SOBOLEV-HERMITE VERSUS SOBOLEV NONPARAMETRIC DENSITY ESTIMATION ON $\mathbb{R}$

D. BELOMESTNY[(1)], F. COMTE[(2)] & V. GENON-CATALOT[(3)]

Abstract. In this paper, our aim is to revisit the nonparametric estimation of f assuming that f is square integrable on $\mathbb{R}$, by using projection estimators on a Hermite basis. These estimators are defined and studied from the point of view of their mean integrated squared error on $\mathbb{R}$. A model selection method is described and proved to perform an automatic bias variance compromise. Then, we present another collection of estimators, of deconvolution type, for which we define another model selection strategy. Considering Sobolev and Sobolev-Hermite spaces, the asymptotic rates of these estimators can be computed and compared: they are mainly proved to be equivalent. However, complexity evaluations prove that the Hermite estimators have a much lower computational cost than their deconvolution (or kernel) counterparts. These results are illustrated through a small simulation study.

(1) Duisburg-Essen Universität, Germany. email: denis.belomestny@uni-due.de
(2) MAP5, UMR CNRS 8145, Université Paris Descartes, France. email: fabienne.comte@parisdescartes.fr
(3) MAP5, UMR CNRS 8145, Université Paris Descartes, France. email: valentine.genon-catalot@parisdescartes.fr.

Keywords. Complexity. Density estimation. Hermite basis. Model selection. Projection estimator.
Mathematics Subject Classification. 62G05-62G07.

1. Introduction

Consider $X_1, \ldots, X_n$ n *i.i.d.* random variables with unknown density f. The nonparametric estimation of f has been the subject of such a huge number of contributions in the past decades that it is difficult to make an exhaustive list of references. Roughly speaking, there are two approaches, kernel or projection method. In the projection method which is our concern here, for f belonging to $\mathbb{L}^2(\mathbb{R})$, considering an orthonormal basis of this space, estimators are built by estimating a finite number of coefficients of the development of f on the basis. Fourier and wavelet bases, for instance, are commonly used. Bases of orthogonal polynomials are also used for compactly supported densities (see *e.g.* Donoho *et al.* (1996), Birgé and Massart (2007), and Efromovich (1999), Massart (2007), Tsybakov (2009) for reference books). For densities with non compact support included in $\mathbb{R}^+$, recent contributions use bases composed of Laguerre functions (see *e.g.* Comte and Genon-Catalot (2015), Belomestny *et al.* (2016), Mabon (2015)).

To our knowledge, for densities on $\mathbb{R}$, the use of a Hermite basis is only considered in Schwarz (1967) and Walter (1977). In this paper, our aim is to revisit the nonparametric estimation of f assuming that $f \in \mathbb{L}^2(\mathbb{R})$ by using projection estimators on a Hermite basis. To find asymptotic rates of convergence and optimize the risk bound, authors generally assume that the unknown density belongs to a function space specifying some regularity properties of f. Here, we consider the Sobolev-Hermite spaces which are naturally associated with the Hermite basis

Date: September 27, 2016.

and are defined in Bongioanni and Torrea (2006). It turns out that the Sobolev-Hermite space of regularity index s is included in the classical Sobolev space with same index. Therefore, we are led to compare the performances of the projection estimators on the Hermite basis with those of the deconvolution estimators which are projection estimators on the sine cardinal basis. Deconvolution estimators have been widely studied mainly for observations with additive noise and also for direct observations (see *e.g.* Comte *et al.* (2008)). The optimal $\mathbb{L}^2$-risk for density estimation on a Sobolev ball with regularity index s is of order $O(n^{-2s/(2s+1)})$, see Schipper (1996), Efromovich (2008) and Efromovich (2009) For densities having a fifth-order moment belonging to a Sobolev Hermite ball with the same regularity index s, we obtain the same rate. Therefore, from the asymptotic point of view, no difference can be made between these two classes of estimators at least for non heavy tailed densities. Other examples and counter-examples are discussed.

While most papers focus on deriving minimax convergence rates, the computational efficiency of the proposed estimator is not often considered. This issue is especially important for densities with non compact support. We prove that the Hermite estimators have a much lower complexity than the deconvolution estimators, resulting in a noteworthy computational gain.

In Section 2, we present the Hermite basis, and the $\mathbb{L}^2$-risk of the associated projection estimators is studied together with the possible orders for the variance term. A data-driven choice of the dimension is proposed and the associated estimator is proved to be realize adequately the bias-variance tradeoff. In Section 3, results on deconvolution estimators are presented. Section 4 is devoted to the study of asymptotic rates of convergence. From this point of view, the two approaches of the previous sections are proved to be equivalent, except in some special cases. Then, we compare the complexity of the procedures and conclude that the Hermite method has a substantial advantage from this point of view. Section 5 is devoted to numerical simulation results, and aims at illustrating the previous findings. Proofs are gathered in Section 6.

2. Projection estimators on the Hermite basis.

2.1. **Hermite basis.** Below, we denote by $\|.\|$ the $\mathbb{L}^2$-norm on $\mathbb{R}$ and by $\langle \cdot, \cdot \rangle$ the $\mathbb{L}^2$-scalar product.

The Hermite polynomial of order j is given, for $j \geq 0$, by:

$$H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j}(e^{-x^2}).$$

Hermite polynomials are orthogonal with respect to the weight function e^{-x^2} and satisfy: $\int_{\mathbb{R}} H_j(x) H_\ell(x) e^{-x^2} dx = 2^j j! \sqrt{\pi} \delta_{j,\ell}$ (see *e.g.* Abramowitz and Stegun (1964)). The Hermite function of order j is given by:

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

The sequence $(h_j, j \geq 0)$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R})$. The density f to be estimated can be developed in the Hermite basis $f = \sum_{j \geq 0} a_j(f) h_j$ where $a_j(f) = \int_{\mathbb{R}} f(x) h_j(x) dx = \langle f, h_j \rangle$.

We define $S_m = \text{span}(h_0, h_1, \ldots, h_{m-1})$ the linear space generated by the m functions $h_0, \ldots, h_{m-1}$ and $f_m = \sum_{j=0}^{m-1} a_j(f) h_j$ the orthogonal projection of f on S_m.

2.2. **Hermite estimator and risk bound.** Consider a sample $X_1, \ldots, X_n$ of *i.i.d.* random variables with density f, belonging to $\mathbb{L}^2(\mathbb{R})$. We define for each $m \geq 0$, $\hat{f}_m = \sum_{j=0}^{m-1} \hat{a}_j h_j$ a projection estimator of f, with $\hat{a}_j = n^{-1} \sum_{i=1}^n h_j(X_i)$, that is, an unbiased estimator of $f_m = \sum_{j=0}^{m-1} a_j(f) h_j$.

These estimators are considered in Schwartz (1967) and then in Walter (1977). As usual, the $\mathbb{L}^2$-risk is split into a variance and a square bias term. We give a more accurate rate for the variance term than in the latter papers. Indeed, we have the classical decomposition

$$\begin{aligned}\mathbb{E}(\|\hat{f}_m - f\|^2) &= \|f - f_m\|^2 + \sum_{j=0}^{m-1} \mathrm{Var}(\hat{a}_j) = \|f - f_m\|^2 + \frac{1}{n}\sum_{j=0}^{m-1} \mathrm{Var}(h_j(X_1)) \\ &\leq \|f - f_m\|^2 + \frac{V_m}{n}, \qquad (2)\end{aligned}$$

where

$$V_m = \int_{\mathbb{R}} \left(\sum_{j=0}^{m-1} h_j^2(x)\right) f(x)dx = \mathbb{E}(\sum_{j=0}^{m-1} h_j^2(X_1)). \tag{3}$$

The infinite norm of h_j satisfies (see Abramowitz and Stegun (1964), Szegö (1975) p.242):

$$\|h_j\|_\infty \leq \Phi_0, \qquad \Phi_0 \simeq 1,086435/\pi^{1/4} \simeq 0.8160. \tag{4}$$

Therefore, we have $V_m \leq \Phi_0^2 m$, as usual for projection density estimator, see Massart (2007), Chapter 7. However, more precise properties of the Hermite functions provide refined bounds:

Proposition 2.1.
(i) There exists constant c such that, for any density f and for any integer m,

$$V_m \leq c m^{5/6}.$$

(ii) If $\mathbb{E}|X|^5 < +\infty$, then there exists constant c' such that for any integer m,

$$V_m \leq c' m^{1/2}.$$

(iii) Assume that there exists $K > 0$ with

$$|f(x)| \leq g(x) := \alpha \frac{1}{(1+|x|)^a}, \text{ for } |x| \geq K \text{ and } \alpha > 0, a > 1.$$

Then, there exists c'' such that, for m large enough, $V_m \leq c'' m^{\frac{a+2}{2(a+1)}}$.

Proposition 2.1 (i) shows that V_m is at most of order $m^{5/6}$, a property obtained in Walter (1977). However (ii)-(iii) show that this order can be improved depending on additional assumptions on f.

In the next paragraph, we make no assumption on the regularity properties of f and propose a data-driven choice of the dimension m leading to an estimator whose $\mathbb{L}^2$-risk automatically realizes the bias-variance trade-off in a non asymptotic way.

2.3. **Model selection.** For model selection, we must estimate the bias and the variance term. Define $\mathcal{M}_n = \{1, \ldots, m_n\}$, where m_n is the largest integer such that $m_n^{5/6} \leq n/\log(n)$ and set

$$\hat{m} = \arg\min_{m \in \mathcal{M}_n} \{-\|\hat{f}_m\|^2 + \widehat{\mathrm{pen}}(m)\}, \quad \widehat{\mathrm{pen}}(m) = \kappa \frac{\widehat{V}_m}{n}, \quad \widehat{V}_m = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{m-1} h_j^2(X_i), \tag{5}$$

where κ is a numerical constant. The quantity $-\|\hat{f}_m\|^2$ estimates $-\|f_m\|^2 = \|f - f_m\|^2 - \|f\|^2$, and we can ignore the (unknown) constant term $\|f\|^2$. Usually, the penalty is chosen equal to

$\kappa\Phi_0^2 m/n$, which is the known upper bound of the variance term, where Φ_0 is defined by (4). Here, the fact that the order of V_m varies according to the assumptions on f justifies that we rather use $\widehat{V}_m$, an unbiased estimator of V_m. We can prove the following result.

Theorem 2.1. *Assume that f is bounded and that $\inf_{a\le x\le b} f(x) > 0$ for some interval $[a,b]$. Then there exists κ_0 such that, for $\kappa \ge \kappa_0$, the estimator $\hat f_{\hat m}$ where $\hat m$ is defined by (5) satisfies*

$$\mathbb{E}\left(\|\hat f_{\hat m} - f\|^2\right) \le C \inf_{m\in\mathcal{M}_n}\left(\|f-f_m\|^2 + \kappa\frac{V_m}{n}\right) + \frac{C'}{n},$$

where C is a numerical constant($C = 4$ suits) and C' is a constant depending on $\|f\|_\infty$.

The estimator $\hat f_{\hat m}$ is adaptive in the sense that its risk bound achieves automatically the bias-variance compromise, up to a negligible term of order $O(1/n)$. It follows from the proof that $\kappa_0 = 8$ is possible. This value of κ_0 is certainly not optimal; finding the optimal theoretical value of κ in the penalty is not an easy task, even in simple models (see for instance Birgé and Massart (2007) in a Gaussian regression model). This is why it is standard to calibrate the value κ in the penalty by preliminary simulations, as we do in Section 5.

Actually, the assumption $\inf_{a\le x\le b} f(x) > 0$ is due to the fact that the proof requires the condition

$$\forall m \ge m_0, V_m \ge 1, \text{ and } \forall a > 0, \sum_{m\in\mathcal{M}_n} e^{-a\sqrt{V_m}} \le A < +\infty. \tag{6}$$

Condition (6) holds, as we can prove:

Proposition 2.2. *If $\inf_{a\le x\le b} f(x) > 0$ for some interval $[a,b]$, then, for m large enough, $V_m \ge c'' m^{1/2}$ where c'' is a constant.*

3. Deconvolution estimators.

As we want to compare the performances of projection estimators on the Hermite basis to those of projection estimators on the sine cardinal basis, we recall the definition of the latter estimators, *i.e.* the deconvolution estimators. Let $\varphi(x) = \sin(\pi x)/(\pi x)$ which satisfies $\varphi^*(t) = 1_{[-\pi,\pi]}(t)$, where φ^* denotes the Fourier transform of φ. The functions $(\varphi_{\ell,j}(x) = \sqrt{\ell}\varphi(\ell x - j), j \in \mathbb{Z})$ constitute an orthonormal system in $\mathbb{L}^2(\mathbb{R})$. The space Σ_ℓ generated by this system is exactly the subspace of $\mathbb{L}^2(\mathbb{R})$ of functions having Fourier transforms with compact support $[-\pi\ell, \pi\ell]$. The orthogonal projection $\bar f_\ell$ of f on Σ_ℓ satisfies $\bar f_\ell^* = f^* 1_{[-\pi\ell,\pi\ell]}$. Therefore,

$$\|f - \bar f_\ell\|^2 = \frac{1}{2\pi}\int_{|t|\ge\pi\ell} |f^*(t)|^2 dt. \tag{7}$$

The projection estimator $\widetilde{f}_\ell$ of f is defined by:

$$\widetilde{f}_\ell(x) = \frac{1}{2\pi}\int_{-\pi\ell}^{\pi\ell} e^{-itx}\frac{1}{n}\sum_{k=1}^n e^{itX_k} dt = \frac{1}{n}\sum_{k=1}^n \frac{\sin(\pi\ell(X_k - x))}{\pi(X_k - x)}. \tag{8}$$

This expression corresponds to the fact that:

$$\bar f_\ell = \frac{1}{2\pi}\int_{-\pi\ell}^{\pi\ell} e^{-itx} f^*(t) dt = \sum_{j\in\mathbb{Z}} a_{\ell,j}\varphi_{\ell,j}(x), \qquad a_{\ell,j} = \langle f, \varphi_{\ell,j}\rangle.$$

Contrary to $\hat{f}_m$, the estimator $\widetilde{f}_\ell$ cannot be expressed as the corresponding sum with the estimated coefficients $\tilde{a}_{\ell,j} = \frac{1}{n}\sum_{k=1}^n \varphi_{\ell,j}(X_k)$ as this sum would be infinite and not defined. To compute it in concrete, one can use (8) or a truncated version

$$\widetilde{f}_\ell^{(n)}(x) = \sum_{|j|\le K_n} \tilde{a}_{\ell,j}\varphi_{\ell,j}(x), \qquad \tilde{a}_{\ell,j} = \frac{1}{n}\sum_{k=1}^n \varphi_{\ell,j}(X_k).$$

which creates an additional bias but is comparable to the previous Hermite estimator. We give the results for $\widetilde{f}_\ell$ and $\widetilde{f}_\ell^{(n)}$.

Proposition 3.1. *The estimator $\widetilde{f}_\ell$ satisfies*

$$\mathbb{E}(\|\widetilde{f}_\ell - f\|^2) \quad \le \quad \|f - \bar{f}_\ell\|^2 + \frac{\ell}{n}.$$

If moreover $M_2 = \int x^2 f^2(x)dx < +\infty$, then the estimator $\widetilde{f}_\ell^{(n)}$ satisfies

$$\mathbb{E}(\|\widetilde{f}_\ell^{(n)} - f\|^2) \quad \le \quad 2\|f - \bar{f}_\ell\|^2 + \frac{\ell}{n} + 4\frac{\ell^2(M_2+1)}{K_n}.$$

If $\ell \le n$ and $K_n \ge n^2$, the last term is of order $O(\ell/n)$ and can be associated to the variance term ℓ/n. Note that condition $K_n \ge n^2$ implies that the computation of a large number of coefficients is required for $\widetilde{f}_\ell^{(n)}$, for large n. In practice, we take K_n even smaller than n in order to keep reasonable computation times.
As in the previous case, we can define a data-driven choice of the cutoff parameter ℓ and build adaptive estimators:

$$\tilde{\ell} = \arg\min_{\ell\le n}\{-\|\widetilde{f}_\ell\|^2 + \tilde{\kappa}\frac{\ell}{n}\}, \quad \tilde{\ell}_n = \arg\min_{\ell\le n}\{-\|\widetilde{f}_\ell^{(n)}\|^2 + \tilde{\kappa}\frac{\ell}{n}\}, \tag{9}$$

where $\tilde{\kappa}$ is a numerical constant. Note that

$$\|\widetilde{f}_\ell\|^2 = \frac{1}{n^2}\sum_{1\le j,k\le n} \frac{\sin(\pi\ell(X_k - X_j))}{\pi(X_k - X_j)}, \quad \|\widetilde{f}_\ell^{(n)}\|^2 = \sum_{|j|\le K_n} |\tilde{a}_{\ell,j}|^2.$$

We give the result for $\widetilde{f}_\ell^{(n)}$ only, as $\|\widetilde{f}_\ell^{(n)}\|^2$ is faster to compute if K_n is chosen in a restricted range, $K_n \le n$, see Section 4.4 and Section 5.
The following result holds.

Theorem 3.1. *If $K_n \ge n^2$ and $M_2 = \int x^2 f^2(x)dx < +\infty$, then there exists a numerical constant $\tilde{\kappa}_0$ such that, for $\tilde{\kappa} \ge \tilde{\kappa}_0$, the estimator $\widetilde{f}_{\tilde{\ell}_n}^{(n)}$ where $\tilde{\ell}_n$ is defined by (9) satisfies*

$$\mathbb{E}\left(\|\widetilde{f}_{\tilde{\ell}_n}^{(n)} - f\|^2\right) \le C_1 \inf_{\ell\le n}\left(\|f - f_\ell\|^2 + \tilde{\kappa}\frac{\ell}{n} + \frac{\ell(M_2+1)}{n}\right) + \frac{C_2}{n},$$

where C_1 is a numerical constant and C_2 is a constant depending on $\|f\|_\infty$.

For $\widetilde{f}_{\tilde{\ell}}$, an analogous risk bound may be obtained, without condition $M_2 < +\infty$ and without the term $\ell(M_2+1)/n$ in the bound.
For Theorem 3.1, we refer to Comte *et al.* (2008), Proposition 5.1, p.97.

4. Comparison of rates of convergence and discussion.

In this section, we compute the rates of convergence that can be deduced from the optimization of the upper bounds of $\mathbb{L}^2$-risks. This requires to assess the rate of decay of the bias terms $\|f - f_m\|^2$ in the Hermite case, $\|f - \bar{f}_\ell\|^2$ in the deconvolution framework. The latter is usually obtained by assuming that the unknown density f belongs to a Sobolev space. For the former, we consider the Sobolev-Hermite spaces which are naturally linked with the Hermite basis.

4.1. **Sobolev and Sobolev-Hermite regularity.** For $s > 0$, the Sobolev-Hermite space with regularity s may be defined by:

$$W^s = \{f \in \mathbb{L}^2(\mathbb{R}), \|f\|^2_{s,sobherm} = \sum_{n\geq 0} n^s a_n^2(f) < +\infty\} \tag{10}$$

where $a_n(f) = \langle f, h_n\rangle$ is the n-th component of f in the Hermite basis. We refer to Bongioanni and Torrea (2006) for a definition using operator theory. Let $\mathcal{F} = \{\sum_{j\in J} a_j h_j, J \subset \mathbb{N}, \text{ finite }\}$ be the set of finite linear combinations of Hermite functions and C_c^∞ the set of infinitely derivable functions with compact support. The sets C_c^∞ and $\mathcal{F}$ are dense in W^s. As the Fourier transform of h_n satisfies

$$h_n^* = \sqrt{2\pi} i^n h_n, \tag{11}$$

$f \in W^s$ if and only if $f^* \in W^s$. We now describe W^s when s is integer. Let

$$A_+ f = f' + xf, \quad A_- f = -f' + xf$$

The following result is proved in Bongioanni and Torrea (2006). For sake of clarity, we give a simplified proof.

Proposition 4.1. *For s integer, the Sobolev-Hermite space W^s is equal to:*

$$\begin{aligned} W^s \; = \; & \{f \in \mathbb{L}^2(\mathbb{R}), f \textit{ admits derivatives up to order } s, \\ & |||f|||_{s,sobherm} = \sum_{\substack{j_1,\ldots,j_m \in \{-,+\},\\ 1\leq m\leq s}} \|A_{j_1}\ldots A_{j_m} f\| + \|f\| < +\infty\}. \end{aligned}$$

Moreover, the following statements are equivalent: for s integer,

1. *$f \in W^s$,*
2. *f admits derivatives up to order s which satisfy $f, f', \ldots, f^{(s)}, x^{s-\ell} f^{(\ell)}, \ell = 0, \ldots, s-1$ belong to $\mathbb{L}^2(\mathbb{R})$.*

The two norms $\|f\|_{s,sobherm}$ and $|||f|||_{s,sobherm}$ are equivalent.
Now, we recall the definition of usual Sobolev spaces. The Sobolev space with regularity index s is defined by

$$\mathcal{W}^s = \{f \in \mathbb{L}^2(\mathbb{R}), \|f\|^2_{s,sob} = \int_{\mathbb{R}} (1 + t^{2s})|f^*(t)|^2 dt < +\infty\} \tag{12}$$

If s is integer, then

$$\begin{aligned} \mathcal{W}^s \; = \; & \{f \in \mathbb{L}^2(\mathbb{R}), f \text{ admits derivatives up to order } s \\ & \text{such that } |||f|||^2_{s,sob} = \|f\|^2 + \|f'\|^2 + \ldots + \|f^{(s)}\|^2 < +\infty\}. \end{aligned}$$

The two norms $|||.|||_{s,sob}$ and $\|.\|_{s,sob}$ are equivalent. Therefore, for s integer, $W^s \subset \mathcal{W}^s$. Morevover, the following properties are proved in Bongioanni and Torrea (2006): for all $s > 0$,

- $W^s \subsetneq \mathcal{W}^s$. If $f \in \mathcal{W}^s$ has compact support, then $f \in W^s$.

•

$$f \in W^s \Rightarrow x^s f \in \mathbb{L}^2(\mathbb{R}). \tag{13}$$

4.2. **Rates of convergence.** Now, we look at asymptotic rates of convergence. We first consider rates for Hermite projection estimators. We already studied the variance rate V_m/n (see the bounds for V_m in Proposition 2.1). If f belongs to

$$W^s(L) = \{f \in \mathbb{L}^2(\mathbb{R}), \sum_{n\geq 0} n^s a_n^2(f) \leq L\},$$

then $\|f - f_m\|^2 \leq Lm^{-s}$. Plugging this and the bounds of Proposition 2.1 in Inequality (2) gives the following rates of the $\mathbb{L}^2(\mathbb{R})$-risk.

Proposition 4.2. *Assume that $f \in W^s(L)$ and consider the three cases (i), (ii), (iii) of Proposition 2.1.*
Case (i) (general case). For $m_{\rm opt} = [n^{1/(s+(5/6))}]$, $\mathbb{E}(\|\hat{f}_{m_{\rm opt}} - f\|^2) \lesssim n^{-\frac{s}{s+(5/6)}}$.
Case (ii). For $m_{\rm opt} = [n^{1/(s+(1/2))}]$, $\mathbb{E}(\|\hat{f}_{m_{\rm opt}} - f\|^2) \lesssim n^{-\frac{s}{s+1/2}}$.
Case (iii). For $m_{\rm opt} = [n^{1/(s+(a+2)/(2(a+1)))}]$, $\mathbb{E}(\|\hat{f}_{m_{\rm opt}} - f\|^2) \lesssim n^{-\frac{s}{s+(a+2)/[2(a+1)]}}$.

Case (ii) gives the best rate. Note that the rate in case (iii) is strictly better than in case (i) as $(a+2)/(a+1) < 5/3$ as soon as $a > 1/2$. Cases $(ii) - (iii)$ improve the results of Schwarz (1967) and Walter (1977).

Now, we can compare the rates to those of projection estimators in the sine cardinal basis. The following result is deduced from Proposition 3.1 and (7).

Proposition 4.3. *If $f \in \mathcal{W}^s(R) = \{f \in \mathbb{L}^2(\mathbb{R}), \|f\|^2_{s,sob} = \int_{\mathbb{R}}(1+t^{2s})|f^*(t)|^2dt \leq R\}$, and $\ell_{\rm opt} = n^{1/(2s+1)}$, we have $\mathbb{E}(\|\tilde{f}_{\ell_{\rm opt}} - f\|^2) \lesssim n^{-2s/(2s+1)}$.*
If moreover $K_n \geq n^2$, $\mathbb{E}(\|\tilde{f}^{(n)}_{\ell_{\rm opt}} - f\|^2) \lesssim n^{-2s/(2s+1)}$.

In Schipper (1996) it is proved that this rate is minimax optimal on Sobolev balls (at least for an integer s), see also Efromovich (2002) for $s < 1/2$ and other references.
Let us compare results of Proposition 4.3 and of Proposition 4.2. As $W^s \subset \mathcal{W}^s$, see Section 4.1, the comparison is relevant. In case (i), we see that the estimator $\tilde{f}_{\ell_{\rm opt}}$ has a better rate than $\hat{f}_{m_{\rm opt}}$. In case (ii), the estimators have the same rate. In case (iii), the estimator $\tilde{f}_{\ell_{\rm opt}}$ is slightly better than $\hat{f}_{m_{\rm opt}}$. In view of case (ii), the Hermite method is competitive. Indeed the moment condition for (ii) is not very strong.

4.3. **Rates of convergence in some special cases.** When the density f belongs to W^s for all s, we must obtain directly the exact rate of decay of the bias term. This is possible for centered Gaussian and some related densities as one can make an exact computation of the coefficients $a_j(f)$. Let

$$f_{p,\sigma}(x) = \frac{x^{2p}}{\sigma^{2p}C_{2p}} f_\sigma(x) \quad \text{with} \quad f_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad \text{and} \quad C_{2p} = \mathbb{E}X^{2p}, \tag{14}$$

for X a standard Gaussian variable. The distribution $f_{p,\sigma}(x)dx$ is equal to $\varepsilon G^{1/2}$ for ε a symmetric Bernoulli variable, G a Gamma$(p+(1/2), 1/(2\sigma^2))$ variable, independent of ε.

Proposition 4.4. *Assume that $f = f_\sigma$. Then for $m_{\rm opt} = [(\log n)/\lambda]$ where $\lambda = \log\left(\frac{\sigma^2+1}{\sigma^2-1}\right)^2$, we have*

$$\mathbb{E}(\|\hat f_{m_{\rm opt}} - f\|^2) \lesssim \sqrt{\log n}/n.$$

The same result folds for $f = f_{p,\sigma}$ or any finite mixture of such distributions. We can compare the rate of Proposition 4.4 with the rate $\log(n)/n$, which is optimal in the class of analytical densities (see Ibragimov and Has'minskii (1980)). So the Hermite based approach outperforms the kernel method in the case of finite normal mixtures.

For $f = f_\sigma$, the estimator $\widetilde{f}_\ell$ satisfies,

$$\mathbb{E}(\|\widetilde{f}_\ell - f\|^2) \quad \lesssim \quad \frac{1}{\ell}\exp\left(-\ell^2/2\sigma^2\right) + n^{-1}\ell.$$

For $\ell_{\rm opt} = \sigma\sqrt{2\log n}$, the rate of $\widetilde{f}_{\ell_{\rm opt}}$ is $\sqrt{\log n}/n$. The rate is identical to the one obtained in Proposition 4.4. The result is analogous for $f = f_{p,\sigma}$.
Finally, the Cauchy density will provide a counter-example. Let

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

From Proposition 2.1, case *(iii)*, we take $a = 2$ and obtain for the variance term $V_m \lesssim m^{2/3}$. Using Proposition 4.1, we check that $f \in W^1$, $f \notin W^2$. Moreover, by (13), $x^s f \notin W^s$ for $s \geq 3/2$. Therefore, $f \notin W^{3/2}$, so the best rate we can obtain is $n^{-s/s+(2/3)}$ with $s < 3/2$, for $m_{\rm opt} = [n^{1/(s+(2/3))}]$.

For the sinus cardinale method, $f^*(t) = \exp\left(-|t|\right)$, so that $\|f - f_\ell\| \lesssim \exp\left(-2\pi\ell\right)$. Therefore, for $\ell_{\rm opt} = \log n/2\pi$, the estimator $\widetilde{f}_{\ell_{\rm opt}}$ has a risk with rate $\log n/n$. This is much better than for the Hermite estimator.

This discussion on rates of convergence points out the interest of the adaptive method. Indeed, it automatically realizes the bias-variance compromise and thus the previous rates are reached without any specific knowledge on f.

4.4. **Complexity.** In this paragraph, we compare the Hermite and deconvolution estimators from another point of view: the computational efficiency.

Consider an estimator $\hat f_n$ of a function f whose $\mathbb{L}^2$-risk can be evaluated on a ball $B(L)$ of some functional space. Define its complexity $\mathcal{C}_{\hat f_n}(\varepsilon)$ as the minimal cost of computing $\hat f_n$ at the observation points $X_1, \ldots, X_n$, given that

$$\sup_{f\in B(L)} \mathbb{E}(\|\hat f_n - f\|^2) \leq \varepsilon^2.$$

Let us compute the complexity of the estimate $\widetilde{f}_{\ell_{\rm opt}}$ on the Sobolev ball $\mathcal{W}^s(L)$. As we need to evaluate the function $\frac{\sin(\pi\ell\cdot)}{\pi\cdot}$ at all points $(X_k - X_j)$, $1 \leq k, j \leq n$, the cost of computing $\widetilde{f}_{\ell_{\rm opt}}$ is of order n^2. Thus $\varepsilon^2 \asymp n^{-2s/(2s+1)}$ yields $n \asymp \varepsilon^{-2-1/s}$ so that $C_{\widetilde{f}_{\ell_{\rm opt}}}(\varepsilon) \asymp \varepsilon^{-4-2/s}$ as $\varepsilon \to 0$. So even in the case of infinitely smooth densities, the complexity of the deconvolution estimate can not be (asymptotically) lower than ε^{-4}. A natural question is whether one can find an estimate with lower order of complexity. Note that the complexity would be the same for a kernel estimator on a ball of a Nikol'ski class with regularity s, see Tsybakov (2009), at least for kernels with a non compact support used in Ibragimov and Has'minskii (1980).

For the truncated estimator $\widetilde{f}^{(n)}_{\ell_{\rm opt}}$, the cost is of order nK_n: indeed, one must compute the $\varphi_{\ell,j}(X_i)$ for $i = 1, \ldots, n$ and $|j| \leq K_n$. Consequently, compared to the previous one, this estimate is competitive in term of computational cost as soon as $K_n < n$ (however this choice would contradict Theorem 3.1 where $K_n \geq n^2$).

Now, let us look at the projection estimator $\hat{f}_{m_{\text{opt}}}$ for $f \in W^s(L)$. The cost of computing a projection estimator $\hat{f}_m$ at observation points $X_1, \dots, X_n$ corresponds to the cost of computing $h_j(X_i)$ for $i = 1, \dots, n$ and $j = 0, \dots, m-1$, i.e. is of order nm. Thus we derive the following proposition.

Proposition 4.5. *Assume that $f \in W^s(L)$ and consider the three cases (i), (ii), (iii) of Proposition 2.1. The the complexity of the estimate $\hat{f}_{m_{\text{opt}}}$ is given by $\mathcal{C}_{\hat{f}}(\varepsilon) \sim \varepsilon^{-2-\frac{2(\alpha+1)}{s}}$ with $\alpha = 5/6, 1/2, (a+2)/[2(a+1)]$, respectively.*

Proof of Proposition 4.5. Taking $\varepsilon^2 \asymp n^{-2s/(2s+1)}$, hence $n \asymp \varepsilon^{-2-1/s}$, and the three values of m_{opt} given Proposition 4.2 yield the result. □

As can be seen, the complexity order of the Hermite-based estimate $\hat{f}_{m_{\text{opt}}}$ is lower than the complexity order of the deconvolution estimate $\widetilde{f}_{\ell_{\text{opt}}}$ provided $s > \alpha$. So in the case of densities with finite fifth moment already for $s > 1/2$, our approach leads to estimates with much lower complexity. The difference between the estimates $\hat{f}_{m_{\text{opt}}}$ and $\widetilde{f}_{\ell_{\text{opt}}}$ becomes especially pronounced in the limiting case $s \to \infty$, where $\mathcal{C}_{\hat{f}_{m_{\text{opt}}}}(\varepsilon) \asymp \varepsilon^{-2}$ while $\mathcal{C}_{\widetilde{f}_{\ell_{\text{opt}}}}(\varepsilon) \asymp \varepsilon^{-4}$ as $\varepsilon \to 0$, resulting in a huge computational gain.

Projection estimator on a compact set A Besov ball of $B_{2,s,\infty}(A)$	Projection estimator on $\mathbb{R}$ Sobolev-Hermite ball $W^s(L)$	Projection estimator on $\mathbb{R}^+$ Sobolev-Laguerre ball of index s	Deconvolution estimator on $\mathbb{R}$ Sobolev ball $\mathcal{W}^s(L)$
$\varepsilon^{-2-2/s}$	$\varepsilon^{-2-3/s}$ (best case $\alpha = 1/2$)	$\varepsilon^{-2-4/s}$	$\varepsilon^{-4-2/s}$

Table 1. Complexity for density estimation in different contexts.

For any projection estimator, the cost of computation if of order nm_{opt} where m_{opt} is the optimal dimension. In the case of a density with compact support A, if we evaluate the $\mathbb{L}^2$-risk of a projection estimator on a Besov ball of $B_{2,s,\infty}(A)$, we have $\varepsilon^2 \asymp n^{-2s/(2s+1)}$ with $m_{\text{opt}} \asymp n^{1/(2s+1)}$, thus a cost of order $\varepsilon^{-2-2/s}$, see Barron *et al.* (1999) for rates and definition of Besov spaces.

In the case of a density of $\mathbb{R}^+$, the best $\mathbb{L}^2$-risk of a projection estimator on a Sobolev-Laguerre ball of index s is of order $n^{-s/(s+1)}$ with $m_{\text{opt}} \asymp n^{1/(s+1)}$, hence a cost of order $\varepsilon^{-2-4/s}$, see Belomestny *et al.* (2016).

All these results are summarized in Table 1.

5. Simulation results

In this Section, we propose a few illustrations of the previous theoretical findings. To that aim, we consider several densities, fitting different assumptions of our setting.

(i) A Gaussian $\mathcal{N}(0,1)$,

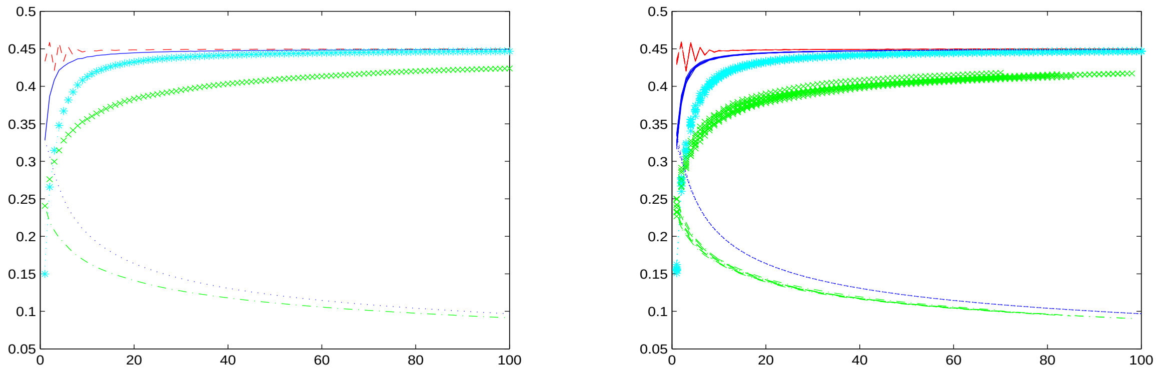

FIGURE 1. Left: $m \mapsto \widehat{V}_m/\sqrt{m}$ for 1 sample of densities (i) (blue line), (iv) (cyan stars), (vi) (red dashed), (ix) (green x marks) and $m \mapsto \widehat{V}_m/m^{5/6}$ for 1 sample of densities (i) (blue dashed) and (ix) (green dash-dot). Right: the same as previously for 10 samples.

(ii) A Gaussian $\mathcal{N}(0, \sigma^2)$, $\sigma = 0.5$,
(iii) A mixed Gaussian density $0.4\mathcal{N}(-3, \sigma^2) + 0.6\mathcal{N}(3, \sigma^2)$, $\sigma = 0.5$,
(iv) A Gamma $\gamma(3, 0.5)$ density,
(v) A mixed Gamma $0.4\gamma(2, 1/2) + 0.6\gamma(16, 1/4)$
(vi) A beta density $\beta(3, 3)$,
(vii) A beta density $\beta(3, 6)$,
(viii) Laplace density $f(x) = e^{-|x|}/2$,
(ix) A Cauchy density, $f(x) = 5/[\pi(1 + (5x)^2)]$.

Density (i) is proportional to the first basis function h_0 and should be perfectly estimated in the Hermite procedure, densities (vi) and (vii) are compactly supported and density (ix) does not admit any moment (in particular no fifth moment, so it does not fit case (ii) of Proposition 2.1). Hermite functions are recursively computed via (26) and with normalization (1).

We plot in Figure 1 the representation of $m \mapsto \widehat{V}_m/\sqrt{m}$ for 1 and 10 samples drawn from densities (i), (iv), (vi), (ix) (see (5)). It seems that the ratio is stable along the repetitions, and converges to a fixed value, which is the same in the first three cases. On the contrary, $m \mapsto \widehat{V}_m/m^{5/6}$ given for (i) and (ix) seems to decrease and to tend to zero in any case. It is tempting to conclude from these plots that the order of V_m is $O(m^{1/2})$ in a rather general case.

We have implemented the Hermite projection estimator $\hat{f}_{\hat{m}}$ with $\hat{m}$ given in (5), $\widetilde{f}^{(n)}_{\tilde{\ell}_n}$ with $\tilde{\ell}_n$ given by (9) and the kernel estimator given by the function `ksdensity` of Matlab. For the model selection steps of the first two estimators, the two constants κ and $\tilde{\kappa}$ of the procedures have been both calibrated by preliminary simulations including other densities than the ones mentioned above (to avoid overfitting): the selected values were $\kappa = \tilde{\kappa} = 4$. We considered two sample sizes $n = 250$ and $n = 1000$, but as the sine cardinale procedure is rather slow, we only took $K_{250} = K_{1000} = 100$. The theoretical value $K_n = n^2$ is unreachable in practice (the computing time becomes much too large), and our choice of K_n is consistent with the complexity considerations of Section 4.4.

For m_n, we should take $(n/\log(n))^{6/5}$, which is of order 100 for $n = 250$ and 400 for $n = 1000$. We took $m_{250} = m_{1000} = 200$ as a compromise. The cutoff $\ell\pi$ is selected among 100 equispaced values between 0 and 10. For each distribution, we present in Table 2 the MISE computed over 200 repetitions, together with the standard deviation. In the three cases, we provide also the mean (and standard deviations in parenthesis) of the selected dimension (Hermite), cutoff (Sine cardinale) or bandwidth (kernel).

	$n = 250$			$n = 1000$		
density	Hermite	Sin. Card.	Kernel	Hermite	Sin. Card.	Kernel
(i)	0.5 (1.4)	2.0 (1.8)	2.9 (2.1)	0.1 (0.4)	0.6 (0.5)	1.1 (0.6)
	1.08 (0.32)	0.77 (0.08)	0.35 (0.03)	1.08 (0.53)	0.87 (0.08)	0.27 (0.01)
(ii)	4.7 (4.7)	4.7 (4.9)	6.0 (4.2)	1.4 (1.3)	1.4 (1.4)	2.1 (1.3)
	8.87 (3.69)	1.46 (0.21)	0.17 (0.01)	11.6 (5.1)	1.65 (0.24)	0.13 (0.005)
(iii)	4.9 (2.6)	5.5 (2.5)	24.2 (14.5)	1.5 (0.9)	1.5 (0.9)	11.1 (3.1)
	11.6 (1.5)	1.28 (0.13)	0.51 (0.12)	14.3 (2.0)	1.53 (0.10)	0.39 (0.04)
(iv)	5.6 (3.4)	5.3 (3.4)	4.6 (2.9)	1.8 (1.0)	1.8 (1.0)	1.9 (1.0)
	6.28 (2.64)	1.25 (0.19)	0.27 (0.02)	11.9 (4.5)	1.7 (0.25)	0.21 (0.01)
(v)	7.2 (3.6)	6.6 (2.8)	17.1 (2.6)	2.4 (0.9)	2.7 (0.8)	10.2 (1.2)
	15.1 (2.0)	1.13 (0.20)	0.74 (0.07)	18.2 (2.6)	1.67 (0.27)	0.57 (0.02)
(vi)	7.2 (7.2)	7.3 (7.4)	12.8 (8.3)	3.2 (2.6)	3.3 (2.7)	4.8 (2.7)
	46.5 (10.5)	3.14 (0.29)	0.07 (0.005)	63.3 (27.3)	3.6 (0.65)	0.05 (0.002)
(vii)	17.2 (11.3)	19.3 (15.6)	19.4 (12.2)	5.8 (3.1)	6.3 (4.1)	7.0 (4.0)
	104.2 (25.5)	4.77 (0.91)	0.05 (0.004)	143.3 (13.5)	5.89 (0.85)	0.04 (0.002)
(viii)	7.4 (2.0)	6.7 (3.0)	5.5 (3.2)	2.6 (0.8)	2.5 (0.8)	2.3 (1.1)
	2.5 (2.96)	1.03 (0.25)	0.36 (0.04)	7.8 (4.4)	1.42 (0.32)	0.27 (0.01)
(ix)	21.6 (9.1)	21.5 (9.3)	18.6 (10.6)	6.9 (2.9)	6.9 (3.0)	7.6 (3.9)
	65 (25)	3.7 (0.7)	0.10 (0.01)	97 (21)	4.71 (0.77)	0.08 (0.004)

TABLE 2. Results after 200 iterations of simulations of density (i) to (ix). For each density (i)-(ix), first line: MISE$\times$ 1000 with (std $\times$ 1000) in parenthesis; second line: mean of selected dimension (Hermite), cutoff (sinus cardinale) or bandwidth (kernel) with std in parenthesis.

We can see from the results of Table 2 that the Hermite and sinus cardinale methods give very similar results, except for the $\mathcal{N}(0,1)$ where the Hermite projection is much better as expected, as the procedure most of the time chooses $m = 1$. The kernel method seems globally less satisfactory. The noteworthy difference between the first two methods is the computation time: as the models are nested in the Hermite projection strategy, all coefficients can be computed once for all, and then the dimension is selected. In the sinus cardinale strategy, each time ℓ is changed, all the coefficients have to be recalculated. For instance, when the maximal dimension proposed m_n is 50, and K_n is 100, the elapsed times for 100 simulations is: for $n = 250$, around 0.5s for Hermite, 41s for sinus cardinale; for $n = 1000$, around 1.2s for Hermite, 137s for sinus cardinale, all times measured on the same personal computer to give an order of the difference. This is coherent with the lower complexity property of the Hermite method.

Table 2 also provides the selected dimensions, cutoffs and bandwidths. As could be expected, $\hat{m}$, $\tilde{\ell}$ vary in opposite way, compared to $\hat{h}$. Without surprise also, the selected dimensions and cutoffs increase when the sample size increases. What is remarkable is the values of the selected dimensions for β-distributions, which are very large. Globally, we can see that these values are very different from one distribution to the other. Contrary to the theoretical result, the Cauchy density is estimated with similar MISEs in the Hermite and sinus cardinale methods.

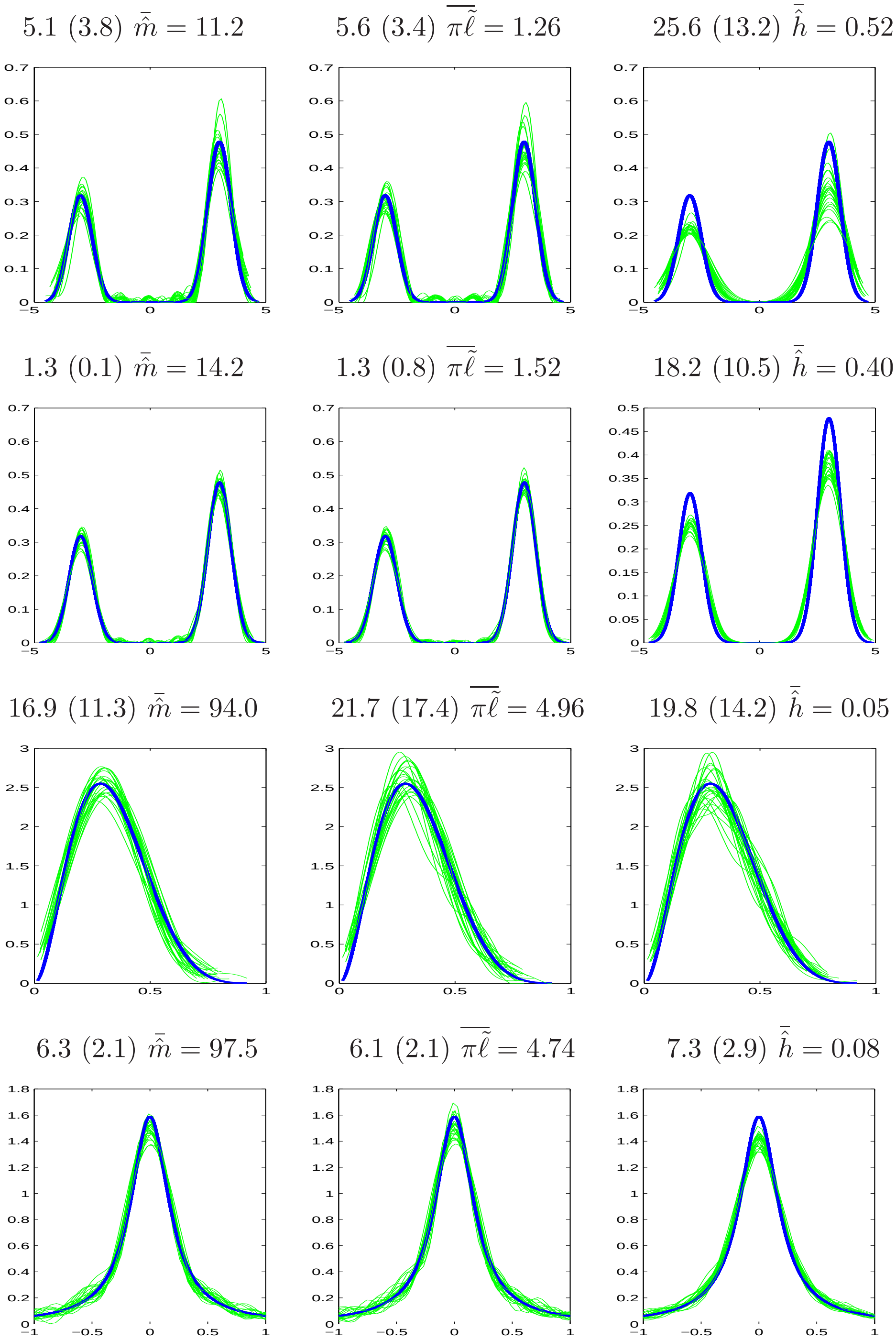

FIGURE 2. True density f in bold blue for Model (iii) (first two lines), Model (vii) (third line) and Model (ix) (fourth line), together with 25 estimates (green/grey) with $n = 250$ (lines 1 and 3) or $n = 1000$ (lines 2 and 4). First column: Hermite; second column: Sinus Cardinale; third column: kernel. Above each plot: MISE$\times$ 1000 and std $\times$ 1000 in parenthesis, followed par the mean of selected dimensions, cutoffs and bandwidths (all means over the 25 samples).

In Figure (2), density and 25 estimators are plotted for models (iii), (vii) and (ix). Risks and standard deviation for the 25 curves are given above each plot, together with the mean of the selected dimension, cutoff or bandwidth. The methods are comparable, even for the Cauchy distribution, except for the mixtures, where the kernel method fails. The first two lines illustrate the improvement obtained when increasing n. We note again that the selected dimensions in the Hermite method are possibly rather high (see the beta and the Cauchy densities). However, computation time remains very short.

6. Proofs

6.1. **Proof of Propositions 2.1 and 2.2.** We start by proving Proposition 2.1.

(i). The following bound comes from Szegö (1959, p.242) where an expression of C_∞ is given:

$$\forall x \in \mathbb{R}, \quad |h_j(x)| \le C_\infty (j+1)^{-(1/12)}, \quad j = 0, 1, \dots. \tag{15}$$

Therefore,

$$V_m \le C_\infty^2 \sum_{j=0}^{m-1} (j+1)^{-(1/6)} \le \frac{6}{5} C_\infty^2 m^{5/6}. \tag{16}$$

(ii). Now, as in Walter (1977), we use the following expression for the Hermite function h_n (see Szegö (1959, p.248)):

$$h_j(x) = \lambda_j \cos\left((2j+1)^{1/2}x - \frac{j\pi}{2}\right) + \frac{1}{(2j+1)^{1/2}}\xi_j(x) \tag{17}$$

where $\lambda_j = |h_j(0)|$ if j is even, $\lambda_j = |h_j'(0)|/(2j+1)^{1/2}$ if j is odd and

$$\xi_j(x) = \int_0^x \sin\left[(2j+1)^{1/2}(x-t)\right] t^2 h_j(t) dt.$$

By the Cauchy-Schwarz inequality,

$$\xi_j^2(x) \le \int_0^{|x|} t^4 dt \int_0^{|x|} h_j^2(t) dt \le \frac{|x|^5}{5} \times \frac{1}{2}.$$

Moreover,

$$\lambda_{2j} = \frac{(2j)!^{1/2}}{2^j j! \pi^{1/4}}, \quad \lambda_{2j+1} = \lambda_{2j} \frac{\sqrt{2j+1}}{\sqrt{2j+3/2}}.$$

By the Stirling formula and its proof, $\lambda_{2j} \sim \pi^{-1/2} j^{-1/4}$, $\lambda_{2j+1} \sim \pi^{-1/2} j^{-1/4}$ and for all j, there exists constants c_1, c_2 such that, for all $j \ge 1$,

$$\frac{c_1}{\pi^{1/2} j^{1/4}} \le \lambda_j \le \frac{c_2}{\pi^{1/2} j^{1/4}}.$$

Therefore,

$$h_j^2(x) \le 2\frac{c_2^2}{\pi j^{1/2}} + \frac{1}{2j+1}\frac{|x|^5}{5}.$$

This yields:

$$\int h_j^2(x) f(x) dx \le 2\frac{c_2^2}{\pi j^{1/2}} + \frac{1}{5(2j+1)} \mathbb{E}|X|^5,$$

which implies $V_m \lesssim m^{1/2}$.

Now we study case (iii). The following bound for h_j is given in Markett (1984, p.190): There exist positive constants C, γ, independent of x and j, such that, for $J = 2j + 1$,

$$\begin{aligned} |h_j(x)| &\leq C(J^{1/3} + |x^2 - J|)^{-1/4}, \quad x^2 \leq 2J, \\ &\leq C\exp{(-\gamma x^2)}, \quad x^2 > 2J. \end{aligned}$$

Consider a sequence (a_j) such that $a_j \to +\infty$, $a_j/\sqrt{j} \to 0$ with $J = 2j+1$ large enough to ensure $\frac{a_J}{\sqrt{J}} \leq 1/\sqrt{2}, a_J \geq K$. As $\int h_j^2(x)dx = 1$, $a_J < \sqrt{J}$, $a_J \geq K$ and g is decreasing,

$$\int h_j^2(x)f(x)dx \leq 2C\|f\|_\infty \int_0^{a_J} (J^{1/3} + J - x^2)^{-1/2}dx + g(a_J).$$

Set $x = (J^{1/3} + J)^{1/2}y$ in the integral. This yields:

$$\int_0^{a_J} (J^{1/3} + J - x^2)^{-1/2}dx = \int_0^{a_J/(J^{1/3}+J)^{1/2}} \frac{dy}{\sqrt{1-y^2}} = \text{Arcsin}(\frac{a_J}{(J^{1/3}+J)^{1/2}}) \leq 2\frac{a_J}{\sqrt{J}}.$$

as for $0 \leq x \leq 1/\sqrt{2}$, Arcsin$x \leq 2x$. Now we choose the sequence (a_j) and consider $a_j = j^{1/(2(a+1))}$. We deduce

$$\int h_j^2(x)f(x)dx \lesssim j^{-a/(2(a+1))},$$

which leads to

(18) $$V_m \lesssim m^{\frac{a+2}{2(a+1)}}.\square$$

Now we turn to the proof of Proposition 2.2 and we look at the lower bound. We have, setting $c = \inf_{a \leq x \leq b} f(x)$, and using (17),

$$\begin{aligned} \int h_j^2(x)f(x)dx &\geq c\int_a^b h_j^2(x)dx \\ &\geq c\lambda_j^2 \int_a^b \cos^2\left((2j+1)^{1/2}x - \frac{j\pi}{2}\right)dx \\ &+ c\frac{2\lambda_j}{(2j+1)^{1/2}} \int_a^b \cos\left((2j+1)^{1/2}x - \frac{j\pi}{2}\right)\xi_j(x)dx. \end{aligned}$$

We have $j^{-3/4}c_1/\sqrt{\pi} \leq \frac{2\lambda_j}{(2j+1)^{1/2}} \leq j^{-3/4}\sqrt{2/\pi}c_2$ and

$$|\int_a^b \cos\left((2j+1)^{1/2}x - \frac{j\pi}{2}\right)\xi_j(x)dx| \leq \int_a^b \frac{|x|^{5/2}}{\sqrt{10}}dx := C.$$

Thus, the second term is lower bounded by $-Cj^{-3/4}c_1/\sqrt{\pi}$. For the first term, $\lambda_j^2 \geq j^{-1/2}c_1^2/\pi$ and

$$\int_a^b \cos^2\left((2j+1)^{1/2}x - \frac{j\pi}{2}\right)dx = \frac{1}{2}(b-a) + \int_a^b \cos\left(2(2j+1)^{1/2}x - j\pi\right)dx = \frac{1}{2}(b-a) + O(\frac{1}{j^{1/2}}).$$

Therefore,

$$\int h_j^2(x)f(x)dx \geq cj^{-1/2}c_1^2/\pi\left[\frac{b-a}{2} + O(\frac{1}{j^{1/2}})\right] - Cj^{-3/4}c_1/\sqrt{\pi}.$$

Consequently, for j large enough, $\int h_j^2(x)f(x)dx \geq c'j^{-1/2}$. This implies, $V_m \geq c''m^{1/2}$. $\square$

6.2. **Proof of Theorem 2.1.** Let S_m be the space spanned by $\{h_0, \dots, h_{m-1}\}$ and $B_m = \{t \in S_m, \|t\| = 1\}$. We have $\hat{f}_m = \arg\min_{t \in S_m} \gamma_n(t)$ where $\gamma_n(t) = \|t\|^2 - 2n^{-1}\sum_{i=1}^n t(X_i)$ and $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2$. Now, we write, for two functions $t, s \in \mathbb{L}^2(\mathbb{R})$,

$$\gamma_n(t) - \gamma_n(s) = \|t - f\|^2 - \|s - f\|^2 - 2\nu_n(t - s)$$

where

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^n [t(X_i) - \langle t, f\rangle].$$

Then, for any $m \in \mathcal{M}_n = \{1 \le m \le m_n\}$, $m_n \le n/\log n$, and any $f_m \in S_m$,

$$\gamma_n(\hat{f}_{\hat{m}}) + \widehat{\mathrm{pen}}(\hat{m}) \le \gamma_n(f_m) + \widehat{\mathrm{pen}}(m).$$

This yields

$$\|\hat{f}_{\hat{m}} - f\|^2 \le \|f - f_m\|^2 + \widehat{\mathrm{pen}}(m) - \widehat{\mathrm{pen}}(\hat{m}) + 2\nu_n(\hat{f}_{\hat{m}} - f_m).$$

We use that

$$2\nu_n(\hat{f}_{\hat{m}} - f_m) \le 4 \sup_{t \in B_{m\vee\hat{m}}} \nu_n^2(t) + \frac{1}{4}\|\hat{f}_{\hat{m}} - f_m\|^2,$$

and some classical algebra to obtain:

$$\begin{aligned} \frac{1}{2}\|\hat{f}_{\hat{m}} - f\|^2 \quad &\le \quad \frac{3}{2}\|f - f_m\|^2 + \widehat{\mathrm{pen}}(m) + 4\left(\sup_{t \in B_{m\vee\hat{m}}} \nu_n^2(t) - p(m \vee \hat{m})\right) \\ &\quad + (4p(m \vee \hat{m}) - \mathrm{pen}(\hat{m})) + (\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m})). \end{aligned} \tag{19}$$

We can choose $p(m)$ such that

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\left(\sup_{t \in B_{m\vee m'}} \nu_n^2(t) - p(m \vee m')\right)_+ \le \frac{c}{n}. \tag{20}$$

Indeed, for this, we apply the Talagrand Inequality:

$$\mathbb{E}\left(\sup_{t \in B_m} \nu_n^2(t) - 4H^2\right)_+ \le \frac{C_1}{n}\left(v^2 e^{-C_2\frac{nH^2}{v}} + \frac{M_1}{n} e^{-C_3\frac{nH}{M_1}}\right)$$

where $\mathbb{E}\left(\sup_{t\in B_m} \nu_n^2(t)\right) \le \frac{V_m}{n} := H^2$, $\sup_{t\in B_m} \mathrm{Var}(t(X_1)) \le \sup_{t\in B_m} \mathbb{E}(t^2(X_1)) \le \|f\|_\infty := v^2$ and $\sup_{t\in B_m} \sup_x |t(x)| \le \sqrt{\sup_x \sum_{j=0}^{m-1} h_j^2(x)} \le C'_\infty m^{5/12} \le C'_\infty\sqrt{n} := M_1$ (see (15)). Therefore we obtain

$$\mathbb{E}\left(\sup_{t \in B_m} \nu_n^2(t) - 4\frac{V_m}{n}\right)_+ \le \frac{C_1}{n}\left(\|f\|_\infty e^{-C'_2 V_m} + \frac{1}{\sqrt{n}} e^{-C'_3\sqrt{V_m}}\right).$$

Therefore, with the choice $p(m) = 4V_m/n$, (20) holds under condition (6) which is ensured by Proposition 2.2.

Taking expectation in (19) yields

$$\begin{aligned} \frac{1}{2}\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2) \quad &\le \quad \frac{3}{2}\|f - f_m\|^2 + \mathrm{pen}(m) + \mathbb{E}(4p(m \vee \hat{m}) - \mathrm{pen}(\hat{m})) \\ &\quad + \mathbb{E}(\mathrm{pen}(\hat{m}) - \widehat{\mathrm{pen}}(\hat{m}))_+ + \frac{c}{n}. \end{aligned} \tag{21}$$

Let us define

$$Y_i^{(m)} := \sum_{j=0}^{m-1} h_j^2(X_i), \quad \widehat{V}_m = \frac{1}{n}\sum_{i=1}^n Y_i^{(m)},$$

and the set inspired by Bernstein Inequality

$$\Omega=\left\{\forall m\in\mathcal{M}_n,\quad \frac{1}{n}\left|\sum_{i=1}^n(Y_i^{(m)}-\mathbb{E}(Y_i^{(m)}))\right|\leq\sqrt{2V_mC''_\infty m^{5/6}\frac{\log(n)}{n}}+4C''_\infty m^{5/6}\frac{\log(n)}{3n}\right\}.$$

with $C''_\infty:=(C'_\infty)^2$ and C'_∞ is the constant appearing in M_1 above. We split the term to study in (21) as follows:

$$\mathbb{E}(\mathrm{pen}(\hat m)-\widehat{\mathrm{pen}}(m))_+\leq\mathbb{E}\left[(\mathrm{pen}(\hat m)-\widehat{\mathrm{pen}}(m))_+\mathbf{1}_\Omega\right]+\mathbb{E}\left[(\mathrm{pen}(\hat m)-\widehat{\mathrm{pen}}(m))_+\mathbf{1}_{\Omega^c}\right].$$

On Ω,

$$\begin{aligned}|\widehat V_{\hat m}-V_{\hat m}| &\leq \sqrt{2V_{\hat m}C''_\infty\hat m^{5/6}\log(n)/n}+4C''_\infty\hat m^{5/6}\log(n)/(3n)\\ &\leq \frac{1}{2}V_{\hat m}+\frac{7}{3}C''_\infty\frac{\hat m^{5/6}\log(n)}{n},\end{aligned}$$

using that $2xy\leq x^2+y^2$ applied to $\sqrt{2VA}=2\sqrt{V/2}\sqrt{A}\leq V/2+A$ with $V=V_{\hat m}$ and $A=C"_\infty\hat m^{5/6}\log(n)/n$.
and thus, by definition of $\mathcal{M}_n$,

$$\mathbb{E}\left[(\mathrm{pen}(\hat m)-\widehat{\mathrm{pen}}(\hat m))_+\mathbf{1}_\Omega\right]+\leq\frac{1}{2}\mathbb{E}(\mathrm{pen}(\hat m))+\frac{c}{n}.$$

On the other hand,

$$\mathbb{E}\left[(\mathrm{pen}(\hat m)-\widehat{\mathrm{pen}}(\hat m))_+\mathbf{1}_{\Omega^c}\right]\leq 2\kappa\mathbb{P}(\Omega^c).$$

Now,

$$\mathbb{P}(\Omega^c)\leq\sum_{m\in\mathcal{M}_n}2e^{-2\log(n)}\leq\frac{c}{n}$$

as we apply Bersntein inequality: $\mathbb{P}(|S_n/n|\geq\sqrt{2v^2x/n}+bx/(3n))\leq 2e^{-x}$ for $S_n=\sum_{i=1}^n(U_i-\mathbb{E}(U_i))$, $\mathrm{Var}(U_1)\leq v^2$, $|U_i|\leq b$. In our case $U_=Y_i^{(m)}$ and $v^2=V_mC''_\infty m^{5/6}$, $b=C''_\infty m^{5/6}$ and we took $x=2\log(n)$.

So Equation (19) becomes

$$\begin{aligned}\frac{1}{2}\mathbb{E}(\|\hat f_{\hat m}-f\|^2) &\leq \frac{3}{2}\|f-f_m\|^2+\mathrm{pen}(m)+\mathbb{E}(4p(m\vee\hat m)-\mathrm{pen}(\hat m)) &(22)\\ &\quad+\frac{1}{2}\mathbb{E}(\mathrm{pen}(\hat m))+\frac{c}{n} &(23)\\ &\leq \frac{3}{2}\|f-f_m\|^2+\mathrm{pen}(m)+\mathbb{E}(4p(m\vee\hat m)-\frac{1}{2}\mathrm{pen}(\hat m))+c\frac{c}{n} &(24)\end{aligned}$$

Now we note that, for $\kappa\geq 8:=\kappa_0$,

$$4p(m\vee\hat m)-\frac{1}{2}\mathrm{pen}(\hat m)\leq\mathrm{pen}(m).$$

Finally, we get, for all $m\in\mathcal{M}_n$,

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

which ends the proof. □

6.3. **Proof of Proposition 3.1.** The first inequality is standard. Let us study $\widetilde{f}_\ell^{(n)}(x)$. We write that

$$\begin{aligned}\|\widetilde{f}_\ell^{(n)} - f\|^2 &= \|\widetilde{f}_\ell^{(n)} - \mathbb{E}\widetilde{f}_\ell^{(n)}\|^2 + \|\mathbb{E}(\widetilde{f}_\ell^{(n)}) - f\|^2 \\ &\leq \|\widetilde{f}_\ell^{(n)} - \mathbb{E}\widetilde{f}_\ell^{(n)}\|^2 + 2\|\mathbb{E}(\widetilde{f}_\ell^{(n)}) - \bar{f}_\ell\|^2 + 2\|\bar{f}_\ell - f\|^2.\end{aligned}$$

The term $\|\bar{f}_\ell - f\|^2$ is the usual bias term. moreover

$$\begin{aligned}\mathbb{E}\left(\|\widetilde{f}_\ell^{(n)} - \mathbb{E}\widetilde{f}_\ell^{(n)}\|^2\right) &= \sum_{|j|\leq K_n} \mathrm{Var}(\tilde{a}_{\ell,j}) = \frac{1}{n}\sum_{|j|\leq K_n} \mathrm{Var}(\varphi_{\ell,j}(X_1)) \\ &\leq \frac{1}{n}\sum_{|j|\leq K_n} \mathbb{E}[\varphi_{\ell,j}^2(X_1)] \leq \frac{\ell}{n}\end{aligned}$$

because $\sum_{j\in\mathbb{Z}} |\varphi_{\ell,j}(x)|^2 \leq \ell$. This is the standard variance term order.
The new term is

$$\|\mathbb{E}(\widetilde{f}_\ell^{(n)}) - \bar{f}_\ell\|^2 = \sum_{|j|\geq K_n} |a_{\ell,j}|^2 \leq 2\sup_j |ja_{\ell,j}|^2 \sum_{j>K_n} j^{-2} \leq \frac{2}{K_n}\sup_j |ja_{\ell,j}|^2. \tag{25}$$

We write that $ja_{\ell,j} = j\sqrt{\ell}\int \varphi(\ell x - j)f(x)dx = \sqrt{\ell}(I_1 + I_2)$ where

$$I_1 = \ell\int x\varphi(\ell x - j)f(x)dx, \quad I_2 = -\int(\ell x - j)\varphi(\ell x - j)f(x)dx$$

and we bound I_1 and I_2.

$$|I_1| \leq \ell\sqrt{\int |\varphi(\ell x - j)|^2 dx \int x^2 f^2(x)dx} = \sqrt{\ell}\sqrt{M_2}, \qquad \text{where } M_2 = \int x^2 f^2(x)dx.$$

On the other hand, $|I_2| \leq \sup_{u\in\mathbb{R}} |u\varphi(u)| \int f(x)dx \leq 1$. We obtain:

$$|ja_{\ell,j}| \leq \ell\sqrt{M_2} + \sqrt{\ell} \leq \ell(\sqrt{M_2} + 1).$$

Plugging this in (25), we find the bound: $\|\mathbb{E}(\widetilde{f}_\ell^{(n)}) - \bar{f}_\ell\|^2 \leq 4\ell^2(M_2+1)/K_n$. This term is $O(\ell/n)$ if $\ell \leq n$ and $K_n \geq n^2$. □

6.4. **Proof of Proposition 4.1.** Using the relations (see *e.g.* Abramowitz and Stegun (1964))

$$2xH_n(x) = H_{n+1}(x) + 2nH_{n-1}(x), \quad H_n'(x) = 2nH_{n-1}(x), \quad n \geq 1. \tag{26}$$

we get:

$$A_+h_n = \sqrt{2n}h_{n-1}, \quad A_-h_n = \sqrt{2(n+1)}h_{n+1}.$$

We deduce:

$$\sqrt{2}h_n' = \sqrt{n}\, h_{n-1} - \sqrt{n+1}h_{n+1}, \quad 2x\, h_n = \sqrt{2(n+1)}\, h_{n+1} + \sqrt{2n}\, h_{n-1}, \tag{27}$$

Assume first that $f \in \mathbb{L}^2(\mathbb{R})$, f admits derivatives up to order s, and for $j_1, \dots, j_m \in \{-,+\}$ and $1 \leq m \leq s$, $A_{j_1}\dots A_{j_m}f \in \mathbb{L}^2(\mathbb{R})$. We prove that $\sum_{n\geq 0} n^s a_n^2(f) < +\infty$. We do the proof only for f compactly supported and refer to Bongioanni and Torrea (2006) otherwise.

For the proof, set $A_{-1} = A_-$, $A_{+1} = A_+$ so that, for $n - j \geq 0$, $A_jh_n = \sqrt{2(n+d_j)}h_{n-j}$, $d_j = 0$ if $j = 1$, $d_j = 1$ if $j = -1$. Thus, for $n - j_1 - j_2 - \dots - j_m \geq 0$,

$$A_{j_1}\dots A_{j_m}h_n = \sqrt{2(n+d_{j_1})} \times \dots \times \sqrt{2(n+d_{j_m})}h_{n-j_1-j_2-\dots-j_m}.$$

Now, for f compactly supported,

$$\langle A_j f, h_n\rangle = \langle f, A_{-j} h_n\rangle = \sqrt{2(n+d_{-j})}\langle f, h_{n+j}\rangle.$$

Iterating yields, for $n + j_1 + j_2 + \ldots + j_m \geq 0$,

$$\langle A_{j_1}\ldots A_{j_m} f, h_n\rangle = \langle f, A_{-j_m}A_{-j_{m-1}}\ldots A_{-j_1}h_n\rangle = \prod_{1\leq k\leq m}\sqrt{2(n+d_{-j_k})}\langle f, h_{n+j_1+j_2+\ldots+j_m}\rangle.$$

Therefore, $\sum_{n\geq 0}(\langle A_{j_1}\ldots A_{j_m} f, h_n\rangle)^2 < +\infty$ is equivalent to

$$\sum_{n+j_1+j_2+\ldots+j_m\geq 0} n^m a^2_{n+j_1+j_2+\ldots+j_m}(f) < \infty.$$

Now assume that $\sum_{n\geq 0} n a_n^2(f) < +\infty$.

We have $f = \sum_{n\geq 0} a_n(f) h_n$. We can write for n_1 large enough:

$$|\sum_{n=n_1}^{n_1+n_2} a_n(f)h_n(x)| \leq \left(\sum_{n=n_1}^{n_1+n_2} n^{1+1/6}a_n^2(f)h_n^2(x)\sum_{n=n_1}^{n_1+n_2} n^{-(1+1/6)}\right)^{1/2} \leq C\sum_{n=n_1}^{n_1+n_2} n a_n^2(f).$$

Thus, the series for f converges uniformly, f is continuous and satisfies for all x, $f(x) = \sum_{n\geq 0} a_n(f)h_n(x)$. Therefore, we have:

$$\begin{aligned} f(y)-f(x) &= \sum_{n\geq 0} a_n(f)\int_x^y h_n'(t)dt \\ &= a_0(f)(h_0((x)-h_0(y)) + 2^{-1/2}\sum_{n\geq 1} a_n(f)\int_x^y(\sqrt{n}\, h_{n-1}(t) - \sqrt{n+1}h_{n+1}(t))dt \end{aligned}$$

Set $S_N(t) = \sum_{n=1}^N a_n(f)(\sqrt{n}\, h_{n-1}(t) - \sqrt{n+1}h_{n+1}(t))$ and $S(t) = \sum_{n\geq 1} a_n(f)(\sqrt{n}\, h_{n-1}(t) - \sqrt{n+1}h_{n+1}(t))$. The function $S(t)$ is well defined by assumption and S_N converges to S in $\mathbb{L}^2(\mathbb{R})$. Therefore, as N tends to infinity.

$$\int_x^y |S_N(t) - S(t)|dt \leq \sqrt{y-x}\|S_N - S\| \to 0.$$

We have proved that

$$f(y) - f(x) = a_0(f)\int_x^y h_0'(t)dt + \int_x^y S(t)dt.$$

Thus, f is absolutely continuous and $f' = S$ belongs to $\mathbb{L}^2(\mathbb{R})$. Analogously, we prove that xf belongs to $\mathbb{L}^2(\mathbb{R})$. Thus, A_+f, A_-f belong to $\mathbb{L}^2(\mathbb{R})$.

Next, by the same reasoning as above, using that $\sum_n n^2 a_n(f) < +\infty$ the series for $f'(t) = S(t)$ is uniformly convergent and $f'(t)$ is continuous. We proceed analogously to prove that f' is absolutely continuous and that xf' and f'' belong to $\mathbb{L}^2(\mathbb{R})$. Iterating the reasoning, we obtain that f admits continuous derivatives up to $s-1$ and that $f^{(s-1)}$ is absolutely continuous and that $f, f', \ldots, f^{(s)}, x^{k-m}f^{(k-m)}, m = 0, \ldots, s-1$ all belong to $\mathbb{L}^2(\mathbb{R})$. This shows that, for $j_1, \ldots, j_m \in \{-,+\}, 1 \leq m \leq s$, $A_{j_1}\ldots A_{j_m} f$ belongs to $\mathbb{L}^2(\mathbb{R})$. □

6.5. **Proof of Proposition 4.4.** To prove the result, we use the following proposition.

Proposition 6.1. *Recall that $a_j(f) = \int f(x)h_j(x)dx$. For $j \geq 0$, we have:*

$$a_{2j}(f_\sigma) \quad = \quad c_{2j}\left(\frac{1}{1+\sigma^2}\right)^{1/2}\frac{(2j)!}{j!}\left(\frac{\sigma^2-1}{\sigma^2+1}\right)^j, \quad a_{2j+1}(f_\sigma) = 0.$$

For $n \geq p$, $j \geq 0$,

$$|a_{2j}(f_{p,\sigma})| \leq C(p,\sigma^2)c_{2j}\frac{(2j)!}{(j-p)!}\left|\frac{\sigma^2-1}{\sigma^2+1}\right|^{j-p}, \quad a_{2j+1}(f_{p,\sigma}) = 0.$$

We can now deduce the risk of $\hat{f}_m$ when $f = f_\sigma$. We have:

$$a_{2j}^2(f_\sigma) \sim \pi^{-1}j^{-1/2}\frac{1}{1+\sigma^2}\left(\frac{\sigma^2-1}{\sigma^2+1}\right)^{2j}.$$

Therefore, setting $\lambda = \log\left[\left(\frac{\sigma^2+1}{\sigma^2-1}\right)^2\right]$ yields $\|f - f_m\|^2 \lesssim \frac{1}{\sqrt{m}}\exp(-\lambda m)$. Combining with Proposition 2.1, we obtain $\mathbb{E}(\|\hat{f}_m - f\|^2) \lesssim \frac{1}{\sqrt{m}}\exp(-\lambda m) + n^{-1}\sqrt{m}$, and thus Proposition 4.4. □

Proof of Proposition 6.1. We first compute the coefficients of the centered Gaussian density. As Hermite polynomials of odd index are odd, the coefficients with odd index are null. We compute the coefficients with even index. Let

$$\bar{\sigma}^2 = (1+\sigma^{-2})^{-1} = \frac{\sigma^2}{1+\sigma^2}. \tag{28}$$

Note that if $2\bar{\sigma}^2 = 1$, *i.e.* $\sigma^2 = 1$, the coefficients are null except for $n = 0$.
We have

$$\int x^{2p}f_{\bar{\sigma}}(x)dx = C_{2p}\bar{\sigma}^{2p} \quad \text{with} \quad C_{2p} = 3\times 5\times 7\times\ldots\times(2p-1) = \frac{(2p)!}{2^p p!}.$$

Using that (see e.g. Lebedev (1972), formula (4.9.2) p.60)

$$H_{2j}(x) = \sum_{k=0}^{j}\frac{(-1)^k(2j)!}{k!(2j-2k)!}(2x)^{2j-2k},$$

we obtain:

$$\begin{aligned} a_{2j}(f_\sigma) &= (2j)!c_{2j}\frac{\bar{\sigma}}{\sigma}\sum_{k=0}^{j}\frac{(-1)^k}{k!(2j-2k)!}2^{2j-2k}C_{2(j-k)}\bar{\sigma}^{2(j-k)} = c_{2j}\frac{\bar{\sigma}}{\sigma}\frac{(2j)!}{j!}(2\bar{\sigma}^2-1)^j \\ &= c_{2j}\left(\frac{1}{1+\sigma^2}\right)^{1/2}\frac{(2j)!}{j!}\left(\frac{\sigma^2-1}{\sigma^2+1}\right)^j \end{aligned}$$

Note that $|\sigma^2-1)/(1+\sigma^2|<1$.
Analogously,

$$\begin{aligned}
a_{2j}(f_{p,\sigma}) &= \frac{(2j)!}{j!}c_{2j}\left(\frac{\bar\sigma}{\sigma}\right)^{2p+1}\sum_{k=0}^{j}\frac{(-1)^k j!\, 2^{2j-2k}\bar\sigma^{2(j-k)}}{k!(2j-2k)!C_{2p}}C_{2(j-k+p)}\\
&= \frac{(2j)!}{j!}c_{2j}\left(\frac{\bar\sigma}{\sigma}\right)^{2p+1}\sum_{k=0}^{j}\frac{(-1)^k j!}{k!(j-k)!}(2\bar\sigma^2)^{j-k}\frac{C_{2(j-k+p)}}{C_{2(j-k)}C_{2p}}\\
&= \frac{(2j)!}{j!}c_{2j}\left(\frac{\bar\sigma}{\sigma}\right)^{2p+1}\sum_{m=0}^{j}\frac{(-1)^{j-m} j!}{m!(j-m)!}(2\bar\sigma^2)^{m}\frac{C_{2(m+p)}}{C_{2m}C_{2p}}.
\end{aligned}$$

Now, we use the following result which is proved in Chaleyat-maurel and Genon-Catalot (2006, Lemma 3.1, p.1459):

$$\frac{C_{2(m+p)}}{C_{2m}C_{2p}}=\sum_{r=0}^{p} m(m-1)\ldots(m-r+1)\binom{p}{r}\frac{2^r}{C_{2r}}.$$

After some computations, we get:

$$a_{2j}(f_{p,\sigma}) = \frac{(2j)!}{(j-p)!}c_{2j}\left(\frac{\bar\sigma}{\sigma}\right)^{2p+1}(2\bar\sigma^2-1)^{j-p}S_p \quad \text{where}$$

$$S_p=\sum_{r=0}^{p}\frac{(n-p)!\; p!\; 2^r}{(n-r)!\; j!(p-r)!C_{2r}}(2\bar\sigma^2)^r(2\bar\sigma^2-1)^{p-r}.$$

Therefore,

$$|S_p|\le c(p)\left(2\bar\sigma^2+|2\bar\sigma^2-1|\right)^p,$$

which allows to bound $|a_{2j}|$ and ends the proof. □

References

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

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

[3] Belomestny, D., Comte, F. and Genon-Catalot, V. (2016) Nonparametric Laguerre estimation in the multiplicative censoring model. Preprint hal-01252143, version 3 and MAP5 2016-1.

[4] Birgé, L. and Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138**, 33-73.

[5] Bongioanni, B. and Torrea, J.L. (2006). Sobolev spaces associated to the harmonic oscillator. *Proc.Indian Acad. Sci. (math. Sci.)* **116** (3), 337-360.

[6] Chaleyat-maurel, M. and Genon-Catalot, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stoch. Proc. and Appl.* **116**, 1447-1467.

[7] Comte, F., Dedecker, J. and Taupin, M.L. (2008). Adaptive Density Deconvolution with Dependent Inputs. *Mathematical methods of Statistics*, **17**, 87-112.

[8] Comte, F. and Genon-Catalot, V. (2015) Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.* **9**, 1113-1149.

[9] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.*, **24**, 508-539.

[10] Efromovich, S. (1999) *Nonparametric curve estimation. Methods, theory, and applications.* Springer Series in Statistics. Springer-Verlag, New York.

[11] Efromovich, S. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Ann. Statist.* **36**, 1127-1155.

[12] Efromovich, S. (2009). Lower bound for estimation of Sobolev densities of order less 1/2. *J. Statist. Plann. Inference* **139**, 2261-2268.

[13] Ibragimov, I. A. and Has'minskii, R. Z. (1980) An estimate of the density of a distribution. (Russian) Studies in mathematical statistics, IV. Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **98**, 61-85, 161-162.

[14] Lebedev, N. N. (1972) *Special functions and their applications.* Revised edition, translated from the Russian and edited by Richard A. Silverman. Unabridged and corrected republication. Dover Publications, Inc., New York.

[15] Mabon, G. (2015). Adaptive deconvolution on the nonnegative real line. Preprint HAL hal-01076927, V2.

[16] Markett, C. (1984). Norm estimates for (C, δ) means of Hermite expansions and bounds for δ_{eff}. *Acta math. Hung.*, **43**, 187-198.

[17] Massart, P. (2007) *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6-23, 2003. With a foreword by Jean Picard. Lecture Notes in Mathematics, 1896. Springer, Berlin, 2007.

[18] Schipper, M. (1996). Optimal rates and constants in L2-minimax estimation of probability density functions. Mathematical Methods of Statistics, 5(3), 253-274.

[19] Schwartz, S.C. (1967). Estimation of a probability density by an orthogonal series. *Ann. Math. Statist.* **38**, 1261-1265.

[20] Shen, J. (2000). Stable and efficient spectral methods in unbounded domains using Laguerre functions. *SIAM J. Numer. Anal.* **38**, 1113-1133.

[21] Tsybakov, A. B. (2009) Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York.

[22] Szegö, G. (1975) *Orthogonal polynomials.* Fourth edition. American Mathematical Society, Colloquium Publications, Vol. XXIII. American mathematical Society, Providence, R.I.

[23] Walter, G.G. (1977). Properties of Hermite series estimation of probability density. *Annals of Statistics*, **5**, 1258-1264.