk. We remark that we have a usual bias term, the same variance term as in the i.i.d. case with an additional term $c' \sqrt{m}/n$ which is clearly specific to the β -mixing case. As \sqrt{m} is negligible compared to $\frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}$, inequality (21) implies that the risk of \hat{f}_m here has the same order as in the i.i.d. case. We have therefore the same rates of convergence. Thus, the same an adaptive estimation procedure as previously would work. We do no investigate further, methods and results would be similar to Comte *et al.* (2008).

7 PROOFS

7.1 Proof of proposition 3.1. We start by the part (i). For \hat{f}_m given by (8), we have:

$$(22) \quad \mathbb{E} \left[\|\hat{f}_m - f\|^2 \right] = \|f - f_m\|^2 + \mathbb{E} \left[\|\hat{f}_m - f_m\|^2 \right] = \|f - f_m\|^2 + \sum_{j=0}^{m-1} \text{Var}(\hat{a}_j).$$

Now with the definition of \hat{a}_j given by (8) we have

$$\begin{aligned} \text{Var}(\hat{a}_j) &= \text{Var} \left(\frac{(-i)^j}{\sqrt{2\pi n}} \int_{\mathbb{R}} \sum_{k=1}^n e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right) = \frac{1}{2\pi n} \text{Var} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right) \\ &\leq \frac{1}{2\pi n} \mathbb{E} \left[\left| (-i)^j \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right]. \end{aligned}$$

Plugging this in (22) yields

$$\mathbb{E} \left[\|\widehat{f}_m - f\|^2 \right] \leq \|f - f_m\|^2 + \frac{1}{2\pi n} \sum_{j=0}^{m-1} \mathbb{E} \left[\left| \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right].$$

Using $|a + b|^2 \leq 2|a|^2 + 2|b|^2$, we deduce

$$\begin{aligned} \mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right] &\leq 2\mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{|u|>\sqrt{lm}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right] \\ &\quad + 2\mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{|u|\leq\sqrt{lm}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right]. \end{aligned}$$

We evaluate the two right-hand side terms of the previous inequality. By Bessel inequality we have, for the last term:

$$(23) \quad \mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{|u|\leq\sqrt{lm}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right] = \mathbb{E} \left[\sum_{j=0}^{m-1} \left| \left\langle \frac{e^{iZ_1 \bullet}}{f_\varepsilon^*} \mathbb{1}_{|\bullet|\leq\sqrt{lm}}, \varphi_j \right\rangle \right|^2 \right] \leq \int_{|u|\leq\sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}.$$

Moreover, let $\psi_j(u) = \frac{\varphi_j(u)}{f_\varepsilon^*(u)} \mathbb{1}_{|u|>\sqrt{lm}}$, we get for the other term

$$\begin{aligned} \mathbb{E} \left[\sum_{j=0}^{m-1} \left| \int_{|u|>\sqrt{lm}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right] &= \sum_{j=0}^{m-1} \int_{\mathbb{R}} \left| \int_{|u|>\sqrt{lm}} e^{iuz} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 f_Z(z) dz \\ &\leq \|f_Z\|_\infty \sum_{j=0}^{m-1} \int_{\mathbb{R}} \left| \int_{|u|>\sqrt{lm}} e^{iuz} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 dz \\ (24) \quad &= \|f_Z\|_\infty \sum_{j=0}^{m-1} \|\psi_j^*\|^2 = 2\pi \|f_Z\|_\infty \sum_{j=0}^{m-1} \|\psi_j\|^2. \end{aligned}$$

Putting (23), (24) in (22), we have the part (i).

Let us prove the part (ii). We have using (5) that:

$$\sum_{j=0}^{m-1} \int_{|u|>\sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du \leq c_1 \sum_{j=0}^{m-1} \int_{|u|>\sqrt{lm}} (1+u^2)^\gamma |\varphi_j(u)|^2 e^{\mu|u|^\delta} du.$$

By (4), we have $|\varphi_j(x)| < Ce^{-\xi x^2}$ if $|x| \geq \sqrt{2j+1}$. For $j \in \{0, \dots, m-1\}$, thus it is in particular true for $|x| \geq \sqrt{lm}$, with $l \geq 2$. Therefore, for $j \leq m-1$,

$$\begin{aligned} \int_{|u|>\sqrt{lm}} (1+u^2)^\gamma |\varphi_j(u)|^2 e^{\mu|u|^\delta} du &\leq C^2 \int_{|u|>\sqrt{lm}} (1+u^2)^\gamma e^{-2\xi u^2} e^{\mu|u|^\delta} du \\ &\leq C^2 e^{-\xi lm} \int_{\mathbb{R}} (1+u^2)^\gamma e^{-\xi u^2} e^{\mu|u|^\delta} du. \end{aligned}$$

And $\int_{\mathbb{R}} (1+u^2)^\gamma e^{-\xi u^2} e^{\mu|u|^\delta} du < +\infty$ if $\delta < 2$ or if $\delta = 2$, $\mu < \xi$, which corresponds to our

assumptions. Therefore: $\sum_{j=0}^{m-1} \int_{|u|>\sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du = \mathcal{O}(me^{-\xi lm})$. Hence the result. \square .

7.2 Proof of proposition 3.2. By (9) and (10), we have:

$$(25) \quad \mathbb{E}[|\widehat{f}_m - f|^2] \leq \|f - f_m\|^2 + \frac{1}{\pi n} \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{c}{n}.$$

Using Lemma 1 in Comte and Lacour (2011) p.586, we have

$$(26) \quad \int_{|u| \leq \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} \asymp m^{\gamma + \frac{1-\delta}{2}} e^{\mu l^{\frac{\delta}{2}} m^{\frac{\delta}{2}}}.$$

We denote for two functions u and v , $u(x) \asymp v(x)$, if $u(x) \lesssim v(x)$ and $v(x) \lesssim u(x)$.

From Belomestny *et al.* (2017) the bias term is exponentially small (see Proposition 7, 8 and 9), thus, the rate of convergence is given by the order of variance term. As f_ε is ordinary smooth, $\delta = 0$ in (26) and replacing m by $m_{opt} = \lceil \log(n)/C_1 \rceil$, with C_1 is given in Proposition 3.2, we have the result. \square

7.3 Proof of theorem 4.1. By definition of \widehat{m} we have: $\gamma_n(\widehat{f}_{\widehat{m}}) + pen(\widehat{m}) \leq \gamma_n(f_m) + pen(m)$. Moreover, for two functions s, t in $\mathbb{L}^2(\mathbb{R})$, $\gamma_n(t) - \gamma_n(s) = \|t - f\|^2 - \|s - f\|^2 + 2\nu_n(t - s)$, where

$$\nu_n(t) = \frac{1}{n} \sum_{k=1}^n (\phi_t(Z_k) - \langle t, f \rangle),$$

where ϕ_t is defined in (16). Thus, $\|\widehat{f}_{\widehat{m}} - f\|^2 \leq \|f_m - f\|^2 + pen(m) + 2\nu_n(\widehat{f}_{\widehat{m}} - f_m) - pen(\widehat{m})$. As the function $t \mapsto \nu_n(t)$ is linear, we deduce

$$(27) \quad \begin{aligned} \|\widehat{f}_{\widehat{m}} - f\|^2 &\leq \|f_m - f\|^2 + pen(m) + 2\|\widehat{f}_{\widehat{m}} - f_m\| \nu_n \left(\frac{\widehat{f}_{\widehat{m}} - f_m}{\|\widehat{f}_{\widehat{m}} - f_m\|} \right) - pen(\widehat{m}) \\ &\leq \|f_m - f\|^2 + pen(m) + 2\|\widehat{f}_{\widehat{m}} - f_m\| \sup_{t \in S_m + S_{\widehat{m}}, \|t\|=1} \nu_n(t) - pen(\widehat{m}). \end{aligned}$$

For all $x, y \geq 0$ we have: $2xy \leq x^2/4 + 4y^2$, therefore we obtain

$$(28) \quad 2\|\widehat{f}_{\widehat{m}} - f_m\| \sup_{t \in S_m + S_{\widehat{m}}, \|t\|=1} \nu_n(t) \leq \frac{1}{4}\|\widehat{f}_{\widehat{m}} - f_m\|^2 + 4 \sup_{t \in S_m + S_{\widehat{m}}, \|t\|=1} (\nu_n(t))^2.$$

Now, $\|\widehat{f}_{\widehat{m}} - f_m\|^2 \leq 2\|\widehat{f}_{\widehat{m}} - f\|^2 + 2\|f_m - f\|^2$ and plugging this and (28) in (27), we have

$$(29) \quad \frac{1}{2}\|\widehat{f}_{\widehat{m}} - f\|^2 \leq \frac{3}{2}\|f_m - f\|^2 + pen(m) + 4 \sup_{t \in S_m + S_{\widehat{m}}, \|t\|=1} (\nu_n(t))^2 - pen(\widehat{m}).$$

We decompose the empirical process $\nu_n(t)$ in two processes. We set $m^* = \widehat{m} \vee m$. For $t \in S_{m^*}$, we have using Plancherel-Parseval

$$(30) \quad \begin{aligned} \nu_n(t) &= \frac{1}{n} \sum_{k=1}^n (\phi_t(Z_k) - \langle t, f \rangle) \\ &= \frac{1}{n} \sum_{k=1}^n \left(\frac{1}{2\pi} \int_{|u| \leq \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du - \mathbb{E} \left[\frac{1}{2\pi} \int_{|u| \leq \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du \right] \right) \\ &\quad + \frac{1}{n} \sum_{k=1}^n \left(\frac{1}{2\pi} \int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du - \mathbb{E} \left[\frac{1}{2\pi} \int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du \right] \right) \\ &= \frac{1}{n} \sum_{k=1}^n (\phi_{t,1}(Z_k) - \mathbb{E}[\phi_{t,1}(Z_k)]) + \frac{1}{2\pi} \int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} (\widehat{f}_Z^*(u) - f_Z^*(u)) du, \end{aligned}$$

with $\phi_{t,1}(x) = \frac{1}{2\pi} \int_{|u| \leq \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iux} du$. Therefore, we write $\nu_n(t) = \nu_{n,1}(t) + \nu_{n,2}(t)$ where $\nu_{n,1}(t) = \frac{1}{n} \sum_{k=1}^n (\phi_{t,1}(Z_k) - \mathbb{E}[\phi_{t,1}(Z_k)])$ and $\nu_{n,2}(t) = \frac{1}{2\pi} \int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} (\widehat{f}_Z^*(-u) - f_Z^*(-u)) du$. Using that $(\nu_{n,1}(t) + \nu_{n,2}(t))^2 \leq 2(\nu_{n,1}(t))^2 + 2(\nu_{n,2}(t))^2$ and by (29), (30) we deduce

$$\frac{1}{2} \|\widehat{f}_{\widehat{m}} - f\|^2 \leq \frac{3}{2} \|f_m - f\|^2 + pen(m) + 8 \sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,1}(t))^2 + 8 \sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 - pen(\widehat{m}).$$

We introduce the function $p(m, m') = \frac{\kappa \Delta(m \vee m')}{8n}$ if f_ε is ordinary smooth or super smooth with $\delta \leq 1/2$ and $p(m, m') = 2\kappa(1 + \varepsilon(m, m')) \frac{\Delta(m \vee m')}{8n}$ otherwise, where $\varepsilon(m, m')$ is given below, which verifies $8p(m, m') \leq pen(m) + pen(m')$. We obtain:

$$\begin{aligned} \|\widehat{f}_{\widehat{m}} - f\|^2 &\leq 3\|f_m - f\|^2 + 4pen(m) + 16 \sum_{m' \in \mathcal{M}_n} \left(\sup_{t \in S_{m \vee m'}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \\ &\quad + 16 \sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2. \end{aligned}$$

By taking expectation, we get

$$\begin{aligned} \mathbb{E} \left[\|\widehat{f}_{\widehat{m}} - f\|^2 \right] &\leq 3\|f_m - f\|^2 + 4pen(m) + 16 \sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in S_{m \vee m'}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \\ &\quad + 16 \mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right]. \end{aligned}$$

The two followings lemmas lead to the result of Theorem 4.1:

Lemma 7.1. *There exist a constant Σ_1 such that*

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in S_{m \vee m'}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \leq \frac{\Sigma_1}{n}.$$

Lemma 7.2. *There exist a constant Σ_2 such that*

$$\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \leq \frac{\Sigma_2}{n}.$$

Using lemmas 7.1 and 7.2, we have the result choosing $C = 4$ and $C' = 16(\Sigma_1 + \Sigma_2)$. \square

7.3.1 Proof of lemma 7.1. To prove this lemma, we use Talagrand's inequality given in Appendix 8.1, and compute H^2 , M_1 , v defined there. Denote by $m'' = m \vee m'$. We start by computing H^2 . As the map $t \mapsto \nu_{n,1}(t)$ is linear, for $t = \sum_{j=0}^{m''-1} a_j \varphi_j$ such that $\|t\| = 1$, we have

$$(\nu_{n,1}(t))^2 = \left(\sum_{j=0}^{m''-1} a_j \nu_{n,1}(\varphi_j) \right)^2 \leq \sum_{j=0}^{m''-1} a_j^2 \sum_{j=0}^{m''-1} \nu_{n,1}(\varphi_j)^2 = \sum_{j=0}^{m''-1} \nu_{n,1}(\varphi_j)^2.$$

Therefore,

$$\begin{aligned} \mathbb{E} \left[\left(\sup_{t \in S_{m''}, \|t\|=1} (\nu_{n,1}(t))^2 \right) \right] &\leq \mathbb{E} \left[\sum_{j=0}^{m''-1} \nu_{n,1}(\varphi_j)^2 \right] = \sum_{j=0}^{m''-1} \frac{1}{n} \text{Var}(\phi_{\varphi_j,1}(Z_1)) \\ &\leq \frac{1}{n} \sum_{j=0}^{m''-1} \mathbb{E} [|\phi_{\varphi_j,1}(Z_1)|^2]. \end{aligned}$$

It comes using (23) that,

$$(31) \quad \mathbb{E} \left[\sum_{j=0}^{m''-1} |\phi_{\varphi_j}(Z_1)|^2 \right] = \frac{1}{(2\pi)^2} \mathbb{E} \left[\sum_{j=0}^{m''-1} \left| \int_{|u| \leq \sqrt{lm''}} \frac{\varphi_j^*(u) e^{-iuZ_1}}{f_\varepsilon^*(-u)} du \right|^2 \right] \leq \frac{\Delta(m'')}{n} =: H^2.$$

Now we look for M_1 . Using Cauchy-Schwarz inequality and Parseval's theorem

$$\begin{aligned} |\phi_{t,1}(x)| &= \frac{1}{2\pi} \left| \int_{|u| \leq \sqrt{lm''}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iux} du \right| \leq \frac{1}{2\pi} \int_{|u| \leq \sqrt{lm''}} \left| \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iux} \right| du \\ &\leq \frac{1}{2\pi} \sqrt{\int |t^*(u)|^2 du \int_{|u| \leq \sqrt{lm''}} \frac{du}{|f_\varepsilon^*(-u)|^2}} = \frac{1}{2\pi} \sqrt{2\pi \|t\|^2 \int_{|u| \leq \sqrt{lm''}} \frac{du}{|f_\varepsilon^*(-u)|^2}} \\ &\leq \sqrt{\Delta(m'')}. \end{aligned}$$

Thus,

$$(32) \quad \sup_{t \in S_m + S_{m'}, \|t\|=1} \|\phi_{t,1}\|_\infty \leq \sqrt{\Delta(m'')} =: M_1.$$

The case of v is more tedious,

$$\begin{aligned} \text{Var}(\phi_{t,1}(Z_1)) &\leq \mathbb{E} [|\phi_{t,1}(Z_1)|^2] = \frac{1}{2\pi} \int \left| \int_{|u| \leq \sqrt{lm''}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuz} du \right|^2 f_Z(z) dz \\ &= \frac{1}{2\pi} \iiint \frac{t^*(u)}{f_\varepsilon^*(-u)} \frac{t^*(-v)}{f_\varepsilon^*(v)} e^{-i(u-v)z} f_Z(z) \mathbf{1}_{|u| \leq \sqrt{lm''}} \mathbf{1}_{|v| \leq \sqrt{lm''}} du dv dz \\ &= \frac{1}{2\pi} \iint \frac{t^*(u)}{f_\varepsilon^*(-u)} \frac{t^*(-v)}{f_\varepsilon^*(v)} f_Z^*(v-u) \mathbf{1}_{|u| \leq \sqrt{lm''}} \mathbf{1}_{|v| \leq \sqrt{lm''}} du dv \\ &\leq \frac{1}{2\pi} \iint \left| \frac{t^*(u)}{f_\varepsilon^*(-u)} \right|^2 |f_Z^*(v-u)| \mathbf{1}_{|u| \leq \sqrt{lm''}} \mathbf{1}_{|v| \leq \sqrt{lm''}} du dv \\ &\leq \frac{1}{2\pi} \int |f_Z^*(z)| dz \int \left| \frac{t^*(u)}{f_\varepsilon^*(-u)} \right|^2 \mathbf{1}_{|u| \leq \sqrt{lm''}} du. \end{aligned}$$

Using the Cauchy-Schwarz inequality and Parseval's theorem we have:

$$\int |f_Z^*(z)| dz = \int |f^*(z) f_\varepsilon^*(z)| dz \leq 2\pi \|f_\varepsilon\| \cdot \|f\|.$$

Thus, we get: $\text{Var}(\phi_{t,1}(Z_1)) \lesssim \int \left| \frac{t^*(u)}{f_\varepsilon^*(-u)} \right|^2 \mathbf{1}_{|u| \leq \sqrt{lm''}} du$. We consider separately two cases.

- (1) **Ordinary smooth case:** In this case, we have by (31) and by (26) that $H^2 \asymp \frac{m''^{\gamma+1/2}}{n}$.

Moreover,

$$\begin{aligned} \text{Var}(\phi_{t,1}(Z_1)) &\leq \int |t^*(u)|^2 (1+t^2)^\gamma \mathbf{1}_{|u| \leq \sqrt{lm''}} du \leq (1+l^\gamma m''^\gamma) \int |t^*(u)|^2 du \\ &= 2\pi(1+l^\gamma m''^\gamma) \|t\|^2 = 2\pi(1+l^\gamma m''^\gamma). \end{aligned}$$

We can set $v = cm''^\gamma$, with $c > 0$. Thus, using Talagrand's inequality we have:

$$(33) \quad \mathbb{E} \left[\left(\sup_{t \in S_{m''}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \lesssim [U(m'') + V(m'')],$$

with $p(m, m') = \frac{\kappa \Delta(m'')}{8} = \frac{\kappa}{8} H^2 \geq 2(1 + 2\varepsilon)H^2$, we take $\kappa_0 = 17$, $\varepsilon = 1/2$, and

$$U(m'') = \frac{v}{n} \exp \left(-\frac{K_1 n H^2}{2} \frac{1}{v} \right) = \frac{cm''^\gamma}{n} \exp \left(-\frac{K_1}{2} n \frac{m''^{\gamma+\frac{1}{2}}}{cm''^\gamma} \right) \lesssim \frac{m''^\gamma}{n} e^{-\frac{K_1}{2c} m''^{\frac{1}{2}}},$$

$$V(m'') = \frac{M_1^2}{C(\varepsilon)^2 n^2} \exp \left(-K'_1 C(\varepsilon) \frac{1}{\sqrt{2}} \frac{nH}{M_1} \right) = C_1 \frac{\Delta(m'')}{n^2} \exp \left(-C_2 n \frac{\sqrt{\frac{\Delta(m'')}{n}}}{\sqrt{\Delta(m'')}} \right) \lesssim \frac{1}{n} e^{-C_2 \sqrt{n}},$$

because for $m \in \mathcal{M}_n$, $\Delta(m) \leq n$. Therefore, we deduce by (33) that:

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in S_{m''}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \lesssim \sum_{m' \in \mathcal{M}_n} [U(m'') + V(m'')].$$

As

$$\begin{aligned} \sum_{m'} U(m'') &\lesssim \frac{1}{n} \sum_{m'} m''^\gamma e^{-\frac{K_1}{2c} m''} = \frac{1}{n} \left[\sum_{m'=0}^m m''^\gamma e^{-\frac{K_1}{2c} m''} + \sum_{m'=m}^{n^2} m''^\gamma e^{-\frac{K_1}{2c} m''} \right] \\ &= \frac{1}{n} \left[m^{\gamma+1} e^{-\frac{K_1}{2c} m} + \sum_{m'=m}^{+\infty} m''^\gamma e^{-\frac{K_1}{2c} m''} \right] \leq \frac{C'_1}{n}, \end{aligned}$$

and

$$\sum_{m' \in \mathcal{M}_n} V(m'') \lesssim \frac{1}{n} \sum_{m' \in \mathcal{M}_n} e^{-C_2 \sqrt{n}} = \frac{1}{n} |\mathcal{M}_n| e^{-C_2 \sqrt{n}} \lesssim n e^{-C_2 \sqrt{n}} \leq \frac{C''_1}{n}.$$

We deduce that

$$(34) \quad \sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in S_{m''}, \|t\|=1} (\nu_{n,1}(t))^2 - p(m, m') \right)_+ \right] \leq \frac{\Sigma_1}{n}, \quad \Sigma_1 = C'_1 + C''_1.$$

(2) **Super smooth case:** In this case the order of H^2 is given by (26): $H^2 \asymp \frac{m''^{\frac{1-\delta}{2}} e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}}}{n}$,

$$\begin{aligned} \text{Var}(\phi_{t,1}(Z_1)) &\leq c_1 \int |t^*(u)|^2 e^{\mu|u|^\delta} \mathbb{1}_{|u| \leq \sqrt{tm''}} du \leq c_1 e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}} \int |t^*(u)|^2 du \\ &= 2\pi c_1 e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}} \|t\|^2 \lesssim e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}} = v. \end{aligned}$$

We use Talagrand's inequality again, we must compute $U(m'')$ and $V(m'')$.

$$U(m'') = \frac{v}{n} \exp \left(-K_1 \varepsilon \frac{nH^2}{v} \right) = \frac{ce^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}}}{n} \exp \left(-K_1 \varepsilon n \frac{m''^{\frac{1-\delta}{2}} e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}}}{e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}}}} \right) \lesssim \frac{1}{n} e^{\mu \frac{\delta}{2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}},$$

$$\begin{aligned} V(m'') &= \frac{M_1^2}{C^2(\varepsilon) n^2} \exp \left(-K'_1 C(\varepsilon) \sqrt{\varepsilon} \frac{nH}{M_1} \right) = \frac{\Delta(m'')}{C^2(\varepsilon) n^2} \exp \left(-K'_1 C(\varepsilon) \sqrt{\varepsilon} \sqrt{n} \right) \\ &\leq \frac{1}{C^2(\varepsilon) n} \exp \left(-K'_1 C(\varepsilon) \sqrt{\varepsilon} \sqrt{n} \right). \end{aligned}$$

• **Study of $\sum_{m'' \in \mathcal{M}_n} U(m'')$:** we have $\sum_{m'' \in \mathcal{M}_n} U(m'') \lesssim \frac{1}{n} \sum_{m'' \in \mathcal{M}_n} e^{\mu l^{\frac{\delta}{2}} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}}$. We are going to study this term according the value of δ .

(i) **Case $0 < \delta < 1/2$:** In this case $\delta/2 < (1 - \delta)/2$. Thus the choice $\varepsilon = 1$ implies that $m e^{\mu l^{\delta} m^{\frac{\delta}{2}} - K_1 \varepsilon m^{\frac{1-\delta}{2}}}$ is bounded by a constant independent of m' , and $e^{\mu l^{\delta} m'^{\frac{\delta}{2}} - K_1 \varepsilon m'^{\frac{1-\delta}{2}}}$ is integrable in m' . We deduce that:

$$(35) \quad \begin{aligned} \frac{1}{n} \sum_{m'' \in \mathcal{M}_n} e^{\mu l^{\delta} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}} &= \frac{1}{n} \left[\sum_{m'=1}^m e^{\mu l^{\delta/2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}} + \sum_{m'=m}^{n^2} e^{\mu l^{\delta/2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}} \right] \\ &\leq \frac{1}{n} \left[m e^{\mu l^{\delta/2} m^{\frac{\delta}{2}} - K_1 \varepsilon m^{\frac{1-\delta}{2}}} + \sum_{m' \in \mathcal{M}_n} e^{\mu l^{\delta/2} m'^{\frac{\delta}{2}} - K_1 \varepsilon m'^{\frac{1-\delta}{2}}} \right] \leq \frac{C_1''}{n}. \end{aligned}$$

(ii) **Case $\delta \geq 1/2$:** We choose ε such that $\mu l^{\delta/2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}} = -\mu l^{\frac{\delta}{2}} m''^{\frac{\delta}{2}}$, that is $\varepsilon = \frac{2\mu l^{\delta/2}}{K_1} m''^{\delta - \frac{1}{2}}$. This implies

$$(36) \quad \frac{1}{n} \sum_{m'' \in \mathcal{M}_n} e^{\mu l^{\delta/2} m''^{\frac{\delta}{2}} - K_1 \varepsilon m''^{\frac{1-\delta}{2}}} = \frac{1}{n} \sum_{m'' \in \mathcal{M}_n} e^{-\mu l^{\delta/2} m''^{\frac{\delta}{2}}} \leq \frac{1}{n} \sum_{m'} e^{-\mu l^{\delta/2} m'^{\frac{\delta}{2}}} \leq \frac{C_1''}{n}.$$

In the all cases, we have : $\sum_{m'' \in \mathcal{M}_n} U(m'') \leq \frac{C_1''}{n}$.

• **Study of $\sum_{m'' \in \mathcal{M}_n} V(m'')$**

As $|\mathcal{M}_n| = \mathcal{O}(n^2)$ and for all choice of ε in the study of $U(m'')$, we have $C(\varepsilon) = 1$, $\varepsilon \geq 1$. Thus,

$$(37) \quad \sum_{m'' \in \mathcal{M}_n} V(m'') \leq \frac{|\mathcal{M}_n|}{C^2(\varepsilon)n} \exp(-K_1' C(\varepsilon) \varepsilon \sqrt{n}) \leq \frac{n}{C^2(\varepsilon)} \exp(-K_1' C(\varepsilon) \sqrt{\varepsilon} \sqrt{n}) \leq \frac{C_1'}{n}.$$

Therefore, (34) holds and the result of Lemma 7.1 is proven. \square

7.3.2 Proof of lemma 7.2. Here $m^* = m \vee \widehat{m}$. Using the Cauchy-Schwarz inequality for $t = \sum_{j=0}^{m^*-1} a_j \varphi_j$ such that $\|t\|^2 = \sum_{j=0}^{m^*-1} a_j^2 = 1$ we have:

$$\begin{aligned} \nu_{n,2}(t)^2 &= \frac{1}{(2\pi)^2} \left(\int_{|u| > \sqrt{lm^*}} \frac{t^*(u)}{f_\varepsilon^*(-u)} (\widehat{f}_Z^*(-u) - f_Z^*(-u)) du \right)^2 \\ &\leq \frac{1}{(2\pi)^2} \left(\sum_{j=0}^{m^*-1} \left| \int_{|u| > \sqrt{lm^*}} \frac{\varphi_j^*(u)}{f_\varepsilon^*(-u)} (\widehat{f}_Z^*(-u) - f_Z^*(-u)) du \right|^2 \right). \end{aligned}$$

By (3)-(4) and using the Cauchy-Schwarz inequality we have:

$$\begin{aligned}
\sum_{j=0}^{m^*-1} \left| \int_{|u|>\sqrt{lm^*}} \frac{\varphi_j^*(u)}{f_\varepsilon^*(-u)} (\widehat{f}_Z^*(u) - f_Z^*(u)) du \right|^2 &= 2\pi \sum_{j=0}^{m^*-1} \left| \int_{|u|>\sqrt{lm^*}} \frac{\varphi_j(u)}{f_\varepsilon^*(-u)} (\widehat{f}_Z^*(-u) - f_Z^*(-u)) du \right|^2 \\
&\lesssim \sum_{j=0}^{m^*-1} \left(\int_{|u|>\sqrt{lm^*}} \frac{|\widehat{f}_Z^*(-u) - f_Z^*(-u)|}{|f_\varepsilon^*(-u)|} |\varphi_j(u)| du \right)^2 \\
&\lesssim \sum_{j=0}^{m^*-1} \left(\int_{|u|>\sqrt{lm^*}} \frac{|\widehat{f}_Z^*(-u) - f_Z^*(-u)|}{|f_\varepsilon^*(-u)|} e^{-\xi u^2} du \right)^2 \\
&\lesssim \sum_{j=0}^{m^*-1} \left(\int_{|u|>\sqrt{lm^*}} \frac{|\widehat{f}_Z^*(-u) - f_Z^*(-u)|^2}{|f_\varepsilon^*(-u)|^2} e^{-\xi u^2} du \right) \\
&\quad \times \int_{|u|>\sqrt{lm^*}} e^{-\xi u^2} du.
\end{aligned}$$

As $\int_{|u|>\sqrt{lm^*}} e^{-\xi u^2} du \leq ce^{-\xi m^*}$ and the function $x \mapsto xe^{-\xi x}$ reaches its maximum $(1/\xi)e^{-1}$ in $x = 1/\xi$, it implies $\nu_{n,2}(t)^2 \lesssim \int_{\mathbb{R}} \frac{|\widehat{f}_Z^*(-u) - f_Z^*(-u)|^2}{|f_\varepsilon^*(-u)|^2} e^{-\xi u^2} du$. Therefore,

$$\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \lesssim \int_{\mathbb{R}} \frac{\mathbb{E} \left[|\widehat{f}_Z^*(-u) - f_Z^*(-u)|^2 \right]}{|f_\varepsilon^*(-u)|^2} e^{-\xi u^2} du.$$

Now, we have

$$\mathbb{E} \left[|\widehat{f}_Z^*(-u) - f_Z^*(-u)|^2 \right] = \text{Var}[\widehat{f}_Z^*(-u)] = \frac{1}{n} \text{Var}[e^{-iuZ_1}] = \frac{1}{n} (1 - |f_Z^*(-u)|^2) \leq \frac{1}{n}.$$

Thus, by this last inequality we deduce $\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \lesssim \frac{1}{n} \int_{\mathbb{R}} \frac{1}{|f_\varepsilon^*(-u)|^2} e^{-\xi u^2} du$. If f_ε is ordinary smooth, the integral is convergent and the previous bound is of order $1/n$. Assume now f_ε super smooth, we have by (5):

$$\mathbb{E} \left[\sup_{t \in S_{m^*}, \|t\|=1} (\nu_{n,2}(t))^2 \right] \lesssim \frac{1}{n} \int_{\mathbb{R}} e^{\mu|u|^\delta} e^{-\xi u^2} du \leq \frac{\Sigma_2}{n},$$

if $\delta < 2$, or if $\delta = 2$, and $\mu < \xi$. This gives the announced result.

7.4 Proof of proposition 6.1. As in the i.i.d. case, we have the bias-variance decomposition given by (22). Now,

$$\begin{aligned}
\text{Var}(\widehat{a}_j) &= \text{Var} \left(\frac{(-i)^j}{\sqrt{2\pi n}} \int_{\mathbb{R}} \sum_{k=1}^n e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(t)} du \right) \\
&= \frac{1}{2\pi n^2} \sum_{k=1}^n \text{Var} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right) \\
&\quad + \frac{1}{2\pi n^2} \sum_{1 \leq k, l \leq n, k \neq l} \text{Cov} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du, (-i)^j \int_{\mathbb{R}} e^{iuZ_l} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right).
\end{aligned}$$

As $\text{Var}(X) \leq \mathbb{E}|X|^2$, it comes

$$(38) \quad \mathbb{E} \left[\|\widehat{f}_m - f\|^2 \right] \leq \|f - f_m\|^2 + \frac{1}{2\pi n} \sum_{j=0}^{m-1} \mathbb{E} \left[\left| \int_{\mathbb{R}} e^{iuZ_1} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right|^2 \right] \\ + \frac{1}{2\pi n^2} \sum_{j=0}^{m-1} \sum_{1 \leq k, l \leq n, k \neq l} \text{Cov} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du, (-i)^j \int_{\mathbb{R}} e^{iuZ_l} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right).$$

The first two right hand side terms are the same as in the independent case and are dealt with as in Proposition 3.1. We compute the covariance term. First,

$$(39) \quad \text{Cov} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du, (-i)^j \int_{\mathbb{R}} e^{iuZ_l} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right) \\ = \mathbb{E} \left[\int_{\mathbb{R}} \int_{\mathbb{R}} e^{i(uZ_k - vZ_l)} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dudv \right] - \mathbb{E} \left[\int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right] \mathbb{E} \left[\int_{\mathbb{R}} e^{-ivZ_l} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dv \right].$$

The first expectation is equal to

$$(40) \quad \mathbb{E} \left[\int_{\mathbb{R}} \int_{\mathbb{R}} e^{i(uZ_k - vZ_l)} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dudv \right] = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{E} \left[e^{i(uX_k + u\varepsilon_k - vX_l - v\varepsilon_l)} \right] \frac{\varphi_j(u)}{f_\varepsilon^*(u)} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dudv \\ = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{E} \left[e^{i(uX_k - vX_l)} \right] \varphi_j(u) \varphi_j(v) dudv,$$

and the second to:

$$(41) \quad \mathbb{E} \left[\int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right] \mathbb{E} \left[\int_{\mathbb{R}} e^{-ivZ_l} \frac{\varphi_j(v)}{f_\varepsilon^*(-v)} dv \right] = \left| \int_{\mathbb{R}} f^*(u) \varphi_j(u) du \right|^2.$$

Thus, from (39), (40) and (41) we deduce

$$\text{Cov} \left((-i)^j \int_{\mathbb{R}} e^{iuZ_k} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du, (-i)^j \int_{\mathbb{R}} e^{iuZ_l} \frac{\varphi_j(u)}{f_\varepsilon^*(u)} du \right) = \text{Cov} \left(\int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du, \int_{\mathbb{R}} e^{iuX_l} \varphi_j(u) du \right).$$

As a consequence

$$\sum_{1 \leq k, l \leq n, k \neq l} \text{Cov} \left(\int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du, \int_{\mathbb{R}} e^{iuX_l} \varphi_j(u) du \right) \leq \text{Var} \left(\sum_{k=1}^n \int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du \right).$$

Using Viennet's covariance inequality (1997) and equality (3), we have

$$(42) \quad \text{Var} \left(\sum_{k=1}^n \int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du \right) = \text{Var} \left(\sum_{k=1}^n \varphi_j^*(X_k) \right) \leq 8\pi n \int_{\mathbb{R}} b(u) \varphi_j(u)^2 f(u) du,$$

with $b = \sum_{k=0}^n b_k$ and b_k , a sequence of measurable functions such that $b_0 = 1$, $\int b_k(u) f(u) du = \beta_k$ (see Theorem 2.1 in Viennet (1997)).

Lemma 7.3. *Under the assumptions and notations of Proposition 6.1, there exist a constant $c > 0$ such that:*

$$(43) \quad \int_{\mathbb{R}} b(x) \varphi_j^2(x) f(x) dx \leq \frac{c}{\sqrt{j}}, \quad \forall j \geq 1.$$

By Lemma 7.3 and (42), we deduce

$$(44) \quad \begin{aligned} \sum_{j=0}^{m-1} \text{Var} \left(\sum_{k=1}^n \int_{\mathbb{R}} e^{iuX_k} \varphi_j(u) du \right) &\leq 8\pi n \left[\int_{\mathbb{R}} b(u) \varphi_0^2(u) f(u) du + \sum_{j=1}^m \int_{\mathbb{R}} b(u) \varphi_j(u)^2 f(u) du \right] \\ &\leq 8\pi n \left[\phi_0^2 \sum_{k \geq 0} \beta_k + \sum_{j=1}^m \frac{c}{\sqrt{j}} \right]. \end{aligned}$$

Using (44), Proposition 3.1 and in view of (38), we obtain the announced result \square .

7.4.1 Proof of Lemma 7.3. To prove this lemma, we use the decomposition formula of the Hermite basis in Laguerre basis (see Comte and Genon-Catalot (2017), Lemma 7.4, page 16) and the approximation formula of Askey and Wainger (1965). To evaluate $\int_{\mathbb{R}} b(u) \varphi_j(u)^2 f(u) du$, we distinguish the cases under the parity of j .

For j even, $j = 2k$, set to $\nu = 4k + 1$ and we have

$$\int_{\mathbb{R}} b(x) \varphi_{2k}^2(x) f(x) dx = \frac{1}{2} \int_0^\infty x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 f(x) b(x) dx := \sum_{l=1}^6 J_l,$$

where the bounds of $\psi_k^{(\delta)}$ are given in Section 7.1.3 (see Comte and Genon-Catalot (2017), page 15), we have six terms to evaluate.

$$\begin{aligned} J_1 &= \frac{1}{2} \int_0^{1/\sqrt{\nu}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x) f(x) dx \leq \frac{C}{2} \int_0^{1/\sqrt{\nu}} x \left[(x^2 \nu)^{-1/4} \right]^2 b(x) f(x) dx \\ &\leq \frac{C}{2\sqrt{\nu}} \int_{\mathbb{R}} b(x) f(x) dx = \frac{C}{2\sqrt{\nu}} \mathbb{E}[b(X_1)] \leq \frac{C}{2\sqrt{\nu}} \sum_{k \geq 0} \beta_k \end{aligned}$$

$$J_2 = \frac{1}{2} \int_{1/\sqrt{\nu}}^{\sqrt{\nu/2}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x) f(x) dx \leq \frac{C}{2} \int_{1/\sqrt{\nu}}^{\sqrt{\nu/2}} x \left[(x^2 \nu)^{-1/4} \right]^2 b(x) f(x) dx \leq \frac{C}{2\sqrt{\nu}} \sum_{k=0} \beta_k.$$

$$\begin{aligned} J_3 &= \frac{1}{2} \int_{\sqrt{\nu/2}}^{(\nu - \nu^{1/3})^{1/2}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x) f(x) dx \leq \frac{C}{2} \int_{\sqrt{\nu/2}}^{(\nu - \nu^{1/3})^{1/2}} x \left(\nu^{-1/4} (\nu - x^2)^{-1/4} \right)^2 b(x) f(x) dx \\ &= \frac{C}{2} \int_{\sqrt{\nu/2}}^{(\nu - \nu^{1/3})^{1/2}} x^{1/3} x^{2/3} \nu^{-1/2} (\nu - x^2)^{-1/2} b(x) f(x) dx \leq \frac{C}{2\sqrt{\nu}} \int_{\mathbb{R}} |x|^{2/3} b(x) f(x) dx. \end{aligned}$$

Using the Hölder inequality, we have

$$\int_{\mathbb{R}} |x|^{2/3} b(x) f(x) dx \leq \left(\int_{\mathbb{R}} |x|^{2q/3} f(x) dx \right)^{1/q} \left(\int_{\mathbb{R}} b^p(x) f(x) dx \right)^{1/p} = \mathbb{E} \left[|X_1|^{2q/3} \right]^{1/q} \mathbb{E} [b(X_1)^p]^{1/p},$$

with $\frac{1}{p} + \frac{1}{q} = 1$. By Lemma 4.2 in Viennet (1997), page 481, we have: $\mathbb{E}[b(X_1)^p] \leq p \sum_{k \geq 0} (k+1)^{p-1} \beta_k$. It comes: $J_3 \leq \frac{C}{2\sqrt{\nu}} \mathbb{E}[|X_1|^{2q/3}]^{1/q} (p \sum_{k \geq 0} (k+1)^{p-1} \beta_k)^{1/p}$.

$$\begin{aligned} J_4 &= \frac{1}{2} \int_{(\nu-\nu^{1/3})^{1/2}}^{(\nu+\nu^{1/3})^{1/2}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x) f(x) dx \leq \frac{C}{2} \int_{(\nu-\nu^{1/3})^{1/2}}^{(\nu+\nu^{1/3})^{1/2}} x (\nu^{-1/3})^2 b(x) f(x) dx \\ &= \frac{C}{2} \int_{(\nu-\nu^{1/3})^{1/2}}^{(\nu+\nu^{1/3})^{1/2}} x^{1/3} x^{2/3} \nu^{-2/3} b(x) f(x) dx \\ &\leq \frac{C}{\sqrt{\nu}} \int_{\mathbb{R}} |x|^{2/3} b(x) f(x) dx. \end{aligned}$$

By the same computation as for (J_3) we deduce: $J_4 \leq \frac{C}{\sqrt{\nu}} \mathbb{E}[|X_1|^{2q/3}]^{1/q} (p \sum_{k \geq 0} (k+1)^{p-1} \beta_k)^{1/p}$.

$$\begin{aligned} J_5 &= \frac{1}{2} \int_{(\nu+\nu^{1/3})^{1/2}}^{\sqrt{3\nu/2}} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x) f(x) dx \\ &\leq \frac{C}{2} \int_{(\nu+\nu^{1/3})^{1/2}}^{\sqrt{3\nu/2}} x^{1/3} x^{2/3} \left(\nu^{-1/4} (x^2 - \nu)^{-1/4} e^{-\gamma_1 \nu^{-1/2} (x^2 - \nu)^{3/2}} \right)^2 b(x) f(x) dx \\ &\leq \frac{C}{2} \int_{(\nu+\nu^{1/3})^{1/2}}^{\sqrt{3\nu/2}} \nu^{-1/2} x^{1/3} (x^2 - \nu)^{-1/2} e^{-2\gamma_1 \nu^{-1/2} (x^2 - \nu)^{3/2}} x^{2/3} b(x) f(x) dx \leq \frac{C}{\sqrt{\nu}} \int_{\mathbb{R}} |x|^{2/3} b(x) f(x) dx. \end{aligned}$$

Again by the Hölder inequality we get: $J_5 \leq \frac{C}{\sqrt{\nu}} \mathbb{E}[|X_1|^{2q/3}]^{1/q} (p \sum_{k \geq 0} (k+1)^{p-1} \beta_k)^{1/p}$.

Finally,

$$\begin{aligned} J_6 &= \frac{1}{2} \int_{\sqrt{3\nu/2}}^{\infty} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 b(x) f(x) dx \leq \frac{C}{2} \int_{\sqrt{3\nu/2}}^{\infty} x e^{-\gamma_2 x^2} b(x) f(x) dx \\ &\leq C' e^{-3\frac{\gamma_2 \nu}{4}} \int_{\mathbb{R}} b(x) f(x) dx = C' e^{-3\frac{\gamma_2 \nu}{4}} \mathbb{E}[b(X_1)] = C' e^{-3\frac{\gamma_2 \nu}{4}} \sum_{k=0}^n \beta_k \leq C' e^{-3\frac{\gamma_2 \nu}{4}} \sum_{k \geq 0} \beta_k. \end{aligned}$$

For j odd, $j = 2k + 1$, set to $\nu = 4k + 3$ we have,

$$\int_{\mathbb{R}} b(x) \varphi_{2k+1}^2(x) f(x) dx = \frac{1}{2} \int_0^{\infty} x \left(\psi_k^{(-1/2)}(x^2/2) \right)^2 f(x) b(x) dx := \sum_{l=1}^6 K_l.$$

Only the first term changes, thus we just compute K_1 and the other terms are such that the bounds coincide for $i = 2 \dots 6$.

$$K_1 = \frac{1}{2} \int_0^{1/\sqrt{\nu}} x \left(\psi_k^{(1/2)}(x^2/2) \right)^2 b(x) f(x) dx \leq \frac{C}{2} \int_0^{1/\sqrt{\nu}} x \left[(x^2 \nu)^{1/4} \right]^2 b(x) f(x) dx \leq \frac{C}{2\sqrt{\nu}} \sum_{k \geq 0} \beta_k$$

By gathering all these inequalities according to the parity of j we have the announced result.

8 APPENDIX

8.1 Talagrand's inequality. Let $(X_i)_{1 \leq i \leq n}$ be independent real random variables, \mathcal{F} a class at most countable of measurable functions and $\nu_n(f) = \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)])$ for all $f \in \mathcal{F}$.

We assume there exist third strictly positive constants M_1, H, v such that: $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M_1$,

$\mathbb{E}[\sup_{f \in \mathcal{F}} |\nu_n(f)|] \leq H$, and $\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \text{Var}(f(X_i)) \leq v$. Then for $\varepsilon > 0$,

$$\mathbb{E}[(\sup_{f \in \mathcal{F}} |\nu_n^2(f)| - 2(1+2\varepsilon)H^2)_+] \leq \frac{4}{K_1} \left(\frac{v}{n} \exp\left(-K_1 \varepsilon \frac{nH^2}{v}\right) + \frac{49M_1^2}{K_1 C^2(\varepsilon) n^2} \exp\left(-K_1' C(\varepsilon) \sqrt{\varepsilon} \frac{nH}{M_1}\right) \right),$$

where $C(\varepsilon) = (\sqrt{1+\varepsilon} - 1) \wedge 1$, $K_1 = 1/6$ and K_1' a universal constant. The Talagrand inequalities has been proven in Talagrand (1996), reworded by Ledoux (1997). This version is given in Klein and Rio (2005).

Acknowledgements: The author thanks F. Comte and C. Duval for helpful advices and kind proofreadings.

REFERENCES

- [1] Abramowitz M, Stegun IA (1965) Handbook of mathematical functions: with formulas, graphs, and mathematical tables. Courier Corporation 55
- [2] Askey R, Wainger (1965) Mean convergence of expansion in Laguerre and Hermite Series. Amer. J. Math 87:695–708
- [3] Belomestny D, Comte F, Génon-Catalot V (2017) Sobolev-Hermite versus Sobolev nonparametric density estimation on \mathbb{R} . Annals of the Institute of Statistical Mathematics, pages 1–34.
- [4] Bongioanni B, Torrea JL (2006) Sobolev spaces associated to the harmonic oscillator Proc. Indian. Acad. Sci (math. Sci.) 116:337–360
- [5] Butucea C (2004) Deconvolution of supersmooth densities with smooth noise. Canad. J. Statist 32(2):181–192
- [6] Butucea C, Tsybakov AB (2007) Sharp optimality in density deconvolution with dominating bias. I. Teor. Veroyatn. Primen 52(1):111–128
- [7] Carroll RJ, Hall P (1988) Optimal rates of convergence for deconvolving a density. J. Amer. Statist. Assoc. 83(404):1184–1186
- [8] Comte F, Dedecker J, Taupin ML (2008) Adaptive density deconvolution with dependent inputs. Math. Methods Statist. 17(2):87–112
- [9] Comte F, Genon-Catalot V (2018) Laguerre and Hermite bases for inverse problems. J. Korean Statist. Soc. 47(3):273–296
- [10] Comte F, Lacour C (2011) Data-driven density estimation in the presence of additive noise with unknown distribution. J. R. Stat. Soc. Ser B Stat. Methodol 73(4):601–627
- [11] Fan J (1991) On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. 19(3):1257–1272
- [12] Fan J (1993) Adaptively local one-dimensional subproblems with application to a deconvolution problem. Ann. Statist. 21(2):600–610
- [13] Kappus J, Mabon G (2014) Adaptive density estimation in deconvolution problems with unknown error distribution. Electron. J. Stat. 8(2):2879–2904
- [14] Klein T, Rio E (2005) Concentration around the mean for maxima of empirical processes. Ann. Probab. 33(3):1060–1077
- [15] Ledoux M (1997) On Talagrand’s deviation inequalities for product measures. ESAIM: Probability and statistics 1:63–87
- [16] Mabon G (2017) Adaptive deconvolution on the non-negative real line. Scand. J. Stat. 44(3):707–740
- [17] Pensky M, Vidakovic B (1999) Adaptive wavelet estimator for nonparametric density deconvolution. Ann. Statist. 27(6):2033–2053
- [18] Talagrand M (1996) New concentration inequalities in product spaces. Invent. Math 126(3):505–563
- [19] Viennet G (1997) Inequalities for absolutely regular sequences: application to density estimation. Probab. Theory Related Fields 107(4):467–492