Projection estimation for inverses problems on Laguerre and Hermite spaces

Ousmane Bereke Sacko

Université de Paris

École doctorale de Sciences Mathématiques de Paris Centre

Laboratoire Mathématiques Appliquées à Paris 5, CNRS UMR 8145

THÈSE DE DOCTORAT

Spécialité : Mathématiques Appliquées

Présentée par

Ousmane SACKO

Projection estimation for inverse problems on Laguerre and Hermite spaces

---

Estimation par projection pour des problèmes inverses sur des espaces de Laguerre et d'Hermite

Dirigée par :

Fabienne COMTE et Céline DUVAL

Soutenue publiquement le 15 novembre 2021 devant un jury composé de :

Cristina BUTUCEA	Professeure, Université Gustave Eiffel et ENSAE	Rapporteure
Clément MARTEAU	Professeur, Université Claude Bernard Lyon 1	Rapporteur
Jérôme DEDECKER	Professeur, Université de Paris	Examinateur
Jan JOHANNES	Professeur, University of Heidelberg	Examinateur
Jean-Michel LOUBES	Professeur, Université Paul Sabatier	Examinateur
Thanh Mai PHAM NGOC	Maître de conférence-HDR, Université Paris Saclay	Examinatrice
Fabienne COMTE	Professeure, Université de Paris	Directrice
Céline DUVAL	Professeure, Université de Lille	Directrice

Résumé

Dans cette thèse, nous développons des procédures d'estimation non paramétrique dans un cadre d'observation indirecte (problème inverse). La nouveauté dans ce travail est l'utilisation des bases de Laguerre et d'Hermite qui sont à support $\mathbb{R}^+$ ou $\mathbb{R}$ respectivement. Elles ont des propriétés spécifiques que nous exploiterons pour construire des estimateurs pour divers problèmes inverses. Ces bases sont à support non compact et ne nécessitent donc pas de connaître le support de l'objet à estimer, cela est bien adapté dans notre contexte de problème inverse. Quand on utilise une base à support compact pour faire de l'estimation, on considère en théorie ce support comme fixé ; pourtant, en pratique, on le détermine grâce aux données. Les bases de Laguerre et d'Hermite ne sont pas concernées par le choix préliminaire du support d'estimation. Lorsque les variables sont positives, il est naturel d'utiliser la base de Laguerre. La base d'Hermite permet de construire des estimateurs de faible complexité donc résumant la fonction estimée à un petit nombre de coefficients estimés. Toutefois, l'utilisation de ces bases soulève des difficultés mathématiques nécessitant des outils spécifiques.

Cette thèse comporte deux parties.

1. La première partie étudie le problème d'estimation des dérivées d'une fonction de densité f en base de Laguerre et d'Hermite à partir d'observations indépendantes et identiquement distribuées (i.i.d.) de densité f. C'est l'objet du Chapitre 2 de cette thèse. Nous introduisons un estimateur par projection en utilisant les relations de récursivité entre les fonctions de Laguerre et d'Hermite et leurs dérivées. Nous proposons une procédure adaptative en utilisant un critère de sélection de modèles par pénalisation, avec une pénalité qui est indépendante de la base utilisée et nous démontrons qu'elle est optimale au sens du minimax si la densité appartient à une classe de régularité de Sobolev-Laguerre ou Sobolev-Hermite. Une étude numérique est réalisée et des comparaisons avec l'approche à noyau illustrent la bonne performance de notre procédure. La méthode fournit aussi une description parcimonieuse des dérivées d'une densité, puisqu'un petit nombre de coefficients suffit pour avoir des résultats très satisfaisants.
2. La deuxième partie est dédiée à l'estimation d'une densité et d'une fonction de régression dans un modèle convolution. Elle est divisée en deux chapitres.

   Le Chapitre 3 est consacré à l'estimation d'une densité en base d'Hermite dans un modèle à bruit additif. Nous observons les variables $(Z_k)_{1\leqslant k\leqslant n}$ issues du modèle suivant :

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

   où $(X_k)_{1\leqslant k\leqslant n}$ et $(\varepsilon_k)_{1\leqslant k\leqslant n}$ sont indépendantes. Le but est donc d'estimer la densité commune f des variables $(X_k)_{1\leqslant k\leqslant n}$ qui sont supposées soit i.i.d. soit strictement stationnaires et β-mélangeantes. On suppose en outre que les erreurs ε_k sont i.i.d. de densité connue. En utilisant que la transformée de Fourier d'une fonction de la base d'Hermite est cette même fonction à une constante près, nous construisons un estimateur par projection fondée sur un développement en base d'Hermite. La méthode a l'avantage d'être parcimonieuse car elle ne requiert qu'un petit nombre

de coefficients à estimer. L'estimateur est adaptatif et atteint les vitesses classiques dans ce contexte pour une large classe d'erreur. Des illustrations numériques sont réalisées pour illustrer la performance de l'estimateur.

Le Chapitre 4 s'intéresse à l'estimation d'une fonction f à partir des données $(y(x_k), x_k)$ issues du modèle de convolution-régression suivant :

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

où $h(x) = f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy$; $0 < T < \infty$ est fixé ; g est supposée connue et f est la fonction inconnue que l'on cherche à estimer ; les erreurs $(\varepsilon_k)_{-n \leqslant k \leqslant n-1}$ sont i.i.d. avec $\mathbb{E}[\varepsilon_k] = 0$ et $\mathrm{Var}(\varepsilon_k) = \sigma_\varepsilon^2 < \infty$, connu. Nous proposons deux procédures d'estimation. La première est une approche *déconvolution-projection* fondée sur une décomposition de h en base d'Hermite et une transformation de Fourier inverse de f. L'estimateur obtenu par cette première méthode est convergent et adaptatif. Ensuite, nous présentons une approche *projection-projection* consistant à décomposer f et h en base d'Hermite. On obtient un estimateur de f en injectant l'estimateur des moindres carrés de h dans la formule des coefficients de la décomposition de f. Il atteint la même vitesse que la première méthode d'estimation. La fin du chapitre est consacrée à des études numériques pour illustrer les résultats théoriques.

Abstract

In this thesis, we develop non parametric estimation procedures in an indirect observation framework (inverse problem). The novelty of this work is the use of Laguerre and Hermite bases which are $\mathbb{R}^+$ or $\mathbb{R}$ supported respectively. They have specific properties that we will exploit to build estimators for various inverse problems. These bases are non compact supported and therefore do not require to know the support of the object to estimate, this is well adapted to our context of inverse problem. When we use a compactly supported basis for estimation, we consider its support as fixed ; however, in practice, we determine it from the data. Laguerre and Hermite bases are not concerned by the preliminary choice of the estimation support. If the variables are positive, it is natural to use the Laguerre basis. The Hermite basis allows the construction of estimators of low complexity, thus reducing the estimated function to a small number of estimated coefficients. However, their use creates mathematical difficulties requiring specific tools.

This thesis contains two parts.

1. The first part studies the problem of estimating the derivatives of a density f in Laguerre and Hermite bases from independent and identically distributed (i.i.d.) observations with density f. This is the subject of Chapter 2 of this thesis. Using the recursive relations between the Laguerre and Hermite functions and their derivatives, we introduce a projection estimator. We propose an adaptive procedure by using penalization model selection criteria, with a penalty which is independent of the basis considered and we prove that it is minimax optimal if the density belongs to a Sobolev-Laguerre or Sobolev-Hermite regularity class. Numerical studies are realized and comparison with a kernel approach illustrates the good performances of our procedure. The method also provides a sparse description of the derivatives of a density, since a small number of coefficients is sufficient to obtain very satisfactory results.
2. The second part is devoted to the estimation of a density and a regression function in a convolution model. It is split in two chapters.

   Chapter 3 is dedicated to the estimation of a density in Hermite basis with additive noise. We observe the variables $(Z_k)_{1\leqslant k\leqslant n}$ from the following model :

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

   where $(X_k)_{1\leqslant k\leqslant n}$ and $(\varepsilon_k)_{1\leqslant k\leqslant n}$ are independent. The aim is then to reconstruct the common density of variables $(X_k)_{1\leqslant k\leqslant n}$ which are assumed to be either i.i.d. or strictly stationary and β-mixing. We also suppose that the errors $(\varepsilon_k)_{1\leqslant k\leqslant n}$ are i.i.d. of known density. Using that the Fourier transform of a function of the Hermite basis is also the same function up to a constant factor, we construct a projection estimator based on a development in the Hermite basis. The method has the advantage of being parsimonious because it requires only a small number of coefficients for the estimation. The estimator is adaptive and achieves classical rates in this context for a large class of errors. Numerical studies are performed to illustrate the performance of the estimator.

In Chapter 4, we are interested in the estimation of a function f from the data $(y(x_k), x_k)$ obtained from the following convolution-regression model :

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

where $h(x) = f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy$; $0 < T < \infty$ is fixed ; g is assumed to be known and f is the function of interest to be estimated ; the errors $(\varepsilon_k)_{-n \leqslant k \leqslant n-1}$ are i.i.d. such that $\mathbb{E}[\varepsilon_k] = 0$ and $\mathrm{Var}(\varepsilon_k) = \sigma_\varepsilon^2 < +\infty$, known. We propose two estimation procedures. The first is a *deconvolution-projection* approach based on the decomposition of h in the Hermite basis and recover f using an inverse Fourier transform. This estimator obtained by this first method is consistent and adaptive. Then, we present a *projection-projection* approach consisting in decomposing f and h in the Hermite basis. We obtain an estimator of f by injecting the least squares estimator of h into the formula of f's coefficients. It reaches the same rate as the first estimation method. The end of the chapter is devoted to numerical studies to illustrate the theoretical results.

Merci !

Tout d'abord, j'adresse toute ma gratitude à mes deux directrices de thèse, Fabienne Comte et Céline Duval, de m'avoir fait honneur d'encadrer cette thèse. Je souhaite également les remercier pour leur bienveillance, disponibilité, gentillesse. Merci aussi pour vos conseils, encouragements qui m'ont permis d'aller au bout de cette aventure. Je vous suis très reconnaissant de m'avoir initié et fait découvrir le monde de la recherche, j'ai énormément appris avec vous et j'espère que cela n'est que le début d'une fructueuse collaboration. Fabienne, merci d'avoir accepté ma candidature spontanée pour faire mon Stage de Master 2 avec toi.

J'exprime aussi mes sincères et chaleureux remerciements à Cristina Butucea et Clément Marteau, de me faire l'honneur de rapporter cette thèse. Merci pour l'intérêt que vous avez porté à mon travail, pour vos remarques, questions pertinentes et pour vos rapports de qualité. Merci également à Jérôme Dedecker, Jan Johannes, Jean-Michel Loubes et Thanh Mai Pham Ngoc de constituer mon jury. Je souhaite encore remercier Cristina Butucea et Jérôme Dedecker d'avoir fait partie de mon comité du mi-parcours.

Je suis très reconnaissant envers la FSMP de m'avoir accordé une bourse pour mes deux années de Master. J'en profite pour remercier sincèrement Flora Alarcon, ma tutrice pendant mes deux années de Master. Merci aussi à Sylvain Durand, Georges Koepfler et Rachid Lounes d'avoir soutenu mon dossier.

Je voudrais remercier tous les membres permanents du MAP5 et de l'UFR Maths-info de l'Université de Paris (ex Paris Descartes). J'ai eu beaucoup d'entre vous en cours ou TD pour mon L3 et mes années de Master. Je remercie particulièrement Lionel Moison (de m'avoir confié un groupe de TD pendant mes trois années de monitorat dans le cadre de mon contrat doctoral), Manon Defosseux, Julie Délon, Thierry Cabanal, Antoine Chambaz (merci pour ta bienveillance), Marc Briant, Etienne Birmelé, Nicolas Meunier, Florent Benaych-Gerorges, Joan Glaunès, Christine Graffigne, Raphaël Lachieze-Rey. Je pense aussi à mes professeurs du Mali qui m'ont donné l'envie de faire des Mathématiques : Moussa Coulibaly, Oumar Sougané, Aliou Singaré, Diadouba Diabaté, Sigidogho.

Je remercie aussi les équipes administratives et Informatiques du MAP5 et l'UFR Math-info de Paris Descartes. Merci encore à Fabienne et à Anne Estrade (directrice du MAP5) de permettre aux doctorants d'avoir un cadre très agréable de travail. Un merci très spécial aussi à Marie-Hèlene pour sa bonne humeur, sa gentillesse et sa disponibilité. J'en profite pour remercier Sandrine Fallu, Isabelle Valéro, Christophe Castellani et Augustin Hangat. Merci également à Max Paisley, Arnaud Meunier et Rémy Abregel pour leur aide et les dépannages informatiques.

Un grand merci à toutes et à tous les doctorant(e)s, postdorant(e)s, ATER du MAP5 : Andréa, Allesandro, Arthur, Marta, Julie, Juliana, Sinda, Cécile, Laurent, Mehdi, Diala, Sonia, Claire, Alexandre, Anton, Pierre-Louis, Rémi B, Rémi L, Zoé, Fabien, Allan, Vivien, Florian, Antoine M, Yen, Safa, Mariem, Warith, Vincent, Ariane, Charlie, Apolline, Chabane. Merci pour ces 3 belles années.

Merci à tous et à toutes mes camarades et ami(e)s avec qui j'ai partagé énormément des choses : Abdourahmane Sarr, Alpha Diop, Bakary Sidibé, Checik Kanté, Kamila Kare,

Herve, Steaven, Terence, Seydou Keïta, Ibrahim Traoré, Idrissa Touré, Niffa Diaby, Kassim Berthé, Moussa Dembélé, Bineta Diack, Mambi Kébé, Nama Sangaré, Cheichné Dieffaga.

Je souhaite aussi remercier ma famille. Merci à mes oncles et tantes : Boubou Baba, Amina Diombéra, Sita Soumaré (merci beaucoup de t'occuper du pot de ma thèse), Youssouf Dimbéra, N'Kissima Toulé, Aliou Traoré (le directeur) pour ne citer qu'eux. Merci aussi à mes cousines et cousins : Djibril, Nourou, Kaou, Koita et Niouma, Aliou Adama, Oumou Adama, Anthioumane Adama, Salia Dramé, Kaou Sidibé, Amara (shaba), Bouna, Mamédi, Ousmane ("président"), Kandjoura ("Takosss"), Oumar, Harouna Sacko (milanais), Mamédi Dramane (Diouf), Ichaka Diombéra, Tahirou Bintou, Tahirou Souleymane, Ibrahim Soumaré, Moussa Bodo, Nayé Diombéra, Chouaïbou (Gabonais).
Enfin, je remercie mes parents, Diarétou Kaba et Béréké Ladji, pour vos encouragements, vos bénédictions et votre soutien. Je termine par remercier et embrasser tendrement mes frères et sœurs : Mamadou, Aboubakar, Kadiatou, Oumarou, Cissé, Abou, Mégui et notre "Lakharé" Maimouna.

Table des matières

Chapitre 1

Introduction

Sommaire

1.1 Motivations

En statistiques non paramétriques, les bases de Laguerre et d'Hermite sont des outils efficaces pour construire des estimateurs par projection. Elles ont des propriétés spécifiques qui ont été exploitées dans divers contextes : estimation d'une fonction de densité en observation directe, estimation d'une fonction de régression, problème inverse en présence d'un bruit additif ou multiplicatif (voir Comte and Genon-Catalot (2015), Mabon (2017), Comte and Genon-Catalot (2018), Belomestny et al. (2019), Comte and Genon-Catalot (2020a)). Ces bases ont la particularité de ne pas être à support compact. Cela peut avoir des intérêts pratiques ou théoriques, notamment dans le cas où les variables d'intérêt ne sont pas directement observées (problème inverse). Cependant, la caractéristique de non compacité peut créer des difficultés théoriques, en particulier dans le cas du problème d'estimation non paramétrique d'une fonction de régression dans un modèle de régression simple. Cela a été longtemps un obstacle. Récemment, les travaux de Comte and Genon-Catalot (2020a) ont permis de généraliser des résultats longtemps obtenus pour des bases à support compact aux bases de Laguerre et d'Hermite. Quand on utilise une base à support compact pour faire de l'estimation, on considère en théorie ce support comme fixé. Pourtant, en pratique, on le détermine grâce aux données. Les bases de Laguerre et d'Hermite ne sont pas touchées par ce paradoxe, car elles ne requièrent pas de choix préliminaire du support. Si les variables d'intérêt sont positives, il est naturel d'utiliser la base de Laguerre. Des travaux récents de Belomestny et al. (2019) montrent que la base d'Hermite permet de construire des estimateurs de faible complexité, donc rapides numériquement, ne nécessitant l'estimation qu'un petit nombre de coefficients (parcimonieux). D'où l'intérêt d'utiliser la méthode de projection puisqu'elle permet de résumer l'estimation d'une fonction inconnue à un petit nombre de coefficients. Notons également que les fonctions de Laguerre ou d'Hermite ont de bonnes propriétés mathématiques : par exemple les dérivées s'expriment simplement comme une combinaison linéaire des autres fonctions de la base grâce au caractère récursif des polynômes de Laguerre ou d'Hermite, voir Abramowitz and Stegun (1964), Szegö (1959). Ces expressions seront exploitées pour étudier théoriquement et numériquement les estimateurs proposés dans ce manuscrit.

L'objectif de cette thèse est d'utiliser les bases de Laguerre et d'Hermite pour traiter différents problèmes inverses :

1. Estimation des dérivées d'une densité,
2. Estimation de densité dans un modèle de convolution,
3. Estimation d'une fonction de régression dans un modèle de régression-convolution.

Les problèmes considérés se situent dans le cadre de *l'estimation fonctionnelle* ou *estimation non paramétrique*. Ainsi, nous présentons dans ce chapitre introductif la notion d'estimation non paramétrique par méthode de projection, le problème d'adaptation, préliminaires nécessaires à la lecture de ce manuscrit et les résultats obtenus dans la thèse sont ensuite détaillés.

La résultats de la section suivante sur les méthodes de projections s'inspirent en partie des travaux de Massart (2007), Tsybakov (2009) et Comte (2017). Cette section permet aussi de fixer les notations utilisées dans cette thèse.

1.2 Généralités sur l'estimation non paramétrique

En statistique, le cadre classique d'estimation est la statistique paramétrique. Elle consiste à mettre un a priori sur la forme ou la nature du modèle statistique. Le modèle est donc décrit par un nombre fini de paramètres. L'inférence statistique vise à estimer les paramètres du modèle. Toutefois, il est rare d'avoir des informations précises sur la forme du modèle ou le problème qu'on souhaite traiter. De plus, les modèles paramétriques n'expliquent pas toujours bien la complexité des données dont les inconnues sont des fonctions. L'objet que l'on cherche à estimer n'est donc plus un ou plusieurs paramètres mais une fonction qui appartient à une certaine classe de régularité qui est en général de dimension infinie. On parle *d'estimation non paramétrique ou d'estimation fonctionnelle.* Concrètement, on dispose des observations $X_1, \dots, X_n$ et on cherche à estimer une fonction inconnue notée f.

On appelle estimateur noté $\widehat{f}_n = \widehat{f}_n(X_1, \dots, X_n)$, toute fonction mesurable et calculable en fonction des données.

On supposera que f est à valeurs réelles. L'estimateur $\widehat{f}_n$ reflète avec une certaine *erreur* la vraie fonction f inconnue. Pour quantifier cette erreur dite erreur d'estimation, on considère la quantité $\mathbb{E}[\ell(d(\widehat{f}_n, f))]$, où d est la distance qui mesure l'écart entre $\widehat{f}_n$ et f, ℓ une fonction de perte et $\mathbb{E}$ est l'espérance mathématique. En moyenne, elle mesure l'erreur que l'on commet en estimant f par $\widehat{f}_n$ pour la distance d et la perte ℓ. On le nomme aussi risque, et c'est le terme qu'on utilisera dans la suite. Les questions mathématiques "classiques" suivantes se posent : est-ce que $\mathbb{E}[\ell(d(\widehat{f}_n, f))]$ tend vers 0 lorsque le nombre d'observations augmente ? Si oui à quelle vitesse ? Peut-on faire mieux en terme de rapidité du point de vue *général* ? Pour répondre à ces interrogations, nous commençons d'abord par établir des majorations pour $\mathbb{E}[\ell(d(\widehat{f}_n, f))]$, c'est-à-dire des inégalités du type :

$$\sup_{f \in \mathcal{F}} \mathbb{E}\left[\ell(d(\widehat{f}_n, f))\right] \leqslant C\psi_n, \tag{1.1}$$

avec C une constante positive indépendante de n, ψ_n une suite décroissante de n que l'on espère converger vers 0 quand n tend vers $+\infty$ et $\mathcal{F}$ une classe de fonctions contenant la fonction inconnue. On ne se contentera pas de majorations mais on cherchera aussi à obtenir des minorations qui garantissent *l'optimalité* de la vitesse du point de vue général. Pour ce faire, nous prouvons des résultats de type bornes inférieures, c'est-à-dire de montrer que :

$$\inf_{\widehat{f}_n} \sup_{f \in \mathcal{F}} \mathbb{E}\left[\ell(d(\widehat{f}_n, f))\right] \geqslant c\phi_n, \tag{1.2}$$

où l'infimum est pris sur tous les estimateurs possibles, $c > 0$ une constante indépendante de n et ϕ_n une suite décroissante tendant vers 0 quand n tend vers l'infini.

Définition 1.2.1. *On dira qu'un estimateur est optimal au sens du minimax s'il vérifie* (1.1) *et* (1.2)*, avec* $\phi_n = \psi_n$.

La notion de vitesse optimale n'est définie qu'à une constante multiplicative près. Dans cette thèse, on considère une perte quadratique $\ell(u) = u^2$ et une distance de type $\mathbb{L}^2(A)$

définie par

$$d(s,t) = \|t-s\| = \left(\int_A (t-s)^2(x)dx\right)^{\frac{1}{2}},$$

où $A = \mathbb{R}^+$ en base Laguerre et $A = \mathbb{R}$ en base d'Hermite. Dans ce cas, la quantité $\mathbb{E}\left[\ell(d(\widehat{f}_n, f))\right] = \mathbb{E}[\|\widehat{f}_n - f\|^2]$ est appelée risque quadratique intégré, MISE (*Mean Integrated Squared Error*) en anglais. Ces choix sont motivés par les méthodes d'estimation utilisées, notamment la méthode de projection. D'autres exemples de distance et perte sont donnés dans le Chapitre 2, p. 65 de Tsybakov (2009).

1.2.1 Estimation non paramétrique par projection

Il existe deux grandes familles d'estimateurs non paramétrique : estimateur non paramétrique par projection fondé sur la minimisation d'un contraste et les estimateurs à noyau. Nous ne présentons que la méthode de projection, seule utilisée dans ce travail. Nous exposons pour préciser la méthode, en détails, le cas du problème d'estimation d'une densité en observation directe. Le lecteur peut se référer aux livres de Tsybakov (2009) et Comte (2017) pour plus de détails sur les estimateurs à noyau.

Principe d'estimation par projection

Considérons des observations $X = (X_1, \ldots, X_n)$ et f une fonction inconnue que l'on cherche à estimer à partir des données. La méthode de projection consiste à décomposer la fonction inconnue dans une base orthonormée. Un estimateur de f est obtenu en substituant un nombre fini de coefficients de cette décomposition par un estimateur. L'estimateur appartient alors à un espace de dimension finie, d'où le terme *projection*. Plus précisément, soit $(\varphi_j)_{j\geqslant 0}$ une base orthonormée de $\mathbb{L}^2(A)$ où A est une partie de $\mathbb{R}$. On définit l'espace de projection pour chaque $m \geqslant 1$:

$$S_m = \text{Vect}(\varphi_0, \ldots, \varphi_{m-1}), \tag{1.3}$$

l'espace vectoriel engendré par $(\varphi_0, \ldots, \varphi_{m-1})$ de dimension m dans $\mathbb{L}^2(A)$. Pour $f \in \mathbb{L}^2(A)$, nous avons que : $f = \sum_{j\geqslant 0} a_j(f)\varphi_j$, $\quad a_j(f) = \langle f, \varphi_j\rangle = \int_A \varphi_j(x)f(x)dx$. Comme on ne peut pas estimer un nombre infini de coefficient, on considère le projeté de f sur S_m. La projection orthogonale de f dans S_m est directement donnée par les m premiers termes de la décomposition de f :

$$\Pi_{S_m}(f) = f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j.$$

On obtient alors un estimateur de f en remplaçant les m premiers coefficients de f par des estimateurs $\widehat{a}_j$: $\widehat{f}_m = \sum_{j=0}^{m-1} \widehat{a}_j\varphi_j$. Pour calculer les coefficients $\widehat{a}_j$, on utilise deux méthodes classiques d'estimation : les méthodes de moment ou le principe de minimisation de contraste. On privilégie la deuxième approche car elle permet d'obtenir des estimateurs optimaux au sens de l'oracle que nous définirons dans la suite. À cette fin, on introduit

un contraste γ_n qui change selon le problème étudié et on écrira $\widehat{f}_m$ comme le minimiseur du contraste :

$$\widehat{f}_m = \arg\min_{t \in S_m} \gamma_n(t). \tag{1.4}$$

Nous donnons dans le paragraphe suivant un exemple d'application du principe de projection.

Estimation d'une fonction de densité dans le cas d'observation directe

Considérons un échantillon $X_1, \ldots, X_n$ de variables aléatoires réelles indépendantes et identiquement distribuées (i.i.d.) de densité commune f par rapport à la mesure de Lebesgue. Le but est d'estimer la fonction f. On part de la définition d'une projection orthogonale pour déterminer le contraste de minimisation, on a que $f_m = \arg\min_{t \in S_m} \|t - f\|^2$. Posons $\gamma(t) = \|t - f\|^2$: la quantité $\gamma(t)$ n'est pas calculable car elle dépend de f qui est inconnue. Il nous faut donc la remplacer par une quantité empirique calculable. Ainsi, on remplace la norme $\mathbb{L}^2(A)$ par sa version empirique (à une constante près). On commence par la décomposition suivante : $\|t - f\|^2 = \|t\|^2 - 2\langle t, f\rangle + \|f\|^2$, cela implique $\arg\min_{t \in S_m} \|t - f\|^2 = \arg\min_{t \in S_m} \left(\|t\|^2 - 2\langle t, f\rangle\right)$, car la quantité $\|f\|^2$ est indépendante de t. En remarquant que $\langle t, f\rangle = \int_A t(x) f(x) dx = \mathbb{E}[t(X_1)]$, on définit une version empirique du contraste à minimiser :

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n} t(X_i), \quad \|t\|^2 = \int_A t^2(x) dx. \tag{1.5}$$

Par la loi des grands nombres $\gamma_n(t) \xrightarrow[n\to\infty]{p.s.} \|t\|^2 - 2\langle t, f\rangle$, où « *p.s.*» est l'abréviation de presque sûrement. Donc minimiser $\gamma_n(t)$ pour n grand revient donc à minimiser $\gamma(t) = \|t - f\|^2$ pour $t \in S_m$. L'estimateur par projection de la densité est donc donné par :

$$\widehat{f}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j, \qquad \widehat{a}_j = \frac{1}{n}\sum_{i=1}^{n} \varphi_j(X_i). \tag{1.6}$$

On remarque que $\widehat{f}_m$ est un estimateur sans biais de f_m : $\mathbb{E}[\widehat{f}_m] = f_m$. Notons aussi $\widehat{f}_m$ coïncide avec l'estimateur obtenu par la méthode des moments puisque $a_j(f) = \mathbb{E}[\varphi_j(X_1)]$. La question qui se pose ensuite c'est le choix pertinent de l'espace d'approximation S_m. Dans cette optique, on fixe le choix de la base et on s'intéresse au choix de la dimension m. Ce choix relève de la sélection de modèles que nous décrivons un peu plus loin (voir Section 1.3).

Exemples de base orthonormée

Nous donnons ici quelques exemples de bases orthonormées. Nous commençons par deux bases classiques à supports compacts qui ne seront pas étudiées dans ce manuscrit.

- Base trigonométrique. On définit une famille de fonction sur un compact $[a, b]$ par : $\varphi_1(\cdot) = \frac{1}{\sqrt{b-a}} \mathbb{1}_{[a,b]}(\cdot)$, et pour $k \in [1, d]$

$$\varphi_{2k}(\cdot) = \sqrt{\frac{2}{b-a}} \cos(2\pi k \frac{\cdot - a}{b-a}) \mathbb{1}_{[a,b]}(\cdot), \quad \varphi_{2k+1}(\cdot) = \sqrt{\frac{2}{b-a}} \sin(2\pi k \frac{\cdot - a}{b-a}) \mathbb{1}_{[a,b]}(\cdot).$$

On a donc que $m = 2d+1$ et la famille $(\varphi_1, \varphi_2, \ldots, \varphi_m)$ forme une base orthonormée sur $[a, b]$.

- Base d'histogrammes réguliers. La construction se fonde sur le fait de scinder l'intervalle $[a, b]$ en m morceaux de tailles égales. Cette base est donnée par

$$\varphi_k(\cdot) = \sqrt{\frac{m}{b-a}} \mathbb{1}_{[a+\frac{(k-1)(b-a)}{m}, a+\frac{k(b-a)}{m}[}(\cdot)$$

Pour $k \neq j$, le produit $\varphi_k \varphi_j = 0$. La quantité $\sqrt{\frac{m}{b-a}}$ est un facteur de normalisation qui permet d'avoir $\int_a^b \varphi_k^2 = 1$. Notons que les estimateurs obtenues en utilisant cette base sont constants par morceaux. On peut également construire des bases d'histogrammes irréguliers. Mais dans ce cas, la sélection de modèle (voir Section 1.3) est difficile à implémenter, voir le papier de Comte and Rozenholc (2004) pour plus de détails.

On rencontre souvent des bases sur $[0, 1]$, on obtient une base sur $[a, b]$ en faisant des transformations affines. Le choix d'une base sur $[0, 1]$ est fait pour des raisons de simplicité. On peut aussi citer d'autres bases à support compact : les polynômes par morceaux, la base de Legendre, les bases de Splines, les bases d'ondelettes (voir aussi le Chapitre 1 de Comte (2017)).

En réalité, l'intervalle d'estimation n'est pas fixe et change selon le modèle. Il est déterminé grâce aux données du problème. Dans ce travail, nous utiliserons les deux bases suivantes qui sont à support non compact pour construire des estimateurs pour différents problèmes inverses (qui apparaissent quand on n'a pas d'information directe sur les données).

- Base de Laguerre. Les fonctions de Laguerre sont définies sur $[0, +\infty[$ par l'expression

$$\ell_j(x) = \sqrt{2} L_j(2x) e^{-x}, \quad L_j(x) = \sum_{k=0}^{j} \binom{j}{k} (-1)^k \frac{x^k}{k!}, \quad x \geqslant 0, \quad j \geqslant 0, \tag{1.7}$$

où $(L_j)_{j \geqslant 0}$ est le polynôme de Laguerre de degré j. Il vérifie : $\int_0^{+\infty} L_j(x) L_k(x) e^{-x} dx = \delta_{j,k}$ (voir Abramowitz and Stegun (1964), chap 22.2.13) où $\delta_{j,k}$ est le symbole de Kronecher,

$$\delta_{j,k} = \begin{cases} 1 & \text{si } j = k \\ 0 & \text{sinon .} \end{cases}$$

Ainsi, la famille $(\ell_j)_{j \geqslant 0}$ forme une base orthonormée sur $\mathbb{L}^2(\mathbb{R}^+)$ qui satisfait

$$\|\ell_j\|_\infty = \sup_{x \in \mathbb{R}^+} |\ell_j(x)| = \sqrt{2}. \tag{1.8}$$

- Base d'Hermite. De façon analogue, la base d'Hermite est définie à partir des polynômes d'Hermite $(H_j)_{j \geqslant 0}$:

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

Les polynômes d'Hermite sont orthogonaux par rapport à la fonction de poids e^{-x^2} : $\int_{\mathbb{R}} H_j(x)H_k(x)e^{-x^2} = 2^j(j!)\sqrt{\pi}\delta_{j,k}$ (voir Abramowitz and Stegun (1964), chap 22.2.14). On en déduit que la base d'Hermite $(h_j)_{j\geqslant 0}$ est une base orthonormée sur $\mathbb{L}^2(\mathbb{R})$:

$$h_j(x) = c_j H_j(x)e^{-x^2/2}, \quad c_j = (2^j j!\sqrt{\pi})^{-1/2}, \quad x \in \mathbb{R}. \tag{1.9}$$

La base d'Hermite est une base bornée satisfaisant

$$\|h_j\|_\infty = \sup_{x\in\mathbb{R}} |h_j(x)| \leqslant \pi^{-1/4}, \tag{1.10}$$

(voir Abramowitz and Stegun (1964), chap 22.14.17 et Indritz (1961)).

1.2.2 Compromis biais-variance et vitesse de convergence de l'estimateur par projection

Nous disposons d'une collection d'estimateurs $(\widehat{f}_m)_{m\geqslant 1}$ associés aux espaces S_m. On veut valider le principe de projection. Pour ce faire, on calcule son risque quadratique intégré $\mathbb{E}[\|\widehat{f}_m - f\|^2]$ communément appelé MISE en anglais (*Mean Integrated Squared Error*), pour chaque $m \geqslant 1$, afin de choisir la meilleure dimension parmi tous les choix possibles. Ce choix conduit à une vitesse de convergence. En utilisant le théorème de Pythagore, on considère la décomposition suivante :

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] = \|f - f_m\|^2 + \mathbb{E}[\|\widehat{f}_m - f_m\|^2]. \tag{1.11}$$

- Le premier terme $\|f - f_m\|^2 = \sum_{j\geqslant m} a_j(f)^2$ de la décomposition (1.11) est appelé le biais. Il quantifie l'erreur d'approximation, c'est l'erreur qu'on commet en remplaçant f par son projeté f_m,
- Le second terme $\mathbb{E}[\|\widehat{f}_m - f_m\|^2] = \sum_{j=0}^{m-1} \mathbb{E}[(\widehat{a}_j - a_j(f))^2]$ est appelé la variance ou erreur stochastique qui vient du fait que l'on remplace les coefficients $a_j(f)$ par $\widehat{a}_j$ pour $j = 0, \ldots, m-1$.

Les deux termes ont des comportements antagonistes par rapport à la dimension m. Le biais décroit lorsque m augmente puisque f se rapproche de plus en plus de f_m alors que la variance augmente avec m. Il faut donc trouver un équilibre entre ces deux termes. D'où le terme *compromis biais-variance*. Soit m_{opt} la dimension qui fait le compromis biais-variance dans (1.11), c'est-à-dire

$$m_{opt} = \arg\min_{m\geqslant 1}\{\|f - f_m\|^2 + \mathbb{E}[\|\widehat{f}_m - f_m\|^2]\}.$$

Pour calculer la dimension pertinente m_{opt}, on examine d'abord les ordres de grandeur du biais et de la variance. L'ordre du biais est obtenu en mettant une condition de régularité sur la fonction inconnue. En général, chaque base est associée à un espace de régularité spécifique. Les bases d'histogrammes, de polynômes par morceaux sont associées à des espaces de Besov. La base trigonométrique est associée aux Sobolev classiques. Les bases de Laguerre et d'Hermite qui sont utilisées dans cette thèse, sont associées respectivement à des espaces de Sobolev-Laguerre et Sobolev-d'Hermite. La classe Sobolev-d'Hermite est

un sous espace du Sobolev classique (voir Bongioanni and Torrea (2006)) et donc des classes de Besov si la régularité est un entier positif. Dans la suite, nous rappellerons les définitions des classes de Sobolev, Sobolev-Laguerre et Sobolev-Hermite. Le lecteur peut se référer à DeVore and Lorentz (1993) pour plus de détails sur les classes de Besov. Très généralement, le biais est d'ordre m^{-2s} où $s > 0$ est la régularité de la fonction inconnue. En ce qui concerne la variance, son ordre est une certaine fonction croissante de m divisée par le nombre d'observations. Cette fonction de m change selon le problème étudié. Nous précisons dans le paragraphe suivant son ordre exact dans le cas du problème direct de l'estimation de densité, cela conduit à une vitesse de convergence.

Vitesse classique dans le cas d'estimation d'une densité en observations directes

Dans ce cas particulier, l'estimateur par projection $\widehat{f}_m$ est donné en (1.6), la méthode usuelle consiste à utiliser la majoration suivante : $\mathbb{E}[\|\widehat{f}_m - f_m\|^2] \leqslant C_\varphi^2 m/n$ où $C_\varphi^2 > 0$ est une constante qui dépend de la base utilisée ($C_\varphi = 1$ pour une base trigonométrique sur $[0,1]$). On obtient alors

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leqslant C_f m^{-2s} + C_\varphi^2 \frac{m}{n}, \tag{1.12}$$

où C_f est une constante qui dépend de la classe de régularité considérée. En cherchant l'argmin du terme à droite de (1.12), on trouve $m_{opt} = (2C_f/C_\varphi^2)^{1/(2s+1)} n^{1/(2s+1)}$. Cela implique la borne suivante

$$\mathbb{E}[\|\widehat{f}_{m_{opt}} - f\|^2] \leqslant C(s, C_f) n^{-\frac{2s}{2s+1}}, \tag{1.13}$$

où $C(s, C_f)$ est une constante dépendant uniquement de s et C_f et non de m. L'estimateur $\widehat{f}_{m_{opt}}$, appelé, *oracle* converge à la vitesse $n^{-2s/(2s+1)}$. Cette vitesse est d'autant meilleure que s est grand. On peut donc interpréter cela par «Plus la fonction qu'on cherche à estimer est régulière plus elle est "facile" à estimer». C'est la vitesse classique dans le contexte d'estimation non paramétrique de densité. Elle est connue pour être optimale au sens minimax pour des espaces de Hölder, Sobolev... (voir Tsybakov (2009)). De plus, elle coïncide avec celle obtenue dans le cas où on estime f dans un modèle de régression non paramétrique (voir Barron et al. (1999)). Elle est toujours moins bonne que la vitesse obtenue dans le cadre d'estimation paramétrique qui est souvent $1/n$ à erreur fixée, il faut plus de données pour avoir une performance équivalente au cas paramétrique. C'est le prix à payer pour l'utilisation d' approches non paramétriques.

Que se passe t-il avec les bases de Laguerre ou d'Hermite ? Avec les bases de Laguerre et d'Hermite, les ordres de grandeur du biais et de la variance ne sont plus donnés par (1.12). Comte and Genon-Catalot (2018) établissent que l'ordre exact (majoration et minoration moins un reste près) du terme de variance est $\sqrt{m}/n$. Elles obtiennent cet ordre en utilisant les formules d'Askey and Wainger (1965) et de Szegö (1959) sous des conditions de moment faibles. Pour contrôler le biais, on définit les espaces de régularité associés à ces deux bases.

Classe de Sobolev-Hermite.

Définition 1.2.2. *Soient $s > 0$ et $D > 0$, on définit l'espace Sobolev-Hermite de régularité s par (voir Bongioanni and Torrea (2006)) :*

$$W_H^s = \left\{ \theta \in \mathbb{L}^2(\mathbb{R}), |\theta|_s^2 = \sum_{k \geqslant 0} k^s a_k^2(\theta) < +\infty \right\}, \quad a_k(\theta) = \int_{\mathbb{R}} \theta(x) h_k(x) dx,$$

où h_j est la base d'Hermite donnée en (1.9)*. La boule de Sobolev-Hermite est donc définie par :*

$$W_H^s(D) = \left\{ \theta \in \mathbb{L}^2(\mathbb{R}), |\theta|_s^2 = \sum_{k \geqslant 0} k^s a_k^2(\theta) \leqslant D \right\}, \quad D > 0,$$

où D est le rayon de la boule.

Cette régularité s peut-être vue comme l'ordre de dérivabilité de la fonction. En effet : si $s \geqslant 1$ est entier, θ appartient à W_H^s si et seulement si θ admet des dérivées jusqu'à l'ordre s et les fonctions $\theta, \ldots, \theta^{(s)}, x^{s-l}\theta^{(l)}$ pour $l = 0, \ldots, s-1$ sont de carrés intégrables. Pour $l = 0, \ldots, s-1$, on a

$$\int_{\mathbb{R}} |x^{s-l}\theta^{(l)}(x)|^2 dx \leqslant \int_{\mathbb{R}} |(1+x^s)\theta^{(l)}(x)|^2 dx.$$

Il vient donc que si $s \geqslant 1$, « $\theta \in W_H^s$ » est équivalente à « θ est s fois dérivable et les fonctions $\theta, \ldots, \theta^{(s)}, x^s\theta^{(l)}$ sont de carrés intégrables sur $\mathbb{R}$ pour $l = 0, \ldots, s-1$». On peut donc comparer cet espace à la classe de Sobolev classique d'indice de régularité s, définie par :

$$W^s = \left\{ \theta \in \mathbb{L}^2(\mathbb{R}), \int (1+u^2)^s |\theta^*(u)|^2 du < \infty \right\}, \quad \theta^*(u) = \int e^{iux}\theta(x) dx.$$

Si s est entier W^s devient

$$W^s = \left\{ \begin{array}{c} \theta \in \mathbb{L}^2(\mathbb{R}),\ \theta \text{ admet des dérivées jusqu'à l'ordre } s, \text{ tel que} \\ |||\theta|||_{s,sob} := \sum_{j=0}^s |\theta^{(j)}|^2 < +\infty \end{array} \right\}.$$

Pour s entier, on en déduit alors que $W_H^s \subset W^s$ (au sens large). De plus, il est prouvé dans Bongioanni and Torrea (2006) que $W_H^s \subsetneq W^s$ (au sens strict) pour tout $s > 0$ et si $\theta \in W^s$ est à support compact, c'est-à-dire $\text{supp}(\theta) \subset [-a, a]$ avec $a > 0$ alors $\theta \in W_H^s$. Ainsi, il y a donc une équivalence entre $\theta \in W^s$ et $\theta \in W_H^s$ pour des fonctions à support compact. Notons aussi que θ et toutes ses dérivées jusqu'à l'ordre $s-1$ s'annulent aux bords : $\theta(a) = \theta(-a) = \ldots \theta^{(s-1)}(a) = \theta^{(s-1)}(-a) = 0$ s'il est à support compact.

Classe de Sobolev-Laguerre. La classe de Sobolev-Laguerre est définie de façon analogue à la précédente.

Définition 1.2.3. *La classe de Sobolev-Laguerre de régularité s est définie par (voir Bongioanni and Torrea (2006)) :*

$$W_L^s = \left\{ \theta \in \mathbb{L}^2(\mathbb{R}^+), |\theta|_s^2 = \sum_{k \geqslant 0} k^s a_k^2(\theta) < +\infty \right\}, \quad a_k(\theta) = \int_{\mathbb{R}^+} \theta(x) \ell_k(x) dx,$$

ℓ_j est la base de Laguerre, définie en (1.7)*. La boule de Sobolev-Laguerre est donc donnée par :*

$$W_L^s(D) = \left\{ \theta \in \mathbb{L}^2(\mathbb{R}^+), |\theta|_s^2 = \sum_{k\geqslant 0} k^s a_k^2(\theta) \leqslant D \right\}, \quad D > 0.$$

Comme pour les W_H^s, la propriété $\theta \in W_L^s$ est liée à la régularité de θ au sens des dérivées. En effet : si $s \geqslant 1$ est entier, Comte and Genon-Catalot (2015) (voir Section 7) montrent que «$\theta \in W_L^s$» est équivalente à « θ admet des dérivées jusqu'à l'ordre $s-1$ tel que $f^{(s-1)}$ est absoluement continue et les fonctions $x^{(l+1)/2}\sum_{j=0}^{l+1}\binom{l+1}{j}\theta^{(j)}$ sont de carrés intégrables pour $l = 0, \ldots, s-1$». Ainsi, si θ est à support compact alors « $\theta \in W_L^s$ » est équivalente à «$\theta \in W^s$».

Pour $f \in W_H^s(D)$ ou $f \in W_L^s(D)$, on a que $\|f - f_m\|^2 = \sum_{j\geqslant m} a_j(f)^2 \leqslant Dm^{-s}$. On en déduit alors la borne suivante

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leqslant Dm^{-s} + c\frac{\sqrt{m}}{n}, \quad c > 0. \tag{1.14}$$

Cette borne est la même que celle obtenue en (1.12) si on remplace m par $\sqrt{m}$. Une première conséquence est que le rôle de la dimension est joué par $\sqrt{m}$ et non m pour les fonctions de Laguerre ou d'Hermite. En sélectionnant $m_{opt} \propto n^{\frac{1}{s+1/2}}$, on retrouve ainsi la vitesse classique $n^{-2s/(2s+1)}$. Dans le chapitre 2, nous démontrons que c'est la vitesse optimale pour les classes de Sobolev-Laguerre et Sobolev-Hermite qui s'étendent aussi aux Sobolev classiques puisque les fonctions de test utilisées sont à support compact. Nous retrouverons le même phénomène dans le Chapitre 3. Cette spécificité d'avoir un biais en m^{-s} affecte les résultats de convergence dans d'autres contextes, notamment dans le cas du problème de régression non paramétrique dont la variance est exactement égal en m/n (voir Baraud (2000), Baraud (2002)), et cela est indépendant de la base utilisée. Elle conduit ainsi à la vitesse $n^{-s/(s+1)}$ (voir Comte and Genon-Catalot (2020a) dont la procédure est fondée sur une approche des moindres carrés). C'est la vitesse optimale pour les classes de Sobolev-Laguerre ou Sobolev-Hermite dans ce contexte pour un bruit gaussien. Elle n'est pas standard et est spécifique aux bases de Laguerre et d'Hermite. En effet, dans Baraud (2000), Baraud (2002), l'estimateur des moindres carrés converge à la vitesse $n^{-2s/(2s+1)}$ si la fonction de régression appartient aux classes de Besov. On aura un résultat similaire dans le contexte du chapitre 4.

L'oracle $\widehat{f}_{m_{opt}}$ n'est cependant pas calculable car il dépend de la fonction inconnue f, on dira que le choix m_{opt} n'est pas adaptatif. Il faut donc trouver un autre stratégie pour choisir la dimension m. C'est l'objet de la Section 1.3.

1.3 Adaptation

On a vu dans la section précédente que le risque de l'oracle $\widehat{f}_{m_{opt}}$ dépend de la régularité de la fonction inconnue que l'on cherche à estimer. Cette régularité n'est donc évidemment pas connue puisque la fonction est elle même inconnue. Ainsi, des méthodes d'estimations adaptatives ont été développées dans les années 90. Ces méthodes de construction sont

dites de *"data driven"*, c'est-à-dire conduites par les données. Les estimateurs obtenus par ces méthodes sont calculables uniquement en fonction des données du problème sans aucun a priori sur la régularité de la fonction inconnue. Les estimateurs résultants de ces méthodes sont appelés estimateurs adaptatifs : en effet, sans connaître la régularité de la fonction inconnue, l'estimateur adaptatif atteint la même vitesse que si la régularité était connue, dans le sens où le compromis biais-variance est automatiquement réalisé.

Dans ce travail, nous utiliserons deux méthodes pour faire l'adaptation. La première est la *sélection de modèles par pénalisation*, utilisée pour faire la sélection de dimension pour un estimateur par projection. La deuxième méthode dite de *Goldenshluger et Lepski* (notée sélection GL dans la suite), utilisée à l'origine pour faire la sélection de fenêtre pour un estimateur à noyau, a été étendue aux méthodes de projection en dimension supérieure. Mentionnons, à titre indicatif, que d'autres approches de sélection de modèle existent, telles que le seuillage en ondelettes (voir Donoho et al. (1996) et Härdle et al. (1998)), les méthodes d'agrégation (voir Rigollet and Tsybakov (2007)).

Dans cette thèse, nous adoptons le point de vue oracle. Il s'agit d'établir, pour une dimension $\widehat{m}$ résultant d'une des procédures de sélection, des inégalités du type :

$$\mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] \leqslant C \inf_{m \in \mathcal{M}_n} \left(\mathbb{E}[\|\widehat{f}_m - f\|^2] \right) + R_n, \tag{1.15}$$

où C est une constante numérique strictement positive et R_n un terme qui dépend de n, que l'on espère négligeable devant $C \inf_{m \in \mathcal{M}_n} \left(\mathbb{E}[\|\widehat{f}_m - f\|^2] \right)$. Le terme R_n est très souvent d'ordre à $1/n$ dans le meilleur des cas ou $(\log(n)^p)/n$ où p est un entier positif que l'on espère pas trop grand. L'inégalité (1.15) est appelée *type-oracle*, son obtention est basée sur des inégalités de concentration que l'on rappellera dans la suite (voir Section 1.3.3). Cette borne est d'autant meilleure que la constante C est proche de 1. Les estimateurs qui vérifient (1.15) sont dits adaptatifs, dans le sens où le compromis biais-variance est automatiquement réalisé. L'estimateur final $\widehat{f}_{\widehat{m}}$ est donc aussi performant que l'oracle à une constante multiplicative C près plus un reste R_n négligeable.

Les résultats du type oracle que nous établirons dans cette thèse sont de nature non asymptotique c'est-à-dire vrais pour n'importe quelle valeur du nombre d'observations.

1.3.1 Sélection de modèles par pénalisation

Commençons par une brève chronologie. Les procédures de sélection de modèle par pénalisation ont été développées dans les années 90. Les premiers travaux ont été introduits par Akaike (1973) et Mallows (1995) qui suggéreraient de pénaliser par la dimension du modèle. Ces travaux ont depuis été généralisés par Birgé and Massart (1997), Barron et al. (1999) et Massart (2007) pour différents problèmes classiques : estimation d'une densité, d'une fonction de régression par exemple. La version que nous décrivons s'inspire de ces travaux.

Cette méthode de sélection est utilisée dans les chapitres 2 et 3 du manuscrit.

Description de la procédure de sélection par pénalisation

Considérons une suite croissante d'espace d'approximation $(S_m)_{m\in\mathbb{N}}$ défini en (1.3) où l'on rappelle que m est la dimension de S_m. Soit $(\widehat{f}_m)_{m\in\mathcal{M}_n}$ une collection d'estimateurs définis en (1.4) associée à la sous collection $(S_m)_{m\in\mathcal{M}_n}$ où $\mathcal{M}_n$ est une collection finie de modèles, mais assez massive pour permettre à la procédure d'avoir suffisamment de marge de manœuvre. En effet, plus on donne de choix à la procédure moins il y a de chance qu'elle se trompe. Elle est en général choisie pour borner l'ordre de la variance. On peut choisir par exemple $\mathcal{M}_n = \{1, 2, \ldots, n\}$. Dans le meilleur des mondes, on cherche m tel que le risque quadratique intégré est minimum : $\widehat{m} := \arg\min_{m\in\mathcal{M}_n}\{\mathbb{E}\,\|\widehat{f}_m - f\|^2\}$. Ce minimiseur dépend de la fonction inconnue donc n'est pas calculable. Une approche naturelle serait de minimiser la version empirique du risque $\gamma_n(\widehat{f}_m)$ en m où γ_n est par exemple donné en (1.5) pour l'estimation en observation directe d'une densité. Supposons que les modèles (S_m) sont emboîtés c'est-à-dire que si $m \leqslant m'$ alors on a $S_m \subset S_{m'}$. Cette hypothèse est assez forte mais est naturellement vérifiée par les espaces associés aux deux bases utilisées dans ce travail. Une condition plus faible consiste à considérer qu'il existe un espace englobant (voir Birgé and Massart (1998) et Baraud (2000)) : il existe $m_n \in \mathcal{M}_n$ tel que pour tout $m \in \mathcal{M}_n$, on a $S_m \subset S_{m_n}$. La condition d'avoir des espaces S_m emboîtés ou un englobant est naturelle dans le cadre de la sélection de modèles. Ainsi, pour $S_{m'} \subset S_m$, on a $\gamma_n(\widehat{f}_m) \leqslant \gamma_n(\widehat{f}_{m'})$. Le risque empirique $\gamma_n(\widehat{f}_m)$ décroit quand la dimension augmente, ce n'est pas le cas pour le vrai risque. On introduit alors le critère suivant :

$$\mathrm{crit}(m) = \gamma_n(\widehat{f}_m) + \mathrm{pen}(m),$$

où γ_n est un contraste de minimisation et $\mathrm{pen} : \mathcal{M}_n \mapsto \mathbb{R}^+$ est une fonction croissante du modèle introduite pour équilibrer le terme $\gamma_n(\widehat{f}_m)$. On choisit donc la dimension pertinente en minimisant le critère :

$$\widehat{m} = \arg\min_{m\in\mathcal{M}_n} \mathrm{crit(m)}.$$

Expliquons maintenant comment on choisit la fonction de pénalité. La fonction $\mathrm{pen}(m)$ est très souvent choisie comme étant au moins la variance :

$$\mathrm{pen}(m) \geqslant \kappa\mathbb{E}[\|\widehat{f}_m - f_m\|^2,$$

où κ est une constante numérique strictement positive et indépendante des données. Une valeur doit être attribuée à la constante κ. Les résultats théoriques obtenus grâce aux inégalités de déviation que nous donnerons dans la suite (Section 1.3.3) donnent une minoration de κ. Cependant, ces résultats sont obtenues après des majorations qui ne sont pas toujours les plus fines possibles. En pratique, κ est pris plus petit que la valeur théorique très généralement. Cette valeur est calibrée une fois pour toutes avec des simulations préliminaires, ou en utilisant des méthodes spécifiques. Il existe deux stratégies (l'heuristique de pente et la méthode de saut de dimension) qui permettent de déterminer une valeur de κ en fonction des données, décrites dans Baudry et al. (2012), et implémentées dans des programmes Matlab et R ("Capushe") librement accessibles.

L'objectif final est qu'on puisse établir pour l'estimateur résultant de cette sélection une

inégalité du type (1.15) pour une pénalité bien choisie.

Nous détaillons dans le paragraphe suivant le choix de la pénalité et la valeur de $\gamma_n(\widehat{f}_m)$ dans un cas particulier.

Application à l'estimation d'une fonction de densité en observation directe

Concernant ce cas particulier, la quantité $\gamma_n(\widehat{f}_m)$ est égale à $-\|\widehat{f}_m\|^2$ et estime $\|f-f_m\|^2 = \|f\|^2 - \|f_m\|^2$ à la constante (par rapport à m) $\|f\|^2$ près. La pénalité est en général choisie comme déterministe et d'ordre de la variance (voir Équations (1.12) et (1.14)) : $\mathrm{pen}(m) = \kappa C_\varphi^2 \frac{D_m}{n}$ où C_φ est une constante explicitement connue mais change en fonction de la base utilisée ($C_\varphi = 1$ pour les bases trigonométriques ou d'histogramme régulier sur $[0,1]$) et D_m une fonction croissante de m. Cette quantité D_m est égale m pour les bases trigonométriques et histogrammes réguliers ou $\sqrt{m}$ pour les bases de Laguerre et d'Hermite. Une stratégie plus générale introduite par Comte and Genon-Catalot (2018) consiste à estimer directement la quantité $\sum_{j=0}^{m-1} \mathbb{E}[(\varphi_j(X_1))^2]$. En effet, dans le cas de la densité, on a

$$\mathbb{E}[\|\widehat{f}_m - f_m\|^2] = \sum_{j=0}^{m-1} (\widehat{a}_j - a_j(f))^2 = \sum_{j=0}^{m-1} \mathrm{Var}(\widehat{a}_j).$$

Comme $\mathrm{Var}(\widehat{a}_j) = \frac{1}{n}\mathrm{Var}(\varphi_j(X_1))$ car les $\varphi_j(X_i)$ pour $i = 1, \dots, n$ sont i.i.d., il vient alors

$$\mathbb{E}[\|\widehat{f}_m - f_m\|^2] \leqslant \frac{1}{n} \sum_{j=0}^{m-1} \mathbb{E}[(\varphi_j(X_1))^2].$$

En substituant $\sum_{j=0}^{m-1} \mathbb{E}[(\varphi_j(X_1))^2]$ par sa version empirique, on introduit la quantité

$$\widehat{V}_m = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{m-1} (\varphi_j(X_i))^2.$$

Finalement, on sélectionne $\widehat{m}$ comme suit :

$$\widehat{m} := \arg\min_{m \in \mathcal{M}_n} \{-\|\widehat{f}_m\|^2 + \widehat{\mathrm{pen}}(m)\}, \quad \text{où} \quad \widehat{\mathrm{pen}}(m) = \kappa \frac{\widehat{V}_m}{n}, \tag{1.16}$$

où κ est une constante à calibrer. Nous avons une pénalité aléatoire et indépendante de la base utilisée. Elle est heuristiquement d'ordre de la variance puisque $\mathbb{E}[\widehat{V}_m] = \sum_{j=0}^{m-1} \mathbb{E}[(\varphi_j(X_1))^2]$. Cela a l'avantage de rendre la calibration de κ moins fastidieuse mais le prix à payer est qu'on a besoin d'autres outils mathématiques. En considérant les bases de Laguerre et d'Hermite, on peut établir l'inégalité oracle suivante (voir Comte and Genon-Catalot (2018) et Belomestny et al. (2019)) pour $\widehat{m}$ sélectionné en (1.16) et $\kappa \geqslant 8$:

$$\mathbb{E}[[\|\widehat{f}_{\widehat{m}} - f\|^2] \leqslant C \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|^2 + \kappa \frac{\sum_{j=0}^{m-1} \mathbb{E}[(\varphi_j(X_1)^2)]}{n} \right) + \frac{C'}{n},$$

où C est une constante numérique ($C = 4$ convient) et C' une constante qui dépend de $\|f\|_\infty$. Les outils théoriques pour établir cette inégalité oracle sont les inégalités de Talagrand et de Bernstein (voir Section 1.3.3), contrairement au cas d'une pénalité déterministe où on utilise uniquement l'Inégalité de Talagrand. Elles montrent en particulier qu'on peut choisir $\kappa = 8$. Cependant, en pratique la valeur pertinente est $\kappa = 4$ (voir Comte and Genon-Catalot (2018) et Belomestny et al. (2019)).

Nous généralisons cette approche qui consiste à estimer directement un majorant de la variance dans le Chapitre 2 pour estimer les dérivées d'une densité.

Dans la Figure 1.1, nous donnons une illustration de la procédure de sélection par pénalisation. On observe que le choix de la méthode est pertinent, parmi les 50 possibilités proposées. De plus, la figure de gauche permet de comprendre le compromis biais-variance, en effet : pour des dimensions petites, l'estimateur est biaisé par contre pour des dimensions grandes, on voit apparaître des oscillations (variance).

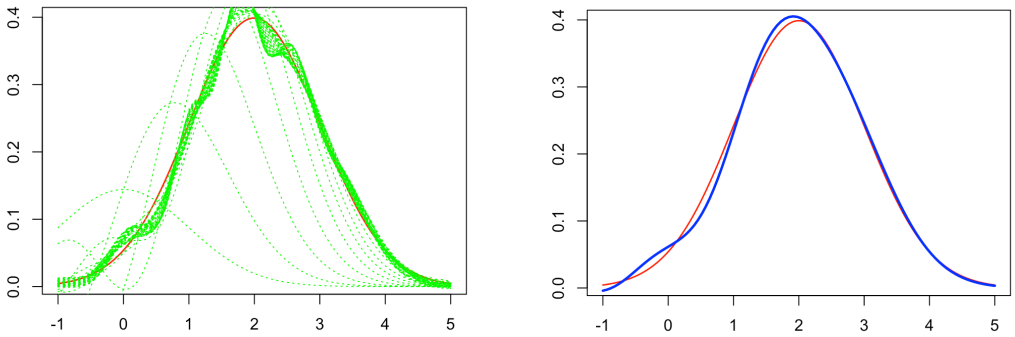


FIGURE 1.1 – Collection d'estimateurs par projection en base d'Hermite de la densité d'une $\mathcal{N}(2,1)$ en trait épais (rouge), la courbe sélectionnée à droite parmi 50 propositions données à gauche (en vert pointillés). La procédure par pénalisation sélectionne $\widehat{m} = 9$ pour $n = 1000$.

1.3.2 Sélection de modèles inspirée de Goldenshluger et Lepski

La méthode de Goldenshluger and Lepski (2011) a été introduite pour faire la sélection du paramètre de lissage (*fenêtre*) dans le cas de l'estimation d'une fonction de densité en dimension d quelconque. Cette dernière méthode s'adapte aussi à la sélection de modèle, on parlera de la sélection GL. Elle est fondée principalement sur la comparaison des estimateurs deux à deux. La différence principale avec la sélection de modèle par pénalisation réside sur l'estimation du biais. Les premières inégalités de type oracle ont été établies par Kerkyacharian et al. (2001), Goldenshluger and Lepski (2009) et Goldenshluger and Lepski (2008). Le lien entre la sélection de modèle avec la méthode de Goldenshluger et Lepski a été fait par Birgé (2001). L'auteur s'inspire d'une stratégie antérieure des méthodes de Lepski. Cette approche peut être utile dans le cas où on ne procède pas par minimisation

de contraste pour construire un estimateur par projection. Elle permet de faire de la sélection de modèle anisotropique en dimension supérieure. La version que nous allons décrire ici est utilisée dans des travaux de Comte and Johannes (2012), Chagny (2013) et récemment par Mabon (2016) dans divers contextes d'estimation. C'est une version simplifiée et adaptée à la sélection de modèle.

Cette méthode est également utilisée dans le chapitre 4 pour un estimateur obtenu en inversant une transformée de Fourier dans un modèle de convolution.

Description de la procédure de sélection GL

Considérons $(\widehat{f}_m)_{m\in\mathcal{M}_n}$ la collection d'estimateurs définie en (1.4) et associée au sous espace $(S_m)_{m\in\mathcal{M}_n}$ défini en (1.3) avec $\mathcal{M}_n$ une collection de modèle finie, $\mathcal{M}_n = \{1, 2, \dots, n\}$ par exemple. Comme pour la sélection de modèle par pénalisation, le but est de trouver une dimension pertinente $\widehat{m}$ telle que l'estimateur résultant $\widehat{f}_{\widehat{m}}$ vérifie une inégalité d'oracle (Inégalité (1.15)). Pour ce faire, on part d'une décomposition biais-variance, plus précisément de (1.11). Nous allons donc remplacer le biais $\|f - f_m\|^2 = \|f - \Pi_{S_m}(f)\|^2$ par un estimateur. En substituant cette fois ci f par $f_{m'}$ avec m' non nécessairement égal à m, on peut donc substituer au biais le terme $\|\Pi_{S_{m'}}(f) - \Pi_{S_m}(\Pi_{S_{m'}}(f))\|^2 = \|\Pi_{S_{m'}}(f) - \Pi_{S_{m\wedge m'}}(f)\|^2$ en utilisant que les espaces d'approximation sont emboîtés où $m \wedge m'$ désigne le minimum entre m et m'. On dispose ainsi de l'estimateur *auxiliaire* :

$$\widehat{f}_{m'\wedge m} = \sum_{j=0}^{m\wedge m'-1} \widehat{a}_j \varphi_j,$$

où $\widehat{a}_j$ estime $a_j(f) = \int_A f(x)\varphi_j(x)dx$. Pour $\kappa_1 > 0$, on définit donc l'estimateur du biais :

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

où $V(m)$ est de l'ordre du terme de variance (dans le meilleur des cas). Le terme $V(m')$ dans l'expression de $\widehat{A}(m)$ est introduit pour corriger la quantité $\|\widehat{f}_{m'} - \widehat{f}_{m\wedge m'}\|^2$ qui contient de l'aléa alors que le terme de biais n'en contient pas. Heuristiquement, on souhaite que le terme $\widehat{A}(m)$ soit d'ordre du bais plus un reste négligeable d'ordre $1/n$ par exemple. Ainsi, on sélectionne $\widehat{m}$ en posant

$$\widehat{m} = \arg\min_{m\in\mathcal{M}_n} \left\{ \widehat{A}(m) + \kappa_2 V(m) \right\}, \quad \kappa_2 \geqslant \kappa_1.$$

Pour une variance $V(m)$ bien choisie et κ_1 au dessus d'un certain seuil, on veut que l'estimateur $\widehat{f}_{\widehat{m}}$ satisfasse une inégalité oracle non asymptotique du type :

$$\mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] \leqslant C \inf_{m\in\mathcal{M}_n} \left(\|f - f_m\|^2 + V(m) \right) + R_n. \tag{1.17}$$

De façon analogue au choix de pen(m), on prend la variance $V(m) = C_\varphi D_m/n$ ($D_m = m$ pour les bases trigonométriques et d'histogrammes ou $D_m = \sqrt{m}$ pour les bases de Laguerre et d'Hermite) où $C_\varphi > 0$ est une constante qui dépend de la base considérée et $R_n = 1/n$ pour le cas de l'estimation d'une densité en observation directe. Le même ordre

de variance apparait dans les travaux de Chagny (2013) (avec $D_m = m$) qui compare les deux approches de sélection de modèles en régression non paramétrique. Comme pour la sélection de modèles par pénalisation, une valeur doit être attribuée à chacune des constantes κ_1 et κ_2 pour piloter la procédure GL en pratique. On peut choisir $\kappa_2 = \kappa_1$, de façon à n'avoir qu'une constante à calibrer. Un calcul d'erreur empirique pour différentes valeurs de κ_1 permet par exemple de faire un bon compromis parmi plusieurs valeurs. Toutefois, cette calibration de κ_1 s'est révélée difficile en pratique dans le cas de la sélection de fenêtre pour un estimateur à noyau car on a une même constante qui apparait sur deux termes qui ont des comportements très différentes. Lacour and Massart (2016) développent l'idée de considérer deux constantes distinctes dans l'expression $\widehat{A}(h)$ et $\widehat{h}$ pour un estimateur à noyau d'une densité de fenêtre h. Ils proposent de prendre $\kappa_2 = 2\kappa_1$. D'après leurs études $\kappa_1 = 1$ et $\kappa_2 = 2$ sont deux valeurs pertinentes en densité. Cette stratégie n'est pas toujours pertinente pour la méthode mixte. En effet, Mabon (2016) compare la qualité des estimations en choisissant $\kappa_1 = \kappa_2$ et $\kappa_2 = 2\kappa_1$ dans le cadre d'un modèle de convolution pour des variables positives pour l'estimation des fonctionnelles linéaires d'une densité. Il en ressort que les résultats étaient légèrement meilleurs pour le premier choix ($\kappa_1 = \kappa_2$). Il faut peut-être des tests adaptés au problème traité pour discriminer entre $\kappa_1 = \kappa_2$ ou $\kappa_1 \neq \kappa_2$.

En outre, un travail récent de Lacour et al. (2017) propose l'idée de ne plus utiliser l'estimateur auxiliaire pour effectuer la sélection de fenêtre. En réadaptant cette idée ici, on sélectionne alors m par :

$$\widehat{m} := \arg \min_{m \in \mathcal{M}_n} \{\|\widehat{f}_m - \widehat{f}_{m_{\max}}\|^2 + \kappa V(m)\}, \quad m_{\max} = \max \mathcal{M}_n,$$

où $\kappa > 0$ une constante à calibrer numériquement. Cela a l'avantage d'être plus rapide du point de vue numérique car on n'a plus besoin de comparer tous les estimateurs deux à deux pour calculer $\widehat{A}(m)$. Par ailleurs, on a juste une constante à optimiser et la question de savoir comment calibrer les deux constantes ne se pose évidemment plus.

Dans le chapitre 2, nous généralisons en pratique la version dérivée de Goldenshluger et Lepski fondée sur les idées de Lacour et al. (2017) pour effectuer la sélection de fenêtre dans le cas d'estimations des dérivées d'une densité.

Comme pour la sélection de modèles par pénalisation, nous donnons dans la Figure 1.2, un exemple d'illustration de la sélection GL. On remarque que le choix de la procédure est aussi pertinent comparé aux 50 possibilités.

1.3.3 Inégalités de déviation

Nous présentons dans cette section deux inégalités de concentration qui sont des outils efficaces et incontournables pour obtenir des inégalités de type oracle. Ces deux inégalités sont : l'Inégalité de Bernstein et l'Inégalité de Talagrand. Elles permettent de contrôler les déviations de supremum de processus empiriques autour de leur moyenne.

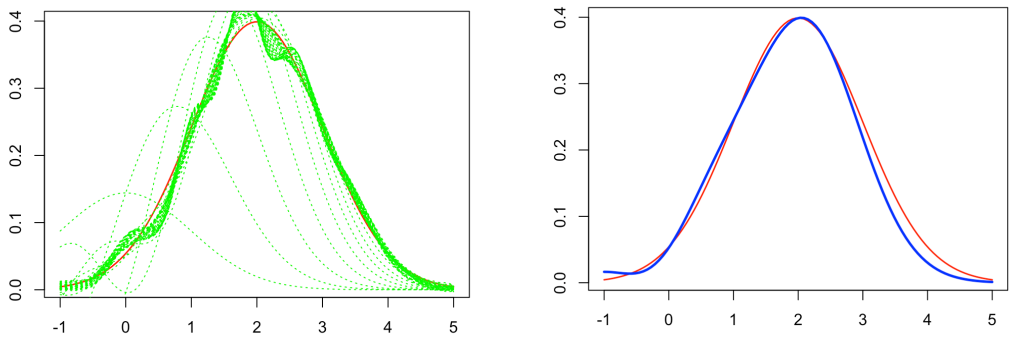


FIGURE 1.2 – Collection d'estimateurs par projection en base d'Hermite de la densité d'une $\mathcal{N}(2,1)$ en trait épais (rouge), la courbe sélectionnée à droite parmi 50 propositions données à gauche (en vert pointillés). La procédure GL sélectionne $\widehat{m} = 7$ pour $n = 1000$.

Inégalité de Talagrand

Les inégalités de Talagrand ont été prouvées dans Talagrand (1996), reformulées par Ledoux (1997). Le résultat ci-dessous en est une version donnée dans Klein and Rio (2005). Elle est utilisée dans les trois chapitres de la thèse.

Lemme 1.3.1. *Soient $(X_i)_{i\in\{1,\dots,n\}}$ une famille de variables aléatoires réelles indépendantes et $\mathcal{F}$ une classe dénombrable de fonctions mesurables. On définit pour tout $s \in \mathcal{F}$*

$$\nu_n(s) = \frac{1}{n}\sum_{i=1}^{n}(s(X_i) - \mathbb{E}[s(X_i)]).$$

Supposons qu'il existe trois constantes strictement positives M_1, H et v telles que :

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

alors pour tout $\delta > 0$, $C(\delta) = (\sqrt{1+\delta} - 1 \wedge 1)$ et $K_1 = 1/6$, nous avons

$$\mathbb{E}\left[\left(\sup_{s\in\mathcal{F}}|\nu_n^2(s)| - 2(1+2\delta)H^2\right)_+\right] \leqslant \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\delta\frac{nH}{v}} + \frac{49M_1^2}{K_1C^2(\delta)n^2}e^{-\frac{\sqrt{2}K_1C(\delta)\sqrt{\delta}}{7}\frac{nH}{M_1}}\right).$$

Sous certaines conditions, on peut étendre le résultat précédent à des classes de fonctions non dénombrables.

Inégalité de Bernstein

Les inégalités de Bernstein fournissent des bornes de déviation d'une somme de variables centrées. La preuve du résultat qui suit se trouve dans Massart (2007). Le lemme suivant est utilisée dans le chapitre 2.

Lemme 1.3.2. *Soit $X_1, \dots X_n$ des variables aléatoires réelles indépendantes et posons*

$$S_n = \sum_{i=1}^{n} (X_i - \mathbb{E}[X_i]).$$

On suppose qu'il existe deux constantes s^2 et b, telles que $\mathrm{Var}(X_i) \leqslant s^2$ et $|X_i| \leqslant b$ presque sûrement (p.s.), alors pour tout x positif, nous avons

$$\mathbb{P}\left(|S_n| \geqslant \sqrt{2ns^2x} + \frac{bx}{3}\right) \leqslant 2e^{-x}.$$

Classiquement, le nombre x est choisi comme étant un multiple de $\log(n)$.

1.4 Les résultats obtenus

Dans cette section, nous présentons un résumé des principaux résultats obtenus dans cette thèse. Nous commençons d'abord par faire un point des résultats existants pour chaque problème traité puis donner à la fin nos résultats.

1.4.1 Estimation adaptative optimale des dérivées d'une densité sur $\mathbb{R}$ ou $\mathbb{R}^+$.

Contexte Le chapitre 2 est consacré à l'estimation de la dérivée d'ordre d d'une densité, notée $f^{(d)}$, à partir des observations $X_1, \dots, X_n$ i.i.d. de densité f par rapport à la mesure de Lebesgue. Le problème d'estimation de la densité f est largement étudié dans la littérature. Beaucoup d'informations pertinentes sont contenues dans les dérivées d'une densité. Des exemples d'applications sont données dans Singh (1977a) et Sasaki et al. (2016). Les cas le plus communs sont pour $d = 0, 1, 2$. La première dérivée peut être utilisée pour la recherche de modes dans le cas du modèle de mélange et en analyse des données voir Cheng (1995), Chacón and Duong (2013). La dérivée seconde quand à elle peut être utilisée pour estimer une paramètre d'une famille exponentielle (voir Genovese et al. (2016)), de développer un test pour des modes (voir Cheng (1995)), de sélectionner la fenêtre optimale pour l'estimation d'une densité (voir Silverman (1978)). Enfin, les dérivées d'une densité donnent aussi de l'information sur la pente d'une courbe, les extrema locaux, les points selles.... Deux exemples spécifiques sont détaillés dans l'introduction du chapitre.

Des méthodes à noyau, par projection et bayésiennes ont été développées pour estimer les dérivées d'une densité : voir Bhattacharya (1967), Silverman (1978) ,Chacón et al. (2011) pour les méthodes à noyau ; Efromovich (1998), Rao (1996), Schmisser (2013) (pour le cas dépendant) pour les procédures de projection ; Shen and Ghosal (2017) pour l'approche bayésienne.

Dans ce chapitre, nous proposons un estimateur par projection en exploitant les relations entre les fonctions de Laguerre ou Hermite et leur dérivées.

Hypothèses et estimateur Nous considérons les hypothèses suivantes :

(**H**1) La densité f est d-fois dérivable et $f^{(d)}$ appartient à $\mathbb{L}^2(\mathbb{R}^+)$ pour le cas Laguerre ou $\mathbb{L}^2(\mathbb{R})$ pour le cas Hermite.

(**H**2) Pour tout r entier, $0 \leqslant r \leqslant d-1$, nous avons $\|f^{(r)}\|_\infty = \sup_{x\in\mathbb{R}}|f^{(r)}(x)| < +\infty$.

(**H**3) Pour tout r entier, $0 \leqslant r \leqslant d-1$, $\lim_{x\to 0} f^{(r)}(x) = 0$

L'hypothèse (**H**3) est spécifique au cas Laguerre, elle évite un problème au bord de l'intervalle. Elle exclut des densités classiques comme la distribution exponentielle qui correspond parfois à la première fonction de la base de Laguerre, $\varphi_0 = \sqrt{2}e^{-x}\mathbb{1}_{x\geqslant 0}$, dont une dimension ($m = 1$) suffit pour l'estimer. On verra dans la suite qu'on peut s'affranchir de cette hypothèse mais avec des performances moins bonnes. Sous les hypothèses (**H**1) à (**H**3), en utilisant la méthode des moments, on introduit l'estimateur suivant

$$\widehat{f}_{m,(d)} = \sum_{j=0}^{m-1} \widehat{a}_j^{(d)}\varphi_j, \quad \text{avec} \quad \widehat{a}_j^{(d)} = \frac{(-1)^d}{n}\sum_{i=1}^{n}\varphi_j^{(d)}(X_i), \tag{1.18}$$

où φ_j la base Laguerre sur $\mathbb{R}^+$ ou d'Hermite sur $\mathbb{R}$ (voir (1.7) et (1.9) respectivement pour les définitions).

Notons que pour $d = 0$ dans (1.18), on retrouve un estimateur classique de la densité.

Généralement pour la méthode à noyau, on obtient un estimateur des dérivées de f en dérivant l'estimateur à noyau de f. Avec les méthodes de projection, ce n'est pas bien adapté car la dérivée d'une base orthonormée n'est pas une base orthonormée. C'est pourquoi, notre stratégie est différente : on projette directement la dérivée d'ordre d dans une base orthonormée. On note dans la suite la projection orthogonale de $f^{(d)}$dans $S_m = \text{Vect}(\varphi_0, \ldots, \varphi_{m-1})$ par

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

Borne pour le risque Dans l'optique de calculer la vitesse de convergence, on établit d'abord la majoration suivante pour le risque de f :

Théorème 1.4.1. *Sous les hypothèses* (**H**1) *à* (**H**3) *et si de plus*

$$\mathbb{E}[X_1^{-d-1/2}] < +\infty \text{ pour le cas Laguerre et } \mathbb{E}[|X_1|^{2/3}] < +\infty \text{ pour le cas Hermite.} \tag{1.19}$$

Alors, pour $m \geqslant d$ assez grand, nous avons que

$$\mathbb{E}\big[\|\widehat{f}_{m,(d)} - f^{(d)}\|^2\big] \leqslant \|f_{m,(d)} - f^{(d)}\|^2 + C\frac{m^{d+\frac{1}{2}}}{n} - \frac{\|f_m^{(d)}\|^2}{n}, \tag{1.20}$$

où C est une constante qui ne dépend que des conditions de moment données en (1.19).

Pour le cas Hermite, la condition de moment $\mathbb{E}[|X_1|^{2/3}] < +\infty$ n'est pas nécessaire (voir "Lemma 1" dans Comte and Lacour (2021)).

Donc pour $f \in W_L^s(D)$, boule de Sobolev-Laguerre (un peu modifié voir Section 2.2.2 pour plus de détails) et $f \in W_H^s(D)$ (boule de Sobolev-Hermite), on en déduit pour le choix $m_{opt} = [n^{2/(2s+1)}]$

$$\mathbb{E}\big[\|\widehat{f}_{m_{opt},(d)} - f^{(d)}\|^2\big] \leqslant C_{(s,d,D)} n^{-\frac{2(s-d)}{2s+1}}, \tag{1.21}$$

où $C_{(s,d,D)}$ dépend uniquement de s, d et D. Cette vitesse est la même que celle obtenue par Schmisser (2013) pour le cas dépendant et par Giné and Nickl (2016). Notons que les ordres de grandeur du biais et de la variance sont spécifiques à notre méthode : le rôle de la dimension est joué ici par $\sqrt{m}$. Notons également que notre hypothèse de régularité est naturellement faite sur f et non sur $f^{(d)}$ (contrairement à ce qui apparait dans Rao (1996) and Lepski (2018)). Cette vitesse est meilleure que celle obtenue par Rao (1996) dans le cas i.i.d. si on considère la même condition de régularité. De façon intéressante, la dimension m_{opt} ne dépend pas de d. Cela coïncide avec la stratégie de Lepski (2018) qui injectait la fenêtre optimale de l'estimateur à noyau d'une densité dans l'estimation de ces dérivées. Pour $d = 0$ dans (1.21), on retrouve la vitesse optimale pour l'estimation de f pour des classes d'Hölder, Sobolev voir Tsybakov (2009).

Nous soulignons pour des densités gaussiennes (cas Hermite) ou gamma (cas Laguerre) que le biais est à décroissance exponentielle. On obtient donc une vitesse plus rapide que le cas général : $\log(n)^{d+1/2}/n$.

Dans la Section 2.2.3, nous prouvons également des minorations qui assurent l'optimalité de la vitesse au sens du minimax pour les classes considérées. Le résultat est résumé dans le théorème suivant :

Théorème 1.4.2. *Soit $s > d$ un entier et $\widetilde{f}_{n,d}$ un estimateur quelconque de $f^{(d)}$. Alors pour n assez grand, nous avons*

$$\inf_{\widetilde{f}_{n,d}} \sup_{f \in W^s(D)} \mathbb{E}[\|\widetilde{f}_{n,d} - f^{(d)}\|^2] \geqslant c n^{-\frac{2(s-d)}{2s+1}},$$

où $W^s(D) = W_L^s(D)$ pour le cas Laguerre ou $W^s(D) = W_H^s(D)$ pour le cas Hermite et $c > 0$ une constante qui dépend de s et d uniquement.

Précisons qu'un résultat similaire pour des espaces de Lipschitz périodique est donné dans Efromovich (1999), voir aussi Lepski (2018) pour les classes de Nikol'ski.

Procédure adaptative Dans la Section 2.2.4, nous décrivons une procédure de sélection de modèle par pénalisation pour choisir la dimension pertinente. On introduit une pénalité aléatoire qui est en fait un estimateur d'un majorant de la variance. Notons que la pénalité est indépendante de la base utilisée. En outre, nous démontrons une inégalité oracle en utilisant les inégalités de Talagrand et de Bernstein permettant de réaliser automatiquement le compromis biais-variance dans (1.20).

Analogie avec les techniques noyaux et autres résultats Ensuite, dans la Section 2.3 nous comparons pour $d = 1$, notre estimateur $(\widehat{f}_{m,(1)})$ à la dérivée de l'estimateur par

projection de la densité :

$$(\widehat{f}_m)^{(d)} = \sum_{k=0}^{m-1} \widehat{a}_k^{(0)} \varphi_k^{(d)}, \quad \widehat{a}_k^{(0)} = \frac{1}{n} \sum_{i=1}^{n} \varphi_k(X_i).$$

Nous donnons des éléments pour s'affranchir de l'hypothèse (**H**3) (c'est-à-dire que $f(0) = 0$ pour $d = 1$). Par ailleurs, nous obtenons que notre estimateur converge à la même vitesse que l'estimateur $(\widehat{f}_m)^{(1)}$ pour le cas Hermite mais concernant le cas Laguerre notre stratégie est plus performant mais avec (**H**3). Ansi, une correction de l'estimateur introduit en (1.18) sans (**H**3) est aussi effectuée dans la section 2.3. On définit l'estimateur suivant avec $f(0) \neq 0$:

$$\widetilde{f}'_{m,K} = \sum_{j=0}^{m-1} \widehat{a}_{j,K}^{(1)} \ell_j, \quad \widehat{a}_{j,K}^{(1)} = -\frac{1}{n} \sum_{i=1}^{n} \ell'_j(X_i) - \widehat{f}_K(0)\ell_j(0),$$

où ℓ_j désigne la base de Laguerre définie en (1.7) et $\widehat{f}_K(0)$ estime $f(0)$ avec

$$\widehat{f}_K = \sum_{j=0}^{K-1} \widehat{a}_j^{(0)} \ell_j, \quad \widehat{a}_j^{(0)} = \frac{1}{n} \sum_{i=1}^{n} \ell_j(X_i).$$

Pour $\widetilde{f}'_{m,K}$, on obtient la même vitesse que $(\widehat{f}_m)^{(1)}$. De plus, nous prouvons une inégalité oracle pour $K := K_n$ fixé pour l'estimateur $\widetilde{f}'_{\widehat{m}_K,K}$ où $\widehat{m}_K$ dépend de K en utilisant une procédure par pénalisation.

Simulations En Section 2.4, nous présentons des simulations numériques pour illustrer les résultats théoriques et une comparaison avec la procédure à noyau pour une fenêtre choisie en s'inspirant de la méthode de Lacour et al. (2017) est effectuée.

Enfin, le chapitre se termine ainsi par la démonstration des résultats énoncés et des outils théoriques d'analyse et de probabilité.

1.4.2 Déconvolution d'une densité sur $\mathbb{R}$ en base d'Hermite

Contexte et modèle Dans ce chapitre, nous considérons le modèle de convolution

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

où

(**H**1) les $(X_k)_{k\geqslant 1}$ sont i.i.d. de densité f inconnue par rapport à la mesure de Lebesgue,

(**H**2) les $(\varepsilon_k)_{k\geqslant 1}$ sont i.i.d. de densité f_ε connue, par rapport à la mesure de Lebesgue,

(**H**3) les suites $(X_k)_{k\geqslant 1}$ et $(\varepsilon_k)_{k\geqslant 1}$ sont indépendantes.

On cherche à estimer f à partir des données $Z_1, \dots, Z_n$. Notons f_Z la densité de Z_1. Sous (**H**3) nous avons que $f_Z = f * f_\varepsilon$, où $g * h$ désigne le produit de convolution entre g et h, $g * h(x) = \int g(x-y)h(y)dy$, c'est ce qui explique le terme de *"déconvolution"* pour l'estimation de f. Outre la régularité de f, la régularité de f_ε influe sur la vitesse de convergence des estimations. Elle est décrite par la décroissance de la transformée de Fourier de f_ε. Nous considérons les deux hypothèses classiques en déconvolution sur la loi du bruit.

(**H4**) $\forall u \in \mathbb{R}$, $f_\varepsilon^*(u) \neq 0$, où t^* désigne la transformée de Fourier de t, $t^*(u) = \int e^{ixu} t(x)dx$. On suppose aussi qu'il existe $c_1 \geqslant c_1' > 0$, et $\gamma \geqslant 0, \mu \geqslant 0, \delta \geqslant 0$ (avec $\gamma > 0$ si $\delta = 0$) tels que

$$c_1'(1+u^2)^\gamma e^{\mu|u|^\delta} \leqslant \frac{1}{|f_\varepsilon^*(u)|^2} \leqslant c_1(1+u^2)^\gamma e^{\mu|u|^\delta}. \tag{1.22}$$

Si $\delta = 0$ dans (1.22), la loi du bruit f_ε et les erreurs sont appelées ordinairement régulières ("ordinary smooth"), sinon elles sont dites super régulières ("super smooth"). Soulignons que la condition (1.22) implique (**H4**) et est vérifiée par certaines distributions classiques : on peut citer à titre d'exemple Laplace (avec $\delta = 0$ et $\gamma = 2$), Gamma ($\delta = 0$ et $\gamma = p$, où p est le paramètre de forme), Gaussienne ($\gamma = 0$ et $\delta = 2$), Cauchy ($\gamma = 0$ et $\delta = 1$).

Le problème de déconvolution a été beaucoup étudié dans la littérature. Les premiers travaux ont été menés dans le cas non adaptatif par : Carroll and Hall (1988), Fan (1991), Fan (1993) et le livre de Meister (2009) sur le sujet. La procédure d'estimation adaptative a été considérée en premier par : Pensky and Vidakovic (1999) puis massivement utilisée entre autres par Comte and Lacour (2011) (cas bruit inconnu), Mabon (2017) dans le cas où $X_k \geqslant 0$.... Fan (1993), Butucea (2004), Butucea and Tsybakov (2007) établissent l'optimalité des vitesses au sens minimax. Des généralisations multidimensionnelles (Comte and Lacour (2013)) ou sur la sphère (Kerkyacharian et al. (2011)) ont ensuite été considérées.

Nous présentons ici une nouvelle procédure fondée sur un développement en base d'Hermite, inspirée d'idées développées dans Comte and Genon-Catalot (2018). L'originalité de ce travail est d'utiliser la base d'Hermite qui est notée φ_j. La transformée de Fourier d'une base d'Hermite est à constante près une base d'Hermite, $\varphi_j^* = \sqrt{2\pi}(i)^j\varphi_j$, cela nous permet de définir un estimateur par projection.

Définition de l'estimateur On considère le contraste de minimisation suivant :

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

En supposant que le ratio $\varphi_j/f_\varepsilon^*$ est intégrable pour $j = 0, \ldots, m-1$, on définit un estimateur de f en minimisant le contraste ci-dessus pour tout $t \in S_m$:

$$\widehat{f}_m = \sum_{j=0}^{m-1} \widehat{a}_j\varphi_j, \qquad \widehat{a}_j = \frac{(-i)^j}{\sqrt{2\pi}}\int \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)}\varphi_j(u)du, \tag{1.23}$$

où $\widehat{f}_Z^*(t) = \frac{1}{n}\sum_{k=1}^n e^{itZ_k}$ estime sans biais la fonction caractéristique de Z_1. La spécificité de la base d'Hermite qui décroit en e^{-cx^2} rend la quantité $\varphi_j/f_\varepsilon^*$ intégrable pour beaucoup de fonctions f_ε. Notons que le coefficient $\widehat{a}_j$ est réel, en effet :

$$\overline{\widehat{a}_j} = ((-i)^j/\sqrt{2\pi})\int \widehat{f}_Z^*(-u)\varphi_j(u)/f_\varepsilon^*(-u)du = ((i)^j/\sqrt{2\pi})\int \widehat{f}_Z^*(u)\varphi_j(-u)/f_\varepsilon^*(u)du = \widehat{a}_j$$

puisque $\varphi_j(-x) = (-1)^j\varphi_j(x)$. C'est aussi un estimateur naturel de $a_j(f) = \langle f, \varphi_j\rangle = \frac{1}{2\pi}\int f_Z^*(f_\varepsilon^*)^{-1}\varphi_j^*$ d'après le Théorème de Plancherel-Parseval. Donc $\widehat{f}_m$ estime sans biais

le projeté de f : $f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j$ contrairement à l'estimateur obtenu par Comte and Genon-Catalot (2018) où $\widehat{a}_j$ est remplacé par :

$$\widetilde{a}_{j,\sqrt{m}} = ((-i)^j/\sqrt{2\pi}) \int_{|u|\leqslant \pi\sqrt{m}} \widehat{f}_Z^*(u)\varphi_j(u)/f_\varepsilon^*(u)du.$$

Borne du risque Sous l'hypothèse additionnelle :

(**H5**) $\|f_Z\|_\infty = \sup_{x\in\mathbb{R}} |f_Z(x)| < \infty$,

on a la décomposition biais-variance suivante (voir Proposition 3.3.1 avec $l = 2$) :

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leqslant \|f - f_m\|^2 + \frac{1}{\pi n}\int_{|u|\leqslant\sqrt{2m}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{c}{n}. \tag{1.24}$$

où c est une constante qui dépend de $\|f_Z\|_\infty$, γ, μ, δ et d'autres constantes liées à la base d'Hermite. Le premier terme à droite de l'inégalité (1.24) ($\|f_m - f\|^2 = \sum_{j\geqslant m} a_j(f)^2$) est le terme de biais, il mesure la distance entre f et f_m au sens de $\mathbb{L}^2(\mathbb{R})$. C'est un terme décroissant en m. Le deuxième terme est le terme principal de la variance, il croît clairement avec m. Le dernier terme vient aussi du calcul de variance, c'est un terme résiduel. Par conséquent le risque minimal s'obtient en faisant un compromis biais-variance. Si la densité f appartient à la boule de Sobolev-Hermite de régularité s et de rayon D notée par $W_H^s(D)$, nous obtenons les vitesses de convergence suivante pour l'oracle $\widehat{f}_{m_{opt}}$. Ces vitesses sont connues pour être optimales. Elles coïncident avec celles obtenues

	$\delta = 0$	$0 < \delta < 2$ or $\delta = 2,\ \mu < \xi \leqslant 1/2$
m_{opt}	$[n^{\frac{2}{2s+2\gamma+1}}]$	$\left[\frac{1}{2}\left(\frac{\log n}{2\mu}\right)^{\frac{2}{\delta}}\right]$
Vitesse	$n^{-\frac{2s}{2s+2\gamma+1}}$	$(\log n)^{-\frac{2s}{\delta}}$

TABLE 1.1 – Vitesse de convergence pour le MISE si $f \in W_H^s(D)$.

par Fan (1993), Pensky and Vidakovic (1999) respectivement calculées sur les classes de Hölder et de Sobolev. Nous fournirons aussi des taux de convergence pour des densités de mélange gaussienne. Les résultats obtenus sont les mêmes que dans Butucea (2004) pour des fonctions f super-régulière.

Ensuite, nous montrons que les résultats obtenus jusqu'ici s'étendent au cas de variables dépendantes. En effet, on peut remplacer l'hypothèse (**H**1) par :

(**H′**1) $(X_k)_{k\geqslant 1}$ est strictement stationnaire et β-mélangeant (voir Section 3.3.5 pour la définition).

Dans la Proposition 3.3.3, on obtient la décomposition biais-variance suivante sous des conditions classiques sur le coefficient de mélange et de moment :

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leqslant \|f - f_m\|^2 + \frac{1}{\pi n}\int_{|u|\leqslant\sqrt{2m}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{c}{n} + c'\frac{\sqrt{m}}{n},$$

où c est une constante positive qui dépend par exemple de $\|f_Z\|_\infty$, γ, μ, δ, et c' est une constante dépendant des coefficients de mélange et d'une condition de moment sur X_1. Le biais et la variance principale sont les mêmes que pour le cas i.i.d. avec un terme additionnel $c'\frac{\sqrt{m}}{n}$ qui est spécifique au cas β-mélangeant. Comme $|f_\varepsilon^*(u)| \leqslant 1$, nous avons $\frac{1}{\pi}\int_{|u|\leqslant\sqrt{2m}} \frac{du}{|f_\varepsilon^*(u)|^2} \geqslant \frac{2\sqrt{2}}{\pi}\sqrt{m}$. Ainsi, le terme $c'\frac{\sqrt{m}}{n}$ est aussi un reste qui est négligeable que le terme principal de la variance. D'où les mêmes vitesses que pour le cas i.i.d.

Adaptation De plus, nous proposons une méthode pour choisir automatiquement le meilleur modèle m dans une collection pour le cas i.i.d.. On utilise la procédure par pénalisation du type Birgé and Massart (1997) et on prouve une inégalité oracle en utilisant l'inégalité de Talagrand (voir Section 1.3.3). La procédure complète est décrite dans la Section 3.4.

Simulations Enfin, nous illustrons numériquement la procédure adaptative et des comparaisons avantageuses avec la stratégie de Comte and Lacour (2011) et le problème direct est effectué.

La preuve des résultats théoriques (Section 3.7) et quelques outils d'analyse et de probabilité terminent ce chapitre.

1.4.3 Estimation adaptative dans un modèle de régression en base d'Hermite

Contexte Le chapitre 4 est dédié à l'estimation d'une fonction de régression dans un modèle de convolution-régression. On considère le modèle de convolution suivant :

$$y(x_k) = h(x_k) + \varepsilon_k, \quad k = -n, \dots, n-1, \tag{1.25}$$

où

$$h(x) = f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy, \tag{1.26}$$

où la fonction g appelée *noyau* est supposée connue et f est la fonction inconnue qu'on cherche à estimer ; les erreurs $(\varepsilon_k)_{-n\leqslant k\leqslant n-1}$ sont i.i.d. avec $\mathbb{E}[\varepsilon_k] = 0$ et $\mathrm{Var}(\varepsilon_k) = \sigma_\varepsilon^2 < \infty$, connu ; les points $(x_k = kT/n)_{-n\leqslant k\leqslant n-1}$ sont déterministes et équirépartis sur $[-T, T]$, où $0 < T < \infty$ est fixé. Ce modèle apparait dans beaucoup de domaines d'applications : en analyse des données d'imagerie dynamique à contraste amélioré (voir (Goh et al. (2005), Cuenod et al. (2006), Goh et al. (2007), et Comte et al. (2017)) et dans l' étude du spectroscopie de fluorescence résolue en temps (voir Gafni et al. (1975), McKinnon et al. (1977), O'Connor et al. (1979), Ameloot and Hendrickx (1983), Abramovich et al. (2013)). L'estimation de la fonction inconnue h à partir des données $(y(x_k), x_k)$ est connue comme problème de régression non paramétrique en "fixed design". Ce problème a été largement étudié dans la littérature, voir Barron et al. (1999), Baraud (2000) par exemple. Lorsqu'on cherche à estimer la densité f d'une variable aléatoire X quand on observe $Z = X + \varepsilon$ avec ε indépendant de X de densité g se réduit à reconstruire f à partir d'un estimateur de $f_Z = f \star g$. Ce problème est connu comme un problème de déconvolution. Il s'agit d'un problème inverse qui est étudié aussi intensivement dans la littérature, voir par exemple Carroll and

Hall (1988), Fan (1991), Pensky and Vidakovic (1999). Nous l'avons aussi étudié dans le Chapitre 3 en utilisant la base d'Hermite (voir Section 1.4.2 pour plus détails). Le modèle (1.25) cumule deux questions « régression et déconvolution», c'est pourquoi il est difficile. Les résultats théoriques sur ces deux problèmes traités *indépendamment* sont bien connus. Notons que les fonctions f et g ne sont pas nécessairement des densités dans ce chapitre. Le problème (1.25) a été étudié dans la littérature pour des fonctions particulières. Si f et g sont à support dans $[0,1]$, Rice and Rosenblatt (1983) ont résolu le problème en supposant que f est de classe C^4 et en utilisant un spline de lissage pour $x_k = k/n$ avec $k = 1, \dots, n$. Le modèle (1.25) peut être vu comme une généralisation de celui de la convolution de Laplace. En effet si on impose que f et g sont à support dans $\mathbb{R}^+$, nous avons que : $h(x) = \int_0^x f(x-y)g(y)dy$ dont la version stochastique discrète est donnée par (1.25) avec $k = 1, \dots, n$. Il a été étudié dans Dey et al. (1998) pour $g(x) = be^{-ax}1_{x\geqslant 0}$ et en utilisant que la solution de (1.26) satisfait une équation différentielle linéaire. Récemment, Abramovich et al. (2013) ont étudié le problème de la déconvolution de Laplace pour g connu : les auteurs résument le problème d'estimation de f à l'estimation des dérivées de h. Ces dérivées sont estimées par une méthode à noyau. Vareschi (2015) a aussi étudié le problème de Laplace déconvolution en utilisant la projection de Galerkin sur les fonctions de Laguerre pour un noyau g contaminé par un bruit blanc. Comte et al. (2017) ont proposé un estimateur par projection, fondé sur un développement des fonctions f, g et h en base de Laguerre. Enfin, Benhaddou et al. (2019) utilise aussi un développement en base de Laguerre pour une fonction f de trois variables (2 variables en espace et une variable en temps) et g une fonction d'une variable.

L'ensemble de ces études ont été menées pour des fonctions f et g à support dans $\mathbb{R}^+$ ou $[0,1]$. Dans ce travail, on considère le Modèle (1.25) avec des fonctions à support dans $\mathbb{R}$. La base de Laguerre n'est clairement pas adaptée à notre problème général. On considère la base d'Hermite qui est à support $\mathbb{R}$ et est bien adaptée à ce contexte.

Nous considérons les hypothèses classiques suivantes en déconvolution sur le noyau g :

(**H1**) La transformée Fourier de g notée g^* est bien définie tel que : $g^* \neq 0$, où $t^*(u) = \int e^{iux}t(x)dx$, et i est le nombre complexe avec $i^2 = -1$.

(**H2**) Il existe des constantes c_1, c_1', γ, $c_1 \geqslant c_1' > 0$ et $\gamma > 0$ telles que

$$c_1'(1+t^2)^\gamma \leqslant |g^*(t)|^{-2} \leqslant c_1(1+t^2)^\gamma, \quad \forall t \in \mathbb{R}. \tag{1.27}$$

Nos objectifs sont les suivants : introduire un estimateur consistant de f ; donner des résultats de convergence ; proposer une procédure adaptative et illustrer numériquement sa performance. Considérons les observations discrètes $(x_k, y(x_k))_{-n\leqslant k\leqslant n-1}$ issues du modèle (1.25). Le but est d'estimer f à partir des données $(x_k, y(x_k))_{-n\leqslant k\leqslant n-1}$. Différentes approches sont proposées dans le chapitre : mixte Fourier-projection et projection-projection.

Premier estimateur : approche "Fourier-Hermite" (notée FH dans la suite).

Procédure d'estimation Pour ce faire, nous avons besoin de l'hypothèse suivante :

(**H3**) La fonction f et sa transformée de Fourier f^* sont intégrables.

Cette hypothèse est spécifique à cette première stratégie et est nécessaire pour utiliser la transformée de Fourier inverse.

Sous (**H3**), (**H1**), en utilisant la transformée de Fourier inverse, nous avons

$$f(x) = \int_{\mathbb{R}} e^{-iux} \frac{h^*(u)}{g^*(u)} du, \quad \forall x \in \mathbb{R}. \tag{1.28}$$

L'équation (1.28) est la clé pour obtenir un estimateur de f en remplaçant h par un estimateur. Pour reconstruire h, on utilise la méthode des moindres carrés fondée sur un développement de h en base d'Hermite. A cette fin, on pose

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

avec Φ_d^t est la transposée de la matrice Φ_d et φ_j est la base d'Hermite (voir (1.9) pour sa définition). Ainsi, on estime h par

$$\widehat{h}_d = \sum_{j=0}^{d-1} \widehat{b}_j^{(d)} \varphi_j, \quad \vec{\widehat{b}}^{(d)} = (\widehat{b}_0^{(d)}, \dots, \widehat{b}_{d-1}^{(d)})^t = (\Phi_d^t \Phi_d)^{-1} \Phi_d^t \vec{y} = \frac{T}{n} \Psi_d^{-1} \Phi_d^t \vec{y}, \tag{1.29}$$

où $\vec{y} = (y(x_{-n}), \dots, y(x_{n-1}))^t$ et $\widehat{h}_d$ est dans $S_d := \text{Vect}\{\varphi_0, \dots, \varphi_{d-1}\}$, l'espace linéaire engendré par $(\varphi_j)_{0 \leqslant j \leqslant d-1}$. Notons que la base d'Hermite rend la matrice Ψ_d inversible pour $n > d$. En prenant la transformée de Fourier de (1.29) et en injectant cela dans (1.28), nous introduisons un premier estimateur de f par :

$$\widehat{f}_{(d)}(x) = \frac{1}{2\pi} \int e^{-iux} \frac{\widehat{h}_d^*(u)}{g^*(u)} du. \tag{1.30}$$

L'estimateur est bien défini car la base d'Hermite, qui décroit en $e^{-\xi x^2}$ rend la fonction $\widehat{h}_d^*/g^*$ intégrable pour toutes les fonctions g telles que (1.27) est vraie $\varphi_j^* = \sqrt{2\pi}(i)^j \varphi_j$. La qualité de $\widehat{f}_{(d)}$ est clairement liée à celle de $\widehat{h}_d^*$. On se doute bien qu'en pratique, il faut introduire un cut-off pour calculer $\widehat{f}_{(d)}$. En outre, on verra que le contrôle du risque de $\widehat{f}_{(d)}$ dépend d'une version tronquée de $\widehat{f}_{(d)}$. Ainsi, on considère l'estimateur suivant

$$\widehat{f}_{(\ell),d}(x) = \frac{1}{2\pi} \int_{-\ell}^{\ell} e^{-iux} \frac{\widehat{h}_d^*(u)}{g^*(u)} du, \text{ for } \ell > 0. \tag{1.31}$$

Régression en fixe design en base d'Hermite En section 4.3, nous présentons une étude complète (théorique et numérique) de la régression non paramétrique en fixe design en base d'Hermite. Contrairement à Baraud (2000), on ne considère pas des bases à support compact. On obtient la borne suivante que nous exploitons dans la suite (voir Proposition 4.3.1) :

$$\mathbb{E}\left[\|\widehat{h}_d - h\|^2\right] \leqslant \sigma_\varepsilon^2 \frac{T}{n} \text{tr}\left(\Psi_d^{-1}\right) + \|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right) \|h - h_d\|_n^2. \tag{1.32}$$

Cette borne est nouvelle à notre connaissance. Il s'agit d'une décomposition biais-variance, de biais égal à $\lambda_{max}\left(\Psi_d^{-1}\right) \|h - h_d\|_n^2$ où h_d est le projeté de h dans S_d et de variance à $\sigma_\varepsilon^2 \text{tr}\left(\Psi_d^{-1}\right) T/n$. C'est en partie grâce à cette borne qu'on a pu établi une inégalité oracle pour l'estimateur de régression pour la norme $\mathbb{L}^2(\mathbb{R})$ (voir Theorem 4.3.1). En particulier, nous établirons une inégalité oracle du risque pour la norme $\mathbb{L}^2(\mathbb{R})$ qui est nouvelle à notre connaissance.

Borne du risque de $\widehat{f}_{(d)}$ et $\widehat{f}_{(\ell),d}$ Pour valider l'estimateur $\widehat{f}_{(d)}$, on étudie son risque. On introduit les notations suivantes

$$\Delta(\ell) = \sup_{|u|\leqslant\ell} |g^*(u)|^{-2}, \qquad f_{(\ell)}(x) = \frac{1}{2\pi}\int_{-\ell}^{\ell} e^{-iux}\frac{h^*(u)}{g^*(u)}du. \tag{1.33}$$

On considère les deux hypothèses suivantes :

(**H4**) Il existe une constante $\lambda > 0$ telle que

$$0 < \lambda_{max}(\Psi_d^{-1}) \leqslant \lambda < +\infty,$$

où $\lambda_{max}(A)$ désigne le maximum des valeurs propres de la matrice A.

(**H5**) $\|h\|_\infty = \sup_{x\in\mathbb{R}} |h(x)| < +\infty$.

La majoration du risque de $\widehat{f}_{(d)}$ et $\widehat{f}_{(\ell),d}$ est donnée dans la proposition suivante :

Proposition 1.4.1. *Supposons que les hypothèses* (**H1**) *et* (**H3**) *sont satisfaites. Alors, l'estimateur $\widehat{f}_{(\ell),d}$ défini en (1.31) vérifie*

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right] \leqslant \|f - f_{(\ell)}\|^2 + \Delta(\ell)\left(\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right) + \|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2\right), \tag{1.34}$$

où $\lambda_{max}(A)$ est le rayon spectral de la matrice A.

Si de plus les hypothèses (**H2**), (**H4**) *et* (**H5**) *sont vérifiées, nous avons pour $\widehat{f}_{(d)}$ donné en* (1.30) *et $\ell \geqslant \sqrt{2d}$*

$$\mathbb{E}\left[\|\widehat{f}_{(d)} - f\|^2\right] \leqslant 2C\lambda T e^{-\xi d} + 2\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right], \tag{1.35}$$

où C est une constante qui dépend de C'_∞, ξ qui sont tels que $|\varphi_j(x)| \leqslant C'_\infty e^{-\xi^2}$ pour $|x| \geqslant \sqrt{2j+1}$ et $\|h\|_\infty$.

(a) Le premier terme à droite de (1.34) ($\|f - f_{(\ell)}\|^2 = \frac{1}{2\pi}\int_{|u|>\ell} |f^*(u)|^2 du$) est le terme classique de biais : il décroît avec le cut-off ℓ.

(b) Le terme $\Delta(\ell)$ correspond à l'aspect déconvolution du problème : il est étudié en utilisant la condition de régularité sur g^* donnée dans (**H2**). Il croît avec le cut-off ℓ.

(c) Enfin, le terme dans la grande parenthèse représente le risque d'estimation de h (voir Équation (1.32))

L'étape d'après est donc de fournir des résultats de convergence pour $\widehat{f}_{(d)}$ et $\widehat{f}_{(\ell),d}$. Un résultat global est résumé dans le théorème suivant

Théorème 1.4.3. *Sous* (**H1**), ..., (**H4**), *pour $f \in W^s(D)$ (Sobolev classique) et $h \in W_H^{s+\gamma}(D')$ (Sobolev-Hermite), nous avons pour les choix $d_{opt} = [n^{1/(s+\gamma+1)}]$ et $\ell_{opt} = n^{\frac{1}{2(s+\gamma+1)}}$ avec $s + \gamma \geqslant 11/6$ que*

$$\sup_{f\in W^s(D)} \mathbb{E}\left[\|\widehat{f}_{(\ell_{opt}),d_{opt}} - f\|^2\right] \leqslant C(s,\gamma,D,T,\sigma_\varepsilon)n^{-\frac{s}{s+\gamma+1}},$$

où $C(s,\gamma,D,T,\sigma_\varepsilon)$ est une constante qui ne dépend que de $s,\gamma,L,T,\sigma_\varepsilon$ et γ donné en (**H2**).

Notons que ce résultat de convergence est valide aussi pour l'estimateur $\widehat{f}_{d_{opt}}$ avec l'hypothèse (**H5**), voir (1.35). Les estimateurs $\widehat{f}_{(\ell_{opt}),d_{opt}}$ et $\widehat{f}_{d_{opt}}$ convergent à une vitesse polynomiale comme en déconvolution d'une densité pour des fonctions ordinairement régulières.

Étonnamment, la difficulté ici est de contrôler le biais de la fonction de régression. On aimerait que l'hypothèse $h \in W_H^{s+\gamma}(D')$ soit une conséquence directe du fait que $f \in W^s(D)$ et g vérifie (**H2**) pour $s+\gamma$ entier. Les conditions satisfaisants cette hypothèse sont déferrées dans la section preuve du chapitre 4.7, Proposition 4.7.1. La difficulté provient du fait qu'on ne connait pas h et que les classes de régularité sont différentes pour f et g. De plus, nous soulignons que si $f \in W^s(D)$ et g vérifie (**H2**), on peut monter par un calcul élémentaire que $h \in W^{s+\gamma}(D/c_1)$ où c_1 est donné en (**H2**). Nous avons aussi calculé des vitesses de convergence pour des fonctions f et g spécifiques. En particulier si le noyau g est gaussien, on retrouve la même vitesse qu' en déconvolution d'une densité voir aussi le Chapitre 3. Les résultats sont présentés dans le tableau 1.2 suivant :

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 1.2 – Vitesse de convergence du MISE pour $\widehat{f}_{(\ell_{opt}),d_{opt}}$ dans les cas spécifiques.

Adaptation pour $\widehat{f}_{(\ell),d}$ D'après le Théorème 1.4.3, le choix pertinent du cut-off est d'ordre $\sqrt{d}$, pour cette raison, on pose $\ell = \sqrt{2d}$ qui est la valeur minimale admissible. Dans la Section 4.4.4, on propose une procédure de sélection de modèle en s'inspirant de la méthode de Goldenshluger and Lepski (2011) et nous démontrons une inégalité oracle : l'estimateur résultant réalise automatiquement un compromis-biais variance à un facteur logarithmique près. La procédure est décrite en Section 4.4.4.

Seconde stratégie. Dans un second temps, nous introduisons une autre approche appelée Hermite stratégie (notée HH) fondée sur un développent de f et h en base d'Hermite.

Définition de l'estimateur On construit l'estimateur suivant en remplaçant h par son estimateur des moindres carrés dans $a_j(f) = \langle h^*(g^*)^{-1}, \varphi_j \rangle$:

$$\widehat{f}_{m,d} = \sum_{j=0}^{m-1} \widehat{a}_{j,d}\varphi_j, \quad \widehat{a}_{j,d} = \frac{(-i)^j}{\sqrt{2\pi}} \int \frac{\widehat{h}_d^*(u)}{g^*(u)} \varphi_j(u)du, \tag{1.36}$$

à condition que $\widehat{h}_d^*\varphi_j/g^*$ est intégrable pour $j = 0, \ldots, m-1$. Notons que les coefficients $\widehat{a}_j$ sont réels et bien définis car la base d'Hermite rend la quantité $\widehat{h}_d^*\varphi_j/g^*$ intégrable. Nous

avons un estimateur qui dépend donc de deux paramètres qui doivent être optimisés. Comme pour la stratégie FH, la qualité de $\widehat{f}_{m,d}$ dépend aussi de celle de $\widehat{h}_d$. La borne du risque obtenue (voir Proposition 4.5.1) est donnée par :

Borne du risque de $\widehat{f}_{m,d}$ et comparaison avec l'approche FH

Proposition 1.4.2. *Supposons f et h sont de carré intégrable et posons*

$$\Sigma(m) = \sup_{|u|\leqslant\sqrt{\rho m}} |g^*(u)|^{-2} + \sum_{j=0}^{m-1} \int_{|u|\geqslant\sqrt{\rho m}} |\varphi_j(u)|^2 |g^*(u)|^{-2} du, \quad \rho > 0. \tag{1.37}$$

Pour $\widehat{f}_{m,d}$ donné en (1.36), nous avons

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] \leqslant \|f - f_m\|^2 + 2\Sigma(m)\left(\|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2 + \sigma_\varepsilon^2 \frac{T}{n}\left(\mathrm{tr}\left(\Psi_d^{-1}\right) \wedge 2\pi^2 m\right)\right). \tag{1.38}$$

La différence avec la borne obtenue pour la stratégie FH se trouve dans le biais de f ($\|f - f_m\|^2$) et $\Sigma(m)$, l'étape de la partie régression ne change bien évidement pas. Le terme $\sum_{j=0}^{m-1} \int_{|u|\geqslant\sqrt{\rho m}} |\varphi_j(u)|^2 |g^*(u)|^{-2} du$ est à décroissance exponentielle si $\rho \geqslant 2$ et donc négligeable par rapport au terme $\sup_{|u|\leqslant\sqrt{\rho m}} |g^*(u)|^{-2} = \Delta(\sqrt{\rho m})$. Donc, pour $f \in W_H^s(D)$ et $\ell \asymp \sqrt{m}$, on retrouve la même vitesse que pour la stratégie FH. L'ensemble des résultats sur la deuxième méthode est dans la Section 4.5.

Simulations La fin du chapitre 4 est dédiées à des simulations numériques dans la section 4.6 pour illustrer la procédure adaptative de l'approche FH. Dans le but de donner un tableau du risque, nous implémentons aussi une stratégie dérivant de la méthode de Goldenshluger et Lepski en s'inspirant de la procédure de Lacour et al. (2017). Les résultats obtenus sont satisfaisants mais la procédure semble trop lente. Il faut certainement la compléter en implémentant la procédure HH.

Enfin, nous donnons la preuve des résultats théoriques et des outils d'analyse et de probabilité à la fin chapitre.

Nous terminons ce chapitre en soulevant quelques points susceptibles de faire l'objet de futurs travaux.

1.5 Perspectives de recherche

Conclusions Dans ces travaux, nous avons construit des estimateurs non paramétriques ayant de bonnes propriétés statistiques. Les études du risque montrent que dans chaque cas nous avons un compromis biais-variance à effectuer pour déduire des vitesses de convergence. Nous avons développé des procédures adaptatives pour choisir les paramètres pertinents de chaque méthode utilisée. De plus, nous prouvons aussi des inégalités de type *oracle* grâce aux inégalités de déviations (voir Section 1.3.3). Les procédures développées

sont facilement implémentables et des comparaisons avec d'autres méthodes existantes sont exposées dans les chapitres 2 et 3. Les résultats obtenus dans ce manuscrit ouvrent des pistes pour des futurs travaux. Énumérons sans tarder quelques pistes.

Estimation des dérivées d'une densité Dans le chapitre 2, nous avons étudié le problème d'estimation des dérivées d'une densité en supposant que les variables étaient i.i.d., ce cas ne reflète pas toujours la complexité des données. Il serait donc intéressant d'enquêter sur le cas où les variables sont dépendantes. Des résultats existent dans le contexte où les variables sont supposées β-mélangeantes (faiblement dépendantes) avec une pénalité qui dépend des coefficients de mélange inconnus. Une piste à explorer serait de voir s'il est possible d'étendre ces résultats pour une pénalité indépendante des coefficients de mélange, un peu comme dans Comte et al. (2008) en contexte de déconvolution. En outre, nous avons généralisé numériquement la méthode dite de PCO introduite dans Lacour et al. (2017) pour un estimateur à noyau de la dérivée d'une densité. Une question intéressante serait d'étudier le risque théorique de l'estimateur résultant de cette sélection de fenêtre.

Déconvolution d'une densité en base d'Hermite Dans le chapitre 3, nous avons introduit un estimateur de projection en base d'Hermite dans un modèle de bruit additif dans le cas i.i.d. et dépendant. Nous avons considéré dans ce chapitre que la densité du bruit était connue. Cette hypothèse est assez restrictive mais nécessaire pour des raisons d'identifiabilité. Une piste à suivre serait de considérer le cas où la densité du bruit est inconnue mais estimée grâce à un échantillon préliminaire du bruit $(\varepsilon_{-M}, \dots, \varepsilon_{-1})$ voir Neumann (2007). Cela permettrait de se comparer par exemple aux stratégies développées dans Comte and Lacour (2011) et Kappus and Mabon (2014).

La base d'Hermite peut être aussi considérée dans le cas d'estimation d'une fonction de régression avec erreur sur les variables. Le modèle est le suivant :

$$Y_k = X_k + \varepsilon_k, \quad Z_k = b(X_k) + \eta_k, \quad \text{où } \varepsilon_k \text{ indépedant de } \eta_k \ .$$

Dans ce cadre, nous observons (Y_k, Z_k) et nous cherchons à estimer la fonction de régression b. Les propriétés de la base d'Hermite pourraient être exploitées pour proposer un estimateur direct sans pour autant faire intervenir des estimateurs quotients.

Estimation dans un modèle de régression en base d'Hermite Dans le chapitre 4, nous avons construit deux estimateurs pour estimer une fonction de régression dans un modèle généralisant celui de "Laplace convolution" (Model (1.25)). Nous avons obtenu des vitesses de convergence non standard comparées aux vitesses obtenues en déconvolution d'une densité (voir Chapitre 3). Ce serait une motivation pour d'établir des bornes inférieures afin d'étudier l'optimalité des vitesses. Nous avons aussi établi une inégalité oracle pour la première stratégie (stratégie Fourier-Hermite) en utilisant la méthode de Goldenshluter et Lepski. Une implémentation de cette dernière et d'une version dérivée inspirée de Lacour et al. (2017) montrent la bonne performance des deux procédures. Mais elles sont cependant relativement lentes en terme de temps de calcul (environ 5h pour avoir un résultat d'erreur après 100 répétitions). Une piste à creuser serait de considérer des

procédures par pénalisation. En outre, nous avons testé numériquement la procédure HH. Nous remarquons une réduction considérable du temps de calcul (35mins environ pour avoir un résultat d'erreur après 100 répétitions) en adaptant la procédure de Lacour et al. (2017). On peut aussi enquêter sur une procédure par pénalisation.

De plus, nous avons supposé globalement que le noyau g était connu, on peut étudier le cas où il est inconnu, par exemple contaminé par un bruit blanc (voir Vareschi (2015)). Enfin, les données x_k dans Modèle (1.25) sont déterministes dans le chapitre 4, il serait intéressant d'étudier le cas des x_k aléatoires i.i.d. admettant une certaine densité commune. Cette dernière pourra faire intervenir des matrices aléatoires dans l'expression de l'estimateur de régression.

Global à tous les chapitres Dans tous ces chapitres, on considère des fonctions d'une variable, une question naturelle est d'essayer d'étendre les résultats obtenus en dimension supérieure. On pourra par exemple utiliser la méthode de Goldenshluster et Lepski pour le choix des paramètres pertinents.

Première partie

Estimation adaptative et optimale des dérivées d'une densité

Chapitre 2

Optimal adaptive estimation of the derivative of a density on $\mathbb{R}$ or $\mathbb{R}^+$

Ce chapitre est issu de 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.

Résumé. Dans ce chapitre, nous considérons le problème d'estimation de la dérivée d'ordre d notée $f^{(d)}$ d'une densité f à partir des observations $X_1, \ldots, X_n$ i.i.d. de densité f à support dans $\mathbb{R}$ ou $\mathbb{R}^+$. Nous proposons des estimateurs par projection définis dans les bases orthonormales d'Hermite ou de Laguerre et étudions leur $\mathbb{L}^2$-risque intégré. Pour f appartenant à des espaces de régularité et pour un espace de projection choisi avec une dimension adéquate, nous obtenons des vitesses de convergence pour nos estimateurs qui sont optimaux au sens minimax. Le choix optimal de la dimension de l'espace de projection dépend de paramètres inconnus. De ce fait, nous décrivons une procédure adaptative pour choisir la dimension pertinente qui conduit à cette vitesse optimale. Nous discutons les hypothèses et nous comparons l'estimateur introduit à celui obtenu en dérivant simplement l'estimateur de densité. Enfin, des simulations sont réalisées. Elles illustrent les bonnes performances de la procédure et permettent de comparer numériquement les estimateurs par projection et par noyau.

Mots-clés. Estimation des dérivées d'une densité, base d'Hermite, base de Laguerre, sélection de modèles, estimateur par projection.

Abstract. In this chapter, we consider the problem of estimating the d-th order derivative $f^{(d)}$ of a density f, relying on a sample of n i.i.d. observations $X_1, \ldots, X_n$ with density f supported on $\mathbb{R}$ or $\mathbb{R}^+$. We propose projection estimators defined in the orthonormal Hermite or Laguerre bases and study their integrated $\mathbb{L}^2$-risk. For the density f belonging to regularity spaces and for a projection space chosen with adequate dimension, we obtain rates of convergence for our estimators, which are optimal in the minimax sense. The optimal choice of the projection space depends on unknown parameters, so a general data-driven procedure is proposed to reach the bias-variance compromise automatically. We discuss the assumptions and the estimator is compared to the one obtained by simply differentiating the density estimator. Simulations are finally performed. They illustrate

the good performances of the procedure and provide numerical comparison of projection and kernel estimators.

Keywords. Estimation of derivatives of a density, Hermite basis, Laguerre basis, model selection, projection estimator.

Sommaire

2.1 Introduction

2.1.1 Motivations and content

Let $X_1, \ldots, X_n$ be n i.i.d. random variables with common density f with respect to the Lebesgue measure. The problem of estimating f in this simple model has been widely studied. In some contexts, it is also of interest to estimate the d-th order derivative $f^{(d)}$ of f, for different values of the integer d. Density derivatives provide information about the slope of the curves, local extrema or saddle points, for instance. Several examples of use of derivatives are developed in Singh (1977a) and Sasaki et al. (2016). The most common cases are those with $d \in \{1, 2\}$. The first order density derivative permits to reach information, such as mode seeking in mixture models and in data analysis, see *e.g.* Cheng (1995), Chacón and Duong (2013). The second order derivative of the density can be used to estimate one parameter scale of exponential families (see Genovese et al. (2016)), to develop tests for mode (see Cheng (1995)), to select the optimal bandwidth parameter for density estimation (see Silverman (1978)). Let us detail two specific contexts.

1. The question arises when considering regression models. The estimation of the so-called "average derivative" defined by $\delta = \mathbb{E}[Y\psi(X)]$, with $\psi(x) = f^{(1)}(x)/f(x)$, and f is the marginal distribution of X (see Härdle and Stoker (1989) and Härdle et al. (1992)) relies on the estimation of the derivative of the density of X. This quantity enables to quantify the relative impact of X on the variable of interest Y. In an econometric context, the average derivative is also used to verify empirically the law of demand : it allows to compare two economies with different price systems (see Härdle et al. (1991) and Härdle et al. (1992), section 3). In Bercu et al. (2019), the study of sea shore water quality leads the authors to estimate the derivative of the regression function, and the derivative of a Nadaraya-Watson estimator involves the derivative of a density estimator. Regression curves (see Park and Kang (2008)) also involve derivatives of densities, consider $r(x) = \mathbb{E}(Y|X = x)$, Singh (1977a) (see Equation (2.1)) establishes that for specific families of conditional distributions of Y given X, on can express $r(x) = \psi(x)$ as $\psi(x) = f^{(1)}(x)/f(x)$ where f is a density (see (2.1) in Singh (1977a)).

2. Derivatives also appear in the study of diffusion processes. Let $(X_t)_{t\geqslant 0}$ be the solution of
$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = \eta,$$
where W_t is a standard Brownian independent of η. There exists a solution under standard assumptions on b and σ. The model is widely used, for example in finance and biology. One related statistical problem is to estimate the drift function b, from discrete time observations of the process X. Under additional conditions (see Schmisser (2013)), the model is stationary, admits a stationary distribution f and it holds that
$$\frac{f^{(1)}(x)}{f(x)} \propto \frac{2b(x)}{\sigma^2(x)} - 2\frac{\sigma'(x)}{\sigma(x)}.$$
If the variance σ is either a constant or known, estimating f and $f^{(1)}$ lead to an estimator of b.

These examples illustrate the interest of the mathematical question of nonparametric estimation of derivatives as a general inverse problem.

Most proposals for estimating the derivative of a density are built as derivatives of kernel density estimators, see Bhattacharya (1967), Schuster (1969), Silverman (1978), Rao (1996), Chacón et al. (2011), Chacón and Duong (2013), Markovich (2016) or Giné and Nickl (2016), either in independent or in α-mixing settings, in univariate or in multivariate contexts. A slightly different proposal still based on kernels can be found in Singh (1979). The question of bandwidth selection is only considered in the more recent papers. For instance, Chacón and Duong (2013) propose a general cross-validation method in the multivariate case for a matrix bandwidth, see also the references therein. Most recently, Lepski (2018) proposed a general original approach to bandwidth selection, and applies it to derivative estimation in a multivariate $\mathbb{L}^p$ setting and for anisotropic Nikol'ski regularity classes. This paper is, to the best of our knowledge, the first to study the risk of an adaptive kernel estimator.

Projection estimators have also been considered for density and derivatives estimation. More precisely, using trigonometric basis, Efromovich (1998) proposes a complete study of optimality and sharpness of such estimators, on Sobolev periodic spaces. Lately, Giné and Nickl (2016) propose a projection estimator and provide an upper bound for its $\mathbb{L}^p$-risk, $p \in [1, \infty]$. In a dependent context, Schmisser (2013) studies projection estimators in a compactly supported basis constrained on the borders or a non compact multi-resolution basis : she considers dependent β-mixing variables and a model selection method is proposed and proved to reach optimal rates on Besov spaces. In most results, the rate obtained for estimating $f^{(d)}$ the d-th order derivative assumed to belong to a regularity space associated to a regularity α, is of order $n^{-2\alpha/(2\alpha+2d+1)}$. Recently, a bayesian approach has been investigated in Shen and Ghosal (2017) relying on a B spline basis expansion, the procedure requires the knowledge of the regularity of the estimated function.

In the present work, we consider projection estimators on projection spaces generated by Hermite or Laguerre basis, which have non compact supports, $\mathbb{R}$ or $\mathbb{R}^+$. When using compactly supported bases, one has to choose the basis support : it is generally considered as a fixed interval say $[a, b]$, but the bounds a and b are in fact determined from the data. Hermite and Laguerre bases do not require this preliminary choice. Moreover, in a recent work, Belomestny et al. (2019) prove that estimators represented in Hermite basis have a low complexity and that few coefficients are required for a good representation of the functions : therefore, the computation is numerically fast and the estimate is parsimonious. If the X_i's are nonnegative, then one should use the Laguerre basis : thus, this basis is of natural use in survival analysis where most functions under study are $\mathbb{R}^+$-supported. Lastly, we mention that derivatives of Laguerre or Hermite functions have interesting mathematical properties : their derivatives are simple and explicit linear combination of other functions of the bases. This property is fully exploited to construct our estimators.

The integrated $\mathbb{L}^2$-risk of such estimators is classically decomposed into a squared bias and a variance term. The specificity of our context is threefold.

1. The bias term is studied on specific regularity spaces, namely Sobolev Hermite and Sobolev Laguerre spaces, as defined in Bongioanni and Torrea (2009), enabling to consider non compact estimation support $\mathbb{R}$ or $\mathbb{R}^+$.

2. The order of the variance term depends on moment assumptions. This explains why, to perform a data driven selection of the projection space, we propose a random empirical estimator of the variance term, which has automatically the adequate order.
3. In standard settings, the dimension of the projection space is the relevant parameter that needs to be selected to achieve the bias-variance compromise. In our context, this role is played by the square root of the dimension.

We also mention that our procedure provides parsimonious estimators, as few coefficients are required to reconstruct functions accurately. Moreover, our regularity assumptions are naturally set on f and not on its derivatives, contrary to what is done in several papers. Our random penalty proposal is new, and most relevant in a context where the representative parameter of the projection space is not necessarily its dimension, but possibly the square root of the dimension. We compare our estimators with those defined as derivatives of projection density estimators, which is the strategy usually applied with kernel methods. Finally, we also propose a numerical comparison between our projection procedure and a sophisticated kernel method inspired by the recent proposal in density estimation of Lacour et al. (2017).

The chapter is organized as follows. In the remaining of this section, we define the Hermite and Laguerre bases and associated projection spaces. In Section 2.2, we define the estimators and establish general risk bounds, from which rates of convergence are obtained, and lower bounds in the minimax sense are proved. A model selection procedure is proposed, relying on a general variance estimate ; it leads to a data-driven bias-variance compromise. Further questions are studied in Section 2.3 : the comparison with the derivatives of the density estimator leads in our setting to different developments depending on the considered basis : interestingly Hermite and Laguerre cases happen to behave differently from this point of view. Lastly, a simulation study is conducted in Section 2.4, in which kernel and projection strategies are compared.

2.1.2 Notations and definition of the basis

The following notations are used in the remaining of this paper. For a, b two real numbers, denote $a \vee b = \max(a, b)$ and $a_+ = \max(0, a)$. For u and v two functions in $\mathbb{L}^2(\mathbb{R})$, denote $\langle u, v\rangle = \int_{-\infty}^{+\infty} u(x)v(x)dx$ the scalar product on $\mathbb{L}^2(\mathbb{R})$ and $\|u\| = \left(\int_{-\infty}^{+\infty} u(x)^2 dx\right)^{1/2}$ the norm on $\mathbb{L}^2(\mathbb{R})$. Note that these definitions remain consistent if u and v are in $\mathbb{L}^2(\mathbb{R}^+)$.

The Laguerre basis.

Define the Laguerre basis by :

$$\ell_j(x) = \sqrt{2}L_j(2x)e^{-x}, \quad L_j(x) = \sum_{k=0}^{j} \binom{j}{k}(-1)^k \frac{x^k}{k!}, \quad x \geqslant 0, \quad j \geqslant 0, \tag{2.1}$$

where L_j is the Laguerre polynomial of degree j. It satisfies : $\int_0^{+\infty} L_k(x)L_j(x)e^{-x}dx = \delta_{k,j}$ (see Abramowitz and Stegun (1964), 22.2.13), where $\delta_{k,j}$ is the Kronecher symbol. The

family $(\ell_j)_{j\geqslant 0}$ is an orthonormal basis on $\mathbb{L}^2(\mathbb{R}^+)$ such that $\|\ell_j\|_\infty = \sup_{x\in\mathbb{R}^+}|\ell_j(x)| = \sqrt{2}$. The derivative of ℓ_j satisfies a recursive formula (see Lemma 8.1 in Comte and Genon-Catalot (2018)) that plays an important role in the sequel :

$$\ell_0' = -\ell_0, \qquad \ell_j' = -\ell_j - 2\sum_{k=0}^{j-1}\ell_k, \quad \forall j \geqslant 1. \tag{2.2}$$

The Hermite basis.

Define the Hermite basis $(h_j)_{j\geqslant 0}$ from Hermite polynomials $(H_j)_{j\geqslant 0}$:

$$h_j(x) = c_j H_j(x)e^{-x^2/2}, \quad H_j(x) = (-1)^j e^{x^2}\frac{d^j}{dx^j}(e^{-x^2}), \quad c_j = (2^j j!\sqrt{\pi})^{-1/2}, \quad x\in\mathbb{R},\, j\geqslant 0. \tag{2.3}$$

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

$$\|h_j\|_\infty = \sup_{x\in\mathbb{R}}|h_j(x)| \leqslant \phi_0, \text{ with } \phi_0 = \pi^{-1/4}, \tag{2.4}$$

(see Abramowitz and Stegun (1964), chap.22.14.17 and Indritz (1961)). The derivatives of h_j also satisfy a recursive formula (see Comte and Genon-Catalot (2018), Equation (52) in Section 8.2),

$$h_0' = -h_1/\sqrt{2}, \qquad h_j' = (\sqrt{j}\; h_{j-1} - \sqrt{j+1}h_{j+1})/\sqrt{2}, \quad \forall j \geqslant 1. \tag{2.5}$$

In the sequel, we denote by φ_j either for h_j in the Hermite case or for ℓ_j in the Laguerre case. Let $g\in\mathbb{L}^2(\mathbb{R})$ or $g\in\mathbb{L}^2(\mathbb{R}^+)$, g develops either in the Hermite basis or the Laguerre basis :

$$g = \sum_{j\geqslant 0} a_j(g)\varphi_j, \quad a_j(g) = \langle g, \varphi_j\rangle.$$

Define, for an integer $m\geqslant 1$, the space

$$S_m = \text{Span}\{\varphi_0, \ldots, \varphi_{m-1}\}.$$

The orthogonal projection of g on S_m is given by : $g_m = \sum_{j=0}^{m-1} a_j(g)\varphi_j$.

2.2 Estimation of the derivatives

2.2.1 Assumptions and projection estimator of $f^{(d)}$.

Let $X_1, \ldots, X_n$ be n i.i.d. random variables with common density f with respect to the Lebesgue measure and consider the following assumptions. Let d be an integer, $d\geqslant 1$.

(**A**1) The density f is d-times differentiable and $f^{(d)}$ belongs to $\mathbb{L}^2(\mathbb{R}^+)$ in the Laguerre case or $\mathbb{L}^2(\mathbb{R})$ in the Hermite case.

(**A**2) For all integer r, $0 \leqslant r \leqslant d-1$, we have $\|f^{(r)}\|_\infty < +\infty$.

(**A**3) For all integer r, $0 \leqslant r \leqslant d-1$, it holds $\lim_{x\to 0} f^{(r)}(x) = 0$.

Assumption (**A**3) is specific to the Laguerre case and avoids boundary issue. In particular, it permits to establish Lemma 2.2.1 below that is central to define our estimator. This assumption can be removed at the expense of additional technicalities, see Section 2.3.

Under (**A**1), we develop $f^{(d)}$ in the Laguerre or Hermite basis, its orthogonal projection on S_m, $m \geqslant 1$, is

$$f_m^{(d)} = \sum_{j=0}^{m-1} a_j(f^{(d)})\varphi_j, \text{ where, } \quad a_j(f^{(d)}) = \langle f^{(d)}, \varphi_j \rangle. \tag{2.6}$$

The estimator is built by using the following result, proved in Appendix 2.6.8.

Lemma 2.2.1. *Suppose that* (**A**1) *and* (**A**2) *hold in the Hermite case and that* (**A**1), (**A**2) *and* (**A**3) *hold in the Laguerre case. Then* $a_j(f^{(d)}) = (-1)^d \mathbb{E}[\varphi_j^{(d)}(X_1)]$, $\forall j \geqslant 0$.

Remark 2.1. *If the support of the density f is a strict compact subset $[a,b]$ of the estimation support (here $\mathbb{R}$ and $a < b$ or $\mathbb{R}^+$ and $0 < a < b$), then the regularity condition* (**A**1) *implies that f must be null in a, b, as well as its derivatives up to order $d-1$(i.e. $f(x_0) = f^{(1)}(x_0) = \cdots = f^{(d-1)}(x_0) = 0$ for $x_0 \in \{a, b\}$). On the contrary, Assumption* (**A**3) *in the Laguerre case can be dropped out (see Section 2.3) and this shows that a specific problem occurs when the density support coincides with the estimation interval. This point presents a real difficulty and is either not discussed in the literature, or hidden by periodicity conditions.*

We derive the following estimator of $f^{(d)}$ (see also Giné and Nickl (2016) p.402) : let $m \geqslant 1$,

$$\widehat{f}_{m,(d)} = \sum_{j=0}^{m-1} \widehat{a}_j^{(d)} \varphi_j, \quad \text{with} \quad \widehat{a}_j^{(d)} = \frac{(-1)^d}{n} \sum_{i=1}^{n} \varphi_j^{(d)}(X_i). \tag{2.7}$$

For $d = 0$, we recover an estimator of the density f. Usually for kernel procedure, we obtain an estimator of $f^{(d)}$ by differentiating the kernel density estimator of f see Rao (1996), Chacón et al. (2011), Chacón and Duong (2013) or Giné and Nickl (2016). With projection estimators, it is not well adapted as the derivative of an orthonormal basis is not a orthonormal basis. This is why our strategy here is different. Moreover, a comparison with the derivative of the density estimator in our context is provided in the sequel (see Section 2.3)

2.2.2 Risk bound and rate of convergence.

We consider the $\mathbb{L}^2$-risk of $\widehat{f}_{m,(d)}$, defined in (2.7),

$$\mathbb{E}\big[\|\widehat{f}_{m,(d)} - f^{(d)}\|^2\big] = \|f_m^{(d)} - f^{(d)}\|^2 + \mathbb{E}\big[\|\widehat{f}_{m,(d)} - f_m^{(d)}\|^2\big], \tag{2.8}$$

where $f_m^{(d)} := \sum_{k=0}^{m-1} a_j(f^{(d)})\varphi_j$. The study of the second right-hand-side term of the equality (variance term) leads to the following result.

Theorem 2.2.1. *Suppose that* (**A**1) *and* (**A**2) *hold in the Hermite case and that* (**A**1), (**A**2) *and* (**A**3) *hold in the Laguerre case. Assume that*

$$\mathbb{E}[X_1^{-d-1/2}] < +\infty \text{ in the Laguerre case and } \mathbb{E}[|X_1|^{2/3}] < +\infty \text{ in the Hermite case.} \tag{2.9}$$

Then, for sufficiently large $m \geqslant d$, it holds that

$$\mathbb{E}\big[\|\widehat{f}_{m,(d)} - f^{(d)}\|^2\big] \leqslant \|f_m^{(d)} - f^{(d)}\|^2 + C\frac{m^{d+\frac{1}{2}}}{n} - \frac{\|f_m^{(d)}\|^2}{n}, \tag{2.10}$$

for a positive constant C depending on the moments in condition (2.9) (but not on m nor n).

Remark 2.2. *In the Laguerre case, condition* (2.9) *is a consequence of* (**A**3) *and $f^{(d)}(0) < +\infty$. Indeed,* (**A1**) *imposes that $f(x) \underset{x\to 0}{\sim} x^d f^{(d)}(x)$ which, under $f^{(d)}(0) < +\infty$, ensures integrability of $x^{-d-1/2}f(x)$ around 0^+ (i.e. $\int_0 x^{-d-1/2}f(x)dx < \infty$); integrability near ∞ is a consequence of $f \in \mathbb{L}^1([0,\infty))$.*

The bound obtained for $\widehat{f}_{m,(d)}$ in Theorem 2.2.1 is sharp. Indeed, we can establish the following lower bound.

Proposition 2.2.1. *Under the Assumptions of Theorem 2.2.1, it holds, for some constant $c > 0$, that*

$$\mathbb{E}\Big[\|\widehat{f}_{m,(d)} - f^{(d)}\|^2\Big] \geqslant \|f_m^{(d)} - f^{(d)}\|^2 + c\frac{m^{d+\frac{1}{2}}}{n} - \frac{\|f_m^{(d)}\|^2}{n}.$$

The next step is to compute the rate of convergence.

Definition of regularity classes and rate of convergence

The first two terms in the right hand side of (2.10) have an antagonistic behavior with respect to m : the first term, $\|f_m^{(d)} - f^{(d)}\|^2$ is a squared bias term which decreases when m increases, while the second $m^{d+1/2}/n$ is a variance term which increases with m. Thus, the optimal choice of m requires a bias-variance compromise which allows to derive the rate of convergence of $\widehat{f}_{m,(d)}$. To evaluate the order of the bias term, we introduce Sobolev-Hermite and Sobolev-Laguerre regularity classes for f (see Bongioanni and Torrea (2009) and Comte and Genon-Catalot (2018)).

Sobolev-Hermite classes

Let $s > 0$ and $D > 0$, define the Sobolev-Hermite ball of regularity s

$$W_H^s(D) = \{\theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k\geqslant 0} k^s a_k^2(\theta) \leqslant D\}, \tag{2.11}$$

where $a_k^2(\theta) = \langle \theta, h_k\rangle$ and k^s is to be understood as $(\sqrt{k})^{2s}$, see Remark 2.3 below. The following Lemma relates the regularity of $f^{(d)}$ and the one of f.

Lemma 2.2.2. *Let $s \geqslant d$ and $D > 0$, assume that f belongs to $W_H^s(D)$ and* (**A2**)*, then there exist a constant $D_d > D$ such that $f^{(d)}$ is in $W_H^{s-d}(D_d)$.*

Sobolev-Laguerre classes

Similarly, consider the Sobolev-Laguerre ball of regularity s

$$W_L^s(D) = \{\theta \in \mathbb{L}^2(\mathbb{R}^+), |\theta|_s^2 = \sum_{k\geqslant 0} k^s a_k^2(\theta) \leqslant D\}, \quad D > 0, \tag{2.12}$$

where $a_k(\theta) = \langle \theta, \ell_k \rangle$. If $s \geqslant 1$ an integer, there is an equivalent norm of $|\theta|_s^2$ (see Section 7.2 of Belomestny et al. (2016)) defined by

$$|||\theta|||_s^2 = \sum_{j=0}^{s} \|\theta\|_j^2, \quad \|\theta\|_j^2 = \|x^{j/2} \sum_{k=0}^{j} \binom{j}{k} \theta^{(k)}\|^2. \tag{2.13}$$

This inspires the definition, for $s \in \mathbb{N}$ and $D > 0$, of the subset $\widetilde{W}_L^s(D)$ as

$$\widetilde{W}_L^s(D) = \{\theta \in \mathbb{L}^2(\mathbb{R}^+),\, \theta^{(j)} \in C([0,\infty)),\, x \mapsto x^{k/2}\theta^{(j)}(x) \in \mathbb{L}^2(\mathbb{R}^+),\, 0 \leqslant j \leqslant k \leqslant s, |\theta|_s^2 \leqslant D\}. \tag{2.14}$$

It is straightforward to see that $\widetilde{W}_L^s(D) \subset W_L^s(D)$. Moreover, we can relate the regularity of $f^{(d)}$ and the one of f.

Lemma 2.2.3. *Let $s \in \mathbb{N}$, $s \geqslant d \geqslant 1$, $D > 0$ and $\theta \in \widetilde{W}_L^s(D)$, then, $\theta^{(d)} \in \widetilde{W}_L^{s-d}(D_d)$ where $D \leqslant D_d < \infty$.*

Rate of convergence of $\widehat{f}_{m,(d)}$

Assume that $f \in W_H^s(D)$ or $f \in \widetilde{W}_L^s(D)$, then Lemmas 2.2.2 and 2.2.3 enable a control of the bias term in (2.10)

$$\|f_m^{(d)} - f^{(d)}\|^2 = \sum_{j\geqslant m} (a_j(f^{(d)}))^2 = \sum_{j\geqslant m} j^{s-d}(a_j(f^{(d)}))^2 j^{-(s-d)} \leqslant D_d m^{-(s-d)}.$$

Injecting this in (2.10) yields

$$\mathbb{E}\big[\|\widehat{f}_{m,(d)} - f^{(d)}\|^2\big] \leqslant D' m^{-(s-d)} + c\frac{m^{d+\frac{1}{2}}}{n}.$$

Remark 2.3. *We stress that the squared bias and variance terms have orders specific to the use of Laguerre or Hermite bases. For instance if $d = 0$, the latter bound becomes $m^{-s} + c\sqrt{m}/n$ showing that the associated spaces are represented by the square root of their dimension and not their dimension. Analogously in the context of derivatives, the role of the dimension in Schmisser (2013) is played in our case by $\sqrt{m}$.*

Consequently, selecting $m_{opt} = [n^{2/(2s+1)}]$ gives the rate of convergence

$$\mathbb{E}\big[\|\widehat{f}_{m_{opt},(d)} - f^{(d)}\|^2\big] \leqslant C(s,d,D) n^{-\frac{2(s-d)}{2s+1}}, \tag{2.15}$$

where $C(s,d,D)$ depends only on s, d and D, not on m. This rate coincides with the one obtained by Schmisser (2013) in the dependent case and by **?**. Contrary to Rao (1996) and Lepski (2018), we set the regularity conditions on the function f and not on its derivatives : for a regularity s of $f^{(d)}$, they obtain a quadratic risk $n^{-2(s-d)/(2s+1)}$ (case $p=2$ in Lepski (2018) and dimension 1). Interestingly, m_{opt} does not depend on d. This is in accordance with Lepski (2018)'s strategy, which consists in plugging in the derivative kernel estimator the bandwidth selected for the direct density estimation problem. Note that, for $d=0$ in (2.15), we recover the optimal rate for estimation of the density f.

Remark 2.4. *If f is a mixture of Gaussian densities in the Hermite case or a mixture of Gamma densities in the Laguerre case, it is known from Section 3.2 in Comte and Genon-Catalot (2018) that the bias decreases with exponential rate. The computations therein can be extended to the present setting and imply in both Hermite and Laguerre cases that m_{opt} is then proportional to $\log(n)$. Therefore the risk has order $[\log(n)]^{d+\frac{1}{2}}/n$: for these collections of densities, the estimator converges much faster than in the general setting.*

2.2.3 Lower bound

Contrary to the lower bound given in Proposition 2.2.1, which ensures that the upper bound derived in Theorem 2.2.1 for the specific estimator $\widehat{f}_{m,(d)}$ is sharp, we provide a general lower bound that guarantees that the rate of the estimator $\widehat{f}_{m,(d)}$ is minimax optimal. The following assertion states that the rate obtained in (2.15) is the optimal rate.

Let $s \geqslant d$ be an integer and $\widetilde{f}_{n,d}$ be any estimator of $f^{(d)}$. Then for n large enough, we have

$$\inf_{\widetilde{f}_{n,d}} \sup_{f \in W^s(D)} \mathbb{E}[\|\widetilde{f}_{n,d} - f^{(d)}\|^2] \geqslant c n^{-\frac{2(s-d)}{2s+1}}, \tag{2.16}$$

where the infimum is taken over all estimator of $f^{(d)}$, c a positive constant depending on s and d, and $W^s(D)$ stands either for $W_L^s(D)$ or for $W_H^s(D)$.

We provide in Section 2.6.3 the key elements to establish (2.16). We emphasize that the proof relies on compactly supported test functions, implying that the lower bound on usual Sobolev spaces and the present one coincide, as these functions belong to both. This had to be checked since Hermite Sobolev spaces are strict subspaces of usual Sobolev spaces. Similar lower bounds were known for this model for different regularity spaces. We mention e.g. (7.3.3) in Efromovich (1999), which considers perdiodic Lispchitz spaces, or Lepski (2018), which examines general Nikol'ski spaces.

2.2.4 Adaptive estimator of $f^{(d)}$.

The choice of $m_{opt} = [n^{2/(2s+1)}]$ leading to the optimal rate of convergence is not feasible in practice. In this section we provide an automatic choice of the dimension m, from the observations $(X_1, \ldots, X_n)$, that realizes the bias-variance compromise in (2.10). Assume that m belongs to a finite model collection $\mathcal{M}_{n,d}$, we look for m that minimizes the bias-

variance decomposition (2.8) rewritten as

$$\mathbb{E}\big[\|\widehat{f}_{m,(d)} - f^{(d)}\|^2\big] = \|f_m^{(d)} - f^{(d)}\|^2 + \frac{1}{n}\sum_{j=0}^{m-1} \mathrm{Var}\Big[\varphi_j^{(d)}(X_1)\Big].$$

Note that the bias is such that $\|f_m^{(d)} - f^{(d)}\|^2 = \|f^{(d)}\|^2 - \|f_m^{(d)}\|^2$ where $\|f^{(d)}\|^2$ is independent of m and can be dropped out. The remaining quantity $-\|f_m^{(d)}\|^2$ is estimated by $-\|\widehat{f}_{m,(d)}\|^2$. The variance term is replaced by an estimator of a sharp upper bound, given by

$$\widehat{V}_{m,d} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{m-1} (\varphi_j^{(d)}(X_i))^2. \tag{2.17}$$

Finally, we set

$$\widehat{m}_n := \underset{m\in\mathcal{M}_{n,d}}{\mathrm{argmin}}\{-\|\widehat{f}_{m,(d)}\|^2 + \widehat{\mathrm{pen}}_d(m)\}, \quad \text{where} \quad \widehat{\mathrm{pen}}_d(m) = \kappa\frac{\widehat{V}_{m,d}}{n}, \tag{2.18}$$

where κ is a positive numerical constant. If we set $V_{m,d} := \sum_{j=0}^{m-1}\mathbb{E}[(\varphi_j^{(d)}(X_1)^2]$, it holds $\mathbb{E}[\widehat{\mathrm{pen}}_d(m)] = \kappa V_{m,d}/n$. In the sequel, we write $\mathrm{pen}_d(m) := \kappa V_{m,d}/n$. To implement the procedure a value for κ has to be set. Theorem 2.2.2 below provides a theoretical lower bound for κ, which is however generally too large. In practice this constant is calibrated by intensive preliminary experiments, see Section 2.4. General calibration methods can be found in Baudry et al. (2012) for theoretical explanations and heuristics, and in the associated package, for practical implementation.

Remark 2.5. *Note that in the definition of the penalty, instead of* (2.18)*, we can plug the deterministic upper bound on the variance and take* $c\,m^{d+\frac{1}{2}}/n$ *as a penalty (see Theorem 2.2.1) as Proposition 2.2.1 ensures its sharpness. However, this upper bound relies on additional assumptions given in* (2.9) *and depends on non explicit constants (see Askey and Wainger (1965)). This is why we choose to estimate directly the variance by* $\widehat{V}_{m,n}$ *and use* $\widehat{V}_{m,n}/n$ *as the penalty term.*

Theorem 2.2.2. *Let* $\mathcal{M}_{n,d} := \{d, \dots, m_n(d)\}$*, where* $m_n(d) \geqslant d$*. Assume that* (**A**1) *and* (**A**2) *hold, and that* (**A**3) *holds in the Laguerre case, and that* $\|f\|_\infty < +\infty$*.*

AL. Set $m_n(d) = \lfloor (n/\log^3(n))^{\frac{2}{2d+1}} \rfloor$*, assume that* $\sup_{x\in\mathbb{R}^+}\frac{f(x)}{x^d} < +\infty$ *in the Laguerre case,*

AH. Set $m_n(d) = \lfloor n^{\frac{2}{2d+1}} \rfloor$ *in the Hermite case.*

Then, for any $\kappa \geqslant \kappa_0 := 32$ *it holds that*

$$\mathbb{E}\Big[\|\widehat{f}_{\widehat{m}_n,(d)} - f^{(d)}\|^2\Big] \leqslant C \inf_{m\in\mathcal{M}_{n,d}}\Big(\|f_m^{(d)} - f^{(d)}\|^2 + \mathrm{pen}_d(m)\Big) + \frac{C'}{n}, \tag{2.19}$$

where C *is a universal constant (*$C = 3$ *suits) and* C' *is a constant depending on* $\sup_{x\in\mathbb{R}^+}\frac{f(x)}{x^d} < +\infty$ *and* $\mathbb{E}[X_1^{-d-1/2}] < +\infty$ *(Laguerre case) or* $\|f\|_\infty$ *(Hermite case).*

The constraint on the the largest element $m_n(d)$ of the collection $\mathcal{M}_{n,d}$ ensures that the variance term, which is upper bounded by $m^{d+\frac{1}{2}}/n$ vanishes asymptotically. The additional log term does not influence the rate of the optimal estimator : the optimal (and unknown) dimension $m_{opt} \asymp n^{\frac{2}{2s+1}}$, with s the regularity index of f, is such that $m_{opt} \ll n^{\frac{2}{2d+1}}$ as soon as $s > d$. For $s = d$, a log-loss in the rate would occur in the Laguerre case, but not in the Hermite case.

Note that, in the Laguerre case, condition $\sup_{x\in\mathbb{R}^+} \frac{f(x)}{x^d} < +\infty$ implies $\mathbb{E}(X_1^{-d-1/2}) < +\infty$ (see condition 2.9)) and is clearly related to (**A**3). Inequality (2.19) is a key result and expresses that $\widehat{f}_{\widehat{m}_n,(d)}$ realizes automatically a bias-variance compromise and is performing as well as the best model in the collection, up to the multiplicative constant C, since clearly, the last term C'/n is negligible. Thus, for f in $\widetilde{W}_L^s(D)$ or $W_H^s(D)$ and under the assumptions of Theorem 2.2.2, we have $\mathbb{E}\big[\|\widehat{f}_{\widehat{m},(d)} - f^{(d)}\|^2\big] = \mathcal{O}(n^{-2(s-d)/(2s+1)})$, which implies that the estimator is adaptive.

2.3 Further questions

We investigate here additional questions, and set for simplicity $d = 1$. Mainly, we compare our estimator to the derivative of a density estimator, and discuss condition (**A**3) in the Laguerre case.

2.3.1 Derivatives of the density estimator

When using kernel strategies, it is classical to build an estimator of the derivative of f by differentiating the kernel density estimator, as already mentioned in the Introduction. For projection estimators, we find more relevant to proceed differently. Indeed, our aim is to obtain an estimator expressed in an orthonormal basis ; unfortunately, the derivative of an orthonormal basis is a collection of functions but not an orthonormal basis. So, our proposal (2.7) is easier to handle. Moreover, our estimator can be seen as a contrast minimizer, which makes model selection possible to settle up.

However, Laguerre and Hermite cases are somehow different and can be more precisely compared. Let us recall that the projetion estimator of f on S_m is defined by (see Comte and Genon-Catalot (2018), or (2.7) for $d = 0$) :

$$\widehat{f}_m := \sum_{k=0}^{m-1} \widehat{a}_k^{(0)} \varphi_k, \quad \text{where} \quad \widehat{a}_k^{(0)} := \frac{1}{n}\sum_{j=0}^{n} \varphi_k(X_j).$$

As the functions $(\varphi_j)_j$ are infinitely differentiable, both in Hermite and Laguerre settings, this leads to the natural estimator of $f^{(d)}$, $d \geqslant 1$,

$$(\widehat{f}_m)^{(d)} = \sum_{k=0}^{m-1} \widehat{a}_k^{(0)} \varphi_k^{(d)}. \tag{2.20}$$

For $d = 1$, we write $(\widehat{f}_m)^{(1)} = (\widehat{f}_m)'$. We want to compare $(\widehat{f}_m)'$ to $\widehat{f}_{m,(1)}$. In both Hermite and Laguerre cases, this estimator is consistent, under adequate regularity assumptions and for adequate choice of m as a function of n.

2.3.2 Comparison of $\widehat{f}_{m,(1)}$ with $(\widehat{f}_m)'$ in the Hermite case.

Using the recursive formula (2.5), in (2.20) and (2.7) respectively, straightforward computations give

$$(\widehat{f}_m)' = \frac{1}{\sqrt{2}}\widehat{a}_1^{(0)}h_0 + \sum_{j=1}^{m-1}\left(\sqrt{\frac{j+1}{2}}\widehat{a}_{j+1}^{(0)} - \sqrt{\frac{j}{2}}\widehat{a}_{j-1}^{(0)}\right)h_j - \sqrt{\frac{m}{2}}\left(\widehat{a}_m^{(0)}h_{m-1} + \widehat{a}_{m-1}^{(0)}h_m\right),$$

whereas
$$\widehat{f}_{m,(1)} = \frac{1}{\sqrt{2}}\widehat{a}_1^{(0)}h_0 + \sum_{j=1}^{m-1}\left(\sqrt{\frac{j+1}{2}}\widehat{a}_{j+1}^{(0)} - \sqrt{\frac{j}{2}}\widehat{a}_{j-1}^{(0)}\right)h_j.$$

Therefore, it holds that $\mathbb{E}[\|(\widehat{f}_m)' - \widehat{f}_{m,(1)}\|^2] = m/2\{\,\mathbb{E}\left[(\widehat{a}_m^{(0)})^2\right] + \mathbb{E}\left[(\widehat{a}_{m-1}^{(0)})^2\right]\}$ and

$$\mathbb{E}[\|(\widehat{f}_m)' - \widehat{f}_{m,(1)}\|^2] \leqslant \frac{m}{2}(a_{m-1}^2(f) + a_m^2(f)) + \frac{m}{2n}\left(\int h_m^2(x)f(x)dx + \int h_{m-1}^2(x)f(x)dx\right).$$

Using Lemma 8.5 in Comte and Genon-Catalot (2018) under $\mathbb{E}[|X_1|^{2/3}] < +\infty$ and for f in $W_H^s(D)$, $s > 1$, it follows for some positive constant C that,

$$\mathbb{E}[\|(\widehat{f}_m)' - \widehat{f}_{m,(1)}\|^2] \leqslant \frac{D}{2}m^{-s+1} + C\frac{\sqrt{m}}{n}.$$

Under the same assumptions, (2.10) for $d = 1$ implies

$$\mathbb{E}[\|(\widehat{f}_m)' - f'\|^2] \leqslant D'm^{-s+1} + c\frac{m^{3/2}}{n}.$$

Therefore, by triangle inequality, this implies that $(\widehat{f}_m)'$ reaches the same (optimal) rate as $\widehat{f}_{m,(1)}$, under the same assumptions.

2.3.3 Comparison of $\widehat{f}_{m,(1)}$ with $(\widehat{f}_m)'$ in the Laguerre case.

In the Laguerre case, assumption (**A**3) is required for the estimator $\widehat{f}_{m,(1)}$ to be consistent, while it is not for the estimator $(\widehat{f}_m)'$.

Proceeding as previously and taking advantage of the recursive formula (2.2) in (2.20) and (2.7) respectively, straightforward computations give, for $m \geqslant 1$,

$$(\widehat{f}_m)' = \sum_{j=0}^{m-1}\left(\widehat{a}_j^{(0)} - 2\sum_{k=j}^{m-1}\widehat{a}_k^{(0)}\right)\ell_j, \quad \text{whereas} \quad \widehat{f}_{m,(1)} = \sum_{j=0}^{m-1}\left(\widehat{a}_j^{(0)} + 2\sum_{k=0}^{j-1}\widehat{a}_k^{(0)}\right)\ell_j. \tag{2.21}$$

Therefore, in the Laguerre case, the coefficients of $\widehat{f}_{m,(1)}$ in the basis $(\ell_j)_j$ do not depend on m while those of $(\widehat{f}_m)'$ do. Moreover, computing the difference between the estimators leads to $\widehat{f}_{m,(1)} - (\widehat{f}_m)' = 2\sum_{j=0}^{m-1}(\sum_{k=0}^{m-1}\widehat{a}_k^{(0)})\ell_j$ and

$$\|\widehat{f}_{m,(1)} - (\widehat{f}_m)'\|^2 = 4m\left(\sum_{k=0}^{m-1}\widehat{a}_k^{(0)}\right)^2.$$

Heuristically, if $f(0) = 0$, as $f(0) = \sqrt{2}\sum_{j\geqslant 0} a_j(f) = 0$, it follows that $\sum_{j=0}^{m-1} a_j(f)$ should be small for m large enough. Consequently, its consistent estimator $\sum_{k=0}^{m-1} \widehat{a}_k^{(0)}$ should also be small. This would imply that, when $f(0) = 0$, the distance $\|\widehat{f}_{m,(1)} - (\widehat{f}_m)'\|^2$ can be small; on the contrary, the distance should tend to infinity with m if $f(0) \neq 0$. This is due to the fact that $\widehat{f}_{m,(1)}$ is not consistent, while $(\widehat{f}_m)'$ is. Indeed, in the general case ($f(0) \neq 0$), the risk bound we obtain for $(\widehat{f}_m)'$ is the following.

Proposition 2.3.1. *Assume that* (**A**1) *and* (**A**2) *hold for* $d = 1$ *and that* f *belongs to* $W_L^s(D)$. *Then, it holds*

$$\mathbb{E}\|(\widehat{f}_m)' - f'\|^2 \leqslant Cm^{-s+2} + \frac{3}{n}\|f\|_\infty m^2. \tag{2.22}$$

Obviously, for suitably chosen m the estimator is consistent and by selecting $m_{\text{opt}} \asymp n^{1/s}$, it reaches the rate : $\mathbb{E}[\|(\widehat{f}_{m_{\text{opt}}})' - f'\|^2] \leqslant C(s, D)n^{-(s-2)/s}$. This rate is worse than the one obtained for $\widehat{f}_{m,(1)}$ but it is valid without (**A**3), and thus $\widehat{f}_{m,(1)}$ is consistent to estimate an exponential density, or any mixture involving exponential densities. Note that both the order of the bias and the variance in (2.22) are deteriorated compared to (2.10), and we believe these orders are sharp.

In the following section, we investigate if the rate can be improved, if (**A**3) is not satisfied, by correcting our estimator (2.6).

2.3.4 Estimation of f' on $\mathbb{R}^+$ with $f(0) > 0$

Assumption (**A1**) excludes some classical distribution such as the exponential distribution or Beta distributions $\beta(a, b)$ with $a = 1$. If $f(0) > 0$, Lemma 2.2.1 no longer holds, and one has $a_j(f') = -f(0)\ell_j(0) - \mathbb{E}[\ell_j'(X_1)]$ instead. Therefore, $f(0)$ has to be estimated and we consider

$$\widehat{a}_{j,K}^{(1)} = -\ell_j(0)\widehat{f}_K(0) - \frac{1}{n}\sum_{i=1}^n \ell_j'(X_i), \text{ with } \widehat{f}_K = \sum_{j=0}^{K-1} \widehat{a}_j^{(0)}\ell_j,\ \widehat{a}_j^{(0)} = \frac{1}{n}\sum_{i=1}^n \ell_j(X_i). \tag{2.23}$$

We estimate f' as follows

$$\widetilde{f}'_{m,K} = \sum_{j=0}^{m-1} \widehat{a}_{j,K}^{(1)}\ell_j, \text{ with } \widehat{a}_{j,K}^{(1)} = -\frac{1}{n}\sum_{i=1}^n \ell_j'(X_i) - \widehat{f}_K(0)\ell_j(0). \tag{2.24}$$

Obviously, $\widehat{a}_{j,K}^{(1)}$ is a biased estimator of $a_j(f')$, implying that $\widetilde{f}'_{m,K}$ is a biased estimator of f'_m. Now there are two dimensions m and K to be optimized. We can establish the following upper bound.

Proposition 2.3.2. *Suppose* (**A**1) *is satisfied for* $d = 1$, *then it holds that*

$$\mathbb{E}\big[\|\widetilde{f}'_{m,K} - f'\|^2\big] \leqslant \|f' - f'_m\|^2 + \frac{2}{n}\sum_{j=0}^{m-1} \mathbb{E}\left[\left(\ell_j'(X_1)\right)^2\right] + 4m(\mathrm{Var}(\widehat{f}_K(0)) + (f(0) - f_K(0))^2), \tag{2.25}$$

where f_K *is the orthogonal projection of* f *on* S_K *defined by :* $f_K = \sum_{j=0}^{K-1} a_j(f)\ell_j$.

The first two terms of the upper bound seem similar to the ones obtained under (**A**3), but as we no longer assume $f(0) = 0$, Assumption (2.9) for $d = 1$ cannot hold and the tools used to bound the variance term $V_{m,1}$ by $m^{3/2}$ no longer apply : we only get an order m^2 for this term, under $\|f\|_\infty < +\infty$.

The last two terms of (2.25) correspond to m times the pointwise risk of $\widehat{f}_K(0)$. Then, using $\|\ell_j\|_\infty \leqslant \sqrt{2}$, we obtain $\mathrm{Var}(\widehat{f}_K(x)) \leqslant 4K^2/n$. If $\|f\|_\infty < \infty$, this can be improved in $\mathrm{Var}(\widehat{f}_K(x)) \leqslant \|f\|_\infty K/n$, using the orthonormality of $(\ell_j)_j$.

To sum up, if $f \in \widetilde{W}_L^s(D)$, and $\|f\|_\infty < \infty$, then

$$\mathbb{E}\big[\|\widetilde{f}'_{m,K} - f'\|^2\big] \leqslant C(s, D, \|f\|_\infty) \left\{ m^{-s+2} + \frac{m^2}{n} + m\left(K^{-s+1} + \frac{K}{n}\right)\right\}.$$

Choosing $K_{\mathrm{opt}} = cn^{1/s}$ and $m_{\mathrm{opt}} = cn^{1/s}$ gives the rate $\mathbb{E}\big[\|\widetilde{f}'_{m_{\mathrm{opt}},K_{\mathrm{opt}}} - f'\|^2\big] \leqslant Cn^{-(s-2)/s}$, that is the same rate as the one obtained for $(\widehat{f}_{m_{\mathrm{opt}}})'$. Then, renouncing to Assumption (**A**3) has a cost, it renders the procedure burdensome and leads to slower rates.

We propose a model selection procedure adapted to this new estimator. Let

$$\widehat{f}'_{m,K} = \arg\min_{t \in S_m} \gamma_n(t) \tag{2.26}$$

where $\gamma_n(t) = \|t\|^2 + \frac{2}{n}\sum_{i=1}^n t'(X_i) + 2t(0)\widehat{f}_K(0)$. Here, we consider that $K = K_n$ is chosen so that $\widehat{f}_{K_n}$ satisfies

$$\Big[\mathbb{E}(\widehat{f}_{K_n}(0)) - f(0)\Big]^2 \leqslant \frac{K_n \log(n)}{n}. \tag{2.27}$$

This assumption is likely to be fulfilled for a K selected in order to provide a squared bias/variance compromise, see the pointwise adaptive procedure for density estimation in Plancade (2009) ; however therein, the choice of K is random while we set K_n as fixed, here. Then, we select m as follows :

$$\widehat{m}_K = \arg\min_{m \in \mathcal{M}_n} \Big\{\gamma_n(\widehat{f}'_{m,K}) + \mathrm{pen}_K(m)\Big\},\ \mathcal{M}_n = \{1, \dots, [\sqrt{n}]\} \tag{2.28}$$

with

$$\mathrm{pen}_K(m) = c_1\|f\|_\infty \frac{m^2\ \log(n)}{n} + c_2(\|f\|_\infty \vee 1)\frac{m\,K\ \log(n)}{n} := \mathrm{pen}_1(m) + \mathrm{pen}_{2,K}(m). \tag{2.29}$$

It is easy to ckeck that $\gamma_n(\widehat{f}'_{m,K}) = -\|\widehat{f}'_{m,K}\|^2$. We prove the following result

Theorem 2.3.1. *Let $\widehat{f}'_{m,K_n}$ be defined by (2.26) with $m = \widehat{m}_{K_n}$ selected by (2.28)-(2.29) and K_n such that (2.27) holds. Then for c_1 and c_2 larger than fixed constants $c_{0,1}, c_{0,2}$, we have*

$$\mathbb{E}\left(\|f' - \widehat{f}'_{\widehat{m},K_n}\|^2\right) \leqslant C\left(\|f' - f'_m\|^2 + m^2\frac{\log(n)}{n} + m\frac{K_n\log(n)}{n}\right) + \frac{C'}{n},$$

where C is a numerical constant and C' depends on f.

Theorem 2.3.1 implies that the adaptive estimator $\widehat{f}'_{m,K_n}$ provides the adequate compromise, up to log terms.

2.4 Numerical examples

In this section, we provide a nonexhaustive illustration of our theoretical results.

2.4.1 Simulation setting and implementation.

We illustrate the performances of the adaptive estimator $\widehat{f}_{\widehat{m}_n,(d)}$ defined in (2.7), with $\widehat{m}$ selected by (2.17)-(2.18), for different distributions and values of d ($d = 1, 2$). In the **Hermite case** we consider the following distributions which are estimated on the interval I, which we fix to ensure reproducibility of our experiments :

(i) Gaussian standard $\mathcal{N}(0, 1)$, $I = [-4, 4]$,

(ii) Mixed Gaussian $0.4\mathcal{N}(-1, 1/4) + 0.6\mathcal{N}(1, 1/4)$, $I = [-2.5, 2.5]$,

(iii) Cauchy standard, density : $f(x) = (\pi(1 + x^2))^{-1}$, $I = [-6, 6]$,

(iv) Gamma $\Gamma(5, 5)/10$, $I = [0, 7]$,

(v) Beta $5\beta(4, 5)$, $I = [0, 5]$.

In the **Laguerre case** we consider densities (iv), (v) and the two following additional distributions

(vi) Weibull $W(4, 1)$, I = $[0, 1.5]$,

(vii) Maxwell with density $\sqrt{2}x^2e^{-x^2/(2\sigma^2)}/(\sigma^3\sqrt{\pi})$, with $\sigma = 2$ and I = $[0, 8]$.

All these distributions satisfy Assumptions (**A**1), (**A**2) and densities (iv)-(vii) satisfy (**A**3). The moment conditions given in (2.9) are fulfilled for $d = 1, 2$, even by the Cauchy distribution (iii) which has finite moments of order $2/3 < 1$. For the adaptive procedure, the model collection considered is $\mathcal{M}_{n,d} = \{d, \dots, m_n(d)\}$, where the maximal dimension is $m_n(d) = 50$ in the Laguerre case and $m_n(d) = 40$ in the Hermite case, for all values of n and d (smaller values may be sufficient and spare computation time). In practice, the adaptive procedure follows the steps :

- For m in $\mathcal{M}_{n,d}$, compute $-\sum_{j=0}^{m-1}(\widehat{a}_j^{(d)})^2 + \widehat{\text{pen}}_d(m)$, with $\widehat{a}_j^{(d)}$ given in (2.7) and $\widehat{\text{pen}}_d(m)$ in (2.18),
- Choose $\widehat{m}_n$ via $\widehat{m}_n = \underset{m\in\mathcal{M}_{n,d}}{\text{argmin}}\ \{-\sum_{j=0}^{m-1}(\widehat{a}_j^{(d)})^2 + \widehat{\text{pen}}_d(m)\}$,
- Compute $\widehat{f}_{\widehat{m}_n,(d)} = \sum_{j=0}^{\widehat{m}-1}\widehat{a}_j^{(d)}\varphi_j$.

Then, we compute the empirical *Mean Integrated Squared Errors (MISE)* of $\widehat{f}_{\widehat{m}_n,(d)}$. For that, we first compute the ISE by Riemann discretization in 100 points : for the j-th path, and the j-th estimate $\widehat{g}_{\widehat{m}}^{(j)}$ of g, where g stands either for the density f or for its derivative f', we set

$$\|g-\widehat{g}_{\widehat{m}}^{(j)}\|^2 \approx \frac{\text{length}(I)}{K}\sum_{k=1}^{K}(\widehat{g}_{\widehat{m}}^{(j)}(x_k))-g(x_k))^2, \quad x_k = \min(I)+k\frac{\text{length}(I)}{K}, \quad k = 1, \dots, K,$$

for $j = 1, \dots R$. To get the MISE, we average over j of these R values of ISEs.
The *constant* κ in the penalty is calibrated by preliminary experiments. A comparison of the MISEs for different values of κ and different distributions (distinct from the previous

ones to avoid overfitting) allows to choose a relevant value. We take $\kappa = 3.5$ for the density and its first derivative and $\kappa = 5$ for the second order derivative in the Laguerre case or $\kappa = 4$ for the density and its first derivative and $\kappa = 6.5$ for the second order derivative in the Hermite case.

Comparison with kernel estimators. We compare the performances of our method with those of kernel estimators, and start by density estimation ($d = 0$). The density kernel estimator is defined as follows

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right), \quad x \in \mathbb{R}$$

where $h > 0$ is the bandwidth and K a kernel such that $\int K(x)dx = 1$. These two quantities (h and K) are user-chosen. For density estimation, we use the function implemented in the statistical software `R` called `density`, where the kernel is chosen Gaussian and the bandwidth selected by plug-in (`R`-function `bw.SJ`), see Tables 2.2 and 2.4.

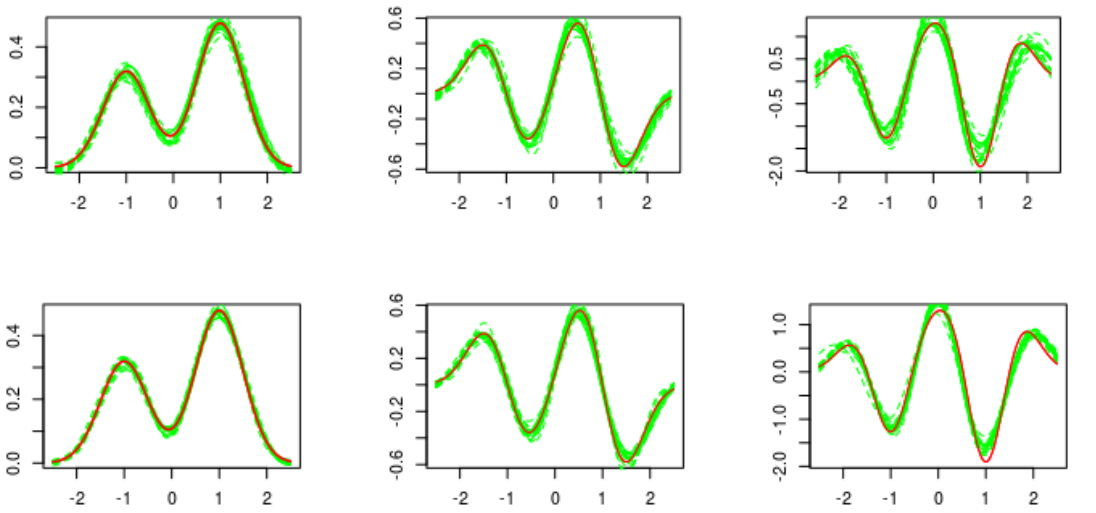


FIGURE 2.1 – 20 estimates $\widehat{f}_{\widehat{m}_n,(d)}$ in the Hermite basis of a Mixed Gaussian distribution (ii), with $n = 500$ (first line) and $n = 2000$ (second line). The true quantity is in bold red and the estimate in dotted lines (left $d = 0$, middle $d = 1$ and right $d = 2$).

For the estimation of the derivative, the kernel estimator we compare with (see Tables 2.3 and 2.5) is defined by :

$$\widehat{f}'_h(x) = -\frac{1}{nh^2} \sum_{i=1}^{n} K'\left(\frac{X_i - x}{h}\right).$$

In that latter case there is no ready-to-use procedure implemented in `R`; therefore, we generalize the adaptive procedure of Lacour et al. (2017) from density to derivative estimation. To that aim, we consider a kernel of order 7 (*i.e.* $\int x^j K(x)dx = 0$, *for* $j = 1, \ldots, 7$) built as a Gaussian mixture defined by :

$$K(x) = 4n_1(x) - 6n_2(x) + 4n_3(x) - n_4(x), \tag{2.30}$$

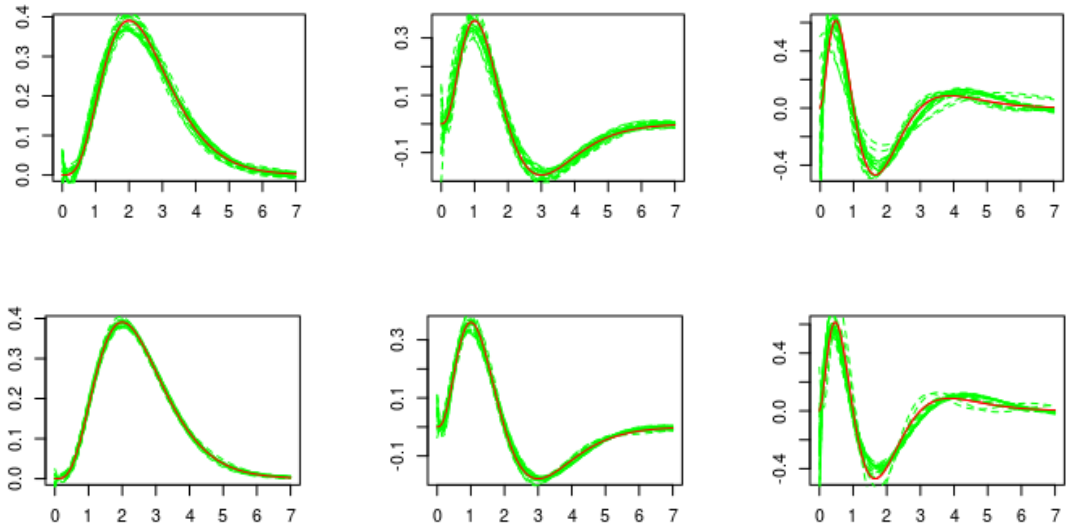


FIGURE 2.2 – 20 estimates $\widehat{f}_{\widehat{m}_n,(d)}$ in the Laguerre basis of a Gamma distribution (iv), with $n = 500$ (first line), and $n = 2000$ (second line). The true quantity is in bold red and the estimate in dotted lines (left $d = 0$, middle $d = 1$ and right $d = 2$).

f		Hermite case		Laguerre case	
Density		(ii)		(vi)	
n		500	2000	500	2000
Mean of m_{opt}	$d = 0$	7.65	9.45	5.85	7.65
	$d = 1$	8.15	9.70	6.15	6.80
	$d = 2$	7.85	8.95	5.15	5.65

TABLE 2.1 – Mean of selected dimensions $\widehat{m}_n$ presented in Figures 2.1 and 2.2.

where $n_j(x)$ is the density of a centered Gaussian with a variance equal to j : the higher the order, the better the results, in theory (see Tsybakov (2009)) and in practice (see Comte and Marie (2020)). By analogy with the proposal of Lacour et al. (2017) for density estimation, we select h by :

$$\widehat{h} = \underset{h\in\mathcal{H}}{\operatorname{argmin}}\{\|\widehat{f}'_h - \widehat{f}'_{h_{min}}\|^2 + \operatorname{pen}(h)\}, \text{ with } \operatorname{pen}(h) = \frac{4}{n}\langle K'_h, K_{h'_{min}}\rangle,$$

where $h_{min} = \min \mathcal{H}$, for $\mathcal{H}$ the collection of bandwidths chosen in $[c/n, 1]$ and $K_h(x) = \frac{1}{h}K(\frac{x}{h})$. Note that

$$\operatorname{pen}(h) = \frac{4}{n}\langle K'_h, K_{h'_{min}}\rangle = \frac{4}{nh^2h^2_{min}}\int K'(\frac{u}{h})K'(\frac{u}{h_{min}})du$$

and this term can be explicitely computed with the definition of K in (2.30).

2.4.2 Results and discussion.

Figures 2.1 and 2.2 show 20 estimated f, f', f'' in case (ii), for two values of n, 500 and 2000. These plots can be read as variability bands illustrating the performance and the stability of the estimator. We observe that increasing n improves the estimation and, on the contrary, that increasing the order of the derivative makes the problem more difficult. The means of the dimensions selected by the adaptive procedure are given in Table 2.1. Unsurprisingly, this dimension increases with the sample size n. In average, these dimensions are comparable for $d \in \{0, 1, 2\}$, this is in accordance with the theory : the optimal value m_{opt} does not depend on d.
Tables 2.2 and 2.4 for $d = 0$ and Tables 2.3 and 2.5 for $d = 1$ allow to compare the MISEs obtained with our method and the kernel method for different sample sizes and densities.The error decreases when the sample size increases for both methods. For density estimation ($d = 0$), the results obtained with our Hermite projection method in Table 2.2 are better in most cases than the kernel competitor, except for smallest sample size $n = 100$ and Gamma (iv) and Beta (v) distributions. Table 2.3 gives the risks obtained for derivative estimation in the Hermite basis : our method is better for densities (i), (ii), (iii) (except for $n = 100$ for Gaussian distribution (i)), but the kernel method is often better for densities (iv) and (v) ; they correspond to Gamma and beta densities which are in fact with support included in $\mathbb{R}^+$.

In Table 2.4, we compare the errors obtained for densities (iv)-(vii) with support in $\mathbb{R}^+$. Our method is always better than the `R`-kernel estimate. For the derivatives, in Table 2.5, our method and the kernel estimator seem equivalent. Lastly, Table 2.6 allows to compare Laguerre and Hermite bases for the estimation of the second order derivatives of functions (iv) and (v), for larger sample sizes. As expected, the risks are larger, because the degree of ill posedness increases and thus the rate deteriorates. For these $\mathbb{R}^+$-supported functions, the Laguerre basis is clearly better. It is possible that scale of the functions themselves also increase (multiplicative factors appearing by derivation). Note that the same phenomenon is observed for the $\mathbb{L}^1$-risk computed in Shen and Ghosal (2017), see their Table 1.

	Our method				Kernel method			
f \ n	100	500	1000	2000	100	500	1000	2000
Gaussian (i)	0.12	0.03	0.02	4.10^{-3}	0.74	0.23	0.13	0.07
Mixed Gaussian (ii)	1.01	0.26	0.13	0.07	1.46	0.44	0.22	0.14
Cauchy (iii)	0.63	0.38	0.19	0.10	4.26	3.42	1.75	0.89
Gamma (iv)	1.46	0.36	0.18	0.09	0.99	0.26	0.14	0.08
Beta (v)	1.09	0.18	0.10	0.05	0.96	0.26	0.151	0.09

TABLE 2.2 – Empirical MISE $100 \times \mathbb{E}\|\widehat{f}_{\widehat{m},(0)} - f\|^2$ (left) and $100 \times \mathbb{E}\|\widehat{f}_{\widehat{h}} - f\|^2$ (right, Kernel Estimator) for $R = 100$ in the Hermite case.

	Our method				Kernel method			
f \ n	100	500	1000	2000	100	500	1000	2000
Gaussian (i)	1.21	0.30	0.15	0.10	1.16	0.81	0.53	0.25
Mixed Gaussian (ii)	10.08	2.39	1.89	1.07	14.13	3.56	2.00	1.2
Cauchy (iii)	2.91	1.28	0.87	0.56	4.14	1.58	1.19	0.88
Gamma (iv)	5.88	1.89	1.43	0.60	2.45	1.25	0.75	0.63
Beta (v)	5.84	1.76	0.91	0.87	5.62	3.19	0.59	0.33

TABLE 2.3 – Empirical MISE $100 \times \mathbb{E}\|\widehat{f}_{\widehat{m},(1)} - f'\|^2$ (left) and $100 \times \mathbb{E}\|\widehat{f}'_{\widehat{h}} - f'\|^2$ (right) for $R = 100$ in the Hermite case.

	Our method				Kernel method			
f \ n	100	500	1000	2000	100	500	1000	2000
Gamma (iv)	0.54	0.16	0.08	0.04	0.99	0.26	0.14	0.08
Beta (v)	0.86	0.20	0.10	0.06	0.96	0.26	0.15	0.09
Weibull (vi)	2.61	0.60	0.33	0.17	3.55	0.80	0.46	0.29
Maxwell (vii)	0.64	0.11	0.06	0.04	0.59	0.16	0.10	0.06

TABLE 2.4 – Empirical MISE $(100 \times \mathbb{E}\|\widehat{f}_{\widehat{m},(0)} - f\|^2$ (left) and $100 \times \mathbb{E}\|\widehat{f}_{\widehat{h}} - f\|^2$ (right) for $R = 100$ in the Laguerre case.

	Our method				Kernel method			
f \ n	100	500	1000	2000	100	500	1000	2000
Gamma (iv)	5.21	0.95	0.48	0.17	2.45	1.25	0.75	0.63
Beta (v)	4.55	1.55	0.95	0.45	5.62	3.19	0.59	0.33
Weibull (vi)	126.95	34.54	22.31	14.10	127.38	38.60	35.47	11.36
Maxwell (vii)	1.46	0.60	0.24	0.13	0.87	0.21	0.18	0.10

TABLE 2.5 – Empirical MISE : $100 \times \mathbb{E}\|\widehat{f}_{\widehat{m},(1)} - f'\|^2$ (left) and $100 \times \mathbb{E}\|\widehat{f}'_{\widehat{h}} - f'\|^2$ (right) for $R = 100$ in the Laguerre case.

	Hermite case				Laguerre case			
f \ n	1000	2000	5000	10000	1000	2000	5000	10000
Gamma (iv)	6.40	4.20	3.39	2.91	3.98	3.70	1.92	1.00
Beta (v)	11.32	9.45	4.14	1.42	7.60	5.05	2.43	1.99

TABLE 2.6 – Empirical MISE $100 \times \mathbb{E}\|\widehat{f}^{(2)}_{\widehat{m},(2)} - f^{(2)}\|^2$ for $R = 100$.

2.5 Concluding remarks

We introduced a projection estimator of $f^{(d)}$ from observations $X_1, \dots, X_n$ of density f, relying on the Laguerre or Hermite basis. Theoretical study prove that our estimator is adaptive and minimax optimal in the minimax sense. We also provide numerical study and the comparison with the kernel method ensure the good performance and the stability of our procedure. For future works, it is interesting to extend these results to the dependent case. In literature, results exist (see *e.g.* Schmisser (2013)) with a penalty which depends on the mixing coefficient, which is unknown. It is therefore interesting to see if it is possible to propose an adaptive procedure with a penalty independent of the mixing coefficients.

2.6 Proofs

In the sequel C denotes a generic constant whose value may change from line to line and whose dependency is sometimes given in indexes.

2.6.1 Proof of Theorem 2.2.1

Following (2.8) we study the variance term, notice that $\mathbb{E}\big[\|\widehat{f}_{m,(d)} - f_m^{(d)}\|^2\big] = \sum_{j=0}^{m-1} \mathrm{Var}(\widehat{a}_j^{(d)})$. By definition of $\widehat{a}_j^{(d)}$ given in (2.7), we have

$$\mathrm{Var}(\widehat{a}_j^{(d)}) = \mathrm{Var}\left(\frac{(-1)^d}{n}\sum_{i=1}^{n}\varphi_j^{(d)}(X_i)\right) = \frac{1}{n}\mathrm{Var}(\varphi_j^{(d)}(X_1)) = \frac{1}{n}\mathbb{E}[(\varphi_j^{(d)}(X_1))^2] - \frac{a_j^2(f^{(d)})}{n}. \tag{2.31}$$

Clearly, $\sum_{j=0}^{m-1} a_j^2(f^{(d)}) = \|f_m^{(d)}\|^2$. In the sequel we denote by $V_{m,d}$ the quantity

$$V_{m,d} = \sum_{j=0}^{m-1} \mathbb{E}[(\varphi_j^{(d)}(X_1))^2]. \tag{2.32}$$

The remaining of the proof consists in showing that under (2.9) we have $V_{m,d} \leqslant cm^{d+1/2}$. For that, write

$$V_{m,d} = \sum_{j=0}^{m-1}\int(\varphi_j^{(d)}(x))^2 f(x)dx = \left(\sum_{j=0}^{d-1}\int(\varphi_j^{(d)}(x))^2 f(x)dx + \sum_{j=d}^{m-1}\int(\varphi_j^{(d)}(x))^2 f(x)dx\right), \tag{2.33}$$

where

$$\sum_{j=0}^{d-1}\int(\varphi_j^{(d)}(x))^2 f(x)dx \leqslant \sum_{j=0}^{d-1}\|\varphi_j^{(d)}\|_\infty^2 := c(d). \tag{2.34}$$

To bound the second term in (2.33), we consider separately Hermite and Laguerre cases.

The Laguerre case.

We derive from (2.1) that

$$\ell_j^{(d)}(x) = \sqrt{2}\sum_{k=0}^{d}(-1)^{d-k}\binom{d}{k}L_j^{(k)}(2x)e^{-x}.$$

Using Koekoek (1990), Equation 2.10, we derive

$$L_j^{(k)}(x) = \frac{d^k}{dx^k}L_j(x) = (-1)^k L_{j-k,(k)}(x), \quad \text{where} \quad L_{p,(\delta)}(x) = \frac{1}{p!}e^x x^{-\delta}\frac{d^p}{dx^p}\left(x^{\delta+p}e^{-x}\right)\mathbf{1}_{\delta\leqslant p}.$$

Moreover, introduce the orthonormal basis on $\mathbb{L}^2(\mathbb{R}^+)$ $(\ell_{k,(\delta)})_{0\leqslant k<\infty}$ by

$$\ell_{k,(\delta)}(x) = 2^{\frac{\delta+1}{2}}\left(\frac{k!}{\Gamma(k+\delta+1)}\right)^{1/2}L_{k,(\delta)}(2x)x^{\frac{\delta}{2}}e^{-x}. \tag{2.35}$$

Therefore, $(L_j(2x))^{(k)} = 2^k L_{j-k,(k)}(2x)\mathbf{1}_{j\geqslant k}$, so that

$$\ell_j^{(d)}(x) = (-1)^d\sum_{k=0}^{d}\binom{d}{k}2^{\frac{k}{2}}x^{-k/2}\left(\frac{j!}{(j-k)!}\right)^{\frac{1}{2}}\ell_{j-k,(k)}(x), \tag{2.36}$$

where $\ell_{j,(\delta)}$ is defined in (2.35). Using the Cauchy Schwarz inequality in (2.36), we derive that

$$\begin{aligned}\sum_{j=d}^{m-1}\int_0^\infty [\ell_j^{(d)}(x)]^2 f(x)dx \leqslant& 3^d\sum_{j=d}^{m-1}\sum_{k=0}^{d}\binom{d}{k}\frac{j!}{(j-k)!}\int_0^{+\infty}x^{-k}[\ell_{j-k,(k)}(x)]^2 f(x)dx\\ \leqslant& C_d\sum_{j=d}^{m-1}\sum_{k=0}^{d}j^d\int_0^{+\infty}x^{-k}(\ell_{j-k,(k)}(x/2))^2 f(x/2)dx.\end{aligned}$$

Now we rely on the following Lemma, proved in Appendix 2.6.8.

Lemma 2.6.1. *Let $j \geqslant k \geqslant 0$ and suppose that $\mathbb{E}[X^{-k-1/2}] < +\infty$, it holds, for a positive constant C depending only on k, that*

$$\int_0^{+\infty}x^{-k}\left[\ell_{j-k,(k)}(x/2)\right]^2 f(x/2)dx \leqslant \frac{C}{\sqrt{j}}.$$

From Lemma 2.6.1, we obtain

$$\sum_{j=d}^{m-1}\int(\ell_j^{(d)}(x))^2 f(x)dx \leqslant C\sum_{j=d}^{m-1}\sum_{k=0}^{d}j^{d-1/2} \leqslant Cm^{d+1/2}.$$

Plugging this and (2.34) in (2.33), gives the result (2.10) and Theorem 2.2.1 in the Laguerre case.

The Hermite case.

We first introduce a useful technical result, its proof is given in Appendix 2.6.8.

Lemma 2.6.2. *Let h_j given in* (4.9)*, the d-th derivative of h_j is such that*

$$h_j^{(d)} = \sum_{k=-d}^{d} b_{k,j}^{(d)} h_{j+k}, \quad \textit{where} \quad b_{k,j}^{(d)} = \mathcal{O}(j^{d/2}), \quad j \geqslant d \geqslant |k|. \tag{2.37}$$

Using successively Lemma 2.6.2, the Cauchy Schwarz inequality and Lemma 8.5 in Comte and Genon-Catalot (2018) (using that $\mathbb{E}[|X_1|^{2/3}] < \infty$), we obtain, for $k+j$ large enough,

$$\begin{aligned}\sum_{j=d}^{m-1} \int (h_j^{(d)}(x))^2 f(x)dx \leqslant & (2d+1) \sum_{j=d}^{m-1} \sum_{k=-d}^{d} (b_{k,j}^{(d)})^2 \int h_{j+k}(x)^2 f(x)dx \leqslant d(2d+1)^2 \sum_{k=-d}^{d} \sum_{j=d}^{m-1} cj^{d-\frac{1}{2}} \\ \leqslant & c'(d) m^{d+\frac{1}{2}}. \end{aligned} \tag{2.38}$$

Plugging (2.38) and (2.34) in (2.33) leads to inequality (2.10) and Theorem 2.2.1 in the Hermite case.

2.6.2 Proof of Proposition 2.2.1

We build a lower bound for (2.8). Recalling (2.31) and notation $V_{m,d} = \sum_{j=0}^{m-1} \mathbb{E}[(\varphi_j^{(d)}(X_1))^2]$, to establish Proposition 2.2.1, we have to build a minorant for $V_{m,d}$. We consider separately the Laguerre and Hermite cases.

The Laguerre case.

Using (2.36), we have

$$\begin{aligned}\ell_j^{(d)}(x) =& (-1)^d 2^{d/2} x^{-d/2} \Big(\frac{j!}{(j-d)!}\Big)^{1/2} \ell_{j-d,(d)}(x) + (-1)^d \sum_{k=0}^{d-1} \binom{d}{k} 2^{\frac{k}{2}} x^{-k/2} \left(\frac{j!}{(j-k)!}\right)^{\frac{1}{2}} \ell_{j-k,(k)}(x) \\ :=& T_1(x) + T_2(x).\end{aligned}$$

It follows that

$$\int_0^{+\infty} (\ell_j^{(d)})^2(x) f(x)dx \geqslant \int_0^{+\infty} T_1(x)^2 f(x)dx + 2\int_0^{+\infty} T_1(x)T_2(x)f(x)dx := E_1 + E_2.$$

For the first term, as (**A**1) ensures that f is a continuous density, there exist $0 \leqslant a < b$ and $c > 0$, such that $\inf_{a \leqslant x \leqslant b} f(x) \geqslant c > 0$. We derive

$$E_1 \geqslant 2^d \frac{j!}{(j-d)!} \int_0^{+\infty} x^{-d} \ell_{j-d,(d)}^2(x) f(x)dx \geqslant c2^d (j-d)^d b^{-d} \int_a^b \ell_{j-d,(d)}^2(x)dx.$$

By Theorem 8.22.5 in Szegö (1959), for $\delta > -1$ an integer, and for $\underline{b}/j \leqslant x \leqslant \bar{b}$, where $\underline{b}$, $\bar{b}$ are arbitrary positive constants, it holds

$$\ell_{j,(\delta)}(x) = \mathfrak{d}(jx)^{-\frac{1}{4}} \left(\cos(2\sqrt{2}\sqrt{jx} - \frac{\delta\pi}{2} - \frac{\pi}{4}) + (jx)^{-\frac{1}{2}}\mathcal{O}(1)\right), \tag{2.39}$$

where $\mathcal{O}(1)$ is uniform on $[\underline{b}/j, \bar{b}]$ and $\mathfrak{d} = 2^{1/4}/\sqrt{\pi}$. It follows that,

$$\ell^2_{j,(\delta)}(x) = \frac{\mathfrak{d}^2}{2}(jx)^{-\frac{1}{2}}\left[1 + \cos(4\sqrt{2}\sqrt{jx} - \delta\pi - \frac{\pi}{2})\right] + (jx)^{-1}\mathcal{O}(1).$$

We derive that $\int_a^b \ell^2_{j-d,(d)}(x)dx \geqslant C(j-d)^{-1/2}$, after a change of variable $y = \sqrt{x}$, for some positive constant C depending on a, b and d. Consequently, it holds

$$E_1 \geqslant C(j-d)^{d-\frac{1}{2}} \geqslant C' j^{d-\frac{1}{2}}, \quad \forall j \geqslant 2d, \tag{2.40}$$

where C' depends on a, b, c and d. For the second term, we have

$$\begin{aligned}|E_2| &\leqslant 2\int_0^{+\infty} |T_1(x)T_2(x)|f(x)dx \\ &\leqslant 2j^{\frac{d}{2}} j^{\frac{d-1}{2}} \sum_{k=0}^{d-1} \binom{d}{k} 2^{\frac{k+d}{2}} \left[\int_0^{+\infty} x^{-d}\ell^2_{j-d,(d)}(x)f(x)dx + \int_0^{+\infty} x^{-k}\ell^2_{j-k,(k)}(x)f(x)dx\right].\end{aligned}$$

By Lemma 2.6.1, it follows that

$$|E_2| \leqslant Cj^{\frac{d}{2}} j^{\frac{d-1}{2}} j^{-\frac{1}{2}} \sum_{k=0}^{d-1} \binom{d}{k} 2^{\frac{k+d}{2}} \leqslant Cj^{d-1}.$$

This together with (2.40), lead to $\int_0^{+\infty} (\ell_j^{(d)})^2(x)f(x)dx \geqslant C' j^{d-\frac{1}{2}}$, $j \geqslant 2d$ where C depends on a, b, c and d. We derive

$$V_{m,d} \geqslant Cm^{d+\frac{1}{2}}, \tag{2.41}$$

which ends the proof in the Laguerre case.

The Hermite Case.

The proof is similar to the Laguerre case. Consider the following expression of h_j (see Szegö (1959), p.248) :

$$h_j(x) = \lambda_j \cos\left((2j+1)^{\frac{1}{2}}x - \frac{j\pi}{2}\right) + \frac{1}{(2j+1)^{\frac{1}{2}}}\xi_j(x), \quad \forall x \in \mathbb{R}, \tag{2.42}$$

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

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

By Stirling Formula, it holds

$$\lambda_{2j} = \frac{(2j)!^{\frac{1}{2}}}{2^j j!\pi^{1/4}} \sim \pi^{-1/2}j^{-1/4} \text{ and } \lambda_{2j+1} = \lambda_{2j}\frac{\sqrt{2j+1}}{\sqrt{2j+3/2}} \sim \pi^{-1/2}j^{-1/4}. \tag{2.43}$$

Differentiating (2.42), we get

$$h_j^{(d)}(x) = \lambda_j (2j+1)^{\frac{d}{2}} \cos\left((2j+1)^{\frac{1}{2}}x - \frac{j\pi}{2} + \frac{d\pi}{2}\right) + \frac{1}{\sqrt{2j+1}}\xi_j^{(d)}(x).$$

Note that if $d = 2$ it holds

$$\xi_j^{(2)}(x) = \sqrt{2j+1}x^2 h_j(x) - (2j+1)\xi_j(x). \tag{2.44}$$

From (**A**1), there exists $a < b$ and $c > 0$ such that $\inf_{a \leqslant x \leqslant b} f(x) \geqslant c > 0$. It follows

$$\begin{aligned}\int_{\mathbb{R}} h_j^{(d)}(x)^2 f(x)dx \geqslant & c(2j+1)^d \lambda_j^2 \int_a^b \cos^2\left((2j+1)^{\frac{1}{2}}x - (j+d)\frac{\pi}{2}\right)dx \\ & + 2c\lambda_j(2j+1)^{\frac{d-1}{2}} \int_a^b \cos\left((2j+1)^{\frac{1}{2}}x - (j+d)\frac{\pi}{2}\right)\xi_j^{(d)}(x)dx := E_1 + E_2.\end{aligned}$$

For the first term, using $\cos^2(x) = (1+\cos(2x))/2$ and (2.43), we get

$$E_1 = c(2j+1)^d \lambda_j^2 \left(\frac{b-a}{2} + \mathcal{O}(\frac{1}{\sqrt{j}})\right) \geqslant c' j^{d-\frac{1}{2}} \left(\frac{b-a}{2} + \mathcal{O}(\frac{1}{\sqrt{j}})\right).$$

For the second term we first show that

$$\forall x \in [a,b],\ \forall j \geqslant 0,\ \forall d \geqslant 0,\ \xi_j^{(d)}(x) = \mathcal{O}(j^{d/2}). \tag{2.45}$$

To establish (2.45) we first note, using (2.44), that for $d \geqslant 2$, $\forall x \in \mathbb{R}$,

$$\xi_j^{(d)}(x) + (2j+1)\xi_j^{(d-2)}(x) = (\xi_j^{(2)}(x) + (2j+1)\xi_j(x))^{(d-2)} = \sqrt{2j+1}(x^2 h_j(x))^{(d-2)} =: \Psi_{j,d}(x).$$

Together with Lemma 2.6.2, one easily obtains by induction that $\forall x \in [a,b]$, $\forall j \geqslant 0$, $\Psi_{j,d}(x) = \mathcal{O}(j^{\frac{d-1}{2}})$. The latter result gives $\xi_j^{(d)} = -j\xi_j^{(d-2)} + \Psi_{j,d}$ and an immediate induction on d leads to (2.45). Injecting this in E_2 gives, together with (2.43), $|E_2| \leqslant C j^{d-\frac{3}{4}}$, for a positive constant C depending on a, b, c and d. Gathering the bound on E_1 and E_2 lead to

$$\int_{\mathbb{R}} h_j^{(d)}(x)^2 f(x)dx \geqslant c' j^{d-\frac{1}{2}} \left(\frac{b-a}{2} + \mathcal{O}(\frac{1}{\sqrt{j}})\right) - \mathcal{O}(j^{d-\frac{3}{4}}) \geqslant C'_d j^{d-\frac{1}{2}},$$

and

$$V_{m,d} \geqslant c_d m^{d+\frac{1}{2}}, \tag{2.46}$$

which ends the proof of the Hermite case.

2.6.3 Proof of (2.16)

We apply Theorem 2.7 in Tsybakov (2009). We start by the construction of a family of hypotheses $(f_\theta)_\theta$. The construction is inspired by Belomestny et al. (2017). Define f_0 by

$$f_0(x) = P(x)\mathbf{1}_{]0,1[}(x) + \frac{1}{2}x\mathbf{1}_{[1,2]}(x) + Q(x)\mathbf{1}_{]2,3]}(x), \tag{2.47}$$

where P and Q are positive polynomials, for $0 \leqslant k \leqslant s$, $P^{(k)}(0) = Q^{(k)}(3) = 0$, $P^{(k)}(1) = \lim_{x\downarrow 1}(x/2)^{(k)}$, $Q^{(k)}(2) = \lim_{x\uparrow 2}(x/2)^{(k)}$ and finally $\int_0^1 P(x)dx = \int_2^3 Q(x)dx = \frac{1}{8}$. Consider f_θ defined as a perturbation of f_0

$$f_\theta(x) = f_0(x) + \delta K^{-(\gamma+d)} \sum_{k=0}^{K-1} \theta_{k+1}\psi\big((x-1)(K+1) - k\big), \text{ with } K \in \mathbb{N}, \tag{2.48}$$

for some $\delta > 0$, $\theta = (\theta_1, \ldots, \theta_K) \in \{0,1\}^K$, $\gamma > 0$ and ψ which is supported on $[1,2]$, admits bounded derivatives up to order s and is such that $\int_1^2 \psi(x)dx = 0$. The lower bound (2.16) is a consequence of the following Lemma.

Lemma 2.6.3. *(i). Let $s \geqslant d$, $\forall\ \theta \in \{0,1\}^K$, there exist δ small enough and $\gamma > 0$ such that f_θ is density. There exists $D > 0$ such that f_θ belongs to $W_H^s(D)$. If in addition $\gamma \geqslant s - d$, f_θ belongs to $W_L^s(D)$.*

(ii). Let M an integer, for all $j < l \leqslant M$, $\forall \theta^{(j)}$, $\theta^{(l)}$ in $\{0,1\}^K$, it holds $\|f_{\theta^{(j)}}^{(d)} - f_{\theta^{(l)}}^{(d)}\|^2 \geqslant C\delta^2 K^{-2\gamma}$.

(iii). For δ small enough, $K = n^{1/(2\gamma+2d+1)}$ and for all $(\theta^{(j)})_{1\leqslant j\leqslant M} \in (\{0,1\}^K)^M$, it holds

$$\frac{1}{M}\sum_{j=1}^{M} \chi^2\left({f_{\theta^{(j)}}}^{\otimes n}, {f_0}^{\otimes n}\right) \leqslant \alpha M,$$

where $0 < \alpha < 1/8$ and $\chi^2(g,h)$ denotes the χ^2 divergence between the distributions g and h.

Choosing $\gamma = s - d$, $K = n^{1/(2\gamma+2d+1)}$ and δ small enough, we derive from Lemma 2.6.3 that,

$$\|f_{\theta^{(j)}}^{(d)} - f_{\theta^{(l)}}^{(d)}\|^2 \geqslant C\delta^2 n^{-2\frac{(s-d)}{2s+1}}, \quad \forall \theta^{(j)}, \theta^{(l)} \in \{0,1\}^K.$$

The announced result is then a consequence of Theorem 2.7 in Tsybakov (2009).

2.6.4 Proof of Theorem 2.2.2

Consider the contrast function defined as follows :

$$\gamma_{n,d}(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n}(-1)^d t^{(d)}(X_i), \quad t \in \mathbb{L}^2(\mathbb{R}),$$

for which $\widehat{f}_{m,(d)} = \underset{t\in S_m}{\operatorname{argmin}}\gamma_{n,d}(t)$ (see (2.7)) and $\gamma_n(\widehat{f}_{m,(d)}) = -\|\widehat{f}_{m,(d)}\|^2$. For two functions $t, s \in \mathbb{L}^2(\mathbb{R})$, consider the decomposition :

$$\gamma_{n,d}(t) - \gamma_{n,d}(s) = \|t - f^{(d)}\|^2 - \|s - f^{(d)}\|^2 - 2\nu_{n,d}(t-s), \tag{2.49}$$

where

$$\nu_{n,d}(t) = \frac{1}{n}\sum_{i=1}^{n}\left((-1)^d t^{(d)}(X_i) - \langle t, f^{(d)}\rangle\right).$$

By (2.18), it holds for all $m \in \mathcal{M}_{n,d}$, that $\gamma_{n,d}(\widehat{f}_{\widehat{m}_n,(d)}) + \widehat{\mathrm{pen}}_d(\widehat{m}_n) \leqslant \gamma_{n,d}(f_m^{(d)}) + \widehat{\mathrm{pen}}_d(m)$. Plugging this in (2.49) yields, for all $m \in \mathcal{M}_{n,d}$,

$$\|\widehat{f}_{\widehat{m}_n,(d)} - f^{(d)}\|^2 \leqslant \|f_m^{(d)} - f^{(d)}\|^2 + \widehat{\mathrm{pen}}_d(m) + 2\nu_{n,d}\left(\widehat{f}_{\widehat{m}_n,(d)} - f_m^{(d)}\right) - \widehat{\mathrm{pen}}_d(\widehat{m}_n). \quad (2.50)$$

Note that for $t \in \mathbb{L}^2(\mathbb{R})$, $\nu_{n,d}(t) = \|t\|\nu_{n,d}\big(t/\|t\|\big) \leqslant \|t\| \sup_{s \in S_m + S_{\widehat{m}}, \|s\|=1} |\nu_{n,d}(s)|$. Consequently, using $2xy \leqslant x^2/4 + 4y^2$, we obtain

$$2\nu_{n,d}\left(\widehat{f}_{\widehat{m}_n,(d)} - f_m^{(d)}\right) \leqslant \frac{1}{2}\|\widehat{f}_{\widehat{m}_n,(d)} - f^{(d)}\|^2 + \frac{1}{2}\|f_m^{(d)} - f^{(d)}\|^2 + 4 \sup_{t \in S_m + S_{\widehat{m}}, ||t||=1} |\nu_{n,d}(t)|^2. \quad (2.51)$$

It follows from (2.50) and (2.51) that :

$$\frac{1}{2}\|\widehat{f}_{\widehat{m}_n,(d)} - f^{(d)}\|^2 \leqslant \frac{3}{2}\|f_m^{(d)} - f^{(d)}\|^2 + \widehat{\mathrm{pen}}_d(m) + 4 \sup_{t \in S_m + S_{\widehat{m}}, ||t||=1} |\nu_{n,d}(t)|^2 - \widehat{\mathrm{pen}}_d(\widehat{m}_n).$$

Introduce the function $p(m, m') = 4\frac{V_{m \vee m',d}}{n}$, we get, after taking the expectation,

$$\begin{aligned}\frac{1}{2}\mathbb{E}\left[\|\widehat{f}_{\widehat{m}_n,(d)} - f^{(d)}\|^2\right] \leqslant \frac{3}{2}\|f_m^{(d)} - f^{(d)}\|^2 + \mathrm{pen}_d(m) \\ + 4\mathbb{E}\left[\left(\sup_{t \in S_m + S_{\widehat{m}}, ||t||=1} |\nu_{n,d}(t)|^2 - p(m, \widehat{m}_n)\right)_+\right] \\ + \mathbb{E}[4p(m, \widehat{m}_n) - \mathrm{pen}_d(\widehat{m}_n)] + \mathbb{E}\left[(\mathrm{pen}_d(\widehat{m}_n) - \widehat{\mathrm{pen}}_d(\widehat{m}_n))_+\right].\end{aligned}$$

The remaining of the proof is a consequence of the following Lemma.

Lemma 2.6.4. *Under the assumptions of Theorem 2.2.2, the following hold.*

(i) There exists a constant Σ_1 such that :

$$\mathbb{E}\left[\left(\sup_{t \in S_m + S_{\widehat{m}}, ||t||=1} |\nu_{n,d}(t)|^2 - p(m, \widehat{m}_n)\right)_+\right] \leqslant \frac{\Sigma_1}{n}.$$

(ii) There exists a constant Σ_2 such that :

$$\mathbb{E}\left[(\mathrm{pen}_d(\widehat{m}_n) - \widehat{\mathrm{pen}}_d(\widehat{m}_n))_+\right] \leqslant \frac{1}{2}\mathbb{E}[\mathrm{pen}_d(\widehat{m}_n)] + \frac{\Sigma_2}{n}.$$

Lemma 2.6.4 yields

$$\frac{1}{2}\mathbb{E}\left[\|\widehat{f}_{\widehat{m}_n,(d)} - f^{(d)}\|^2\right] \leqslant \frac{3}{2}\|f_m^{(d)} - f^{(d)}\|^2 + \mathrm{pen}_d(m) + 4\frac{\Sigma_1}{n} + \mathbb{E}[4p(m, \widehat{m}_n) - \frac{1}{2}\mathrm{pen}_d(\widehat{m}_n)] + \frac{\Sigma_2}{n}.$$

Next, for $\kappa \geqslant 32 =: \kappa_0$, we have, $4p(m, \widehat{m}_n) \leqslant \mathrm{pen}_d(\widehat{m}_n)/2 + \mathrm{pen}_d(m)/2$. Therefore, we derive

$$\mathbb{E}\left[\|\widehat{f}_{\widehat{m}_n,(d)} - f^{(d)}\|^2\right] \leqslant 3\|f_m^{(d)} - f^{(d)}\|^2 + 3\mathrm{pen}_d(m) + 2\frac{4\Sigma_1 + \Sigma_2}{n}, \quad \forall m \in \mathcal{M}_{n,d}.$$

Taking the infimum on $\mathcal{M}_{n,d}$, $C = 3$ and $C' = 2(4\Sigma_1 + \Sigma_2)/n$ completes the proof.

2.6.5 Proof of Proposition 2.3.1.

First, it holds that

$$\mathbb{E}\Big[\|(\widehat{f}_m)' - f'\|^2\Big] \leqslant 2\Big[\|(f_m)' - f'\|^2 + \mathbb{E}[\|(\widehat{f}_m)' - (f_m)'\|^2]\Big]$$
$$= 2\int_0^{+\infty} (\sum_{j\geqslant m} a_j(f)\ell_j'(x))^2 dx + 2\mathbb{E}\left[\|\sum_{j=0}^{m-1} (\widehat{a}_j^{(0)} - a_j(f))\ell_j'\|^2\right].$$

For the first bias term, we derive from (2.2) that $\langle \ell_j', \ell_k'\rangle = 2 + 4j \wedge k$ for $j \neq k$ and $\langle \ell_j', \ell_j'\rangle = 1 + 4j$, and we derive that

$$\int_0^{+\infty} (\sum_{j\geqslant m} a_j(f)\ell_j'(x))^2 dx = \sum_{j\geqslant m} a_j(f)^2(1+4j) + 2\sum_{m\leqslant j<k} a_j(f)a_k(f)(2+4j).$$

First, for f in $W_L^s(D)$, we have

$$\sum_{j\geqslant m} a_j(f)^2(1+4j) \leqslant m^{-s}\sum_{j\geqslant m} j^s a_j(f)^2 + 4m^{-s+1}\sum_{j\geqslant m} j^s a_j(f)^2 \leqslant 5Dm^{-s+1},$$

and by the Cauchy-Schwarz inequality, it holds for a positive constant C,

$$\sum_{m\leqslant j<k} a_j(f)a_k(f) \leqslant \left(\sum_{m\leqslant j<k} j^s a_j(f)^2 k^s a_k(f)^2\right)^{\frac{1}{2}} \left(\sum_{m\leqslant j<k} j^{-s}k^{-s}\right)^{\frac{1}{2}}$$
$$\leqslant \sum_{j\geqslant m} j^s a_j(f)^2 \sum_{j\geqslant m} j^{-s} \leqslant DCm^{-s+1}$$

$$\sum_{m\leqslant j<k} j|a_j(f)a_k(f)| \leqslant \sum_{j\geqslant m} j|a_j(f)|\left(\sum_{k\geqslant j} k^s a_k(f)^2 \sum_{k\geqslant j} k^{-s}\right)^{\frac{1}{2}} \leqslant \sqrt{DC}\sum_{j\geqslant m} j^{\frac{s}{2}-s+\frac{3}{2}}|a_j(f)| \leqslant DCm^{-s+2}.$$

Thus, it comes

$$2\|(f_m)' - f'\|^2 \leqslant Cm^{-(s-2)}, \tag{2.52}$$

where $C > 0$ depends on D. Second, for the variance term, straightforward computations lead to

$$\mathbb{E}\Big[\|\sum_{j=0}^{m-1} (\widehat{a}_j^{(0)} - a_j(f))\ell_j'\|^2\Big] = \frac{1}{n}\int_0^{+\infty} \mathrm{Var}(\sum_{j=0}^{m-1} \ell_j(X_1)\ell_j'(x))dx \leqslant \frac{1}{n}\int_0^{+\infty} \mathbb{E}\left[(\sum_{j=0}^{m-1} \ell_j(X_1)\ell_j'(x))^2\right] dx.$$

By the orthonormality of $(\ell_j)_j$ and (**A3**), we obtain

$$\int_0^{+\infty} \mathbb{E}\left[(\sum_{j=0}^{m-1} \ell_j(X_1)\ell_j'(x))^2\right] dx \leqslant \|f\|_\infty \sum_{j,k=0}^{m-1} \int_0^{+\infty}\int_0^{+\infty} \ell_j(u)\ell_j'(x)\ell_k(u)\ell_k'(x)dudx$$
$$= \|f\|_\infty \sum_{j=0}^{m-1} (1+4j) \leqslant 3\|f\|_\infty m^2.$$

From this and (2.52), the result follows.

2.6.6 Proof of Proposition 2.3.2

By the Pythagoras Theorem, we have the bias-variance decomposition $\mathbb{E}\big[\|\widetilde{f}'_{m,K} - f'\|^2\big] = \|f' - f'_m\|^2 + \mathbb{E}\big[\|\widetilde{f}'_{m,K} - f'_m\|^2\big]$. As $\ell_j(0) = \sqrt{2}$, it follows that

$$\widetilde{f}'_{m,K} - f'_m = \sum_{j=0}^{m-1} \left[-\sqrt{2}(\widehat{f}_K(0) - f(0)) - \frac{1}{n}\sum_{i=1}^{n} (\ell'_j(X_i) - \mathbb{E}[\ell'_j(X_i)]) \right] \ell_j.$$

From the orthonormality of $(\ell_j)_j$, it follows

$$\begin{aligned}\mathbb{E}\big[\|\widetilde{f}'_{m,K} - f'_m\|^2\big] &= \sum_{j=0}^{m-1} \mathbb{E}\left[-\sqrt{2}(\widehat{f}_K(0) - f(0)) - \frac{1}{n}\sum_{i=1}^{n} (\ell'_j(X_i) - \mathbb{E}[\ell'_j(X_i)]) \right]^2 \\ &\leqslant 4m\mathbb{E}\Big[(\widehat{f}_K(0) - f(0))^2\Big] + 2\sum_{j=0}^{m-1} \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} (\ell'_j(X_i) - \mathbb{E}[\ell'_j(X_i)]) \right)^2 \right].\end{aligned}$$

Finally, using that the $(X_i)_i$ are i.i.d. lead to the result in the second variance term.

2.6.7 Proof of Theorem 2.3.1

We have the decomposition :

$$\gamma_n(t) - \gamma_n(s) = \|t - f'\|^2 - \|s - f'\|^2 - 2\langle s - t, f'\rangle - \frac{2}{n}\sum_{i=1}^{n} (s' - t')(X_i) - 2(s(0) - t(0))\widehat{f}_K(0)$$

and as $\langle t, f'\rangle = -t(0)f(0) - \int t' f$, we get

$$\gamma_n(t) - \gamma_n(s) = \|t - f'\|^2 - \|s - f'\|^2 - 2\nu_n(s - t) - 2(s(0) - t(0))(\widehat{f}_K(0) - f(0)), \quad (2.53)$$

$$\text{with} \qquad \nu_n(t) = \frac{1}{n}\sum_{i=1}^{n} (t'(X_i) - \langle t', f\rangle.$$

First note that for

$$f'_{m,K} = \sum_{j=0}^{m-1} a^{(1)}_{j,K}\ell_j, \quad a^{(1)}_{j,K} = \mathbb{E}[\widehat{a}^{(1)}_{j,K}] = \langle f', \ell_j\rangle + \ell_j(0)(f(0) - \mathbb{E}[\widehat{f}_K(0)],$$

it holds that

$$\begin{aligned}\|f' - f'_{m,K}\|^2 &= \left\| \sum_{j=0}^{\infty} \langle f', \ell_j\rangle\ell_j - \sum_{j=0}^{m-1} \langle f', \ell_j\rangle\ell_j - \sum_{j=0}^{m-1} \ell_j(0)\big(f(0) - \mathbb{E}[\widehat{f}_K(0)]\big)\ell_j \right\|^2 \\ &= \sum_{j\geqslant m} \langle f', \ell_j\rangle^2 + 2\sum_{j=0}^{m-1} \big(f(0) - \mathbb{E}[\widehat{f}_K(0)]\big)^2 = \|f' - f'_m\|^2 + 2m\big(f(0) - \mathbb{E}[\widehat{f}_K(0)]\big)^2.\end{aligned}$$

Let us start by writing that, by definition of $\widehat{m}_K$, it holds, $\forall m \in \mathcal{M}_n$,

$$\gamma_n(\widehat{f}'_{\widehat{m}_K,K}) + \mathrm{pen}_K(\widehat{m}_K) \leqslant \gamma_n(f'_{m,K}) + \mathrm{pen}_K(m),$$

which yields, with (2.53) and notations introduced in (2.29),

$$\begin{aligned}\|\widehat{f}'_{\widehat{m}_K,K} - f'\|^2 &\leqslant \|f'_{m,K} - f'\|^2 + \mathrm{pen}_K(m) + 2\nu_n(f'_{m,K} - \widehat{f}'_{\widehat{m}_K,K}) - \mathrm{pen}_1(\widehat{m}_K) \\ &\quad +2(f'_{m,K}(0) - \widehat{f}'_{\widehat{m}_K,K}(0))(\widehat{f}_K(0) - f(0)) - \mathrm{pen}_{2,K}(\widehat{m}_K) \\ &\leqslant \|f'_{m,K} - f'\|^2 + \mathrm{pen}_K(m) + \frac{1}{4}\|f'_{m,K} - \widehat{f}'_{\widehat{m}_K,K}\|^2 + 8\sup_{t\in S_{m\vee\widehat{m}_K}} \nu_n^2(t) - \mathrm{pen}_1(\widehat{m}_K) \\ &\quad +16(m\vee\widehat{m}_K)[\widehat{f}_K(0) - f(0)]^2 - \mathrm{pen}_{2,K}(\widehat{m}_K).\end{aligned}$$

To get the last line, we write that, for any $t \in S_m$,

$$|t(0)| = \sqrt{2}\left|\sum_{j=0}^{m-1} a_j(t)\right| \leqslant \sqrt{2m\sum_{j=0}^{m} a_j^2(t)} \leqslant \sqrt{2m}\|t\|,$$

and we use that $2xy \leqslant x^2/8 + 8y^2$ for all real x, y. We obtain

$$\begin{aligned}\frac{1}{2}\|\widehat{f}'_{\widehat{m}_K,K} - f'\|^2 &\leqslant \frac{3}{2}\|f'_{m,K} - f'\|^2 + \mathrm{pen}_K(m) + 16m(\widehat{f}_K(0) - f(0))^2 \\ &\quad +8\left(\sup_{t\in S_{m\vee\widehat{m}_K},\|t\|=1} \nu_n^2(t) - p_1(m\vee\widehat{m}_K)\right)_+ + 8p_1(m\vee\widehat{m}_K) - \mathrm{pen}_1(\widehat{m}_K) \\ &\quad +16\widehat{m}_K\left[(\widehat{f}_K(0) - f(0))^2 - c_2(\|f\|_\infty\vee 1)K\frac{\log(n)}{n}\right], \qquad (2.54)\end{aligned}$$

where

$$p_1(m) = \mathtt{b}(1 + 2\log(n))\|f\|_\infty\frac{m^2}{n}, \quad \mathtt{b} > 0.$$

The following Lemma can be proved using the Talagrand Inequality (see Section 2.7.2).

Lemma 2.6.5. *Under the assumptions of Theorem 2.3.1, and* $\mathtt{b} \geqslant 6$,

$$\sum_{m\in\mathcal{M}_n} \mathbb{E}\left[\sup_{t\in S_m,\|t\|=1} \nu_n^2(t) - p_1(m)\right]_+ \leqslant \frac{c}{n}.$$

It follows that

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in S_{m\vee\widehat{m}_K},\|t\|=1} \nu_n^2(t) - p_1(m\vee\widehat{m}_K)\right)_+ &\leqslant \sum_{m'\in\mathcal{M}_n} \mathbb{E}\left(\sup_{t\in S_{m'\vee m},\|t\|=1} \nu_n^2(t) - p_1(m\vee m')\right)_+ \\ &\leqslant \frac{c}{n}. \qquad (2.55)\end{aligned}$$

This implies that $8p_1(m\vee\widehat{m}_K) \leqslant \mathrm{pen}_1(m) + \mathrm{pen}_1(\widehat{m}_K)$ for c_1 –defined in (2.29)– large enough.
Moreover, let $\mathtt{a} > 0$ and

$$\Omega_K := \left\{\left|\frac{1}{n}\sum_{i=1}^{n}(Z_i^K - \mathbb{E}(Z_i^K))\right| \leqslant \sqrt{\mathtt{a}(\|f\|_\infty\vee 1)\frac{K\log(n)}{n}}\right\},$$

where $Z_i^K := \sum_{j=0}^{K-1} \ell_j(X_i)$. To apply the Bernstein Inequality (see Section 2.7.3), we compute $s^2 = \|f\|_\infty K$ and $b = \sqrt{2}K$ and note that $K\log(n)/n \leqslant 1$. Thus, we get that there exist constants c_0, c such that

$$\text{For } \mathtt{a} > c_0, \quad \mathbb{P}(\Omega_K^c) \leqslant \frac{c}{n^4}. \tag{2.56}$$

On Ω_K, it holds that

$$(\widehat{f}_K(0) - f_K(0))^2 = \left(\frac{1}{n}\sum_{i=1}^n (Z_i^K - \mathbb{E}(Z_i^K))\right)^2 \leqslant 2\mathtt{a}(\|f\|_\infty \vee 1)K\frac{\log(n)}{n}. \tag{2.57}$$

For any $K_n \leqslant [n/\log(n)]$ satisfying condition (2.27), we have

$$\begin{aligned} & \mathbb{E}\left\{\widehat{m}_{K_n}\left[(\widehat{f}_{K_n}(0) - f(0))^2 - c_2(\|f\|_\infty \vee 1)K_n\frac{\log(n)}{n}\right]\right\} \\ \leqslant\ & \mathbb{E}\left\{\widehat{m}_{K_n}\left[(\widehat{f}_{K_n}(0) - f_{K_n}(0))^2 - (c_2-2)(\|f\|_\infty \vee 1)K_n\frac{\log(n)}{n}\right]\right\} \end{aligned}$$

Now we note that $|\widehat{f}_K(x)| \leqslant 2K$ for all $x \in \mathbb{R}^+$ and any integer K and by using the definition of (2.57), provided that $c_2 > 2\mathtt{a} + 2$, we obtain

$$\begin{aligned} & \mathbb{E}\left\{\widehat{m}_{K_n}\left[(\widehat{f}_{K_n}(0) - f_{K_n}(0))^2 - (c_2-2)(\|f\|_\infty \vee 1)K_n\frac{\log(n)}{n}\right]\right\} \\ \leqslant\ & \mathbb{E}\left\{\widehat{m}_{K_n}\left[(\widehat{f}_{K_n}(0) - f_{K_n}(0))^2 - (c_2-2)(\|f\|_\infty \vee 1)K_n\frac{\log(n)}{n}\right]\mathbf{1}_{\Omega_{K_n}}\right\} \\ & +\mathbb{E}\left\{\widehat{m}_{K_n}\left[(\widehat{f}_{K_n}(0) - f_{K_n}(0))^2 - (c_2-2)(\|f\|_\infty \vee 1)K_n\frac{\log(n)}{n}\right]\mathbf{1}_{\Omega_{K_n}^c}\right\} \\ \lesssim\ & Cn^{5/2}\mathbb{P}(\Omega_{K_n}^c) \lesssim \frac{1}{n}, \end{aligned}$$

the term on Ω_{K_n} being less than or equal to 0. Plugging this and (2.55) into (2.54), we get

$$\mathbb{E}\left(\|\widehat{f}'_{\widehat{m}_K,K} - f'\|^2\right) \leqslant 3\|f'_{m,K} - f'\|^2 + 4\mathrm{pen}_K(m) + 32m(\widehat{f}_K(0) - f(0))^2 + \frac{c}{n}$$

which gives the result of Theorem 2.3.1. □

2.6.8 Proofs of auxiliary results

Proof of Lemma 2.2.1

In the Hermite case $\varphi_j = h_j$ and $f : \mathbb{R} \mapsto [0,\infty)$, allowing d successive integration by parts, it holds that

$$a_j(f^{(d)}) = \int_{\mathbb{R}} f^{(d)}(x)h_j(x)dx = \left[\sum_{k=0}^{d-1}(-1)^k f^{(d-1-k)}(x)h_j^{(k)}(x)\right]_{-\infty}^{+\infty} + (-1)^d\int_{\mathbb{R}} h_j^{(d)}(x)f(x)dx. \tag{2.58}$$

By definition for all $j \geqslant 0$, $h_j(x) = c_j H_j(x) e^{-\frac{x^2}{2}}$ where H_j is a polynomial. Then, its k-th derivative, $0 \leqslant k \leqslant d-1$, is a polynomial multiplied by $e^{-x^2/2}$ and $\lim_{|x|\to+\infty} h_j^{(k)}(x) = 0$. This together with (**A**2), gives that the bracket in (2.58) is null and the result follows.

Similarly in the Laguerre case, (2.58) holds integrating on $[0,\infty)$ instead of $\mathbb{R}$ and replacing h_j by ℓ_j. The term in the bracket is null at 0 from (**A1**). It is also null at infinity using (**A**2) together with the fact that ℓ_j are polynomials multiplied by e^{-x} leading similarly to $\lim_{x\to\infty} f^{(d-1-k)}(x)\ell_j^{(k)}(x) = 0$, $0 \leqslant k \leqslant d-1$, $j \geqslant 0$. The result follows.

Proof of Lemma 2.2.2

We control the quantity

$$\sum_{j\geqslant 0} j^{s-d}\langle f^{(d)}, h_j\rangle^2 = \sum_{j=0}^{d-1} j^{s-d}\langle f^{(d)}, h_j\rangle^2 + \sum_{j\geqslant d} j^{s-d}\langle f^{(d)}, h_j\rangle^2. \tag{2.59}$$

The first term is a constant which depending on d. For the second term using Lemma 2.6.2, we obtain

$$\begin{aligned}
\sum_{j\geqslant d} j^{s-d}\langle f^{(d)}, h_j\rangle^2 &= \sum_{j\geqslant d} j^{s-d}\left(\sum_{k=-d}^{d} b_{k,j}^{(d)} \int h_{j+k}(x)f(x)dx\right)^2 \\
&\leqslant C_d \sum_{j\geqslant d} j^s \sum_{k=-d}^{d}\left(\int h_{j+k}(x)f(x)dx\right)^2 = C_d \sum_{k=-d}^{d}\sum_{j\geqslant d} j^s\langle h_{j+k}, f\rangle^2 \\
&= C_d \sum_{k=-d}^{d}\left(\sum_{j\geqslant d+k} |j-k|^s\langle h_j, f\rangle^2\right) \leqslant C_d \sum_{k=-d}^{d}\left(\sum_{j\geqslant 0} 2^s j^s\langle h_j, f\rangle^2\right) = (2d+1)2^s D C_d.
\end{aligned}$$

Inserting this in (2.59), we obtain the announced result.

Proof of Lemma 2.2.3

We establish the result for $d = 1$, the general case is an immediate consequence. It follows from the definition of $\widetilde{W}_L^s(D)$ that $(\theta')^{(j)}$, $0 \leqslant j \leqslant s-1$ are in $C([0,\infty))$. Moreover, it holds that $x \mapsto x^{k/2}(\theta')^{(j)}(x) \in \mathbb{L}^2(\mathbb{R}^+)$ for all $0 \leqslant j < k \leqslant s-1$. The case $k = j$ is obtained using that $\theta^{(j)}$ is continuous on $C([0,\infty))$ and that $x \mapsto x^{(j+1)/2}(\theta')^{(j)}(x) \in \mathbb{L}^2(\mathbb{R}^+)$. It follows that

$$\begin{aligned}
|||\theta'|||_s^2 &= \sum_{j=0}^{s-1}\left\|x^{j/2}\sum_{k=0}^{j}\binom{j}{k}(\theta')^{(k)}\right\|^2 \leqslant 2\sum_{j=0}^{s-1}\left\|x^{j/2}\sum_{k=0}^{j-1}\binom{j}{k}(\theta')^{(k)}\right\|^2 + 2\sum_{j=0}^{s-1}\left\|x^{j/2}(\theta')^{(j)}\right\|^2 \\
&\leqslant C + 2\sum_{j=0}^{s-1}\|x^{(j+1)/2}(\theta')^{(j)}(x)\|^2 < \infty,
\end{aligned}$$

where C depends on D. Finally, using the equivalence of the norms $|.|_s$ and $|||.|||_s$, the value of D' follows from the latter inequality.

Proof of Lemma 2.6.1.

Consider the decomposition

$$\int_0^{+\infty} x^{-k}(\ell_{j-k,(k)}(x/2))^2 f(x/2)dx = \sum_{i=1}^{6} I_i,$$

where for $\nu = 4j - 2k + 2$, $j \geqslant k$, we used the decomposition $(0,\infty) = (0, \frac{1}{\nu}] \cup (\frac{1}{\nu}, \frac{\nu}{2}] \cup (\frac{\nu}{2}, \nu - \nu^{1/3}] \cup (\nu - \nu^{1/3}, \nu + \nu^{/13}] \cup (\nu + \nu^{1/3}, 3\nu/2] \cup (3\nu/2, \infty)$. Using Askey and Wainger (1965) (see Appendix 2.7.1) and straightforward inequalities give

$$I_1 \lesssim \int_0^{\frac{1}{\nu}} x^{-k}(x\nu)^k f(x/2)dx \leqslant \int_0^{\frac{1}{\nu}} x^{-k}(x\nu)^{-1/2} f(x/2)dx \lesssim \nu^{-1/2}\mathbb{E}[X^{-k-1/2}],$$

$$I_2 \lesssim \int_{1/\nu}^{\frac{\nu}{2}} x^{-k}((x\nu)^{-1/4})^2 f(x/2)dx = \nu^{-1/2}\int_{1/\nu}^{\frac{\nu}{2}} x^{-k-1/2} f(x/2)dx \leqslant \nu^{-1/2}\mathbb{E}[X^{-k-1/2}],$$

$$I_3 \lesssim \int_{\frac{\nu}{2}}^{\nu-\nu^{1/3}} x^{-k}(\nu^{-1/4}(\nu - x)^{-1/4})^2 f(x/2)dx = \nu^{-1/2}\int_{\frac{\nu}{2}}^{\nu-\nu^{1/3}} x^{-k}(\nu - x)^{-1/2} f(x/2)dx \lesssim \nu^{-1/2},$$

$$I_4 \lesssim \int_{\nu-\nu^{1/3}}^{\nu+\nu^{1/3}} x^{-k}(\nu^{-1/3})^2 f(x/2)dx \leqslant \nu^{-2/3}\int_{\frac{\nu}{2}}^{\nu+\nu^{1/3}} x^{-k} f(x/2)dx \lesssim \nu^{-1/2}\nu^{-k} \leqslant \nu^{-1/2},$$

$$I_5 \lesssim \int_{\nu+\nu^{1/3}}^{3\nu/2} x^{-k}\nu^{-1/2}(x - \nu)^{-1/2} e^{-2\gamma_1\nu^{-1/2}(x-\nu)^{3/2}} f(x/2)dx \lesssim \nu^{-1/2}\nu^{-1/6}\nu^{-k}\int f(x/2)dx \lesssim \nu^{-1/2},$$

$$I_6 \lesssim \int_{3\nu/2}^{+\infty} x^{-k} e^{-2\gamma_2 x} f(x/2)dx \lesssim e^{-3\gamma_2\nu/2} = \mathcal{O}(\nu^{-1/2}).$$

Gathering these inequalities give the announced result.

Proof of Lemma 2.6.2.

The result is obtained by induction on d. If $d = 1$, h_j' is given by (2.5), with $b_{-1,j-1}^{(1)} = j^{1/2}/\sqrt{2}$, $b_{0,j} = 0$ and $b_{1,j}^{(1)} = (j+1)^{1/2}/\sqrt{2}$, $\forall j \geqslant 1$. Thus, it holds $b_{k,j}^{(1)} = \mathcal{O}(j^{1/2})$ and (2.37) is satisfied for $d = 1$. Let $\mathrm{P}(d)$ the proposition given by Equation (2.37) and assume $\mathrm{P}(d)$ holds and we establish $\mathrm{P}(d+1)$. It holds using successively $\mathrm{P}(d)$ and (2.5) that

$$\begin{aligned} h_j^{(d+1)}(x) &= \sum_{k=-d}^{d} b_{k,j}^{(d)}\left[\frac{\sqrt{j+k}}{\sqrt{2}}h_{j+k-1} - \frac{\sqrt{j+k+1}}{\sqrt{2}}h_{j+k+1}\right] \\ &= \sum_{k'=-d-1}^{d-1} b_{k'+1,j}^{(d)}\frac{\sqrt{j+k'+1}}{\sqrt{2}}h_{j+k'} - \sum_{k'=-d+1}^{d+1} b_{k'-1,j}^{(d)}\frac{\sqrt{j+k'}}{\sqrt{2}}h_{j+k'} := \sum_{k=-d-1}^{d+1} b_{k,j}^{(d+1)}h_{j+k'}, \end{aligned}$$

where $\quad b_{k,j}^{(d)} = \mathcal{O}(j^{d/2})$, $\forall j \geqslant d \geqslant |k| \quad$ and $\quad b_{k,j}^{(d+1)} = b_{k+1,j}^{(d)}\frac{\sqrt{j+k+1}}{\sqrt{2}}\mathbf{1}_{|k|\leqslant d-1} - b_{k-1,j}^{(d)}\frac{\sqrt{j+k}}{\sqrt{2}}\mathbf{1}_{|k|\leqslant d+1}.$

It follows that $|b_{k,j}^{(d+1)}| \leqslant 2\sqrt{(j+d+1)/2}j^{\frac{d}{2}} \leqslant C_d j^{\frac{d+1}{2}}$, $|k| \leqslant d \leqslant j$, which completes the proof.

2.6.9 Proof of Lemma 2.6.3.

Proof of part (i)**.**

By construction, f_0 is positive and $\forall\theta\in\{0,1\}^K$, $\int f_\theta(x)dx=\int f_0(x)dx=1$. It remains to show that f_θ is nonnegative. The supports of $\big(\psi((.-1)(K+1)-k)\big)_{0\leqslant k\leqslant K-1}$ are disjoint and are in $[1,2]$, then $f_\theta(x)\geqslant 0$ for all $x\in\mathbb{R}\backslash[1,2]$. Now, for all x in $[1,2]$, there exists k_0 such that

$$f_\theta(x)=\frac{x}{2}+\delta K^{-\gamma-d}\theta_{k_0+1}\psi\big((x-1)(K+1)-k_0\big)\geqslant\frac{1}{2}-\delta\|\psi\|_\infty K^{-\gamma-d},$$

which is nonnegative if $\delta\leqslant\|\psi\|_\infty^{-1}/2$. Now, let us show that f_0 and f_θ belong to $W^s(D)$.

The Laguerre case. We use the equivalent norm $|||.|||_s$ of $|.|_s$ (see (2.13)) and start with f_0. As f_0 is s-th differentiable, we have

$$|||f_0|||_s^2=\sum_{j=0}^s\int_0^3\left(x^{j/2}\sum_{k=0}^j\binom{j}{k}f_0^{(k)}(x)\right)^2dx\leqslant\sum_{j=0}^s 2^j\sum_{k=0}^j\binom{j}{k}\int_0^3(x^{j/2}f_0^{(k)}(x))^2dx.$$

As $\int_0^3(x^{j/2}f_0^{(k)}(x))^2dx\leqslant c(s)<+\infty$, $0\leqslant k\leqslant j\leqslant s$, it follows $|f|_s^2\leqslant 3D/4$, D depends on s. For f_θ, we have

$$\begin{aligned}|||f_\theta-f_0|||_s^2=&\delta^2K^{-2\gamma-2d}\sum_{j=0}^s\int_1^2\left(\sum_{l=0}^j\binom{j}{l}\sum_{k=0}^{K-1}x^{j/2}\theta_{k+1}(K+1)^l\psi^{(l)}\big((x-1)(K+1)-k\big)\right)^2dx\\ \leqslant&\delta^2K^{-2\gamma-2d}\sum_{j=0}^s\sum_{l=0}^j 2^j\binom{j}{l}\int_1^2\left(x^{j/2}\sum_{k=0}^{K-1}\theta_{k+1}(K+1)^l\psi^{(l)}\big((x-1)(K+1)-k\big)\right)^2dx.\end{aligned}$$

Using that $\psi^{(l)}\big((x-1)(K+1)-k\big),\psi^{(l)}\big((x-1)(K+1)-k'\big)$ have disjoint supports for $k\neq k'$ and that $\psi^{(l)}$ are bounded by c, we get after the change of variable $y=(x-1)(K+1)-k$,

$$|||f_\theta-f_0|||_s^2\leqslant\delta^2 2^{3s}c^2K^{-2\gamma-2d}\sum_{j=0}^s\sum_{k=0}^{K-1}(K+1)^{2j-1}\leqslant C(s)\delta^2K^{-2\gamma-2d+2s}.$$

For $\gamma\geqslant s-d$ and δ small enough, it holds $|f_\theta-f_0|_s\leqslant D/4$ and therefore $|f_\theta|_s\leqslant|f_\theta-f_0|_s+|f_0|_s\leqslant D$.

The Hermite case. The usual Sobolev space W^s, if s is integer, is defined by

$$W^s=\{f\in\mathbb{L}^2(\mathbb{R}),f\text{ admits derivatives up to order }s,\text{such that }|||f|||_{s,sob}=\sum_{j=0}^s\|f^{(j)}\|^2<+\infty\}.$$

It is proved in Bongioanni and Torrea (2006) that : if $f\in W^s$ has compact support, then f belongs to W_H^s. By construction f_0 and f_θ have a compact support and as they admit derivatives up to order s, they belong to W^s. It follows that f_0 and f_θ belong W_H^s. This completes the proof of (i).

Proof of part (ii).

As for $k \neq k'$, $\psi\big((.-1)(K+1)-k\big), \psi\big((.-1)(K+1)-k'\big)$ have disjoint supports, we have, $\forall \theta^{(j)}, \theta^{(l)} \in \{0,1\}^K$,

$$\begin{aligned}\|f^{(d)}_{\theta^{(j)}} - f^{(d)}_{\theta^{(l)}}\|^2 =& \delta^2 \sum_{k=0}^{K-1} (\theta^{(j)}_{k+1} - \theta^{(l)}_{k+1})^2 K^{-2\gamma-2d}(K+1)^{2d} \int_1^2 \psi^{(d)}\big((x-1)(K+1)-k\big)^2 dx \\ &\geqslant \delta^2 \|\psi^{(d)}\|^2 K^{-2\gamma-1} \rho(\theta^{(j)}, \theta^{(l)}),\end{aligned}$$

where $\rho(\theta^{(j)}, \theta^{(l)}) = \sum_{k=1}^K \mathbf{1}_{\theta^{(j)}_k \neq \theta^{(l)}_k}$ is the Hamming distance. By Lemma 2.7 in Tsybakov (2009), for $K \geqslant 8$, there exist $\{\theta^{(0)}, \dots, \theta^{(M)}\}$ in $\{0,1\}^K$ such that

$$\rho(\theta^{(j)}, \theta^{(l)}) \geqslant \frac{K}{8}, \ \forall \quad 0 \leqslant j < l \leqslant M \text{ and } M \geqslant 2^{\frac{K}{8}}.$$

Thus, it holds, $\forall \theta^{(j)}, \theta^{(l)} \in \{0,1\}^K$, $\|f^{(d)}_{\theta^{(j)}} - f^{(d)}_{\theta^{(l)}}\|^2 \geqslant \delta^2/8 \|\psi^{(d)}\|^2 K^{-2\gamma}$, which gives (ii) if we set $C = \|\psi^{(d)}\|^2/8$.

Proof of part (iii).

For M integer and $(\theta^{(j)})_{1\leqslant j\leqslant M}$ in $(\{0,1\}^K)^M$, we have

$$\sum_{j=1}^M \chi^2\left(f_{\theta^{(j)}}{}^{\otimes n}, f_0{}^{\otimes n}\right) = \sum_{j=1}^M \left(\left(1+\chi^2(f_{\theta^{(j)}}, f_0)\right)^n - 1\right) = \sum_{j=1}^M \left(e^{n\log(1+\chi^2(f_{\theta^{(j)}}, f_0))} - 1\right). \tag{2.60}$$

Since $f_0 \geqslant c > 0$ on $[1,2]$, it holds for any $\theta \in \{0,1\}^K$,

$$\begin{aligned}\chi^2(f_\theta, f_0) = \int_1^2 \frac{(f_\theta(x) - f_0(x))^2}{f_0(x)} dx &\leqslant \frac{\delta^2}{c} K^{-2\gamma-2d} \sum_{k=0}^{K-1} \int_1^2 \left(\psi\big((x-1)(K+1)-k\big)\right)^2 dx \\ &\leqslant \frac{\delta^2}{c} K^{-2\gamma-2d} \|\psi\|^2 \leqslant \frac{8\delta^2}{c\log 2} \log(M) K^{-2\gamma-2d-1},\end{aligned}$$

where we used that $M \geqslant 2^{\frac{K}{8}}$. Consequently, using in (2.60) that $\log(1+x) \leqslant x$, for any $x \geqslant 0$, and the latter inequality, give

$$\frac{1}{M}\sum_{j=1}^M \chi^2\left(f_{\theta^{(j)}}{}^{\otimes n}, f_0{}^{\otimes n}\right) \leqslant e^{n\frac{8\delta^2}{c\log 2}\log(M)K^{-2\gamma-2d-1}} - 1.$$

For δ well chosen and $K = n^{1/(2\gamma+2d+1)}$, comes the result.

Proof of Lemma 2.6.4

Proof of part (i) First, it holds that

$$\mathbb{E}\left[\left(\sup_{t\in S_m + S_{\widehat{m}}, \|t\|=1} |\nu_{n,d}(t)|^2 - p(m,\widehat{m}_n)\right)_+\right] \leqslant \sum_{m'\in\mathcal{M}_{n,d}} \mathbb{E}\left[\left(\sup_{t\in S_m+S_{m'}, \|t\|=1} |\nu_{n,d}(t)|^2 - p(m,m')\right)_+\right], \tag{2.61}$$

which we bound applying a Talagrand Inequality (see Section 2.7.2). Following notations of Section 2.7.2, we have three terms H^2, v and M_1 to compute. Let us denote by $m^* = m \vee m'$, for $t \in S_m + S_{m'}$, $\|t\| = 1$, it holds

$$\|t\|^2 = \| \sum_{j=0}^{m^*-1} a_j \varphi_j \|^2 = \sum_{j=0}^{m^*-1} a_j^2 = 1.$$

Computing H^2. By the linearity of $\nu_{n,d}$ and the Cauchy Schwarz inequality, we have

$$\nu_{n,d}(t)^2 = \left(\sum_{j=0}^{m^*-1} a_j \nu_{n,d}(\varphi_j) \right)^2 \leqslant \sum_{j=0}^{m^*-1} a_j^2 \sum_{j=0}^{m^*-1} \nu_{n,d}^2(\varphi_j) = \sum_{j=0}^{m^*-1} \nu_{n,d}^2(\varphi_j).$$

One can check that the latter is an equality for $a_j = \nu_{n,d}(\varphi_j)$. Therefore, taking expectation, it follows

$$\begin{aligned} \mathbb{E}\left[\sup_{t \in S_m^*, \|t\|=1} \nu_{n,d}^2(t) \right] &= \sum_{j=0}^{m^*-1} \mathrm{Var}(\nu_{n,d}(\varphi_j)) = \frac{1}{n} \sum_{j=0}^{m^*-1} \mathrm{Var}(\varphi_j^{(d)}(X_1)) \\ &\leqslant \frac{1}{n} \sum_{j=0}^{m^*-1} \mathbb{E}\left[\varphi_j^{(d)}(X_1)^2 \right] = \frac{V_{m^*,d}}{n} =: H^2. \end{aligned}$$

Computing v. It holds for $t \in S_m + S_{m'}$, $\|t\| = 1$,

$$\begin{aligned} \mathrm{Var}\left((-1)^d t^{(d)}(X_1) \right) &\leqslant \int t^{(d)}(x)^2 f(x) dx = \int \left(\sum_{j=0}^{m^*-1} a_j \varphi_j^{(d)}(x) \right)^2 f(x) dx \qquad (2.62) \\ &\leqslant 2 \int \left(\sum_{j=0}^{d-1} a_j \varphi_j^{(d)}(x) \right)^2 f(x) dx + 2 \int \left(\sum_{j=d}^{m^*-1} a_j \varphi_j^{(d)}(x) \right)^2 f(x) dx. \end{aligned}$$

The first term of the previous inequality is a constant depending only on d. For the second term, we consider separately the Laguerre and Hermite cases.

The Laguerre Case ($\varphi_j = \ell_j$). Using (2.36) and the Cauchy Schwarz inequality, it holds that

$$\begin{aligned} \int \left(\sum_{j=d}^{m^*-1} a_j \ell_j^{(d)}(x) \right)^2 f(x) dx &\leqslant 3^d \sum_{k=0}^{d} \binom{d}{k} \int \left(\sum_{j=d}^{m^*-1} a_j \left(\frac{j!}{(j-k)!} \right)^{\frac{1}{2}} x^{-\frac{k}{2}} \ell_{j-k,(k)}(x) \right)^2 f(x) dx \\ &\leqslant 3^d \sum_{k=0}^{d} \binom{d}{k} \sup_{x \in \mathbb{R}^+} \frac{f(x)}{x^k} \sum_{j=d}^{m^*-1} a_j^2 \frac{j!}{(j-k)!} \leqslant C(d) (m^*)^d, \end{aligned} \qquad (2.63)$$

where we used the orthonormality of $(\ell_{j,(k)})_{j \geqslant 0}$ and where $C(d)$ is a constant depending only on d and $\sup_{x \in \mathbb{R}^+} \frac{f(x)}{x^k}$.

The Hermite case ($\varphi_j = h_j$). Similarly, using Lemma 2.6.2 and the orthonormality of h_j,

it follows

$$\int \left(\sum_{j=d}^{m^*-1} a_j h_j^{(d)}(x) \right)^2 f(x)dx \leqslant (2d+1) \sum_{k=-d}^{d} \int \left(\sum_{j=d}^{m^*-1} a_j b_{k,j} h_{j+k}(x) \right)^2 f(x)dx$$
$$\leqslant C(d)\|f\|_\infty (m^*)^d. \tag{2.64}$$

Plugging (2.63) or (2.64) in (2.62), we set in the two cases $v := c_1(m^*)^d$ where c_1 depends on d and either on $\sup_{x\in\mathbb{R}^+} \frac{f(x)}{x^k}$ (Laguerre case) or $\|f\|_\infty$ (Hermite case).

Computing M_1. The Cauchy Schwarz Inequality and $\|t\| = 1$ give

$$\|(-1)^d t^{(d)}\|_\infty = \| \sum_{j=0}^{m^*-1} (-1)^d a_j \varphi_j^{(d)} \|_\infty \leqslant \sup_{x\in\mathbb{R}} \sqrt{\sum_{j=0}^{m^*-1} \varphi_j^{(d)}(x)^2}. \tag{2.65}$$

The Laguerre case. We use the following Lemma whose proof is a consequence of (2.2) and an induction on d.

Lemma 2.6.6. *For ℓ_j given in* (2.1)*, the d-th derivative of ℓ_j is such that $\|\ell_j^{(d)}\|_\infty \leqslant C_d(j+1)^d$, $\forall j \geqslant 0$ and where C_d is a positive constant depending on d.*

Using Lemma 2.6.6, we obtain

$$\sum_{j=0}^{m^*-1} \ell_j^{(d)}(x)^2 \leqslant C_d^2 (m^*)^{2d+1}. \tag{2.66}$$

The Hermite case. The d first terms in the sum in (2.65) can be bounded by a constant depending only on d. For the remaining terms, Lemma 2.6.2 and $\|h_j\|_\infty \leqslant \phi_0$ (see (4.10)) give

$$\sum_{j=d}^{m^*-1} [h_j^{(d)}(x)]^2 \leqslant C_d^2 \phi_0^2 \sum_{k=-d}^{d} \sum_{j=d}^{m^*-1} j^d \leqslant C(m^*)^{d+1}, \tag{2.67}$$

where C is a positive constant depending on d and ϕ_0.

Injecting either (2.66) or (2.67) in (2.65), we set $M_1 = \mathcal{O}(m^{d+\frac{1}{2}})$ in the Laguerre case or $M_1 = \mathcal{O}(m^{\frac{d}{2}+\frac{1}{2}})$ in the Hermite case.

Now, we apply the Talagrand Inequality see Appendix 2.7.2 with $\varepsilon = 1/2$, it follows

$$\mathbb{E}\left[\left(\sup_{t\in S_m + S_{m'}, \|t\|=1} |\nu_{n,d}(t)|^2 - 4H^2 \right)_+ \right] \leqslant \frac{C_1}{n} \left(v \exp\left(-C_2 \frac{nH^2}{v} \right) + C_3 \frac{M_1^2}{n} \exp\left(-C_4 \frac{nH}{M_1} \right) \right)$$
$$:= \frac{C_1}{n} \left(U_d(m^*) + V_d(m^*) \right).$$

The Laguerre Case. We have

$$U_d(m^*) = c_1(m^*)^d \exp\left(-C_2 \frac{V_{m^*,d}}{c_1(m^*)^d} \right) \text{ and } V_d(m^*) = C_3 c_2 \frac{(m^*)^{2d+1}}{n} \exp\left(-C_4 \sqrt{n} \frac{\sqrt{V_{m^*,d}}}{c_2 (m^*)^{d+\frac{1}{2}}} \right).$$

From (2.41) and the value of $m_n(d)$, we obtain

$$U_d(m^*) \leqslant c_1 (m^*)^d \exp(-C_2' {m^*}^{\frac{1}{2}}) \quad \text{and} \quad V_d(m^*) \leqslant C_3 c_2 (m^*)^{d+\frac{1}{2}} \exp(-C_4' \sqrt{n} (m^*)^{-\frac{d}{2}-\frac{1}{4}}).$$

Using the value $m_n(d)$, it holds $(m^*)^{d+1/2} \leqslant n/\log^3(n)$, which implies (recall $m^* = m \vee m'$)

$$\sum_{m' \in \mathcal{M}_{n,d}} V_d(m^*) \leqslant C \sum_{m' \in \mathcal{M}_{n,d}} (m^*)^{d+\frac{1}{2}} \exp\left(-C_4 \log^2(n)\right) \leqslant \Sigma_{d,2},$$

where $\Sigma_{d,2}$ is a constant depending only on d. Next, it follows

$$\sum_{m'=1}^{n} U_d(m^*) = \sum_{m'=1}^{m} U_d(m^*) + \sum_{m'=m}^{n} U_d(m^*) = c_1 m^{d+1} \exp(-C_2' m^{\frac{1}{2}}) + \sum_{m'=m}^{n} c_1 (m')^d \exp(-C_2' {m'}^{\frac{1}{2}}).$$

The function $m \mapsto m^{d+1} \exp(-C_2' m^{\frac{1}{2}})$ is bounded and the sum is finite on m', it holds

$$C_1 \sum_{m'=1}^{n} U_d(m^*) \leqslant \Sigma_{d,1}, \text{ where } \Sigma_{d,1} \text{ depends only on } d.$$

The Hermite case. Only the second term $V_d(m^*)$ changes. Here, it is given by

$$\begin{aligned} V_d(m^*) &= C_3 c_2 \frac{(m^*)^{d+1}}{n} \exp\left(-C_3 \sqrt{n} \frac{\sqrt{V_{m^*,d}}}{c_2 (m^*)^{\frac{d}{2}+\frac{1}{2}}}\right) \leqslant C_3 c_2 (m^*)^{1/2} \exp(-C_4' \sqrt{n} (m^*)^{-\frac{1}{4}}) \\ &\leqslant C_3 c_2 (m^*)^{1/2} \exp(-C_4' (m^*)^{\frac{d}{2}}), \end{aligned}$$

where we used (2.46) and the value of $m_n(d)$. We derive that $\sum_{m' \in \mathcal{M}_{n,d}} V_d(m^*) \leqslant \Sigma_{d,2}$.

Gathering all terms, it follows

$$\mathbb{E}\left[\left(\sup_{t \in S_m + S_{m'}, \|t\|=1} |\nu_{n,d}(t)|^2 - 4H^2\right)_+\right] \leqslant \frac{\Sigma}{n}, \text{ where } \Sigma = \Sigma_{d,1} + \Sigma_{d,2}$$

Plugging this in (2.61) gives the announced result.

Proof of part (ii). We use the Bernstein Inequality (see Appendix 2.7.3) to prove the result. Define

$$Z_i^{(m)} = \sum_{j=0}^{m-1} (\varphi_j^{(d)}(X_i))^2, \quad \text{then,} \quad \widehat{V}_{m,d} = \frac{1}{n} \sum_{i=1}^{n} Z_i^{(m)}$$

We select s^2 and b such that $\mathrm{Var}(Z_i^{(m)}) \leqslant s^2$ and $|Z_i^{(m)}| \leqslant b$. By the computation of M_1 (see Proof of part (i)), we set $b := C^* m^\alpha$, with $\alpha = 2d+1$ (Laguerre case) or $\alpha = d+1$ (Hermite case), where C^* depends on d. For s^2, using that $\mathrm{Var}(Z_i^{(m)}) \leqslant \mathbb{E}[(Z_i^{(m)})^2] \leqslant b \sum_{j=0}^{m-1} \mathbb{E}\left[(\varphi_j^{(d)}(X_i))^2\right] = C^* m^\alpha V_{m,d} =: s^2$. Applying the Bernstein Inequality, we have for $S_n = n(\widehat{V}_{m,d} - V_{m,d})$

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geqslant \sqrt{\frac{2xC^* m^\alpha V_{m,d}}{n}} + \frac{C^* m^\alpha x}{3n}\right) \leqslant 2e^{-x}, \quad \forall x > 0. \tag{2.68}$$

Choose $x = 2\log(n)$ and define the set

$$\Omega := \left\{ m \in \mathcal{M}_{n,d},\ \frac{1}{n}|S_n| \leqslant 2\sqrt{\frac{C^* m^\alpha \log(n) V_{m,d}}{n}} + \frac{2C^* m^\alpha \log(n)}{3n} \right\}.$$

Consider the decomposition,

$$\begin{aligned}\mathbb{E}\left[(\mathrm{pen}_d(\widehat{m}_n) - \widehat{\mathrm{pen}}_d(\widehat{m}_n))_+\right] \leqslant \mathbb{E}\left[(\mathrm{pen}_d(\widehat{m}_n) - \widehat{\mathrm{pen}}_d(\widehat{m}_n))_+ \mathbf{1}_\Omega\right] \\ + \mathbb{E}\left[(\mathrm{pen}_d(\widehat{m}_n) - \widehat{\mathrm{pen}}_d(\widehat{m}_n))_+ \mathbf{1}_{\Omega^c}\right]\end{aligned}$$

Using $2xy \leqslant x^2 + y^2$, we have on Ω

$$|\widehat{V}_{\widehat{m},d} - V_{\widehat{m},d}| \leqslant \frac{V_{\widehat{m},d}}{2} + \frac{2C^* \widehat{m}^\alpha \log(n)}{n} + \frac{2C^* \widehat{m}^\alpha \log(n)}{3n} = \frac{V_{\widehat{m},d}}{2} + \frac{8}{3}\frac{C^* \widehat{m}^\alpha \log(n)}{n}.$$

The constraint on m_n gives $\widehat{m}^{d+1/2} \leqslant Cn/(\log(n))^2$ together with (2.41) giving $V_{\widehat{m},d} \geqslant c^* \widehat{m}^{d+1/2}$ give for $\alpha = 2d+1$ (Laguerre case) that $\frac{8C^*}{3}\frac{\widehat{m}^\alpha \log(n)}{n} \leqslant \frac{8CC^*}{3c^*}\frac{V_{\widehat{m},d}}{\log(n)} \leqslant \frac{V_{\widehat{m},d}}{4}$, for n large enough and

$$\mathbb{E}\left[(\mathrm{pen}_d(\widehat{m}_n) - \widehat{\mathrm{pen}}_d(\widehat{m}_n))_+ \mathbf{1}_\Omega\right] \leqslant \frac{3}{4}\mathbb{E}[\mathrm{pen}_d(\widehat{m}_n)]. \tag{2.69}$$

In the Hermite case ($\alpha = d+1$) computations are similar as $\widehat{m}^{d+1} \leqslant \widehat{m}^{2d+1}$. For the control on Ω^c, we write, using (2.68),

$$\mathbb{E}\left[(\mathrm{pen}_d(\widehat{m}_n) - \widehat{\mathrm{pen}}_d(\widehat{m}_n))_+ \mathbf{1}_{\Omega^c}\right] \leqslant 2\kappa \mathbb{P}(\Omega^c) \leqslant 2\kappa \sum_{m \in \mathcal{M}_{n,d}} 2e^{-2\log(n)} := \frac{\Sigma_2}{n}. \tag{2.70}$$

Gathering (2.69) and (2.70), we get the desired result.

2.7 Some inequalities

2.7.1 Asymptotic Askey and Wainger formula

From Askey and Wainger (1965), we have for $\nu = 4k + 2\delta + 2$, and k large enough

$$|\ell_{k,(\delta)}(x/2)| \leqslant C \begin{cases} a) & (x\nu)^{\delta/2} & \text{if } 0 \leqslant x \leqslant 1/\nu \\ b) & (x\nu)^{-1/4} & \text{if } 1/\nu \leqslant x \leqslant \nu/2 \\ c) & \nu^{-1/4}(\nu - x)^{-1/4} & \text{if } \nu/2 \leqslant x \leqslant \nu - \nu^{1/3} \\ d) & \nu^{-1/3} & \text{if } \nu - \nu^{1/3} \leqslant x \leqslant \nu + \nu^{1/3} \\ e) & \nu^{-1/4}(x - \nu)^{-1/4} e^{-\gamma_1 \nu^{-1/2}(x-\nu)^{3/2}} & \text{if } \nu + \nu^{1/3} \leqslant x \leqslant 3\nu/2 \\ f) & e^{-\gamma_2 x} & \text{if } x \geqslant 3\nu/2 \end{cases}$$

where γ_1 and γ_2 are positive and fixed constants.

2.7.2 A Talagrand Inequality.

The Talagrand inequalities have been proven in Talagrand (1996) and reworked by Ledoux (1997). This version is given in Klein and Rio (2005). Let $(X_i)_{1\leqslant i\leqslant n}$ be independent real random variables and

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^{n}(t(X_i) - \mathbb{E}[t(X_i)]),$$

for t in $\mathcal{F}$ a class of measurable functions. If there exist M_1, H and v such that :

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

then, for $\varepsilon > 0$,

$$\mathbb{E}\left[\left\{\sup_{t\in\mathcal{F}}|\nu_n^2(t)| - 2(1+2\varepsilon)H^2\right)_+\right] \leqslant \frac{4}{K_1}\left\{\frac{v}{n}e^{-K_1\varepsilon\frac{nH^2}{v}} + \frac{49M_1^2}{K_1C^2(\varepsilon)n^2}e^{-K_1'C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{M_1}}\right\},$$

where $C(\varepsilon) = (\sqrt{1+\varepsilon} - 1) \wedge 1$, $K_1 = 1/6$ and K_1' a universal constant.

2.7.3 Bernstein Inequality (Massart (2007)).

Let $X_1, \dots X_n$, n independent real random variables. Assume there exist two constants s^2 and b, such that $\mathrm{Var}(X_i) \leqslant s^2$ and $|X_i| \leqslant b$. Then, for all x positive, we have

$$\mathbb{P}\left(|S_n| \geqslant \sqrt{2ns^2x} + \frac{bx}{3}\right) \leqslant 2e^{-x}, \text{ with } S_n = \sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]).$$

Deuxième partie

Déconvolution en base d'Hermite

Chapitre 3

Hermite density deconvolution

Ce chapitre est une version modifiée de l'article Sacko, O. (2020). Hermite density deconvolution. *ALEA Lat. Am. J. Probab. Math. Stat.*, 17(1) :419–443.

Résumé. Considérons le modèle à bruit additif : $Z = X + \varepsilon$, où X et ε sont indépendantes. Nous construisons un nouvel estimateur de la densité de X à partir d'observations de Z, fondé sur une méthode de projection en base d'Hermite sur $\mathbb{R}$. Nous étudions le risque quadratique intégré de notre estimateur. Nous prouvons qu'il est consistant et atteint les vitesses classiques dans ce contexte. Nous montrons aussi que les résultats peuvent s'étendre au cas de variables dépendantes. Nous proposons une procédure d'estimation adaptative, c'est-à-dire une méthode pour sélectionner un modèle pertinent. L'estimateur résultant réalise automatiquement un compromis biais-variance. Des simulations numériques sont proposées et des comparaisons avec la méthode proposée dans Comte and Lacour (2011) et le cas direct ($\varepsilon = 0$ presque sûrement) sont effectuées.

Mots-clés. Déconvolution, base d'Hermite, sélection de modèle, estimateur non paramétrique.

Abstract. We consider the additive model : $Z = X + \varepsilon$, where X and ε are independent. We construct a new estimator of the density of X from n observations of Z. We propose a projection method which exploits the specific properties of the Hermite basis. We study the quality of the resulting estimator by proving a bound on the integrated quadratic risk. We show also that the results can be easily extended to dependent variables. We then propose an adaptive estimation procedure, that is a method of selecting a relevant model. The resulting estimator realizes automatically a bias-variance compromise. We check that our estimator reaches the classical convergence speeds of deconvolution. Numerical simulations are proposed and comparisons with the results of the method proposed in Comte and Lacour (2011) and with the direct case ($\varepsilon = 0$ almost surely) are performed.

Keywords. Deconvolution, Hermite basis, nonparametric estimator, model selection.

Sommaire

3.1 Introduction

Consider the additive noise model :

$$Z_k = X_k + \varepsilon_k, \qquad k = 1, \dots, n \tag{3.1}$$

where

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

(**H**2) $(\varepsilon_k)_{k\geqslant 1}$ are i.i.d. with known common density f_ε, with respect to the Lebesgue measure,

(**H**3) $(X_k)_{k\geqslant 1}$ and $(\varepsilon_k)_{k\geqslant 1}$ are independent.

We observe n copies $Z_1, \dots, Z_n$. We want to estimate f, the distribution of X_1, using $Z_1, \dots, Z_n$ only. Under (**H**3), if we denote by f_Z the density of Z_1, we can write

$$f_Z = f * f_\varepsilon, \tag{3.2}$$

where $g * h(x) = \int_{\mathbb{R}} g(u)h(x-u)du$ is the convolution product of the functions g and h under adequate assumptions. Formula (3.2) explains the term of "deconvolution" for density estimation in model (3.1). The deconvolution problem has been widely studied in the literature. It appears that two factors influence the rate of convergence : the regularity of f and the asymptotic decay of the Fourier transform of the errors f_ε, with slower rate of convergence if this decay is faster. Two types of errors are usually considered : "ordinary smooth" errors, when the Fourier transform of f_ε is polynomially decaying near infinity, and "super smooth" errors, when it is exponentially decaying near infinity. The first works proposed kernel nonadaptive estimators assuming that f is ordinary smooth and that f_ε is ordinary or super smooth. We can cite Carroll and Hall (1988), Fan (1991), Fan (1993), among others, see also the monograph of Meister (2009) on the topic. Adaptive estimation, based on a wavelet method, was first considered by Pensky and Vidakovic (1999). Butucea (2004) establishes the minimax rate in the case where f is super smooth and f_ε is ordinary smooth while Butucea and Tsybakov (2007) study optimality in the very difficult case where both functions are super smooth. Some more recent works were dedicated to this problem : Comte and Lacour (2011) consider the case where the noise density is unknown, and propose an adaptive estimator in this setting, later improved by Kappus and Mabon (2014). Mabon (2017) builds a projection estimator in Laguerre basis in the case where the variable of interest is positive.

Recently, Comte and Genon-Catalot (2018) and Belomestny et al. (2019) described nice properties of Hermite basis. Projection methods allow to summarize the information available on the unknown function through a small number of coefficients. This is why we go further in this direction, and we define a new estimator taking advantage of these convenient properties of Hermite basis. We propose also an adaptive model selection procedure. We obtain a simple, fast and powerful procedure, which preserves standard deconvolution rates. Moreover, its numerical performances are very good.

The chapter is organized as follows : we define our estimator in Section 3.2.4. We prove a bound on the risk in both the independent and β-dependent cases in Section 3.3, and

discuss rates of convergence in Section 3.3.2. In Section 3.4, an adaptive estimation procedure is proposed in the independent case and a risk control of the resulting estimator is provided. We then illustrate the performance and stability of the adaptive estimation procedure in Section 3.5, and we compare our results with Comte and Lacour (2011). Proofs of most theoretical results are gathered in Section 3.7.

3.2 Estimation procedure and Hermite basis

3.2.1 Notations.

For a, $b \in \mathbb{R}$, let $a \vee b = \max(a,b)$, and $a_+ = \max(0,a)$. For f, g in $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$, we denote by $\langle f,g\rangle = \int_{\mathbb{R}} f(u)\overline{g(u)}du$, $\|f\|^2 = \int_{\mathbb{R}} |f(u)|^2 du$, $f^*(x) = \int_{\mathbb{R}} e^{itu} f(u)du$ and $f*g(x) = \int_{\mathbb{R}} f(x-u)g(u)du \ \forall x \in \mathbb{R}$. Lastly, we recall Plancherel-Parseval formula $\langle f,g\rangle = (2\pi)^{-1}\langle f^*, g^*\rangle$.
Before proposing an estimator, we start by recalling the definition of the Hermite basis.

3.2.2 Hermite basis

The Hermite basis $(\varphi_j)_{j\geqslant 0}$ is a basis on $\mathbb{L}^2(\mathbb{R})$ defined from Hermite polynomials $(H_j)_{j\geqslant 0}$:

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

The Hermite polynomials are orthogonal with respect to the weight function e^{-x^2} :

$$\int_{\mathbb{R}} H_j(x)H_k(x)e^{-x^2}dx = 2^j j!\sqrt{\pi}\delta_{j,k}$$

(see Abramowitz and Stegun (1964), chap 22.2.14), where $\delta_{j,k}$ is the Kronecher symbol. Thus, we deduce that the basis :

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

is orthonormal in $\mathbb{L}^2(\mathbb{R})$. The Hermite basis $(\varphi_j)_{j\geqslant 0}$ is a bounded basis verifying

$$\|\varphi_j\|_\infty = \sup_{x\in\mathbb{R}} |\varphi_j(x)| \leqslant \phi_0, \text{ with } \phi_0 = 1/\pi^{1/4} \tag{3.3}$$

(see Abramowitz and Stegun (1964), chap 22.14.17 and Indritz (1961)). The Fourier transform of $(\varphi_j)_{j\geqslant 0}$ verifies :

$$\varphi_j^* = \sqrt{2\pi}(i)^j \varphi_j. \tag{3.4}$$

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

$$|\varphi_j(x)| < Ce^{-\xi x^2}, \qquad |x| \geqslant \sqrt{2j+1}, \qquad C > 0, \tag{3.5}$$

where ξ is a positive constant independent of x, $0 < \xi < \frac{1}{2}$.

3.2.3 Assumptions on the noise.

For the definition of our estimator, we assume the following :

(**H4**) the noise density f_ε is such that $f_\varepsilon^* \neq 0$.

We also assume that f_ε satisfies :

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

$$c_1'(1+t^2)^\gamma e^{\mu|t|^\delta} \leqslant \frac{1}{|f_\varepsilon^*(t)|^2} \leqslant c_1(1+t^2)^\gamma e^{\mu|t|^\delta}, \text{ for all } t \in \mathbb{R}. \tag{3.6}$$

It is standard to assume a condition like (3.6) in the deconvolution setting. When $\delta = 0$ in (3.6), the function f_ε and the errors are called "ordinary smooth". When $\delta > 0$ (with the convention that $\delta > 0$ if and only if $\mu > 0$), they are called "super smooth". Note that (3.6) implies (**H4**) and is checked by some classical distributions : we can cite for example Laplace (with $\delta = 0$ and $\gamma = 2$), Gamma ($\delta = 0$ and $\gamma = p$, where p is the shape), Gaussian ($\gamma = 0$ and $\delta = 2$), Cauchy distributions ($\gamma = 0$ and $\delta = 1$).

Remark 3.1. *According to Lukacs (1970), Theorem 4.1.1, the only characteristic function ϕ with $\phi(t) = 1 + o(t^2)$, as $t \to 0$, is the function $\phi(t) = 1$ for all t. That rules out characteristic functions of the form $e^{-\mu|t|^\delta}$ with $\delta > 2$. This implies that in definition (3.6), when $\gamma = 0$, if $|f_\varepsilon^*(t)|^2 = ce^{-\mu|t|^\delta}$ then necessarily $\delta \leqslant 2$. Indeed, $|f_\varepsilon^*(t)|^2$ is also the characteristic function of a probability density function (it is a characteristic function of $\varepsilon_1 - \varepsilon_1'$ where ε_1 and ε_1' are i.i.d.).*

3.2.4 Estimation procedure.

We denote by S_m=span$\{\varphi_0, \ldots, \varphi_{m-1}\}$, the linear space generated by $(\varphi_0, \ldots, \varphi_{m-1})$ in $\mathbb{L}^2(\mathbb{R})$. Now, we construct an estimator of f relying on the data $Z_1, \ldots, Z_n$, from model (3.1). We suppose that f belongs to $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$, thus we can write $f = \sum_{j=0}^{+\infty} a_j\varphi_j$ with $a_j = \langle f, \varphi_j\rangle$ and the orthogonal projection of f on S_m is given by : $f_m = \sum_{j=0}^{m-1} a_j\varphi_j$. In fact, we estimate f_m and therefore, we build m estimators $\widehat{a}_j$ of $a_j, j = 0, \ldots, m-1$. Under (**H4**) and using (3.2), we have $f^* = \frac{f_Z^*}{f_\varepsilon^*}$. Therefore, using Parseval's Theorem and (3.4), we have :

$$a_j = \langle f, \varphi_j\rangle = \frac{1}{2\pi}\langle f^*, \varphi_j^*\rangle = \frac{(-i)^j}{\sqrt{2\pi}}\langle f^*, \varphi_j\rangle = \frac{(-i)^j}{\sqrt{2\pi}}\int \frac{f_Z^*(u)}{f_\varepsilon^*(u)}\varphi_j(u)du. \tag{3.7}$$

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

$$\widehat{f}_Z^*(t) = \frac{1}{n}\sum_{k=1}^{n} e^{itZ_k}. \tag{3.8}$$

Plugging (3.8) into (3.7), we can propose an estimator of f_m, provided that $\varphi_j/f_\varepsilon^*$ is integrable on $\mathbb{R}$, for $j = 0, \ldots, m-1$:

$$\widehat{f}_m = \sum_{j=0}^{m-1} \widehat{a}_j\varphi_j, \qquad \widehat{a}_j = \frac{(-i)^j}{\sqrt{2\pi}}\int \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)}\varphi_j(u)du. \tag{3.9}$$

Note that the coefficients $\widehat{a}_j$ are real. Indeed, using that $\varphi_j(-x) = (-1)^j\varphi_j(x)$, it holds :

$$\overline{\widehat{a}_j} = \frac{(i)^j}{\sqrt{2\pi}}\int \frac{\widehat{f}_Z^*(-u)}{f_\varepsilon^*(-u)}\varphi_j(u)du = \frac{(i)^j}{\sqrt{2\pi}}\int \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)}\varphi_j(-u)du = \widehat{a}_j,$$

where $\overline{z}$ denotes the complex conjugate of the complex number z. The Hermite basis has the specificity of leading to integrable $\varphi_j/f_\varepsilon^*$ in a large number of cases. This estimator is different from the one proposed by Comte and Genon-Catalot (2018), who propose to take instead of $\widehat{a}_j$, the estimator

$$\widetilde{a}_{j,\sqrt{m}} = ((-i)^j/\sqrt{2\pi})\int_{|u|\leqslant\pi\sqrt{m}} \widehat{f}_Z^*(u)\varphi_j(u)/f_\varepsilon^*(u)du.$$

The drawback of the latter estimator is that it is biased and the coefficients depend on m, making the choice of m untractable in the sequel. Our estimator is an unbiased estimator of f_m and is easy to handle.

3.3 Risk study of the estimator

3.3.1 Risk of the estimator for fixed m.

Under the additional assumption :

(**H5**) f_Z is bounded,

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

Proposition 3.3.1. *(i) Under* (**H**1)*, ...,* (**H5**) *and for* $\widehat{f}_m$ *given by (3.9), we have for any* $l > 0$

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leqslant \|f - f_m\|^2 + \frac{1}{\pi n}\int_{|u|\leqslant\sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{2}{n}\|f_Z\|_\infty \sum_{j=0}^{m-1}\int_{|u|>\sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2}du, \tag{3.10}$$

(ii) If in addition we choose $l \geqslant 2$ *and if* f_ε *satisfies (3.6) with* $0 \leqslant \delta < 2$ *or* ($\delta = 2$*, with* $\mu < \xi$*), where* ξ *is defined in (3.5), then*

$$\frac{2}{n}\|f_Z\|_\infty \sum_{j=0}^{m-1}\int_{|u|>\sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2}du = \mathcal{O}(\frac{1}{n}). \tag{3.11}$$

Note that the constant l does not depend on m or n. The first right-hand side term of (3.10) is the bias term, it is decreasing with m as $\|f - f_m\|^2 = \sum_{j\geqslant m} a_j^2$. The second term is the main variance term, it is clearly increasing with m. The last term also comes from the variance computation, but we give in Proposition 3.3.1, part (ii) conditions ensuring that it is negligible. Thus, choosing m that minimizes the risk requires a bias-variance compromise.

So under the assumptions of Proposition 3.3.1, part (ii), (3.10) becomes :

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

Remark 3.2. *Condition* (**H5**) *is not very strong and holds if f or f_ε is bounded, or if both functions are square integrable. Indeed : we both have* $\forall x \in \mathbb{R}$, $|f_Z(x)| = |f * f_\varepsilon(x)| \leqslant \min(\|f\|_\infty, \|f_\varepsilon\|_\infty)$ *and* $|f_Z(x)| \leqslant \|f\|.\|f_\varepsilon\|$.

3.3.2 Rate of convergence on a Sobolev-Hermite space.

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

Definition 3.3.1. *For $s > 0$, the Sobolev-Hermite space of regularity s (see Bongioanni and Torrea (2006)) is given by :*

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

and the Sobolev-Hermite ball by :

$$W_H^s(D) = \left\{ \theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k \geqslant 0} k^s a_k^2(\theta) \leqslant D \right\}, \quad D > 0. \tag{3.12}$$

For s integer, θ belongs to W_H^s if and only if θ admits derivatives up to order s and the functions $\theta, \theta', \ldots, \theta^{(s)}, x^{(s-k)}\theta^{(k)}$ belong to $\mathbb{L}^2(\mathbb{R})$, with $k = 0, \ldots, s-1$. We can compare this space with the classical Sobolev space with regularity s, defined by :

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

Actually, Bongioanni and Torrea (2006) prove that, for $s > 0$, $W_H^s \subsetneq W^s$. It is also proved therein and in Belomestny et al. (2019) that, for s integer,

$$W^s = \left\{ \begin{array}{c} \theta \in \mathbb{L}^2(\mathbb{R}),\ \theta \text{ admits derivatives up to order } s, \text{ such that} \\ |||\theta|||_{s,sob} := \sum_{j=0}^s |\theta^{(j)}|^2 < +\infty \end{array} \right\}.$$

Consequently, for s integer, it follows that $W_H^s \subset W^s$. For more details on these regularity spaces, the reader is referred to Section 4.1 in Belomestny et al. (2019).

Thus, for f in $W_H^s(D)$, we have $\|f - f_m\|^2 = \sum_{j \geqslant m} j^s a_j^2 j^{-s} \leqslant D m^{-s}$. Under the assumptions of Proposition 3.3.1 and for $f \in W_H^s(D)$, we get :

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leqslant D m^{-s} + \frac{1}{\pi n} \int_{|u| \leqslant \sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} + \frac{C}{n}, \tag{3.13}$$

where $C > 0$, for two functions u, v, we denote $u(x) \lesssim v(x)$ if $u(x) \leqslant cv(x)$, with c is constant independent of x. This inequality is similar to the one in Comte and Lacour (2011), with m therein replaced now by $\sqrt{m}$. It is worth underlining that the role of the dimension m in projection methods is played here by $\sqrt{m}$: this is a specificity of the

Hermite basis. The result is similar in density estimation when X_k are directly observed, (see Comte and Genon-Catalot (2018), Belomestny et al. (2019)). Let us denote by m_{opt} the value of m for which the bias-variance compromise is obtained, relying on the same calculations as in Comte and Lacour (2011), the rates and the dimension m_{opt} are given in Table 3.1.

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

TABLE 3.1 – Rate of convergence for the MISE if $f \in W_H^s(D)$.

These rates coincide with the ones obtained by Fan (1993), Pensky and Vidakovic (1999). They are known to be optimal : lower bounds corresponding to these rates for f_ε verifying (3.6) are proved by Fan (1993) when f belongs to a Hölder class, and by Pensky and Vidakovic (1999) for f in a Sobolev class.

3.3.3 Rates of convergence for specific function classes

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

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

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

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

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

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

with ϕ the density of standard Gaussian and $\mathcal{P}(\mathbb{R})$ the set of probability measures on $\mathbb{R}$. The following results are based on bias evaluation obtained in Belomestny et al. (2019). The rate is given by the order of variance term, since in all these cases, the bias term is exponentially small. We can prove the following proposition.

Proposition 3.3.2. *Assume the assumptions* (**H**1)*, ...,* (**H5**) *hold and f_ε is ordinary smooth, that is f_ε satisfies 3.6 with $\delta = 0$. For the choice $m_{opt} = [\log(n)/C_1]$, with $C_1 =$*

$\log(2) + e\mu^2$ *if* $f = f_{\mu,1}$, $C_1 = \log\left(\frac{\sigma^2+1}{\sigma^2-1}\right)^2$ *where* $\sigma^2 \neq 1$ *if* $f = f_{0,\sigma}$, $C_1 = \frac{1}{(eC+1/\log(2))}$ *if* $f \in \mathcal{F}(C)$, $C_1 = \left(\frac{v^2-1}{v^2+1}\right)$ *if* $f \in \mathcal{G}(v)$, *we have*

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

where γ *is given in (3.6).*

The same result holds for $f = g_{p,\sigma}$. This rate is similar to the one obtained in Butucea (2004) for super-smooth functions f. Note that if $\sigma^2 = 1$, the bias is null and our estimator reaches the parametric rate.

However in all previous cases the choice $m = m_{opt}$ depends on the regularity of f and associated parameters, which are unknown. This is why we have to look for another method to make the bias-variance compromise, in a data-driven way (see Section 3.4).

3.3.4 Comparison with the classical estimator in deconvolution.

The "standard" deconvolution estimator (see Fan (1991), and choose sinus cardinal kernel) is given by :

$$\check{f}_\ell(x) = \frac{1}{2\pi}\int_{-\pi\ell}^{\pi\ell} e^{-ixu}\frac{\widehat{f_Z^*}(u)}{f_\varepsilon^*(-u)}du, \text{ where } \widehat{f_Z^*} \text{ is defined by (3.8).} \tag{3.14}$$

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

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

A finite (computable) development would require an additional approximation (truncation of the sum as in Comte et al. (2008)) to $k_n \geqslant n$ coefficients. From computation point of view, the low complexity of $\widehat{f}_m$ in the Hermite basis is an advantage (see Belomestny et al. (2019), Section 4.5). The risk of $\check{f}_\ell$ verifies

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

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

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

Note that it is proved in Belomestny et al. (2019) that $W_H^s(D) \subset W^s(D')$, for D and D' related constants. For $f \in W^s(D)$ the bias term is such that $\frac{1}{2\pi}\int_{|t|>\pi\ell}|f^*(u)|^2\, du \leqslant \frac{D}{2\pi}(\pi\ell)^{-2s} = C\ell^{-2s}$, where $C = \frac{D}{2\pi}\pi^{-2s}$. Therefore, for $\ell = \sqrt{m}$, the risks of the two estimators have the same order on $W_H^s(D)$. This implies that they have the same rates of convergence.

3.3.5 Extension to the dependent case.

Proposition 3.3.1 (and its consequences) may be extended to the context of dependent X_i's. We first define the mixing coefficients.

Definition 3.3.2. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $\mathcal{U}$, $\mathcal{V}$ two σ-algebras of $\mathcal{A}$. The β-mixing coefficient is defined by*

$$\beta(\mathcal{U}, \mathcal{V}) = \frac{1}{2} \sup\{\sum_{i=1}^{\mathcal{I}} \sum_{j=1}^{\mathcal{J}} |\mathbb{P}(U_i \cap V_j) - \mathbb{P}(U_i)\mathbb{P}(V_j)|\}, \tag{3.15}$$

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

Let $(X_k)_{k\in\mathbb{Z}}$ a strictly stationary process. Let $\mathcal{F}_0 = \sigma(X_i, i \leqslant 0)$ and $\mathcal{F}_k = \sigma(X_i, i \geqslant k)$ for all $k \in \mathbb{Z}$, where $\mathcal{F}_0$ is the σ-algebra generated by the X_i for $i \leqslant 0$ and $\mathcal{F}_k$ generated by X_i for $i \geqslant k$. The mixing coefficient β_k is defined by $\beta_k = \beta(\mathcal{F}_0, \mathcal{F}_k)$, where β is defined by (3.15).

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

In this section, we still consider model (3.1), but we replace (**H**1) by :

(**H**′1) $(X_k)_{k\geqslant 1}$ is strictly stationary and β-mixing.

The estimator is the same as in the independent case and we can prove a bound on the risk.

Proposition 3.3.3. *Let assumptions (**H**′1), (**H**2), ..., (**H5**) hold. Let $1 \leqslant p, q < +\infty$ two real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$. Then, if $\mathbb{E}[|X_1|^{2q/3}] < +\infty$ and the mixing coefficient are such that $\sum_{k=0}^{+\infty}(k+1)^{p-1}\beta_k < +\infty$, we have*

$$\begin{aligned}\mathbb{E}[\|\widehat{f}_m - f\|^2] \leqslant & \|f - f_m\|^2 + \frac{1}{\pi n}\int_{|u|\leqslant\sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} \\ & + \frac{2}{n}\|f_Z\|_\infty \sum_{j=0}^{m-1} \int_{|u|>\sqrt{lm}} \frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2} du + c'\frac{\sqrt{m}}{n},\end{aligned} \tag{3.16}$$

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

Now, we comment this bound of risk. We remark that we have the same bias and variance terms as in the i.i.d. case with an additional term $c'\sqrt{m}/n$ which is clearly specific to the β-mixing case. As $|f_\varepsilon^*(u)| \leqslant 1$, we have, $\frac{1}{\pi}\int_{|u|\leqslant\sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2} \geqslant \frac{2\sqrt{l}}{\pi}\sqrt{m}$. Consequently, $\sqrt{m}/n$ has smaller order than $\frac{1}{\pi n}\int_{|u|\leqslant\sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}$ and Inequality (3.16) implies that the risk of $\widehat{f}_m$ here has the same order as in the i.i.d. case. We have therefore the same rates of convergence.

We compare the result given in Proposition 3.3.3 to Proposition 4.1 in Comte et al. (2008). The first two right-hand side terms of (3.16) ($\|f-f_m\|^2+\frac{1}{\pi n}\int_{|u|\leqslant\sqrt{lm}} \frac{du}{|f_\varepsilon^*(u)|^2}$) are the same as

in Comte et al. (2008) with $\sqrt{lm}$ replaced by πm (see Section 3.3). Under the assumptions of Proposition 3.3.1 (ii) the other terms (residual terms) are order $\mathcal{O}(n^{-1}) + \mathcal{O}(\sqrt{m}n^{-1})$. This order is smaller than the order of the residual term stated in (4.4) of Comte et al. (2008), which is $n^{-1}m^2$. Note that all estimators of their collection require to compute $k_n \geqslant n$ coefficients, which can make the procedure slow when n is large.

3.4 Adaptive estimation and model selection

For sake of brevity and simplicity, we only study the independent case (i.i.d case) hereafter.

From now on, l given in Proposition 3.3.1, part (ii) is assumed to be fixed. In this section we propose an automatic selection of m which performs the bias-variance compromise. The procedure does not depend on the regularity of the density f, but only on data $Z_1, \ldots, Z_n$. Consider the contrast function defined by

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{k=1}^{n}\phi_t(Z_k), \qquad \phi_t(x) = \frac{1}{2\pi}\int \frac{t^*(u)}{f_\varepsilon^*(-u)}e^{-ixu}du. \tag{3.17}$$

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

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

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

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

This collection is finite and contains models with bounded variance. More precisely, as already noticed, $|f_\varepsilon^*(u)| \leqslant 1$, implies $\Delta(m) \geqslant \frac{1}{\pi}\int_{|u|\leqslant\sqrt{lm}} du = \frac{2\sqrt{lm}}{\pi}$. Therefore, the elements m of $\mathcal{M}_n$ satisfy $m \lesssim n^2$. The cardinal of $\mathcal{M}_n$ is therefore at most of order $\mathcal{O}(n^2)$. Our aim is to find the best model $\widehat{m}$ in $\mathcal{M}_n$, that is, to select $\widehat{m}$ such that, the risk of $\widehat{f}_{\widehat{m}}$ approximately performs the bias-variance trade-off, without any information on f. We set :

$$\widehat{m} = \underset{m\in\mathcal{M}_n}{\operatorname{argmin}}\{\gamma_n(\widehat{f}_m) + \operatorname{pen}(m)\}, \tag{3.18}$$

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

$$\operatorname{pen}(m) = \begin{cases} \kappa\frac{\Delta(m)}{n}, & \text{if } f_\varepsilon \text{ is ordinary smooth or super smooth with } \delta < \frac{1}{2}, \\ 2\kappa\left(1 + 24\mu l^{\delta/2}m^{\delta-\frac{1}{2}}\right)\frac{\Delta(m)}{n} & \text{if } f_\varepsilon \text{ is super smooth with } \frac{1}{2} \leqslant \delta \leqslant 2, \end{cases} \tag{3.19}$$

where $\kappa > 0$ is a numerical constant, μ is the constant given in (3.6) and $l \geqslant 2$ given in Proposition 3.3.1, fixed. As $\gamma_n(\widehat{f}_m) = -\|\widehat{f}_m\|^2 = -\sum_{j=0}^{m-1}\widehat{a}_j^2$, it is worth emphasizing that computing $\widehat{m}$ is numerically fast. Clearly the choice of m given by (2.18) is entirely determined by the data. The constant κ is independent of the data. The theoretical results show that $\kappa > 17$ is suitable (see the proof of Lemma 3.7.2). In practice this value is too large and is calibrated by preliminary simulation experiments. They confirm that (see Section 3.5) smaller practical values must be chosen.

We can prove the following oracle inequality.

Theorem 3.4.1. *Assume* (**H**1), …, (**H5**) *hold and* f_ε *is square integrable. Let* $\mathrm{pen}(m)$ *defined by (3.19),* $\widehat{f}_m = \underset{t\in S_m}{\mathrm{argmin}}\gamma_n(t)$ *and* $\widehat{m}$ *selected by (3.18). Then, there exists a constant* κ_0 *such that, for all* $\kappa > \kappa_0 = 17$*, the estimator* $\widehat{f}_{\widehat{m}}$ *satisfies*

$$\mathbb{E}\left[\|\widehat{f}_{\widehat{m}} - f\|^2\right] \leqslant C \inf_{m\in\mathcal{M}_n}\left(\|f - f_m\|^2 + \mathrm{pen}(m)\right) + \frac{C'}{n}, \tag{3.20}$$

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

Remark 3.3. *Assume that the assumptions of Theorem 3.4.1 are satisfied. Then if* $f \in W_H^s(D)$ *the estimator* $\widehat{f}_{\widehat{m}}$ *converges to* f *with the rates obtained in Table 3.1. Indeed, the term* C'/n *in (3.20) does not change the order of the rate, and is negligible compared to the term* $\|f - f_m\|^2 + \mathrm{pen}(m)$*. Moreover, (3.19) induces a loss in the order of* $\mathrm{pen}(m)$ *compared to the variance term when* $\delta > 1/2$*, but this does not change the rate which is governed by the bias term in this case (see Table 3.1 and choice of* m_{opt} *of order* $(\log n)$*).*

3.5 Simulation and numerical results

3.5.1 Implementation of the adaptive estimator.

In this section, we propose some illustrations of the theoretical results. More precisely, we implement the projection estimator given by (3.9). To do this, we consider data simulated according to (3.1). For the density f, we choose the distributions (following Comte and Lacour (2011)) :

(i) Gaussian standard $\mathcal{N}(0,1)$, $I = [-4,4]$

(ii) Cauchy standard : $f(x) = \left(\pi\left(1+x^2\right)\right)^{-1}$, $I = [-10,10]$

(iii) Laplace density : $f(x) = e^{-\sqrt{2}|x|}/\sqrt{2}$, $I = [-5,5]$

(iv) Gamma density $\Gamma(4, 1/\sqrt{3})/\sqrt{12}$, $I = [0,6]$

(v) Mixed-Gaussian density $(0.5\mathcal{N}(-2,1) + 0.5\mathcal{N}(2,1))/\sqrt{5}$, $I = [-3,3]$

where I is the interval on which we compute the risks. Except the Cauchy density, all the densities are normalized to have variance equal to 1. Note also densities (i) and (v) belong to W_H^s with $s = +\infty$, (iv) has regularity $s = 3-\eta$, $0 < \eta < 3$, (ii) and (iii) admit a regularity $s = 1-\eta$, $0 < \eta < 1$ (but (ii) is infinitely differentiable).
For noise distributions, we consider two cases with the same variance $1/10$ and thus, except for the Cauchy density the signal to noise ratio is equal to 10.

- **Case 1 : Laplace noise ("ordinary smooth")**
  We consider the density f_ε :

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

  The penalty term is given by :

$$\mathrm{pen}(m) = \frac{\kappa}{n}\Delta(m) = \frac{\kappa}{\pi n}\int_{|u|\leqslant\sqrt{lm}}(1+\frac{u^2}{\lambda^2})^2 du = \frac{2\kappa}{\pi n}\left(\sqrt{lm} + \frac{2}{3\lambda^2}(\sqrt{lm})^3 + \frac{(\sqrt{lm})^5}{5\lambda^4}\right),$$

  where $l = 6$.

- **Case 2 : Gaussian noise ("super smooth")**
  We have :

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

  The penalty proposed is :

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

  where $l = 4$ here and the integral is computed by a Riemann sum discretized in 300 points.

Then, we have to calibrate the penalty constant κ. This constant is fixed through preliminary simulations, by testing set of values on different densities f with a large number of repetitions. The comparison of the risks for these different values of κ makes it possible to make a reasonable choice. We choose $\kappa = 0.4$ for a Laplace noise, $\kappa = 10^{-3}$ for a Gaussian noise. We fix the maximum dimension equal to 50 and consider the following collection of models for the two cases : $\mathcal{M}_n = \{1, \ldots, 50\}$.

The estimation procedure is described as follows :

- For m in $\mathcal{M}_n$, compute $-\sum_{j=0}^{m-1} \widehat{a}_j^2 + \text{pen}(m) = \text{Cr}(m)$, with $\widehat{a}_j$ given by (3.9),
- Select $\widehat{m}$ such that $\widehat{m} = \underset{m\in\mathcal{M}_n}{\text{argmin}} \quad \text{Cr}(m)$,
- Compute $\widehat{f}_{\widehat{m}} = \sum_{j=0}^{\widehat{m}-1} \widehat{a}_j \varphi_j$, and $\int_I (\widehat{f}_{\widehat{m}}(u) - f(u))^2 du$ by discretization.

Note that the coefficients $\widehat{a}_j$ are computed by elementary discretization of the Riemann integrals. We used that

$$\begin{aligned}\widehat{a}_j = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)} \overline{\varphi_j^*(u)} du &= \frac{1}{2\pi}\left(\int_0^{+\infty} \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)} \varphi_j^*(-u) du + \int_{-\infty}^0 \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)} \varphi_j^*(-u) du\right) \\ &= \frac{1}{\pi} \Re\left(\int_0^{+\infty} \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)} \varphi_j^*(-u) du\right),\end{aligned}$$

where $\Re$ is the real part of a complex number. We approximate the integral $\frac{1}{\pi}\Re\left(\int_0^{+\infty} \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)} \varphi_j^*(-u) du\right)$ by a Riemann sum on $[0, b]$, for b large enough, $\widehat{a}_j \approx \frac{b}{\pi K} \sum_{p=1}^K \Re\left[\frac{\widehat{f}_Z^*(\frac{pb}{K})}{f_\varepsilon^*(\frac{pb}{K})} \varphi_j^*(-\frac{pb}{K})\right]$. We take $b = 50$.

Comparaison with the direct case. We also implemented the adaptive estimator $\widetilde{f}_m$ in the direct case (*i.e.* the case where we observe directly $(Z_i = X_i)_{1\leqslant i\leqslant n}$ with $\varepsilon_i = 0$ almost surely in (3.1)). Recall that, it is given by (see also Chapter 2) :

$$\widetilde{f}_m := \sum_{k=0}^{m-1} \widetilde{a}_k \varphi_k, \quad \text{where} \quad \widetilde{a}_k := \frac{1}{n}\sum_{j=0}^n \varphi_k(X_j).$$

In this case, to choose m, we set

$$\widetilde{m} := \underset{m\in\{1,\ldots,n\}}{\text{argmin}} \{-\|\widetilde{f}_m\|^2 + \widetilde{\text{pen}}_d(m)\}, \quad \text{where} \quad \widetilde{\text{pen}}_(m) = \kappa \frac{\widetilde{V}_{m,d}}{n},$$

where $\widetilde{V}_m = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{m-1}(\varphi_j(X_i))^2$ and κ is a positive numerical constant which is calibrated to 4 after numerical test.

3.5.2 Simulations results.

Simulation results are given in Tables 3.2, 3.3 and 3.4. The columns of Table 3.2 indicate the values of the MISE (Mean Integrated Squared Error) multiplied by 100 for a Laplace noise or a Gaussian noise, Table 3.3 gives the ratio of the risk values obtained in Comte and Lacour (2011) divided by the risk values obtained by our method : the larger it is, the better our method is. The errors obtained by our method are computed by a discretization of the integral as Riemann sums and averaged over 100 independent simulations. We remark that increasing the sample size makes the error smaller and thus improves the estimation. Globally the results of our simulations are satisfactory and our method is often better than Comte and Lacour (2011) for both noise densities. The main exception concerns the Gamma density (iv) which has $\mathbb{R}^+$ supported. Some failures for Cauchy density (ii) and super smooth noise are also observed, especially when n increases. In Table 3.4, we give the MISEs for the direct observation. Except the case (ii), it seems that the estimator obtained in direct case is better than the deconvolution cases. This illustrates the fact that we are in the context of an inverse problem.

We also illustrate our method by some figures. Figure 3.1 and 3.2 display the density and its 20 estimates in direct and deconvolution cases. For each graph, the first column corresponds to the direct case, the middle to Laplace noise and the right to Gaussian noise. These graphs illustrate the performance of estimation in the direct and deconvolution cases : it seems that the estimator obtained in direct case is better than the deconvolution cases (see also Table 3.4). This can be clearly seen in Figure 3.1 ($n = 1000$) and 3.2. Moreover, they show the stability of both cases. We provide in Table 3.5, the mean of $\widehat{m}$ or $\widetilde{m}$ selected by the algorithm. In average, it is increasing when n is increasing.

	$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
f \ Noise	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.
Gaussian	0.44	0.37	0.12	0.06	0.1	0.04	0.07	0.04
Cauchy	0.28	0.89	0.20	0.56	0.14	0.37	0.10	0.29
Laplace	1.65	2.18	1.06	1.34	0.75	1.16	0.57	0.87
Gamma	1.70	1.27	0.98	0.97	0.50	0.90	0.28	0.83
Mixed-Gaussian	2.82	1.91	1.09	0.87	0.66	0.69	0.41	0.53

TABLE 3.2 – Empirical integrated mean squared errors computed from $(100 \times \mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|^2)$ over 100 independent simulations for $n = 100, 250, 500, 1000$.

	$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
f \ Noise	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.	Lap.	Gauss.
Gaussian	1.95	1.27	5.67	5.00	5.01	5.11	2.41	3.41
Cauchy	4.07	1.07	2.45	0.79	2.43	0.70	1.40	0.52
Laplace	1.47	1.40	1.13	1.34	1.12	1.02	1.04	0.89
Gamma	0.67	0.88	0.66	0.73	0.82	0.49	1	0.37
Mixed-Gaussian	1.26	2.17	1.45	2.24	1.17	1.68	0.95	1.15

TABLE 3.3 – Ratio of the risks obtained in Comte and Lacour (2011) divided by those of Table 3.2.

	Direct case			
f \ n	100	250	500	1000
Gaussian	0.17	0.05	0.04	6.05×10^{-3}
Cauchy	0.62	0.50	0.36	0.18
Laplace	2.43	1.09	0.69	0.42
Gamma	1.20	0.54	0.23	0.13
Mixed-Gaussian	1.35	0.49	0.26	0.15

TABLE 3.4 – Empirical MISE $100 \times \mathbb{E}\|\widetilde{f}_{\widetilde{m}} - f\|^2$ over 100 independent simulations for $n = 100, 250, 500, 1000$.

Density		Gamma		Mixed-Gaussain	
n		250	1000	250	1000
Mean of $\widehat{m}$	Direct case	6.55	8.85	9.10	10.4
	Lap. noise	5.80	8.10	7.30	8.5
	Gauss. noise	5.55	7.15	6.75	7.00

TABLE 3.5 – Mean of selected dimensions $\widehat{m}$ or $\widetilde{m}$ presented in Figures 3.1 and 3.2.

3.6 Concluding remarks

We proposed a projection estimator of the density of X in the convolution model (3.1), relying on the Hermite basis. The estimator has the advantage to be kernel-free, as the integral is over the entire real line and not truncated as in the previous works by Comte and Genon-Catalot (2018). The method provides a parsimonious description of the function under estimation : indeed the function is relevantly estimated thanks to a small number of coefficients. This has also the advantage of making the method numerically fast and convenient. We prove a bound on the quadratic risk in the independent and β-dependent cases which shows that the relevant parameter is not the dimension m but rather $\sqrt{m}$. A data driven estimator is proposed : the model can be automatically chosen and the resulting estimator reaches optimal rates in most cases. We also provide numerical simulation

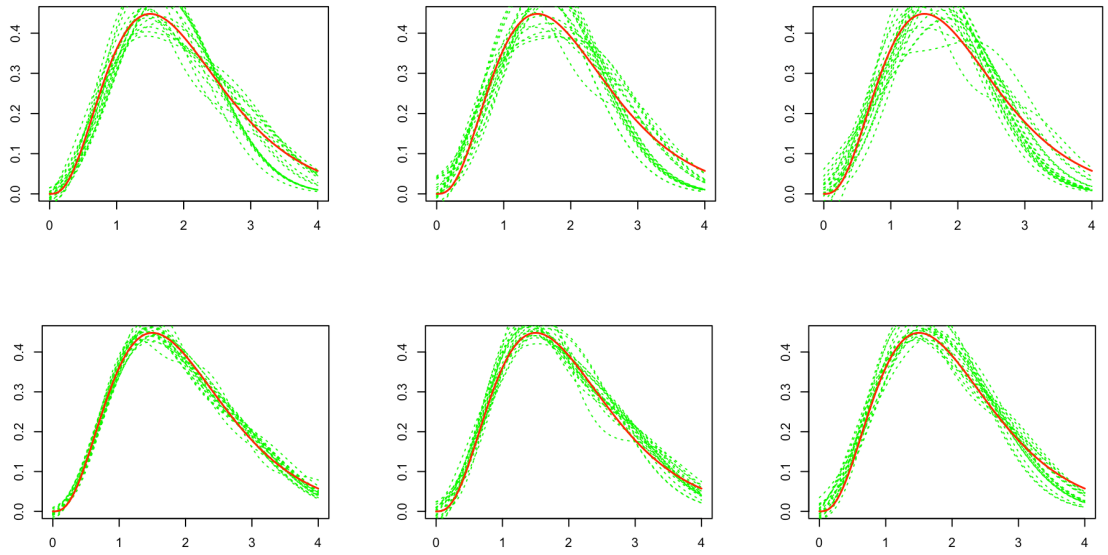


FIGURE 3.1 – 20 estimates of (iii), with $n = 250$ (first line) and $n = 1000$ (second line). The true quantity is in bold red and the estimate in dotted lines (left : direct case, middle : Laplace noise, right : Gaussian noise).

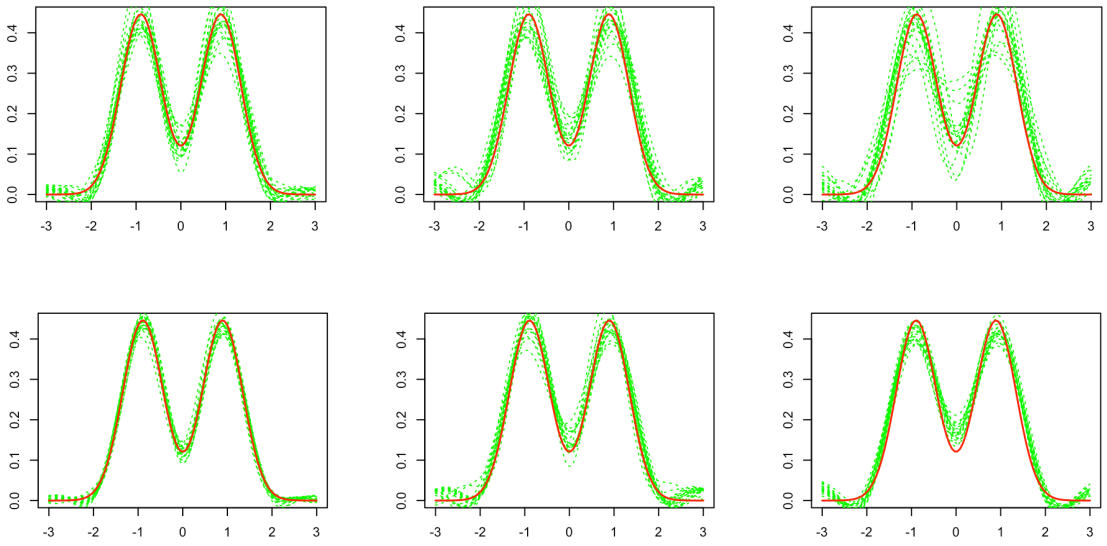


FIGURE 3.2 – 20 estimates of (iii), with $n = 250$ (first line) and $n = 1000$ (second line). The true quantity is in bold red and the estimate in dotted lines (left : direct case, middle : Laplace noise, right : Gaussian noise).

results, and the comparison with Comte and Lacour (2011) ensures the good performances

of our method.

3.7 Proofs

3.7.1 Proof of Proposition 3.3.1.

- **Proof of Part (i).** For $\widehat{f}_m$ given by (3.9), we have :

$$\mathbb{E}\left[\|\widehat{f}_m - f\|^2\right] = \|f - f_m\|^2 + \mathbb{E}\left[\|\widehat{f}_m - f_m\|^2\right] = \|f - f_m\|^2 + \sum_{j=0}^{m-1} \mathrm{Var}(\widehat{a}_j). \quad (3.21)$$

Now with the definition of $\widehat{a}_j$ given by (3.9) we have

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

Plugging this in (3.21) yields

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

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

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

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

$$\mathbb{E}\left[\sum_{j=0}^{m-1}\left|\int_{|u|\leqslant\sqrt{lm}} e^{iuZ_1}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right|^2\right] = \mathbb{E}\left[\sum_{j=0}^{m-1}\left|\left\langle\frac{e^{iZ_1\bullet}}{f_\varepsilon^*}\mathbb{1}_{|\bullet|\leqslant\sqrt{lm}}, \varphi_j\right\rangle\right|^2\right]$$
$$\leqslant \int_{|u|\leqslant\sqrt{lm}}\frac{du}{|f_\varepsilon^*(u)|^2}. \quad (3.22)$$

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

$$\begin{aligned}\mathbb{E}\left[\sum_{j=0}^{m-1}\left|\int_{|u|>\sqrt{lm}} e^{iuZ_1}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right|^2\right] &= \sum_{j=0}^{m-1}\int_{\mathbb{R}}\left|\int_{|u|>\sqrt{lm}} e^{iuz}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right|^2 f_Z(z)dz \\ &\leqslant \|f_Z\|_\infty \sum_{j=0}^{m-1}\int_{\mathbb{R}}\left|\int_{|u|>\sqrt{lm}} e^{iuz}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right|^2 dz \\ &= \|f_Z\|_\infty \sum_{j=0}^{m-1}\|\psi_j^*\|^2 = 2\pi\|f_Z\|_\infty\sum_{j=0}^{m-1}\|\psi_j\|^2. \quad (3.23)\end{aligned}$$

Putting (3.22), (4.60) in (3.21), we have the part (i).

- **Proof of Part (ii).** We have, using (3.6), that :

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

By (3.5), we have $|\varphi_j(x)| < Ce^{-\xi x^2}$ if $|x| \geqslant \sqrt{2j+1}$, for $j \in \{0,\dots,m-1\}$. Thus it is in particular true for $|x| \geqslant \sqrt{lm}$, with $l \geqslant 2$. Therefore, for $j \leqslant m-1$, we have

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

And $\int_{\mathbb{R}}(1+u^2)^\gamma e^{-\xi u^2}e^{\mu|u|^\delta}du < +\infty$ if $\delta < 2$ or if $\delta = 2$, $\mu < \xi$, which corresponds to our assumptions. Therefore, we get $\sum_{j=0}^{m-1}\int_{|u|>\sqrt{lm}}\frac{|\varphi_j(u)|^2}{|f_\varepsilon^*(u)|^2}du = \mathcal{O}(me^{-\xi lm})$. Hence the result. □.

3.7.2 Proof of Proposition 3.3.2.

By (3.10) and (3.11), we have :

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leqslant \|f - f_m\|^2 + \frac{1}{\pi n}\int_{|u|\leqslant\sqrt{lm}}\frac{du}{|f_\varepsilon^*(u)|^2} + \frac{c}{n}. \quad (3.24)$$

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

$$\int_{|u|\leqslant\sqrt{lm}}\frac{du}{|f_\varepsilon^*(u)|^2} \asymp m^{\gamma+\frac{1-\delta}{2}}e^{\mu l^{\frac{\delta}{2}}m^{\frac{\delta}{2}}}. \quad (3.25)$$

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

From Belomestny et al. (2019) the bias term is exponentially small (see Proposition 7, 8 and 9), thus, the rate of convergence is given by the order of variance term. As f_ε is ordinary smooth, $\delta = 0$ in (3.25) and replacing m by $m_{opt} = [\log(n)/C_1]$, with C_1 is given in Proposition 3.3.2, we have the result. □

3.7.3 Proof of Proposition 3.3.3.

As in the i.i.d. case, we have the bias-variance decomposition given by (3.21). Now,

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

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

$$\begin{aligned}\mathbb{E}\left[\|\widehat{f}_m - f\|^2\right] &\leqslant \|f-f_m\|^2 + \frac{1}{2\pi n}\sum_{j=0}^{m-1}\mathbb{E}\left[\left|\int_{\mathbb{R}} e^{iuZ_1}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right|^2\right]\\ &\quad+\frac{1}{2\pi n^2}\sum_{j=0}^{m-1}\sum_{1\leqslant k,l\leqslant n,k\neq l}\mathrm{Cov}\left((-i)^j\int_{\mathbb{R}} e^{iuZ_k}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du,(-i)^j\int_{\mathbb{R}} e^{iuZ_l}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right).\end{aligned} \tag{3.26}$$

The first two right hand side terms are the same as in the independent case and are dealt with as in Proposition 3.3.1. We compute the covariance term. First,

$$\begin{aligned}&\mathrm{Cov}\left((-i)^j\int_{\mathbb{R}} e^{iuZ_k}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du,(-i)^j\int_{\mathbb{R}} e^{iuZ_l}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right)\\ &\qquad=\mathbb{E}\left[\int_{\mathbb{R}}\int_{\mathbb{R}} e^{i(uZ_k-vZ_l)}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}\frac{\varphi_j(v)}{f_\varepsilon^*(-v)}dudv\right]\\ &\qquad\quad-\mathbb{E}\left[\int_{\mathbb{R}} e^{iuZ_k}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right]\mathbb{E}\left[\int_{\mathbb{R}} e^{-ivZ_l}\frac{\varphi_j(v)}{f_\varepsilon^*(-v)}dv\right].\end{aligned} \tag{3.27}$$

The first expectation is equal to

$$\begin{aligned}\mathbb{E}\left[\int_{\mathbb{R}}\int_{\mathbb{R}} e^{i(uZ_k-vZ_l)}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}\frac{\varphi_j(v)}{f_\varepsilon^*(-v)}dudv\right] &= \int_{\mathbb{R}}\int_{\mathbb{R}}\mathbb{E}\left[e^{i(uX_k+u\varepsilon_k-vX_l-v\varepsilon_l)}\right]\frac{\varphi_j(u)}{f_\varepsilon^*(u)}\frac{\varphi_j(v)}{f_\varepsilon^*(-v)}dudv\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}}\mathbb{E}\left[e^{i(uX_k-vX_l)}\right]\varphi_j(u)\varphi_j(v)dudv,\end{aligned} \tag{3.28}$$

and the second to :

$$\mathbb{E}\left[\int_{\mathbb{R}} e^{iuZ_k}\frac{\varphi_j(u)}{f_\varepsilon^*(u)}du\right]\mathbb{E}\left[\int_{\mathbb{R}} e^{-ivZ_l}\frac{\varphi_j(v)}{f_\varepsilon^*(-v)}dv\right] = \left|\int_{\mathbb{R}} f^*(u)\varphi_j(u)du\right|^2. \tag{3.29}$$

Thus, from (3.27), (3.28) and (3.29) we deduce

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

As a consequence

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

Using Viennet (1997)'s covariance inequality and equality (4.4.1), we have

$$\mathrm{Var}\left(\sum_{k=1}^{n}\int_{\mathbb{R}} e^{iuX_k}\varphi_j(u)du\right) = \mathrm{Var}\left(\sum_{k=1}^{n}\varphi_j^*(X_k)\right) \leqslant 8\pi n\int_{\mathbb{R}} b(u)\varphi_j(u)^2 f(u)du, \quad (3.30)$$

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

Lemma 3.7.1. *Under the assumptions and notations of Proposition 3.3.3, there exists a constant $c^*>0$ depending on $\mathbb{E}[|X_1|^{2q/3}]$ and $\sum_{k=0}^{+\infty}(k+1)^{p-1}\beta_k<+\infty$ such that :*

$$\int_{\mathbb{R}} b(x)\varphi_j^2(x)f(x)dx \leqslant \frac{c^*}{\sqrt{j}}, \quad \forall j\geqslant 1. \quad (3.31)$$

By Lemma 3.7.1 and (3.30), we deduce

$$\begin{aligned}\sum_{j=0}^{m-1}\mathrm{Var}\left(\sum_{k=1}^{n}\int_{\mathbb{R}} e^{iuX_k}\varphi_j(u)du\right) &\leqslant 8\pi n\left[\int_{\mathbb{R}} b(u)\varphi_0^2(u)f(u)du + \sum_{j=1}^{m-1}\int_{\mathbb{R}} b(u)\varphi_j(u)^2 f(u)du\right]\\ &\leqslant 8\pi n\left[\phi_0^2\sum_{k\geqslant 0}\beta_k + \sum_{j=1}^{m-1}\frac{c^*}{\sqrt{j}}\right]. \quad (3.32)\end{aligned}$$

Using (3.32), Proposition 3.3.1 and in view of (3.26), we obtain the announced result □.

Proof of Lemma 2.6.1.

To prove this lemma, we first use the decomposition formula of the Hermite basis in the Laguerre basis (see Comte and Genon-Catalot (2018), Lemma 8.4, p. 287) given by :

$$\varphi_{2k}(x) = (-1)^k\sqrt{x/2}\psi_k^{(-1/2)}(x^2/2), \quad \varphi_{2k+1}(x) = (-1)^k\sqrt{x/2}\psi_k^{(1/2)}(x^2/2), \quad x\geqslant 0$$

where $(\psi_k^{(\delta)})_{k\geqslant 0}$ is the Laguerre function with index $\delta>-1$ defined from the Laguerre polynomial $(L_k^{(\delta)})_{k\geqslant 0}$ with index $\delta>-1$ and degree k given by :

$$\psi_k^{(\delta)}(x) = 2^{\frac{\delta+1}{2}}\left(\frac{k!}{\Gamma(k+\delta+1)}\right)^{1/2}L_k^{(\delta)}(2x)x^{\frac{\delta}{2}}e^{-x}, L_k^{(\delta)}(x) = \frac{1}{k!}e^x x^{-\delta}\frac{d^k}{dx^k}\left(x^{\delta+k}e^{-x}\right).$$

Note that $(\psi_k^{(\delta)})_{k\geqslant 0}$ is an orthonormal basis on $\mathbb{L}^2(\mathbb{R}^+)$. Next, using the asymptotic formula of Askey and Wainger (1965) recalled in Section 3.8.2, we get a bound of $(\psi_k^{(\delta)})_{k\geqslant 0}$, for k large enough. We distinguish two cases depending on the parity of j and we study only the first term of the following decomposition :

$$\int_{\mathbb{R}} b(x)\varphi_j^2(x)f(x)dx = \int_0^{\infty} b(x)\varphi_j^2(x)f(x)dx + \int_0^{\infty} b(-x)\varphi_j^2(x)f(-x)dx,$$

since $(\varphi_j)_{j\geqslant 0}$ is even for j even and odd for j odd. The study of the other term is similar and its bound is the same as the one on the first term.
For j even, $j = 2k$, we have :

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

where J_l are integrals on disjoint domains specified below, see also Section 3.8.2. Setting $\nu = 4k+1$, we have six terms to evaluate.

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

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

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

Using the Hölder inequality, we have

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

with $\frac{1}{p}+\frac{1}{q}=1$. By Lemma 4.2 in Viennet (1997), page 481, we have :

$$\mathbb{E}\left[b(X_1)^p\right] \leqslant p\sum_{k\geqslant 0}(k+1)^{p-1}\beta_k.$$

It comes : $J_3 \leqslant \frac{C}{2\sqrt{\nu}}\mathbb{E}\left[|X_1|^{2q/3}\right]^{1/q}(p\sum_{k\geqslant 0}(k+1)^{p-1}\beta_k)^{1/p}$.

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

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

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

Again by the Hölder inequality we get : $J_5 \leqslant \frac{C}{\sqrt{\nu}}\mathbb{E}\left[|X_1|^{2q/3}\right]^{1/q}(p\sum_{k\geqslant 0}(k+1)^{p-1}\beta_k)^{1/p}$. Finally, it holds

$$\begin{aligned}
J_6 &= \frac{1}{2}\int_{\sqrt{3\nu/2}}^{\infty} x\left(\psi_k^{(-1/2)}(x^2/2)\right)^2 b(x)f(x)dx \leqslant \frac{C}{2}\int_{\sqrt{3\nu/2}}^{\infty} xe^{-\gamma_2x^2}b(x)f(x)dx \\
&\leqslant C'e^{-3\frac{\gamma_2\nu}{4}}\int_{\mathbb{R}} b(x)f(x)dx = C'e^{-3\frac{\gamma_2\nu}{4}}\mathbb{E}\left[b(X_1)\right] \leqslant C'e^{-3\frac{\gamma_2\nu}{4}}\sum_{k\geqslant 0}\beta_k.
\end{aligned}$$

For j odd, $j = 2k+1$, and setting $\nu = 4k+3$, we have :

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

Only the first term changes, thus, we just compute K_1 and the other terms are such that the bounds coincide with the case where j is even for $l = 2,\dots,6$.

$$\begin{aligned}
K_1 = \frac{1}{2}\int_0^{1/\sqrt{\nu}} x\left(\psi_k^{(1/2)}(x^2/2)\right)^2 b(x)f(x)dx &\leqslant \frac{C}{2}\int_0^{1/\sqrt{\nu}} x\left[(x^2\nu)^{1/4}\right]^2 b(x)f(x)dx \\
&\leqslant \frac{C}{2\sqrt{\nu}}\sum_{k\geqslant 0}\beta_k
\end{aligned}$$

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

3.7.4 Proof of Theorem 3.4.1.

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

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

where ϕ_t is defined in (3.17). Thus, for m any element of $\mathcal{M}_n$, we have

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

As the function $t \mapsto \nu_n(t)$ is linear, we deduce

$$\begin{aligned}\|\widehat{f}_{\widehat{m}} - f\|^2 &\leqslant \|f_m - f\|^2 + \mathrm{pen}(m) + 2\|\widehat{f}_{\widehat{m}} - f_m\|\nu_n\left(\frac{\widehat{f}_{\widehat{m}} - f_m}{\|\widehat{f}_{\widehat{m}} - f_m\|}\right) - \mathrm{pen}(\widehat{m}) \\ &\leqslant \|f_m - f\|^2 + \mathrm{pen}(m) + 2\|\widehat{f}_{\widehat{m}} - f_m\| \sup_{t\in S_m + S_{\widehat{m}}, \|t\|=1} \nu_n(t) - \mathrm{pen}(\widehat{m}). \qquad (3.33)\end{aligned}$$

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

$$2\|\widehat{f}_{\widehat{m}} - f_m\| \sup_{t\in S_m + S_{\widehat{m}}, \|t\|=1} \nu_n(t) \leqslant \frac{1}{4}\|\widehat{f}_{\widehat{m}} - f_m\|^2 + 4 \sup_{t\in S_m + S_{\widehat{m}}, \|t\|=1} (\nu_n(t))^2. \qquad (3.34)$$

Now, $\|\widehat{f}_{\widehat{m}} - f_m\|^2 \leqslant 2\|\widehat{f}_{\widehat{m}} - f\|^2 + 2\|f_m - f\|^2$ and plugging this and (3.34) in (3.33), we have

$$\frac{1}{2}\|\widehat{f}_{\widehat{m}} - f\|^2 \leqslant \frac{3}{2}\|f_m - f\|^2 + \mathrm{pen}(m) + 4 \sup_{t\in S_m + S_{\widehat{m}}, \|t\|=1} (\nu_n(t))^2 - \mathrm{pen}(\widehat{m}). \qquad (3.35)$$

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

$$\begin{aligned}\nu_n(t) &= \frac{1}{n}\sum_{k=1}^n (\phi_t(Z_k) - \langle t, f\rangle) \\ &= \frac{1}{n}\sum_{k=1}^n \left(\frac{1}{2\pi}\int_{|u|\leqslant\sqrt{lm^\star}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du - \mathbb{E}\left[\frac{1}{2\pi}\int_{|u|\leqslant\sqrt{lm^\star}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du\right]\right) \\ &\quad + \frac{1}{n}\sum_{k=1}^n \left(\frac{1}{2\pi}\int_{|u|>\sqrt{lm^\star}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du - \mathbb{E}\left[\frac{1}{2\pi}\int_{|u|>\sqrt{lm^\star}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iuZ_k} du\right]\right) \\ &= \frac{1}{n}\sum_{k=1}^n (\phi_{t,1}(Z_k) - \mathbb{E}\left[\phi_{t,1}(Z_k)\right]) + \frac{1}{2\pi}\int_{|u|>\sqrt{lm^\star}} \frac{t^*(u)}{f_\varepsilon^*(-u)} (\widehat{f}_Z^*(u) - f_Z^*(u)) du, \qquad (3.36)\end{aligned}$$

with $\phi_{t,1}(x) = \frac{1}{2\pi}\int_{|u|\leqslant\sqrt{lm^\star}} \frac{t^*(u)}{f_\varepsilon^*(-u)} e^{-iux} du$. Therefore, we write $\nu_n(t) = \nu_{n,1}(t) + \nu_{n,2}(t)$ where

$$\nu_{n,1}(t) = \frac{1}{n}\sum_{k=1}^n (\phi_{t,1}(Z_k) - \mathbb{E}\left[\phi_{t,1}(Z_k)\right])$$

and

$$\nu_{n,2}(t) = \frac{1}{2\pi}\int_{|u|>\sqrt{lm^\star}} \frac{t^*(u)}{f_\varepsilon^*(-u)} (\widehat{f}_Z^*(-u) - f_Z^*(-u)) du.$$

Using that $(\nu_{n,1}(t) + \nu_{n,2}(t))^2 \leqslant 2(\nu_{n,1}(t))^2 + 2(\nu_{n,2}(t))^2$ and by (3.35), (3.36) we deduce

$$\begin{aligned}\frac{1}{2}\|\widehat{f}_{\widehat{m}} - f\|^2 \leqslant \frac{3}{2}\|f_m - f\|^2 + \mathrm{pen}(m) &+ 8 \sup_{t\in S_{m^\star}, \|t\|=1} (\nu_{n,1}(t))^2 \\ &+ 8 \sup_{t\in S_{m^\star}, \|t\|=1} (\nu_{n,2}(t))^2 - \mathrm{pen}(\widehat{m}).\end{aligned}$$

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

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

By taking expectation, we get

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

The two followings lemmas lead to the result of Theorem 3.4.1 :

Lemma 3.7.2. *Under the assumtions of Theorem 3.4.1, there exists a constant Σ_1 such that*

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

Lemma 3.7.3. *Under the assumtions of Theorem 3.4.1, there exists a constant Σ_2 such that*

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

Using Lemmas 3.7.2 and 3.7.3, we have the result choosing $C = 4$ and $C' = 16(\Sigma_1 + \Sigma_2)$. □

Proof of Lemma 3.7.2.

To prove this lemma, we use Talagrand's inequality given in Appendix 3.8.3, and compute H^2, M_1, v defined there. Denote by $m'' = m \vee m'$. We start by computing H^2. As the map $t \mapsto \nu_{n,1}(t)$ is linear, for $t = \sum_{j=0}^{m''-1} a_j\varphi_j$ such that $\|t\| = 1$, we have

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

Therefore,

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

It comes using (3.22) that,

$$\mathbb{E}[\sum_{j=0}^{m''-1}|\phi_{\varphi_j}(Z_1)|^2] = \frac{1}{(2\pi)^2}\mathbb{E}[\sum_{j=0}^{m''-1}\left|\int_{|u|\leqslant\sqrt{lm''}}\frac{\varphi_j^*(u)e^{-iuZ_1}}{f_\varepsilon^*(-u)}du\right|^2] \leqslant \frac{\Delta(m'')}{n} := H^2. \quad (3.37)$$

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

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

Thus, it follows

$$\sup_{t\in S_m+S_{m'},\|t\|=1}\|\phi_{t,1}\|_\infty \leqslant \sqrt{\Delta(m'')} := M_1. \quad (3.38)$$

The case of v is more tedious,

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

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

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

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

1. **Ordinary smooth case :** In this case, we have by (3.37) and by (3.25) that $H^2 \asymp \frac{m''^{\gamma+1/2}}{n}$. Moreover,

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

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

$$\mathbb{E}\left[\left(\sup_{t\in S_{m''},\ \|t\|=1}(\nu_{n,1}(t))^2 - \ \mathrm{p}(m,m')\right)_+\right] \lesssim \left[U(m'') + V(m'')\right], \tag{3.39}$$

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

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

$$\begin{aligned} V(m'') =& \frac{M_1^2}{C(\varepsilon)^2 n^2}\exp\left(-K_1'C(\varepsilon)\frac{1}{\sqrt{2}}\frac{nH}{M_1}\right) = C_1\frac{\Delta(m'')}{n^2}\exp\left(-C_2 n\frac{\sqrt{\frac{\Delta(m'')}{n}}}{\sqrt{\Delta(m'')}}\right) \\ \lesssim& \frac{1}{n}e^{-C_2\sqrt{n}}, \end{aligned}$$

because for $m \in \mathcal{M}_n$, $\Delta(m) \leqslant n$. Therefore, we deduce by (3.39) that :

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

As

$$\begin{aligned} \sum_{m'}U(m'') \lesssim \frac{1}{n}\sum_{m'}m''^{\gamma}e^{-\frac{K_1}{2c}\sqrt{m''}} =& \frac{1}{n}\left[\sum_{m'=0}^{m}m''^{\gamma}e^{-\frac{K_1}{2c}\sqrt{m''}} + \sum_{m'=m}^{n^2}m''^{\gamma}e^{-\frac{K_1}{2c}\sqrt{m''}}\right] \\ =& \frac{1}{n}\left[m^{\gamma+1}e^{-\frac{K_1}{2c}\sqrt{m}} + \sum_{m'=m}^{+\infty}m'^{\gamma}e^{-\frac{K_1}{2c}\sqrt{m'}}\right] \leqslant \frac{C_1'}{n}, \end{aligned}$$

and

$$\sum_{m'\in\mathcal{M}_n}V(m'') \lesssim \frac{1}{n}\sum_{m'\in\mathcal{M}_n}e^{-C_2\sqrt{n}} = \frac{1}{n}|\mathcal{M}_n|e^{-C_2\sqrt{n}} \lesssim ne^{-C_2\sqrt{n}} \leqslant \frac{C_1''}{n}.$$

We deduce that

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in S_{m''},\ ||t||=1}(\nu_{n,1}(t))^2 - \mathrm{p}(m,m')\right)_+\right] \leqslant \frac{\Sigma_1}{n}, \quad \Sigma_1 = C_1' + C_1''. \tag{3.40}$$

2. **Super smooth case :** In this case the order of H^2 is given by (3.25) : $H^2 \asymp \frac{m''^{\frac{1-\delta}{2}}e^{\mu l^{\frac{\delta}{2}}m''^{\frac{\delta}{2}}}}{n}$,

$$\begin{aligned} \mathrm{Var}(\phi_{t,1}(Z_1)) \leqslant c_1\int|t^*(u)|^2e^{\mu|u|^{\delta}}\mathbb{1}_{|u|\leqslant\sqrt{lm''}}du \leqslant c_1e^{\mu l^{\frac{\delta}{2}}m''^{\frac{\delta}{2}}}\int|t^*(u)|^2du \\ = 2\pi c_1e^{\mu l^{\frac{\delta}{2}}m''^{\frac{\delta}{2}}}\|t\|^2 \lesssim e^{\mu l^{\frac{\delta}{2}}m''^{\frac{\delta}{2}}} = v. \end{aligned}$$

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

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

$$V(m'') = \frac{M_1^2}{C^2(\varepsilon)n^2}\exp\left(-K_1'C(\varepsilon)\sqrt{\varepsilon}\frac{nH}{M_1}\right) = \frac{\Delta(m'')}{C^2(\varepsilon)n^2}\exp\left(-K_1'C(\varepsilon)\sqrt{\varepsilon}\sqrt{n}\right)$$
$$\leqslant \frac{1}{C^2(\varepsilon)n}\exp\left(-K_1'C(\varepsilon)\sqrt{\varepsilon}\sqrt{n}\right).$$

• **Study of** $\sum_{m'\in\mathcal{M}_n}U(m'')$: we have

$$\sum_{m'\in\mathcal{M}_n}U(m'') \lesssim \frac{1}{n}\sum_{m'\in\mathcal{M}_n}e^{\mu l^{\frac{\delta}{2}}m''^{\frac{\delta}{2}}-K_1\varepsilon m''^{\frac{1-\delta}{2}}}.$$

We are going to study this term according the value of δ.

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

$$\frac{1}{n}\sum_{m'\in\mathcal{M}_n}e^{\mu l^{\delta}m''^{\frac{\delta}{2}}-K_1\varepsilon m''^{\frac{1-\delta}{2}}} = \frac{1}{n}\left[\sum_{m'=1}^{m}e^{\mu l^{\delta/2}m''^{\frac{\delta}{2}}-K_1\varepsilon m''^{\frac{1-\delta}{2}}} + \sum_{m'=m}^{n^2}e^{\mu l^{\delta/2}m''^{\frac{\delta}{2}}-K_1\varepsilon m''^{\frac{1-\delta}{2}}}\right]$$
$$\leqslant \frac{1}{n}\left[me^{\mu l^{\delta/2}m^{\frac{\delta}{2}}-K_1\varepsilon m^{\frac{1-\delta}{2}}} + \sum_{m'\in\mathcal{M}_n}e^{\mu l^{\delta/2}m'^{\frac{\delta}{2}}-K_1\varepsilon m'^{\frac{1-\delta}{2}}}\right]$$
$$\leqslant \frac{C_1''}{n}. \tag{3.41}$$

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

$$\frac{1}{n}\sum_{m'\in\mathcal{M}_n}e^{\mu l^{\delta/2}m''^{\frac{\delta}{2}}-K_1\varepsilon m''^{\frac{1-\delta}{2}}} = \frac{1}{n}\sum_{m'\in\mathcal{M}_n}e^{-\mu l^{\delta/2}m''^{\frac{\delta}{2}}} \leqslant \frac{1}{n}\sum_{m'}e^{-\mu l^{\delta/2}m'^{\frac{\delta}{2}}} \leqslant \frac{C_1''}{n}. \tag{3.42}$$

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

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

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

$$\sum_{m'\in\mathcal{M}_n}V(m'') \leqslant \frac{|\mathcal{M}_n|}{C^2(\varepsilon)n}\exp\left(-K_1'C(\varepsilon)\varepsilon\sqrt{n}\right) \leqslant \frac{n}{C^2(\varepsilon)}\exp\left(-K_1'C(\varepsilon)\sqrt{\varepsilon}\sqrt{n}\right)$$
$$\leqslant \frac{C_1'}{n}. \tag{3.43}$$

Therefore, (3.40) holds and the result of Lemma 3.7.2 is proven. □

Proof of Lemma 3.7.3.

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

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

By (3.4)-(3.5) and using the Cauchy-Schwarz inequality, we have :

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

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

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

Now, we have

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

Thus, by this last inequality we deduce

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

If f_ε is ordinary smooth, the integral is convergent and the previous bound is of order $1/n$. Assume now f_ε super smooth, we have by (3.6) :

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

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

3.8 Appendix

3.8.1 Covariance inequality (Viennet (1997))

Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary absolutely process with $\beta-$missing sequence $(\beta_k)_{k\geqslant 0}$. Then, there exists a sequence of measurable functions $(b_k)_{k\geqslant 0}$, with $b_0 \equiv 1$, $0 \leqslant b_k \leqslant 1$, $\mathbb{E}_P[b_k] = \beta_k$ such that for any measurable function f in $\mathbb{L}^2(P)$ and any positive integer n, we have

$$\mathrm{Var}(\sum_{i=1}^{n} f(X_i)) \leqslant 4n\int b(x)f^2(x)dP(x),$$

where $b = \sum_{k=0}^{n} b_k$ is such that $\mathbb{E}_P(b^p) \leqslant p\sum_{k\geqslant 0}(k+1)^{p-1}\beta_k$, for $1 \leqslant p < +\infty$ (see Lemma 4.2 in Viennet (1997)) p. 481).

3.8.2 Asymptotic Askey and Wainger formula

From Askey and Wainger (1965), we have for $\nu = 4k + 2\delta + 2$, and k large enough

$$|\psi_k^{(\delta)}(x/2)| \leqslant C\begin{cases} a) & (x\nu)^{\delta/2} & \text{if } 0 \leqslant x \leqslant 1/\nu \\ b) & (x\nu)^{-1/4} & \text{if } 1/\nu \leqslant x \leqslant \nu/2 \\ c) & \nu^{-1/4}(\nu - x)^{-1/4} & \text{if } \nu/2 \leqslant x \leqslant \nu - \nu^{1/3} \\ d) & \nu^{-1/3} & \text{if } \nu - \nu^{1/3} \leqslant x \leqslant \nu + \nu^{1/3} \\ e) & \nu^{-1/4}(x-\nu)^{-1/4}e^{-\gamma_1\nu^{-1/2}(x-\nu)^{3/2}} & \text{if } \nu + \nu^{1/3} \leqslant x \leqslant 3\nu/2 \\ f) & e^{-\gamma_2 x} & \text{if } x \geqslant 3\nu/2 \end{cases}$$

where γ_1 and γ_2 are positive and fixed constants.

3.8.3 Talagrand's inequality.

Let $(X_i)_{1\leqslant i\leqslant n}$ be independent real random variables, $\mathcal{F}$ a class at most countable of measurable functions and $\nu_n(f) = \frac{1}{n}\sum_{i=1}^{n}(f(X_i) - \mathbb{E}[f(X_i)])$ for all $f \in \mathcal{F}$. We assume there exist third strictly positive constants M_1, H, v such that :

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

Then, for $\varepsilon > 0$,

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

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

Chapitre 4

Hermite estimation in noisy convolution model

Un article écrit à partir des éléments de ce chapitre est prévu.

Résumé. Dans ce chapitre, nous étudions le problème d'estimation d'une fonction de régression dans un modèle de convolution. Nous considérons le modèle suivant : $y(x_k) = h(x_k)+\varepsilon_k, h(x) = f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy, k = -n, \dots, n-1$ où g est supposée connue et f est la fonction inconnue que l'on cherche à estimer ; les erreurs $(\varepsilon_k)_{-n\leqslant k\leqslant n-1}$ sont indépendantes et identiquement distribuées (i.i.d.) avec $\mathbb{E}[\varepsilon_k] = 0$ et $\mathrm{Var}(\varepsilon_k) = \sigma_\varepsilon^2 < \infty$, connu ; les points $(x_k = kT/n)_{-n\leqslant k\leqslant n-1}$ sont déterministes et équirépartis sur $[-T, T]$, où $0 < T < \infty$ est fixé. Nous introduisons deux procédures d'estimation de f en exploitant les propriétés de la base d'Hermite. Nous proposons une étude du risque quadratique de chaque estimateur. Nous obtenons des vitesses de convergence pour f dans une boule de Sobolev pour la première approche ou dans une boule de Sobolev-Hermite pour la deuxième méthode. Nous présentons aussi une procédure de sélection de modèles en s'inspirant des méthodes de Goldenshluter et Lepski pour la première approche d'estimation : l'estimateur résultant satisfait une inégalité oracle pour ε sous-gaussienne. Enfin, nous illustrons numériquement cette méthode et une nouvelle procédure dérivée des méthodes de Goldenshluter et Lepski en s'inspirant de l'approche Lacour et al. (2017).

Abstract. In this chapter, 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 function of interest 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 are considered to estimate f 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 for the first approach inspired by Goldenshluter and Lepski methods, the

resulting estimator satisfies an oracle inequality for ε's sub-Gaussian. Finally, we illustrate numerically this approach and a novel method inspired by Lacour et al. (2017)'s procedure.

Sommaire

4.1 Introduction

Consider the convolution model

$$y(x_k) = h(x_k) + \varepsilon_k, \quad k = -n, \dots, n-1, \tag{4.1}$$

where

$$h(x) = f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy, \tag{4.2}$$

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 (2020a) 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). Model (4.1) cumulates the two questions of regression and deconvolution, and this is why it is difficult. We mention that in Model (4.1), 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 (4.1) 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 (4.1) 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 (4.1) 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 (4.2) 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 previous studies were conducted for $\mathbb{R}^+$ supported f and g. The novelty of present work, is that we consider Model (4.1) with $\mathbb{R}$-supported function 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 chapter, 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 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.

The plan of the chapter is the following : In Section 4.2, we present a first naive approach to estimate f and explain why it is not consistent. The study of the estimation of regression function h in the Hermite basis for fixed design is described in Section 4.3. Those results are exploited to study the FH and HH strategies. Section 4.4 is devoted to

the FH strategy. In particular, we define the FH estimator in Section 4.4.1. A bias-variance decomposition is given in Section 4.4.2. In Section 4.4.3, we provide rates of convergence. Section 4.4.4 is devoted to selection of model for the FH procedure and an oracle inequality is proved for the resulting estimator therein. In Section 4.5, we describe the HH estimation strategy and a comparison with the FH method is performed. Section 4.6 is devoted to the numerical study to illustrate the performance of the adaptive procedure. Finally, all the proofs are presented in Section 4.7 and some useful results are given in the Appendix.

4.2 A first naive approach

Consider discrete observations $(x_k, y(x_k))_{-n\leqslant k\leqslant n-1}$ from model (4.1). Our aim is to estimate f using Fourier analysis tools. 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

$$c_1'(1+t^2)^{\gamma} \leqslant |g^*(t)|^{-2} \leqslant c_1(1+t^2)^{\gamma}, \quad \forall t \in \mathbb{R}. \tag{4.3}$$

(**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. As $h = f \star g$ (see (4.2)), under (**A1**), (**A2**) and using the Fourier inversion formula, we have :

$$f(x) = \int_{\mathbb{R}} e^{-iux}\frac{h^*(u)}{g^*(u)}du, \quad \forall x \in \mathbb{R} \tag{4.4}$$

Equation (4.4) leads to an estimator of f by replacing h by an estimator. A simple idea is to use the following approximation :

$$\frac{T}{n}\sum_{j=-n}^{n-1} e^{itx_j}h(x_j) = \int_{-T}^{T} e^{itx}h(x)dx + \mathcal{O}\left(\frac{T^2}{n}\right). \tag{4.5}$$

Then, we can estimate h^* by :

$$\widetilde{h}^*(t) = \frac{T}{n}\sum_{j=-n}^{n-1} e^{itx_j}y(x_j). \tag{4.6}$$

This brings us to the definition of the following estimator :

$$\widetilde{f}_\ell(x) = \frac{1}{2\pi}\int_{-\ell}^{\ell} e^{-iux}\frac{\widetilde{h}^*(u)}{g^*(u)}du, \text{ for } \ell > 0. \tag{4.7}$$

The parameter of cut-off ℓ is introduced to make the ratio $\widetilde{h}^*/g^*$ integrable.

In the following, $\|s\|_\infty = \sup_{x\in\mathbb{R}}|s(x)|$ denotes the supremum norm of s on $\mathbb{R}$, s' the first derivative of s. We can state the following bound for the risk of $\widetilde{f}_\ell$.

Proposition 4.2.1. *Let Assumptions* (**A1**) *and* (**A2**) *hold and assume that* $\|h\|_\infty < +\infty$, $\|h'\|_\infty < +\infty$. *Let* $\widetilde{f}_\ell$ *be defined in* (4.7) *and set*

$$\Lambda(\ell) = \frac{1}{\pi}\int_{-\ell}^{\ell}\frac{du}{|g^*(u)|^2}.$$

Then, we have

$$\mathbb{E}[\|\widetilde{f}_\ell - f\|^2] \leqslant \int_{|u|>\ell}|f^*(u)|^2du + \Lambda(\ell)\sigma_\varepsilon^2\frac{T^2}{n} + \Lambda(\ell)\frac{T^4}{n^2}(\ell\|h\|_\infty + \|h'\|_\infty)^2 + \Lambda(\ell)\left(\int_{|u|>T}|h(x)|dx\right)^2. \tag{4.8}$$

— The first term of the bound (4.8), $\int_{|u|>\ell}|f^*(u)|^2du$ is the classical bias term.

— The second ($\sigma_\varepsilon^2 T\Lambda(\ell)/n$) is the standard variance term for deconvolution problems.

— The third term $\Lambda(\ell)\frac{T^4}{n^2}\left(\ell\|h\|_\infty + \|h'\|_\infty\right)^2$ comes from of the approximation error given in (4.5). If we consider the following collection of models $\{\ell : \frac{T^2}{n}\Lambda(\ell) \lesssim 1\}$ it has the order of variance term. Indeed, under (**A3**), we have $\Lambda(\ell) \gtrsim \ell^2$ for $\gamma \geqslant \frac{1}{2}$, then $\Lambda(\ell)\left(\frac{T^4}{n^2}\left(\ell\|h\|_\infty + \|h'\|_\infty\right)^2\right) \lesssim \frac{T^2}{n}\ell^2 \lesssim \frac{T^2}{n}\Lambda(\ell)$, where, for two functions u, v, we denote by $u(x) \lesssim v(x)$ if $u(x) \leqslant cv(x)$, with c is a constant independent of x.

— Finally, for fixed T, the term $\Lambda(\ell)(\int_{|u|>T}|h(x)|dx)^2$ does not tend to zero when n tends infinity, whatever the choice of ℓ is. Consequently, $\widetilde{f}_\ell$ is not consistent for fixed T.

Then, we propose in sequel two estimations procedures : Fourier-Hermite strategy and Hermite-Hermite strategy. First, we study the estimation of h by using the least squares methods.

4.3 Hermite fixed design regression

We present a complete study of the regression function in Hermite basis. Let us start by recalling the definition of Hermite basis and the associated regularity space.

4.3.1 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 $\langle \varphi, \psi \rangle = (2\pi)^{-1}\langle \varphi^*, \psi^* \rangle$.

4.3.2 The Hermite basis

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

$$\varphi_j(x) = c_j H_j(x) e^{-x^2/2}, \quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j}(e^{-x^2}), \quad c_j = (2^j j! \sqrt{\pi})^{-1/2}, \quad x \in \mathbb{R},\ j \geqslant 0. \tag{4.9}$$

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

$$\|\varphi_j\|_\infty = \sup_{x\in\mathbb{R}} |\varphi_j(x)| \leqslant \phi_0, \text{ with } \phi_0 = \pi^{-1/4}, \tag{4.10}$$

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

$$\|\varphi_j\|_\infty \leqslant \frac{C_\infty}{(j+1)^{\frac{1}{12}}}, \tag{4.11}$$

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

$$\varphi_j^* = \sqrt{2\pi}(i)^j \varphi_j. \tag{4.12}$$

From Askey and Wainger (1965) or Markett (1984), it holds :

$$|\varphi_j(x)| \leqslant C'_\infty e^{-\xi x^2}, \quad |x| \geqslant \sqrt{2j+1}, \tag{4.13}$$

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) :

$$\|\varphi'_j\|_\infty \leqslant C''_\infty (j+1)^{\frac{5}{12}}, \quad j \geqslant 0, \tag{4.14}$$

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

4.3.3 Regularity spaces

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

Definition 4.3.1. *Let $s, L > 0$, define the Sobolev-Hermite ball of regularity s by*

$$W_H^s(L) = \{\theta \in \mathbb{L}^2(\mathbb{R}), \sum_{k\geqslant 0} k^s a_k^2(\theta) \leqslant L\}, \text{ where } a_k(\theta) = \int \theta(x)\varphi_k(x)dx. \tag{4.15}$$

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

$$W^s(L) = \{\theta \in \mathbb{L}^2(\mathbb{R}), \int (1+u^2)^s |\theta^*(u)|^2 du < L\}. \tag{4.16}$$

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.

4.3.4 Definition of the regression estimator

Let $d \geqslant 1$ an integer and

$$S_d := \text{span}\{\varphi_0, \ldots, \varphi_{d-1}\}, \tag{4.17}$$

the linear space generated by $\varphi_0, \ldots, \varphi_{d-1}$, where φ_j is the Hermite basis defined in (4.9). 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 :

$$\Phi_d = (\varphi_j(x_i))_{-n\leqslant i\leqslant n-1, 0\leqslant j\leqslant d-1}, \quad \Psi_d = \frac{T}{n}\Phi_d^t\Phi_d, \tag{4.18}$$

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

Lemma 4.3.1. *For all $n \geqslant d$, Ψ_d is invertible.*

By the least squares method and Lemma 4.3.1, we derive the following projection estimator of h on S_d :

$$\widehat{h}_d = \sum_{j=0}^{d-1} \widehat{b}_j^{(d)}\varphi_j, \text{ where } \vec{\widehat{b}}^{(d)} = (\widehat{b}_0^{(d)}, \ldots, \widehat{b}_{d-1}^{(d)})^t = (\Phi_d^t\Phi_d)^{-1}\Phi_d^t\vec{y} = \frac{T}{n}\Psi_d^{-1}\Phi_d^t\vec{y}, \tag{4.19}$$

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

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 $h \in \mathbb{L}^2(\mathbb{R})$.

4.3.5 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 4.3.1. *Let $(x_i, y(x_i))_{-n\leqslant i\leqslant n-1}$ be observations from model (4.1). Assume that h belongs to $\mathbb{L}^2(\mathbb{R})$ and consider the estimator $\widehat{h}_d$ defined in (4.19).*

(i) Then, it holds that

$$\mathbb{E}\left[\|\widehat{h}_d - h\|_n^2\right] = \inf_{t\in S_d}\|t-h\|_n^2 + \sigma_\varepsilon^2 T\frac{d}{n}. \tag{4.20}$$

(ii) Moreover, we have

$$\mathbb{E}[\|\widehat{h}_d - h\|^2] \leqslant \|h-h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h-h_d\|_n^2 + \sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right), \tag{4.21}$$

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 4.3.1 corresponds to a classical bias-variance decomposition for the empirical norm $\|\cdot\|_n$. The first term in the right-hand side of (4.20) 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 (4.21) 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.

When the points $(x_i)_{-n\leqslant i\leqslant n-1}$ are i.i.d. random variables with common density μ (see Comte and Genon-Catalot (2020a)), we have $\mathbb{E}[\inf_{t\in S_d}\|t-h\|_n^2] \leqslant \inf_{t\in S_d}\left[\int(h-t)^2(x)\mu(x)dx\right]$ instead of $\inf_{t\in S_d}\|t-h\|_n^2$.
In our context 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 4.3.2. *Assume that h belongs to $W_H^\alpha(L)$ (Sobolev-Hermite ball defined in (4.15)).*

(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

(**A4**) There exists a constant $\lambda_2 > 0$ such that the maximum eigenvalue of Ψ_d^{-1} satisfies

$$\lambda_{\max}(\Psi_d^{-1}) \leqslant \lambda_2 < +\infty,$$

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 4.8.1 in Appendix). It follows that Assumption (**A4**) holds asymptotically with λ_2 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 4.3.2. *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

$$\sup_{h\in W_H^\alpha(L)} \mathbb{E}\left[\|\widehat{h}_{d_{opt}} - h\|_n^2\right] \leqslant C(\alpha, L, T, \sigma_\varepsilon) n^{-\frac{\alpha}{\alpha+1}}, \tag{4.22}$$

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

*(ii) If in addition (**A4**) is satisfied, it yields that*

$$\sup_{h\in W_H^\alpha(L)} \mathbb{E}\left[\|\widehat{h}_{d_{opt}} - h\|^2\right] \leqslant C(\alpha, L, T, \sigma_\varepsilon, \lambda_2) n^{-\frac{\alpha}{\alpha+1}}. \tag{4.23}$$

Our estimator reaches the same rate as in the case where (x_i) are random variables (see Comte and Genon-Catalot (2020a)). 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 4.1. *The constraint $\alpha > 11/6$ or $\alpha > 17/6$ comes from the study of $\|h - h_d\|_n^2$ (see the Proof of Lemma 4.3.2). 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 (4.10). 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 (4.22). 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.

4.3.6 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 :

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

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 :

$$\widehat{d} := \underset{d \in \mathcal{M}_n}{\arg\min} \{\gamma_n(\widehat{h}_d) + \text{pen}(d)\}, \quad \text{where} \quad \text{pen}(d) = \kappa T \frac{d}{n} \sigma_\varepsilon^2, \quad \kappa > 1 \tag{4.24}$$

where $\mathcal{M}_n = \{1, \dots, d_{\max}\}$, $d_{\max} \leqslant n$ 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 4.3.1. *Let $(x_i, y(x_i))_{-n \leqslant i \leqslant n-1}$ be observations, from model (4.1). Assume that $\mathbb{E}[\varepsilon_1^8] < \infty$.*

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

$$\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|_n^2] \leqslant C(\kappa) \inf_{d \in \mathcal{M}_n} \left(\inf_{t \in S_d} \|t - h\|_n^2 + \sigma_\varepsilon^2 T \frac{d}{n} \right) + \frac{C'T}{n}, \tag{4.25}$$

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 (**A4**) holds, we have*

$$\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|^2] \leqslant C_1 \inf_{d \in \mathcal{M}_n} \left((2\lambda_2^2 + 1)\|h - h_d\|_n^2 + \|h_d - h\|^2 + \sigma_\varepsilon^2 T \frac{d}{n} \right) + \frac{C_1' \lambda_2 T}{n}, \tag{4.26}$$

*where λ_2 is given in (**A4**), $C_1 = \max(1, 2\lambda_2^2 C(\kappa))$ 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 $\lambda_2 C_1' T/n$ are residual terms. Indeed, for $h \in W_H^\alpha(L)$, we deduce from Proposition 4.3.2 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 4.3.1 is a consequence of Theorem 3.1 given in Baraud (2000) and the bound given in (4.21).

Remark 4.2. *The variance σ_ε^2 of the noise which appears in (4.24) is assumed to be known but is in general unknown and must be estimated. A classical estimator is the residual least squares estimator :*

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

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

4.3.7 Illustration for the regression estimator in Hermite basis

We illustrate the adaptive procedure of h. We compute the estimator $\widehat{h}_{\widehat{d}}$ defined in (4.19) with $\widehat{d}$ selected in (4.24). 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) $f(x) = \frac{4}{\sqrt{2\pi}}(0.4\exp(-8(x+1)^2) + 0.6\exp(-8(x-1)^2))$, $I = [-2.5, 2.5]$,

(iv) $f(x) = x^2\exp(-x)1_{x\geqslant 0}$, $I = [0, 8]$.

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 values of the sample sizes $n = 250, 1000$. The maximal dimension of d is $d_{\max} = 50$. The regression $h = f \star g$ is computed for each function test f and kernel g by Riemann sum discretization in 500 points. Computations of risk for different values of κ allow to fix $\kappa = 2.25$. The adaptive procedure is implemented as follows :

- For each d in $\mathcal{M}_n$, compute $-\|\widehat{h}_d\|_n^2 + \mathrm{pen}(d)$, with $\mathrm{pen}(d)$ given in (4.24),
- Select $\widehat{d}$ such that $\widehat{d} = \underset{d\in\mathcal{M}_n}{\arg\min}\{-\|\widehat{h}_d\|_n^2 + \mathrm{pen}(d)\}$,
- Compute $\widehat{h}_{\widehat{d}} = \sum_{j=0}^{\widehat{d}-1}\widehat{b}_j^{(\widehat{d})}\varphi_j$, where $\widehat{b}_j^{(\widehat{d})}$ is given in (4.19),

Simulation results and comments

We present in Table 4.1 simulation results. For each function f, we provide the MISE (with the standard deviations in the parenthesis) on the first line computed over 200 repetitions using Riemann's sum discretization for the estimation of h. The second line corresponds to the average of $\widehat{d}$ selected by the algorithm. In the third line, we give also the mean of the theoretical value of Signal-to-noise ratio $s2n$ which 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 remark that increasing n and $s2n$ improve estimations. We also observe that the mean of $\widehat{d}$ increases

with n. Note also in view of the bias-variance decomposition given in (4.21), we can select d as follows :

$$\widehat{d}^{(1)} := \arg\min_{d\in\mathcal{M}_n}\{-\|\widehat{h}_d\|^2 + \mathrm{pen}(d)\}, \quad \text{where} \quad \mathrm{pen}(d) = \kappa\sigma_\varepsilon^2 T\frac{d}{n}. \tag{4.27}$$

The term $-\|\widehat{h}_d\|^2$ estimates $\|h-h_d\|^2 = \|h\|^2 - \|h_d\|^2$ since $\|h\|^2$ is independent on d and can be dropped. Numerically, the two choices of d given in (4.24) and (4.27) provide exactly the same result in average. That is in accordance with the theory (see Lemma 4.3.2).
We also illustrate the results by some graphs. Figure 4.1 and 4.2 show 25 beams of estimators of $h = f \star g$. For each graph, we plot the true function h and 25 estimates for $n = 250$ on the first line and $n = 1000$ on the last. We choose $\sigma_\varepsilon = 1/4$ on the first column and $\sigma_\varepsilon = 1/8$ on the second. Moreover, they ensure the stability of the procedure with some variance for $n = 250$ and $\sigma_\varepsilon = 1/4$. This confirms the results obtained in Table 4.1.

	$\sigma_\varepsilon = \frac{1}{8}$		$\sigma_\varepsilon = \frac{1}{4}$		
f \ n	250	1000	250	1000	
(i)	$0.61_{(0.33)}$	$0.17_{(0.08)}$	$2.24_{(1.24)}$	$0.60_{(0.31)}$	MISE
	7.44	8.98	5.49	7.42	$\bar{\widehat{d}}$
	2.37	2.36	0.60	0.59	$s2n$
(ii)	$0.34_{(0.19)}$	$0.11_{(0.05)}$	$1.27_{(0.84)}$	$0.37_{(0.24)}$	MISE
	9.39	13.02	6.78	9.52	$\bar{\widehat{d}}$
	1.58	1.57	0.39	0.40	$s2n$
(iii)	$1.14_{(0.34)}$	$0.29_{(0.14)}$	$2.87_{(1.15)}$	$1.13_{(0.35)}$	MISE
	11.93	21.12	8.24	12.95	$\bar{\widehat{d}}$
	1.17	1.17	0.30	0.29	$s2n$
(iv)	$0.89_{(0.26)}$	$0.24_{(0.07)}$	$3.05_{(1.02)}$	$0.89_{(0.29)}$	MISE
	23.24	30.60	17.51	23.23	$\bar{\widehat{d}}$
	2.32	2.30	0.58	0.58	$s2n$

TABLE 4.1 – First line : empirical 100× MISE (with 100×sd) for the estimation of h ; second line : mean of $\widehat{d}$; third line : mean of Signal/Noise ratio computed over 200 independent simulations for $\widehat{h}_{\widehat{d}}$.

4.4 Fourier-Hermite approach for the regression-deconvolution model

In this section, we construct an estimator of f using the Fourier inverse transform and then the least squares estimator.

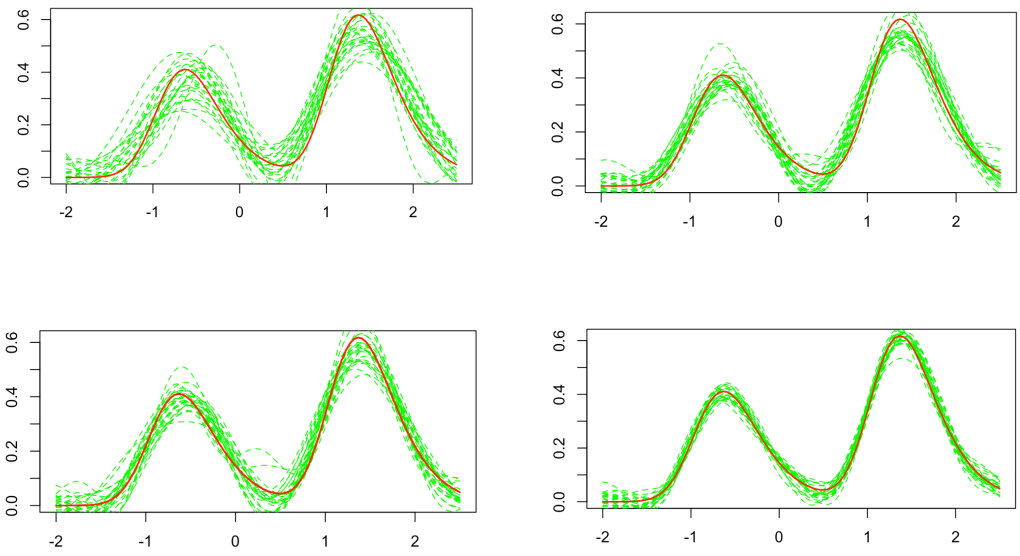


FIGURE 4.1 – 25 estimates of h for (iii), with $n = 250$ (first line) and $n = 1000$ (second line). The true quantity is in bold red and the estimate in dotted lines (left $\sigma_\varepsilon = 1/4$, right $\sigma_\varepsilon = 1/8$).

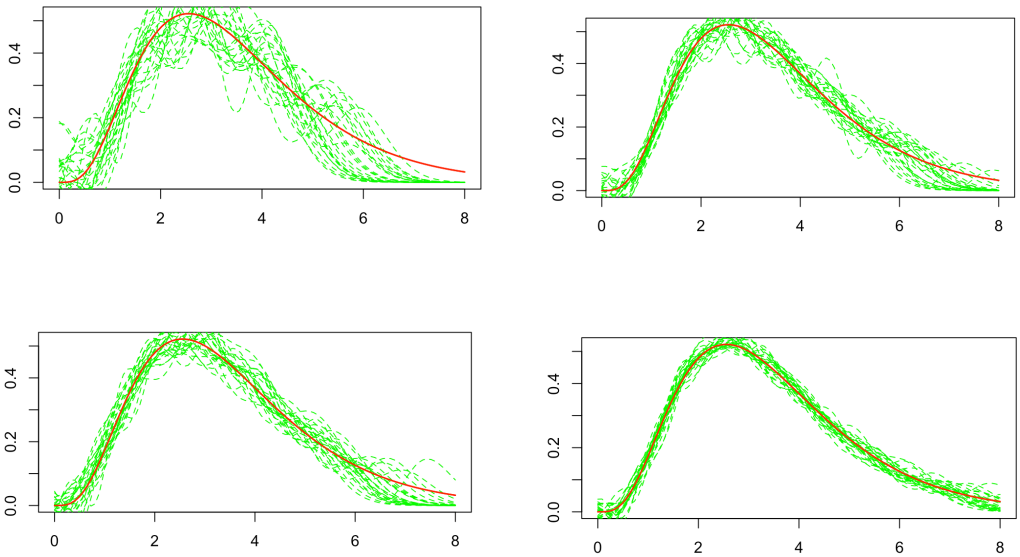


FIGURE 4.2 – 25 estimates of h for (iv), with $n = 250$ (first line) and $n = 1000$ (second line). The true quantity is in bold red and the estimate in dotted lines (left $\sigma_\varepsilon = 1/4$, right $\sigma_\varepsilon = 1/8$).

4.4.1 Estimation procedure

Consider discrete observations $(x_k, y(x_k))_{-n\leqslant k\leqslant n-1}$ from model (4.1). By taking the Fourier transform of (4.19), it yields

$$\widehat{h}_d^*(u) = \sum_{j=0}^{d-1} \widehat{b}_j^{(d)} \varphi_j^*(u). \tag{4.28}$$

Plugging (4.28) in (4.4), we introduce the following estimator of f :

$$\widehat{f}_{(d)}(x) = \frac{1}{2\pi} \int e^{-iux} \frac{\widehat{h}_d^*(u)}{g^*(u)} du. \tag{4.29}$$

The estimator is well defined because the Hermite basis decreases as $e^{-\xi x^2}$ (see (4.13)), 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 4.3. 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

$$\widehat{f}_{(\ell),d}(x) = \frac{1}{2\pi} \int_{-\ell}^{\ell} e^{-iux} \frac{\widehat{h}_d^*(u)}{g^*(u)} du, \text{ for } \ell > 0. \tag{4.30}$$

4.4.2 Risk bound for the deconvolution estimator

Now, we study the integrated quadratic risk of $\widehat{f}_{(d)}$ given by (4.29). Define

$$\Delta(\ell) = \sup_{|u|\leqslant \ell} |g^*(u)|^{-2}, \qquad f_{(\ell)}(x) = \frac{1}{2\pi} \int_{-\ell}^{\ell} e^{-iux} \frac{h^*(u)}{g^*(u)} du, \tag{4.31}$$

Consider also the following assumption :

(**A5**) $\|h\|_\infty = \sup_{x\in\mathbb{R}} |h(x)| < \infty$.

We recall that, by the Cauchy-Schwarz inequality, $\|h\|_\infty \leqslant \|f\|\|g\|$. Therefore, if f and g are square integrable then, condition (**A5**) is automatically satisfied.

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

Proposition 4.4.1. *Suppose that the assumptions* (**A1**) *to* (**A5**) *hold. For* $\widehat{f}_{(d)}$ *given in (4.29),* $\widehat{f}_{(\ell),d}$ *defined in (4.30) and* $\ell \geqslant \sqrt{2d}$*, we have*

$$\mathbb{E}\left[\|\widehat{f}_{(d)} - f\|^2\right] \leqslant 2C\lambda_2 T e^{-\xi d} + 2\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right], \tag{4.32}$$

where C *is a constant depending on* C'_∞, ξ *given in* (4.13), c_1 *in* (**A3**) *and* $\|h\|_\infty$*. For* $\widehat{f}_{(\ell),d}$ *defined in (4.30) and any* $\ell > 0$*, it holds that*

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right] \leqslant \|f - f_{(\ell)}\|^2 + \Delta(\ell)\left(\|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2 + \sigma_\varepsilon^2 \frac{T}{n} \mathrm{tr}\left(\Psi_d^{-1}\right)\right). \tag{4.33}$$

(a) The first term on the right-hand side of (4.33) ($\|f - f_{(\ell)}\|^2 = \frac{1}{2\pi}\int_{|u|>\ell} |f^*(u)|^2 du$) 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 4.3.1 (ii)).

We also mention that the term $C\lambda_2 T e^{-\xi d}$ 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 (4.16) 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.

4.4.3 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 (4.23)) and then for the deconvolution part, by substituting this value in (4.33) and optimizing in ℓ to get the rates of $\widehat{f}_{(\ell),d}$ and $\widehat{f}_{(d)}$. The following result of convergence holds.

Theorem 4.4.1. *Let assumptions* (**A1**) *to* (**A4**) *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* (4.16) *and* γ *is given in* (**A3**).

The same result holds for the estimator $\widehat{f}_{(d_{opt})}$ with the assumption (**A5**), see (4.32). 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.

The hypothesis $h \in W_H^{s+\gamma}(\cdot)$ is related to the regularity of f and g. Conditions ensuring that this assumption is fulfilled are given in Section 4.7.

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 4.4.1. 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 4.3). Therefore, it yields under (**A1**) to (**A4**) 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 4.3. *Assume that f belongs to $W^s(L)$ (see Section 4.3.3) and g is ordinary smooth (i.e. g satisfies (4.3)). Then, h belongs to $W^{s+\gamma}(L/c_1')$, where c_1' is given in (4.3). Indeed, we have*

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

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 $\eta = 0,\dots,s+\gamma-1$ (see Section 4.3.3). The latter is discussed in the proof section, see Proposition 4.7.1.

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 (4.32)) considering (**A5**).

Rate of convergence for f Gaussian

Let

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{4.34}$$

we can establish the following result.

Proposition 4.4.2. *Let assumptions (**A1**) to (**A4**) hold and $f = f_\sigma$ where f_σ is defined in (4.34). 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).

Rate of convergence for Gaussian kernel

By reversing the role of f and g in Proposition 4.4.2, 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 4.4.3. *Let Assumptions (**A1**), (**A3**) and (**A4**) 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*

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

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 4.4.4. *Assume that* (**A1**), (**A2**) *and* (**A4**) *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*

$$\mathbb{E}\left[\|\widehat{h}_{d_{opt}} - h\|^2\right] \lesssim \frac{\log(n)}{n}. \tag{4.35}$$

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 (4.9), 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 4.4.4.

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 4.4.5. *Let* (**A1**),. . .,(**A4**) *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 4.2 :

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

4.4.4 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

$$\widehat{\widehat{f}}_d(x) := \widehat{f}_{(\sqrt{2d}),d}(x) = \frac{1}{2\pi}\int_{-\sqrt{2d}}^{\sqrt{2d}} e^{-iux}\frac{\widehat{h}_d^*(u)}{g^*(u)}du. \tag{4.36}$$

This choice of ℓ is motivated by the results obtained in Proposition 4.4.1 and Theorem 4.4.1. 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

$$\mathcal{M}_n^{(1)} := \{1 \leqslant d \leqslant n, \quad \frac{\sigma_\varepsilon^2 \lambda_2 T d \Delta(\sqrt{2d})}{n} \leqslant 1\}$$

where $\Delta(d)$ is given by (4.31) and λ_2 in (**A4**). Define

$$\widehat{A}(d) := \max_{d' \in \mathcal{M}_n^{(1)}} \left\{ \left(\|\widehat{\widehat{f}}_{d'} - \widehat{\widehat{f}}_{d\wedge d'}\|^2 - \kappa_1 V(d') \right)_+ \right\}, \tag{4.37}$$

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

$$V(d) = 2\left(1 + 24\log(n)\right)\sigma_\varepsilon^2 \Delta(\sqrt{2d})\frac{\lambda_2 dT}{n}. \tag{4.38}$$

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

$$\widehat{d} := \arg \min_{d \in \mathcal{M}_n^{(1)}} \left\{ \widehat{A}(d) + \kappa_2 V(d) \right\}, \tag{4.39}$$

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

(**A6**) 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 (**A6**) is also satisfied if ε_1 is bounded.

The following non asymptotic result holds for $\widehat{\widehat{f}}_{\widehat{d}}$.

Theorem 4.4.2. *Let assumptions* (**A1**) *to* (**A4**) *and* (**A6**) *hold,* $\widehat{\widehat{f}}_d$ *be defined by (4.36),* $\widehat{d}$ *selected by* (4.39)*. Then, for* $\kappa_1 \geqslant 12$*, we have*

$$\mathbb{E}[\|\widehat{\widehat{f}}_{\widehat{d}} - f\|^2] \leqslant C \inf_{d \in \mathcal{M}_n^{(1)}} \left(\|f - f_{(\sqrt{2d})}\|^2 + R_b(d) + V(d) \right) + C' \frac{\log(n)}{n}, \tag{4.40}$$

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* $C' = C'(\mathbb{E}[\varepsilon_1^4], \gamma, c_1, \xi, \lambda_2, C'_\infty)$ *with* c'_1, γ *given in* (**A3**)*,* ξ, C'_∞ *in* (4.13) *and* λ_2 *in* (**A4**)*.*

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*

$$\mathbb{E}[\|\widehat{\widehat{f}}_{\widehat{d}} - f\|^2] \leqslant C_1 \inf_{d \in \mathcal{M}_n^{(1)}} \left(d^{-s} + V(d) \right) + C'_1 \frac{\log(n)}{n}, \tag{4.41}$$

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 (4.40) and (4.41) 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 $\widehat{\widehat{f}}_{\widehat{d}}$ 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 4.4.1 with n replaced by $n/\log(n)$ that under the assumptions of Theorem 4.4.2

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

where $C > 0$ is a numerical constant.

4.5 Hermite-Hermite strategy for the regression-deconvolution model

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.

4.5.1 Estimation strategy

Let $(x_k, y(x_k))_{-n\leqslant k\leqslant n-1}$ from model (4.1), $m \geqslant 1$, integer and consider S_m defined in (4.17). 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 :

$$a_j(f) = \frac{1}{2\pi}\langle \frac{h^*}{g^*}, \varphi_j^*\rangle = \frac{1}{2\pi}\int \frac{h^*(u)}{g^*(u)}\overline{\varphi_j^*(u)}dx = \frac{(-i)^j}{\sqrt{2\pi}}\int \frac{h^*(u)}{g^*(u)}\varphi_j(u)du. \tag{4.42}$$

Replacing h^* by $\widehat{h}_d^*$ defined in (4.28) and plugging this in (4.42), we define the following estimator :

$$\widehat{f}_{m,d} = \sum_{j=0}^{m-1} \widehat{a}_{j,d}\varphi_j, \quad \widehat{a}_{j,d} = \frac{(-i)^j}{\sqrt{2\pi}}\int \frac{\widehat{h}_d^*(u)}{g^*(u)}\varphi_j(u)du, \tag{4.43}$$

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 (4.13)). 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 4.3).

4.5.2 Risk bound for the projection estimator of f

The following risk bound holds for $\widehat{f}_{m,d}$.

Proposition 4.5.1. *Assume that f and h belong to $\mathbb{L}^2(\mathbb{R})$ and set*

$$\Sigma(m) = \sup_{|u|\leqslant\sqrt{\rho m}} |g^*(u)|^{-2} + \sum_{j=0}^{m-1}\int_{|u|\geqslant\sqrt{\rho m}} |\varphi_j(u)|^2|g^*(u)|^{-2}du, \quad \rho > 0. \tag{4.44}$$

For $\widehat{f}_{m,d}$ given in (4.43), we have

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] \leqslant \|f - f_m\|^2 + 2\Sigma(m)\left(\|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2 + \sigma_\varepsilon^2\frac{T}{n}\left(\mathrm{tr}\left(\Psi_d^{-1}\right) \wedge 2\pi^2 m\right)\right). \tag{4.45}$$

Note that the constant $\rho > 0$ is independent from n, m and d. The same comments given after Proposition 4.4.1 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 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 (4.31). 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).

4.5.3 Rate of convergence of $\widehat{f}_{m,d}$

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

Theorem 4.5.1. *Suppose that* (**A3**), (**A4**) *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* (4.15).

The estimator $\widehat{f}_{m_{opt},d_{opt}}$ achieves the same rate as $\widehat{f}_{(d_{opt})}$ obtained in Theorem 4.4.1. Note that the results for some special functions obtained for $\widehat{f}_{(d_{opt})}$ in Proposition 4.4.2, 4.4.3 and 4.4.4 apply here. For the Gamma case (see Proposition 4.4.5), 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 $\|f - f_\ell\| \asymp \ell^{-2p+1}$ 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

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

4.6 Numerical illustration

4.6.1 Practical implementation

In this section, we present the results of a simulation study to illustrate the performances of our strategies. We take the same test functions and kernel given in regression part (see Section 4.3.7). We set $T = 10$ and consider two values of the samples sizes $n = 250, 1000$. We simulate a Gaussian noise for the error with two noise levels $\sigma_\varepsilon \in \{1/8, 1/4\}$. We recall that the function regression $h = f \star g$ is computed by Riemann sum discretization. We consider the following collection of models $\mathcal{M}_n^{(1)} = \{1, 2, \ldots, 30\}$. 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 :

$$V(d) = 2\,(1 + 24\log(n))\,\sigma_\varepsilon^2(1 + \frac{2d}{\theta^2})\frac{\lambda_2 dT}{n}, \quad \theta = 4, \quad \lambda_2 = 1. \tag{4.46}$$

The adaptive estimator is implemented as follows :

— For each $d \in \mathcal{M}_n^{(1)}$, compute $\widehat{A}(d) = \max_{d' \in \mathcal{M}_n^{(1)}} \left\{ \left(\|\widehat{\widehat{f}}_{d'} - \widehat{\widehat{f}}_{d\wedge d'}\|^2 - \kappa_1 V(d') \right)_+ \right\}$, where the integral $\|\widehat{\widehat{f}}_{d'} - \widehat{\widehat{f}}_{d\wedge d'}\|^2$ is computed by Riemann's approximation and $V(d)$ given in (4.46),

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

In the sequel, this procedure is called « **GL** ».

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) suggest the idea to consider 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 $\kappa_1 = 8 \times 10^{-4}$ and $\kappa_2 = 16 \times 10^{-4}$. Of course, the values given to κ_1 and κ_2 may not be the most relevant but it gives quite satisfactory results. Then, we illustrate the procedure by some graphs.
As **GL** method is slow and therefore difficult to calibrate, we implement the improvement proposed by Lacour et al. (2017), which allows us to perform repetitions and propose risk tables. Lacour et al. (2017)'s strategy is introduced to perform bandwidth selection in the case of kernel estimation of a density, and has the advantage to be faster. Furthermore, we must only calibrate one constant denoted $\kappa^{(1)}$. More precisely, the method (called « **PCO** ») is described as follows

— For each $d \in \mathcal{M}_n^{(1)}$, compute $\widetilde{A}(d) = \left\{ \|\widehat{\widehat{f}}_d - \widehat{\widehat{f}}_{d_{\max}}\|^2 \right\}$, by Riemann's approximation where $d_{\max} = 30$.

— Choose $\widetilde{d}$ via $\widetilde{d} = \arg\min_{d \in \mathcal{M}_n^{(1)}} \left\{ \widetilde{A}(d) + \kappa^{(1)} V(d) \right\}$.

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

Calibration of constant $\kappa^{(1)}$. To find the value of $\kappa^{(1)}$, we have evaluated the MISE for different test functions by varying $\kappa^{(1)}$. These preliminary studies helped us to make a good compromise and then to fix $\kappa^{(1)} = 1.5 \times 10^{-3}$.

4.6.2 Numerical simulation results

First, we illustrate the methods by presenting some pictures. Figure 4.3 presents the true unknown function (the bold red line), thirty estimators of f for all possible dimensions proposed for selection to **GL** method in green dotted lines, and the estimator chosen by

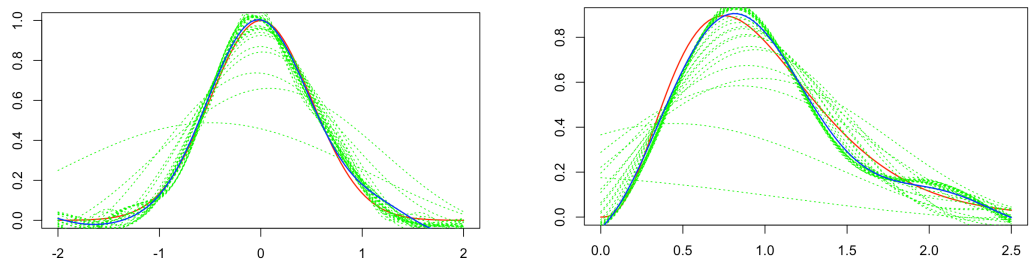
the **GL** procedure in blue bold line for each test function (i), (ii), (iii) and (iv) from left to right. The dimension selected by the procedure is given under each graph. We observe that this choice is often relevant compared to the 30 estimators plots.

In Figures 4.4 and 4.5, we plot the true function in bold red line with 25 estimators in dotted lines for the two last test functions (iii), (iv) by considering the **PCO** algorithm. The first line illustrates the influence of sample size and the second line shows how the noise levels can affect the performance of the estimates. We observe that increasing n improves the estimation and, on the contrary, that increasing the variance of 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 4.1), this effects decreases when the sample size increases. The mean of selected dimensions are given in Table 4.4. We observe that these averages are comparable to the dimension obtained in Figure 4.3 for (iii) and (iv).

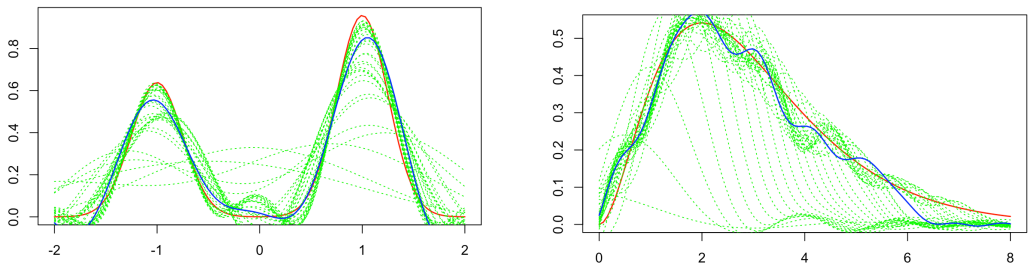
In Table 4.3, we report the values of the MISEs with standard deviation in parentheses multiplied by 100 computed from 100 simulated samples for the estimator $\hat{\hat{f}}_{\widetilde{d}}$ with $\widetilde{d}$ selected using the **PCO** algorithm. We also provide the average of $\widetilde{d}$ selected by the procedure. As for graphical study, we see that increasing the sample size and decreasing the variance of noise (which corresponds to a $s2n$ large see Table 4.1) improve the estimation. When n increases, the average of $\widetilde{d}$ is increasing except in the case of function (i) with $\sigma_\varepsilon = 1/4$. This case corresponds to a $s2n$ equal to 0.60 see Table 4.1 and then the estimation is most difficult ; this can explain why the procedure will seek a large dimension for $n = 250$. Clearly, the influence of signal-to-noise ratio $s2n$ is important, see in Table 4.1 and Figures for graphical analysis.

	$\sigma_\varepsilon = \frac{1}{8}$		$\sigma_\varepsilon = \frac{1}{4}$		
f \ n	250	1000	250	1000	
(i)	$1.37_{(1.19)}$	$0.34_{(0.20)}$	$5.25_{(6.47)}$	$1.07_{(0.64)}$	MISE
	11.49	11.51	10.53	9.47	$\bar{\widetilde{d}}$
(ii)	$1.41_{(1.12)}$	$0.64_{(0.27)}$	$4.55_{(3.99)}$	$1.49_{(0.83)}$	MISE
	14.04	15.36	11.67	12.38	$\bar{\widetilde{d}}$
(iii)	$4.18_{(1.86)}$	$1.22_{(0.58)}$	$8.84_{(3.74)}$	$4.30_{(1.76)}$	MISE
	18.85	23.08	12.99	16.17	$\bar{\widetilde{d}}$
(iv)	$3.69_{(1.56)}$	$1.43_{(0.39)}$	$9_{(0.42)}$	$3.97_{(1.32)}$	MISE
	18.81	22.01	17.44	15.85	$\bar{\widetilde{d}}$

TABLE 4.3 – First line : empirical 100× MISE (with 100×sd) for the estimation of unknown function f computed over 100 independent simulations for $\hat{\hat{f}}_{\widetilde{d}}$; second line : mean of $\widetilde{d}$ selected by the **PCO** algorithm.



$\widehat{d} = 15$ $\widehat{d} = 15$



$\widehat{d} = 21$ $\widehat{d} = 23$

FIGURE 4.3 – Estimators for all possible dimensions in dotted line (green), the chosen estimator by algorithm in bold blue, the true function in red with $n = 1000$, $\sigma_\varepsilon = 1/8$ for the **GL** algorithm.

f	(iii)		(iv)	
n	250	1000	1000	4000
$\sigma_\varepsilon = \frac{1}{8}$	16.20	23.32	22.24	28.24
$\sigma_\varepsilon = \frac{1}{4}$	12.68	16.60	17.88	21.12

TABLE 4.4 – Mean of selected dimensions $\widehat{d}$ presented in Figures 4.4 and 4.5.

4.7 Proofs

In the sequel C denotes a generic constant whose value may change from line to line and whose dependency is sometimes given in indexes.

Proposition 4.7.1. *(Regularity of h if g is ordinary smooth)*
Let s, γ and α integer such that $\alpha = s + \gamma$. Assume that :

(i) f is in $W_H^s(\cdot)$, $f^{(l)}$ and $x^{\alpha-l}f^{(l)}$ for $l = 0, \dots, s$ belong to $\mathbb{L}^1(\mathbb{R})$,

(ii) g is ordinary smooth with parameter γ such that $g \in \mathbb{L}^1(\mathbb{R})$, $x^{\alpha-l}g^{(p)}$ for $p = 0, \dots, \gamma - 2$ belong to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ with $\gamma \geqslant 2$.

Then, $h = f \star g$ belongs to $W_H^{\alpha-1}(\cdot)$. Furthermore, if

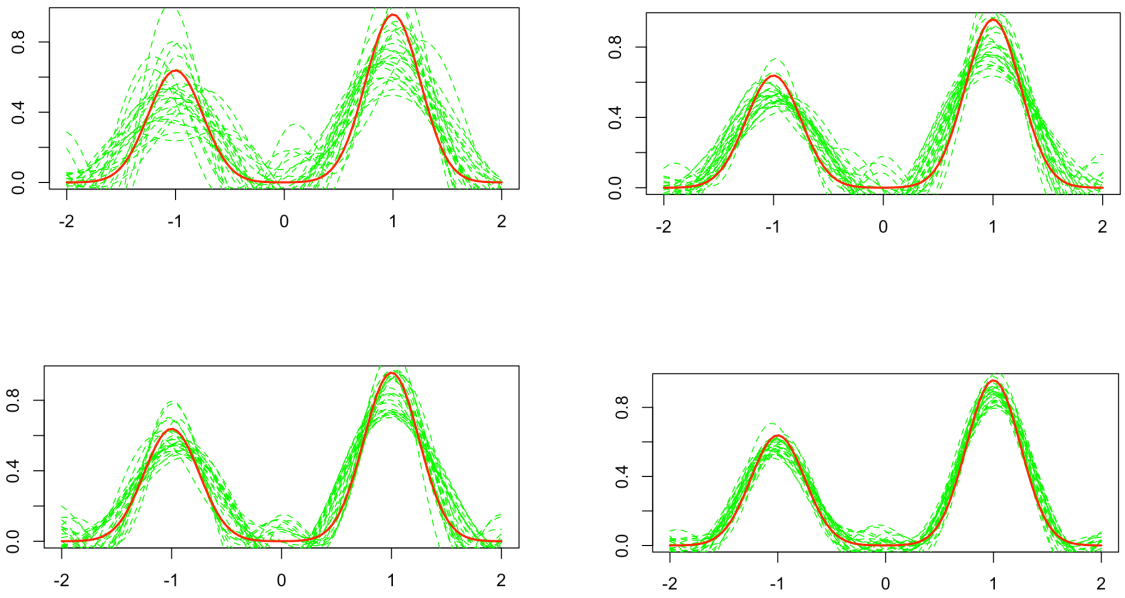


FIGURE 4.4 – 25 estimates of (iii), with $n = 250$ (first line) and $n = 1000$ (second line) for the **PCO** algorithm. The true quantity is in bold red and the estimate in dotted lines (left $\sigma_\varepsilon = 1/4$, right $\sigma_\varepsilon = 1/8$).

(iii) the map $x \mapsto xh^{(\alpha-1)}(x)$ *is square integrable.*

It follows that $h = f \star g \in W_H^\alpha(\cdot)$.

— The assumptions given in (i) are not very restrictive and are checked by many functions : Gaussian, Gamma, Mixed-Gaussian, Mixed-Gamma, Cauchy, Laplace with $s = 0$.

— The conditions on g given in (ii) are verified for the classical ordinary-smooth function : $g(x) = x^{\gamma-1}e^{-\theta x}\mathbb{1}_{x\geqslant 0}$ with $\theta > 0$, $g(x) = e^{-b|x|}$, $b > 0$ and all the symmetric Gamma distributions.

— However, the assumption (iii) is difficult to verify if the regression function cannot be known explicitly. Of course, this is linked to the fact that it is difficult to explicitly compute a convolution product. Indeed, in the simple cases where we can compute $h = f \star g$, this assumption is checked (*e.g.* both functions f and g are of type Gamma with the same scale parameter, Laplace and Gaussian).

Proof of Proposition 4.7.1. Notice that as $f \in \mathbb{L}^1(\mathbb{R})$ and g is bounded, therefore, $h = f \star g$ is well defined and bounded. We known also that h belongs to $W^{s+\gamma}(\cdot)$ which is equivalent to h is $s+\gamma$ times differentiable and $h, h', \ldots, h^{(s+\gamma)}$ are square integrable see Remark 4.3. Then, $h \in W_H^\alpha(\cdot)$ with $\alpha = s+\gamma$ if only and if the functions $x^{\alpha-l}h^{(l)}$ are squared integrable

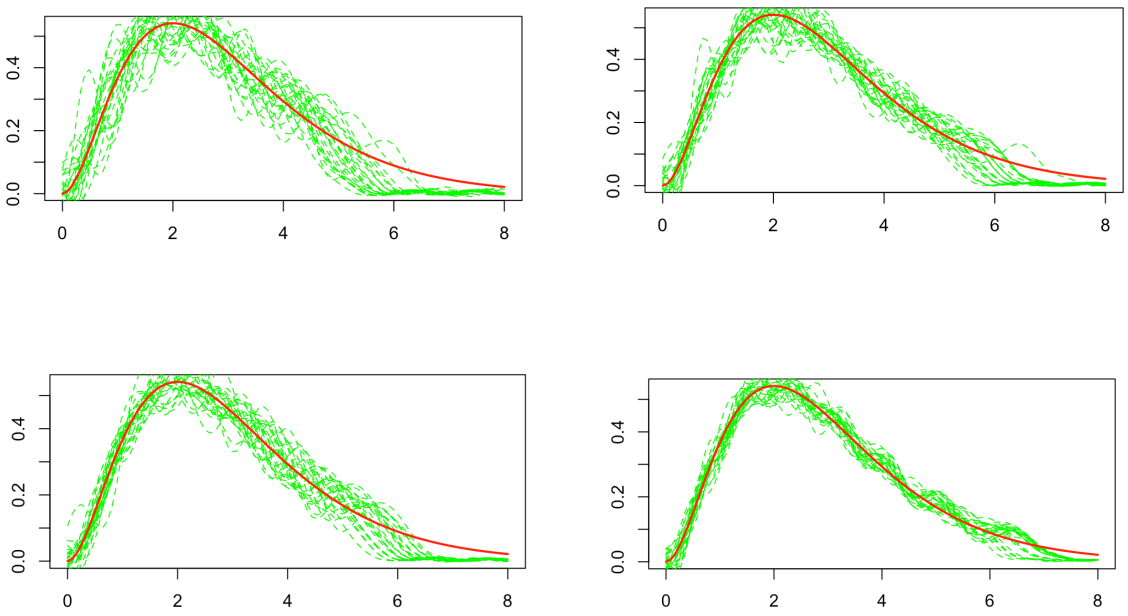


FIGURE 4.5 – 25 estimates of (iv), with $n = 1000$ (first line) and $n = 4000$ (second line) for the **PCO** algorithm. The true quantity is in bold red and the estimate in dotted lines (left $\sigma_\varepsilon = 1/4$, right $\sigma_\varepsilon = 1/8$).

for $l = 0, \dots, \alpha - 1$ (see Section 4.3.3). As

$$\int (x^{\alpha-l} h^{(l)})^2 dx \leqslant \int (h^{(l)}(x))^2 dx + \int (x^\alpha h^{(l)}(x))^2 dx,$$

then, this equivalent to $x^\alpha h^{(\alpha)}$ is squared integrable for $l = 0, \dots, \alpha - 1$ since h^l is squared integrable. We assume now s and γ are integers so that :

$$h^{(l)} = (f \star g)^{(l)} = \begin{cases} f^{(l)} \star g \text{ for } l = 0, \dots, s \\ f^{(s)} \star g^{(p)} \text{ for } l = s + p \text{ and } p = 0, \dots, \gamma - 2. \end{cases} \tag{4.47}$$

Now, we will prove that $h \in W_H^\alpha(L)$. Let us start by the case where $l = 0, \dots, s$. For $l \in \{0, \dots, s\}$, using the Parseveval-Plancherel Equality, we have,

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

For $x^\alpha g \in \mathbb{L}^1(\mathbb{R})$ and $x^\alpha f^{(l)} \in \mathbb{L}^1(\mathbb{R})$, the Fourier transforms of g and $f^{(l)}$ are α differentiable. Using successively Leibniz Formula and Cauchy-Schwarz Inequality, it follows

that,

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

Recall that, let $\chi\in\mathbb{L}^1(\mathbb{R})$ such that $x^p\chi\in\mathbb{L}^1(\mathbb{R})$, then $x^{k'}\chi$ is integrable for $k'=0,\ldots,p$. Indeed, for all $x\in\mathbb{R}$, we have $|x^{k'}\chi(x)|\leqslant(1+|x|^p)\,|\chi(x)|$ and it comes that $\frac{d^{k'}}{du^{k'}}\chi^* = (\chi^*)^{(k')} = (i)^{k'}\left(x^{k'}\chi\right)^*$. The Riemann-Lebesgue Theorem implies that $\|(\chi^*)^{(k')}\|_\infty\leqslant\|x\mapsto x^{k'}\chi\|_1<\infty$. Then, we derive that

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

As $x^\alpha g\in\mathbb{L}^2(\mathbb{R})$, it comes for $0\leqslant k\leqslant\alpha$ that

$$\int|(g^*)^{(k)}(u)|^2du=\frac{1}{2\pi}\int|x^kg(x)|^2dx\leqslant\int_{|x|\leqslant1}|g(x)|^2dx+\int_{|x|\geqslant1}|x^{(\alpha)}g(x)|^2dx<+\infty. \tag{4.48}$$

Consequently, it follows

$$\|x^\alpha h^{(l)}\|^2<+\infty,\quad l=0,\ldots,s. \tag{4.49}$$

Now, we consider the case $l=s+p$, with $p=0,\ldots,\beta-2$. As $x^\alpha g^{(p)}\in\mathbb{L}^1(\mathbb{R})$ and $x^\alpha f^{(s)}\in\mathbb{L}^1(\mathbb{R})$, we derive by the same computations as the case $l=0,\ldots s$, and from (4.47) :

$$\|x^\alpha h^{(l)}\|^2=2\pi\|[(g^{(p)})^*(f^{(s)})^*]^{(\alpha)}\|^2\leqslant C(s,\beta)\sum_{k=0}^{\alpha}\|[(f^{(s)})^*]^{(\alpha-k)}\|_\infty^2\binom{\alpha}{k}\int|[(g^{(p)})^*]^{(k)}|^2du,$$

where $f^{(s)}$ and $g^{(p)}$ play the role of $f^{(l)}$ and g respectively. Using the same decomposition as in (4.48), we get $\int|[(g^{(p)})^*]^{(k)}|^2du<+\infty$ provided that $x^\alpha g^{(p)}\in\mathbb{L}^2(\mathbb{R})$ which is the case here and it holds that

$$\|x^\alpha h^{(l)}\|^2<+\infty,\quad l=s+p,\quad p=0,\ldots,\beta-2. \tag{4.50}$$

We derive from (4.49) and (4.50) that h belongs to $W_H^{\alpha-1}(L')$. As by hypothesis, the maps $x\mapsto xh^{(\alpha-1)}(x)$ is square integrable, we deduce that $h\in W_H^\alpha(L')$. □

4.7.1 Proofs of Section 4.2

Proof of Proposition 4.2.1. We start by introducing f_ℓ defined by :

$$f_\ell(x) = \frac{1}{2\pi}\int_{-\ell}^{\ell} e^{ixu}\frac{h^*(u)}{g^*(u)}du.$$

Then, we have

$$\begin{aligned}\mathbb{E}[\|\widetilde{f}_\ell - f\|^2] &= \|f - f_\ell\|^2 + \mathbb{E}[\|\widehat{f}_\ell - f_\ell\|^2] \\ &= \|f - f_\ell\|^2 + \mathbb{E}[\|\widetilde{f}_\ell - \mathbb{E}[\widetilde{f}_\ell]\|^2] + \|\,\mathbb{E}[\widetilde{f}_\ell] - f_\ell\|^2. \end{aligned} \tag{4.51}$$

We study the two last term of (4.51). For the first one, we have

$$\mathbb{E}[\|\widetilde{f}_\ell - \mathbb{E}[\widetilde{f}_\ell]\|^2] = \frac{1}{2\pi}\int_{-\ell}^{\ell}\frac{\mathbb{E}\left|\widetilde{h}^*(u) - \mathbb{E}[\widetilde{h}^*(u)]\right|^2}{|g^*(u)|^2}du.$$

By the definition of $\widetilde{h}^*$ given in (4.6), it holds that

$$\begin{aligned}\mathbb{E}\left|\widetilde{h}^*(u) - \mathbb{E}[\widetilde{h}^*(u)]\right|^2 &= \mathbb{E}\left[\left|\frac{T}{n}\sum_{j=-n}^{n-1} e^{ix_j u}(y_j - h(x_j))\right|^2\right] = \frac{T^2}{n^2}\,\mathbb{E}\left[\sum_{-n\leqslant j,k\leqslant n-1} e^{iu(x_j - x_k)}\varepsilon_j\varepsilon_k\right] \\ &= \frac{2T^2}{n}\sigma_\varepsilon^2,\end{aligned}$$

since the $(\varepsilon_j)_{-n\leqslant j\leqslant n-1}$ are i.i.d., centered of variance σ_ε^2. From this, it comes that

$$\mathbb{E}[\|\widetilde{f}_\ell - \mathbb{E}[\widetilde{f}_\ell]\|^2] = \frac{\Lambda(\ell)}{n}T^2\sigma_\varepsilon^2. \tag{4.52}$$

Now, we study $\|\,\mathbb{E}[\widetilde{f}_\ell] - f_\ell\|^2 = \frac{1}{2\pi}\int_{-\ell}^{\ell}\frac{|\,\mathbb{E}\,\widehat{h}^*(u) - h^*(u)|^2}{|g^*(u)|^2}du$. We examine $|\,\mathbb{E}\,\widehat{h}^*(u) - h^*(u)|^2$ and write

$$\begin{aligned}\left|\mathbb{E}\,\widetilde{h}^*(u) - h^*(u)\right|^2 &= \left|\frac{T}{n}\sum_{j=-n}^{n-1} e^{iux_j}h(x_j) - \int_{-T}^{T} e^{iux}h(x)dx - \int_{|x|\geqslant T} e^{iux}h(x)dx\right|^2 \\ &\leqslant 2\left|\frac{T}{n}\sum_{j=-n}^{n-1} e^{iux_j}h(x_j) - \int_{-T}^{T} e^{iux}h(x)dx\right|^2 + 2\left|\int_{|x|\geqslant T} e^{iux}h(x)dx\right|^2.\end{aligned}$$

From Lemma 4.8.2, we derive that

$$\left|\frac{T}{n}\sum_{j=-n}^{n-1} e^{iux_j}h(x_j) - \int_{-T}^{T} e^{iux}h(x)dx\right| \leqslant \|\psi_u'\|_\infty\frac{T^2}{n},$$

where $\psi_u(x) = e^{ixu}h(x)$. As $\psi_u'(x) = (iuh(x) + h'(x))e^{iux}$, then, it follows that $|\psi_u'(x)| \leqslant |u|\|h\|_\infty + \|h'\|_\infty$, because $\|h\|_\infty < \infty$ and $\|h'\|_\infty < +\infty$. This implies that

$$\|\,\mathbb{E}[\widetilde{f}_\ell] - f_\ell\|^2 \leqslant \Lambda(\ell)\left[\frac{T^4}{n^2}(\ell\|h\|_\infty + \|h'\|_\infty)^2 + \Big(\int_{|x|\geqslant T} e^{iux}h(x)dx\Big)^2\right]. \tag{4.53}$$

Plugging (4.52) and (4.53) in (4.51) ends the proof. □

4.7.2 Proofs of Section 4.3

Proof of Lemma 4.3.1. Let $\vec{w} = (w_0, \dots w_{d-1})^t$, with $\Psi_d \vec{w} = 0$. Then, it holds

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

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 4.3.1. Denote $\Pi_d h = \Phi_d \vec{b}^{(d)} = \Phi_d(\Phi_d^t \Phi_d)^{-1}\Phi_d^t h(\vec{x})$ with $h(\vec{x}) = (h(x_{-n}), \dots, h(x_{n-1}))^t$ 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

$$\mathbb{E}\left[\|\widehat{h}_d - h\|_n^2\right] = \inf_{t \in S_d}\|t - h\|_n^2 + \mathbb{E}\left[\|\widehat{h}_d - \Pi_d h\|_n^2\right]. \tag{4.54}$$

Then, for $\vec{\widehat{b}}^{(d)}$ given in (4.19), we can write

$$\widehat{h}_d(\vec{x}) = \left(\widehat{h}_d(x_{-n}), \dots, \widehat{h}_d(x_{n-1})\right)^t = \Phi_d \vec{\widehat{b}}^{(d)},$$

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

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

We have

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

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 (4.54) ends the proof of (4.20).

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 (4.28), we get

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

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

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

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

$$\mathbb{E}\left[\|\widehat{h}_d - \mathbb{E}[\widehat{h}_d]\|^2\right] = \sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right). \tag{4.55}$$

For the other term, we have

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

Now, we remark that

$$\begin{pmatrix}h_d(x_{-n})\\ \vdots\\ h_d(x_{n-1})\end{pmatrix} = \sum_{k=0}^{d-1}\langle h,\varphi_k\rangle\begin{pmatrix}\varphi_k(x_{-n})\\ \vdots\\ \varphi_k(x_{n-1})\end{pmatrix} = \Phi_d\begin{pmatrix}\langle h,\varphi_0\rangle\\ \vdots\\ \langle h,\varphi_{d-1}\rangle\end{pmatrix}$$

and therefore,

$$(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\begin{pmatrix}h_d(x_{-n})\\ \vdots\\ h_d(x_{n-1})\end{pmatrix} = \begin{pmatrix}\langle h,\varphi_0\rangle\\ \vdots\\ \langle h,\varphi_{d-1}\rangle\end{pmatrix}.$$

Thus, it follows

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

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

$$\|(\Phi_d^t\Phi_d)^{-1}\Phi_d^t\|_{op}^2 = \lambda_{max}(\Phi_d(\Phi_d^t\Phi_d)^{-1}(\Phi_d^t\Phi_d)^{-1}\Phi_d^t) = \frac{T}{n}\lambda_{max}\left(\Psi_d^{-1}\right) \tag{4.56}$$

This implies

$$\|h_d - \mathbb{E}[\widehat{h}_d]\|^2 \leqslant \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2, \tag{4.57}$$

From (4.55) and (4.7.2), we derive

$$\mathbb{E}[\|\widehat{h}_d - h\|^2] \leqslant \sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right) + \|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2. \tag{4.58}$$

□

Proof of Lemma 4.3.2. 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 4.8.2 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 (4.11), (4.14) and the Cauchy-Schwarz inequality, we have for $h \in W_H^\alpha(L)$ that

$$\sum_{j\geqslant d}a_j(h)\varphi_j'(x) \leqslant \left(\sum_{j\geqslant d}j^\alpha|a_j(h)|^2\right)^{\frac{1}{2}}\left(\sum_{j\geqslant d}j^{-\alpha+\frac{5}{6}}\right)^{\frac{1}{2}} \lesssim \left(d^{-\alpha+\frac{5}{6}+1}\right)^{\frac{1}{2}} = d^{-\frac{\alpha}{2}+\frac{11}{12}},$$

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 (4.11) and the Cauchy-Schwarz inequality, we have for $h \in W_H^\alpha(L)$ that

$$\sum_{j\geqslant d}|a_j(h)\varphi_j(x)| \leqslant \sum_{j\geqslant d}j^{\frac{\alpha}{2}}|a_j(h)|j^{-\frac{\alpha}{2}-\frac{1}{12}} \leqslant \left(\sum_{j\geqslant d}j^\alpha|a_j(h)|^2\right)^{\frac{1}{2}}\left(\sum_{j\geqslant d}j^{-\alpha-\frac{1}{6}}\right)^{\frac{1}{2}} \lesssim d^{-\frac{\alpha}{2}+\frac{5}{12}}.$$

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 4.8.2 (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 (4.11), 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 $\alpha>17/6$

$$|\sum_{j>d}a_j(h)\varphi_j''(x)|\leqslant\left(\sum_{j>d}j^\alpha a_j(h)^2\right)^{\frac12}\left(\sum_{j>d}j^{-\alpha+\frac{11}{6}}\right)^{\frac12}\lesssim(d^{-\alpha+\frac{11}{6}+1})^{\frac12}=d^{-\frac{\alpha}{2}+\frac{17}{12}}.$$

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

$$\|h-h_d\|_n^2\leqslant2\|h-h_d\|^2+C\frac{T^3}{12n^2}.$$

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

Proof of Proposition 4.3.2. For $h\in W_H^\alpha(L)$ with $\alpha>11/6$, we have from Proposition 4.3.1 (i) and Lemma 4.3.2 that

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

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 4.3.2. The part (ii) is similar considering (**A4**). □

Proof of Theorem 4.3.1. Inequality (4.25) 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 (4.26). We recall that (see Equation (17) in Baraud (2000))

$$\forall d \in \mathbb{N} \quad , \sup_{t \in S_d, t \neq 0} \frac{\|t\|}{\|t\|_n} = \lambda_{max}(\Psi_d^{-1}). \tag{4.59}$$

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 (**A4**) 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 (4.59) that $\|\widehat{h}_{\widehat{d}} - h_d\|^2 \leqslant 2\lambda_2^2\|\widehat{h}_{\widehat{d}} - h\|_n^2 + 2\lambda_2^2\|h - h_d\|_n^2$. Thus, for any $d \geqslant 1$,

$$\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|^2] \leqslant 2\lambda_2^2\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|_n^2] + 2\lambda_2^2\|h - h_d\|_n^2 + \|h_d - h\|^2.$$

From (4.25), we derive that

$$\begin{aligned}\mathbb{E}[\|\widehat{h}_{\widehat{d}} - h\|^2] \leqslant & 2\lambda_2^2\left[C(\kappa)\inf_{d \in \mathcal{M}_n}\left(\inf_{t \in S_d}\|t - h\|_n^2 + \sigma_\varepsilon^2 T\frac{d}{n}\right) + \frac{C'}{n}\right] + 2\lambda_2^2\|h - h_d\|_n^2 + \|h_d - h\|^2 \\ & \leqslant \max(1, 2\lambda_2^2 C(\kappa))\inf_{d \in \mathcal{M}_n}\left((2\lambda_2^2 + 1)\|h - h_d\|_n^2 + \|h_d - h\|^2 + \sigma_\varepsilon^2 T\frac{d}{n}\right) + 2\lambda_2^2\frac{C'T}{n}.\end{aligned}$$

This gives (4.26) and ends the proof of Theorem 4.3.1. □

4.7.3 Proofs of Section 4.4

Proof of Proposition 4.4.1.

Proof of Equation (4.32)**.** 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, (4.13) and under (**A3**), we deduce that

$$\begin{aligned}\|\widehat{f}_{(d)} - \widehat{f}_{(\ell),d}\|^2 = & \frac{1}{2\pi}\int_{|u|>\ell}\frac{|\widehat{h}_d^*(u)|^2}{|g^*(u)|^2}du = \frac{1}{2\pi}\int_{|u|>\ell}\frac{\left|\widehat{h}_d^*(u) - \mathbb{E}\,\widehat{h}_d^*(u) + \mathbb{E}\,\widehat{h}_d^*(u)\right|^2}{|g^*(u)|^2}du \\ & = \int_{|u|>\ell}\frac{\left|\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)} - \mathbb{E}\,\widehat{b}_j^{(d)} + \mathbb{E}\,\widehat{b}_j^{(d)}\right)\varphi_j(u)\right|^2}{|g^*(u)|^2}du \\ & \leqslant \sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)} - \mathbb{E}\,\widehat{b}_j^{(d)} + \mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\sum_{j=0}^{d-1}\int_{|u|>\ell}\frac{\varphi_j(u)^2}{|g^*(u)|^2}du \\ & \leqslant c_1 C_\infty'^2\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)} - \mathbb{E}\,\widehat{b}_j^{(d)} + \mathbb{E}\,\widehat{b}_j^{(d)}\right)^2 de^{-\frac{\xi\ell^2}{2}}\int e^{-\frac{\xi u^2}{2}}(1+u^2)^\gamma du.\end{aligned}$$

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

$$\mathbb{E}\left[\|\widehat{f}_{(d)} - \widehat{f}_{(\ell),d}\|^2\right] \leqslant c_1 c_1' C_\infty^2\,\mathbb{E}\left[\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)} - \mathbb{E}\,\widehat{b}_j^{(d)} + \mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\right]de^{-\xi d}. \tag{4.60}$$

By the definition $\widehat{h}_d$ given in (4.19), 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 (4.55) and (4.56) (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 (**A4**) and (**A5**), we have $\sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}(\Psi_d^{-1}) + \lambda_{\max}\left(\Psi_d^{-1}\right)\|h\|_n^2 \leqslant \max(\sigma_\varepsilon^2, 2\|h\|_\infty^2)\lambda_2 T$. It comes that $\mathbb{E}\left[\sum_{j=0}^{d-1}\left(\widehat{b}_j^{(d)} - \mathbb{E}\,\widehat{b}_j^{(d)} + \mathbb{E}\,\widehat{b}_j^{(d)}\right)^2\right] \leqslant \max(\sigma_\varepsilon^2, 2\|h\|_\infty^2)\lambda_2 T$. Injecting this in (4.60), we obtain

$$\mathbb{E}[\|\widehat{f}_{(d)} - \widehat{f}_{(\ell),d}\|^2] \leqslant c_1 c_1' C_\infty'^2 \max(\sigma_\varepsilon^2, 2\|h\|_\infty^2) d e^{-\xi d} = C\lambda_2 T e^{-\frac{\xi d}{2}},$$

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

$$\mathbb{E}[\|\widehat{f}_{(d)} - \widehat{f}_{(\ell),d}\|^2] \leqslant C\lambda_2 T e^{-\xi d} + 2\,\mathbb{E}[\|\widehat{f}_{(\ell),d} - f\|^2].$$

Proof of Equation (4.33)**.** For all $\ell > 0$, $d \geqslant 1$, we have the following decomposition :

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right] = \|f - f_{(\ell)}\|^2 + \mathbb{E}\left[\|f_{(\ell)} - \widehat{f}_{(\ell),d}\|^2\right]. \tag{4.61}$$

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 (4.21) in the above bound and in (4.61) gives (4.33). □

Proof of Theorem 4.4.1. Under (**A3**), (**A4**) and the assumptions of Proposition 4.7.1 (*i.e.* h belongs to W_H^α with $\alpha = s + \gamma$ for some condition on f and g), we get from Lemma 4.3.2 :

$$\mathbb{E}\left[\|\widehat{f}_{(\ell),d} - f\|^2\right] \leqslant L\ell^{-2s} + (1+\ell^2)^\gamma\left[\sigma_\varepsilon^2\lambda_2\frac{dT}{n} + (1+\lambda_2)L' d^{-\alpha} + C(\alpha, L)\frac{T^2}{n}\right].$$

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

Proof of Proposition 4.4.2. As f is Gaussian, then it belongs to $W_H^\alpha(D)$ (see (4.15)) with α as large as desired, since f is infinitely differentiable and $f, \ldots, f^{(\alpha)}, x^{\alpha-l}f^{(l)}$ for $l = 0 \ldots \alpha - 1$, see Section 4.3.3. 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 (same computation as the proof of Proposition 4.7.1 given in Section proof) and the Cauchy-Schwarz inequality that :

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

Moreover, it holds $\int |(g^*)^{(k)}(u)|^2 du = \frac{1}{2\pi}\int |x^k g(x)|^2 dx \leqslant \int_{|x|\leqslant 1} |g(x)|^2 dx + \int_{|x|\geqslant 1} |x^{(\alpha)}g(x)|^2 dx < +\infty$. Therefore, $\|x^\alpha h^{(l)}\|^2 < +\infty$ and h belongs to $W_H^\alpha(\cdot)$. Proposition 4.3.2 *(ii)* gives $\mathbb{E}[\|\widehat{h}_{d_{opt}} - h\|^2] \lesssim n^{-\frac{\alpha}{\alpha+1}}$. Plugging this in (4.33) 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 4.4.3. The proof is similar to that of Proposition 4.4.2. 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 4.4.2) 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 $g^*(u) = \exp(-\frac{\sigma^2t^2}{2})$ and $\|f - f_{(\ell)}\|^2 = \frac{1}{2\pi}\int_{|u|>\ell} |f^*(u)|^2 du \leqslant \ell^{-2s}$ for $f \in W^s(\cdot)$ (see (4.16)). From the previous results, we derive from (4.33)

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

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

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

□

Proof of Proposition 4.4.4. 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 :

$$\mathbb{E}[\|\widehat{h}_d - h\|^2] \lesssim \sigma_\varepsilon^2 \lambda_2 \frac{T}{n} d + \frac{1}{\sqrt{d}} \exp(-\lambda_{\sigma,\theta} d) + \mathcal{O}(\frac{T^2}{n}) \tag{4.62}$$

Injecting $d_{opt} = [\log(n)/\lambda_{\sigma,\theta}]$ in (4.62), we have (4.35). Injecting this in (4.33), 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 (4.33)

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

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 4.4.5. 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

$$\mathbb{E}[\|\widehat{h}_d - h\|^2] \leqslant C d^{-(p+q-2)} + \sigma_\varepsilon^2\lambda_2\frac{T}{n}d + \mathcal{O}(\frac{T^2}{n}).$$

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 $\|f - f_{(\ell)}\|^2 = \frac{1}{2\pi}\int_{|u|>\ell}|f^*(u)|^2 du \asymp (\frac{\ell}{\theta})^{-2p+1}$. Plugging the previous results in (4.33) 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 4.4.2. Let us start by the proof of Inequality (4.40). First, we have by definition of $\widehat{A}$, $\widehat{d}$ and $\forall d \in \mathcal{M}_n^{(1)}$,

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

Taking the expectation in the previous inequality, we get

$$\mathbb{E}\left[\|\widehat{\widehat{f}}_{\widehat{d}} - f\|^2\right] \leqslant 6\,\mathbb{E}[\widehat{A}(d)] + 6\kappa_2 V(d)) + 3\,\mathbb{E}\left[\|\widehat{\widehat{f}}_{d} - f\|^2\right]. \tag{4.63}$$

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

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

Using this, it comes

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

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

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

Besides, by definition of $\widehat{f}_{(\sqrt{2d}),d}$ given in (4.30) and $f_{(\sqrt{2d})}$ in (4.31), 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 $d' \geqslant d$

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

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 4.7.1. *Under the assumptions of Theorem 4.4.2, 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 $C_0 = C_0(\mathbb{E}[\varepsilon_1^4], \gamma, c_1, \xi, \lambda_2, C'_\infty)$.

By Phythagoras Therorem, we have

$$\|h - \mathbb{E}[\widehat{h}_d]\| = \|h - h_d\|^2 + \|h_d - \mathbb{E}[\widehat{h}_d]\|^2. \tag{4.64}$$

Then, we deduce from Lemma 4.7.1, (4.63), (4.61) and (4.55) that :

$$\mathbb{E}[\|\widehat{\widehat{f}}_{\widehat{d}} - f\|^2] \leqslant 57\|f - f_{(\sqrt{2d})}\|^2 + 7\kappa_2 V(d) + 54C_0\frac{\log(n)}{n} + 57R_b(d), \tag{4.65}$$

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 (4.65) ends the proof of Inequality (4.40).

Now, we prove Inequality (4.41). From (4.64)-(4.57) and (**A4**), it holds

$$\Delta(\sqrt{2d'})\|h - \mathbb{E}[\widehat{h}_{d'}]\|^2 \leqslant (1+\lambda_2)\Delta(\sqrt{2d'})\|h - h_{d'}\|^2 + \lambda_2\Delta(\sqrt{2d'})\|h - h_{d'}\|_n^2.$$

Under (**A3**), it comes from Lemma 4.3.2 (ii) and for $h \in W_H^{s+\gamma}(L')$,

$$\begin{aligned}\Delta(\sqrt{2d'})\|h - \mathbb{E}[\widehat{h}_{d'}]\|^2 &\leqslant c_1(1+\lambda_2)(1+2d')^\gamma L'(d')^{-s-\gamma} + C\Delta(\sqrt{2d'})\frac{T^3}{n^2}\\ &\leqslant C\left(d'^{-s} + \frac{T^2}{n}\right).\end{aligned}$$

Then, for $d' \geqslant d$, we derive that $R_b(d) \leqslant C\left(d^{-s} + \frac{T^2}{n}\right)$. Plugging this in (4.65) and using $\|f - f_{\sqrt{2d}}\|^2 \leqslant 2^{-s}Ld^{-s}$ because $f \in W^s(L)$ concludes the proof of Theorem 4.4.2. □

Proof of Lemma 4.7.1. 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 (4.19), 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

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

This implies that

$$\begin{aligned}\sum_{d\in\mathcal{M}_n^{(1)}}\mathbb{E}\left[\left(\|\widehat{f}_{(\sqrt{2d}),d} - \mathbb{E}[\widehat{f}_{(\sqrt{2d}),d}]\|^2 - \frac{\kappa_1}{6}V(d)\right)_+\right] \leqslant 2\sum_{d\in\mathcal{M}_n^{(1)}}\mathbb{E}\left[\left(\sup_{t\in\mathcal{S}_d,\|t\|=1}|\nu_n^{(1)}(t)|^2 - \frac{\kappa_1}{12}V(d)\right)_+\right] \\ + 2n\,\mathbb{E}\left[\sup_{t\in\mathcal{S}_d,\|t\|=1}|\nu_n^{(2)}(t)|^2\right].\end{aligned} \tag{4.66}$$

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 (4.19)) $\check{h}_d = \sum_{j=0}^{d-1} \check{b}_j^{(d)} \varphi_j$, $\vec{\check{b}}^{(d)} = (\check{b}_0^{(d)}, \dots, \check{b}_{d-1}^{(d)})^t = (\Phi_d^t \Phi_d)^{-1} \Phi_d^t \vec{y} = \frac{T}{n} \Psi_d^{-1} \Phi_d^t \vec{y}$, $\vec{y} = (y(x_{-n}), \dots, y(x_{n-1}))^t$ with $y(x_i) = h(x_i) + \xi_i$, 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 (4.55) (with σ_ε^2 replaced by $\mathbb{E}[\xi_1^2]$) and under (**A4**) that $\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(2)}(t)|^2\right] \leqslant \Delta(\sqrt{2d}) \lambda_2 d\, \mathbb{E}[\xi_1^2] \frac{T}{n}$. 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 4.7.2. *Under* (**A6**)*, 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 4.7.2 with $p = 4$ and choosing

$$k_n = 2\sqrt{2} b \sqrt{\log(n)}, \tag{4.67}$$

we get

$$n\,\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(2)}(t)|^2\right] \leqslant n \Delta(\sqrt{2d}) \lambda_2 d \sqrt{\mathbb{E}[\varepsilon_1^4]} \sqrt{\mathbb{P}(|\varepsilon_1| \geqslant k_n)} \frac{T}{n} \leqslant \frac{C}{n}, \tag{4.68}$$

since $\Delta(\sqrt{2d}) \lambda_2 dT \lesssim n$ by definition of $\mathcal{M}_n^{(1)}$.

Upper bound for $\sum_{\boldsymbol{d} \in \boldsymbol{\mathcal{M}_n^{(1)}}} \mathbb{E}\left[\left(\boldsymbol{\sup_{t \in \mathcal{S}_d, \|t\|=1}} |\boldsymbol{\nu_n^{(1)}(t)}|^{\boldsymbol{2}} - \boldsymbol{\frac{\kappa_1}{12}} \boldsymbol{V(d)}\right)_+\right]$. We bound this term applying the Talagrand inequality given in Appendix 4.8.4. 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 (**A4**) and from (4.55),

$$\mathbb{E}\left[\sup_{t \in \mathcal{S}_d, \|t\|=1} |\nu_n^{(1)}(t)|^2\right] \leqslant \lambda_2\, \mathbb{E}[\zeta_1^2] \Delta(\sqrt{2d}) \frac{dT}{n} \leqslant \lambda_2 \sigma_\varepsilon^2 \Delta(\sqrt{2d}) \frac{dT}{n} := H^2.$$

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 (4.12) 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 (**A4**), it yields

$$\frac{T}{n}\sum_{i=-n}^{n-1}\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2 = \frac{T}{n}\mathrm{tr}\left[\Phi_d\Psi_d^{-1}\Psi_d^{-1}\Phi_d^t\right] = \mathrm{tr}\left[\Psi_d^{-1}\right] \leqslant \lambda_2 d,$$

which implies

$$\sup_{t\in\mathcal{S}_d,\|t\|=1}\frac{1}{2n}\sum_{i=-n}^{n-1}\mathrm{Var}(\alpha_{t,d,i}(\zeta_i)) \leqslant 2\sigma_\varepsilon^2 T\lambda_2 d\Delta(\sqrt{2d}) =: v.$$

Computing of M_1. Using successively (4.12), the Cauchy-Schwarz inequality and the

orthonormality of φ_j, we have on the process $\nu_n^{(1)}$

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

$$\leqslant \sup_{t\in\mathcal{S}_d,\|t\|=1}\left(2k_n\frac{T}{\pi}\|t^*\|\sqrt{2\pi}\left(\int_{-\sqrt{2d}}^{\sqrt{2d}}\frac{|\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}\varphi_j(u)|^2}{|g^*(u)|^2}du\right)^{\frac{1}{2}}\right)$$

$$\leqslant 4k_nT\left(\Delta(\sqrt{2d})\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2\right)^{\frac{1}{2}}.$$

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 $\vec{w} = P^t\Psi_d^{-\frac{1}{2}}\Phi_d^t\vec{e}_i$

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

The definition of operator norm implies,

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

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

$$\sum_{j=0}^{d-1}\left[\Psi_d^{-1}\Phi_d^t\right]_{j,i}^2 \leqslant \lambda_{max}(\Psi_d^{-1})\frac{n}{T}. \tag{4.69}$$

Under (**A4**), we obtain

$$\sup_{t\in\mathcal{S}_d,\|t\|=1}\|\alpha_{t,d,i}\|_\infty \leqslant 4k_nT^{\frac{1}{2}}(\Delta(\sqrt{2d})\lambda_2n)^{\frac{1}{2}} := M_1.$$

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

$$T_d = \frac{2\sigma_\varepsilon^2 \lambda_2 T d \Delta(\sqrt{2d})}{n} \exp\left(-\frac{K_1\delta}{2}\right) \text{ and } U_d = \frac{196\lambda_2 k_n^2 T \Delta(\sqrt{2d})}{K_1 C^2(\delta) n} \exp\left(-K_1' C(\delta)\sqrt{\delta}\frac{\sqrt{\sigma_\varepsilon^2}}{k_n}\sqrt{d}\right),$$

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

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

As $Td\Delta(\sqrt{2d}) \lesssim n$, then, it yields $\sum_{d \in \mathcal{M}_n^{(1)}} T_d \lesssim n \exp(-K_1\delta/2)$. 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 (4.67), 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 (4.68) in (4.66) concludes the proof. □

Proof of Lemma 4.7.2. 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 $\inf_{t \geqslant 0} r(t) = -s^2/(2b^2)$ 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 $\mathbb{E}[|\varepsilon_1|^p] = (2b^2)^{\frac{p}{2}} p\Gamma(\frac{p}{2})$. □

4.7.4 Proofs of Section 4.5

Proof of Proposition 4.5.1. By the Pythagoras Theorem, we have

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] = \|f - f_m\|^2 + \mathbb{E}\left[\|\widehat{f}_{m,d} - f_m\|^2\right] \tag{4.70}$$

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

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

Injecting this in (4.70) and using Proposition 4.3.1 (ii), we get

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] \leqslant \|f - f_m\|^2 + 2\Sigma(m)\left(\|h - h_d\|^2 + \lambda_{max}\left(\Psi_d^{-1}\right)\|h - h_d\|_n^2 + \sigma_\varepsilon^2\frac{T}{n}\mathrm{tr}\left(\Psi_d^{-1}\right)\right). \tag{4.71}$$

On the other hand, from (4.70), we have

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] = \|f - f_m\|^2 + \|\,\mathbb{E}[\widehat{f}_{m,d}] - f_m\|^2 + \mathbb{E}\left[\|\widehat{f}_{m,d} - \mathbb{E}[\widehat{f}_{m,d}]\|^2\right]. \tag{4.72}$$

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 (4.19), 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

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

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 4.4.2 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 :

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

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 4.72, 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 (4.71) ends the proof. □

Proof of Theorem 4.5.1. Under (**A3**), (**A4**) and the assumptions of Proposition 4.7.1 (*i.e.* h belongs to W_H^α for some condition on f and g), it holds from Lemma 4.3.2 :

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] \leqslant Lm^{-s} + 2\Sigma(m)\left[(1+\lambda_2)L'd^{-\alpha} + \lambda_2\sigma_\varepsilon^2 T\frac{d}{n} + C(\alpha, L)\frac{T^2}{n}\right]$$

Besides, under (**A3**) and from (4.13) with $\rho \geqslant 2$, we have

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

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 $\Sigma(m) \leqslant C_1 m^\gamma$. Then, we obtain

$$\mathbb{E}\left[\|\widehat{f}_{m,d} - f\|^2\right] \leqslant Lm^{-s} + 2C_1 m^\gamma \left[(1+\lambda_2)L' d^{-\alpha} + \lambda_2 \sigma_\varepsilon^2 T \frac{d}{n} + C(\alpha, L)\frac{T^2}{n}\right],$$

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

4.8 Appendix

4.8.1 Study of $\mathrm{tr}(\Psi_d)$ and discussion on Assumption (A4)

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

$$|||\Psi_d - I_d||| \leqslant C_1 e^{-2\xi T^2} d + \phi_0 d^{\frac{17}{12}} \frac{T^2}{n}, \tag{4.73}$$

where C_1 depends on ξ, C'_∞ given in (4.13) 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 4.8.1 (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 (**A4**) holds and is not a strong condition.

In Table 4.5 and 4.6, we report the matrix norm of $\Psi_d - I_d$ and $\Psi_d^{-1} - I_d$ for

$$|||A||| = |||A|||_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^{n} |A_{ij}|, \quad A = (A_{ij})_{1\leqslant i,j\leqslant n}. \tag{4.74}$$

Comment on Table 4.5 and 4.6. 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 (4.73), 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

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 4.5 – 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 4.6 – Matrix norm of $A - I_d$ with $A = \Psi_d$ without parentheses and $A = \Psi_d^{-1}$ in parentheses for $T = 10$.

excepted. The results of Table 4.6 are better than those of Table 4.5. This is due to the choice $T = 10$ larger than $T = \sqrt{2d-1}$ for the choices of n et d given in Table 4.5 (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 $|||\Psi_d^{-1} - I_d|||$.

Proof of Lemma 4.8.1. We prove the result only for the particular norm defined in (4.74) 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 4.8.2, 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 (4.13) 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

$$\|\!|\Psi_d - I_d|\!\|_1 \leqslant d\left[C_1 e^{-\xi T^2} + \phi_0 d^{\frac{5}{12}} \frac{T^2}{n}\right]. \tag{4.75}$$

□

4.8.2 Estimating error in Riemann sums

We give in this section the approximate errors of Riemann sum.

Lemma 4.8.2. *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 4.8.2. 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 interpoltor 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. □

4.8.3 Useful tools and inequalities

The proof of the following Theorem can be found in Stewart and Sun (1990) (see also Equation (1.2) in Stewart (1990)).

Theorem 4.8.1. *Let A and E be two square matrices. If A is nonsingular and for some norm $\|A^{-1}E\| < 1$, then we have*

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

4.8.4 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).

Bibliographie

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.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281.

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. (2016). Nonparametric Laguerre estimation in the multiplicative censoring model. *Electron. J. Stat.*, 10(2) :3114–3152.

Belomestny, D., Comte, F., and Genon-Catalot, V. (2017). Correction to : Nonparametric Laguerre estimation in the multiplicative censoring model [MR3571964]. *Electron. J. Stat.*, 11(2) :4845–4850.

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.

Bercu, B., Capderou, S., and Durrieu, G. (2019). Nonparametric recursive estimation of the derivative of the regression function with application to sea shores water quality. *Stat. Inference Stoch. Process.*, 22(1) :17–40.

Bhattacharya, P. (1967). Estimation of a probability density function and its derivatives. *Sankhyā : The Indian Journal of Statistics, Series A*, pages 373–382.

Birgé, L. (2001). An alternative point of view on Lepski's method. In *State of the art in probability and statistics (Leiden, 1999)*, volume 36 of *IMS Lecture Notes Monogr. Ser.*, pages 113–133. Inst. Math. Statist., Beachwood, OH.

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York.

Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves : exponential bounds and rates of convergence. *Bernoulli*, 4(3) :329–375.

Birgé, L. and Massart, P. (2001). Gaussian model selection. *Journal of the European Mathematical Society*, 3(3) :203–268.

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

Bongioanni, B. and Torrea, J. L. (2009). What is a Sobolev space for the Laguerre function systems ? *Studia Math.*, 192(2) :147–172.

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

Butucea, C. and Tsybakov, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatn. Primen.*, 52(2) :336–349.

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.

Chacón, E. and Duong, T. (2013). Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electron. J. Stat.*, 7 :499–532.

Chacón, E., Duong, T., and Wand, M. P. (2011). Asymptotics for general multivariate kernel density derivative estimators. *Statist. Sinica*, 21(2) :807–840.

Chagny, G. (2013). Penalization versus Goldenshluger-Lepski strategies in warped bases regression. *ESAIM Probab. Stat.*, 17 :328–358.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence*, 17(8) :790–799.

Comaniciu, D. and Meer, P. (2002). Mean shift : A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (5) :603–619.

Comte, F. (2017). *Estimation non-paramétrique*. Spartacus-Idh.

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., Dedecker, J., and Taupin, M. L. (2008). Adaptive density deconvolution with dependent inputs. *Math. Methods Statist.*, 17(2) :87–112.

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. (2015). Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.*, 9(1) :1113–1149.

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. (2020a). Regression function estimation as a partly inverse problem. *Ann. Inst. Statist. Math.*, 72(4) :1023–1054.

Comte, F. and Genon-Catalot, V. (2020b). Regression function estimation on non compact support in an heteroscesdastic model. *Metrika*, 83(1) :93–128.

Comte, F. and Johannes, J. (2012). Adaptive functional linear regression. *Ann. Statist.*, 40(6) :2765–2797.

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. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(2) :569–609.

Comte, F. and Lacour, C. (2021). Non compact estimation of the conditional density from direct or noisy data.

Comte, F. and Marie, N. (2020). Bandwidth selection for the Wolverton-Wagner estimator. *J. Statist. Plann. Inference*, 207 :198–214.

Comte, F. and Rozenholc, Y. (2004). A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.*, 56(3) :449–473.

Comte, F., Rozenholc, Y., and Taupin, M. (2007). Finite sample penalization in adaptive density deconvolution. *Journal of Statistical Computation and Simulation*, 77(11) :977–1000.

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.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin.

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.

Dmitriev, Y. and Tarasenko, F. (1974). On the estimation of functionals of the probability density and its derivatives. *Theory of Probability & Its Applications*, 18(3) :628–633.

Dobrovidov, A. V. and Markovich, L. A. (2013a). Data-driven bandwidth choice for gamma kernel estimates of density derivatives on the positive semi-axis. *IFAC Proceedings Volumes*, 46(11) :500–505.

Dobrovidov, A. V. and Markovich, L. A. (2013b). Nonparametric gamma kernel estimators of density derivatives on positive semi-axis. *IFAC Proceedings Volumes*, 46(9) :910–915.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2) :508–539.

Efromovich, S. (1998). Simultaneous sharp estimation of functions and their derivatives. *Ann. Statist.*, 26(1) :273–278.

Efromovich, S. (1999). *Nonparametric curve estimation : methods, theory, and applications*. Springer Series in Statistics.

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.

Farrell, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.*, 43 :170–180.

Fukunaga, K. and Hostetler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inform. Theory*, IT-21 :32–40.

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.

Gendre, X. (2009). *Estimation par sélection de modèle en régression hétéroscédastique*. PhD thesis.

Genovese, C. R., Perone-Pacifico, M., Verdinelli, I., and Wasserman, L. (2016). Non-parametric inference for density modes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(1) :99–126.

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

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. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4) :1150–1190.

Goldenshluger, A. and Lepski, O. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities. *Probab. Theory Related Fields*, 143(1-2) :41–71.

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

Härdle, W., Hildenbrand, W., and Jerison, M. (1991). Empirical evidence on the law of demand. *Econometrica : Journal of the Econometric Society*, pages 1525–1549.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Härdle, W., Marron, J. S., and Tsybakov, A. B. (1992). Bandwidth choice for average derivative estimation. *J. Amer. Statist. Assoc.*, 87(417) :218–226.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.*, 84(408) :986–995.

Hosseinioun, N., Doosti, H., and Niroumand, H. A. (2009). Wavelet-based estimators of the integrated squared density derivatives for mixing sequences. *Pakistan J. Statist.*, 25(3) :341–350.

Hosseinioun, N., Doosti, H., and Nirumand, H. A. (2012). Nonparametric estimation of the derivatives of a density by the method of wavelet for mixing sequences. *Statist. Papers*, 53(1) :195–203.

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

Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electron. J. Stat.*, 8(2) :2879–2904.

Kerkyacharian, G., Lepski, O., and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121(2) :137–170.

Kerkyacharian, G., Pham Ngoc, T. M., and Picard, D. (2011). Localized spherical deconvolution. *Ann. Statist.*, 39(2) :1042–1068.

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

Koekoek, R. (1990). Generalizations of Laguerre polynomials. *J. Math. Anal. Appl.*, 153(2) :576–590.

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.

Lacour, C., Massart, P., and Rivoirard, V. (2017). Estimator selection : a new method with applications to kernel density estimation. *Sankhya A*, 79(2) :298–335.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338.

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

Lepski, O. V. (2018). A new approach to estimator selection. *Bernoulli*, 24(4A) :2776–2810.

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

Lukacs, E. (1970). *Characteristic functions*. Hafner Publishing Co., New York. Second edition, revised and enlarged.

Mabon, G. (2016). Adaptive deconvolution of linear functionals on the nonnegative real line. *J. Statist. Plann. Inference*, 178 :1–23.

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

Mallows, C. L. (1995). More comments on cp. *Technometrics*, 37(4) :362–372.

Manski, C. F. and McFadden, D., editors (1981). *Structural analysis of discrete data with econometric applications*. MIT Press, Cambridge, Mass.-London.

Markett, C. (1984). Norm estimates for (c, δ) means of hermite expansions and bounds for δeff. *Acta Mathematica Hungarica*, 43(3-4) :187–198.

Markovich, L. A. (2016). Gamma kernel estimation of the density derivative on the positive semi-axis by dependent data. *REVSTAT*, 14(3) :327–348.

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

Mazurová, L. (2018). Risk theory lecture notes.

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.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.*, 98(10) :1955–1968.

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.

Park, C. and Kang, K. (2008). SiZer analysis for the comparison of regression curves. *Comput. Statist. Data Anal.*, 52(8) :3954–3970.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33 :1065–1076.

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

Pham Ngoc, T. M. (2019). Adaptive optimal kernel density estimation for directional data. *J. Multivariate Anal.*, 173 :248–267.

Plancade, S. (2009). Estimation of the density of regression errors by pointwise model selection. *Math. Methods Statist.*, 18(4) :341–374.

Rao, B. L. S. P. (1996). Nonparametric estimation of the derivatives of a density by the method of wavelets. *Bull. Inform. Cybernet.*, 28(1) :91–100.

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

Rigollet, P. and Tsybakov, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.*, 16(3) :260–280.

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

Sasaki, H., Noh, Y.-K., Niu, G., and Sugiyama, M. (2016). Direct density derivative estimation. *Neural Comput.*, 28(6) :1101–1140.

Schmisser, E. (2013). Nonparametric estimation of the derivatives of the stationary density for stationary processes. *ESAIM Probab. Stat.*, 17 :33–69.

Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.*, 40 :1187–1195.

Shen, W. and Ghosal, S. (2017). Posterior contraction rates of density derivative estimation. *Sankhya A*, 79(2) :336–354.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, 6(1) :177–184.

Singh, R. S. (1976). Nonparametric estimation of mixed partial derivatives of a multivariate density. *J. Multivariate Anal.*, 6(1) :111–122.

Singh, R. S. (1977a). Applications of estimators of a density and its derivatives to certain statistical problems. *J. Roy. Statist. Soc. Ser. B*, 39(3) :357–363.

Singh, R. S. (1977b). Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density. *Ann. Statist.*, 5(2) :394–399.

Singh, R. S. (1978). Nonparametric estimation of derivatives of average of μ-densities with convergence rates and applications. *SIAM J. Appl. Math.*, 35(4) :637–649.

Singh, R. S. (1979). Mean squared errors of estimates of a density and its derivatives. *Biometrika*, 66(1) :177–180.

Stewart, G. W. (1990). Matrix perturbation theory.

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.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

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

Varet, S., Lacour, C., Massart, P., and Rivoirard, V. (2019). Numerical performance of penalized comparison to overfitting for multivariate kernel density estimation. *Preprint hal-02002275, version 1 and ARXIV : 1902.01075*.

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

Viennet, G. (1997). Inequalities for absolutely regular sequences : application to density estimation. *Probab. Theory Related Fields*, 107(4) :467–492.

Zhang, C. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, 18(2) :806–831.

Table des figures

Liste des tableaux

Estimation par projection pour des problèmes inverses sur des espaces de Laguerre et d'Hermite

Résumé. Dans cette thèse, nous développons des procédures d'estimation non paramétrique pour divers problèmes inverses sur des espaces de Laguerre et d'Hermite. La première partie est consacrée à l'estimation des dérivées d'une fonction de densité en base de Laguerre et d'Hermite. La deuxième partie est dédiée à l'estimation d'une densité et d'une fonction de régression dans un modèle de convolution en base d'Hermite. Différentes méthodes d'estimation sont présentées : méthode de projection fondée sur un développement de la fonction d'intérêt (densité, dérivée d'une densité, fonction de régression) en base de Laguerre ou d'Hermite ; mixte déconvolution-projection basée sur un développement en base d'Hermite et une transformation de Fourier inverse. Pour chacune de ces méthodes, nous établissons des bornes pour le risque quadratique. Pour des choix adéquats des paramètres (dimension de l'espace de projection ou cut-off), nous obtenons des vitesses de convergence de nos estimateurs. Ces paramètres dépendent cependant de quantités inconnues. Ainsi, nous proposons des procédures adaptatives pour les choisir de façon pertinente en s'inspirant des critères de sélection de modèles par pénalisation du type Birgé and Massart (1997) ou des méthodes de Goldenshluger and Lepski (2011) et nous démontrons des inégalités oracles non asymptotiques en utilisant les inégalités de concentration. Des études numériques et des comparaisons avec d'autres stratégies sont exposées pour illustrer les bonnes performances des méthodes proposées.

Mots-clés : Estimation non paramétrique ; Estimation par projection ; Problème inverse ; Base d'Hermite ; Base de Laguerre ; Sélection de modèle ; Méthode de Goldenshluger et Lepski.

Projection estimation for inverse problems on Laguerre and Hermite spaces

Abstract. In this thesis, we develop non parametric estimation procedures for various inverse problems on Laguerre and Hermite classes. The first part is dedicated to the estimation of the derivatives of a density function in the Laguerre and Hermite basis. The second part is devoted to the estimation of a density function and a regression function in the Hermite basis in a convolution model. Different estimation methods are presented : projection method based on a development of the function of interest (density, derivative of a density, regression function) in the Laguerre or Hermite basis ; mixed deconvolution-projection based on a development in the Hermite basis and an inverse Fourier transformation. For each of these methods, we establish bounds for quadratic risk. For adequate choices of parameters (dimension of projection space or cut-off), we obtain rates of convergence of our estimators. However, theses parameters depend on unknown quantities. Thus, we propose adaptive procedures to select them in a relevant way inspired by model selection by penalization of Birgé and Massart (1997)'s type or Goldenshluger and Lepski

(2011)'s methods criterions and we prove non-asymptotic oracle inequalities using concentration inequalities. Numerical studies and comparisons with other strategies are presented to illustrate the good performance of the proposed methods.

Keywords : Nonparmetric estimation ; Projection estimation ; Inverse problem ; Hermite basis, Laguerre basis ; Model selection ; Goldenshluger and Lepski method.