PROJECTION ESTIMATION OF A QUADRATIC FUNCTIONAL FROM INDIRECT OBSERVATIONS

Ousmane Sacko

PROJECTION ESTIMATION OF A QUADRATIC FUNCTIONAL FROM INDIRECT OBSERVATIONS

OUSMANE SACKO[1]

ABSTRACT. We consider the convolution model: $Y = X + \varepsilon$, where X and ε are independent. We aim to estimate $\int_{\mathbb{R}} f^2(x)dx$, where f is the unknown density of the signal X from n observations of Y. We introduce a new projection estimator based on expanding f in the Hermite basis. Convergence rates for f within the Sobolev-Hermite ball are provided for various error types. We also present an adaptive procedure inspired by Goldenshluger and Lepski (2011) to select the appropriate space, and we demonstrate an oracle inequality for the adaptive estimator. Numerical experiments are conducted to illustrate the effectiveness of our methodology.

Keywords: Deconvolution; Hermite basis; Model selection; Projection estimator; Quadratic functionals; U-statistics.

1. INTRODUCTION

1.1. Bibliographical context and application.

Consider the convolution model given by:

$$Y_k = X_k + \varepsilon_k, \quad k = 1, \dots, n, \tag{1}$$

and the following assumptions

- (**H**1) The variables $(X_k)_{k\geqslant 1}$ are independent and identically distributed (i.i.d.) with an *unknown density* f, with respect to the Lebesgue measure,
- (**H**2) The variables $(\varepsilon_k)_{k\geqslant 1}$ are i.i.d. with a *known* density f_ε, with respect to the Lebesgue measure,
- (**H**3) The variables$(X_k)_{k\geqslant 1}$ and $(\varepsilon_k)_{k\geqslant 1}$ are independent.

Our aim is to propose an *adaptive estimator* of the quadratic functional of a density

$$\theta(f) = \int_{\mathbb{R}} f^2(x)dx$$

from n copies $Y_1, \dots, Y_n$ with common density $f_Y = f \star f_\varepsilon$, where $g \star h(x) = \int_{\mathbb{R}} g(x-y)h(y)dy$ denotes the convolution product between g and h.

The quadratic functional plays an important role in mathematical statistics. A typical application example is the goodness-of-fit testing problem, which is based on an estimation of $\|f - f_0\|^2$ for f_0 some given density (see Fromont and Laurent (2006), Butucea (2007)). It also appears in the estimation of integral functionals (see Laurent (1996), Giné and Nickl (2016) Section 5.3.1), in particular for the estimation of entropy (see Han et al. (2020)).

When the variables of interest are available, that is $\varepsilon_k \equiv 0$ almost surely $(a.s.)$ in Model (1), the question of nonparametric estimation of quadratic functionals has been studied extensively. For instance, Bickel and Ritov (1988) investigate the estimation of $\int_{\mathbb{R}}(f^{(d)})^2(x)dx$, where $f^{(d)}$ denotes the d-th derivative of f by using a kernel method. They establish that the parametric rate can be reached if the regularity of $f^{(d)}$ is large enough. However, their procedure is non adaptive. Efromovich and Low (1996) propose a projection estimator of $\int_{\mathbb{R}}(f^{(d)})^2(x)dx$ on the Fourier basis. The authors provide convergence rates for $f^{(d)}$ belonging to the Lipschitz class of order β with a logarithmic loss for small β values. But β is unknown.

[1] MODAL'X, UMR 9023, Université Paris Nanterre. Email: osacko@parisnanterre.fr.

Laurent (2005) develops an adaptive estimator of $\theta(f)$ by considering the Haar basis. The author obtains the same convergence rate as in Efromovich and Low (1996) when the density f belongs to a Besov class. Recently, Goldenshluger and Lepski (2022a,b) establish minimax rates for the $\mathbb{L}_p$-norm of a multivariate density function with $p \geqslant 1$ on the Nikolskii space. Furthermore, the problem of estimating a general functional can be found in Birgé and Massart (1995), Laurent (1996), Kerkyacharian and Picard (1996), Tribouley (2000). In privacy constraints framework, Butucea et al. (2023) investigate minimax rates and propose adaptive estimators (up to a logarithmic factor) of $\theta(f) = \int_0^1 f^2(x)dx$ based on the Haar wavelet basis.

In convolution models (1), to our knowledge, Butucea (2007) is the first to consider the problem of estimating of $\theta(f) = \int_{\mathbb{R}} f^2(x)dx$ from Model (1). The author proposes a kernel estimator and proves minimax results under certain assumptions on the density of noise f_ε. Loubes and Marteau (2014) apply the estimation of $\theta(f) = \int_{\mathbb{R}} f^2(x)dx$ in context of goodness-of-fit tests. Chesneau (2011) considers the estimation of quadratic functional of f in the cases where the convolution of the signal f with some known function is contaminated by Gaussian noise. Recently, Schluttenhofer and Johannes (2020) provide minimax results for quadratic functional estimation problems in the case of a circular version of Model (1).

However, the projection method and adaptation issues from Model (1) have not yet been addressed in the literature. In the present paper, our main goal is to provide an adaptive projection estimator of $\int_{\mathbb{R}} f^2(x)dx$, based on the decomposition of f in the Hermite basis. The Hermite basis is $\mathbb{R}$-supported and then leads to the construction of an unconstrained estimator on f's support. When using compactly supported bases with support $[a, b]$, the bounds a and b are determined in practice from the dataset. The Hermite basis does not require this preliminary choice. Our methodology is mainly based on the various good properties of the Hermite basis. For instance, the Fourier transform of a function of the Hermite is almost the same Hermite function up to a constant and it decreases as $e^{-\xi x^2}$ for sufficiently large x and ξ a constant (see below Equations (4) and (5)). This is important in the context of our inverse problem for solving integrability issues. Recently, Belomestny et al. (2019) prove that the Hermite basis has low complexity, requiring only a few coefficients for accurate estimation. The estimate is therefore parsimonious.

1.2. **Contributions.** Our main contributions are non asymptotic and described as follows:

- First, we introduce a new projection estimator of $\int_{\mathbb{R}} f^2(x)dx$ constructed from the expansion of f in the Hermite basis. Our procedure has the advantage to be kernel-free and reduces the problem of estimating $\int_{\mathbb{R}} f^2(x)dx$ to the estimation of a small number of coefficients. Moreover, the proposed estimator generalizes the projection estimator in the direct observation case ($\varepsilon_k \equiv 0$ almost surely in Model (1)).
- Then, we provide convergence rates for f belonging to the Sobolev-Hermite ball and for various types of error. These rates coincide with the one obtained by Butucea (2007) and Laurent (2005), respectively in convolution Model (1) and direct observation cases.
- Following Goldenshluger and Lepski (2011)'s methods, we propose an innovative adaptive procedure to select the relevant dimension of the projection space. We demonstrate that the adaptive estimator satisfies a non asymptotic oracle inequality and achieves the optimal rate, with at least a logarithmic factor. To the best of our knowledge, this result is new in the literature for the estimation of $\int_{\mathbb{R}} f^2(x)dx$ from Model (1).
- Finally, we illustrate the good performance of our procedure through numerical experiments and compare our method to the penalization approach in the direct observation case.

1.3. **Organization of paper.** In Section 2, we recall the definition of the Hermite basis, regularity spaces, classical assumption on the noise and we describe our methodology. In Section 3, we discuss the

rates of convergence over Sobolev Hermite ball. Section 4 is devoted to the adaptive estimation. The results of a simulation study are detailed in Section 5. All the main proofs are presented in Section 7, while the Proofs of auxiliary results and concentration tools are postponed to Section A and B of the supplementary material.

2. Methodology, hermite basis and regularity spaces

2.1. Hermite basis and regularity spaces.

2.1.1. *Notation.* For a, b two real numbers, denote $a\vee b=\max(a,b)$, $a\wedge b=\min(a,b)$ and $a_+=\max(0,a)$. Let z be the complex number, we denote by $\overline{z}$ the conjugate of z. For g and h two functions in $\mathbb{L}^2(\mathbb{R})$, we denote $\langle g,h\rangle=\int_{-\infty}^{+\infty}g(x)\overline{h(x)}dx$ the scalar product on $\mathbb{L}^2(\mathbb{R})$ and $\|g\|=\left(\int_{-\infty}^{+\infty}|g(x)|^2dx\right)^{1/2}$ the norm on $\mathbb{L}^2(\mathbb{R})$. The Fourier transform of g is defined by $g^*(u)=\int e^{iux}g(x)dx$. We recall the Plancherel-Parseval equality $\langle g,h\rangle=(2\pi)^{-1}\langle g^*,h^*\rangle$. Let u_n and v_n be two real sequences, denote $u_n\lesssim v_n$ if there exists a positive constant c such that for all $n\in\mathbb{N}$, $u_n\leqslant cv_n$ and $u_n\asymp v_n$ if $u_n\lesssim v_n$ and $v_n\lesssim u_n$.

We begin by recalling the definition of the Hermite basis and regularity space.

2.1.2. *The Hermite basis.* Define the Hermite basis $(\varphi_j)_{j\geqslant 0}$ from Hermite polynomials $(H_j)_{j\geqslant 0}$ by:

$$\varphi_j(x)=c_jH_j(x)e^{-x^2/2},\quad H_j(x)=(-1)^je^{x^2}\frac{d^j}{dx^j}(e^{-x^2}),\quad c_j=(2^jj!\sqrt{\pi})^{-1/2},\quad x\in\mathbb{R},\ j\geqslant 0. \tag{2}$$

The family $(H_j)_{j\geqslant 0}$ is orthogonal with respect to the weight function e^{-x^2}: $\int_{\mathbb{R}}H_j(x)H_k(x)e^{-x^2}dx=2^jj!\sqrt{\pi}\delta_{j,k}$, where $\delta_{k,j}$ is the Kronecker symbol (see Abramowitz and Stegun (1964), Chap. 22.2.14). It follows that the sequence $(\varphi_j)_{j\geqslant 0}$ is an orthonormal basis on $\mathbb{R}$. The infinite norm of $(\varphi_j)_{j\geqslant 0}$ is such that (see Abramowitz and Stegun (1964), chap. 22.14.17 and Indritz (1961))

$$\|\varphi_j\|_\infty=\sup_{x\in\mathbb{R}}|\varphi_j(x)|\leqslant\phi_0,\ \text{with}\ \phi_0=\pi^{-1/4}. \tag{3}$$

From Askey and Wainger (1965), it yields

$$|\varphi_j(x)|<C'_\infty e^{-\xi x^2},\qquad |x|\geqslant\sqrt{2j+1},\qquad C'_\infty>0, \tag{4}$$

where ξ is a positive constant independent on x and $0<\xi<\frac{1}{2}$. The Fourier transform of $(\varphi_j)_{j\geqslant 0}$ satisfies:

$$\varphi_j^*=\sqrt{2\pi}(i)^j\varphi_j. \tag{5}$$

2.1.3. *Regularity spaces.* We consider the regularity spaces associated to the Hermite basis called Sobolev-Hermite spaces. In the sequel, it is used to the evaluate the bias of our estimator.

Definition 2.1 (Sobolev-Hermite ball)**.** *For $s>0$, define the Sobolev-Hermite ball of regularity s and radius D by:*

$$W_H^s(D)=\{\theta\in\mathbb{L}^2(\mathbb{R}),\sum_{k\geqslant 0}k^sa_k^2(\theta)\leqslant D\},\quad a_k(\theta)=\langle\theta,\varphi_k\rangle=\int_{\mathbb{R}}\theta(x)\varphi_k(x)dx. \tag{6}$$

When $s\geqslant 1$ is an integer, Bongioanni and Torrea (2006) showed that $\theta\in W_H^s(D)$ if and only if θ is s-times differentiable and the functions $\theta,\dots,\theta^{(\ell)},x\mapsto x^{s-\ell}\theta^{(\ell)}$ belong to $\mathbb{L}^2(\mathbb{R})$ for $\ell=0,\dots,s-1$.

Recall also the definition of the classical Sobolev ball

Definition 2.2 (Usual Sobolev-ball)**.** *A Sobolev class with smoothness $s>0$ and radius $D>0$ is defined by:*

$$W^s(D)=\{\theta\in\mathbb{L}^2(\mathbb{R}),\int_{\mathbb{R}}(1+u^2)^s|\theta^*(u)|^2du<D\}.$$

For any $s > 0$, we know in Bongioanni and Torrea (2006) that $W_H^s(D) \subset W^s(D')$. Moreover, if $s \geqslant 1$ is an integer the space $W^s(\cdot)$ is given by:

$$W^s(\cdot) = \{\theta \in \mathbb{L}^2(\mathbb{R}), \quad \theta \text{ is } s\text{-times differentiable and } \sum_{\ell=1}^{s} \|\theta^{(\ell)}\|^2 < \infty\}.$$

Therefore, if θ has a compact support, $\theta \in W_H^s(D)$ is equivalent to $\theta \in W^s(D')$, where D and D' are related constants.

2.1.4. *Assumption on the noise.* The following assumptions on the noise are considered in convolution context.

(**H4**) The Fourier transform of the noise density is such that $f_\varepsilon^* \neq 0$.

Assume also that: there exist $c_1 \geqslant c_1' > 0$, and $\gamma \geqslant 0, \mu \geqslant 0, \delta \geqslant 0$ (with $\gamma > 0$ if $\delta = 0$) such that

$$c_1'(1+t^2)^\gamma e^{\mu|t|^\delta} \leqslant |f_\varepsilon^*(t)|^{-2} \leqslant c_1(1+t^2)^\gamma e^{\mu|t|^\delta}, \quad \forall t \in \mathbb{R}. \tag{7}$$

In the sequel, we consider two types of errors.

- If $\delta = 0$, we say that ε and f_ε are ordinary smooth.
- When $\delta > 0$, they are called super smooth noises or functions.

Let us note that Condition (7) implies Assumption (**H4**) and is fulfilled by some classical densities. Examples of ordinary smooth noises are: Gamma, Laplace and Gamma symmetric distributions. For super smooth noises, we can cite Gaussian, Mixed Gaussian and Cauchy distributions.

Let us now present our methodology.

2.2. **Methodology.** Let $(Y_k)_{k\in\{1,\dots,n\}}$ be n i.i.d. observations drawn from Model (1). Our purpose is to estimate $\theta(f) = \int_{\mathbb{R}} f^2(x)dx$ from indirect observations $Y_1, \dots, Y_n$. For any integer $m \geqslant 1$, define $S_m = \text{Span}\{\varphi_0, \dots, \varphi_{m-1}\}$, the linear spaces generated by $\varphi_0, \dots, \varphi_{m-1}$. Assuming that $f \in \mathbb{L}^2(\mathbb{R})$, we decompose f in an orthonormal basis (here in the Hermite basis) and we define its orthogonal projection on S_m by:

$$f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j, \quad a_j(f) = \langle f, \varphi_j\rangle = \int_{\mathbb{R}} f(x)\varphi_k(x)dx.$$

Set

$$\theta_m(f) = \|f_m\|^2 = \sum_{j=0}^{m-1} a_j^2(f).$$

We will estimate $\theta_m(f)$. Under (**H4**) and using Plancherel-Parseval's theorem, we have

$$a_j(f) = \frac{1}{2\pi}\langle \frac{f_Y^*}{f_\varepsilon^*}, \varphi_j^*\rangle. \tag{8}$$

Since $f_Y^*(u) = \int_{\mathbb{R}} e^{iux} f_Y(x)dx = \mathbb{E}[e^{iuY_1}]$, we estimate $f_Y^*(u)$ by:

$$\widehat{f_Y^*}(u) = \frac{1}{n}\sum_{k=1}^{n} e^{iuY_k}. \tag{9}$$

Consider the random variables $(U_t^*(Y_k))_{k\in\{1,\dots,n\}}$ defined as:

$$U_t^*(Y_k) = \frac{1}{2\pi}\langle \frac{e^{i\cdot Y_k}}{f_\varepsilon^*}, t^*\rangle, \qquad U_t(x) = \frac{1}{2\pi}\frac{t^*(-x)}{f_\varepsilon^*(x)}. \tag{10}$$

Hence, injecting (9) into (8) and assuming that the ratio $\varphi_j/f_\varepsilon^*$ is integrable over $\mathbb{R}$ for $j \in \{0,\dots,m-1\}$, we define the following estimator:

$$\widehat{\theta}_m(f) = \sum_{j=0}^{m-1} \widehat{a_j^2(f)}, \qquad \text{where} \qquad \widehat{a_j^2(f)} = \frac{1}{n(n-1)} \sum_{k\neq \ell=1}^{n} U^*_{\varphi_j}(Y_k) U^*_{\varphi_j}(Y_\ell). \tag{11}$$

The Hermite basis ensures that the ratio $\varphi_j/f_\varepsilon^*$ is integrable in many cases, in particular, for any ordinary smooth noise or super smooth noise with $\delta \leqslant 2$. Consequently, $\widehat{\theta}_m(f)$ is well defined for several function classes satisfying Condition (7). Furthermore, by the Plancherel-Parseval theorem, it is evident that $\mathbb{E}[U^*_{\varphi_j}(Y_k)] = a_j(f)$. Hence, $\widehat{\theta}_m(f)$ is an unbiased estimator of $\theta_m(f) = \sum_{j=0}^{m-1} a_j^2(f)$. Moreover, for any real function t, it hold that $\overline{U_t^*(x)} = U_t^*(x)$. Thus, our estimator $\widehat{\theta}_m(f)$ is a real random variable, which is crucial since we are estimating a real number.

3. Rates of convergence

3.1. Risk bound for fixed m.

Let $m \geqslant 1$ an integer and $\rho > 0$ an absolute constant. Define

$$\Delta(m) := \int_{|u|\leqslant\sqrt{\rho m}} \frac{du}{|f_\varepsilon^*(u)|^2}, \qquad \Lambda(m) := \sum_{j=0}^{m-1} \int_{|u|>\sqrt{\rho m}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du, \tag{12}$$

and

$$\Delta^{(1)}(m) := \sup_{|u|\leqslant\sqrt{\rho m}} |f_\varepsilon^*(u)|^{-2}. \tag{13}$$

We can establish the following bound on the quadratic risk.

Proposition 3.1. *Let $(Y_k)_{k\in\{1,\dots,n\}}$ be n i.i.d. observations drawn from Model (1) and assume that assumptions (**H1**) to (**H4**) hold. Assume that f is bounded ($\|f\|_\infty < \infty$) and f_ε is square integrable. Consider the estimator $\widehat{\theta}_m(f)$ defined in (11).*

(i) Then, for any $\rho > 0$ and $C_0 = 4\|f\|_\infty^{\frac{1}{2}}\|f_\varepsilon\| < \infty$, it holds

$$\mathbb{E}[(\widehat{\theta}_m(f) - \theta(f))^2] \leqslant \|f - f_m\|^4 + \frac{C_0}{n(n-1)}\Delta(m)\left(\Delta^{(1)}(m) + \Lambda(m)\right) + \frac{C_0^2}{n(n-1)}\Lambda^2(m) + \frac{4\|f\|_\infty^2}{n(n-1)} + \frac{4}{n}\mathbb{E}[|U^*_{f_m}(Y_1)|^2].$$

(ii) Moreover, if $\rho \geqslant 2$ and f_ε satisfies (7) with $0 \leqslant \delta < 2$ or ($\delta = 2$, with $\mu < \xi$), where ξ is defined in (4), then, we get

$$\mathbb{E}[(\widehat{\theta}_m(f) - \theta(f))^2] \leqslant \|f - f_m\|^4 + \frac{C_1}{n(n-1)}\Delta(m)\Delta^{(1)}(m) + \frac{4}{n}\mathbb{E}[|U^*_{f_m}(Y_1)|^2] + \frac{C_2}{n},$$

where C_1 and C_2 are constants depending on C_0, $\|f_Y\|_\infty, \rho$ and constants appearing in (7)-(4).

Let us make some comments on the bounds obtained in Proposition 3.1. The first term in the right-hand side ($\|f - f_m\|^4$) is the classical bias term: it is decreasing with m. The others are variance terms. It has been established (refer to Proposition 3.1 in Sacko (2020)) that $\Lambda(m)$ becomes negligible under Condition (7). More precisely, under the assumptions outlined in Proposition 3.1 (ii), we have:

$$\Lambda(m) = \sum_{j=0}^{m-1} \int_{|u|>\sqrt{\rho m}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du \lesssim e^{-\frac{\xi\rho m}{2}}. \tag{14}$$

Equation (14) is a key property of the Hermite basis and what makes it so relevant in the context of inverse problems. Consequently, the main variance term is $\Delta(m)\Delta^{(1)}(m)/n(n-1) + \mathbb{E}[|U^*_{f_m}(Y_1)|^2]/n$. The first is clearly increasing with m contrary of $\mathbb{E}[|U^*_{f_m}(Y_1)|^2]/n$, which is a mixed of bias and variance (see Lemma 3.1). It can be negligible for ordinary smooth noises with an adequate choice of m. The examination of $\mathbb{E}[|U^*_{f_m}(Y_1)|^2]$ entails technical intricacies. Let us proceed to bound it in two distinct ways. Consider the following assumption:

(**H5**) The density f_Y is bounded.

Observe that if (f and f_ε belong to $\mathbb{L}^2(\mathbb{R})$) or (f or f_ε is bounded), (**H5**) is automatically satisfied.

Lemma 3.1. *Assume that (**H5**) holds, $\|f\|_\infty < \infty$ and set $C_0' = \|f\|_\infty \|f_Y\|_\infty < \infty$.*

(i) For any $m \geqslant 1$, it holds

$$\mathbb{E}[|U^*_{f_m}(Y_1)|^2] \leqslant 2\|f_Y\|_\infty \left(\Delta^{(1)}(m)\|f - f_m\|^2 + \frac{1}{2\pi} \int_{|u|\leqslant\sqrt{\rho m}} \frac{|f^*(-u)|^2}{|f^*_\varepsilon(u)|^2} du \right) + C_0' \Lambda(m).$$

(ii) Furthermore, it yields

$$\mathbb{E}[|U^*_{f_m}(Y_1)|^2] \leqslant C_0' \left(\Delta^{(1)}(m) + \Lambda(m) \right).$$

Lemma 3.1 (i) is the key to achieve the parametric rate: $1/n$. It also indicates that the term $\mathbb{E}[|U^*_{f_m}(Y_1)|^2]/n$ will influence the rate of convergence if f_ε is ordinary smooth. Moreover, using

$$2\frac{\|f_Y\|_\infty \Delta^{(1)}(m)}{n}\|f - f_m\|^2 \leqslant \|f - f_m\|^4 + \frac{\|f_Y\|^2_\infty (\Delta^{(1)}(m))^2}{n^2},$$

it holds by Lemma 3.1 (i):

$$\frac{\mathbb{E}[|U^*_{f_m}(Y_1)|^2]}{n} \leqslant \|f - f_m\|^4 + \frac{\|f_Y\|^2_\infty}{n(n-1)}\Delta(m)\Delta^{(1)}(m) + \frac{\|f_Y\|_\infty}{\pi n}\int_{|u|\leqslant\sqrt{\rho m}} \frac{|f^*(-u)|^2}{|f^*_\varepsilon(u)|^2} du + \frac{C_0'}{n}\Lambda(m). \tag{15}$$

The first two terms are the same as in Proposition 3.1 (ii)) with two additional terms $\frac{\|f_Y\|_\infty}{\pi n}\int_{|u|\leqslant\sqrt{\rho m}} \frac{|f^*(-u)|^2}{|f^*_\varepsilon(u)|^2} du + \frac{C_0'}{n}\Lambda(m)$. Under regularity conditions on f, the term $\frac{1}{n}\int_{|u|\leqslant\sqrt{\rho m}} \frac{|f^*(-u)|^2}{|f^*_\varepsilon(u)|^2} du$ is negligible for an adequate choice of m compared to $\Delta(m)\Delta^{(1)}(m)/n(n-1)$. The last term $\Lambda(m)/n$ has order $1/n$ for any value of $m \in \mathbb{N}$ and is therefore negligible. The second part is applicable in the super smooth case and shows that the rate depends on $\Delta^{(1)}(m)/n$.

Remark 1 (Rate of convergence from direct observations)**.** *When X_k are available, $\varepsilon_k = 0$ a.s. and the terms $\Delta^{(1)}(m)$ and $\Lambda(m)$ are bounded uniformly in m. Thus, $\mathbb{E}[|U^*_{f_m}(Y_1)|^2]/n$ is a residual term compared to the term $C_1\Delta(m)\Delta^{(1)}(m)/n(n-1) \lesssim \sqrt{m}/n(n-1)$. For $f \in W^s_H(L)$ (see Definition 2.1), we recover the rate $n^{-8s/(4s+1)} \vee n^{-1}$ computed respectively over Lipschitz and Besov regularity spaces by Bickel and Ritov (1988), Laurent (2005). Our results contain the case of direct observation.*

3.2. **Rate of convergence on Sobolev-Hermite ball.** In order to derive rates of convergence, we consider the Sobolev-Hermite regularity spaces, defined in Definition 2.1.

Proposition 3.2 (Order of bias and variance)**.** *Let $(Y_k)_{k\in\{1,\dots,n\}}$ be n i.i.d. observations drawn from Model (1) and assume that assumptions (**H1**) to (**H5**) hold. Assume that $\|f\|_\infty < \infty$, f belongs to $W^s_H(L)$ and f_ε is square integrable. Consider $\widehat{\theta}_m(f)$ be defined in (11).*

(i) If $\rho \geqslant 2$ and f_ε is ordinary smooth (that is f_ε satisfies (7) with $\delta = 0$), we have

$$\mathbb{E}[(\widehat{\theta}_m(f) - \theta(f))^2] \leqslant 2Lm^{-2s} + \mathfrak{C}\left(\frac{1}{n(n-1)}m^{2\gamma+\frac{1}{2}} + \frac{1}{n}m^{-(s-\gamma)}\mathbb{1}_{s\leqslant\gamma} + \frac{1}{n}\right),$$

where $\mathfrak{C}$ is a constant depending on f_ε, L, $\|f\|_\infty$ and $\|f_Y\|_\infty$.

(ii) If $\rho \geqslant 2$ and f_ε is super smooth (that is f_ε satisfies (7) with $0 \leqslant \delta < 2$ or ($\delta = 2$, with $\mu < \xi$)), where ξ is defined in (4), it holds

$$\mathbb{E}[(\widehat{\theta}_m(f) - \theta(f))^2] \leqslant Lm^{-2s} + \mathfrak{C}_0\left(\frac{1}{n(n-1)}m^{2\gamma+\frac{1}{2}}e^{2\mu(m\rho)^{\frac{\delta}{2}}} + \frac{1}{n}m^{\gamma}e^{\mu(m\rho)^{\frac{\delta}{2}}} + \frac{1}{n}\right),$$

where $\mathfrak{C}_0$ is also a constant depending on f_ε, L, $\|f\|_\infty$ and $\|f_Y\|_\infty$.

Proposition 3.2 is a consequence of Proposition 3.1 and Lemma 3.1. Part(i) of Proposition 3.2 can be seen as an equivalent version of Proposition 1 obtained by Butucea (2007) in the kernel strategy. By realizing a bias-variance compromise with respect to m in the risk bound in Proposition 3.2, we obtain rates of convergences.

Theorem 3.3 (Rate of convergence for ordinary smooth noise)**.** *Let $(Y_k)_{k\in\{1,\dots,n\}}$ be n i.i.d. observations from drawn Model (1) and assume that assumptions (**H**1) to (**H5**) hold. Suppose that $\|f\|_\infty < \infty$, f belongs to $W_H^s(L)$, and f_ε is square integrable and ordinary smooth. Consider $\widehat{\theta}_m(f)$ be defined in (11). Take $m_{opt} = [n^{\frac{4}{4s+4\gamma+1}}]$. Then, we have*

$$\sup_{f\in W_H^s(L)} \mathbb{E}[(\widehat{\theta}_{m_{opt}}(f) - \theta(f))^2] \lesssim \begin{cases} n^{-\frac{8s}{4s+4\gamma+1}} & \text{if } s < \gamma + \frac{1}{4}, \\ n^{-1} & \text{if } s \geqslant \gamma + \frac{1}{4}. \end{cases} \tag{16}$$

We attain the classical rate in the deconvolution scenario when f_ε exhibits ordinary smoothness, see Butucea (2007). This rate is known to be optimal under additional condition on the derivative of f_ε^* and for densities belonging to the classical Sobolev space. If $\gamma = 0$, corresponding to direct observation case, we achieve the rate $n^{-8s/(4s+1)} \vee n^{-1}$, see Bickel and Ritov (1988), Laurent (2005).

Theorem 3.4 (Rate of convergence for super smooth noise)**.** *Let $(Y_k)_{k\in\{1,\dots,n\}}$ be n i.i.d. observations drawn from Model (1) and assume that assumptions (**H**1) to (**H5**) hold. Suppose that $\|f\|_\infty < \infty$ and f belongs to $W_H^s(L)$, and f_ε is square integrable and super smooth with $0 < \delta \leqslant 2$. Let $\widehat{\theta}_m(f)$ be defined in (11). By selecting $m_{opt} = [\frac{1}{\rho}(\frac{\log n}{4\mu})^{\frac{2}{\delta}}]$, we obtain the following bound*

$$\sup_{f\in W_H^s(L)} \mathbb{E}[(\widehat{\theta}_{m_{opt}}(f) - \theta(f))^2] \lesssim (\log n)^{-\frac{4s}{\delta}}.$$

If f_ε is super smooth, we attain the optimal rate, computed over the classical Sobolev spaces $(\log n)^{-\frac{4s}{\delta}}$ (see Butucea (2007)). For the lower bound, Butucea (2007) considers test functions with compact support. Thus, the rate $(\log n)^{-\frac{4s}{\delta}}$ also represents the optimal rate over Sobolev-Hermite spaces, as these spaces coincide with the ones mentioned above (see Section 2.1.3) for compactly supported functions.
Let us summarize the previous rates in Table 1. Let us note that the rate obtained for the classes of mean mixture or variance mixtures of the Gaussian distribution in Sacko (2020) can be extended in our context. Specifically, if f is a Gaussian density and f_ε is ordinary smooth, we deduce from Proposition 3.1 and Lemma 3.1 that:

$$\sup_{f\in W_H^s(L)} \mathbb{E}[(\widehat{\theta}_{m_{opt}}(f) - \theta(f))^2] \lesssim \frac{\log^{2\gamma+\frac{1}{2}}(n)}{n^2} \vee \frac{\log^{\gamma-s}(n)}{n}\mathbb{1}_{\gamma>s},$$

	$\delta = 0$ ordinary smooth	$0 < \delta \leqslant 2$ super smooth
rate	$n^{-\frac{8s}{4s+4\gamma+1}} \vee n^{-1}$	$(\log n)^{-\frac{4s}{\delta}}$

TABLE 1. Rate of convergence for $\mathbb{E}[(\widehat{\theta}_{m_{opt}}(f) - \theta(f))^2]$ if $f \in W_H^s(D)$.

where $m_{opt} \asymp \log(n)$.
When f and f_ε are both exponentially decaying, the rate can be better than any power of logarithm. For instance, if $f(x) = \exp(-x^2/(2\sigma^2))$ and $g(x) = \exp(-x^2/(2\zeta^2))$, selecting $m_{opt} = (\sigma^2 + \zeta^2)^{-1} \log(n)$ yields $\sup_{f \in W_H^s(L)} \mathbb{E}[(\widehat{\theta}_{m_{opt}}(f) - \theta(f))^2] \lesssim n^{-\frac{\sigma^2}{\sigma^2+\zeta^2}}$.

4. Adaptive estimation

In this section, we propose a selection procedure for the estimator $\widehat{\theta}_m(f)$. First, let us note that the adaptive procedure from Model (1) has not been explored in the existing literature. We also note that the procedure by penalization introduced by Laurent (2005) for the direct observation is not adapted to our inverse problem. Indeed, it is based on the fact $\mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^n f(X_k) - \theta(f)\right)^2\right] \leqslant \frac{1}{n}\int_{\mathbb{R}} f^3(x)dx \leqslant \frac{\|f\|_\infty^2}{n}$ is a residual term. For the inverse problem case, by definition of U_t given in (10),

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^n U_f^*(Y_k) - \theta(f)\right)^2\right] \leqslant \frac{\|f_Y\|_\infty}{n}\int_{\mathbb{R}} \frac{|f^*(-u)|^2}{|f_\varepsilon^*(u)|^2}du.$$

Thus, we see that $\int_{\mathbb{R}} \frac{|f^*(-u)|^2}{|f_\varepsilon^*(u)|^2}du$ is not necessarily finite, for instance when $f_\varepsilon^*(u)$ decreases faster than $f^*(u)$ near infinity. This is why, we introduce a novel approach inspired by the Goldenshluger and Lepski (2011) method. The method is based on the comparison of estimators of $\widehat{\theta}_m(f)$.
From now on, let us assume that the constant $\rho \geqslant 2$ given in (12) is fixed. Consider the following collections of model:

$$\mathcal{M}_n := \left\{1 \leqslant m \leqslant \lfloor\sqrt{n}\rfloor, \quad \Delta(m)\Delta^{(1)}(m) \leqslant \frac{n(n-1)}{\log^2(n)}, \quad \Delta^{(1)}(m) \leqslant \frac{n}{\log(n)}\right\},$$

where $\Delta(m)$ and $\Delta^{(1)}(m)$ are respectively given in (12)-(13). We aim to identify the $m \in \mathcal{M}_n$ that minimizes the bias-variance decomposition of $\widehat{\theta}_m(f)$ given in Proposition 3.1(ii). The bias $\|f - f_m\|^4$ is estimated by:

$$\widehat{A}(m) = \max_{m' \in \mathcal{M}_n}\left\{\left((\widehat{\theta}_{m'}(f) - \widehat{\theta}_{m\wedge m'}(f))^2 - \kappa_1 V(m')\right)_+\right\}, \tag{17}$$

where $V(m) = V_1(m) + V_2(m)$ is the order of variance term up to $\log(n)$ factor:

$$V_1(m) = \left(\Delta(m)\Delta^{(1)}(m)\right)\frac{\log^2(n)}{n(n-1)}, \tag{18}$$

and

$$V_2(m) = \begin{cases} \left(\int_{|u|\leqslant\sqrt{\rho m}} \frac{|f_m(u)|^2}{|f_\varepsilon^*(u)|^2}du\right)\frac{\log(n)}{n} & \text{if } f_\varepsilon \text{ is ordinary smooth,} \\ \Delta^{(1)}(m)\frac{\log(n)}{n} & \text{if } f_\varepsilon \text{ is super smooth.} \end{cases} \tag{19}$$

Here, κ_1 is a numerical constant to be calibrated.

We shall prove that (see Proof of Theorem 4.1 provided below) for any given noise.

$$\mathbb{E}[\widehat{A}(m)] \leqslant 3\|f - f_m\|^2 + \mathrm{Var}(\widehat{\theta}_m(f)) + \mathcal{O}\left(\frac{1}{n}\right).$$

The presence of $\mathrm{Var}(\widehat{\theta}_m(f))$ is related to the term $V_2(m)$, which is not necessarily increasing with m in the ordinary smooth case and depends on $|f_m(u)|^2$. Note that $|f_m(u)|^2$ will be replaced later by an estimator (see Section 5). For super-smooth noises, we take an upper bound on $\int_{\sqrt{\rho m}} \frac{|f_m(u)|^2}{|f_\varepsilon^*(u)|^2} du$:

$$\int_{\sqrt{\rho m}} \frac{|f_m(u)|^2}{|f_\varepsilon^*(u)|^2} du \leqslant \Delta^{(1)}(m)\|f_m\|^2,$$

and this is why $V_2(m) = \Delta^{(1)}(m)\log(n)/n$ in this case. Moreover, in this case, we have

$$\mathbb{E}[\widehat{A}(m)] \leqslant 3\|f - f_m\|^2 + \mathcal{O}\left(\frac{1}{n}\right).$$

Select:

$$\widehat{m}_n = \arg\min_{m\in\mathcal{M}_n} \left\{\widehat{A}(m) + \kappa_2 V(m)\right\}, \tag{20}$$

where $\kappa_2 \geqslant \kappa_1 > 0$ is also a numerical constant which must be calibrated.

We can prove the following oracle result.

Theorem 4.1 (Oracle inequality)**.** *Let $(Y_k)_{k\in\{1,\dots,n\}}$ be n i.i.d. observations drawn from Model (1) and asume that assumptions (**H**1) to (**H**5) hold. Suppose also that $\|f\|_\infty < \infty$, f_ε is square integrable and Condition (7) is fulfilled. Let $\widehat{\theta}_{\widehat{m}_n}(f)$ be defined in (11) with $\widehat{m}_n$ selected in (20). There exists a constant κ_0 such that for any $\kappa_1 \geqslant \kappa_0$:*

$$\mathbb{E}[(\widehat{\theta}_{\widehat{m}_n}(f) - \theta(f))^2] \leqslant C \inf_{m\in\mathcal{M}_n}\left(\|f - f_m\|^4 + V(m) + \frac{\mathbb{E}[|U_{f_m}^*(Y_1)|^2]}{n}\right) + \frac{C'}{n},$$

where $C = \max(36, (6\kappa_2 + 3C_1))$ with C_1 is given Proposition 3.1 and C' is a numerical constant depending on f_ε, $\|f\|_\infty$ and $\|f_Y\|_\infty$.

The term $\mathbb{E}[|U_{f_m}^*(Y_1)|^2]/n$ has the same order as $\|f - f_m\|^4 + V(m)$, see (15) under adequate regularity and choice of m. The bound given in Theorem 4.1 is non asymptotic and shows that the estimator $\widehat{\theta}_{\widehat{m}_n}(f)$ realizes automatically a bias-variance trade-off up to $\log(n)$ terms in ordinary smooth cases. Unsurprisingly, we recover the optimal rate for super smooth noises. Indeed, it is known that a logarithm factor in the variance does not affect the rate of convergence. Combining Proposition 3.1 and Lemma 3.1, we get

$$\sup_{f\in W_H^s(L)} \mathbb{E}[(\widehat{\theta}_{\widehat{m}_n}(f) - \theta(f))^2] \lesssim \begin{cases} \left(\frac{n}{\log(n)}\right)^{-\frac{8s}{4s+4\gamma+1}} \vee \left(\frac{n}{\log(n)}\right)^{-1} & \text{if } f_\varepsilon \text{ is ordinary smooth,} \\ (\log n)^{-\frac{4s}{\delta}} & \text{if } f_\varepsilon \text{ is super smooth.} \end{cases}$$

Note that the logarithm factor is due to Goldenshluger and Lepski (2011)'s procedure. In addition, when X_k are available ($\gamma = 0$), we recover the rate $n^{-8s/(4s+1)} \vee n^{-1}$ up to logarithm factor contrary to Laurent (2005)'s where the rate n^{-1} can be achieved for n large enough.

The value κ_0 is not explicit here and depends on the unknown quantities such as $\|f\|_\infty$, $\|f_Y\|$. This is due to the control of our U-statistic of order 2 and to the quantity $\Lambda(m)$, which is bounded by a non

explicit constant, see (14). In practice, we set $\kappa_2 = 2\kappa_1$ and calibrate κ_1 through preliminary simulation experiments.

5. Numerical study

5.1. **Implementation.** In this section, we illustrate the performance of our methodology by implementing the estimator $\widehat{\theta}_m(f)$ defined in (11) for both direct and inverse problem scenarios. Consider the following distributions for the signal

(i) $\mathcal{N}(1, 1/4)$, $\theta(f) = \frac{1}{2\sqrt{\pi}} \approx 0.5641896$,

(ii) Gamma $\mathcal{G}(\alpha, \beta)$, with $\alpha = 1$ and $\beta = 4$, $\theta(f) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^2 \frac{\Gamma(2\alpha-1)}{(2\beta)^{2\alpha-1}} \approx 0.15625$,

(ii) Laplace $\mathcal{L}(\sqrt{2})$, $\theta(f) = \frac{1}{2\sqrt{2}} \approx 0.3535534$,

(iv) Cauchy standard, $\theta(f) = \frac{1}{2\pi} \approx 0.1591549$.

From now, we set $\rho = 4$. For the noise, we distinguish three cases: direct case (that is $\varepsilon = 0$ *a.s.*), ordinary smooth and super-smooth cases.

- **The direct and ordinary smooth cases**. We choose the following estimator of the variance $V(m)$:

$$\widetilde{V}(m) = \Delta(m)\Delta^{(1)}(m)\frac{\log^2(n)}{n(n-1)} + \widetilde{V}_2(m), \qquad \rho = 3,$$

where $\widetilde{V}_2(m)$ estimates $V_2(m)$ given in (18). We tested two estimators of $\widetilde{V}_2(m)$. They are all based on the estimation of $|f_m(u)|^2$. The first is an unbiased estimator of $|f_m(u)|^2$:

$$|\widehat{f_m(u)}|^2 = \sum_{j,j'=0}^{m-1} \widehat{a_j a_{j'}}\varphi_j(u)\varphi_{j'}(u), \qquad \widehat{a_j a_{j'}} = \frac{1}{n(n-1)}\sum_{\ell\neq k=1}^{n} U^*_{\varphi_j}(Y_k)U^*_{\varphi_{j'}}(Y_\ell), \tag{21}$$

where U_t is given (10). The second estimator has a bias but is much faster to compute. It is obtained by replacing $f_m(u)$ by the *Hermite density estimator* studied in Sacko (2020). It is defined as:

$$\widehat{f}_m(u) = \sum_{j=0}^{m-1} \widehat{a}_j\varphi_j(u), \quad \widehat{a}_j = \frac{1}{n}\sum_{k=1}^{n} U^*_{\varphi_j}(Y_k).$$

Both estimators exhibit equivalent performances. Then, we only implement the following estimator of $V_2(m)$ for risk evaluations:

$$\widetilde{V}_2(m) = \left(\int_{|u|\leqslant\sqrt{\rho m}} \frac{|\widehat{f}_m(u)|^2}{|f^*_\varepsilon(u)|^2}du\right)\frac{\log(n)}{n}.$$

This estimator is obtained by substituting $|f_m(u)|^2$ with $|\widehat{f}_m(u)|^2$ in (19). Set $\varepsilon = 0$ *a.s.* in the direct case and consider the Laplace noise in the ordinary smooth case with a density given by:

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

For the last case, $\Delta(m)$ and $\Delta^{(1)}(m)$ are respectively given by:

$$\Delta(m) = \int_{|u|\leqslant\sqrt{\rho m}}(1+\frac{u^2}{\lambda^2})^2du = \left(\sqrt{\rho m} + \frac{2}{3\lambda^2}\left(\sqrt{\rho m}\right)^3 + \frac{\left(\sqrt{\rho m}\right)^5}{5\lambda^4}\right)$$

and

$$\Delta^{(1)}(m) = \left(1+\frac{\rho m}{\lambda^2}\right)^2.$$

- **The super smooth case**. Take a Gaussian noise with density:

$$f_\varepsilon(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma_\varepsilon^2}, \quad f_\varepsilon^*(x) = e^{-\sigma_\varepsilon^2 x^2/2}, \quad \sigma_\varepsilon^2 = 1/10.$$

  The variance is defined in (18)-(19) with

$$\Delta(m) = 2\sqrt{\rho m}\left(\int_0^1 e^{u^2\sigma_\varepsilon^2\rho m} du\right), \qquad \Delta^{(1)}(m) = e^{\sigma_\varepsilon^2\rho m},$$

  where $\int_0^1 e^{u^2\sigma_\varepsilon^2\rho m} du$ is computed by a Riemann sum discretized in 100 points.

The adaptive method is described in three steps:

- For any $m \in \mathcal{M}_n$, compute $\widehat{A}(m) = \sup_{m' \in \mathcal{M}_n} \left\{ \left((\widehat{\theta}_{m'}(f) - \widehat{\theta}_{m \wedge m'}(f))^2 - \kappa_1 \mathrm{Va}(m') \right)_+ \right\}$, where $\widehat{\theta}_m(f)$ is defined in (11) and $\mathrm{Va}(m)$ corresponds to the variance term which is specified above.
- Choose $\widehat{m}_n$ via $\widehat{m}_n = \arg\min_{m \in \mathcal{M}_n} \left\{ \widehat{A}(m) + \kappa_2 \mathrm{Va}(m) \right\}$,
- Compute $\widehat{\theta}_{\widehat{m}_n}(f) = \sum_{j=0}^{\widehat{m}_n - 1} \widehat{a_j^2(f)}$, with $\widehat{a_j^2(f)}$ given by (11).

We mention that the coefficient $(\widehat{a_j^2(f)})_{0 \leqslant j \leqslant m-1}$ is computed by Riemann's approximation in the inverse problem case.

Calibration of κ_1 and κ_2. Following the idea developed by Lacour and Massart (2016), we set $\kappa_2 = 2\kappa_1$ and $\kappa_1 = 10^{-3}$ in the direct case, $\kappa_1 = 3 \times 10^{-3}$ in the Laplace noise case, and $\kappa_1 = 8 \times 10^{-4}$ in the Gaussian noise case.

Comparison with penalization method in the direct case. We compare our adaptive procedure with the penalization procedure introduced by Laurent (2005). In this case, the space S_m is selected by:

$$\widehat{m}_n^{(1)} = \arg\min_{m \in \mathcal{M}_n} \left\{ -\widehat{\theta}_m(f) + \mathrm{pen}(m) \right\}, \quad \widehat{\mathrm{pen}}(m) = \frac{\kappa^{(1)}}{n} \sqrt{(\widehat{\theta}_m(f) + 1)\sqrt{m}\log(\sqrt{m} + 1)},$$

and the adaptive estimator is given by

$$\widetilde{\theta} = \widehat{\theta}_{\widehat{m}_n^{(1)}}(f) - \widehat{\mathrm{pen}}(\widehat{m}_n^{(1)}) = \sup_{m \in \mathcal{M}_n} \left\{ \widehat{\theta}_m(f) - \widehat{\mathrm{pen}}(m) \right\},$$

where $\kappa^{(1)}$ is a constant which must be calibrated and $\widehat{\theta}_m(f)$ is given here by (11) with $\varepsilon = 0$ almost surely. In practice, we choose $\kappa^{(1)} = 0.1$.

5.2. **Numerical results.** First, we note that all results are rounded to three decimal places. Moreover, in all cases, we see that increasing the value of n leads to a smaller MSE, thereby improving the estimation. In Table 2-3:

- The first line shows the MSE values with standard deviations in parentheses, multiplied by 1000.
- The second line represents the average of the dimension selected by the adaptive methods, and in the last line, we provide the average over the repetitions of the parameter estimates for each case.
- The value of the true unknown parameter $\theta(f) = \int_{\mathbb{R}} f^2(x)dx$ (and the signal to noise ratio defined by $\mathrm{SNR} = \mathrm{Var}(X)/\mathrm{Var}(\varepsilon)$ in parenthesis for the deconvolution case, except the Cauchy distribution) are provided in bold in the first column.

In Table 2, we compare the proposed procedure to the penalization approach in the direct observation case. In general, both methods have equivalent performance and the results are very satisfactory.

In Table 3, we present simulation results for the deconvolution case, computed over 200 independent simulations. Unsurprisingly, we observe that the MSE obtained in the case $\varepsilon \equiv 0$ almost surely is smaller than the noisy case. These results illustrate the influence of noise through the signal-to-noise ratio values:

the larger this value is, the better the estimate, as shown by distribution (ii). For distribution (i), the procedure gives a poorer result compared to (ii), or (iii) due to the lower signal-to-noise ratio value. We also observe that the MSE obtained in the Laplace noise case is globally smaller than in the Gaussian noise case. These observations are consistent with theoretical results. For all distributions, the procedure tends to overestimate the true parameter when the SNR is lower.

	Our method		Penalization method	
$\theta(f)$ \ n	200	500	200	500
	$\mathbf{0.894}_{(\mathbf{1.440})}$	$0.460_{(0.602)}$	$0.917_{(1.476)}$	$0.460_{(0.610)}$
	14.66	14.17	27.8	28.01
0.564...	0.568	0.563	0.570	0.564
	$0.085_{(0.110)}$	$0.037_{(0.054)}$	$\mathbf{0.083}_{(\mathbf{0.111})}$	$\mathbf{0.033}_{(\mathbf{0.055})}$
	23.63	25.69	35.24	39.8
0.156...	0.155	0.154	0.156	0.155
	$\mathbf{1.126}_{(\mathbf{1.423})}$	$\mathbf{0.372}_{(\mathbf{0.454})}$	$1.158_{(1.462)}$	$0.376_{(0.453)}$
	13.51	15.6	25.8	31.58
0.354...	0.361	0.354	0.362	0.355
	$0.276_{(0.411)}$	$0.116_{(0.152)}$	$0.276_{(0.419)}$	$0.116_{(0.156)}$
	7.97	9.1	24.31	27.19
0.159...	0.158	0.159	0.159	0.160

TABLE 2. First line: empirical MSE $1000 \times \mathbb{E}[(\widehat{\theta}_{\widehat{m}_n}(f) - \theta(f))^2]$, with $1000\times$sd in parentheses (left: our method, right: penalization method); second line: mean of $\widehat{m}_n$ or $\widehat{m}_n^{(1)}$; third line: mean of $\widehat{\theta}_{\widehat{m}_n}(f)$ or $\widehat{\theta}_{\widehat{m}_n^{(1)}}(f)$ computed over 100 independent simulations in direct observation case.

6. Concluding remarks

In this paper, we propose a new estimation procedure of a quadratic functional of a density based on the expansion of f into the Hermite basis from Model (1). We demonstrate that our estimator achieves the optimal rate obtained in Butucea (2007). An adaptive procedure to select the relevant dimension is proposed, and we establish non asymptotic oracle inequality for the resulting estimator. Numerical experiments illustrate the convergence and good performance of our methodology.

To conclude, we outline some perspectives for future work:

- In this study, we consider that $(X_i)_{i\in 1,\dots,n}$ are i.i.d. One possible extension is to investigate the case of dependent X_i. For instance, we can replace Assumption (**H**1) by: $(X_i)_{i\geqslant 1}$ is strictly stationary and β-mixing.
- We can also account for uncertainty in the noise distribution, i.e., study the case where the density f_ε is unknown but estimated through an additional sample.
- One of the motivations for this work is the issue of goodness-of-fit testing. We can investigate whether our results extend to this framework.

These points will be studied in future work.

	$n = 200$		$n = 500$	
Noise / $\theta(f)$	Laplace	Gaussian	Laplace	Gaussian
0.564...(2.5)	$2.384_{(4.066)}$ 6.74 0.579	$2.949_{(8.484)}$ 6.23 0.576	$1.061_{(1.641)}$ 7.135 0.580	$1.144_{(1.908)}$ 6.84 0.580
0.156...(40)	$0.209_{(0.280)}$ 11.645 0.151	$0.714_{(0.824)}$ 9.69 0.135	$0.061_{(0.078)}$ 13.125 0.156	$0.225_{((0.237)}$ 10.975 0.145
0.354...(10)	$1.398_{(2.631)}$ 4 0.368	$1.350_{(2.544)}$ 4.46 0.368	$0.633_{(0.833)}$ 5.260 0.366	$0.780_{(1.189)}$ 5.54 0.368
0.159...(-)	$0.706_{(1.224)}$ 2.96 0.176	$0.610_{(0.926)}$ 2.4 0.174	$0.351_{(0.437)}$ 1.440 0.173	$0.309_{(0.383)}$ 3.715 0.173

TABLE 3. First line: empirical MSE $1000 \times \mathbb{E}[(\widehat{\theta}_{\widehat{m}_n}(f) - \theta(f))^2]$ (with 1000×sd); second line: mean of $\widehat{m}_n$; third line: mean of $\widehat{\theta}_{\widehat{m}_n}(f)$ computed over 200 independent simulations.

7. Proofs

7.1. Proof of Proposition 3.1.

7.1.1. *Proof of Part (i).* We have by the Pythagorean theorem

$$\mathbb{E}[(\widehat{\theta}_m(f) - \theta(f))^2] = \|f - f_m\|^4 + \mathrm{Var}(\widehat{\theta}_m(f)). \tag{22}$$

Let us bound $\mathrm{Var}(\widehat{\theta}_m(f))$. By definition of $\widehat{\theta}_m(f)$ given in (11), it holds

$$\mathrm{Var}(\widehat{\theta}_m(f)) = \sum_{j,k=0}^{m-1} \mathrm{Cov}(\widehat{a_j^2(f)}, \widehat{a_k^2(f)}) = \sum_{j,k=0}^{m-1} \mathbb{E}\left[\left(\widehat{a_j^2(f)} - \mathbb{E}[\widehat{a_j^2(f)}]\right) \left(\widehat{a_k^2(f)} - \mathbb{E}[\widehat{a_k^2(f)}]\right) \right]. \tag{23}$$

The following decomposition holds:

$$\begin{aligned}
\left(\widehat{a_j^2(f)} - \mathbb{E}[\widehat{a_j^2(f)}]\right) &= \frac{1}{n(n-1)} \sum_{\ell \neq \ell'=1}^{n} \frac{1}{4\pi^2} \langle \frac{e^{i\cdot Y_\ell}}{f_\varepsilon^*}, \varphi_j^* \rangle \langle \frac{e^{i\cdot Y_{\ell'}}}{f_\varepsilon^*}, \varphi_j^* \rangle - a_j^2(f) \\
&= \frac{1}{n(n-1)} \sum_{\ell \neq \ell'=1}^{n} \left(U^*_{\varphi_j}(Y_\ell) - \mathbb{E}[U^*_{\varphi_j}(Y_\ell)]\right) \left(U^*_{\varphi_j}(Y_{\ell'}) - \mathbb{E}[U^*_{\varphi_j}(Y_{\ell'})]\right) \\
&\quad + \frac{2}{n} a_j(f) \sum_{\ell=1}^{n} \left(U^*_{\varphi_j}(Y_\ell) - \mathbb{E}[U^*_{\varphi_j}(Y_\ell)]\right) \\
&= T_{1,j} + T_{2,j}. \qquad (24)
\end{aligned}$$

It follows that

$$\mathrm{Var}(\widehat{\theta}_m(f)) = \sum_{j,k=0}^{m-1} \mathrm{Cov}(T_{1,j}, T_{1,k}) + 2\sum_{j,k=0}^{m-1} \mathrm{Cov}(T_{1,j}, T_{2,k}) + \sum_{j,k=0}^{m-1} \mathrm{Cov}(T_{2,j}, T_{2,k}).$$

We state the following Lemma.

Lemma 7.1. *Let j and k be two positive integers. Let U_t be defined in (10). It yields*

$$\mathrm{Cov}(T_{1,j}, T_{1,k}) = \mathbb{E}[T_{1,j}T_{1,k}] = \frac{2}{n(n-1)}\left(\mathrm{Cov}\left(U^*_{\varphi_j}(Y_1), U^*_{\varphi_k}(Y_1)\right)\right)^2,$$

$$\mathrm{Cov}(T_{1,j}, T_{2,k}) = \mathbb{E}[T_{1,j}T_{2,k}] = 0,$$

$$\mathrm{Cov}(T_{2,j}, T_{2,k}) = \mathbb{E}[T_{2,j}T_{2,k}] = \frac{4}{n}a_j(f)a_k(f)\mathrm{Cov}(U^*_{\varphi_j}(Y_1), U^*_{\varphi_k}(Y_1)).$$

The proof of Lemma 7.1 is detailed in Section A of the supplementary material.
Applying Lemma 7.1, we derive that

$$\mathrm{Var}(\widehat{\theta}_m(f)) = \frac{2}{n(n-1)}\sum_{j,k=0}^{m-1}\left(\mathrm{Cov}\left(U^*_{\varphi_j}(Y_1), U^*_{\varphi_k}(Y_1)\right)\right)^2 + \frac{4}{n}\mathrm{Var}(U^*_{f_m}(Y_1)). \tag{25}$$

We bound the $\sum_{j,k=0}^{m-1}\left(\mathrm{Cov}\left(U^*_{\varphi_j}(Y_1), U^*_{\varphi_k}(Y_1)\right)\right)^2$. Let us notice that

$$\sum_{j,k=0}^{m-1}\left(\mathrm{Cov}\left(U^*_{\varphi_j}(Y_1), U^*_{\varphi_k}(Y_1)\right)\right)^2 \leqslant 2\sum_{j,k=0}^{m-1}\left|\mathbb{E}\left[U^*_{\varphi_j}(Y_1)U^*_{\varphi_k}(Y_1)\right]\right|^2 + 2\|f\|^4.$$

Moreover, by definition of U^*_t given (10) and from (5), we have

$$2\sum_{j,k=0}^{m-1}\left|\mathbb{E}\left[U^*_{\varphi_j}(Y_1)U^*_{\varphi_k}(Y_1)\right]\right|^2 = \frac{1}{\pi^2}\sum_{j,k=0}^{m-1}\left|\iint \frac{\varphi_j(x)}{f^*_\varepsilon(x)}\frac{\varphi_k(y)}{f^*_\varepsilon(y)}f^*_Y(x+y)dxdy\right|^2.$$

We decompose the integrals in 3 pieces:

$$\begin{aligned}
2\sum_{j,k=0}^{m-1}\left|\mathbb{E}\left[U^*_{\varphi_j}(Y_1)U^*_{\varphi_k}(Y_1)\right]\right|^2 &= \frac{1}{\pi}\sum_{j,k=0}^{m-1}\left|\iint \frac{\varphi_j(x)}{f^*_\varepsilon(x)}\frac{\varphi_k(y)}{f^*_\varepsilon(y)}f^*_Y(x+y)\mathbb{1}_{|x|,|y|\leqslant\sqrt{\rho m}}dxdy\right|^2 = S_1 \\
&\quad + \frac{2}{\pi}\sum_{j,k=0}^{m-1}\left|\iint \frac{\varphi_j(x)}{f^*_\varepsilon(x)}\frac{\varphi_k(y)}{f^*_\varepsilon(y)}f^*_Y(x+y)\mathbb{1}_{|x|>\sqrt{\rho m}}\mathbb{1}_{|y|\leqslant\sqrt{\rho m}}dxdy\right|^2 = S_2 \\
&\quad + \frac{1}{\pi}\sum_{j,k=0}^{m-1}\left|\iint \frac{\varphi_j(x)}{f^*_\varepsilon(x)}\frac{\varphi_k(y)}{f^*_\varepsilon(y)}f^*_Y(x+y)\mathbb{1}_{|x|,|y|>\sqrt{\rho m}}dxdy\right|^2 = S_3 \\
&= S_1 + S_2 + S_3.
\end{aligned}$$

We bound successively these above terms. Let

$$\Psi_k(x) = \int \frac{f^*_Y(x+y)}{f^*_\varepsilon(y)}\varphi_k(y)\mathbb{1}_{|y|\leqslant\sqrt{\rho m}}dy = \left\langle \frac{\tau_{-x}f^*_Y}{f^*_\varepsilon}\mathbb{1}_{|\cdot|\leqslant\sqrt{\rho m}}, \varphi_k\right\rangle,$$

where $(\tau_x h)(y) = h(y-x)$. Through Bessel's inequality (applied twice), it holds for the first term

$$S_1 = \frac{1}{\pi}\sum_{j,k=0}^{m-1}\left|\left\langle \frac{\Psi_k}{f_\varepsilon^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{\rho m}}, \varphi_j\right\rangle\right|^2 \leqslant \sum_{k=0}^{m-1}\int \frac{|\Psi_k(x)|^2}{|f_\varepsilon^*(x)|^2}\mathbb{1}_{|x|\leqslant\sqrt{\rho m}}dx$$
$$\leqslant \frac{1}{\pi}\iint \frac{|f_Y^*(x+y)|^2}{|f_\varepsilon^*(x)|^2|f_\varepsilon^*(y)|^2}\mathbb{1}_{|x|,|y|\leqslant\sqrt{\rho m}}dxdy.$$

Let us remark that by the Cauchy-Schwarz inequality and Parseval's theorem, we have $\int|f_Y^*(z)|dz \leqslant 2\pi\|f\|\|f_\varepsilon\| \leqslant 2\pi\|f\|_\infty^{\frac{1}{2}}\|f_\varepsilon\|$. Thus, with the substitution $u = x+y$ and $v = x$, we obtain for S_1:

$$S_1 \leqslant \frac{\Delta^{(1)}(m)}{\pi}\iint \frac{|f_Y^*(u)|^2}{|f_\varepsilon^*(v)|^2}\mathbb{1}_{|v|,|u-v|\leqslant\sqrt{\rho m}}dudv \leqslant 2\|f\|_\infty^{\frac{1}{2}}\|f_\varepsilon\|\Delta^{(1)}(m)\Delta(m).$$

Let $\Phi_j(x) = \left\langle \frac{\tau_{-x}f_Y^*}{f_\varepsilon^*}\mathbb{1}_{|\cdot|>\sqrt{\rho m}}, \varphi_k\right\rangle$. Using the Bessel and Cauchy-Schwarz inequalities, it yields for S_2

$$S_2 = \frac{2}{\pi}\sum_{j,k=0}^{m-1}\left|\left\langle \frac{\Phi_j}{f_\varepsilon^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{\rho m}}, \varphi_j\right\rangle\right|^2$$
$$\leqslant \frac{2}{\pi}\sum_{j=0}^{m-1}\int_{|y|\leqslant\sqrt{\rho m}}\frac{1}{|f_\varepsilon^*(y)|^2}\left|\int_{|x|>\sqrt{\rho m}} f_Y^*(x+y)\frac{\varphi_j(x)}{f_\varepsilon^*(x)}dx\right|^2 dy$$
$$\leqslant \frac{2}{\pi}\sum_{j=0}^{m-1}\int_{|y|\leqslant\sqrt{\rho m}}\frac{1}{|f_\varepsilon^*(y)|^2}\left(\int|f_Y^*(x+y)|^2dx\int_{|x|>\sqrt{\rho m}}\frac{|\varphi_j(x)|^2}{|f_\varepsilon^*(x)|^2}dx\right)dy$$
$$\leqslant 4\|f\|_\infty^{\frac{1}{2}}\|f_\varepsilon\|\Delta(m)\Lambda(m),$$

where we use in the last line $\int|f_Y^*(x+y)|^2dx \leqslant \int|f_Y^*(z)|dz \leqslant 2\pi\|f\|_\infty^{\frac{1}{2}}\|f_\varepsilon\|$. Finally, by making the substitution $u = x+y$ and $v = x$, and as $\int|f_Y^*(z)|dz \leqslant 2\pi\|f\|_\infty^{\frac{1}{2}}\|f_\varepsilon\|$, we have from the Cauchy-Schwarz inequality,

$$S_3 \leqslant \frac{1}{\pi}\sum_{j,k=0}^{m-1}\left(\iint|f_Y^*(x+y)|\frac{|\varphi_j(x)|^2}{|f_\varepsilon^*(x)|^2}\mathbb{1}_{|x|,|y|>\sqrt{\rho m}}dxdy\right)\times\left(\iint|f_Y^*(x+y)|\frac{|\varphi_j(y)|^2}{|f_\varepsilon^*(y)|^2}\mathbb{1}_{|x|,|y|>\sqrt{\rho m}}dxdy\right)$$
$$= \frac{1}{\pi}\left(\sum_{j=0}^{m-1}\iint|f_Y^*(x+y)|\frac{|\varphi_j(x)|^2}{|f_\varepsilon^*(x)|^2}dxdy\mathbb{1}_{|x|,|y|>\sqrt{\rho m}}\right)^2$$
$$\leqslant 4\pi\|f\|_\infty\|f_\varepsilon\|^2\Lambda^2(m).$$

Therefore, we get

$$\frac{2}{n(n-1)}\sum_{j,k=0}^{m-1}\left(\mathrm{Cov}\left(U_{\varphi_j}^*(Y_1), U_{\varphi_k}^*(Y_1)\right)\right)^2 \leqslant \frac{C_0}{n(n-1)}\Delta(m)\left(\Delta^{(1)}(m)+\Lambda(m)\right) + \frac{C_0^2}{n(n-1)}\Lambda^2(m)$$
$$+\frac{4\|f\|_\infty^2}{n(n-1)}. \tag{26}$$

Injecting (26) in (25), we deduce

$$\mathrm{Var}(\widehat{\theta}_m(f)) \leqslant \frac{C_0}{n(n-1)}\Delta(m)\left(\Delta^{(1)}(m)+\Lambda(m)\right)+\frac{C_0^2}{n(n-1)}\Lambda^2(m) + \frac{4\|f\|_\infty^2}{n(n-1)}+\frac{4}{n}\mathrm{Var}(U^*_{f_m}(Y_1)). \tag{27}$$

Plugging (27) in (22) completes the proof of Part (i).

7.1.2. *Proof of Part (ii).* It is a consequence of Part (i). It is based on the fact that $\Lambda(m)$ is bounded uniformly in m. Indeed, we have from (4)-(7) and $\rho \geqslant 2$ that (see also Proposition 3.1 (ii) given in Sacko (2020))

$$\Lambda(m)=\sum_{j=0}^{m-1}\int_{|u|>\sqrt{\rho m}}\frac{|\varphi_j(u)|^2}{|f^*_\varepsilon(u)|^2}du \leqslant C^2e^{-\xi\rho m}\sum_{j=0}^{m-1}\int_{|u|>\sqrt{\rho m}}(1+u^2)^\gamma e^{\mu|u|^\delta}e^{-\xi u^2}du.$$

Since $\int_{|u|>\sqrt{\rho m}}(1+u^2)^\gamma e^{\mu|u|^\delta}e^{-\xi u^2}du<\infty$ if $\delta<2$ or $\delta=2$ and $\xi\leqslant\mu$, it follows that

$$\Lambda(m)\lesssim me^{-\xi m}\lesssim 1.$$

This implies Part (ii) and concludes the proof.

7.2. **Proof of Proposition 3.2.** We only prove the ordinary smooth case. The super smooth is handled in the same way using Part (ii) of Lemma 3.1. By Proposition 3.1 (ii) and (15), it yields

$$\mathbb{E}[(\widehat{\theta}_m(f)-\theta(f))^2] \leqslant 2\|f-f_m\|^4+\frac{C_1+16\|f_Y\|_\infty^2}{n(n-1)}\Delta(m)\Delta^{(1)}(m) + \frac{4\|f_Y\|_\infty}{\pi n}\int_{|u|\leqslant\sqrt{\rho m}}\frac{|f^*(-u)|^2}{|f^*_\varepsilon(u)|^2}du+\frac{C}{n},$$

where C is an absolute constant. We evaluate the order of each term. For $f\in W^s_H(L)$, we recall that $\|f-f_m\|^4=\left(\sum_{j\geqslant m}a_j^2(f)\right)^2\leqslant L^2m^{-2s}$. From Lemma 1 in Comte and Lacour (2011) p.586 and elementary computation, it yields under (7):

$$\Delta(m)\asymp m^{\gamma+\frac{1-\delta}{2}}e^{\mu\rho^{\frac{\delta}{2}}m^{\frac{\delta}{2}}}, \qquad \Delta^{(1)}(m)\leqslant c_1(1+\rho m)^\gamma e^{\mu(\rho m)^{\frac{\delta}{2}}}. \tag{28}$$

As $f\in W^s_H(L)\subset W^s(L')$ (see Section 2.1.3) and f_ε is ordinary smooth, it holds

$$\int_{|u|\leqslant\sqrt{\rho m}}\left|\frac{f^*(-u)}{f^*_\varepsilon(u)}\right|^2du\leqslant c_1\int_{|u|\leqslant\sqrt{\rho m}}(1+u^2)^s|f^*(-u)|^2(1+u^2)^{-(s-\gamma)}du\leqslant L'c_1\left(\mathbb{1}_{s\geqslant\gamma}+(1+\rho m)^{-(s-\gamma)}\mathbb{1}_{s<\gamma}\right),$$

where c_1 is given in (7). Denote by $\mathfrak{C}>0$ the maximum among constants that appear in the previous upper bound of $\mathbb{E}[(\widehat{\theta}_m(f)-\theta(f))^2]$, obtained by computing the order of each term. Setting $\delta=0$ in the above bounds, we get $f\in W^s_H(L)$ and (7)

$$\mathbb{E}[(\widehat{\theta}_m(f)-\theta(f))^2]\leqslant 2L^2m^{-2s}+\mathfrak{C}\left(\frac{1}{n(n-1)}m^{2\gamma+\frac{1}{2}}+\frac{1}{n}m^{-(s-\gamma)}\mathbb{1}_{s<\gamma}+\frac{1}{n}\right).$$

Hence the result for the ordinary smooth case.

7.3. **Proof of Theorem 3.3 and 3.4.** The ideas is to replace m by m_{opt} in Proposition 3.2. We distinguish two cases: ordinary smooth and super smooth cases.

- **The ordinary smooth case.** Plugging the value $m_{opt} = [n^{\frac{4}{4s+4\gamma+1}}]$ in the bound given in Proposition 3.2 (i), it yields

$$\mathbb{E}[(\widehat{\theta}_{m_{opt}}(f) - \theta(f))^2] \lesssim \left(n^{-\frac{8s}{4s+4\gamma+1}} + n^{-\frac{8s+1}{4s+4\gamma+1}} + n^{-1}\right) \lesssim \left(n^{-\frac{8s}{4s+4\gamma+1}} \vee n^{-1}\right).$$

  The result is proven for ordinary smooth noise.
- **The super smooth case.** Injecting $m_{opt} = \frac{1}{\rho}(\frac{\log n}{4\mu})^{\frac{2}{\delta}}]$ in Proposition 3.2 (ii), one obtains

$$\mathbb{E}[(\widehat{\theta}_m(f) - \theta(f))^2] \lesssim (\log(n)^{-\frac{4s}{\delta}}).$$

This concludes the proof of Theorem 3.3 and 3.4.

7.4. **Proof of Theorem 4.1.** For the sake of simplicity, we set $\widehat{\theta}_m(f) = \widehat{\theta}_m$. By definition of $\widehat{A}(m)$ given in (17) and $\widehat{m}_n$ in (20) and $\forall m \in \mathcal{M}_n$, we have the following decomposition:

$$\begin{aligned}(\widehat{\theta}_{\widehat{m}_n} - \theta)^2 &\leqslant 3(\widehat{\theta}_{\widehat{m}_n} - \widehat{\theta}_{\widehat{m}_n \wedge m})^2 + 3(\widehat{\theta}_{\widehat{m}_n \wedge m} - \widehat{\theta}_m)^2 + 3(\widehat{\theta}_m - \theta)^2 \\ &\leqslant 3(\widehat{A}(\widehat{m}_n) + \kappa_1 V(\widehat{m}_n)) + 3(\widehat{A}(m) + \kappa_1 V(m)) + 3(\widehat{\theta}_m - \theta)^2 \\ &\leqslant 6\widehat{A}(m) + 6\kappa_2 V(m) + 3(\widehat{\theta}_m - \theta)^2.\end{aligned}$$

By taking the expectation, one obtains

$$\mathbb{E}\left[(\widehat{\theta}_{\widehat{m}_n} - \theta(f))^2\right] \leqslant 6\,\mathbb{E}[\widehat{A}(m)] + 6\kappa_2 V(m) + 3\,\mathbb{E}[(\widehat{\theta}_m - \theta(f))^2]. \tag{29}$$

We bound $\mathbb{E}[\widehat{A}(m)]$. Decompose $(\widehat{\theta}_{m' \wedge m} - \widehat{\theta}_{m'})^2$ as follows:

$$(\widehat{\theta}_{m' \wedge m} - \widehat{\theta}_{m'})^2 \leqslant 3(\widehat{\theta}_{m' \wedge m} - \mathbb{E}[\widehat{\theta}_{m' \wedge m}])^2 + 3(\mathbb{E}[\widehat{\theta}_{m' \wedge m}] - \mathbb{E}[\widehat{\theta}_{m'}])^2 + 3(\widehat{\theta}_{m'} - \mathbb{E}[\widehat{\theta}_{m'}])^2. \tag{30}$$

Consequently, it yields

$$\begin{aligned}\widehat{A}(m) \leqslant & 3 \sup_{m' \in \mathcal{M}_n} \left\{ \left((\widehat{\theta}_{m' \wedge m} - \mathbb{E}[\widehat{\theta}_{m' \wedge m}])^2 - \frac{\kappa_1}{6} V(m') \right)_+ \right\} + 3 \sup_{m' \in \mathcal{M}_n} \left\{ \left((\widehat{\theta}_{m'} - \mathbb{E}[\widehat{\theta}_{m'}])^2 - \frac{\kappa_1}{6} V(m') \right)_+ \right\} \\ & + 3 \sup_{m' \in \mathcal{M}_n} \left\{ (\mathbb{E}[\widehat{\theta}_{m' \wedge m}] - \mathbb{E}[\widehat{\theta}_{m'}])^2 \right\}.\end{aligned}$$

Observe that, if $m' \leqslant m$, the last term is equal to zero. So, by definition of $\widehat{\theta}_m$ given in (11), we have

$$\begin{aligned}\sup_{m' \in \mathcal{M}_n} \left\{ (\mathbb{E}[\widehat{\theta}_{m' \wedge m}] - \mathbb{E}[\widehat{\theta}_{m'}])^2 \right\} &= \sup_{m' \in \mathcal{M}_n, m < m'} \left\{ (\mathbb{E}[\widehat{\theta}_{m'}] - \mathbb{E}[\widehat{\theta}_m])^2 \right\} = \sup_{m' \in \mathcal{M}_n, m < m'} \left\{ (\|f_{m'}\|^2 - \|f_m\|^2)^2 \right\} \\ &\leqslant \sup_{m' \in \mathcal{M}_n, m < m'} \left\{ (\|f\|^2 - \|f_m\|^2)^2 \right\} = \|f - f_m\|^4,\end{aligned}$$

where the last inequality is obtained by using the Pythagorean theorem. It comes that:

$$\begin{aligned}\widehat{A}(m) \leqslant & 3 \sup_{m' \in \mathcal{M}_n} \left\{ \left((\widehat{\theta}_{m'} - \mathbb{E}[\widehat{\theta}_{m'}])^2 - \frac{\kappa_1}{6} V(m') \right)_+ \right\} + 3 \sup_{m' \in \mathcal{M}_n} \left\{ \left((\widehat{\theta}_{m' \wedge m} - \mathbb{E}[\widehat{\theta}_{m' \wedge m}])^2 - \frac{\kappa_1}{6} V(m') \right)_+ \right\} \\ & + 3\|f - f_m\|^4.\end{aligned}$$

As

$$\sup_{m'\in\mathcal{M}_n}\left\{\left((\widehat{\theta}_{m'\wedge m}-\mathbb{E}[\widehat{\theta}_{m'\wedge m}])^2-\frac{\kappa_1}{6}V(m')\right)_+\right\}$$
$$\leqslant \max\left(\sup_{m'\in\mathcal{M}_n,m'\geqslant m}\left\{\left((\widehat{\theta}_m-\mathbb{E}[\widehat{\theta}_m])^2-\frac{\kappa_1}{6}V(m')\right)_+\right\},\sup_{m'\in\mathcal{M}_n,m'<m}\left\{\left((\widehat{\theta}_{m'}-\mathbb{E}[\widehat{\theta}_{m'}])^2-\frac{\kappa_1}{6}V(m')\right)_+\right\}\right)$$
$$\leqslant \sup_{m'\in\mathcal{M}_n,m'<m}\left\{\left((\widehat{\theta}_{m'}-\mathbb{E}[\widehat{\theta}_{m'}])^2-\frac{\kappa_1}{6}V(m')\right)_+\right\}+(\widehat{\theta}_m-\mathbb{E}[\widehat{\theta}_m])^2,$$

we deduce that

$$\mathbb{E}[\widehat{A}(m)]\leqslant 3\|f-f_m\|^4+6\,\mathbb{E}\left[\sup_{m'\in\mathcal{M}_n,m'<m}\left\{\left((\widehat{\theta}_{m'}-\mathbb{E}[\widehat{\theta}_{m'}])^2-\frac{\kappa_1}{6}V(m')\right)_+\right\}\right]+3\mathrm{Var}(\widehat{\theta}_m).$$

By (27), (14), it holds

$$\mathrm{Var}(\widehat{\theta}_m)\leqslant C_1V(m)+\frac{4}{n}\,\mathbb{E}[|U^*_{f_m}(Y_1)|^2]+\frac{C_2}{n},$$

where C_1 and C_2 are constant appearing in Proposition 3.1 (ii) and U_t is given in (10). This implies

$$\mathbb{E}[\widehat{A}(m)]\leqslant 3\|f-f_m\|^4+6C_1V(m)+\frac{24}{n}\,\mathbb{E}[|U^*_{f_m}(Y_1)|^2]+\frac{6C_2}{n}$$
$$+6\,\mathbb{E}\left[\sup_{m'\in\mathcal{M}_n}\left\{\left((\widehat{\theta}_{m'}-\mathbb{E}[\widehat{\theta}_{m'}])^2-\frac{\kappa_1}{6}V(m')\right)_+\right\}\right].$$

Now, we need the following result which leads to control the last term of the above bound.

Proposition 7.1. *Under the assumptions of Theorem 4.1, there exists a constant $\kappa_1\geqslant\kappa_0$ such that*

$$\mathbb{E}\left[\sup_{m'\in\mathcal{M}_n}\left\{\left((\widehat{\theta}_{m'}-\mathbb{E}[\widehat{\theta}_{m'}])^2-\frac{\kappa_1}{6}V(m')\right)_+\right\}\right]\leqslant\frac{C^{(0)}}{n},$$

where $C^{(0)}$ depends on $f_\varepsilon,\|f\|_\infty$ and $\|f_Y\|_\infty$.

Mainly, the constant κ_0 depends on $\|f\|_\infty$, $\|f_Y\|_\infty$, and the density f_ε. Proposition 7.1 and 3.1 (ii) and (29) imply that

$$\mathbb{E}\left[(\widehat{\theta}_{\widehat{m}_n}(f)-\theta(f))^2\right]\leqslant 21\|f-f_m\|^2+(6\kappa_2+9C_1)\,V(m)+\frac{36}{n}\,\mathbb{E}\left[|U^*_{f_m}(Y_1)|^2\right]+\frac{36C^{(0)}+9C_2}{n},$$

where U_t is defined in (10).
Choosing $C=\max(36,(6\kappa_2+3C_1))$ and $C'=36C^{(0)}+3C_2$ and then taking the infimum on $\mathcal{M}_n$ completes the proof.

7.5. **Proof of Proposition 7.1.** Define a degenerate U-statistic of order 2 by:

$$U_n(H_m)=\frac{1}{n(n-1)}\sum_{k\neq\ell=1}^{n}H_m(Y_k,Y_\ell),\quad H_m(x,y)=\sum_{j=0}^{m-1}(U^*_{\varphi_j}(x)-a_j)(U^*_{\varphi_j}(y)-a_j)\tag{31}$$

where $a_j=a_j(f)=\int f\varphi_j$ and U_t is defined in (10) and the empirical measure by

$$P_n(t):=\frac{1}{n}\sum_{i=1}^{n}U^*_t(Y_i)-\int_{\mathbb{R}}tf.\tag{32}$$

From (24), we have:

$$\widehat{\theta}_m - \mathbb{E}[\widehat{\theta}_m] = U_n(H_m) + P_n(2f_m).$$

This implies that

$$\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left((\widehat{\theta}_m - \mathbb{E}[\widehat{\theta}_m])^2 - \frac{\kappa_1}{6}V(m)\right)_+\right\}\right] \leqslant 2\,\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(U_n^2(H_m) - \frac{\kappa_1}{12}V(m)\right)_+\right\}\right] + 2\,\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(P_n^2(2f_m) - \frac{\kappa_1}{12}V(m)\right)_+\right\}\right]. \tag{33}$$

We bound successively the two terms in the right hand side of (33).

Bounding of $\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(U_n^2(H_m) - \frac{\kappa_1}{12}V(m)\right)_+\right\}\right]$**.** First, we decompose the $U_n(H_m)$ given by (31) in three processes by splitting the integrals. By Plancherel-Parseval's formula, it holds

$$\begin{aligned}
H_m(Y_k,Y_\ell) &= \sum_{j=0}^{m-1}\left(\frac{1}{2\pi}\int_{\mathbb{R}} e^{iuY_k}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du - \mathbb{E}[\frac{1}{2\pi}\int_{\mathbb{R}} e^{iuY_k}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du]\right)\\
&\quad\times\left(\frac{1}{2\pi}\int_{\mathbb{R}} e^{iuY_\ell}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du - \mathbb{E}[\frac{1}{2\pi}\int_{\mathbb{R}} e^{iuY_\ell}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du]\right)\\
&= \sum_{j=0}^{m-1}\left\{\left(\begin{array}{l}\frac{1}{2\pi}\int_{|u|\leqslant\sqrt{\rho m}} e^{iuY_k}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du - \mathbb{E}[\frac{1}{2\pi}\int_{|u|\leqslant\sqrt{\rho m}} e^{iuY_k}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du]\\ +\frac{1}{2\pi}\int_{|u|\geqslant\sqrt{\rho m}} e^{iuY_k}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du - \mathbb{E}[\frac{1}{2\pi}\int_{|u|\geqslant\sqrt{\rho m}} e^{iuY_k}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du]\end{array}\right)\right\}\\
&\quad\times\left\{\left(\begin{array}{l}\frac{1}{2\pi}\int_{|u|\leqslant\sqrt{\rho m}} e^{iuY_\ell}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du - \mathbb{E}[\frac{1}{2\pi}\int_{|u|\leqslant\sqrt{\rho m}} e^{iuY_\ell}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du]\\ +\frac{1}{2\pi}\int_{|u|\geqslant\sqrt{\rho m}} e^{iuY_\ell}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du - \mathbb{E}[\frac{1}{2\pi}\int_{|u|\geqslant\sqrt{\rho m}} e^{iuY_\ell}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du]\end{array}\right)\right\}.
\end{aligned}$$

Define

$$v_{\varphi_j}(x) = \frac{1}{2\pi}\int_{|u|\leqslant\sqrt{\rho m}} e^{iux}\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du. \tag{34}$$

and

$$\Psi_{\varphi_j}(Y_\ell) = \frac{1}{2\pi}\int_{|u|\geqslant\sqrt{\rho m}} \left(e^{iuY_\ell} - \mathbb{E}[e^{iuY_\ell}]\right)\frac{\varphi_j^*(-u)}{f_\varepsilon^*(u)}du. \tag{35}$$

Thanks to (34) and (35), we get

$$\begin{aligned}
H_m(Y_k,Y_\ell) &= \sum_{j=0}^{m-1}(v_{\varphi_j}(Y_k) - \mathbb{E}[v_{\varphi_j}(Y_k)])(v_{\varphi_j}(Y_\ell) - \mathbb{E}[v_{\varphi_j}(Y_\ell)]) + \sum_{j=0}^{m-1}(v_{\varphi_j}(Y_\ell) - \mathbb{E}[v_{\varphi_j}(Y_\ell)])\Psi_{\varphi_j}(Y_k)\\
&\quad + \sum_{j=0}^{m-1}(v_{\varphi_j}(Y_k) - \mathbb{E}[v_{\varphi_j}(Y_k)])\Psi_{\varphi_j}(Y_\ell) + \sum_{j=0}^{m-1}\Psi_{\varphi_j}(Y_k)\Psi_{\varphi_j}(Y_\ell).
\end{aligned}$$

Introduce:

$$U_n(I_m) = \frac{1}{n(n-1)} \sum_{k\neq \ell=1}^{n} I_m(Y_k, Y_\ell), \tag{36}$$
$$I_m(x,y) = \sum_{j=0}^{m-1} (v_{\varphi_j}(x) - \mathbb{E}[v_{\varphi_j}(Y_k)])(v_{\varphi_j}(y) - \mathbb{E}[v_{\varphi_j}(Y_\ell)]),$$

$$\nu_{n,1}(\varphi_j) = \frac{2}{n(n-1)} \sum_{k\neq \ell=1}^{n} \sum_{j=0}^{m-1} (v_{\varphi_j}(Y_k) - \mathbb{E}[v_{\varphi_j}(Y_k)])\Psi_{\varphi_j}(Y_\ell), \tag{37}$$

and

$$\nu_{n,2}(\varphi_j) = \frac{1}{n(n-1)} \sum_{k\neq \ell=1}^{n} \sum_{j=0}^{m-1} \Psi_{\varphi_j}(Y_k)\Psi_{\varphi_j}(Y_\ell). \tag{38}$$

We rewrite

$$U_n(H_m) = U_n(I_m) + \nu_{n,1}(\varphi_j) + \nu_{n,2}(\varphi_j),$$

and

$$\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(U_n^2(H_m) - \frac{\kappa_1}{12}V(m)\right)_+\right\}\right] \leqslant 3\,\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(U_n^2(I_m) - \frac{\kappa_1}{36}V(m)\right)_+\right\}\right] + 3\,\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\nu_{n,1}^2(\varphi_j)\right] + 3\,\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\nu_{n,2}^2(\varphi_j)\right].$$

We establish the following result.

Proposition 7.2. *Let $U_n(I_m)$ be defined in (36), $\nu_{n,1}(\varphi_j)$ in (37) and $\nu_{n,2}(\varphi_j)$ in (38). Under the assumptions of Theorem 4.1, the following hold.*

(i) For κ_1 large enough, depending only on f_ε, we have

$$\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(U_n^2(I_m) - \frac{\kappa_1}{36}V(m)\right)_+\right\}\right] \leqslant 4\rho\frac{(1\vee 4\|f\|_\infty)^2}{n}.$$

(ii) There exist two constants $C_{f_\varepsilon,1}$ and $C_{f_\varepsilon,2}$ depending on f_ε such that

$$\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\nu_{n,1}^2(\varphi_j)\right] \leqslant \frac{C_{f_\varepsilon,1}}{n}, \qquad \mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\nu_{n,2}^2(\varphi_j)\right] \leqslant \frac{C_{f_\varepsilon,2}}{n}.$$

The proof of Proposition 7.2 is postponed to Section A of the supplementary materiel. Proposition 7.2 implies

$$\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(U_n^2(H_m) - \frac{\kappa_1}{12}V(m)\right)_+\right\}\right] \leqslant 12\rho\frac{(1\vee 4\|f\|_\infty)^2}{n} + 3\frac{C_{f_\varepsilon,1} + C_{f_\varepsilon,2}}{n}, \tag{39}$$

Bounding of $\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(P_n^2(2f_m)-\frac{\kappa_1}{12}V(m)\right)_+\right\}\right]$. To control the deviation of $P_n(2f_m)$, we apply Bernstein inequality given in Section B.1 of the supplementary material. Let us first decompose the process $P_n(t)$ defined in (32) as follows: for any $t\in S_m$, we have using the Plancherel Parseval identity

$$\begin{aligned}
P_n(t) &= \frac{1}{n}\sum_{k=1}^{n}(U_t^*(Y_k)-\langle t,f\rangle)\\
&= \frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{2\pi}\int_{|u|\leqslant\sqrt{\rho m}}\frac{t^*(-u)}{f_\varepsilon^*(u)}e^{iuY_k}du-\mathbb{E}\left[\frac{1}{2\pi}\int_{|u|\leqslant\sqrt{\rho m}}\frac{t^*(-u)}{f_\varepsilon^*(u)}e^{iuY_k}du\right]\right)\\
&\quad+\frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{2\pi}\int_{|u|>\sqrt{\rho m}}\frac{t^*(-u)}{f_\varepsilon^*(u)}e^{iuY_k}du-\mathbb{E}\left[\frac{1}{2\pi}\int_{|u|>\sqrt{\rho m}}\frac{t^*(-u)}{f_\varepsilon^*(u)}e^{iuY_k}du\right]\right)\\
&= \frac{1}{n}\sum_{k=1}^{n}\left(v_t(Y_k)-\mathbb{E}\left[v_t(Y_k)\right]\right)+\frac{1}{2\pi}\int_{|u|>\sqrt{\rho m}}\frac{t^*(-u)}{f_\varepsilon^*(u)}(\widehat{f}_Y^*(u)-f_Y^*(u))du, \qquad (40)
\end{aligned}$$

with v_t is given (34) and $\widehat{f}_Y^*$ is defined in (9). Therefore, we write $P_n(t)=P_{n,1}(t)+P_{n,2}(t)$ where

$$P_{n,1}(t)=\frac{1}{n}\sum_{k=1}^{n}\left(v_t(Y_k)-\mathbb{E}\left[v_t(Y_k)\right]\right)$$

and

$$P_{n,2}(t)=\frac{1}{2\pi}\int_{|u|>\sqrt{\rho m}}\frac{t^*(-u)}{f_\varepsilon^*(u)}(\widehat{f}_Y^*(u)-f_Y^*(u))du.$$

It follows that

$$\begin{aligned}
\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(P_n^2(2f_m)-\frac{\kappa_1}{12}V(m)\right)_+\right\}\right] &\leqslant 2\,\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(P_{n,1}^2(2f_m)-\frac{\kappa_1}{24}V(m)\right)_+\right\}\right]\\
&\quad+2\,\mathbb{E}[\sup_{m\in\mathcal{M}_n}P_{n,2}^2(2f_m)]. \qquad (41)
\end{aligned}$$

We establish the following result, which is proved in Section A of the supplementary material.

Proposition 7.3. *Under the assumptions of Theorem 4.1, the following hold.*

(i) There exists a constant $C_{f_\varepsilon,3}$ such that for $\kappa_1\geqslant\max(384\|f_Y\|_\infty p, 384\|f\|_\infty p^2/(\alpha^2c_1'))$ with p an integer ($p=4$ suits) and $\alpha\in)0,1($

$$\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(P_{n,1}^2(2f_m)-\frac{\kappa_1}{24}V(m)\right)_+\right\}\right]\leqslant\frac{C_{f_\varepsilon,3}}{n}.$$

(ii) There exists a constant $C_{f_\varepsilon,4}$ such that

$$\mathbb{E}[\sup_{m\in\mathcal{M}_n}P_{n,2}^2(2f_m)]\leqslant\frac{C_{f_\varepsilon,4}}{n}.$$

From Proposition 7.3 and (41), we obtain

$$\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left(P_n^2(2f_m)-\frac{\kappa_1}{12}V(m)\right)_+\right\}\right]\leqslant 2\frac{C_{f_\varepsilon,4}+C_{f_\varepsilon,3}}{n}.$$

Finally, injecting this and (39) in (33), we derive that for κ_1 large enough

$$\mathbb{E}\left[\sup_{m\in\mathcal{M}_n}\left\{\left((\widehat{\theta}_m-\mathbb{E}[\widehat{\theta}_m])^2-\frac{\kappa_1}{6}V(m)\right)_+\right\}\right]\leqslant\frac{C^{(0)}}{n},$$

where $C^{(0)}=24\rho(1\vee 4\|f\|_\infty)^2+6(C_{f_\varepsilon,1}+C_{f_\varepsilon,2}+C_{f_\varepsilon,3}+C_{f_\varepsilon,4})$. Hence the announced result.

Acknowledgements: The author would like to thank F. Comte for helpful advices and kind proof-readings.

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Askey, R. and Wainger, S. (1965). Mean convergence of expansions in Laguerre and Hermite series. *Amer. J. Math.*, 87:695–708. MR182834.

Belomestny, D., Comte, F., and Genon-Catalot, V. (2019). Sobolev-Hermite versus Sobolev nonparametric density estimation on $\mathbb{R}$. *Ann. Inst. Statist. Math.*, 71(1):29–62.

Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A*, 50(3):381–393.

Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.*, 23(1):11–29.

Bongioanni, B. and Torrea, J. L. (2006). Sobolev spaces associated to the harmonic oscillator. *Proc. Indian Acad. Sci. Math. Sci.*, 116(3):337–360.

Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.*, 35(5):1907–1930.

Butucea, C., Rohde, A., and Steinberger, L. (2023). Interactive versus noninteractive locally differentially private estimation: two elbows for the quadratic functional. *Ann. Statist.*, 51(2):464–486.

Chesneau, C. (2011). On adaptive wavelet estimation of a quadratic functional from a deconvolution problem. *Ann. Inst. Statist. Math.*, 63(2):405–429.

Comte, F. and 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.

Efromovich, S. and Low, M. (1996). On Bickel and Ritov's conjecture about adaptive estimation of the integral of the square of density derivative. *Ann. Statist.*, 24(2):682–686.

Fromont, M. and Laurent, B. (2006). Adaptive goodness-of-fit tests in a density model. *Ann. Statist.*, 34(2):680–720.

Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632.

Goldenshluger, A. and Lepski, O. V. (2022a). Minimax estimation of norms of a probability density: I. Lower bounds. *Bernoulli*, 28(2):1120–1154.

Goldenshluger, A. and Lepski, O. V. (2022b). Minimax estimation of norms of a probability density: II. Rate-optimal estimation procedures. *Bernoulli*, 28(2):1155–1178.

Han, Y., Jiao, J., Weissman, T., and Wu, Y. (2020). Optimal rates of entropy estimation over Lipschitz balls. *Ann. Statist.*, 48(6):3228–3250.

Indritz, J. (1961). An inequality for Hermite polynomials. *Proc. Amer. Math. Soc.*, 12:981–983.

Kerkyacharian, G. and Picard, D. (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.*, 24(2):485–507.

Lacour, C. and Massart, P. (2016). Minimal penalty for Goldenshluger-Lepski method. *Stochastic Process. Appl.*, 126(12):3774–3789.

Laurent, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.*, 24(2):659–681.

Laurent, B. (2005). Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM Probab. Stat.*, 9:1–18.

Loubes, J. M. and Marteau, C. (2014). Goodness-of-fit testing strategies from indirect observations. *J. Nonparametr. Stat.*, 26(1):85–99.

Sacko, O. (2020). Hermite density deconvolution. *ALEA Lat. Am. J. Probab. Math. Stat.*, 17(1):419–443.

Schluttenhofer, S. and Johannes, J. (2020). Minimax testing and quadratic functional estimation for circular convolution. *arXiv preprint arXiv:2004.12714*.

Tribouley, K. (2000). Adaptive estimation of integrated functionals. *Math. Methods Statist.*, 9(1):19–38.