HERMITE ESTIMATION IN NOISY CONVOLUTION MODEL

Ousmane Sacko

HERMITE ESTIMATION IN NOISY CONVOLUTION MODEL

OUSMANE SACKO[1]

Abstract. In this paper, we study the problem of estimating a regression function in a convolution model. We consider the following model: $y(x_k) = h(x_k) + \varepsilon_k, h(x) = f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy, k = -n, \dots, n-1$ where g is assumed to be known and f is the unknown function to be estimated; the errors $(\varepsilon_k)_{-n\leqslant k\leqslant n-1}$ are independent and identically distributed (i.i.d.) such that $\mathbb{E}[\varepsilon_k] = 0$ and $\mathrm{Var}(\varepsilon_k) = \sigma_\varepsilon^2 < +\infty$, known; the points $(x_k = kT/n)_{-n\leqslant k\leqslant n-1}$ are deterministic and equispaced on the interval $[-T, T]$, where $0 < T < \infty$ is fixed. Two estimation methods for f are considered by exploiting the properties of the Hermite basis. We study the quadratic risk of each estimator. If f belongs to the Sobolev (first approach) or Sobolev-Hermite (second approach) spaces, we obtain rates of convergence. We also present an adaptive procedure to select the relevant parameter inspired by Goldenshluter and Lepski method, and prove that the resulting estimator satisfies an oracle inequality for sub-Gaussian ε's. Finally, we illustrate numerically these approaches. January 7, 2022

Key words: Convolution model; Hermite basis; Model selection; Nonparametric regression; Projection estimator

AMS Classification: 62G05 - 62G08

1. Introduction

Consider the convolution model

$$y(x_k) = h(x_k) + \varepsilon_k, \quad k = -n, \dots, n-1, \tag{1}$$

where

$$h(x) = f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy, \tag{2}$$

where the kernel function g is supposed to be known and f is the unknown function to be estimated; the errors $(\varepsilon_k)_{-n\leqslant k\leqslant n-1}$ are independent and identically distributed (i.i.d.) such that $\mathbb{E}[\varepsilon_k] = 0$ and $\mathrm{Var}(\varepsilon_k) = \sigma_\varepsilon^2 < +\infty$, known; the points $(x_k = kT/n)_{-n\leqslant k\leqslant n-1}$ are deterministic and equispaced on the interval $[-T, T]$, where $0 < T < \infty$ is fixed. This model appears in several application contexts: in Dynamic Contrast Enhanced (DCE) imaging data analysis (see Goh et al. (2005), Cuenod et al. (2006), Goh et al. (2007), Cao et al. (2010) and Comte et al. (2017)) and in the study of time-resolved measurements in fluorescence spectroscopy (see Gafni et al. (1975), McKinnon et al. (1977), O'Connor et al. (1979), Ameloot and Hendrickx (1983), Abramovich et al. (2013)). If the function of interest is the unknown function h, this problem is known as a fixed design regression model.
Nonparametric estimation of h has been studied at length in the literature, see Barron et al. (1999), Baraud (2000) and recently Comte and Genon-Catalot (2019) for random design. Estimating the density f of a random variable X when observing $Z = X + \varepsilon$ with ε independent of X with density g amounts to reconstruct f from an estimate of $f_Z = f \star g$. This problem is known as a deconvolution problem. It is an inverse problem which has also been studied extensively in the literature, see Carroll and Hall (1988), Fan (1991), Pensky and Vidakovic (1999), Comte et al. (2006), Delaigle et al. (2008), Mabon (2017), Comte and Genon-Catalot (2018), Sacko (2020) among others, see also the monograph of Meister (2009).

[1] Université de Paris, MAP5, UMR CNRS 8145.

Model (1) cumulates the two questions of regression and deconvolution, and this is why it is difficult. We mention that in Model (1), the unknowns f and the kernel are not necessarily densities.
When f and g are $[0,1]$-supported, Rice and Rosenblatt (1983) solved the problem (1) using a smoothing spline approach for $x_k = k/n$ with $k = 1, \dots, n$. They obtain a control of the risk for f of class C^4. However, the question of the smoothing parameter is not considered in their work. Another special case of Model (1) occurs when f and g are $\mathbb{R}^+$-supported, it is called Laplace convolution. Then, we have $h(x) = \int_0^x f(x-y)g(y)dy$, whose discrete noisy version is given by (1) with $k = 1, \dots, n$. It has been studied in Dey et al. (1998) for $g(x) = be^{-ax}1_{x\geqslant 0}$, using that the solution of (2) satisfies a linear differential equation. The authors compute convergence rates for $n \to \infty$, under the assumption that the s-th derivative of f is continuous, the procedure is not adaptive. Abramovich et al. (2013) study the Laplace deconvolution problem for g known: they summarize the estimating problem of f to estimation of the derivative of h. These derivatives are estimated by a kernel method, the procedure is adaptive and minimax optimal for f in a Sobolev class. Note that the rate depends on $T = T_n \to \infty$ as $n \to \infty$. Vareschi (2015) studies also the Laplace deconvolution problem using the Galerkin projection on Laguerre functions for a g kernel contaminated by white noise. More recently, Comte et al. (2017) proposed a projection estimator, based on the development of the functions f, g and h in the Laguerre basis. The coefficients of the decomposition of h are expressed as a linear combination of those of f, the link matrix being invertible. They also propose an adaptive procedure by penalization: the resulting estimator verifies an oracle inequality up to multiplicative $\log n$ factor. We emphasize that the $(x_k)_{1\leqslant k\leqslant n}$ are not necessary equispaced on $[0,T]$ and T is fixed. Finally, if, f is a function of 3 variables and g of one variable, Benhaddou et al. (2019) consider also the projection method on Laguerre and wavelet bases for a Gaussian white noise. Their method is adaptive and asymptotically optimal up to a logarithmic factor when f belongs to a three-dimensional Laguerre-Sobolev ball. Note that regression model and inverse problems can be encountered in different setting, see for instance Loubes and Marteau (2012) who study an econometric model; then, the inverse problem arises from instrumental variables taken as covariate.

However, all of the afore studies were conducted for $\mathbb{R}^+$ supported f and g. The novelty of the present work, is that we consider Model (1) with $\mathbb{R}$-supported functions and our aims are the following: Define a consistent estimator of f; Provide rates of convergence; Propose an adaptive procedure and illustrate numerically its performances. The Laguerre basis which is $\mathbb{R}^+$-supported clearly no longer suits for our problem. We consider here the Hermite basis which has non compact support and is well adapted in our context. When using compactly supported bases, the support is a fixed interval determined in practice from the dataset. Hermite basis does not require this preliminary choice and is well adapted in our context. Recently, Belomestny et al. (2019) show that the Hermite basis allows to build estimators of low complexity and therefore numerically fast.

In this paper, we first propose a Fourier-Hermite (denoted by FH in the sequel) approach to estimate f. It consists in estimating h as regression function by a nonparametric least squares method, based on the development of h in the Hermite basis. Then, we use the inverse Fourier transform to recover f. Contrary to Baraud (2000), we do not consider a compactly supported basis. Moreover, we obtain a new (to our knowledge) bound on the $\mathbb{L}^2(\mathbb{R})$-risk for regression function h. We provide an upper bound on the risk of the estimator of f which shows that a bias-variance compromise must be performed. For f belonging to a Sobolev ball, we obtain rates of convergence for adequate choice of some parameters (cut-off parameter and dimension of the regression function). We also present an adaptive procedure inspired by Goldenshluger and Lepski (2011) method to select the relevant parameters: the resulting estimator satisfies an oracle inequality for ε sub-Gaussian (see below or Vershynin (2012) for more details), and automatically realizes a bias-variance compromise up to a logarithm term. We also introduce another approach, called

the Hermite-Hermite (denoted by HH in the following) strategy. Both functions f and h are decomposed in the Hermite basis. We construct an estimator of f by replacing h by its nonparametric least squares estimator in the formula of the coefficients of f. As for the FH strategy, we provide a risk bound and the rate obtained therein for f belonging to a Sobolev-Hermite ball, and we propose a procedure to select automatically the relevant dimension.

The plan of the paper is the following: The study of the estimation of regression function h in the Hermite basis for fixed design is described in Section 2. Those results are exploited to study the FH and HH strategies. Section 3 is devoted to the FH strategy. In particular, we define the FH estimator in Section 3.1. A bias-variance decomposition is given in Section 3.2. In Section 3.3, we provide rates of convergence. Section 3.4 is devoted to selection of model for the FH procedure and an oracle inequality is proved for the resulting estimator therein. In Section 4, we describe the HH estimation strategy and a comparison with the FH method is performed. As for FH method, we also propose an adaptive procedure and an oracle inequality is proved in Section 4.4. Section 5 is devoted to the numerical study to illustrate the performance of the adaptive procedure and comparisons between FH and HH method are performed. Finally, all the proofs are presented in Section 6, technical Lemmas and some useful results are given in the Appendix.

2. Hermite regression estimation of h

We first present a study concerning the estimation of h. From this point of view, model (1) corresponds to a standard fixed design regression. Nonparametric estimation in this context can be found in Baraud (2000), who consider compactly supported bases. In view of the following steps for "extracting" f, we need to handle the non compactly supported Hermite basis. Let us start by recalling the definition and useful properties of this basis, and the associated regularity spaces.

2.1. **Notations.** For ϕ, ψ belonging to $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$, denote $\langle \varphi, \psi \rangle = \int \varphi(u)\overline{\psi(u)}du$ the scalar product on $\mathbb{L}^2(\mathbb{R})$ and $\|\varphi\|^2 = \int |\varphi(u)|^2 du$ the associated norm on $\mathbb{L}^2(\mathbb{R})$. The Fourier transform of φ is defined by $\varphi^*(u) = \int e^{iux}\varphi(x)dx$. Lastly, we recall the Plancherel-Parseval equality $\langle \varphi, \psi \rangle = (2\pi)^{-1}\langle \varphi^*, \psi^* \rangle$.

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

$$\varphi_j(x) = c_j H_j(x) e^{-x^2/2}, \quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j}(e^{-x^2}), \quad c_j = (2^j j! \sqrt{\pi})^{-1/2}, \quad x \in \mathbb{R},\ j \geqslant 0. \tag{3}$$

The Hermite polynomials $(H_j)_{j\geqslant 0}$ are orthogonal with respect to the weight function e^{-x^2}: $\int_{\mathbb{R}} H_j(x)H_k(x)e^{-x^2}dx = 2^j j!\sqrt{\pi}\delta_{j,k}$ (see Abramowitz and Stegun (1964), 22.2.14), where $\delta_{j,k}$ is the Kronecher symbol. It follows that the sequence $(\varphi_j)_{j\geqslant 0}$ is an orthonormal basis on $\mathbb{R}$. Moreover, φ_j is bounded by

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

(see Abramowitz and Stegun (1964), chap.22.14.17 and Indritz (1961)) and the following bound holds

$$\|\varphi_j\|_\infty \leqslant \frac{C_\infty}{(j+1)^{\frac{1}{12}}}, \tag{5}$$

where C_∞ is a constant given in Szegö (1959). The Fourier transform $(\varphi_j)_{j\geqslant 0}$ is given as follows

$$\varphi_j^* = \sqrt{2\pi}(i)^j \varphi_j. \tag{6}$$

From Askey and Wainger (1965), it holds:

$$|\varphi_j(x)| \leqslant C'_\infty e^{-\xi x^2}, \quad |x| \geqslant \sqrt{2j+1}, \tag{7}$$

where C'_∞ and ξ are constants independent of x and j. The infinity norm of the derivative of φ_j satisfies (see Comte and Genon-Catalot (2018), Lemma 7.3):

$$\|\varphi'_j\|_\infty \leqslant C''_\infty (j+1)^{\frac{5}{12}}, \quad j \geqslant 0, \tag{8}$$

where $C''_\infty > 0$ is a numerical constant.

2.3. **Regularity spaces.** We consider in the sequel the following regularity spaces (see Bongioanni and Torrea (2006)).

Definition 2.1. *Let $s, L > 0$, define the Sobolev-Hermite ball of regularity s by*

$$W_H^s(L) = \{\theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k\geqslant 0} k^s a_k^2(\theta) \leqslant L\}, \text{ where } a_k(\theta) = \int \theta(x)\varphi_k(x)dx. \tag{9}$$

For s an integer, it is proved in Bongioanni and Torrea (2006) and Belomestny et al. (2019) (see Proposition 4) that θ belongs to $W_H^s(L)$ if and only if θ admits derivatives up to order s and if the functions $\theta, \theta', \ldots, \theta^{(s)}, x^{s-l}\theta^{(l)}$ for $l = 0, \ldots, s-1$ belong to $\mathbb{L}^2(\mathbb{R})$. Recall also that the usual Sobolev ball $W^s(L)$ is defined, for $s > 0$ by

$$W^s(L) = \{\theta \in \mathbb{L}^2(\mathbb{R}), \int (1+u^2)^s |\theta^*(u)|^2 du < L\}. \tag{10}$$

If s is an integer and $L > 0$, it holds (see Bongioanni and Torrea (2006) and Belomestny et al. (2019)) then; $\ll \theta \in W^s(L) \gg$ is equivalent to $\ll$ there exists $L^* > 0$ such that $\sum_{j=0}^s \|f^{(j)}\|^2 < L^* \gg$.
Thus, it follows that $W_H^s(L) \subset W^s(L^*)$. Moreover, if $f \in W^s(L)$ has compact support, then $f \in W_H^s(L^*)$. In other words, $W_H^s(L)$ and $W^s(L^*)$ coincide for compactly supported functions.

2.4. **Definition of the regression estimator.** Let $d \geqslant 1$ an integer and

$$S_d := \text{span}\{\varphi_0, \ldots, \varphi_{d-1}\}, \tag{11}$$

the linear space generated by $\varphi_0, \ldots, \varphi_{d-1}$, where φ_j is the Hermite basis defined in (3). Assume that h belongs to $\mathbb{L}^2(\mathbb{R})$. Then, we can write $h = \sum_{j\geqslant 0} b_j(h)\varphi_j$, with $b_j(h) = \langle h, \varphi_j\rangle$. Moreover, we define $h_d = \sum_{j=0}^{d-1} b_j(h)\varphi_j$, the orthogonal projection of h on S_d. Introduce the matrices:

$$\Phi_d = (\varphi_j(x_i))_{-n\leqslant i\leqslant n-1, 0\leqslant j\leqslant d-1}, \quad \Psi_d = \frac{T}{n}\Phi_d^t\Phi_d, \tag{12}$$

where Φ_d^t denotes the transpose of the matrix Φ_d. We need of following Lemma to get an estimator of h.

Lemma 2.1. *For all $d \leqslant n$, Ψ_d is invertible.*

By the least squares method and Lemma 2.1, we derive the following projection estimator of h on S_d:

$$\widehat{h}_d = \sum_{j=0}^{d-1} \widehat{b}_j^{(d)}\varphi_j, \text{ where } \vec{\widehat{b}}^{(d)} = (\widehat{b}_0^{(d)}, \ldots, \widehat{b}_{d-1}^{(d)})^t = (\Phi_d^t\Phi_d)^{-1}\Phi_d^t\vec{y} = \frac{T}{n}\Psi_d^{-1}\Phi_d^t\vec{y}, \tag{13}$$

$\vec{y} = (y(x_{-n}), \ldots, y(x_{n-1}))^t$.

Comment on the assumption $h \in \mathbb{L}^2(\mathbb{R})$. Let $1 \leqslant p, q, r \leqslant \infty$ such that $1/p + 1/q = 1 + 1/r$. Let us recall that with the Young inequality, we have $\|h\|_r = \|f \star g\|_r \leqslant \|f\|_p\|g\|_q$. Thus, for ($f \in \mathbb{L}^2(\mathbb{R})$ and $g \in \mathbb{L}^1(\mathbb{R})$) or ($g \in \mathbb{L}^2(\mathbb{R})$ and $f \in \mathbb{L}^1(\mathbb{R})$), it follows that $h \in \mathbb{L}^2(\mathbb{R})$.

2.5. **Risk bound of $\widehat{h}_d$ and rate of convergence.** For any s, t in $\mathbb{L}^2(\mathbb{R})$, we define:

$$\|t\|_n^2 := \frac{T}{n}\sum_{i=-n}^{n-1} t^2(x_i), \quad \langle s,t\rangle_n := \frac{T}{n}\sum_{i=-n}^{n-1} s(x_i)t(x_i),$$

The following bias-variance decompositions hold.

Proposition 2.1. *Let $(x_i, y(x_i))_{-n\leqslant i\leqslant n-1}$ be observations from model (1). Assume that h belongs to $\mathbb{L}^2(\mathbb{R})$ and consider the estimator $\widehat{h}_d$ defined in (13).*

(i) Then, it holds that

$$\mathbb{E}\left[\|\widehat{h}_d - h\|_n^2\right] = \inf_{t\in S_d}\|t-h\|_n^2 + \sigma_\varepsilon^2 T\frac{d}{n}. \tag{14}$$

(ii) Moreover, we have

$$\mathbb{E}[\|\widehat{h}_d - h\|^2] \leqslant \|h-h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h-h_d\|_n^2 + \sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right), \tag{15}$$

where $\mathrm{tr}(A)$ is the trace of the matrix A and $\lambda_{max}(A)$ denotes the spectral radius of the matrix A.

The part (i) of Proposition 2.1 corresponds to a classical bias-variance decomposition for the empirical norm $\|\cdot\|_n$. The first term in the right-hand side of (14) is the bias term and the second term is the variance term. They behave in the opposite way with respect to d: $\inf_{t\in S_d}\|t-h\|_n^2$ decreases with d while $\sigma_\varepsilon^2 Td/n$ increases with d. The risk bound given in (15) is new to our knowledge and handles the integrated $\mathbb{L}^2$ risk on $\mathbb{R}$. It is a bias-variance decomposition with bias equal to $\|h-h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h-h_d\|_n^2$ and variance $\sigma_\varepsilon^2\mathrm{tr}\left(\Psi_d^{-1}\right)T/n$. In both cases, we have a bias-variance trade-off to make.

The bias term is studied by exploiting the specific property of the Hermite basis. The following Lemma leads to find the order of the bias:

Lemma 2.2. *Assume that h belongs to $W_H^\alpha(L)$ (Sobolev-Hermite ball defined in (9)).*

(i) If $\alpha > 11/6$, we have $\|h-h_d\|_n^2 \leqslant \|h-h_d\|^2 + C(\alpha,L)\frac{T^2}{n}$, where $C(\alpha,L)$ is a positive constant depending only on α and L.

(ii) If $\alpha > 17/6$, it hold that $\|h-h_d\|_n^2 \leqslant \|h-h_d\|^2 + C'(\alpha,L)\frac{T^3}{12n^2}$, where $C'(\alpha,L)$ is a positive constant which depends on α and L.

For fixed T, the additional term T^2/n or T^3/n^2 is a residual term which is negligible compared to the variance term $\sigma_\varepsilon^2 dT/n$ for the empirical norm or $\sigma_\varepsilon^2\mathrm{tr}\left(\Psi_d^{-1}\right)T/n$ for the integral $\mathbb{L}^2(\mathbb{R})$-norm. Furthermore, to get the rate of convergence for the integral norm $\|\cdot\|$, we have to control $\mathrm{tr}\left(\Psi_d^{-1}\right)$ and $\lambda_{\max}(\Psi_d^{-1})$. We consider the following assumption

(**A0**) There exists a constant $\lambda > 0$ such that the maximum eigenvalue of Ψ_d^{-1} satisfies

$$\lambda_{\max}(\Psi_d^{-1}) \leqslant \lambda < +\infty,$$

uniformly in d.

For n large enough and T, d well chosen, we can show that

$$|||\Psi_d^{-1} - I_d|||^2 \underset{n\to+\infty}{\longrightarrow} 0,$$

where $|||\cdot|||$ is any matrix norm (see Section B in Appendix). It follows that Assumption (**A0**) holds asymptotically with λ near of 1. The same type of hypothesis can be found in Comte et al. (2017) (see Assumption 4) and Vareschi (2015) (see Assumption 2.3). Then, we can deduce the rate of convergence.

Proposition 2.2. *Assume that h belongs to $W_H^\alpha(L)$ with $\alpha > 11/6$ and select $d_{opt} = [n^{1/(\alpha+1)}]$.*

(i) Then, we have

$$\sup_{h\in W_H^\alpha(L)} \mathbb{E}\left[\|\widehat{h}_{d_{opt}} - h\|_n^2\right] \leqslant C(\alpha, L, T, \sigma_\varepsilon) n^{-\frac{\alpha}{\alpha+1}}, \tag{16}$$

where $C(\alpha, L, T, \sigma_\varepsilon)$ depends on α, L, T and σ_ε.

*(ii) If in addition (**A0**) is satisfied, it yields that*

$$\sup_{h\in W_H^\alpha(L)} \mathbb{E}\left[\|\widehat{h}_{d_{opt}} - h\|^2\right] \leqslant C(\alpha, L, T, \sigma_\varepsilon, \lambda) n^{-\frac{\alpha}{\alpha+1}}. \tag{17}$$

Our estimator reaches the same rate as in the case where (x_i) are random variables (see Comte and Genon-Catalot (2019)). From the lower bound stated therein, this rate is optimal when we use the Laguerre or the Hermite basis (at least for gaussian ε's). Note that it is not standard and is specific to the Laguerre and Hermite basis: in Baraud (2000), Baraud (2002), Barron et al. (1999), the least squares estimator converges with rate $n^{-2\alpha/(2\alpha+1)}$ if the regression function h belongs to a Besov space with regularity index α. The reason is that the variance order does not depend on the basis used while bias order does and changes according to the regularity spaces associated with the basis.

Remark 1. *The constraint $\alpha > 11/6$ or $\alpha > 17/6$ comes from the study of $\|h - h_d\|_n^2$ (see the Proof of Lemma 2.2). It excludes some functions h (e.g. Cauchy since $\alpha = 3/2 - \eta$ with $0 < \eta < 3/2$ see Section 4 in Belomestny et al. (2019)). Without this constraint, we have for $\alpha \geqslant 1$ and $h \in W_H^\alpha(L)$*

$$\|h - h_d\|_n^2 = \frac{T}{n}\sum_{i=-n}^{n-1}(h_d(x_i) - h(x_i))^2 \leqslant 2T\phi_0^2\left(\sum_{j\geqslant d} j^{\alpha/2}a_j(h)j^{-\alpha/2}\right)^2 \lesssim d^{-\alpha+1},$$

where ϕ_0 is given in (4). It follows for the choice $d_{opt} = [n^{1/(\alpha+2)}]$ that $\mathbb{E}\left[\|\widehat{h}_{d_{opt}} - h\|_n^2\right] = \mathcal{O}(n^{-\frac{\alpha-1}{\alpha+2}})$. This rate is worse than the one obtained in (16). The estimator remains consistent in this case even if the rate deteriorates. In the sequel, we will see that the condition $\alpha > 11/6$ or $\alpha > 17/6$ is often satisfied.

2.6. **Adaptive estimator for h.** However, the choice of $d = d_{opt}$ depends on the regularity of h which is unknown; thus this choice is only theoretical and cannot be used in practice. This is why an adaptive procedure is developed now. It allows to choose the relevant dimension by replacing the bias and variance terms by computable quantities. Let $\gamma_n(\cdot)$ be the empirical contrast:

$$\gamma_n(t) = \frac{T}{n}\sum_{i=-n}^{n-1}\left[y(x_i) - t(x_i)\right]^2.$$

It is easy to see that $\widehat{h}_d = \underset{t\in S_d}{\arg\min}\gamma_n(t)$. The quantity $\gamma_n(\widehat{h}_d) = -\|\widehat{h}_d\|_n^2$ is a classical estimator of the bias term. Then, we select the space S_d by setting:

$$\widehat{d} := \underset{d\in\mathcal{M}_n}{\arg\min}\{\gamma_n(\widehat{h}_d) + \mathrm{pen}(d)\}, \quad \text{where} \quad \mathrm{pen}(d) = \kappa T\frac{d}{n}\sigma_\varepsilon^2, \quad \kappa > 1 \tag{18}$$

where $\mathcal{M}_n = \{1, \ldots, d_{\max}\}$, $d_{\max} \leqslant n$ is the maximal dimension which depends on n and κ is a positive numerical constant. The constant κ is independent of the data and a value must be assigned in practice. Methods are proposed in Baudry et al. (2012) and programs for density estimation are given in the Softwares R and Matlab called "Capushe". The following oracle inequalities hold for the resulting estimator.

Theorem 2.3. *Let $(x_i, y(x_i))_{-n\leqslant i\leqslant n-1}$ be observations, from model (1). Assume that $\mathbb{E}[\varepsilon_1^8] < \infty$.*

(i) Then, the estimator $\widehat{h}_{\widehat{d}}$ satisfies:

$$\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|_n^2] \leqslant C(\kappa) \inf_{d\in\mathcal{M}_n} \left(\inf_{t\in S_d} \|t-h\|_n^2 + \sigma_\varepsilon^2 T \frac{d}{n} \right) + \frac{C'T}{n}, \tag{19}$$

where $C(\kappa) = 2\kappa(1+4/(\kappa-1)) > 1$ (for instance for $\kappa = 2.5$, $C(2.5) = 9.17$) and $C' > 0$ are numerical constants.

*(ii) If in addition (**A0**) holds, we have*

$$\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|^2] \leqslant C_1 \inf_{d\in\mathcal{M}_n} \left((2\lambda^2+1)\|h-h_d\|_n^2 + \|h_d - h\|^2 + \sigma_\varepsilon^2 T \frac{d}{n} \right) + \frac{C_1'\lambda T}{n}, \tag{20}$$

*where λ is given in (**A0**), $C_1 = \max(1, 2\lambda^2 C(\kappa))$ and $C_1' = 2C'$ are positive constants.*

The estimator $\widehat{h}_{\widehat{d}}$ is adaptive and minimax optimal in the sense that the bias-variance compromise is realized automatically, since $C'T/n$ and $\lambda C_1' T/n$ are residual terms. Indeed, for $h \in W_H^\alpha(L)$, we deduce from Proposition 2.2 that $\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|_n^2] \lesssim n^{-\frac{\alpha}{\alpha+1}}$ and $\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|^2] \lesssim n^{-\frac{\alpha}{\alpha+1}}$. Theorem 2.3 is a consequence of Theorem 3.1 given in Baraud (2000) and the bound given in (15).

Remark 2. *The variance σ_ε^2 of the noise which appears in (18) is assumed to be known but is in general unknown and must be estimated. A classical estimator is the residual least squares estimator:*

$$\widehat{\sigma_\varepsilon^2} := \frac{T}{n} \sum_{i=-n}^{n-1} \left[y(x_i) - \widehat{h}_{d^*}(x_i) \right]^2,$$

where d^* is an arbitrarily chosen dimension (for instance $d^* = [\sqrt{n}]$ suits see Baraud (2000)).

3. Fourier-Hermite approach for the estimation of f

In this section, we construct an estimator of f using the Fourier inverse transform and then the least squares estimator. First, we consider the following assumption on the unknown f.

(**A1**) The unknown function f and its Fourier transform f^* belong to $\mathbb{L}^1(\mathbb{R})$.

Assumption (**A1**) is introduced to use the Fourier transform inverse: $t(x) = 1/(2\pi)\int_{\mathbb{R}} e^{iux} t^*(u)du$.
We will also need of the following assumption on the kernel g which are classical in deconvolution context:

(**A2**) The Fourier transform of g denoted g^* is well defined and such that: $g^* \neq 0$, where $t^*(u) = \int e^{iux} t(x)dx$, and i is the complex number with $i^2 = -1$.

(**A3**) There exist $c_1 \geqslant c_1' > 0$, and $\gamma \geqslant 0$, such that

$$c_1'(1+t^2)^\gamma \leqslant |g^*(t)|^{-2} \leqslant c_1(1+t^2)^\gamma, \quad \forall t \in \mathbb{R}. \tag{21}$$

(**A2**) is necessary to define the estimator and (**A3**) is generally useful to study its risk. Under (**A3**), the function g and the errors are called "ordinary smooth". Observe that (**A3**) implies (**A2**) and is verified by some classical distributions: we can cite for example the Laplace distribution (with $\gamma = 2$), Gamma distributions ($\gamma = p$, where p is the shape parameter) and more generally for all symmetric Gamma distributions.

3.1. Estimation procedure.

Consider discrete observations $(x_k, y(x_k))_{-n\leqslant k\leqslant n-1}$ from model (1). As $h = f \star g$ (see (2)), under (**A1**), (**A2**) and using the Fourier inversion formula, we have:

$$f(x) = \int_{\mathbb{R}} e^{-iux} \frac{h^*(u)}{g^*(u)} du, \quad \forall x \in \mathbb{R}. \tag{22}$$

Equation (22) leads to an estimator of f by replacing h by an estimator. By taking the Fourier transform of (13), it yields

$$\widehat{h}_d^*(u) = \sum_{j=0}^{d-1} \widehat{b}_j^{(d)} \varphi_j^*(u). \tag{23}$$

Plugging (23) in (22), we introduce the following estimator of f:

$$\widehat{f}_{(d)}(x) = \frac{1}{2\pi} \int e^{-iux} \frac{\widehat{h}_d^*(u)}{g^*(u)} du. \tag{24}$$

The estimator is well defined because the Hermite basis decreases as $e^{-\xi x^2}$ (see (7)), which makes the ratio $\widehat{h}_d^*/g^*$ integrable for many functions g (see also Sacko (2020)). The quality of $\widehat{f}_{(d)}$ is related to that of $\widehat{h}_d$ which is studied in Section 2. The dimension d must be optimized. In practice, we must introduce a cut-off to compute $\widehat{f}_{(d)}$. Moreover, to control the risk of $\widehat{f}_{(d)}$, we first consider the following estimator

$$\widehat{f}_{(\ell),d}(x) = \frac{1}{2\pi} \int_{-\ell}^{\ell} e^{-iux} \frac{\widehat{h}_d^*(u)}{g^*(u)} du, \text{ for } \ell > 0. \tag{25}$$

3.2. **Risk bound for the deconvolution estimator.** Now, we study the integrated quadratic risk of $\widehat{f}_{(d)}$ given by (24). Define

$$\Delta(\ell) = \sup_{|u| \leqslant \ell} |g^*(u)|^{-2}, \qquad f_{(\ell)}(x) = \frac{1}{2\pi} \int_{-\ell}^{\ell} e^{-iux} \frac{h^*(u)}{g^*(u)} du, \tag{26}$$

Consider also the following assumption:

(**A4**) $\|h\|_\infty = \sup_{x \in \mathbb{R}} |h(x)| < \infty$.

We recall that, by the Cauchy-Schwarz inequality, $\|h\|_\infty \leqslant \|f\| \|g\|$. Therefore, if f and g are square integrable then, condition (**A4**) is automatically satisfied.

Then, we can state the following upper bound on the risk.

Proposition 3.1. *Suppose that the assumptions (**A0**) to (**A4**) hold. For $\widehat{f}_{(d)}$ given in (24), $\widehat{f}_{(\ell),d}$ defined in (25) and $\ell \geqslant \sqrt{2d}$, we have*

$$\mathbb{E}\left[\|\widehat{f}_{(d)} - f\|^2\right] \leqslant 2C\lambda T e^{-\xi d} + 2\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right], \tag{27}$$

*where C is a constant depending on C'_∞, ξ given in (7), c_1 in (**A3**) and $\|h\|_\infty$. For $\widehat{f}_{(\ell),d}$ defined in (25) and any $\ell > 0$, it holds that*

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right] \leqslant \|f - f_{(\ell)}\|^2 + \Delta(\ell) \left(\|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right) \|h - h_d\|_n^2 + \sigma_\varepsilon^2 \frac{T}{n} \mathrm{tr}\left(\Psi_d^{-1}\right) \right). \tag{28}$$

(a) The first term on the right-hand side of (28) ($\|f - f_{(\ell)}\|^2 = \frac{1}{2\pi}\int_{|u|>\ell} |f^*(u)|^2 du$) is the classical bias term: it is decreasing with the cut-off ℓ.
(b) The term $\Delta(\ell)$ corresponds to the deconvolution aspect of problem: it is studied using the regularity condition on g^* given in (**A3**) and is increasing with ℓ.
(c) Finally, the terms in the big parenthesis represent the regression aspect of problem (see Proposition 2.1 (ii)).

We also mention that the term $C\lambda T e^{-\xi d}$ is negligible compared to $\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right]$ for ℓ large enough and $f \in W^s(L)$ (Sobolev ball see (10) for the definition of $W^s(L)$) and under (**A3**). Then, the two estimators ($\widehat{f}_{(\ell),d}$ and $\widehat{f}_{(d)}$) have the same rate of convergence. We can also consider $\widehat{f}_{(\ell),d}$ as an estimator. However, this requires to optimize two parameters, the cut-off ℓ and the dimension d in practice, contrary to $\widehat{f}_{(d)}$ which requires only to optimize d.

3.3. **Rate of convergence of $\widehat{f}_{(\ell),d}$ and $\widehat{f}_{(d)}$.** In this section, we compute rates of convergence in a collection of specified cases. To derive convergence results, we will make two consecutive bias-variance compromises, first for the regression part (compromise in (17)) and then for the deconvolution part, by substituting this value in (28) and optimizing in ℓ to get the rates of $\widehat{f}_{(\ell),d}$ and $\widehat{f}_{(d)}$. The following result of convergence holds.

Theorem 3.2. *Let assumptions (**A0**) to (**A3**) hold. Assume that $h \in W_H^{s+\gamma}(L')$, then we have for $d_{opt} = \lceil n^{1/(s+\gamma+1)} \rceil$ with $s + \gamma > 11/6$ and $\ell_{opt} \propto n^{1/2(s+\gamma+1)}$ that*

$$\sup_{f \in W^s(L)} \mathbb{E}\left[\|\widehat{f}_{(\ell_{opt}),d_{opt}} - f\|^2\right] = \mathcal{O}\left(n^{-\frac{s}{s+\gamma+1}}\right),$$

*where $W^s(L)$ is the classical Sobolev ball of regularity s defined in (10) and γ is given in (**A3**).*

The same result holds for the estimator $\widehat{f}_{(d_{opt})}$ with the assumption (**A4**), see (27). The estimator $\widehat{f}_{(\ell_{opt}),d_{opt}}$ and $\widehat{f}_{(d_{opt})}$ converge at a polynomial rate as in density deconvolution for ordinary smooth noise.

Clearly, the hypothesis $h \in W_H^{s+\gamma}(\cdot)$ can be related to the regularity of f and g.
Note that as $\ell_{opt}^2 \propto d_{opt}$, then, we can just set $\ell = c\sqrt{2d}$ with $c \geqslant 1$ in the constraint $\ell \geqslant \sqrt{2d}$ given in Proposition 3.1. If we had a Fourier bias instead of Hermite bias (*i.e.* we have $\|h - h_{(\sqrt{d})}\|^2$ instead of $\|h - h_d\|^2$), for $f \in W^s(L)$ and under (**A3**), we have by an elementary calculation that $h = f \star g \in W^{s+\gamma}(L/c_1')$ (see Remark 3). Therefore, it yields under (**A0**) to (**A3**) that $\sup_{f \in W^s(L)} \mathbb{E}\left[\|\widehat{f}_{(\ell_{opt}),d_{opt}} - f\|^2\right] = \mathcal{O}\left(n^{-\frac{s}{s+\gamma+1}}\right)$.

Remark 3. *Assume that f belongs to $W^s(L)$ (see Section 2.3) and g is ordinary smooth (i.e. g satisfies (21)). Then, h belongs to $W^{s+\gamma}(L/c_1')$, where c_1' is given in (21). Indeed, we have*

$$\int (1+u^2)^{s+\gamma} |h^*(u)|^2 du = \int (1+u^2)^s |f^*(u)|^2 (1+u^2)^\gamma |g^*(u)|^2 du \leqslant \frac{1}{c_1'} \int (1+u^2)^s |f^*(u)|^2 du \leqslant \frac{L}{c_1}.$$

We derive that h is $s+\gamma$ times differentiable if $s+\gamma$ is assumed integer and these derivatives up to order $s+\gamma$ belong to $\mathbb{L}^2(\mathbb{R})$. Then, it belongs to $W_H^{s+\gamma}(L)$ if and only if the functions $x^{s+\gamma-\eta} h^{(\eta)}$ belong to $\mathbb{L}^2(\mathbb{R})$ for $\eta = 0, \ldots, s+\gamma-1$ (see Section 2.3).

For some classical functions, we can obtain the exact order of bias of the unknown function f and the regression function h. We only calculate the rate for $\widehat{f}_{(\ell),d}$, these results extend naturally to $\widehat{f}_{(d)}$ (see Equation (27)) considering (**A4**).

3.3.1. *Rate of convergence for f Gaussian.* Let

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{29}$$

we can establish the following result.

Proposition 3.3. *Let assumptions* **(A0)** *to* **(A3)** *hold and* $f = f_\sigma$ *where* f_σ *is defined in (29). Further suppose that* $x^\alpha g \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ *for* α *an integer which can be chosen as large as possible and* $l = 0, \dots, \alpha - 1$. *Set* $d_{opt} = [n^{1/(\alpha+1)}]$ *and* $\ell^2_{opt} = \beta \log(n)$ *with* $\beta = \alpha/(\alpha+1)\sigma^2$, *we have*

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

where γ *is given in* **(A3)**.

Note that the condition $x^\alpha g \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ holds for classical ordinary smooth functions (Laplace or Gamma distributions). As α can be chosen large, then, for $\alpha \to +\infty$ (which corresponds to $d_{opt} = 1$), $\widehat{f}_{(\ell_{opt}),d_{opt}}$ is order $\log(n)^\gamma/n$. In this case, the rate $\log(n)^\gamma/n$ is better than the rate obtained in the classical density deconvolution since the rate is order $\log(n)^{\gamma+1/2}/n$, see Butucea (2004).

3.3.2. *Rate of convergence for Gaussian kernel.* By reversing the role of f and g in Proposition 3.3, namely that $g(x) = (2\pi\sigma^2)^{-1/2} e^{-\frac{x^2}{2\sigma^2}}$ and $f \in W^s(L)$, we recover the classical rate of the density deconvolution framework, see Fan (1993) and Pensky and Vidakovic (1999).

Proposition 3.4. *Let Assumptions* **(A0)**, **(A1)** *and* **(A3)** *hold,* $g(x) = (2\pi\sigma^2)^{-1/2} e^{-\frac{x^2}{2\sigma^2}}$, $f \in W^s(L)$ *and* $x^\alpha f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ *for* α *an integer which can be chosen as large as desired and* $l = 0, \dots, \alpha - 1$. *Then, we have for* $d_{opt} = [n^{1/(\alpha+1)}]$ *and* $\ell^2_{opt} = \frac{\sigma^2\alpha}{2(\alpha+1)} \log(n)$ *that*

$$\mathbb{E}\left[\|\widehat{f}_{(\ell_{opt}),d_{opt}} - f\|^2\right] \lesssim \log(n)^{-s}.$$

3.3.3. *Rate of convergence for* f *and* g *Gaussian.* If f and h belong to $W^s_H(L)$ and are of Gaussian-type, the order of the bias term decreases exponentially (see Belomestny et al. (2019), section 4.3 and Lemma 2 in Comte and Lacour (2011)). The rate is therefore imposed by the variance term.

Proposition 3.5. *Assume that* **(A0)**, **(A1)** *and* **(A2)** *hold,* $f(x) = (2\pi\sigma^2)^{-1/2} e^{-\frac{x^2}{2\sigma^2}}$ *and* $g(x) = (2\pi\theta^2)^{-1/2} e^{-\frac{x^2}{2\theta^2}}$ *with* $\sigma^2 + \theta^2 \neq 1$. *Then, for* $d_{opt} = [\log(n)/\lambda_{\sigma,\theta}]$ *with,* $\lambda_{\sigma,\theta} = \log\left[\left(\frac{\sigma^2+\theta^2+1}{\sigma^2+\theta^2-1}\right)^2\right]$, *we have*

$$\mathbb{E}\left[\|\widehat{h}_{d_{opt}} - h\|^2\right] \lesssim \frac{\log(n)}{n}. \tag{30}$$

Consequently, it comes for $\ell^2_{opt} = \frac{1}{\sigma^2+\theta^2}\log(n) - \frac{3}{2}\frac{1}{\theta^2+\sigma^2}\log\log(n)$ *that*

$$\mathbb{E}\left[\|\widehat{f}_{(\ell_{opt}),d_{opt}} - f\|^2\right] \lesssim n^{-\frac{\sigma^2}{\sigma^2+\theta^2}} \log(n)^{\frac{\sigma^2-\frac{\theta^2}{2}}{\sigma^2+\theta^2}}.$$

The same result holds if f is a mixture of Gaussian random variables. It is known that the rates in double super smooth case are of type $n^{-\delta}$ with $\delta > 0$ up to a certain power of $\log(n)$ (see Lacour (2006), Theorem 3.1 in density deconvolution setting).
Note that if $\sigma^2 + \theta^2 = 1$, we have $h = f \star g = (\sqrt{2})^{-1}(\pi)^{-\frac{1}{4}}\varphi_0$ where φ_0 is the first function of the Hermite basis given by (3), in this case $h_d = h$ and $\|h - h_d\| := 0$ which implies that the rate can be better than the one given in Proposition 3.5.

3.3.4. *Rate of convergence for the Gamma case.* When f is $\Gamma(p,\theta)$ and g $\Gamma(q,\theta)$, where $\Gamma(a,b)$ is the Gamma distribution of with shape parameter a and scale b, then, the regression function h is $\Gamma(p+q,\theta)$. If in addition the shape parameter is an integer, we can derive the exact bias order of h and then the rate of convergence.

Proposition 3.6. *Let* (A0) *to* (A3) *hold, p and q be two integers such that $p+q>2$. Assume that $f\sim\Gamma(p,\theta)$ and $g\sim\Gamma(q,\theta)$. For $d_{opt}=[n^{1/(p+q-1)}]$, we have*

$$\mathbb{E}\left[\|\widehat{h}_{d_{opt}}-h\|^2\right]\lesssim n^{-\frac{p+q-2}{p+q-1}}.$$

Therefore, it follows for $\ell_{opt}\propto n^{\frac{p+q-2}{(p+q-1)(2p+2q-1)}}$ that

$$\mathbb{E}\left[\|\widehat{f}_{(\ell_{opt}),d_{opt}}-f\|^2\right]=\mathcal{O}\left(n^{-\frac{(p+q-2)(2p-1)}{(p+q-1)(2p+2q-1)}}\right).$$

The estimator $\widehat{f}_{(\ell_{opt}),d_{opt}}$ converges with rate $n^{-(p+q-2)(2p-1)/(p+q-1)(2p+2q-1)}$ if f and g are Gamma functions. The same results holds if f is a mixture of Gamma function.

Let us now summarize the previous results in the Table 1:

f \ g	Gaussian $\mathcal{N}(0,\theta^2)$	Gamma $\Gamma(q,\theta)$
Gaussian $\mathcal{N}(0,\sigma^2)$	$n^{-\frac{\sigma^2}{\sigma^2+\theta^2}}\log(n)^{\frac{\sigma^2-\frac{\theta^2}{2}}{\sigma^2+\theta^2}}$	$\log(n)^q n^{-\frac{\alpha}{\alpha+1}}$ α large
Gamma $\Gamma(q,\theta)$	$\log(n)^{-p+\frac{1}{2}}$	$n^{-\frac{(p+q-2)(2p-1)}{(p+q-1)(2p+2q-1)}}$

TABLE 1. Rate of convergence for the MISE of $\widehat{f}_{(\ell_{opt}),d_{opt}}$ in the specific cases.

3.4. **Adaptive procedure for Fourier-Hermite strategy.** The objective of this section is to propose a way of selection for the estimator $\widehat{f}_{(\ell),d}$. First, we remark that $\widehat{f}_{(\ell),d}$ cannot be written as a minimizer of a contrast. Thus, we cannot use a procedure by penalization. This is why, we describe an adaptive choice inspired by the ideas developed by Goldenshluger and Lepski (2011). The procedure is mainly based on the comparison of estimators of f. From now, we set $\ell=\sqrt{2d}$ and introduce the following estimator

$$\widetilde{f}_{(d)}(x):=\widehat{f}_{(\sqrt{2d}),d}(x)=\frac{1}{2\pi}\int_{-\sqrt{2d}}^{\sqrt{2d}}e^{-iux}\frac{\widehat{h}^*_d(u)}{g^*(u)}du. \tag{31}$$

This choice of ℓ is motivated by the results obtained in Proposition 3.1 and Theorem 3.2. Indeed: the optimal choice of ℓ is the order of $\sqrt{d}$ and as the minimal admissible choice is $\ell=\sqrt{2d}$; this is why, we set $\ell=\sqrt{2d}$.
Consider the following collection of models

$$\mathcal{M}_n^{(1)}:=\{1\leqslant d\leqslant n,\quad \frac{\sigma_\varepsilon^2\lambda T d\Delta(\sqrt{2d})}{n}\leqslant 1\}$$

where $\Delta(d)$ is given by (26) and λ in (A0). Define

$$\widehat{A}(d):=\max_{d'\in\mathcal{M}_n^{(1)}}\left\{\left(\|\widetilde{f}_{d'}-\widetilde{f}_{d\wedge d'}\|^2-\kappa_1 V(d')\right)_+\right\}, \tag{32}$$

where $\kappa_1>0$ is numerical constant which must be calibrated in practice by simulations and

$$V(d)=2\left(1+24\log(n)\right)\sigma_\varepsilon^2\Delta(\sqrt{2d})\frac{\lambda dT}{n}. \tag{33}$$

Then, we select $\widehat{d}$ as follows

$$\widehat{d} := \arg \min_{d \in \mathcal{M}_n^{(1)}} \left\{ \widehat{A}(d) + \kappa_2 V(d) \right\}, \tag{34}$$

where $\kappa_1 \leqslant \kappa_2$ and κ_2 must be also calibrated. The term $\widehat{A}(d)$ is an estimator of bias of $\widetilde{f}_{(d)}$ and its construction is based on the comparison of estimators of f. We add the following assumption on the noise

(**A5**) ε_1 is sub-Gaussian variable with proxy variance $b > 0$, that is for every $t \in \mathbb{R}$, it holds

$$\mathbb{E}[\exp(t\varepsilon_1)] \leqslant \exp(\frac{b^2 t^2}{2}).$$

It is also said that ε_1 is b-sub-Gaussian or sub-Gaussian with parameter b. The natural example of a sub-Gaussian random variable is a centered Gaussian. If ε_1 has $\mathcal{N}(0, \sigma^2)$ distribution, it is easy to check $\mathbb{E}[\exp(t\varepsilon_1)] \leqslant \exp(\frac{\sigma^2 t^2}{2})$, then, ε_1 is sub-gaussian with parameter σ^2. Assumption (**A5**) is also satisfied if ε_1 is bounded.
The following non asymptotic result holds for $\widetilde{f}_{(\widehat{d})}$.

Theorem 3.7. *Let assumptions (**A0**) to (**A3**) and (**A5**) hold, $\widetilde{f}_{(d)}$ be defined by (31), $\widehat{d}$ selected by (34). Then, for $\kappa_1 \geqslant 12$, we have*

$$\mathbb{E}[\|\widetilde{f}_{(\widehat{d})} - f\|^2] \leqslant C \inf_{d \in \mathcal{M}_n^{(1)}} \left(\|f - f_{(\sqrt{2d})}\|^2 + R_b(d) + V(d) \right) + C' \frac{\log(n)}{n}, \tag{35}$$

*where $R_b(d) := \max_{d' \in \mathcal{M}_n^{(1)}, d \leqslant d'} \left(\Delta(\sqrt{2d'}) \|h - \mathbb{E}[\widehat{h}_{d'}]\|^2 \right)$ C is a numerical constant and $C' = C'(\mathbb{E}[\varepsilon_1^4], \gamma, c_1, \xi, \lambda, C'_\infty)$ with c'_1, γ given in (**A3**), ξ, C'_∞ in (7) and λ in (**A0**).*
In addition, if f belongs to $W^s(L)$ and h to $W_H^{s+\gamma}(L')$ with $s + \gamma \geqslant 17/6$, it holds

$$\mathbb{E}[\|\widetilde{f}_{(\widehat{d})} - f\|^2] \leqslant C_1 \inf_{d \in \mathcal{M}_n^{(1)}} \left(d^{-s} + V(d) \right) + C'_1 \frac{\log(n)}{n}, \tag{36}$$

where C_1 is a constant depending on C, L, L', s, γ and C'_1 depending on C', s and γ.

The term $R_b(d)$ has the same order as the classical bias of f ($\|f - f_{(\sqrt{2d})}\|^2$) under adequate regularity conditions on f and g. Inequalities (35) and (36) are non asymptotic. In the assumptions of regularity, the values of s (for f) and γ ($s + \gamma$ for h) need not to be known for implementing the procedure or computing the estimator. The two inequalities show that $\widetilde{f}_{(\widehat{d})}$ realizes automatically a bias-variance trade-off up to log term, and an additional residual term $C' \frac{\log(n)}{n}$, which is negligible in general. Moreover, we derive from Theorem 3.2 with n replaced by $n/\log(n)$ that under the assumptions of Theorem 3.7

$$\mathbb{E}[\|\widetilde{f}_{(\widehat{d})} - f\|^2] \leqslant C (\frac{n}{\log(n)})^{-\frac{s}{s+\gamma+1}},$$

where $C > 0$ is a numerical constant.

4. Hermite-Hermite strategy for the estimation of f

Our aim is to build a projection estimator of the unknown function f using the Hermite basis. The ideas is to decompose both functions f and h in the Hermite basis.

4.1. **Estimation strategy.** Let $(x_k, y(x_k))_{-n\leqslant k\leqslant n-1}$ from model (1), $m \geqslant 1$, integer and consider S_m defined in (11). Assuming that f belongs to $\mathbb{L}^2(\mathbb{R})$, we decompose f in the Hermite basis $(\varphi_j)_{j\geqslant 0}$: $f = \sum_{j=0}^{\infty} a_j(f)\varphi_j, a_j(f) = \langle f, \varphi_j\rangle = \int f(x)\varphi_j(x)dx$ and the orthogonal projection of f on S_m is given by: $f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j$. To estimate f, we build m estimators of the coefficients $a_j(f)$. Under (**A2**), using the Plancherel theorem and as $h = f \star g$, it follows that:

$$a_j(f) = \frac{1}{2\pi}\langle \frac{h^*}{g^*}, \varphi_j^*\rangle = \frac{1}{2\pi}\int \frac{h^*(u)}{g^*(u)}\overline{\varphi_j^*(u)}dx = \frac{(-i)^j}{\sqrt{2\pi}}\int \frac{h^*(u)}{g^*(u)}\varphi_j(u)du. \tag{37}$$

Replacing h^* by $\widehat{h}_d^*$ defined in (23) and plugging this in (37), we define the following estimator:

$$\widehat{f}_{m,d} = \sum_{j=0}^{m-1} \widehat{a}_{j,d}\varphi_j, \quad \widehat{a}_{j,d} = \frac{(-i)^j}{\sqrt{2\pi}}\int \frac{\widehat{h}_d^*(u)}{g^*(u)}\varphi_j(u)du, \tag{38}$$

provided that $\widehat{h}_d^*\varphi_j/g^*$ is integrable for $j = 0, \dots, m-1$. The coefficients $\widehat{a}_{j,d}$ are real. Indeed, using that $\varphi_j(x) = (-1)^j\varphi_j(-x)$ (since $H_j(-x) = (-1)^j H_j(x)$), we have

$$\overline{\widehat{a}_{j,d}} = \frac{(i)^j}{\sqrt{2\pi}}\int \overline{\frac{\widehat{h}_d^*(u)}{g^*(u)}\varphi_j(u)}du = \frac{(-i)^j}{\sqrt{2\pi}}\int \frac{\widehat{h}_d^*(u)}{g^*(u)}\varphi_j(u)du = \widehat{a}_{j,d},$$

where $\overline{z}$ is the complex conjugate of the complex number z. Under (**A3**), the integrability condition of the ratio $\widehat{h}_d^*\varphi_j/g^*$ is ensured (see Equation (7)). The two dimensions m and d must be optimized. As for $\widehat{f}_{(d)}$ or $\widehat{f}_{(\ell),d}$, the performance of $\widehat{f}_{m,d}$ depends on $\widehat{h}_d$ which has good statistical properties (see Section 2).

4.2. **Risk bound for the projection estimator of f.** The following risk bound holds for $\widehat{f}_{m,d}$.

Proposition 4.1. *Assume that f and h belong to $\mathbb{L}^2(\mathbb{R})$ and set*

$$\Sigma(m) = \sup_{|u|\leqslant\sqrt{\rho m}} |g^*(u)|^{-2} + \sum_{j=0}^{m-1}\int_{|u|\geqslant\sqrt{\rho m}} |\varphi_j(u)|^2|g^*(u)|^{-2}du, \quad \rho > 0. \tag{39}$$

For $\widehat{f}_{m,d}$ given in (38), we have

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] \leqslant \|f - f_m\|^2 + 2\Sigma(m)\left(\|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2 + \sigma_\varepsilon^2\frac{T}{n}\left(\mathrm{tr}\left(\Psi_d^{-1}\right) \wedge 2\pi^2 m\right)\right). \tag{40}$$

Note that the constant $\rho > 0$ is independent from n, m and d. The same comments given after Proposition 3.1 for the deconvolution estimator $\widehat{f}_{(\ell),d}$ hold here. The difference with $\widehat{f}_{(\ell),d}$ can be found on the bias of $\widehat{f}_{m,d}$ and the term $\Sigma(m)$, the regression part does not change. Moreover, the term $\sum_{j=0}^{m-1}\int_{|x|\geqslant\sqrt{\rho m}} |\varphi_j(x)|^2|g^*(x)|^{-2}dx$ is exponentially decaying in m for $\rho \geqslant 2$ (see Proposition 3.1 in Sacko (2020)) and thus negligible with respect of $\sup_{|x|\leqslant\sqrt{\rho m}}(|g^*(x)|^{-2}) = \Delta(\sqrt{\rho m})$, where $\Delta(\ell)$ is given in (26). Thus, for $f \in W_H^s(L)$ and choosing $\ell \asymp \sqrt{m}$, the estimator $\widehat{f}_{(\ell),d}$ and $\widehat{f}_{m,d}$ have the same order and then rate of convergence (see also Comte and Genon-Catalot (2018) and Sacko (2020) in the framework of density deconvolution).

4.3. **Rate of convergence of $\widehat{f}_{m,d}$.** As for $\widehat{f}_{(d)}$, we propose a two-step bias-variance trade-off.

Theorem 4.2. *Suppose that (A0), (A3) and h belongs to $W_H^{s+\gamma}(L)$. For $d_{opt} = m_{opt} = [n^{1/(s+\gamma+1)}]$ with $s+\gamma > 11/6$, we derive that*

$$\sup_{f\in W_H^s(L)} \mathbb{E}\left[\|\widehat{f}_{m_{opt},d_{opt}} - f\|^2\right] = \mathcal{O}\left(n^{-\frac{s}{s+\gamma+1}}\right),$$

where $W_H^s(L)$ is the classical Sobolev-Hermite ball defined in (9).

The estimator $\widehat{f}_{m_{opt},d_{opt}}$ achieves the same rate as $\widehat{f}_{(d_{opt})}$ obtained in Theorem 3.2. Note that the results for some special functions obtained for $\widehat{f}_{(d_{opt})}$ in Proposition 3.3, 3.4 and 3.5 apply here. If f is a Gamma function (see Proposition 3.6), we have a loss on the order of the bias of f, $\|f - f_m\|^2$ which is linked to the Hermite basis. Indeed, for $\ell^2 \asymp m$, $\widehat{f}_{m,d}$ and $\widehat{f}_{(\ell),d}$ have the same variance order but the bias is order: $\|f - f_m\|^2 \leqslant \ell^{-2p+4}$ contrary to the Fourier bias where $\|f - f_\ell\|^2 \asymp \ell^{-2p+1}$ where p is the shape parameter of Gamma function. For $\widehat{f}_{m,d}$, we get for $m_{opt} = d_{opt}^2 = [n^{\frac{1}{2(p+q-1)}}]$ the following rate of convergence

$$\mathbb{E}\left[\|\widehat{f}_{m_{opt},d_{opt}} - f\|^2\right] = \mathcal{O}\left(n^{-\frac{p-2}{p+q-1}}\right).$$

4.4. **Adaptive procedure for Hermite-Hermite approach.** As for the Fourier-Hermite method and in view of Theorem 4.2, we set $d = m$ and we consider the following estimator

$$\widetilde{f}_m = \widehat{f}_{m,m}, \tag{41}$$

where $\widehat{f}_{m,m}$ is given in (38) Now, we are interested in the choice of m. Let us define the collection of models $\mathcal{M}_n^{(2)}$ by

$$\mathcal{M}_n^{(2)} := \Big\{1 \leqslant m \leqslant n, \quad \frac{\sigma_\varepsilon^2 \lambda T m \Sigma(m)}{n} \leqslant 1\Big\},$$

where $\Sigma(m)$ is given by (39) and λ in (A0). Analogously to FH approach, we estimate the bias by

$$\widehat{B}(m) = \max_{m'\in\mathcal{M}_n^{(2)}} \left\{\left(\|\widetilde{f}_{m'} - \widetilde{f}_{m\wedge m'}\|^2 - \kappa_1' W(m')\right)_+\right\}, \quad W(m) = 4\left(1 + 24\log(n)\right)\sigma_\varepsilon^2 \Sigma(m)\frac{\lambda m T}{n}, \tag{42}$$

where $\kappa_1' > 0$ is a numerical constant which must be adjusted in practice. Then, we set

$$\widetilde{m} := \arg\min_{m\in\mathcal{M}_n^{(2)}} \left\{\widehat{B}(m) + \kappa_2' W(m)\right\}, \tag{43}$$

with $\kappa_2' \geqslant \kappa_1' > 0$ must be also calibrated in practice. We can prove the following oracle inequality

Theorem 4.3. *Let Assumptions (A0) to (A3) and (A5) hold, $\widetilde{f}_m$ be defined by (41), $\widetilde{m}$ selected by (43). Then, for $\kappa_1' \geqslant 12$, it yields*

$$\mathbb{E}[\|\widetilde{f}_{\widetilde{m}} - f\|^2] \leqslant C \inf_{m\in\mathcal{M}_n^{(2)}} \left(\|f - f_m\|^2 + R_b'(m) + W(m)\right) + C'\frac{\log(n)}{n}, \tag{44}$$

where $R_b'(m) = \max_{m'\in\mathcal{M}_n^{(2)}, m\leqslant m'} \Sigma(m')\|h - \mathbb{E}[\widehat{h}_{m'}]\|^2$, C is a numerical constant and $C' = C'(\mathbb{E}[\varepsilon_1^4], \gamma, c_1, \xi, \lambda, C_\infty')$ with c_1', γ given in (A3), ξ, C_∞' in (7) and λ in (A0).
In addition, if f belongs to $W_H^s(L)$ and h to $W_H^{s+\gamma}(L')$ with $s+\gamma \geqslant 17/6$, we derive

$$\mathbb{E}[\|\widetilde{f}_{\widetilde{m}} - f\|^2] \leqslant C_1 \inf_{m\in\mathcal{M}_n^{(2)}} \left(m^{-s} + W(m)\right) + C_1'\frac{\log(n)}{n}, \tag{45}$$

where C_1 is a constant depending on C, L, L', s, γ and C_1' depending on C', s and γ.

The same comments for HH strategy given after Theorem 3.7 hold for $\widetilde{f}_{\widetilde{m}}$. In particular, we deduce from Theorem 4.2 (with $n/\log(n)$ playing the role of n),

$$\mathbb{E}[\|\widetilde{f}_{\widetilde{m}} - f\|^2] \leqslant C(\frac{n}{\log(n)})^{-\frac{s}{s+\gamma+1}},$$

where C is an universal constant.

5. Numerical illustration

5.1. **Practical implementation.** In this section, we present the results of a simulation study to illustrate the performances of our strategies. We compute the estimator $\widetilde{f}_{(\widehat{d})}$ given in (31) with $\widehat{d}$ selected by (34) and $\widetilde{f}_{\widetilde{m}}$ defined in (41) with $\widetilde{m}$ chosen in (43). We consider the following test functions which are estimated on the interval I

(i) $f(x) = \exp(-2x^2)$, $I = [-2, 2]$,
(ii) Gamma distribution $\Gamma(4, 4)$, $I = [0, 2.5]$,
(iii) $f(x) = \frac{4}{\sqrt{2\pi}}(0.4\exp(-8(x+1)^2) + 0.6\exp(-8(x-1)^2))$, $I = [-2, 2]$,
(iv) $f(x) = -2x(1+x^2)^{-2}$, $I = [-2, 2]$.

For the kernel g, we choose a $\Gamma(2, \theta)$ distribution *i.e.* $g(x) = \theta^2 x\exp(-\theta x)1_{x\geqslant 0}$ with $\theta = 4$. The errors (ε_k) are centered Gaussian with standard deviation $\sigma_\varepsilon \in \{1/8, 1/4\}$. We also choose $T = 10$ and consider two sample sizes $n = 250, 1000$. The regression $h = f \star g$ is computed for each test function f and kernel g by Riemann sum discretization in 500 points. We consider the following collection of models $\mathcal{M}_n^{(2)} = \mathcal{M}_n^{(1)} = \{1, 2, \ldots, 25\}$. The Fourier transform of g is equal to $g^*(t) = (1 - i\frac{t}{\theta})^{-2}$ with $\theta = 4$ then, we consider the following variance term in practice for the FH method:

$$V(d) = 2\,(1 + 24\log(n))\,\sigma_\varepsilon^2(1 + \frac{2d}{\theta^2})^2\frac{\lambda dT}{n}, \quad \theta = 4, \quad \lambda = 1. \tag{46}$$

For the HH method, we take $W(m) = 2V(m)$. The adaptive procedure is implemented as follows:

– For each $d \in \mathcal{M}_n^{(1)}$, compute $\widehat{A}(d) = \max_{d'\in\mathcal{M}_n^{(1)}}\left\{\left(\|\widetilde{f}_{(d')} - \widetilde{f}_{(d\wedge d')}\|^2 - \kappa_1 V(d')\right)_+\right\}$, where the integral $\|\widetilde{f}_{(d')} - \widetilde{f}_{(d\wedge d')}\|^2$ is computed by Riemann's approximation and $V(d)$ given in (46),
– Select $\widehat{d}$ such that $\widehat{d} = \arg\min_{d\in\mathcal{M}_n^{(1)}}\left\{\widehat{A}(d) + \kappa_2 V(d)\right\}$,
– Compute $\widetilde{f}_{(\widehat{d})}(x) = \frac{1}{2\pi}\int_{-\sqrt{2\widehat{d}}}^{\sqrt{2\widehat{d}}} e^{-iux}\frac{\widehat{h}^*_{\widehat{d}}(u)}{g^*(u)}du$.

This description remains valid for the HH strategy by setting $\kappa_1' = 2\kappa_1$ and $\kappa_2' = 2\kappa_2$ where $2V(m)$ plays the role of $V(d)$. In the sequel, this procedure is called « **GLM** » (for the Goldenshluger and Lepski method).

Choice of constants κ_1 and κ_2. We can choose $\kappa_1 = \kappa_1$ and have just one constant to calibrate, it is in this kind that the procedure (Goldenshluger and Lepski) was developed. Recently, Lacour and Massart (2016) suggested the idea of considering two different constants ($\kappa_1 \neq \kappa_2$) and propose to take $\kappa_2 = 2\kappa_1$ for kernel density estimation using Goldenshluger and Lepski (2011) method. Here, we adopt the same idea to find the values of κ_1 and κ_2. In a rather "rough" way and after some numerical tests, we choose $\kappa_1 = 1.5 \times 10^{-3}$ and $\kappa_2 = 3 \times 10^{-3}$. Then, we illustrate the procedure by some graphs.

As **GLM** method is slow and therefore difficult to calibrate, we implement the penalization method, which allows us to perform repetitions and propose risk tables. The penalization strategy has the advantage to be faster. Furthermore, we must only calibrate one constant denoted $\kappa^{(1)}$. More precisely, the method (called « **PM** » for the penalization method) is described (only for the FH method) as follows

- For each $d \in \mathcal{M}_n^{(1)}$, compute $\widetilde{A}(d) = -\|\widetilde{f}_d\|^2$, by Riemann's approximation.
- Choose $\widetilde{d}$ via $\widetilde{d} = \arg\min_{d\in\mathcal{M}_n^{(1)}} \left\{ \widetilde{A}(d) + \kappa^{(1)} V(d) \right\}$.
- Compute $\widetilde{f}_{(\widetilde{d})}(x) = \frac{1}{2\pi} \int_{-\sqrt{2\widetilde{d}}}^{\sqrt{2\widetilde{d}}} e^{-iux} \frac{\widehat{h}_{\widetilde{d}}^*(u)}{g^*(u)} du$.

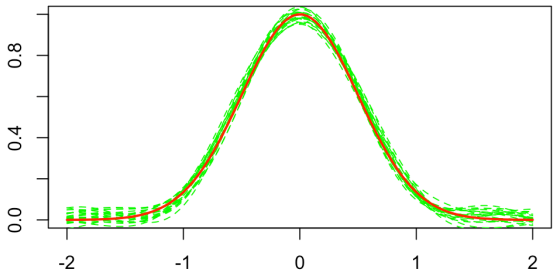
Calibration of constant $\kappa^{(1)}$. To find the value of $\kappa^{(1)}$, we have evaluated the MISE for different test functions and different proposals for $\kappa^{(1)}$. This preliminary study leads to fix $\kappa^{(1)} = 1.5 \times 10^{-3}$ for the FH procedure and 10^{-3} for the HH method.

5.2. **Numerical simulation results.** First, we illustrate the methods by presenting some pictures. Figure 1 presents the true unknown function (the bold red line), and twenty estimators chosen by the **GLM** procedure in green dotted lines, for each test function (i), (ii), (iii) and (iv). The dimension selected by the procedure and the value of Signal-to-noise ratio $s2n$ are given under each graph. Note that $s2n$ is defined here by:

$$s2n := \frac{\frac{1}{2n}\sum_{i=-n}^{n-1} h(x_i)^2}{\frac{1}{2n}\sum_{i=-n}^{n-1} \varepsilon_i^2} \approx \frac{\frac{1}{2n}\sum_{i=-n}^{n-1} y(x_i)^2 - \sigma_\varepsilon^2}{\sigma_\varepsilon^2},$$

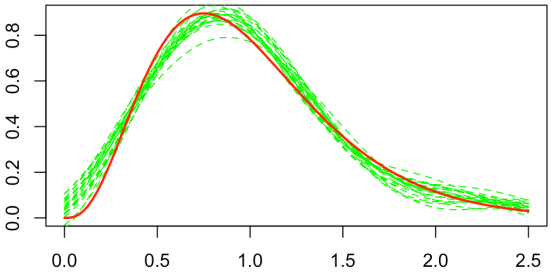
where the above approximation is obtained using the law of large numbers. We observe that the **GLM** procedure give very satisfactory results, visually.
In Figures 2 to 5, we plot the true function in bold red line with 20 estimators in dotted lines for the test functions (iii), (iv) by considering the **PM** algorithm for the two estimation procedures. The first line illustrates the influence of sample size and the second line shows how the noise level can affect the performance of the estimates. We observe that increasing n improves the estimation and, on the contrary, that increasing the noise makes the problem more difficult. We can also see some oscillations when $\sigma_\varepsilon = 1/4$ which corresponds to a $s2n$ ratio less than 1 (see Table 3), this effect decreases when the sample size increases. The mean of selected dimensions are given in Table 2. We observe that these averages are comparable to the dimension obtained in Figure 1 for (iii) and (iv) with **GLM** algorithm.
In Table 3, we report the values of the MISEs with standard deviation in parentheses multiplied by 100 computed from 100 simulated samples for the estimator $\widetilde{f}_{(\widetilde{d})}$ and $\widetilde{f}_{\widetilde{m}}$ with $\widetilde{d}$ and $\widetilde{m}$ selected using the **PM** algorithm. We also provide the average of $\widetilde{d}$ or $\widetilde{m}$ selected by each procedure. As for graphical study, we see that increasing the sample size or decreasing the variance of noise (which corresponds to a larger $s2n$, see Table 3) improves the estimation. When n increases, the average of $\widetilde{d}$ or $\widetilde{m}$ is increasing except in the case of function (i) with $\sigma_\varepsilon = 1/4$. This case corresponds to a $s2n$ equal to 0.58 see Table 3 and then the estimation is most difficult; this can explain why the procedure chooses a large dimension for $n = 250$. Clearly, the influence of signal-to-noise ratio $s2n$ is important, see Figures for graphical analysis.
Comparisons between FH and HH procedures. We observe that the two estimation methods seem to be equivalent. The computing time is the main difference between the two procedures. For example: we need about 25 minutes to obtain the MISEs for $n = 250$ and almost 1 hour for $n = 1000$ for the FH method, while the HH procedure takes only about 4 minutes for $n = 250$ and less than 10 minutes for $n = 1000$. This difference in computation time is probably related to the fact that the Hermite basis allows to build low complexity estimators, see Belomestny et al. (2019).

(i)



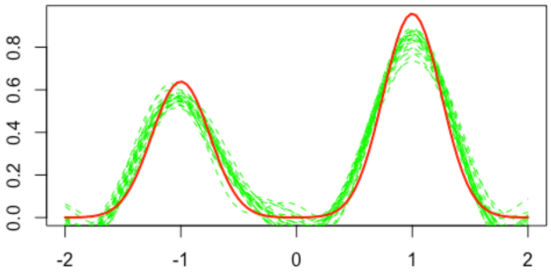
$\overline{\widehat{d}} = 11.20$, $\overline{s2n} = 2.38$, $s \longrightarrow \infty$

(ii)



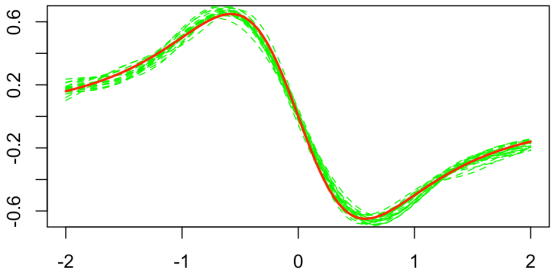
$\overline{\widehat{d}} = 13.85$, $\overline{s2n} = 1.60$, $s = 3$

(iii)



$\overline{\widehat{d}} = 20.05$, $\overline{s2n} = 1.18$, $s \longrightarrow \infty$

(iv)



$\overline{\widehat{d}} = 12.65$, $\overline{s2n} = 1.87$, $s \longrightarrow \infty$

FIGURE 1. 20 estimates of $\widetilde{f}_{(\widehat{d})}$ for the **GLM** algorithm. The true function is in bold red and the estimate in green dotted lines for $n = 1000$.

	$\sigma_\varepsilon = \frac{1}{8}$				$\sigma_\varepsilon = \frac{1}{4}$			
n	250		1000		250		1000	
f	FH	HH	FH	HH	FH	HH	FH	HH
(iii)	13.30	13.75	22.20	21.45	14.05	11.75	17.20	13.05
(iv)	12.65	10.25	12.95	11.40	10.05	9.65	12.90	10.45

TABLE 2. Mean of selected dimensions $\widetilde{d}$ or $\widetilde{m}$ presented in Figures 2 to 5.

6. PROOFS

6.1. **Proofs of Section 2.**

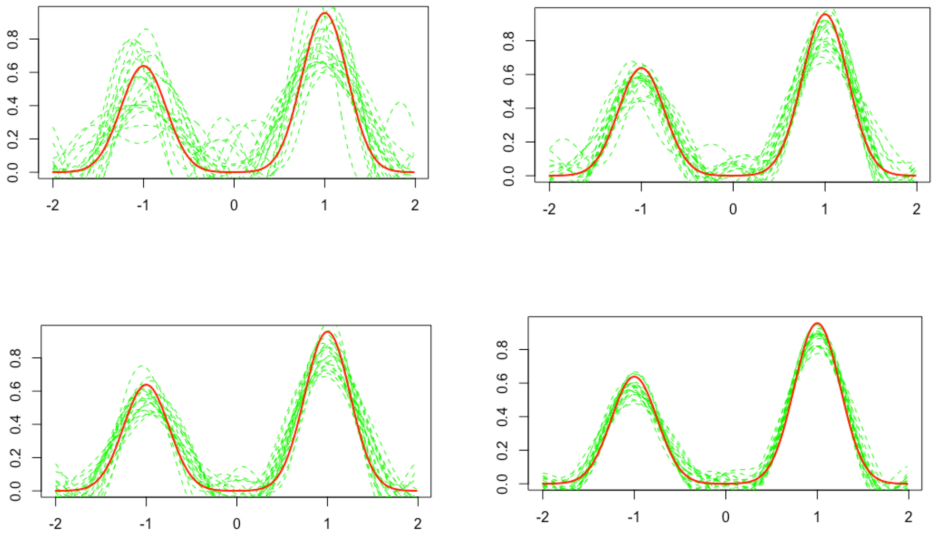


FIGURE 2. 20 estimates of (iii) with FH method, and $n = 250$ (first line) or $n = 1000$ (second line) using the **PM** algorithm. The true function is in bold red and the estimates in green dotted lines (left $\sigma_\varepsilon = 1/4$, right $\sigma_\varepsilon = 1/8$).

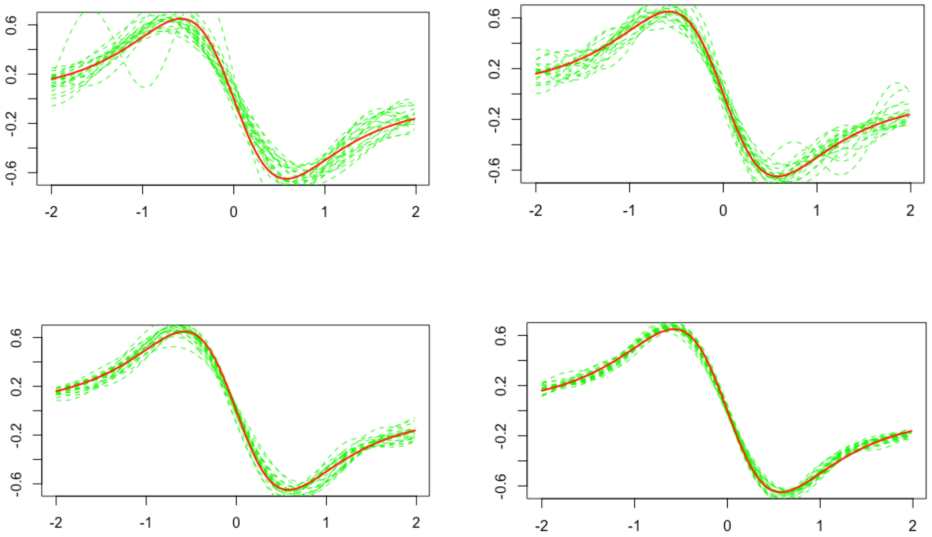


FIGURE 3. 20 estimates of (iv) with FH method, and with $n = 250$ (first line) or $n = 1000$ (second line) using the **PM** algorithm. The true function is in bold red and the estimates in green dotted lines (left $\sigma_\varepsilon = 1/4$, right $\sigma_\varepsilon = 1/8$).

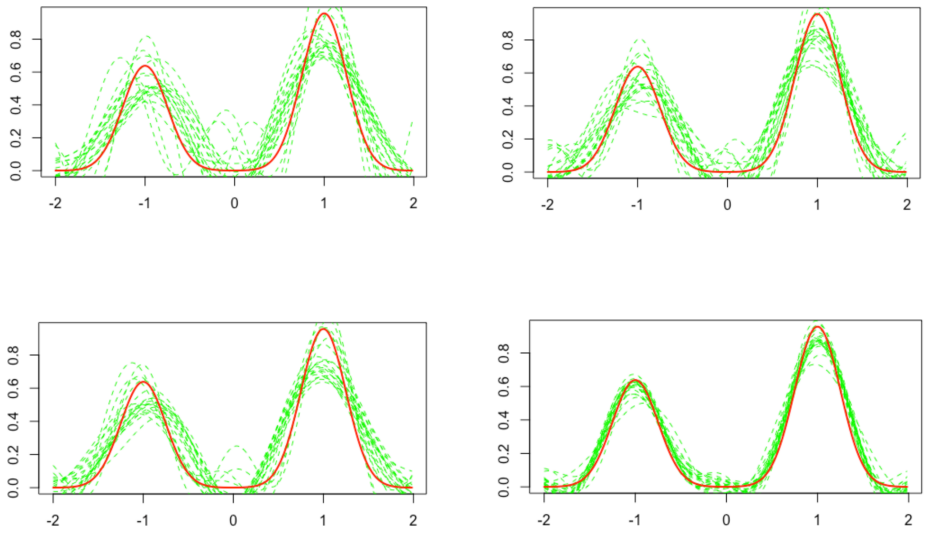


FIGURE 4. 20 estimates of (iii) with HH method, and $n = 250$ (first line) or $n = 1000$ (second line) using the **PM** algorithm. The true function is in bold red and the estimates in green dotted lines (left $\sigma_\varepsilon = 1/4$, right $\sigma_\varepsilon = 1/8$).

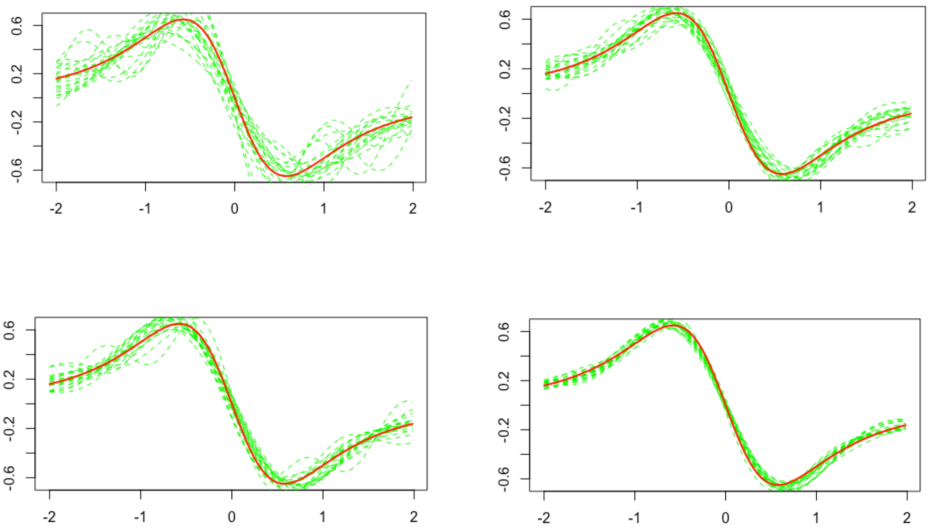


FIGURE 5. 20 estimates of (iv) with HH method, and $n = 250$ (first line) or $n = 1000$ (second line) using the **PM** algorithm. The true function is in bold red and the estimates in green dotted lines (left $\sigma_\varepsilon = 1/4$, right $\sigma_\varepsilon = 1/8$).

Proof of Lemma 2.1. Let $\vec{w} = (w_0, \dots, w_{d-1})^t$, with $\Psi_d \vec{w} = 0$. Then, it holds

$$\frac{T}{n} \vec{w}^T \Phi_d^T \Phi_d \vec{w} = \frac{T}{n} \|\Phi_d \vec{w}\|_2^2 = \frac{T}{n} \sum_{i=-n}^{n} \left(\sum_{j=0}^{d-1} w_j \varphi_j(x_i) \right)^2 = 0.$$

	$\sigma_\varepsilon = \frac{1}{8}$				$\sigma_\varepsilon = \frac{1}{4}$			
f \ n	250		1000		250		1000	
	FH	HH	FH	HH	FH	HH	FH	HH
(i)	$1.28_{(0.99)}$ 11.75 2.37	$\mathbf{1.24_{(0.67)}}$ 9.53 2.36	$\mathbf{0.35_{(0.26)}}$ 11.35 2.36	$0.43_{(0.17)}$ 10.83 2.36	$4.66_{(4.33)}$ 10.10 0.58	$\mathbf{3.11_{(2.05)}}$ 7.51 0.58	$\mathbf{1.12_{(0.88)}}$ 9.95 0.58	$1.16_{(0.71)}$ 8.65 0.59
(ii)	$\mathbf{1.40_{(0.82)}}$ 13.23 1.58	$1.51_{(0.85)}$ 11.87 1.57	$\mathbf{0.61_{(0.28)}}$ 15.31 1.57	$0.65_{(0.25)}$ 14.66 1.57	$4.48_{(3.04)}$ 11.94 0.37	$\mathbf{3.11_{(1.90)}}$ 9.07 0.39	$\mathbf{0.46_{(0.22)}}$ 12.21 0.39	$1.47_{(0.63)}$ 10.67 0.39
(iii)	$\mathbf{3.95_{(1.61)}}$ 18.96 1.19	$4.96_{(1.83)}$ 15.39 1.18	$\mathbf{1.36_{(0.71)}}$ 22.33 1.17	$1.52_{(1.09)}$ 21.93 1.18	$9.85_{(3.95)}$ 13.42 0.28	$\mathbf{9.65_{(3.92)}}$ 11.20 0.30	$\mathbf{4.27_{(1.61)}}$ 16.59 0.29	$5.41_{(1.45)}$ 13.60 0.30
(vi)	$1.60_{(1.05)}$ 12.71 1.88	$\mathbf{1.38_{(0.83)}}$ 10.41 1.89	$0.49_{(0.27)}$ 13.11 1.90	$\mathbf{0.45_{(0.24)}}$ 12.74 1.90	$5.10_{(4.12)}$ 10.45 0.48	$\mathbf{4.53_{(3.69)}}$ 8.87 0.48	$1.46_{(0.78)}$ 10.65 0.48	$\mathbf{1.41_{(0.58)}}$ 9.39 0.47

TABLE 3. First line: empirical 100× MISE (with 100×sd) for the estimation of unknown function f computed over 100 independent simulations; second line: mean of $\widetilde{d}$ in the for $\widetilde{f}_{(\widetilde{d})}$ or mean of $\widetilde{m}$ for $\widetilde{f}_{\widetilde{m}}$, selected by the **PM** algorithm; third line: mean of Signal/Noise ratio.

Therefore, for all $-n \leqslant i \leqslant n-1$, we have, $\sum_{j=0}^{d-1} w_j \varphi_j(x_i) = 0$. As $\varphi_j(x) = c_j H_j(x) e^{-x^2/2}$, we derive $P_d(x_i) := \sum_{j=0}^{d-1} w_j c_j H_j(x_i) = 0$ *i.e.*, P_d is a polynomial of degree $d-1$ admitting $n > d$ distinct roots. Consequently, it follows $P_d \equiv 0$ and thus $\vec{w} \equiv \vec{0}$. □

Proof of Proposition 2.1. Denote $\Pi_d h = \Phi_d \vec{b}^{(d)} = \Phi_d(\Phi_d^t \Phi_d)^{-1} \Phi_d^t h(\vec{x})$ with $h(\vec{x}) = (h(x_{-n}), \ldots, h(x_{n-1}))^t$ the orthogonal projection of h on S_d for the empirical norm $\|\cdot\|_n^2$.

Proof of part (i). We have

$$\|\widehat{h}_d - h\|_n^2 = \|\Pi_d h - h\|_n^2 + \|\widehat{h}_d - \Pi_d h\|_n^2 = \inf_{t \in S_d} \|t - h\|_n^2 + \|\widehat{h}_d - \Pi_d h\|_n^2.$$

Taking the expectation gives

$$\mathbb{E}\left[\|\widehat{h}_d - h\|_n^2\right] = \inf_{t \in S_d} \|t - h\|_n^2 + \mathbb{E}\left[\|\widehat{h}_d - \Pi_d h\|_n^2\right]. \tag{47}$$

Then, for $\widehat{\vec{b}}^{(d)}$ given in (13), we can write $\widehat{h}_d(\vec{x}) = \left(\widehat{h}_d(x_{-n}), \ldots, \widehat{h}_d(x_{n-1})\right)^t = \Phi_d \widehat{\vec{b}}^{(d)}$ and $\Pi_d h = \mathbb{E}[\widehat{h}_d(\vec{x})]$. Setting $P(\vec{x}) = \Phi_d(\Phi_d^t \Phi_d)^{-1} \Phi_d^t$, we have

$$\|\widehat{h}_d - \Pi_d h\|_n^2 = \|P(\vec{x})\vec{\varepsilon}\|_n^2 = \frac{T}{n} \vec{\varepsilon}^t P(\vec{x})^t P(\vec{x}) \vec{\varepsilon} = \frac{T}{n} \vec{\varepsilon}^t P(\vec{x}) \vec{\varepsilon}.$$

Moreover, it yields

$$\mathbb{E}[\vec{\varepsilon}^t P(\vec{x})\vec{\varepsilon}] = \mathbb{E}\Big[\sum_{-n\leqslant i,k\leqslant n-1} \varepsilon_i\varepsilon_k[P(\vec{x})_{i,k}]\Big] = \sigma_\varepsilon^2 \sum_{i=-n}^{n-1} \mathbb{E}[P(\vec{x})_{i,i}] = \sigma_\varepsilon^2 \mathrm{tr}(P(\vec{x})) = \sigma_\varepsilon^2 \mathrm{tr}(I_d) = \sigma_\varepsilon^2 d.$$

Consequently, it holds $\mathbb{E}\left[\|\widehat{h}_d - \Pi_d h\|_n^2\right] = \sigma_\varepsilon^2 T\frac{d}{n}$. Plugging this in (47) ends the proof of (14).

Proof of part (ii). By Phythagoras Theorem, we have

$$\begin{aligned}\mathbb{E}[\|\widehat{h}_d - h\|^2] &= \mathbb{E}[\|\widehat{h}_d - h_d\|^2] + \|h - h_d\|^2 \\ &= \mathbb{E}[\|\widehat{h}_d - \mathbb{E}[\widehat{h}_d]\|^2] + \|\mathbb{E}[\widehat{h}_d] - h_d\|^2 + \|h - h_d\|^2.\end{aligned}$$

We study the two first terms in the right hand side of the previous equality. For the first term, using the definition of $\widehat{h}_d$ given in (23), we get

$$\mathbb{E}\left[\|\widehat{h}_d - \mathbb{E}[\widehat{h}_d]\|^2\right] = \mathbb{E}\|\vec{\widehat{b}}^{(d)} - \mathbb{E}\vec{\widehat{b}}^{(d)}\|_{\mathbb{R}^d}^2 = 2\pi\mathbb{E}\left[(\vec{\widehat{b}}^{(d)} - \mathbb{E}\vec{\widehat{b}}^{(d)})^t(\vec{\widehat{b}}^{(d)} - \mathbb{E}\vec{\widehat{b}}^{(d)})\right].$$

Note that $\vec{\widehat{b}}^{(d)} - \mathbb{E}\vec{\widehat{b}}^{(d)} = (\Phi_d^t\Phi_d)^{-1}\Phi_d^t\vec{\varepsilon}$, this implies

$$\mathbb{E}\left[\|\widehat{h}_d - \mathbb{E}[\widehat{h}_d]\|^2\right] = \mathbb{E}\left[\vec{\varepsilon}^t\Phi_d(\Phi_d^t\Phi_d)^{-1}(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\vec{\varepsilon}\right] = \mathbb{E}\left[\vec{\varepsilon}^t M(\vec{x})\vec{\varepsilon}\right],$$

where $M(\vec{x}) = \Phi_d(\Phi_d^t\Phi_d)^{-1}(\Phi_d^t\Phi_d)^{-1}\Phi_d^t$. As ε_i are i.i.d. of variance σ_ε^2, it holds

$$\begin{aligned}\mathbb{E}\big[\vec{\varepsilon}^t M(\vec{x})\vec{\varepsilon}\big] &= \mathbb{E}\Big[\sum_{-n\leqslant i,k\leqslant n-1} \varepsilon_i\varepsilon_k[M(\vec{x})_{i,k}]\Big] = \sigma_\varepsilon^2 \sum_{i=-n}^{n-1} \mathbb{E}[M(\vec{x})_{i,i}] \\ &= \sigma_\varepsilon^2 \mathrm{tr}(M(\vec{x})) = \sigma_\varepsilon^2 \mathrm{tr}((\Phi_d^t\Phi_d)^{-1}).\end{aligned}$$

We derive that $\mathbb{E}\left[\vec{\varepsilon}^t M(\vec{x})\vec{\varepsilon}\right] = \sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right)$ and

$$\mathbb{E}\left[\|\widehat{h}_d - \mathbb{E}[\widehat{h}_d]\|^2\right] = \sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right). \tag{48}$$

For the other term, we have

$$\|h_d - \mathbb{E}[\widehat{h}_d]\|^2 = \left\|(\langle h,\varphi_0\rangle,\dots,\langle h,\varphi_{d-1}\rangle)^t - (\Phi_d^t\Phi_d)^{-1}\Phi_d^t\left(h(x_{-n}),\dots,h(x_n)\right)^t\right\|_{\mathbb{R}^d}^2.$$

Now, we remark that

$$\left(h_d(x_{-n}),\dots,h_d(x_{n-1})\right)^t = \sum_{k=0}^{d-1}\langle h,\varphi_k\rangle\left(\varphi_k(x_{-n}),\dots,\varphi_k(x_{n-1})\right)^t = \Phi_d\left(\langle h,\varphi_0\rangle,\dots,\langle h,\varphi_{d-1}\rangle\right)^t$$

and therefore,

$$(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\left(h_d(x_{-n}),\dots,h_d(x_{n-1})\right)^t = \left(\langle h,\varphi_0\rangle,\dots,\langle h,\varphi_{d-1}\rangle\right)^t.$$

Thus, it follows

$$\|h_d - \mathbb{E}[\widehat{h}_d]\|^2 = \|(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\left(h_d(\vec{x}) - h(\vec{x})\right)\|_{\mathbb{R}^d}^2 \leqslant \|(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\|_{op}^2\sum_{i=-n}^{n-1}(h_d(x_i) - h(x_i))^2,$$

where $\|A\|_{op}^2$ is the operator norm of the matrix A defined as the square root of the largest eigenvalue of A^tA. Then, it yields

$$\|(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\|_{op}^2 = \lambda_{max}(\Phi_d(\Phi_d^t\Phi_d)^{-1}(\Phi_d^t\Phi_d)^{-1}\Phi_d^t) = \frac{T}{n}\lambda_{max}\left(\Psi_d^{-1}\right) \tag{49}$$

This implies

$$\|h_d - \mathbb{E}[\widehat{h}_d]\|^2 \leqslant \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2, \tag{50}$$

From (48) and (6.1), we derive

$$\mathbb{E}[\|\widehat{h}_d - h\|^2] \leqslant \sigma_\varepsilon^2 \frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right) + \|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2. \tag{51}$$

□

Proof of Proposition 2.2. For $h \in W_H^\alpha(L)$ with $\alpha > 11/6$, we have from Proposition 2.1 (i) and Lemma 2.2 that

$$\mathbb{E}\left[\|\widehat{h}_d - h\|_n^2\right] \leqslant \|h_d - h\|^2 + (\sigma_\varepsilon^2 T + C(\alpha, L)T^2)\frac{d}{n} \leqslant Ld^{-\alpha} + (\sigma_\varepsilon^2 + C(\alpha, L)T^2)\frac{d}{n},$$

where $C(\alpha, L) > 0$ depends on α and L. The choice $d = d_{opt} = [n^{1/(\alpha+1)}]$ yields

$$\mathbb{E}\left[\|\widehat{h}_{d_{opt}} - h\|_n^2\right] = \mathcal{O}(n^{-\frac{\alpha}{\alpha+1}}).$$

Hence the part (i) of Proposition 2.2. The part (ii) is similar considering **(A0)**. □

Proof of Theorem 2.3. Inequality (19) follows from Corollary 3.1 in Baraud (2000), where all terms are multiplied by T with $q = 1$ and $p = 8$. The constant C' is given by:

$$C' = C''(\kappa)\frac{\mathbb{E}[\varepsilon_1^8]}{\sigma_\varepsilon^6}\left(1 + \sum_{d \in \mathcal{M}_n} d^{-2}\right) < +\infty.$$

Let us now prove (20). We recall that (see Equation (17) in Baraud (2000))

$$\forall d \in \mathbb{N} \quad , \sup_{t \in S_d, t \neq 0} \frac{\|t\|}{\|t\|_n} = \lambda_{max}(\Psi_d^{-1}). \tag{52}$$

Using that

$$\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|^2] \leqslant 2\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h_d\|^2] + 2\|h_d - h\|^2$$

Under **(A0)** and as $\widehat{h}_{\widehat{d}} - h_d \in S_{d_n}$, where $d_n \leqslant n$ is the maximum dimension of the collections of models $\mathcal{M}_n$, it holds from (52) that $\|\widehat{h}_{\widehat{d}} - h_d\|^2 \leqslant 2\lambda^2\|\widehat{h}_{\widehat{d}} - h\|_n^2 + 2\lambda^2\|h - h_d\|_n^2$. Thus, for any $d \geqslant 1$,

$$\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|^2] \leqslant 2\lambda^2\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|_n^2] + 2\lambda^2\|h - h_d\|_n^2 + \|h_d - h\|^2.$$

From (19), we derive that

$$\begin{aligned}\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|^2] \leqslant & 2\lambda^2\left[C(\kappa)\inf_{d \in \mathcal{M}_n}\left(\inf_{t \in S_d}\|t - h\|_n^2 + \sigma_\varepsilon^2 T\frac{d}{n}\right) + \frac{C'}{n}\right] + 2\lambda^2\|h - h_d\|_n^2 + \|h_d - h\|^2 \\ \leqslant & \max(1, 2\lambda^2 C(\kappa))\inf_{d \in \mathcal{M}_n}\left((2\lambda^2 + 1)\|h - h_d\|_n^2 + \|h_d - h\|^2 + \sigma_\varepsilon^2 T\frac{d}{n}\right) + 2\lambda^2\frac{C'T}{n}.\end{aligned}$$

This gives (20) and ends the proof of Theorem 2.3. □

6.2. **Proofs of Section 3.**

Proof of Proposition 3.1.

Proof of Equation (27). We have

$$\mathbb{E}[\|\widehat{f}_{(d)}-f\|^2]\leqslant 2\,\mathbb{E}[\|\widehat{f}_{(d)}-\widehat{f}_{(\ell),d}\|^2]+2\,\mathbb{E}[\|\widehat{f}_{(\ell),d}-f\|^2].$$

We examine the first term. Using successively the Cauchy-Schwarz inequality, (7) and under (**A3**), we deduce that

$$\begin{aligned}\|\widehat{f}_{(d)}-\widehat{f}_{(\ell),d}\|^2 =&\frac{1}{2\pi}\int_{|u|>\ell}\frac{|\widehat{h}_d^*(u)|^2}{|g^*(u)|^2}du=\frac{1}{2\pi}\int_{|u|>\ell}\frac{\left|\widehat{h}_d^*(u)-\mathbb{E}\,\widehat{h}_d^*(u)+\mathbb{E}\,\widehat{h}_d^*(u)\right|^2}{|g^*(u)|^2}du\\
&=\int_{|u|>\ell}\frac{\left|\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)}-\mathbb{E}\,\widehat{b}_j^{(d)}+\mathbb{E}\,\widehat{b}_j^{(d)}\right)\varphi_j(u)\right|^2}{|g^*(u)|^2}du\\
&\leqslant\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)}-\mathbb{E}\,\widehat{b}_j^{(d)}+\mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\sum_{j=0}^{d-1}\int_{|u|>\ell}\frac{\varphi_j(u)^2}{|g^*(u)|^2}du\\
&\leqslant c_1C_\infty'^2\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)}-\mathbb{E}\,\widehat{b}_j^{(d)}+\mathbb{E}\,\widehat{b}_j^{(d)}\right)^2de^{-\frac{\xi\ell^2}{2}}\int e^{-\frac{\xi u^2}{2}}(1+u^2)^\gamma du.\end{aligned}$$

As $\int e^{-\frac{\xi u^2}{2}}(1+u^2)^\gamma du\leqslant c_1'<\infty$ with $c_1'=c_1'(\gamma,\xi)$ and $\ell\geqslant\sqrt{2d}$, then, it follows that

$$\mathbb{E}\left[\|\widehat{f}_{(d)}-\widehat{f}_{(\ell),d}\|^2\right]\leqslant c_1c_1'C_\infty^2\,\mathbb{E}\left[\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)}-\mathbb{E}\,\widehat{b}_j^{(d)}+\mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\right]de^{-\xi d}.\tag{53}$$

By the definition $\widehat{h}_d$ given in (13), it yields $\mathbb{E}\left[\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)}-\mathbb{E}\,\widehat{b}_j^{(d)}+\mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\right]=\mathbb{E}\left[\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)}-\mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\right]+\sum_{j=0}^{d-1}\left(\mathbb{E}\,\widehat{b}_j^{(d)}\right)^2$ and $\|\,\mathbb{E}\,\widehat{h}_d\|^2=\|\,\mathbb{E}[\vec{\widehat{b}}^{(d)}]\|_{\mathbb{R}^d}^2=\|(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\vec{h}\|_{\mathbb{R}^d}^2\leqslant\|(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\|_{op}^2\sum_{i=-n}^{n-1}(h(x_i))^2$. Using (48) and (49) (where $h_d:=0$), we derive that

$$\mathbb{E}\left[\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)}-\mathbb{E}\,\widehat{b}_j^{(d)}+\mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\right]=\mathbb{E}[\|\widehat{h}_d-\mathbb{E}\,\widehat{h}_d\|^2]+\|\,\mathbb{E}\,\widehat{h}_d\|^2\leqslant\sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}(\Psi_d^{-1})+\lambda_{\max}\left(\Psi_d^{-1}\right)\|h\|_n^2.$$

Under (**A0**) and (**A4**), we have $\sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}(\Psi_d^{-1})+\lambda_{\max}\left(\Psi_d^{-1}\right)\|h\|_n^2\leqslant\max(\sigma_\varepsilon^2,2\|h\|_\infty^2)\lambda T$. It comes that $\mathbb{E}\left[\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)}-\mathbb{E}\,\widehat{b}_j^{(d)}+\mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\right]\leqslant\max(\sigma_\varepsilon^2,2\|h\|_\infty^2)\lambda T$. Injecting this in (53), we obtain

$$\mathbb{E}[\|\widehat{f}_{(d)}-\widehat{f}_{(\ell),d}\|^2]\leqslant c_1c_1'C_\infty'^2\max(\sigma_\varepsilon^2,2\|h\|_\infty^2)de^{-\xi d}=C\lambda Te^{-\frac{\xi d}{2}},$$

where $C=C(C_\infty',c_1,\|h\|_\infty,\xi)$ and therefore that

$$\mathbb{E}[\|\widehat{f}_{(d)}-\widehat{f}_{(\ell),d}\|^2]\leqslant C\lambda Te^{-\xi d}+2\,\mathbb{E}[\|\widehat{f}_{(\ell),d}-f\|^2].$$

Proof of Equation (28). For all $\ell>0$, $d\geqslant1$, we have the following decomposition:

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d}-f\|^2\right]=\|f-f_{(\ell)}\|^2+\mathbb{E}\left[\|f_{(\ell)}-\widehat{f}_{(\ell),d}\|^2\right].\tag{54}$$

We evaluate $\mathbb{E}\left[\|f_{(\ell)} - \widehat{f}_{(\ell),d}\|^2\right]$ using the Plancherel formula :

$$\mathbb{E}[\|\widehat{f}_{(\ell),d} - f_{(\ell)}\|^2] = \frac{1}{2\pi}\mathbb{E}\left[\int_{-\ell}^{\ell}\left|\frac{\widehat{h}_d^*(u) - h^*(u)}{g^*(u)}\right|^2 du\right] \leqslant \Delta(\ell)\mathbb{E}\left[\|\widehat{h}_d - h\|^2\right].$$

Plugging successively (15) in the above bound and in (54) gives (28). □

Proof of Theorem 3.2. Under **(A3)**, **(A0)** and for h belongs to W_H^α with $\alpha = s + \gamma$, we get from Lemma 2.2:

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right] \leqslant L\ell^{-2s} + (1+\ell^2)^\gamma\left[\sigma_\varepsilon^2\lambda\frac{dT}{n} + (1+\lambda)L'd^{-\alpha} + C(\alpha, L)\frac{T^2}{n}\right].$$

The choices $d_{opt} = [n^{1/(\alpha+1)}]$ and $\ell_{opt} = n^{\frac{1}{2(\alpha+1)}}$ end the proof. □

Proof of Proposition 3.3. As f is Gaussian, then it belongs to $W_H^\alpha(D)$ (see (9)) with α as large as desired, since f is infinitely differentiable and $f, \dots, f^{(\alpha)}, x^{\alpha-l}f^{(l)}$ for $l = 0 \dots \alpha - 1$, see Section 2.3. Using the differentiation under the integral sign theorem, we have that $h = f \star g$ is also infinitely differentiable for $g \in \mathbb{L}^1(\mathbb{R})$ and we write $h^{(l)} = f^{(l)} \star g$. Besides, it yields $\|h^{(l)}\| \leqslant \|f^{(l)}\|\|g\|_1$. Then, h belongs to $W^\alpha(\cdot)$ (Sobolev ball) since these derivative up to order α belong to $\mathbb{L}^2(\mathbb{R})$. Thus, $h \in W_H^\alpha(\cdot)$ if the function $x^{\alpha-l}h^{(l)}$ is square integrable. This is equivalent to prove that $x^\alpha h^{(l)}$ is square integrable. Now, we write

$$\|x^\alpha h^{(l)}\|^2 = 2\pi\|\left(x^{(\alpha)}h^{(l)}\right)^*\|^2 = 2\pi\|[(h^{(l)})^*]^{(\alpha)}\|^2 = 2\pi\|[g^*(f^{(l)})^*]^{(\alpha)}\|^2.$$

As $x^\alpha g \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ and $x^\alpha f^{(l)} \in \mathbb{L}^1(\mathbb{R})$, we get by the Leibniz Formula and the Cauchy-Schwarz inequality that:

$$\begin{aligned}
\|x^\alpha h^{(l)}\|^2 =& 2\pi\|\sum_{k=0}^{\alpha}\binom{\alpha}{k}(g^*)^{(k)}[(f^{(l)})^*]^{(\alpha-k)}\|^2 \\
=& 2\pi\int\left|\sum_{k=0}^{\alpha}\binom{\alpha}{k}(g^*)^{(k)}(u)[(f^{(l)})^*]^{(\alpha-k)}(u)\right|^2 du \\
\leqslant& C(\alpha)\max_{0\leqslant k\leqslant\alpha-l}\|[(f^{(l)})^*]^{(\alpha-k)}\|_\infty^2\sum_{k=0}^{\alpha}\binom{\alpha}{k}\int|(g^*)^{(k)}(u)|^2du.
\end{aligned}$$

Moreover, it holds $\int|(g^*)^{(k)}(u)|^2du = \frac{1}{2\pi}\int|x^k g(x)|^2dx \leqslant \int_{|x|\leqslant 1}|g(x)|^2dx + \int_{|x|\geqslant 1}|x^{(\alpha)}g(x)|^2dx < +\infty$. Therefore, $\|x^\alpha h^{(l)}\|^2 < +\infty$ and h belongs to $W_H^\alpha(L)$. Proposition 2.2 *(ii)* gives $\mathbb{E}[\|\widehat{h}_{d_{opt}} - h\|^2] \lesssim n^{-\frac{\alpha}{\alpha+1}}$. Plugging this in (28) and using Lemma 2 in Comte and Lacour (2011) yield

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d_{opt}} - f\|^2\right] \lesssim \ell^{-1}e^{-\sigma\ell^2} + c_1(1+\ell^2)^\gamma n^{-\frac{\alpha}{\alpha+1}}.$$

Replacing ℓ^2 by $\ell_{opt}^2 = \frac{\alpha}{(\alpha+1)\sigma^2}\log(n)$ ends the proof. □

Proof of Proposition 3.4. The proof is similar to that of Proposition 3.3. The regression part does not change *i.e.* for the choice $d_{opt} = [n^{1/(\alpha+1)}]$, we have always that $\mathbb{E}[\|\widehat{h}_{d_{opt}} - h\|^2] \lesssim n^{-\frac{\alpha}{\alpha+1}}$. (see the Proof of Proposition 3.3) with α as large as desired. But for the deconvolution part, the rate change since the order of the bias of f and $\Delta(\ell)$ have changed. Now, these order are: $\Delta(\ell) = \sup_{|u|\leqslant\ell}|g^*(u)|^{-2} \leqslant e^{\sigma^2\ell^2}$

because $g^*(u) = \exp(-\frac{\sigma^2 t^2}{2})$ and $\|f - f_{(\ell)}\|^2 = \frac{1}{2\pi}\int_{|u|>\ell} |f^*(u)|^2 du \leqslant \ell^{-2s}$ for $f \in W^s(L)$ (see (10)). From the previous results, we derive from (28)

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d_{opt}} - f\|^2\right] \lesssim \ell^{-2s} + e^{\sigma^2\ell^2} n^{-\frac{\alpha}{\alpha+1}}.$$

Choosing $\ell^2_{opt} = (\frac{\alpha}{2(\alpha+1)\sigma^2}\log(n))$, it yields that

$$\mathbb{E}\left[\|\widehat{f}_{(\ell_{opt}),d_{opt}} - f\|^2\right] \lesssim \log(n)^{-s}.$$

□

Proof of Proposition 3.5. First, note that as f and g are Gaussian densities, then $h = f \star g$ is it also a Gaussian density with variance $\sigma^2 + \theta^2$. It is proved in Belomestny et al. (2019) (see Proof of Proposition 7, p. 55-56) that the bias for Gaussian density is exponentially decaying and its order is given by $\|h - h_d\|^2 \lesssim \frac{1}{\sqrt{d}}\exp(-\lambda_{\sigma,\theta} d)$, where $\lambda_{\sigma,\theta} = \log\left[\left(\frac{\sigma^2+\theta^2+1}{\sigma^2+\theta^2-1}\right)^2\right] > 0$. We derive that:

$$\mathbb{E}[\|\widehat{h}_d - h\|^2] \lesssim \sigma_\varepsilon^2 \lambda \frac{T}{n} d + \frac{1}{\sqrt{d}}\exp(-\lambda_{\sigma,\theta} d) + \mathcal{O}(\frac{T^2}{n}) \tag{55}$$

Injecting $d_{opt} = [\log(n)/\lambda_{\sigma,\theta}]$ in (55), we have (30). Injecting this in (28), it comes

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d_{opt}} - f\|^2\right] \leqslant \|f - f_{(\ell)}\|^2 + \Delta(\ell)\frac{\log(n)}{n}.$$

As $g^*(u) = \exp(-\frac{\theta^2 t^2}{2})$ then, it holds $\Delta(\ell) = \sup_{|u|\leqslant\ell} |g^*(u)|^{-2} \leqslant e^{\theta^2\ell^2}$. Using Lemma 2 in Comte and Lacour (2011), we have

$$\|f - f_{(\ell)}\|^2 = \frac{1}{2\pi}\int_{|u|>\ell} |f^*(u)|^2 du \asymp \ell^{-1} e^{-\sigma^2\ell^2}.$$

Consequently, we get from (28)

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d_{opt}} - f\|^2\right] \lesssim \ell^{-1} e^{-\sigma^2\ell^2} + e^{\theta^2\ell^2}\frac{\log(n)}{n},$$

Replacing $\ell^2_{opt} = \frac{1}{(\sigma^2+\theta^2}\log(n) - \frac{3}{2(\theta^2+\sigma^2)}\log\log(n)$ gives the announced result. □

Proof of Proposition 3.6. Recall that as f is $\Gamma(p,\theta)$ and g $\Gamma(q,\theta)$, then, the regression function $h = f \star g \sim \Gamma(p+q,\theta)$ and belongs to $h \in W_H^{(p+q-2)}$ since $p + q > 2$. We have

$$\mathbb{E}[\|\widehat{h}_d - h\|^2] \leqslant C d^{-(p+q-2)} + \sigma_\varepsilon^2 \lambda \frac{T}{n} d + \mathcal{O}(\frac{T^2}{n}).$$

Replacing d by $d_{opt} = [n^{1/(p+q-1)}]$, we derive

$$\mathbb{E}[\|\widehat{h}_{d_{opt}} - h\|^2] \lesssim n^{-\frac{p+q-2}{p+q-1}}.$$

Now, we consider the deconvolution part. The Fourier transform of g and its modulus are given by

$$g^*(t) = (1 - i\frac{t}{\theta})^{-q}, \quad |g^*(t)| = (1 + \frac{t^2}{\theta^2})^{-\frac{q}{2}}.$$

Then, it holds that

$$\Delta(\ell) = \sup_{|u|\leqslant\ell} |g^*(u)|^{-2} \leqslant (1 + \frac{\ell^2}{\theta^2})^q = c\ell^{2q}$$

and using Lemma 2 in Comte and Lacour (2011), it follows $\|f - f_{(\ell)}\|^2 = \frac{1}{2\pi}\int_{|u|>\ell}|f^*(u)|^2du \asymp (\frac{\ell}{\theta})^{-2p+1}$. Plugging the previous results in (28) yields

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d_{opt}} - f\|^2\right] \lesssim (\frac{\ell}{\theta})^{-2p+1} + c\ell^{2q}n^{-\frac{p+q-2}{p+q-1}}.$$

Choosing $\ell_{opt} := n^{\frac{p+q-2}{(p+q-1)(2p+2q-1)}}$ gives

$$\mathbb{E}\left[\|\widehat{f}_{(\ell_{opt}),d_{opt}} - f\|^2\right] = \mathcal{O}\left(n^{-\frac{(p+q-2)(2p-1)}{(p+q-1)(2p+2q-1)}}\right).$$

□

Proof of Theorem 3.7. Let us start by the proof of Inequality (35). First, we have by definition of $\widehat{A}$, $\widehat{d}$ and $\forall d \in \mathcal{M}_n^{(1)}$,

$$\begin{aligned}\|\widetilde{f}_{(\widehat{d})} - f\|^2 &= \|\widetilde{f}_{(\widehat{d})} - \widetilde{f}_{(\widehat{d}\wedge d)} + \widetilde{f}_{(\widehat{d}\wedge d)} - \widetilde{f}_{(d)} + \widetilde{f}_{(d)} - f\|^2\\ &\leqslant 3\|\widetilde{f}_{(\widehat{d})} - \widetilde{f}_{(\widehat{d}\wedge d)}\|^2 + 3\|\widetilde{f}_{(\widehat{d}\wedge d)} - \widetilde{f}_{(d)}\|^2 + 3\|\widetilde{f}_{(d)} - f\|^2\\ &\leqslant 3(\widehat{A}(d) + \kappa_1 V(\widehat{d})) + 3(\widehat{A}(\widehat{d}) + \kappa_1 V(d)) + 3\|\widetilde{f}_d - f\|^2\\ &\leqslant 6(\widehat{A}(d) + \kappa_2 V(d)) + 3\|\widetilde{f}_d - f\|^2.\end{aligned}$$

Taking the expectation in the previous inequality, we get

$$\mathbb{E}\left[\|\widetilde{f}_{(\widehat{d})} - f\|^2\right] \leqslant 6\,\mathbb{E}[\widehat{A}(d)] + 6\kappa_2 V(d)) + 3\,\mathbb{E}\left[\|\widetilde{f}_d - f\|^2\right]. \tag{56}$$

Now, we are interested in the study of $\mathbb{E}[\widehat{A}(d)]$. For all $d \in \mathcal{M}_n^{(1)}$, we use the following decomposition

$$\begin{aligned}\|\widetilde{f}_{(d')} - \widetilde{f}_{(d'\wedge d)}\|^2 &= \|\widetilde{f}_{(d')} - \mathbb{E}[\widetilde{f}_{(d')}] + \mathbb{E}[\widetilde{f}_{(d')}] - \mathbb{E}[\widetilde{f}_{(d'\wedge d)}] + \mathbb{E}[\widetilde{f}_{(d'\wedge d)}] - \widetilde{f}_{(d'\wedge d)}\|^2\\ &\leqslant 3\|\widetilde{f}_{(d')} - \mathbb{E}[\widetilde{f}_{(d')}]\|^2 + 3\|\,\mathbb{E}[\widetilde{f}_{(d'\wedge d)}] - \widetilde{f}_{(d'\wedge d)}\|^2 + 3\|\,\mathbb{E}[\widetilde{f}_{(d')}] - \mathbb{E}[\widetilde{f}_{(d'\wedge d)}]\|^2.\end{aligned}$$

Using this, it comes

$$\begin{aligned}\widehat{A}(d) \leqslant& 3\max_{d'\in\mathcal{M}_n^{(1)}}\left\{\left(\|\widetilde{f}_{(d')} - \mathbb{E}[\widetilde{f}_{(d')}]\|^2 - \frac{\kappa_1}{6}V(d')\right)_+\right\} + 3\max_{d'\in\mathcal{M}_n^{(1)}}\left\{\left(\|\,\mathbb{E}[\widetilde{f}_{(d'\wedge d)}] - \widetilde{f}_{(d'\wedge d)}\|^2 - \frac{\kappa_1}{6}V(d')\right)_+\right\}\\ &+ 3\max_{d'\in\mathcal{M}_n^{(1)}}\left\{\|\,\mathbb{E}[\widetilde{f}_{(d')}] - \mathbb{E}[\widetilde{f}_{(d'\wedge d)}]\|^2\right\}.\end{aligned}$$

Let us remark that if $d' \leqslant d$, the last term is equal to zero. We have

$$\begin{aligned}\max_{d'\in\mathcal{M}_n^{(1)}}\|\,\mathbb{E}[\widetilde{f}_{(d')}] - \mathbb{E}[\widetilde{f}_{(d'\wedge d)}]\|^2 &= \max_{d'\in\mathcal{M}_n^{(1)},d<d'}\|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}] - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\|^2\\ &= \max_{d'\in\mathcal{M}_n^{(1)},d<d'}\left\{\|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}] - f_{(\sqrt{2d'})} + f_{(\sqrt{2d'})} - f_{(\sqrt{2d})} + f_{(\sqrt{2d})} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\|^2\right\}\\ &\leqslant 3\max_{d'\in\mathcal{M}_n^{(1)},d<d'}\|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}] - f_{(\sqrt{2d'})}\|^2 + 3\|f_{(\sqrt{2d})} - \|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\|^2\\ &\qquad + 3\max_{d'\in\mathcal{M}_n^{(1)},d<d'}\|f_{(\sqrt{2d'})} - f_{(\sqrt{2d})}\|^2.\end{aligned}$$

Besides, by definition of $\widehat{f}_{(\sqrt{2d}),d}$ given in (25) and $f_{(\sqrt{2d})}$ in (26), we have for all $d \geqslant 1$

$$\|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}] - f_{(\sqrt{2d})}\|^2 = \frac{1}{2\pi}\int_{-\sqrt{2d}}^{\sqrt{2d}}\frac{|\,\mathbb{E}[\widehat{h}_d^*(u)] - h^*(u)|^2}{|g^*(u)|^2}du \leqslant \Delta(\sqrt{2d})\|h - \mathbb{E}[\widehat{h}_d]\|^2,$$

and for $d' \geqslant d$

$$\|f_{(\sqrt{2d'})} - f_{(\sqrt{2d})}\|^2 = \int_{\sqrt{2d}\leqslant|u|\leqslant\sqrt{2d'}} |f^*(u)|^2 du \leqslant \int_{|u|\geqslant\sqrt{2d}} |f^*(u)|^2 du = \|f - f_{(\sqrt{2d})}\|^2.$$

This implies

$$\begin{aligned}\widehat{A}(d) \leqslant & 3 \max_{d'\in\mathcal{M}_n^{(1)}} \left\{ \left(\|\widehat{f}_{(\sqrt{2d'}),d'} - \mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}]\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right\} \\ & + 3 \max_{d'\in\mathcal{M}_n^{(1)}} \left\{ \left(\|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d'}\wedge\sqrt{2d}),d'\wedge d}] - \widehat{f}_{(\sqrt{2d'}\wedge\sqrt{2d}),d'\wedge d}\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right\} \\ & + 9 \max_{d'\in\mathcal{M}_n^{(1)}, d\leqslant d'} \Delta(\sqrt{2d'}) \|h - \mathbb{E}[\widehat{h}_{d'}]\|^2 + 9\|f - f_{(\sqrt{2d})}\|^2.\end{aligned}$$

As

$$\max_{d'\in\mathcal{M}_n^{(1)}} \left\{ \left(\|\widehat{f}_{(\sqrt{2d'}),d'} - \mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}]\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right\} \leqslant \sum_{d'\in\mathcal{M}_n^{(1)}} \left\{ \left(\|\widehat{f}_{(\sqrt{2d'}),d'} - \mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}]\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right\}$$

and $V(d') \geqslant V(d' \wedge d)$, then, we have the following bound

$$\begin{aligned}& \max_{d'\in\mathcal{M}_n^{(1)}} \left\{ \left(\|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d'}\wedge\sqrt{2d}),d'\wedge d}] - \widehat{f}_{(\sqrt{2d'}\wedge\sqrt{2d}),d'\wedge d}\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right\} \\ & \quad \leqslant \max_{d'\in\mathcal{M}_n^{(1)}, d'\leqslant d} \left\{ \left(\|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}] - \widehat{f}_{(\sqrt{2d'}),d'}\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right\} \\ & \qquad + \left\{ \left(\|\,\mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}] - \widehat{f}_{(\sqrt{2d}),d}\|^2 - \frac{\kappa_1}{6} V(d) \right)_+ \right\} \\ & \qquad \leqslant 2 \sum_{d'\in\mathcal{M}_n^{(1)}} \left\{ \left(\|\widehat{f}_{(\sqrt{2d'}),d'} - \mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}]\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right\}.\end{aligned}$$

Consequently, it follows

$$\begin{aligned}\mathbb{E}\left[\widehat{A}(d)\right] \leqslant & 9 \sum_{d'\in\mathcal{M}_n^{(1)}} \mathbb{E}\left[\left(\|\widehat{f}_{(\sqrt{2d'}),d'} - \mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}]\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right] \\ & + 9 \max_{d'\in\mathcal{M}_n^{(1)}, d\leqslant d'} \Delta(\sqrt{2d'}) \|h - \mathbb{E}[\widehat{h}_{d'}]\|^2 + 9\|f - f_{(\sqrt{2d})}\|^2.\end{aligned}$$

Next, we have to control the term $\sum_{d'\in\mathcal{M}_n^{(1)}} \mathbb{E}\left[\left(\|\widehat{f}_{(\sqrt{2d'}),d'} - \mathbb{E}[\widehat{f}_{(\sqrt{2d'}),d'}]\|^2 - \frac{\kappa_1}{6} V(d') \right)_+ \right]$. We use the following technical Lemma.

Lemma 6.1. *Under the assumptions of Theorem 3.7, it holds for $\kappa_1 \geqslant 12$ and C_0 a positive constant,*

$$\sum_{d\in\mathcal{M}_n^{(1)}} \mathbb{E}\left[\left(\|\widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\|^2 - \frac{\kappa_1}{6} V(d) \right)_+ \right] \leqslant \frac{C_0 \log(n)}{n},$$

where $C_0 = C_0(\mathbb{E}[\varepsilon_1^4], \gamma, c_1, \xi, \lambda, C'_\infty)$.

By Phythagoras Therorem, we have

$$\|h - \mathbb{E}[\widehat{h}_d]\| = \|h - h_d\|^2 + \|h_d - \mathbb{E}[\widehat{h}_d]\|^2. \tag{57}$$

Then, we deduce from Lemma 6.1, (56), (54) and (48) that:

$$\mathbb{E}[\|\widetilde{f}_{(\widehat{d})} - f\|^2] \leqslant 57\|f - f_{(\sqrt{2d})}\|^2 + 7\kappa_2 V(d) + 54C_0\frac{\log(n)}{n} + 57R_b(d), \tag{58}$$

where $R_b(d) := \left[\max_{d'\in\mathcal{M}_n^{(1)}, d\leqslant d'} \Delta(\sqrt{2d'})\|h - \mathbb{E}[\widehat{h}_{d'}]\|^2\right]$. Taking the infimum d and choosing $C = \max(57, 7\kappa_2)$, $C' = 54C_0$ in (58) ends the proof of Inequality (35).

Now, we prove Inequality (36). From (57)-(50) and (**A0**), it holds

$$\Delta(\sqrt{2d'})\|h - \mathbb{E}[\widehat{h}_{d'}]\|^2 \leqslant (1+\lambda)\Delta(\sqrt{2d'})\|h - h_{d'}\|^2 + \lambda\Delta(\sqrt{2d'})\|h - h_{d'}\|_n^2.$$

Under (**A3**), it comes from Lemma 2.2 (ii) and for $h \in W_H^{s+\gamma}(L')$,

$$\begin{aligned}\Delta(\sqrt{2d'})\|h - \mathbb{E}[\widehat{h}_{d'}]\|^2 \leqslant& c_1(1+\lambda)(1+2d')^\gamma L'(d')^{-s-\gamma} + C\Delta(\sqrt{2d'})\frac{T^3}{n^2}\\ \leqslant& C\left(d'^{-s} + \frac{T^2}{n}\right).\end{aligned}$$

Then, for $d' \geqslant d$, we derive that $R_b(d) \leqslant C\left(d^{-s} + \frac{T^2}{n}\right)$. Plugging this in (58) and using $\|f - f_{\sqrt{2d}}\|^2 \leqslant 2^{-s}Ld^{-s}$ because $f \in W^s(L)$ concludes the proof of Theorem 3.7. □

6.3. **Proofs of Section 4.**

Proof of Proposition 4.1. By the Pythagoras Theorem, we have

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] = \|f - f_m\|^2 + \mathbb{E}\left[\|\widehat{f}_{m,d} - f_m\|^2\right] \tag{59}$$

Let us study the term $\mathbb{E}\left[\|\widehat{f}_{m,d} - f_m\|^2\right]$. On the one hand, by definition of $\widehat{f}_{m,d}$ and f_m, it yields

$$\begin{aligned}\mathbb{E}\left[\|\widehat{f}_{m,d} - f_m\|^2\right] =& \mathbb{E}[\sum_{j=0}^{m-1}(\widehat{a}_{j,d} - a_j)^2] = \mathbb{E}\left[\sum_{j=0}^{m-1}\frac{1}{2\pi}|\langle\frac{\widehat{h}_d^* - h^*}{g^*}, \varphi_j\rangle|^2\right]\\ \leqslant& \frac{1}{\pi}\mathbb{E}\left[\sum_{j=0}^{m-1}|\langle\frac{\widehat{h}_d^* - h^*}{g^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{\rho m}}, \varphi_j\rangle|^2\right] + \frac{1}{\pi}\mathbb{E}\left[\sum_{j=0}^{m-1}|\langle\frac{\widehat{h}_d^* - h^*}{g^*}\mathbb{1}_{|\cdot|\geqslant\sqrt{\rho m}}, \varphi_j\rangle|^2\right].\end{aligned}$$

By the Bessel Inequality, it holds

$$\begin{aligned}\sum_{j=0}^{m-1}\langle\frac{\widehat{h}_d^* - h^*}{g^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{\rho m}}, \varphi_j\rangle^2 \leqslant& \|\frac{\widehat{h}_d^* - h^*}{g^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{\rho m}}\|^2 = \int_{|x|\leqslant\sqrt{\rho m}}\left|\frac{\widehat{h}_d^*(u) - h^*}{g^*(u)}\right|^2 du\\ \leqslant& \sup_{|u|\leqslant\sqrt{\rho m}}\frac{1}{|g^*(u)|^2}\int|\widehat{h}_d^*(u) - h^*|^2 du.\end{aligned}$$

The Cauchy-Schwarz inequality gives,

$$\langle\frac{\widehat{h}_d^* - h^*}{g^*}\mathbb{1}_{|\cdot|\geqslant\sqrt{\rho m}}, \varphi_j\rangle^2 \leqslant \|\widehat{h}_d^* - h^*\|^2\|\frac{\varphi_j}{g^*}\mathbb{1}_{|\cdot|\geqslant\sqrt{\rho m}}\|^2.$$

Consequently, we get

$$\mathbb{E}\left[\|\widehat{f}_{m,d}-f_m\|^2\right] \leqslant \frac{1}{\pi}\left[\sup_{|u|\leqslant\sqrt{\rho m}}\frac{1}{|g^*(u)|^2}+\sum_{j=0}^{m-1}\int_{|u|\geqslant\sqrt{\rho m}}\frac{|\varphi_j(u)|^2}{|g^*(u)|^2}du\right]\mathbb{E}\left[\|\widehat{h}_d^*-h^*\|^2\right]$$
$$= 2\Sigma(m)\mathbb{E}\left[\|\widehat{h}_d-h\|^2\right].$$

Injecting this in (59) and using Proposition 2.1 (ii), we get

$$\mathbb{E}\left[\|\widehat{f}_{m,d}-f\|^2\right] \leqslant \|f-f_m\|^2+2\Sigma(m)\left(\|h-h_d\|^2+\lambda_{max}\left(\Psi_d^{-1}\right)\|h-h_d\|_n^2+\sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right)\right). \tag{60}$$

On the other hand, from (59), we have

$$\mathbb{E}\left[\|\widehat{f}_{m,d}-f\|^2\right] = \|f-f_m\|^2+\|\,\mathbb{E}[\widehat{f}_{m,d}]-f_m\|^2+\mathbb{E}\left[\|\widehat{f}_{m,d}-\mathbb{E}[\widehat{f}_{m,d}]\|^2\right]. \tag{61}$$

We study the last two terms on the above expression. Start by the second. To do this, we introduce the matrix:

$$M := \left(\int_{\mathbb{R}}\frac{\varphi_j^*\varphi_k^*}{g^*}\right)_{0\leqslant j\leqslant m-1,\ 0\leqslant k\leqslant d-1}$$

By definition $\widehat{h}_d$ given in (13), we remark $\widehat{a}_{j,d}=[M\vec{\widehat{b}}^{(d)}]_j$ with $\vec{\widehat{b}}^{(d)}=(\widehat{b}_0^{(d)},\dots,\widehat{b}_{d-1}^{(d)})^t$. We set

$$\vec{\widehat{f}}_{m,d}=(\widehat{a}_{0,d},\dots,\widehat{a}_{m-1,d})^t=[M\vec{\widehat{b}}^{(d)}]_{0\leqslant j\leqslant m-1}$$

Then, it yields

$$\mathbb{E}\left[\|\widehat{f}_{m,d}-\mathbb{E}[\widehat{f}_{m,d}]\|^2\right]=\mathbb{E}\left[\|\vec{\widehat{f}}_{m,d}-\mathbb{E}[\vec{\widehat{f}}_{m,d}]\|^2_{\mathbb{R}^{(m)}}\right]=\mathbb{E}\left[\|M(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\vec{\varepsilon}\|^2_{\mathbb{R}^{(m)}}\right]$$
$$=\sigma_\varepsilon^2\mathrm{tr}\left[\Phi_d(\Phi_d^t\Phi_d)^{-1}M^tM(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\right]$$
$$=\frac{\sigma_\varepsilon^2T}{n}\mathrm{tr}[\Psi_d^{-1}M^tM].$$

As Ψ_d^{-1} is a definite symmetric positive matrix, then, it is diagonalizable $\Psi_d^{-1}=PDP^t$ with $D=\mathrm{diag}(\mu_1,\dots,\mu_d)$, where the $\mu_i>0$ are eigenvalues of matrix Ψ_d^{-1} and $PP^t=P^tP=I_d$. We can define the root square of Ψ_d^{-1} and derive (see Proof of Theorem 3.7 when we compute M_1) $\mathrm{tr}[\Psi_d^{-1}M^tM]\leqslant\lambda_{max}(\Psi_d^{-1})\mathrm{tr}[M^tM]$. The Frobenuis norm and Bessel inequality give:

$$\|M\|_F^2=\mathrm{tr}[M^tM]=\sum_{j=0}^{m-1}\sum_{k=0}^{d-1}\left|\int_{\mathbb{R}}\frac{\varphi_j^*(u)\varphi_k^*(u)}{g^*(u)}du\right|^2\leqslant 2\pi\sum_{j=0}^{m-1}\int_{\mathbb{R}}\frac{|\varphi_j^*(u)|^2}{|g^*(u)|^2}du$$
$$\leqslant 4\pi^2\left(m\Delta(\sqrt{\rho m})+\sum_{j=0}^{m-1}\int_{|u|\geqslant\sqrt{\rho m}}\frac{|\varphi_j^*(u)|^2}{|g^*(u)|^2}du\right).$$

Consequently, it holds

$$\mathbb{E}\left[\|\widehat{f}_{m,d}-\mathbb{E}[\widehat{f}_{m,d}]\|^2\right]\leqslant 4\pi^2\left(m\Delta(\sqrt{\rho m})+\sum_{j=0}^{m-1}\int_{|u|\geqslant\sqrt{\rho m}}\frac{|\varphi_j^*(u)|^2}{|g^*(u)|^2}du\right)\sigma_\varepsilon^2\frac{T}{n}.$$

Similarly to the study of quantity $\mathbb{E}\left[\|\widehat{f}_{m,d}-f_m\|^2\right]$ where $\widehat{f}_{m,d}$ is replaced by $\mathbb{E}[\widehat{f}_{m,d}]$, we have

$$\|\,\mathbb{E}[\widehat{f}_{m,d}]-f_m\|^2\leqslant 2\Sigma(m)\left(\|h-h_d\|^2+\lambda_{max}\left(\Psi_d^{-1}\right)\|h-h_d\|_n^2\right).$$

Plugging the two last terms on the above bound in 61, we obtain

$$\mathbb{E}\left[\|\widehat{f}_{m,d}-f\|^2\right]\leqslant\|f-f_m\|^2+2\Sigma(m)\left(\|h-h_d\|^2+\lambda_{max}\left(\Psi_d^{-1}\right)\|h-h_d\|_n^2+2\pi^2\sigma_\varepsilon^2\lambda_{max}\left(\Psi_d^{-1}\right)m\frac{T}{n}\right).$$

Combining this and (60) ends the proof. □

Proof of Theorem 4.2. Under **(A3)**, **(A0)** and for h belongs to $W_H^\alpha(L')$, it holds from Lemma 2.2:

$$\mathbb{E}\left[\|\widehat{f}_{m,d}-f\|^2\right]\leqslant Lm^{-s}+2\Sigma(m)\left[(1+\lambda)L'd^{-\alpha}+\lambda\sigma_\varepsilon^2T\frac{d}{n}+C(\alpha,L)\frac{T^2}{n}\right]$$

Besides, under **(A3)** and from (7) with $\rho\geqslant 2$, we have

$$\sum_{j=0}^{m-1}\int_{|u|\geqslant\sqrt{\rho m}}|\varphi_j(u)|^2|g^*(u)|^{-2}dx\leqslant\sum_{j=0}^{m-1}C^2e^{-\xi\rho m}\int(1+u^2)e^{-\xi u^2}du\leqslant C(\xi)me^{-\xi\rho m}.$$

As $\sup_{|x|\leqslant\sqrt{\rho m}}|g^*(x)|^{-2}\leqslant c_1(1+(m\rho)^\gamma)$, then, there exits a constant, denoted C_1 such that $\Sigma(m)\leqslant C_1m^\gamma$. Then, we obtain

$$\mathbb{E}\left[\|\widehat{f}_{m,d}-f\|^2\right]\leqslant Lm^{-s}+2C_1m^\gamma\left[(1+\lambda)L'd^{-\alpha}+\lambda\sigma_\varepsilon^2T\frac{d}{n}+C(\alpha,L)\frac{T^2}{n}\right],$$

and the choices $m_{opt}=d_{opt}=[n^{1/(\alpha+1)}]$ with $\alpha=s+\gamma>11/6$ end the proof. □

Proof of Theorem 4.3.

Proof of Inequality (44). By definition of $\widehat{B}$, $\widetilde{m}$ and $\forall m\in\mathcal{M}_n^{(2)}$, we have

$$\|\widetilde{f}_{\widetilde{m}}-f\|^2\leqslant 6\Big(\widehat{B}(m)+\kappa_2'W(m)\Big)+3\|\widetilde{f}_m-f\|^2,$$

by analogy with the proof of Theorem 3.7. Then, it yields

$$\mathbb{E}\left[\|\widetilde{f}_{\widetilde{m}}-f\|^2\right]\leqslant 6\,\mathbb{E}[\widehat{B}(m)]+\kappa_2'W(m)+3\,\mathbb{E}\left[\|\widetilde{f}_m-f\|^2\right]\tag{62}$$

Next, we study the term $\mathbb{E}[\widehat{B}(m)]$. For all $m,m'\in\mathcal{M}_n^{(2)}$, we have the following decomposition

$$\|\widetilde{f}_{m'}-\widetilde{f}_{m'\wedge m}\|^2\leqslant 3\|\widetilde{f}_{m'}-\mathbb{E}[\widetilde{f}_{m'}]\|^2+3\|\widetilde{f}_{m'\wedge m}-\mathbb{E}[\widetilde{f}_{m'\wedge m}]\|^2+3\|\,\mathbb{E}[\widetilde{f}_{m'}]-\mathbb{E}[\widetilde{f}_{m'\wedge m}]\|^2,$$

which implies

$$\begin{aligned}\widehat{B}(m)\leqslant&3\max_{m'\in\mathcal{M}_n^{(2)}}\left\{\left(\|\widetilde{f}_{m'}-\mathbb{E}[\widetilde{f}_{m'}]\|^2-\frac{\kappa_1'}{6}W(m')\right)_+\right\}+3\max_{m'\in\mathcal{M}_n^{(2)}}\left\{\left(\|\widetilde{f}_{m'\wedge m}-\mathbb{E}[\widetilde{f}_{m'\wedge m}]\|^2-\frac{\kappa_1'}{6}W(m')\right)_+\right\}\\&+3\max_{m'\in\mathcal{M}_n^{(2)}}\left\{\|\,\mathbb{E}[\widetilde{f}_{m'}]-\mathbb{E}[\widetilde{f}_{m'\wedge m}]\|^2\right\}\end{aligned}$$

Note that for $m'\leqslant m$, the last term is null. More precisely, we have

$$\begin{aligned}\max_{m'\in\mathcal{M}_n^{(2)}}\left\{\|\,\mathbb{E}[\widetilde{f}_{m'}]-\mathbb{E}[\widetilde{f}_{m'\wedge m}]\|^2\right\}&=\max_{m'\in\mathcal{M}_n^{(2)},m<m'}\left\{\|\,\mathbb{E}[\widetilde{f}_{m'}]-\mathbb{E}[\widetilde{f}_{m'\wedge m}]\|^2\right\}\\&\leqslant3\max_{m'\in\mathcal{M}_n^{(2)},m<m'}\left\{\|\,\mathbb{E}[\widetilde{f}_{m'}]-f_{m'}\|^2\right\}+3\|f_m-\mathbb{E}[\widetilde{f}_m]\|^2\\&\quad+3\max_{m'\in\mathcal{M}_n^{(2)},m<m'}\|f'_m-f_m\|^2.\end{aligned}$$

Moreover, it yields

$$\|f_m - \mathbb{E}[\widetilde{f}_m]\|^2 = \frac{1}{2\pi}\left|\sum_{j=0}^{m-1}\int \frac{h^*(u)-\mathbb{E}[h_m^*(u)]}{g^*(u)}\varphi_j(u)du\right|^2$$
$$\leqslant \frac{1}{\pi}\left|\sum_{j=0}^{m-1}\int_{|u|\leqslant\sqrt{\rho m}} \frac{h^*(u)-\mathbb{E}[h_m^*(u)]}{g^*(u)}\varphi_j(u)du\right|^2 + \frac{1}{\pi}\left|\sum_{j=0}^{m-1}\int_{|u|>\sqrt{\rho m}} \frac{h^*(u)-\mathbb{E}[h_m^*(u)]}{g^*(u)}\varphi_j(u)du\right|^2$$

From Bessel, the Cauchy-Schwarz inequalities and Parseval equality, we obtain $\|f_m - \mathbb{E}[\widetilde{f}_m]\|^2 \leqslant 2\Sigma(m)\|h - \mathbb{E}[\widehat{h}_m]\|^2$. For $m < m'$, we have $\|f'_m - f_m\|^2 \leqslant \|f - f_m\|^2$. Consequently, it holds

$$\widehat{B}(m) \leqslant 3\max_{m'\in\mathcal{M}_n^{(2)}}\left\{\left(\|\widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}]\|^2 - \frac{\kappa'_1}{6}W(m')\right)_+\right\} + 3\max_{m'\in\mathcal{M}_n^{(2)}}\left\{\left(\|\widetilde{f}_{m'\wedge m} - \mathbb{E}[\widetilde{f}_{m'\wedge m}]\|^2 - \frac{\kappa'_1}{6}W(m')\right)_+\right\}$$
$$+ 18\max_{m'\in\mathcal{M}_n^{(2)}, m\leqslant m'}\Sigma(m')\|h - \mathbb{E}[\widehat{h}_{m'}]\|^2 + 9\|f - f_m\|^2.$$

Notice that

$$\max_{m'\in\mathcal{M}_n^{(2)}}\left\{\left(\|\widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}]\|^2 - \frac{\kappa'_1}{6}W(m')\right)_+\right\} \leqslant \sum_{m'\in\mathcal{M}_n^{(2)}}\left\{\left(\|\widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}]\|^2 - \frac{\kappa'_1}{6}W(m')\right)_+\right\},$$

and $W(m') \geqslant W(m'\wedge m)$, then, we have the following bound

$$\max_{m'\in\mathcal{M}_n^{(2)}}\left\{\left(\|\widetilde{f}_{m'\wedge m} - \mathbb{E}[\widetilde{f}_{m'\wedge m}]\|^2 - \frac{\kappa'_1}{6}W(m')\right)_+\right\} \leqslant 2\sum_{m'\in\mathcal{M}_n^{(2)}}\left\{\left(\|\widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}]\|^2 - \frac{\kappa'_1}{6}W(m')\right)_+\right\}.$$

Finally, after taking expectation, we get

$$\mathbb{E}[\widehat{B}(m)] \leqslant 9\sum_{m'\in\mathcal{M}_n^{(2)}}\mathbb{E}\left[\left(\|\widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}]\|^2 - \frac{\kappa'_1}{6}W(m')\right)_+\right]$$
$$+ 18\max_{m'\in\mathcal{M}_n^{(2)}, m\leqslant m'}\Sigma(m')\|h - \mathbb{E}[\widehat{h}_{m'}]\|^2 + 9\|f - f_m\|^2.$$

Lemma 6.2. *Under the assumption of Theorem 4.3, for $\kappa'_1 \geqslant 12$, we have*

$$\sum_{m'\in\mathcal{M}_n^{(2)}}\mathbb{E}\left[\left(\|\widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}]\|^2 - \frac{\kappa'_1}{6}W(m')\right)_+\right] \leqslant C_0\frac{\log(n)}{n}.$$

Lemma 6.2 implies that

$$\mathbb{E}[\widehat{B}(m)] \leqslant 9C_0\frac{\log(n)}{n} + 18\max_{m'\in\mathcal{M}_n^{(2)}, m\leqslant m'}\Sigma(m')\|h - \mathbb{E}[\widehat{h}_{m'}]\|^2 + 9\|f - f_m\|^2.$$

Injecting this in (62) and from (57)-(50), we obtain

$$\mathbb{E}\left[\|\widetilde{f}_{\widetilde{m}} - f\|^2\right] \leqslant 57\|f - f_m\|^2 + 108R_b(m') + 7\kappa'_2W(m) + 54C_0\frac{\log(n)}{n}. \tag{63}$$

Choosing $C = \max(108, 7\kappa'_2)$, $C' = 54C_0$ and taking the infimum on $m \in \mathcal{M}_n^{(2)}$ concludes the proof of Inequality (44).

Proof of Inequality (45). Recall that for $f \in W^s(L)$, it holds $\|f - f_m\|^2 \leqslant Lm^{-s}$. Similarly to

the proof of (36), we derive for $h \in W_H^{s+\gamma}(L')$ with $s+\gamma \geqslant 17/6$ that $R'_b(m) \leqslant C(S,\gamma,\lambda,\sigma_\varepsilon^2)(m^{-s}+T^2/n)$. Plugging the two previous inequalities into (63) gives (45). □

References

Abramovich, F., Pensky, M., and Rozenholc, Y. (2013). Laplace deconvolution with noisy observations. *Electron. J. Stat.*, 7:1094–1128.

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.

Ameloot, M. and Hendrickx, H. (1983). Extension of the performance of laplace deconvolution in the analysis of fluorescence decay curves. *Biophysical journal*, 44(1):27–38.

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

Baraud, Y. (2000). Model selection for regression on a fixed design. *Probab. Theory Related Fields*, 117(4):467–493.

Baraud, Y. (2002). Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6:127–146.

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413.

Baudry, J., Maugis, C., and Michel, B. (2012). Slope heuristics: overview and implementation. *Stat. Comput.*, 22(2):455–470.

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.

Benhaddou, R., Pensky, M., and Rajapakshage, R. (2019). Anisotropic functional Laplace deconvolution. *J. Statist. Plann. Inference*, 199:271–285.

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

Cao, M., Liang, Y., and Stantz, K. M. (2010). Response to letter regarding article:"developing dce-ct to quantify intra-tumor heterogeneity in breast tumors with differing angiogenic phenotype". *IEEE Transactions on Medical Imaging*, 29(4):1089–1092.

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

Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2017). Laplace deconvolution on the basis of time domain data and its application to dynamic contrast-enhanced imaging. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(1):69–94.

Comte, F., Duval, C., and Sacko, O. (2020). Optimal adaptive estimation on $\mathbb{R}$ or $\mathbb{R}^+$ of the derivatives of a density. *Math. Methods Statist.*, 29(1):1–31.

Comte, F. and Genon-Catalot, V. (2018). Laguerre and Hermite bases for inverse problems. *J. Korean Statist. Soc.*, 47(3):273–296.

Comte, F. and Genon-Catalot, V. (2019). Regression function estimation as a partly inverse problem. *Annals of the Institute of Statistical Mathematics*, pages 1–32.

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.

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

Cuenod, C., Fournier, L., Balvay, D., and Guinebretiere, J.-M. (2006). Tumor angiogenesis: pathophysiology and implications for contrast-enhanced mri and ct assessment. *Abdominal imaging*, 31(2):188–193.

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

Dey, A. K., Martin, C. F., and Ruymgaart, F. H. (1998). Input recovery from noisy output data, using regularized inversion of the laplace transform. *IEEE Transactions on Information Theory*, 44(3):1125–1130.

Fan, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A*, 53(1):97–110.

Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.*, 21(2):600–610.

Gafni, A., Modlin, R. L., and Brand, L. (1975). Analysis of fluorescence decay curves by means of the laplace transformation. *Biophysical journal*, 15(3):263–280.

Goh, V., Halligan, S., Hugill, J.-A., Gartner, L., and Bartram, C. I. (2005). Quantitative colorectal cancer perfusion measurement using dynamic contrast-enhanced multidetector-row computed tomography: effect of acquisition time and implications for protocols. *Journal of computer assisted tomography*, 29(1):59–63.

Goh, V., Padhani, A. R., and Rasheed, S. (2007). Functional imaging of colorectal cancer angiogenesis. *The Lancet Oncology*, 8(3):245–255.

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

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

Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077.

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

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

Ledoux, M. (1997). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, 1:63–87.

Loubes, J.-M. and Marteau, C. (2012). Adaptive estimation for an inverse regression model with unknown operator. *Stat. Risk Model.*, 29(3):215–242.

Mabon, G. (2017). Adaptive deconvolution on the non-negative real line. *Scand. J. Stat.*, 44(3):707–740.

McKinnon, A., Szabo, A., and Miller, D. (1977). The deconvolution of photoluminescence data. *The Journal of Physical Chemistry*, 81(16):1564–1570.

Meister, A. (2009). On testing for local monotonicity in deconvolution problems. *Statist. Probab. Lett.*, 79(3):312–319.

O'Connor, D., Ware, W., and Andre, J. (1979). Deconvolution of fluorescence decay curves. a critical comparison of techniques. *Journal of Physical Chemistry*, 83(10):1333–1343.

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

Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Ann. Statist.*, 11(1):141–156.

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

Stewart, G. W. and Sun, J. G. (1990). *Matrix perturbation theory*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA.

Szegö, G. (1959). *Orthogonal polynomials*. American Mathematical Society Colloquium Publications, Vol. 23. American Mathematical Society, Providence, R.I. Revised ed.

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

Vareschi, T. (2015). Noisy Laplace deconvolution with error in the operator. *J. Statist. Plann. Inference*, 157/158:16–35.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, Cambridge.

Appendix A. Proof of technical Lemmas

Proof of Lemma 2.2. Recall that $\|h_d - h\|_n^2 = \frac{T}{n}\sum_{i=-n}^{n-1}(h_d(x_i) - h(x_i))^2$.

Proof of part (i). We write $\frac{T}{n}\sum_{i=-n}^{n-1}(h_d(x_i)-h(x_i))^2 = \frac{T}{n}\sum_{i=-n}^{n-1}(h_d(x_i)-h(x_i))^2 - \int_{-T}^{T}(h-h_d)^2(u)du + \int_{-T}^{T}(h-h_d)^2(u)du$. Using Lemma B.2 given in the Appendix yields

$$\left|\frac{T}{n}\sum_{i=-n}^{n-1}(h_d(x_i)-h(x_i))^2 - \int_{-T}^{T}(h-h_d)^2(u)du\right| \leqslant \|\psi'\|_\infty \frac{T^2}{n},$$

where $\psi(x) = (\sum_{j\geqslant d} a_j(h)\varphi_j(x))^2$. Using (5), (8) and the Cauchy-Schwarz inequality, we have for $h \in W_H^\alpha(L)$ that

$$\sum_{j\geqslant d} a_j(h)\varphi_j'(x) \leqslant \left(\sum_{j\geqslant d} j^\alpha |a_j(h)|^2\right)^{\frac12}\left(\sum_{j\geqslant d} j^{-\alpha+\frac56}\right)^{\frac12} \lesssim \left(d^{-\alpha+\frac56+1}\right)^{\frac12} = d^{-\frac{\alpha}{2}+\frac{11}{12}},$$

provided $-\alpha+5/6+1<0$, that is $\alpha > 11/6$. Then, ψ is differentiable and $\psi'(x) = 2\sum_{j\geqslant d} a_j(h)\varphi_j'(x)\sum_{j\geqslant d} a_j(h)\varphi_j(x)$. Again, using (5) and the Cauchy-Schwarz inequality, we have for $h\in W_H^\alpha(L)$ that

$$\sum_{j\geqslant d}|a_j(h)\varphi_j(x)| \leqslant \sum_{j\geqslant d} j^{\frac{\alpha}{2}}|a_j(h)|j^{-\frac{\alpha}{2}-\frac{1}{12}} \leqslant \left(\sum_{j\geqslant d} j^\alpha|a_j(h)|^2\right)^{\frac12}\left(\sum_{j\geqslant d} j^{-\alpha-\frac16}\right)^{\frac12} \lesssim d^{-\frac{\alpha}{2}+\frac{5}{12}}.$$

Consequently, it follows for $\alpha > 11/6$ that $\frac{T}{n}\sum_{i=-n}^{n-1}(h_d(x_i)-h(x_i))^2 - \int_{-T}^{T}(h-h_d)^2(u)du \leqslant C\frac{T^2}{n}$ and therefore $\|h-h_d\|_n^2 \leqslant C(\alpha,L)\frac{T^2}{n} + \|h-h_d\|^2$. This gives the part (i).

Proof of part (ii). Let us start by writing

$$\begin{aligned}
\|h-h_d\|_n^2 &= \frac{T}{n}\sum_{i=-n}^{n-1}(h-h_d)^2(x_i)\\
&\leqslant 2\frac{T}{n}\sum_{i=-n}^{n-1}\left[\frac{(h-h_d)^2(x_i)+(h-h_d)^2(x_{i+1})}{2}\right]\\
&= 2\frac{T}{n}\sum_{i=-n}^{n-1}\left[\frac{(h-h_d)^2(x_i)+(h-h_d)^2(x_{i+1})}{2}\right] - 2\int_{-T}^{T}(h-h_d)^2(x)dx + 2\int_{-T}^{T}(h-h_d)^2(x)dx.
\end{aligned}$$

From Lemma B.2 (ii) given in Appendix, we have

$$\left|\frac{T}{n}\sum_{i=-n}^{n-1}\left[\frac{(h-h_d)^2(x_i)+(h-h_d)^2(x_{i+1})}{2}\right] - \int_{-T}^{T}(h-h_d)^2(x)dx\right| \leqslant \|\psi''\|_\infty\frac{T^3}{12n^2},$$

where $\psi(x) = (h-h_d)^2(x) = (\sum_{j>d} a_j(h)\varphi_j(x))^2$ with $a_j(h) = \langle h, \varphi_j\rangle$. Next, we evaluate the term $\|\psi''\|_\infty$. By induction on d, the d-th derivative of φ_j is given by (see Lemma 5.2 in Comte et al. (2020) for the proof)

$$\varphi_j^{(d)} = \sum_{k=-d}^{d} b_{k,j}^{(d)} \varphi_{j+k}, \quad \text{where} \quad b_{k,j}^{(d)} = \mathcal{O}(j^{d/2}), \quad j \geqslant d \geqslant |k|.$$

Using this for $d = 2$ and (5), it follows $|\varphi_j''(x)| \lesssim j(j+k)^{-\frac{1}{12}} \lesssim j^{\frac{11}{12}}$ and then we get for $W_H^\alpha(L)$ and $\alpha > 17/6$

$$|\sum_{j>d} a_j(h)\varphi_j''(x)| \leqslant \left(\sum_{j>d} j^\alpha a_j(h)^2\right)^{\frac{1}{2}} \left(\sum_{j>d} j^{-\alpha+\frac{11}{6}}\right)^{\frac{1}{2}} \lesssim (d^{-\alpha+\frac{11}{6}+1})^{\frac{1}{2}} = d^{-\frac{\alpha}{2}+\frac{17}{12}}.$$

This implies that ψ is differentiable of order 2. Then, for any $j > d$, it holds

$$\psi''(x) = 2\left[\sum_{j>d} a_j(h)\varphi_j''(x) \sum_{j\geqslant d} a_j(h)\varphi_j(x) + \left(\sum_{j>d} a_j(h)\varphi_j'(x)\right)^2\right],$$

where the bound of last term is $d^{-\alpha+\frac{11}{6}}$ for $h \in W_H^\alpha(L)$ (see Proof of part (i)). Besides, the order of $\sum_{j\geqslant d} a_j(h)\varphi_j(x)$ is $d^{-\frac{\alpha}{2}+\frac{5}{12}}$. Therefore, it comes $\|\psi''\|_\infty \lesssim d^{-\alpha+\frac{11}{6}}$ and then

$$\|h-h_d\|_n^2 \leqslant 2\|h-h_d\|^2 + C\frac{T^3}{12n^2}.$$

This ends the proof of part (ii) and then the proof of Lemma. □

Proof of Lemma 6.1. Consider the process $\nu_n(t) = \langle t, \widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\rangle$. Let us denote by $\mathcal{S}_d := \{t \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}),\ \text{supp}\ (t^*) \subset [-\sqrt{2d}, \sqrt{2d}]\}$. We have, $|\nu_n(t)|^2 \leqslant \|t\|^2 \|\widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\|^2$ with equality in $t = \widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]/(\widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}])$, then, it holds

$$\|\widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\|^2 = \sup_{t\in\mathcal{S}_d, \|t\|=1} |\nu_n(t)|^2.$$

By definition of $\widehat{h}_d$ given in (13), we write,

$$\begin{aligned}\nu_n(t) =& \langle t, \widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\rangle = \frac{1}{2\pi}\int_{-\sqrt{2d}}^{\sqrt{2d}} \frac{\widehat{h}_d^*(u) - \mathbb{E}[\widehat{h}_d^*(u)]}{g^*(u)} t^*(-u)du \\ =& \frac{1}{2\pi}\frac{T}{n}\int_{-\sqrt{2d}}^{\sqrt{2d}} \frac{\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\vec{\varepsilon}\right]_j \varphi_j^*(u)}{g^*(u)} t^*(-u)du\end{aligned}$$

Using that $\left[\Psi_d^{-1}\Phi_d^t\vec{\varepsilon}\right]_{0\leqslant j\leqslant d-1} = \left[\sum_{i=-n}^{n-1}[\Psi_d^{-1}\Phi_d^t]_{j,i}\varepsilon_i\right]_{0\leqslant j\leqslant d-1}$, it holds

$$\nu_n(t) = \frac{1}{2\pi}\frac{T}{n}\sum_{i=-n}^{n-1}\varepsilon_i\langle t^*, \frac{\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\varphi_j^*}{g^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{2d}}\rangle = \frac{1}{2n}\sum_{i=-n}^{n-1}\alpha_{t,d,i}(\varepsilon_i)$$

where

$$\alpha_{t,d,i}(x) = x\frac{T}{\pi}\langle t^*, \frac{\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\varphi_j^*}{g^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{2d}}\rangle.$$

As the noise is not bounded, we cannot apply directly the Talagrand's inequality to the process $\nu_n(t)$. In this respect, we use the following decomposition

$$\varepsilon_i = \zeta_i + \xi_i, \quad \zeta_i = \varepsilon_i \mathbb{1}_{|\varepsilon_i| \leqslant k_n} - \mathbb{E}[\varepsilon_i \mathbb{1}_{|\varepsilon_i| \leqslant k_n}], \qquad \xi_i = \varepsilon_i \mathbb{1}_{|\varepsilon_i| > k_n} - \mathbb{E}[\varepsilon_i \mathbb{1}_{|\varepsilon_i| > k_n}],$$

where k_n is chosen in the sequel. Then, it follows that

$$\nu_n(t) = \nu_n^{(1)}(t) + \nu_n^{(2)}(t), \quad \nu_n^{(1)}(t) = \frac{1}{2n} \sum_{i=-n}^{n-1} \alpha_{t,d,i}(\zeta_i), \quad \nu_n^{(2)}(t) = \frac{1}{2n} \sum_{i=-n}^{n-1} \alpha_{t,d,i}(\xi_i),$$

and

$$\mathbb{E}\left[\left(\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n(t)|^2 - \frac{\kappa_1}{6} V(d)\right)_+\right] \leqslant 2\,\mathbb{E}\left[\left(\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(1)}(t)|^2 - \frac{\kappa_1}{12} V(d)\right)_+\right] + 2\,\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(2)}(t)|^2\right].$$

This implies that

$$\sum_{d \in \mathcal{M}_n^{(1)}} \mathbb{E}\left[\left(\|\widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\|^2 - \frac{\kappa_1}{6} V(d)\right)_+\right] \leqslant 2 \sum_{d \in \mathcal{M}_n^{(1)}} \mathbb{E}\left[\left(\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(1)}(t)|^2 - \frac{\kappa_1}{12} V(d)\right)_+\right] + 2n\,\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(2)}(t)|^2\right]. \tag{64}$$

Now, we study the last two terms. We start by the second.

Upper bound for $n\,\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(2)}(t)|^2\right]$. For $t \in \mathcal{S}_d$, we remark that

$$\nu_n^{(2)}(t) = \frac{1}{2\pi} \frac{T}{n} \int_{-\sqrt{2d}}^{\sqrt{2d}} \frac{\sum_{j=0}^{d-1} \left[\Psi_d^{-1} \Phi_d^t \vec{\xi}\right]_j \varphi_j^*(u)}{g^*(u)} t^*(-u) du = \frac{1}{2\pi} \int_{-\sqrt{2d}}^{\sqrt{2d}} \frac{\check{h}_d^*(u) - \mathbb{E}[\check{h}_d^*(u)]}{g^*(u)} t^*(-u) du,$$

where (see Equation (13)) $\check{h}_d = \sum_{j=0}^{d-1} \check{b}_j^{(d)} \varphi_j$, $\vec{\check{b}}^{(d)} = (\check{b}_0^{(d)}, \ldots, \check{b}_{d-1}^{(d)})^t = (\Phi_d^t \Phi_d)^{-1} \Phi_d^t \vec{y} = \frac{T}{n} \Psi_d^{-1} \Phi_d^t \vec{\check{y}}$, $\vec{\check{y}} = (\check{y}(x_{-n}), \ldots, \check{y}(x_{n-1}))^t$ with $\check{y}(x_i) = h(x_i) + \xi_i$, here and only for the study of $n\,\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(2)}(t)|^2\right]$. It comes that

$$\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(2)}(t)|^2\right] \leqslant \|t\|^2 \Delta(\sqrt{2d})\,\mathbb{E}\left[\|\check{h}_d - \mathbb{E}[\check{h}_d]\|^2\right].$$

The bounds obtained for $\widehat{h}_d$ extend to $\check{h}_d$. Then, it yields from (48) (with σ_ε^2 replaced by $\mathbb{E}[\xi_1^2]$) and under **(A0)** that $\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(2)}(t)|^2\right] \leqslant \Delta(\sqrt{2d}) \lambda d\, \mathbb{E}[\xi_1^2] \frac{T}{n}$. By the Cauchy-Schwarz inequality, we have

$$\mathbb{E}[\xi_1^2] \leqslant \mathbb{E}[\varepsilon_1^2 \mathbb{1}_{|\varepsilon_1| \geqslant k_n}] \leqslant \sqrt{\mathbb{E}[\varepsilon_1^4]} \sqrt{\mathbb{P}(|\varepsilon_1| \geqslant k_n)}.$$

We introduce the following technical Lemma to obtain a bound of $\mathbb{E}[\xi_1^2]$.

Lemma A.1. *Under (A5), it yields* $\mathbb{P}(|\varepsilon_1| \geqslant k_n) \leqslant 2e^{-\frac{k_n^2}{2b^2}}$. *Moreover,* ε_1 *admits a finite moment of any order,* $\mathbb{E}[|\varepsilon_1|^p] \leqslant (2b^2)^{\frac{p}{2}}]p\Gamma(\frac{p}{2})$, *where* $\Gamma(\cdot)$ *denotes the gamma function defined by:*

$$\Gamma(t) = \int_0^{+\infty} x^{t-1}e^{-x}dx, \quad \forall t \in \mathbb{R}.$$

Using Lemma A.1 with $p = 4$ and choosing

$$k_n = 2\sqrt{2}b\sqrt{\log(n)}, \tag{65}$$

we get

$$n\,\mathbb{E}\left[\sup_{t\in\mathcal{S}_d,\|t\|=1} |\nu_n^{(2)}(t)|^2\right] \leqslant n\Delta(\sqrt{2d})\lambda d\sqrt{\mathbb{E}[\varepsilon_1^4]}\sqrt{\mathbb{P}(|\varepsilon_1| \geqslant k_n)}\frac{T}{n} \leqslant \frac{C}{n}, \tag{66}$$

since $\Delta(\sqrt{2d})\lambda dT \lesssim n$ by definition of $\mathcal{M}_n^{(1)}$.

Upper bound for $\sum_{d\in\mathcal{M}_n^{(1)}} \mathbb{E}\left[\left(\sup_{t\in\mathcal{S}_d,\|t\|=1} |\nu_n^{(1)}(t)|^2 - \frac{\kappa_1}{12}V(d)\right)_+\right]$. We bound this term applying the Talagrand inequality given in Appendix C.2. Let us first compute the three constants H^2, M_1 and v.

Computing of H^2**.** Similarly to the study of $\mathbb{E}\left[\sup_{t\in\mathcal{S}_d,\|t\|=1} |\nu_n^{(2)}(t)|^2\right]$, we have under (A0) and from (48),

$$\mathbb{E}\left[\sup_{t\in\mathcal{S}_d,\|t\|=1} |\nu_n^{(1)}(t)|^2\right] \leqslant \lambda\,\mathbb{E}[\zeta_1^2]\Delta(\sqrt{2d})\frac{dT}{n} \leqslant \lambda\sigma_\varepsilon^2\Delta(\sqrt{2d})\frac{dT}{n} := H^2.$$

Computing of v**.** For $t \in \mathcal{S}_d$, it holds by the Cauchy-Schwarz inequality, $\mathbb{E}[\zeta_1^2] \leqslant \sigma_\varepsilon^2$ and as $\|t\|^2 = 1$,

$$\begin{aligned}
\frac{1}{2n}\sum_{i=-n}^{n-1}\mathrm{Var}(\alpha_{t,d,i}(\zeta_i)) =& \frac{1}{2n}\sum_{i=-n}^{n-1}\mathbb{E}\left[\alpha_{t,d,i}(\zeta_i)\overline{\alpha_{t,d,i}(\zeta_i)}\right] \\
=& \sigma_\varepsilon^2\frac{T^2}{2n\pi^2}\sum_{i=-n}^{n-1}\left|\left\langle t^*, \frac{\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\varphi_j^*}{g^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{2d}}\right\rangle\right|^2 \\
\leqslant& \sigma_\varepsilon^2\frac{T^2}{2n\pi^2}\sum_{i=-n}^{n-1}\|t^*\|^2\int_{-\sqrt{2d}}^{\sqrt{2d}}\frac{\left|\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\varphi_j^*(u)\right|^2}{|g^*(u)|^2}du \\
=& \sigma_\varepsilon^2\frac{T^2}{n\pi}\Delta(\sqrt{2d})\sum_{i=-n}^{n-1}\int_{-\sqrt{2d}}^{\sqrt{2d}}\left|\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\varphi_j^*(u)\right|^2 du.
\end{aligned}$$

The Fourier transform of (φ_j), see (6) gives,

$$\begin{aligned}
\frac{1}{2n}\sum_{i=-n}^{n-1}\mathrm{Var}(\alpha_{t,d,i}(\zeta_i)) &\leqslant 2\sigma_\varepsilon^2\frac{T^2}{n}\Delta(\sqrt{2d})\sum_{i=-n}^{n-1}\int_{\mathbb{R}}|\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\varphi_j(u)|^2du \\
&= 2\sigma_\varepsilon^2\frac{T^2}{n}\Delta(\sqrt{2d})\sum_{i=-n}^{n-1}\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2
\end{aligned}$$

where we use the orthonormality (φ_j). Recall that for $A=(a_{i,j})_{1\leqslant i\leqslant m,1\leqslant j\leqslant n}$ a matrix with real coefficients, the Frobenius norm of A is defined by

$$\|A\|_F^2=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{i,j}^2=\mathrm{tr}\left[A^tA\right].$$

Then, under **(A0)**, it yields

$$\frac{T}{n}\sum_{i=-n}^{n-1}\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2=\frac{T}{n}\mathrm{tr}\left[\Phi_d\Psi_d^{-1}\Psi_d^{-1}\Phi_d^t\right]=\mathrm{tr}\left[\Psi_d^{-1}\right]\leqslant\lambda d,$$

which implies

$$\sup_{t\in\mathcal{S}_d,\|t\|=1}\frac{1}{2n}\sum_{i=-n}^{n-1}\mathrm{Var}(\alpha_{t,d,i}(\zeta_i))\leqslant 2\sigma_\varepsilon^2T\lambda d\Delta(\sqrt{2d})=:v.$$

Computing of M_1. Using successively (6), the Cauchy-Schwarz inequality and the orthonormality of φ_j, we have on the process $\nu_n^{(1)}$

$$\begin{aligned}\sup_{t\in\mathcal{S}_d,\|t\|=1}\|\alpha_{t,d,i}\|_\infty&=\sup_{t\in\mathcal{S}_d,\|t\|=1}\sup_{x\in\mathbb{R}}|\alpha_{t,d,i}(x)|=\sup_{t\in\mathcal{S}_d,\|t\|=1}\sup_{x\in\mathbb{R}}\left|x\mathbb{1}_{x\leqslant k_n}\frac{T}{\pi}\langle t^*,\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\frac{\varphi_j^*}{g^*}\mathbb{1}_{|\cdot|\leqslant\sqrt{2d}}\rangle\right|\\&\leqslant\sup_{t\in\mathcal{S}_d,\|t\|=1}\left(2k_n\frac{T}{\pi}\|t^*\|\sqrt{2\pi}\left(\int_{-\sqrt{2d}}^{\sqrt{2d}}\frac{|\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\varphi_j(u)|^2}{|g^*(u)|^2}du\right)^{\frac12}\right)\\&\leqslant 4k_nT\left(\Delta(\sqrt{2d})\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2\right)^{\frac12}.\end{aligned}$$

To bound the term $\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2$, we use:

$$\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2=\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\left[\Psi_d^{-1}\Phi_d^t\right]_{i,j}=\left[\Phi_d\Psi_d^{-1}\Psi_d^{-1}\Phi_d^t\right]_{-n\leqslant i,i\leqslant n-1}=\vec{e}_i^{\,t}\Phi_d\Psi_d^{-1}\Psi_d^{-1}\Phi_d^t\vec{e}_i,$$

where $(\vec{e}_i)_{-n\leqslant i\leqslant n-1}$ is the vector of the canonical basis of $\mathbb{R}^{2n}$. The matrix Ψ_d^{-1} is a definite symmetric, then diagonalizable and we can write

$$\Psi_d^{-1}=PDP^t,\quad P^tP=PP^t=I_d,\quad D=\mathrm{Diag}(\mu_1,\dots,\mu_d),$$

where $(\mu_i)_{1\leqslant i\leqslant d}$ are the eigenvalues of matrix Ψ_d^{-1}. We can define its square root and we have for $\vec{w}=P^t\Psi_d^{-\frac12}\Phi_d^t\vec{e}_i$

$$\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2=\vec{e}_i^{\,t}\Phi_d\Psi_d^{-\frac12}\Psi_d^{-1}\Psi_d^{-\frac12}\Phi_d^t\vec{e}_i=\sum_{j=0}^{d-1}\mu_jw_j^2\leqslant\lambda_{max}(\Psi_d^{-1})\vec{w}^t\vec{w},$$

The definition of operator norm implies,

$$\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2\leqslant\lambda_{max}(\Psi_d^{-1})\sup_{\|\vec{x}\|=1}\left(\vec{x}^t\Phi_d\Psi_d^{-\frac12}\Psi_d^{-\frac12}\Phi_d^t\vec{x}\right)=\lambda_{max}(\Psi_d^{-1})\lambda_{max}(\Phi_d\Psi_d^{-1}\Phi_d^t).$$

Furthermore, the matrix $\Phi_d\Psi_d^{-1}\Phi_d^t = \frac{n}{T}\Phi_d(\Phi_d^t\Phi_d)^{-1}\Phi_d^t$ is an orthogonal projection matrix, then, it comes

$$\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2 \leqslant \lambda_{max}(\Psi_d^{-1})\frac{n}{T}. \tag{67}$$

Under (**A0**), we obtain

$$\sup_{t\in\mathcal{S}_d,\|t\|=1}\|\alpha_{t,d,i}\|_\infty \leqslant 4k_nT^{\frac{1}{2}}(\Delta(\sqrt{2d})\lambda n)^{\frac{1}{2}} := M_1.$$

For $\delta>0$, the Talagrand inequality gives,

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{S}_d,\|t\|=1}\left|\nu_n^{(1)}(t)\right|^2 - 2(1+2\delta)H^2\right)_+\right] \leqslant \frac{4}{K_1}(T_d+U_d),$$

where

$$T_d = \frac{2\sigma_\varepsilon^2\lambda T d\Delta(\sqrt{2d})}{n}\exp\left(-\frac{K_1\delta}{2}\right) \text{ and } U_d = \frac{196\lambda k_n^2 T\Delta(\sqrt{2d})}{K_1C^2(\delta)n}\exp\left(-K_1'C(\delta)\sqrt{\delta}\frac{\sqrt{\sigma_\varepsilon^2}}{k_n}\sqrt{d}\right),$$

$K_1=1/3$, $C(\delta)=(\sqrt{1+\delta}-1)\wedge 1$. It follows that

$$\sum_{d\in\mathcal{M}_n^{(1)}}\mathbb{E}\left[\left(\sup_{t\in\mathcal{S}_d,\|t\|=1}|\nu_n^{(1)}(t)|^2 - 2(1+2\delta)H^2\right)_+\right] \lesssim \sum_{d\in\mathcal{M}_n^{(1)}}\left[T_d+U_d\right].$$

As $Td\Delta(\sqrt{2d})\lesssim n$, then, it yields $\sum_{d\in\mathcal{M}_n^{(1)}}T_d \lesssim n\exp\left(-K_1\delta/2\right)$. The choice $\delta=\frac{4}{K_1}\log(n)$ ensures that $\sum_{d\in\mathcal{M}_n^{(1)}}T_d\leqslant\frac{C}{n}$. With this choice of δ and k_n given by (65), we derive $C(\delta)=1$ and

$$\sum_{d\in\mathcal{M}_n^{(1)}}U_d \leqslant C\frac{\log(n)}{n}\sum_{d\in\mathcal{M}_n^{(1)}}\Delta(\sqrt{2d})\exp\left(-C\sqrt{d}\right)\leqslant C\frac{\log(n)}{n},$$

since

$$\sum_{d\in\mathcal{M}_n^{(1)}}\left(\Delta(\sqrt{2d})\exp\left(-C\sqrt{d}\right)\right)\lesssim\sum_{d\in\mathcal{M}_n^{(1)}}d^\gamma\exp\left(-C\sqrt{d}\right)<\infty.$$

Finally, it holds for $\kappa_1\geqslant 12$ that

$$\sum_{d\in\mathcal{M}_n^{(1)}}\mathbb{E}\left[\left(\sup_{t\in\mathcal{S}_d,\|t\|=1}|\nu_n^{(1)}(t)|^2-\frac{\kappa_1}{12}V(d)\right)_+\right]\leqslant C\frac{\log(n)}{n}.$$

Plugging this and (66) in (64) concludes the proof. □

Proof of Lemma A.1. Let us prove the first bound. Using the Markov inequality, we have for any $t,s>0$

$$\mathbb{P}(\varepsilon_1>s)\leqslant\mathbb{P}(e^{t\varepsilon_1}>e^{st})\leqslant\frac{\mathbb{E}[e^{t\varepsilon_1}]}{e^{st}}\leqslant e^{\frac{b^2t^2}{2}-st},$$

where the last bound is obtained using the fact that ε_1 is b-sub-Gaussian. The above inequality holds for any $t>0$, then, for the t which minimizes the bound. Set $r(t)=\frac{b^2t^2}{2}-st$, we have $r'(t)=0$ in $t=s/b^2$ and $r''(t)>0$ for any $t>0$. It follows that $t=s/b^2$ is the minimizer of $r(t)$ and $\inf_{t\geqslant0}r(t)=-s^2/(2b^2)$ and then, $\mathbb{P}(\varepsilon_1>s)\leqslant e^{-\frac{s^2}{2b^2}}$. Likewise, it yields $\mathbb{P}(\varepsilon_1<-s)\leqslant e^{-\frac{s^2}{2b^2}}$. Consequently, we get $\mathbb{P}(|\varepsilon_1|>s)\leqslant$

$\mathbb{P}(\varepsilon_1 > s) + \mathbb{P}(\varepsilon_1 < -s) \leqslant 2e^{-\frac{s^2}{2b^2}}$. This prove the first part by setting $s = k_n$. For the second part, we have by the definition of the expectation for non negative variable

$$\begin{aligned}\mathbb{E}[|\varepsilon_1|^p] &= \int_0^{+\infty} \mathbb{P}(|\varepsilon_1|^p \geqslant x)dx \leqslant 2\int_0^{+\infty} e^{-\frac{x^{\frac{2}{p}}}{2b^2}}dx \\ &= (2b^2)^{\frac{p}{2}} p \int_0^{+\infty} e^{-y} y^{\frac{p}{2}-1} dy.\end{aligned}$$

Using the definition of the gamma function, we get $\mathbb{E}[|\varepsilon_1|^p] = (2b^2)^{\frac{p}{2}} p\Gamma(\frac{p}{2})$. □

Proof of Lemma 6.2. Define the linear process $\nu_n(s) = \langle s, \widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}]\rangle$. For all function s, it holds $\|\widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}]\|^2 = \sup_{s\in\mathcal{S}_m, \|s\|=1} |\nu_n(s)|^2$. By definition of $\widetilde{f}_m$ given in (41), we have

$$\begin{aligned}\nu_n(s) &= \int_{\mathbb{R}} s(u)(\widetilde{f}_{m'} - \mathbb{E}[\widetilde{f}_{m'}])(u)du = \int_{\mathbb{R}} \sum_{j=0}^{m-1} \big(\widehat{a}_{j,m} - \mathbb{E}[\widehat{a}_{j,m}]\big) s(u)\varphi_j(u)du \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \sum_{j=0}^{m-1} \left(\int_{\mathbb{R}} \frac{\widehat{h}^*_m(v) - \mathbb{E}[\widehat{h}^*_m(v)]}{g^*(v)} \varphi_j^*(v)dv\right) \varphi_j(u)s(u)du \\ &= \frac{1}{2\pi}\frac{T}{n}\sum_{i=-n}^{n-1} \varepsilon_i \int_{\mathbb{R}} \sum_{j=0}^{m-1} \left(\int_{\mathbb{R}} \frac{\sum_{k=0}^{m-1} \left[\Psi_m^{-1}\Phi_m^t\right]_{k,i} \varphi_k^*(v)}{g^*(v)} \varphi_j^*(v)dv\right) \varphi_j(u)s(u)du \\ &= \frac{1}{2\pi}\frac{T}{n}\sum_{i=-n}^{n-1} \varepsilon_i \left\langle s, \sum_{j=0}^{m-1} \varphi_j \left(\int_{\mathbb{R}} \frac{\sum_{k=0}^{m-1} \left[\Psi_m^{-1}\Phi_m^t\right]_{k,i} \varphi_k^*(v)}{g^*(v)} \varphi_j^*(v)dv\right)\right\rangle \\ &= \frac{1}{2n}\sum_{i=-n}^{n-1} \beta_{s,m,i}(\varepsilon_i)\end{aligned}$$

where

$$\beta_{s,m,i}(x) = x\frac{T}{\pi}\left\langle s, \sum_{j=0}^{m-1} \varphi_j \left(\int_{\mathbb{R}} \frac{\sum_{k=0}^{m-1} \left[\Psi_m^{-1}\Phi_m^t\right]_{k,i} \varphi_k^*(v)}{g^*(v)} \varphi_j^*(v)dv\right)\right\rangle.$$

As the noise is not necessarily bounded, we cannot used the Talagrand inequality directly to the process ν_n, then, we split the noise as follows:

$$\varepsilon_i = \zeta_i + \xi_i, \quad \zeta_i = \varepsilon_i \mathbb{1}_{|\varepsilon_i|\leqslant k_n} - \mathbb{E}[\varepsilon_i \mathbb{1}_{|\varepsilon_i|\leqslant k_n}], \quad \xi_i = \varepsilon_i \mathbb{1}_{|\varepsilon_i|> k_n} - \mathbb{E}[\varepsilon_i \mathbb{1}_{|\varepsilon_i|> k_n}],$$

where k_n is chosen in the sequel. Thus, it comes

$$\nu_n^{(1)}(s) = \nu_n^{(1)}(s) + \nu_n^{(2)}(s), \quad \nu_n^{(1)}(s) = \frac{1}{2n}\sum_{i=-n}^{n-1} \beta_{s,m,i}(\zeta_i), \quad \nu_n^{(2)}(s) = \frac{1}{2n}\sum_{i=-n}^{n-1} \beta_{s,m,i}(\xi_i),$$

and

$$\begin{aligned}\sum_{m\in\mathcal{M}_n^{(2)}} \mathbb{E}\left[\left(\|\widetilde{f}_m - \mathbb{E}[\widetilde{f}_m]\|^2 - \frac{\kappa_1'}{6}W(m)\right)_+\right] &\leqslant 2\sum_{m\in\mathcal{M}_n^{(2)}} \mathbb{E}\left[\left(\sup_{s\in\mathcal{S}_m,\|s\|=1} |\nu_n^{(1)}(s)|^2 - \frac{\kappa_1'}{12}W(m')\right)_+\right] \\ &\quad + 2n\,\mathbb{E}\left[\sup_{s\in\mathcal{S}_m,\|s\|=1} |\nu_n^{(2)}(s)|^2\right].\end{aligned} \tag{68}$$

We study separately the terms on the right-hand side of previous bound. Let us start the last term.

Bounding of $n\,\mathbb{E}\left[\sup_{s\in\mathcal{S}_m,\|s\|=1}|\nu_n^{(2)}(s)|^2\right]$. We first remark that:

$$\nu_n^{(2)}(s)=\frac{1}{2\pi}\int_{\mathbb{R}}\sum_{j=0}^{m-1}\left(\int_{\mathbb{R}}\frac{\breve{h}_m^*(v)-\mathbb{E}[\breve{h}_m^*(v)]}{g^*(v)}\varphi_j^*(v)dv\right)\varphi_j(u)s(u)du,$$

where where (see Equation (13)) $\breve{h}_d=\sum_{j=0}^{d-1}\breve{b}_j^{(d)}\varphi_j$, $\vec{\breve{b}}^{(d)}=(\breve{b}_0^{(d)},\dots,\breve{b}_{d-1}^{(d)})^t=(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\vec{y}=\frac{T}{n}\Psi_d^{-1}\Phi_d^t\breve{y}$, $\vec{\breve{y}}=(\breve{y}(x_{-n}),\dots,\breve{y}(x_{n-1}))^t$ with $\breve{y}(x_i)=h(x_i)+\xi_i$, only here. The Cauchy Schwarz inequality implies

$$|\nu_n^{(2)}(s)|\leqslant\frac{1}{4\pi^2}\|s\|^2\left\|\sum_{j=0}^{m-1}\langle\frac{\breve{h}_m^*-\mathbb{E}[\breve{h}_m^*]}{g^*},\varphi_j^*\rangle\varphi_j\right\|^2=\frac{1}{\pi}\sum_{j=0}^{m-1}\left|\langle\frac{\breve{h}_m^*-\mathbb{E}[\breve{h}_m^*]}{g^*},\varphi_j\rangle\right|^2.$$

Splitting $\int_{\mathbb{R}}(.)=\int_{|\cdot|\leqslant\sqrt{\rho m}}(\cdot)+\int_{|\cdot|>\sqrt{\rho m}}(\cdot)$, using Bessel, the Cauchy Schwarz inequalities and Parseval equality, we derive $|\nu_n^{(2)}(s)|\leqslant 2\|\breve{h}_m-\mathbb{E}[\breve{h}_m]\|^2\Sigma(m)$. Recall that the bounds obtained for $\widehat{h}_m$ remain valid for $\breve{h}_m$. In particular, it holds from (48) with $\mathbb{E}[\xi_1^2]$ plays the role of σ_ε^2 and under **(A0)**, $\mathbb{E}\left[\|\breve{h}_m-\mathbb{E}[\breve{h}_m]\|^2\right]\leqslant\lambda m\,\mathbb{E}[\xi_1^2]\frac{T}{n}$, which implies $\mathbb{E}\left[\sup_{s\in\mathcal{S}_m,\|s\|=1}|\nu_n^{(2)}(s)|^2\right]\leqslant 2\Sigma(m)\lambda m\,\mathbb{E}[\xi_1^2]\frac{T}{n}\leqslant\frac{\mathbb{E}[\xi_1^2]}{\sigma_\varepsilon^2}$ since $2\Sigma(m)\lambda m\frac{T}{n}\leqslant\frac{1}{\sigma_\varepsilon^2}$ by definition of $\mathcal{M}_n^{(2)}$. Next, by the Cauchy Schwarz inequality and as ε_1 is sub-Gaussian, we derive from Lemma A.1

$$\mathbb{E}[\xi_1^2]=\mathbb{E}[\varepsilon_1^2\mathbb{1}_{|\varepsilon_1|\geqslant k_n}]\leqslant\sqrt{\mathbb{E}[\varepsilon_1^4]}\sqrt{\mathbb{P}(|\varepsilon_1|\geqslant k_n)}\leqslant 4\sqrt{2}b^2e^{-\frac{k_n^2}{4b^2}},$$

where $b>0$ is a constant given in **(A5)**. Choosing

$$k_n=2b\sqrt{2}\log(n),\tag{69}$$

we deduce

$$n\,\mathbb{E}\left[\sup_{s\in\mathcal{S}_m,\|s\|=1}|\nu_n^{(2)}(s)|^2\right]\leqslant\frac{C}{n}.\tag{70}$$

Bounding of $\sum_{m\in\mathcal{M}_n^{(2)}}\mathbb{E}\left[\left(\sup_{s\in\mathcal{S}_m,\|s\|=1}|\nu_n^{(1)}(s)|^2-\frac{\kappa_1'}{12}W(m)\right)_+\right]$. We apply the Talagrand inequality given in Appendix C.2. We must compute three constant H^2, M_1 and v.

Computing H^2. Analogously to the study of $n\,\mathbb{E}\left[\sup_{s\in\mathcal{S}_m,\|s\|=1}|\nu_n^{(2)}(s)|^2\right]$, we derive under **(A0)** and as $\mathbb{E}[\zeta_1^2]\leqslant\sigma_\varepsilon^2$

$$\mathbb{E}\left[\sup_{s\in\mathcal{S}_m,\|s\|=1}|\nu_n^{(1)}(s)|^2\right]\leqslant 2\Sigma(m)\lambda m\sigma_\varepsilon^2\frac{T}{n}:=H^2.$$

Computing of v. For all $s\in\mathcal{S}_m$, we have by the Cauchy-Schwarz inequality, $\mathbb{E}[\zeta_1^2]\leqslant\sigma_\varepsilon^2$, as $\varphi_j^*=$

$\sqrt{2\pi}(i)^j\varphi_j$ (see (6)) and $\|s\|^2=1$

$$\begin{aligned}\frac{1}{2n}\sum_{i=-n}^{n-1}\mathrm{Var}(\beta_{s,m,i}(\zeta_i)) =&\frac{1}{2n}\sum_{i=-n}^{n-1}\mathbb{E}\left[\left|\frac{T}{\pi}\zeta_i\left\langle s,\sum_{j=0}^{m-1}\varphi_j\left(\int_{\mathbb{R}}\frac{\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}\varphi_k^*(v)}{g^*(v)}\varphi_j^*(v)dv\right)\right\rangle\right|^2\right]\\ \leqslant&\frac{1}{2n}\sum_{i=-n}^{n-1}\frac{T^2}{\pi^2}\sigma_\varepsilon^2\|s\|^2\left\|\sum_{j=0}^{m-1}\varphi_j\left(\int_{\mathbb{R}}\frac{\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}\varphi_k^*(v)}{g^*(v)}\varphi_j^*(v)dv\right)\right\|^2\\ \leqslant&\frac{T^2}{n\pi}\sigma_\varepsilon^2\sum_{i=-n}^{n-1}\sum_{j=0}^{m-1}\left|\int_{\mathbb{R}}\frac{\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}\varphi_k^*(v)}{g^*(v)}\varphi_j(v)dv\right|^2\end{aligned}$$

By splitting $\int_{\mathbb{R}}$ into $\int_{|\cdot|\leqslant\sqrt{\rho m}}(\cdot)+\int_{|\cdot|>\sqrt{\rho m}}(\cdot)$, the Bessel and Cauchy-Schwarz inequalities, $\varphi_j^*=\sqrt{2\pi}(i)^j\varphi_j$, it yields

$$\sum_{j=0}^{m-1}\left|\int_{\mathbb{R}}\frac{\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}\varphi_k^*(v)}{g^*(v)}\varphi_j(v)dv\right|^2\leqslant 4\pi\Sigma(m)\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}^2, \tag{71}$$

which implies

$$\frac{1}{2n}\sum_{i=-n}^{n-1}\mathrm{Var}(\beta_{s,m,i}(\zeta_i))\leqslant 4\frac{T^2}{n}\sigma_\varepsilon^2\Sigma(m)\sum_{i=-n}^{n-1}\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}^2.$$

By definition of Frobenius norm, it holds under **(A0)**

$$\frac{T}{n}\sum_{i=-n}^{n-1}\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}^2=\frac{T}{n}\mathrm{tr}\left[\Phi_m\Psi_m^{-1}\Psi_m^{-1}\Phi_m^t\right]=\mathrm{tr}\left[\Psi_m^{-1}\right]\leqslant\lambda m.$$

Therefore, we get

$$\sup_{s\in\mathcal{S}_m,\|s\|=1}\frac{1}{2n}\sum_{i=-n}^{n-1}\mathrm{Var}(\beta_{s,m,i}(\zeta_i))\leqslant 4\lambda\sigma_\varepsilon^2T\Sigma(m)m:=v.$$

Computing M_1. Using the Cauchy-Schwarz inequality, (6) and from (71), it holds

$$\begin{aligned}\sup_{s\in\mathcal{S}_m,\|s\|=1}\|\beta_{s,m,i}\|_\infty\leqslant&\sup_{s\in\mathcal{S}_m,\|s\|=1}\left\{2k_n\frac{T}{\pi}\|s\|\left(\sum_{j=0}^{m-1}\left|\int_{\mathbb{R}}\frac{\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}\varphi_k^*(v)}{g^*(v)}\varphi_j(v)dv\right|^2\right)^{\frac12}\right\}\\ \leqslant&\left\{4k_n\frac{T}{\sqrt{\pi}}\left(\Sigma(m)\sum_{k=0}^{m-1}\left[\Psi_m^{-1}\Phi_m^t\right]_{k,i}^2\right)^{\frac12}\right\}.\end{aligned}$$

Under **(A0)** and from (67), we derive

$$\sup_{s\in\mathcal{S}_m,\|s\|=1}\|\beta_{s,m,i}\|_\infty\leqslant\frac{4}{\sqrt{\pi}}k_n\big(\lambda T\Sigma(m)n\big)^{\frac12}:=M_1.$$

Applying the Talagrand inequality, we have for all $\delta>0$,

$$\sum_{m\in\mathcal{M}_n^{(2)}}\mathbb{E}\left[\left(\sup_{s\in\mathcal{S}_m,\|s\|=1}|\nu_n^{(1)}(s)|^2-2(1+2\delta)H^2\right)_+\right]\leqslant\frac{4}{K_1}\left\{\sum_{m\in\mathcal{M}_n^{(2)}}T(m)+\sum_{m\in\mathcal{M}_n^{(2)}}V(m)\right\},$$

where

$$T(m) = \frac{4\lambda\sigma_\varepsilon^2 T m \Sigma(m)}{n} e^{-\frac{K_1\delta}{2}}, \quad U(m) = \frac{784 k_n \lambda T \Sigma(m)}{\pi^2 K_1 C(\delta)^2 n} e^{-K_1' C(\delta)\sqrt{\delta}\frac{\sqrt{2\pi\sigma_\varepsilon^2 m}}{4k_n}},$$

$C(\delta) = (\sqrt{1+\delta}-1)\wedge 1$, $K_1 = 1/3$ and $K_1' > 0$ is a universal constant. By definition of $\mathcal{M}_n^{(2)}$ and choosing $\delta = 4\log(n)/K_1$, it yields $\frac{4}{K_1}\sum_{m\in\mathcal{M}_n^{(2)}} T(m) \leqslant C/n$. From this choice of δ and k_n given in (65), we deduce from **(A1)**

$$\frac{4}{K_1}\sum_{m\in\mathcal{M}_n^{(2)}} U(m) \lesssim \frac{\log(n)}{n}\sum_{m\in\mathcal{M}_n^{(2)}} m^\gamma e^{-C(\sigma,k_1')\sqrt{m}} \leqslant C\frac{\log(n)}{n}.$$

Consequently, we get for $\kappa_1' \geqslant 12$

$$\sum_{m\in\mathcal{M}_n^{(2)}} \mathbb{E}\left[\left(\sup_{s\in\mathcal{S}_m,\|s\|=1} |\nu_n^{(1)}(s)|^2 - \frac{\kappa_1'}{12}W(m)\right)_+\right] \leqslant C\frac{\log(n)}{n}.$$

Plugging this and (70) in (68) ends the proof. □

Appendix B. Study of $\mathrm{tr}(\Psi_d)$ and discussion on Assumption (A0)

In this section, T depends on d. For n large, we have $\mathrm{tr}\left(\Psi_d^{-1}\right) \asymp d$. Indeed, we can prove the following Lemma:

Lemma B.1. *Assume that $T \geqslant \sqrt{2d-1}$, we have*

$$|||\Psi_d - I_d||| \leqslant C_1 e^{-2\xi T^2} d + \phi_0 d^{\frac{17}{12}}\frac{T^2}{n}, \tag{72}$$

where C_1 depends on ξ, C_∞' given in (7) and $|||\cdot|||$ is any matrix norm.

Then, for the choice $d = [n^\omega]$ such that $\omega = 12/17 - \eta$, with $0 < \eta < 12/17$ and $T \asymp \sqrt{2d-1}$, we have $|||\Psi_d - I_d|||^2 \underset{n\to+\infty}{\longrightarrow} 0$. It follows that $|||\Psi_d - I_d||| \leqslant 1/2$ for n large enough. Using Theorem C.1 (see Stewart and Sun (1990)), we get

$$|||\Psi_d^{-1} - I_d||| \leqslant \frac{|||\Psi_d - I_d|||\,|||I_d|||^2}{1 - |||\Psi_d - I_d|||}.$$

This implies

$$|||\Psi_d^{-1} - I_d|||^2 \underset{n\to+\infty}{\longrightarrow} 0.$$

Thus, for n large enough **(A0)** holds and is not a strong condition.

In Table 4 and 5, we report the matrix norm of $\Psi_d - I_d$ and $\Psi_d^{-1} - I_d$ for

$$|||A||| = |||A|||_1 = \max_{1\leqslant j\leqslant n}\sum_{i=1}^n |A_{ij}|, \quad A = (A_{ij})_{1\leqslant i,j\leqslant n}. \tag{73}$$

Comment on Table 4 and 5. Globally, we see that increasing n makes the norm smaller but on the other hand the increase of d increases the norm. This is in accordance with the theory. Indeed in (72), we observe that for d large enough, it is the second term that determines the precision of these two norms. The increase with d of the norms is thus excepted. The results of Table 5 are better than those of Table 4. This is due to the choice $T = 10$ larger than $T = \sqrt{2d-1}$ for the choices of n et d given in Table 4 (for instance for $n = 1000$, $d = [n^{1/2}] = 31$, we have $T \approx 7.81$). Lastly, the norm $|||\Psi_d - I_d|||$ is smaller than $|||\Psi_d^{-1} - I_d|||$.

d \ n	100	500	1000
$[n^{1/2}]$	0.094 (0.101)	0.082 (0.087)	0.079 (0.084)
	10 (4.359)	22 (6.557)	31 (7.810)
$[n^{1/3}]$	0.103 (0.114)	0.094 (0.101)	0.090 (0.097)
	4 (2.646)	7 (3.606)	9 (4.123)
$[n^{1/4}]$	0.109 (0.121)	0.102 (0.111)	0.098 (0.107)
	3 (2.236)	4 (2.646)	5 (3)

TABLE 4. First line: Matrix norm of $A-I_d$ with $A=\Psi_d$ without parentheses and $A=\Psi_d^{-1}$ in parentheses for $T=\sqrt{2d-1}$. Second line: values of d with T in parentheses.

d \ n	100	500	1000
$[n^{1/2}]$	9.19e-16 (9.19e-16)	2.03e-15 (2.14e-15)	9.57e-11 (9.57e-11)
$[n^{1/3}]$	5.02e-16 (5.15e-16)	6.07e-16 (6.07e-16)	4.84e-16 (4.84e-16)
$[n^{1/4}]$	7.29e-16 (8.40e-16)	5.00e-16 (5.00e-16)	3.45e-16 (3.45e-16)

TABLE 5. Matrix norm of $A-I_d$ with $A=\Psi_d$ without parentheses and $A=\Psi_d^{-1}$ in parentheses for $T=10$.

Proof of Lemma B.1. We prove the result only for the particular norm defined in (73) but the result is valid for any matrix norm since we are in finite dimension. The general term of $(\Psi_d - I_d)$ is

$$\left(\frac{T}{n}\Phi_d^T\Phi_d - I_d\right)_{j,k} = \left(\frac{T}{n}\sum_{i=-n}^{n}\varphi_j(iT/n)\varphi_k(iT/n) - \int\varphi_j(u)\varphi_k(u)du\right)_{0\leqslant j,k\leqslant d-1}$$

For $0\leqslant j,k\leqslant d-1$, we write

$$\begin{aligned}\frac{T}{n}\sum_{i=-n}^{n-1}\varphi_j(iT/n)\varphi_k(iT/n) - \int\varphi_j(u)\varphi_k(u)du =& \frac{T}{n}\sum_{i=-n}^{n-1}\varphi_j(iT/n)\varphi_k(iT/n) - \int_{-T}^{T}\varphi_j(u)\varphi_k(u)du\\ &- \int_{|u|\geqslant T}\varphi_j(u)\varphi_k(u)du.\end{aligned}$$

Using Lemma B.2, we get

$$\left|\frac{T}{n}\sum_{i=-n}^{n-1}\varphi_j(iT/n)\varphi_k(iT/n) - \int_{-T}^{T}\varphi_j(u)\varphi_k(u)du\right| \leqslant \|(\varphi_j\varphi_k)'\|_\infty\frac{T^2}{n} \leqslant \phi_0 d^{\frac{5}{12}}\frac{T^2}{n}.$$

From (7) and as $T\geqslant\sqrt{2d-1}$, we have

$$|\int_{|x|\geqslant T}\varphi_j(x)\varphi_k(x)dx| \leqslant \int_{|x|\geqslant T}|\varphi_j(x)\varphi_k(x)|dx \leqslant C_\infty'^2 e^{-\xi T^2}\int e^{-\xi x^2}dx \leqslant C_1 e^{-\xi T^2},$$

where C_1 is a positive constant since $\int e^{-\xi x^2}dx<+\infty$. It comes

$$\|\!|\Psi_d - I_d|\!\|_1 \leqslant d\left[C_1 e^{-\xi T^2} + \phi_0 d^{\frac{5}{12}}\frac{T^2}{n}\right]. \tag{74}$$

□

B.1. **Estimating error in Riemann sums.** We give in this section the approximate errors of Riemann sum.

Lemma B.2. *Let $n \geqslant 1$, $T > 0$, $(x_i = iT/n)_{-n \leqslant i \leqslant n-1}$. Then,*

(i) For ψ be a function of class C^1 on $[-T, T]$, we have

$$\left| \frac{T}{n} \sum_{i=-n}^{n-1} \psi(x_i) - \int_{-T}^{T} \psi(x)dx \right| \leqslant \|\psi'\|_\infty \frac{T^2}{n}.$$

(ii) For ψ be a function of class C^2 on $[-T, T]$,

$$\left| \frac{T}{n} \sum_{i=-n}^{n-1} \frac{\psi(x_i) + \psi(x_{i+1})}{2} - \int_{-T}^{T} \psi(x)dx \right| \leqslant \|\psi''\|_\infty \frac{T^3}{12n^2}.$$

Proof of Lemma B.2. These proof are very classic when we approximate an integral by Rieman's sum.

Proof of part (i). By Chasles's relation, it yields

$$\int_{-T}^{T} \psi(u)du = \sum_{i=-n}^{n+1} \int_{x_i}^{x_{i+1}} \psi(u)du.$$

On the other hand, we write

$$\frac{T}{n} \sum_{i=-n}^{n-1} \psi(x_i) = \sum_{i=-n}^{n-1} \int_{x_i}^{x_{i+1}} \psi(x_i)du.$$

Then, we have by the mean value theorem that

$$\begin{aligned} \left| \frac{T}{n} \sum_{i=-n}^{n-1} \psi(x_i) - \int_{-T}^{T} \psi(x)dx \right| &\leqslant \sum_{i=-n}^{n-1} \int_{x_i}^{x_{i+1}} |\psi(u) - \psi(x_i)|du \\ &\leqslant \|\psi'\|_\infty \sum_{i=-n}^{n-1} \int_{x_i}^{x_{i+1}} (u - x_i)du = \|\psi'\|_\infty \frac{T^2}{n}. \end{aligned}$$

Proof of part (ii). Define the Lagrangian interpolation polynomial of ψ by

$$\psi_i(x) = \psi(x_i) + \frac{\psi(x_{i+1}) - \psi(x_i)}{x_{i+1} - x_i}(x - x_i).$$

This linear function coincide with ψ for $x \in \{x_i, x_{i+1}\}$. We first remark that:

$$\frac{T}{n} \sum_{i=-n}^{n-1} \frac{\psi(x_i) + \psi(x_{i+1})}{2} = \sum_{i=-n}^{n-1} \int_{x_i}^{x_{i+1}} \psi_i(x)dx.$$

Then, it follows that

$$\left| \frac{T}{n} \sum_{i=-n}^{n-1} \frac{\psi(x_i) + \psi(x_{i+1})}{2} - \int_{-T}^{T} \psi(x)dx \right| \leqslant \sum_{i=-n}^{n-1} \int_{x_i}^{x_{i+1}} |\psi_i(x) - \psi(x)|dx.$$

Now, we look for a bound of $\int_{x_i}^{x_{i+1}} |\psi_i(x) - \psi(x)|dx$ for all $x \in \mathbb{R}$. We introduce the following function for fixed x on $[x_i, x_{i+1}]$

$$\phi(t) = \psi(t) - \psi_i(t) - \frac{(t - x_i)(t - x_{i+1})}{(x - x_i)(x - x_{i+1})}(\psi(x) - \psi_i(x))$$

This function is null in $t = x, x_i$ and x_{i+1}. By the Rolle theorem, there exists a constant c_x such that $\phi''(c_x) = \psi''(c_x) - 2\frac{\psi(x)-\psi_i(x)}{(x-x_i)(x-x_{i+1})} = 0$ which gives $\psi(x) - \psi_i(x) = (x - x_i)(x - x_{i+1})\frac{\psi''(c_x)}{2}$. From this, we deduce that

$$\int_{x_i}^{x_{i+1}} |\psi_i(x) - \psi(x)|dx \leqslant \frac{\|\psi\|_\infty}{2} \int_{x_i}^{x_{i+1}} (x - x_i)(x_{i+1} - x)dx \leqslant \frac{\|\psi\|_\infty}{12}(x_{i+1} - x_i)^3,$$

and $\left|\frac{T}{n}\sum_{i=-n}^{n-1} \frac{\psi(x_i)+\psi(x_{i+1})}{2} - \int_{-T}^{T} \psi(x)dx\right| \leqslant \|\psi\|_\infty \frac{T^3}{12n^2}$. This concludes the proof. □

Appendix C. Some inequalities

The proof of the following Theorem can be found in Stewart and Sun (1990).

Theorem C.1. *Let A and E be two square matrices. If A is nonsingular and for some norm $\|A^{-1}E\| < 1$, then we have*

$$\|(A + E)^{-1} - A^{-1}\| \leqslant \frac{\|A\|^2\|E\|}{1 - \|A^{-1}\|\|E\|},$$

Theorem C.2 (Talagrand's inequality)**.** *Let $(X_i)_{-n\leqslant i\leqslant n-1}$ be independent real random variables, $\mathcal{F}$ a class at most countable of measurable functions.*

$$\nu_n(s) = \frac{1}{2n}\sum_{i=-n}^{n-1} (s(X_i) - \mathbb{E}[s(X_i)]), \qquad \forall s \in \mathcal{F}.$$

We assume there exist third strictly positive constants M_1, H, v such that:

$$\sup_{s\in\mathcal{F}} \|s\|_\infty \leqslant M_1, \ \mathbb{E}[\sup_{s\in\mathcal{F}} |\nu_n(s)|] \leqslant H, \ \text{and} \ \sup_{s\in\mathcal{F}} \frac{1}{n}\sum_{i=-n}^{n-1} \mathrm{Var}(s(X_i)) \leqslant v.$$

Then, for all $\delta > 0$,

$$\mathbb{E}\left[\left(\sup_{s\in\mathcal{F}} |\nu_n^2(s)| - 2(1 + 2\delta)H^2\right)_+\right] \leqslant \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\delta\frac{nH^2}{v}} + \frac{49M_1^2}{K_1C^2(\delta)n^2}e^{-K_1'C(\delta)\sqrt{\delta}\frac{nH}{M_1}}\right)$$

where $C(\delta) = (\sqrt{1 + \delta} - 1) \wedge 1$, $K_1 = 1/3$ and K_1' a universal constant.

The Talagrand inequalities has been proven in Talagrand (1996), reworded by Ledoux (1997). This version is given in Klein and Rio (2005).