Estimateurs adaptatifs avec parcimonie structurée

Fabien Navarro

Université de Caen Basse-Normandie

École doctorale SIMEM

Thèse de doctorat

présentée et soutenue le : 27/11/2013

par

Fabien Navarro

pour obtenir le

Doctorat de l'Université de Caen Basse-Normandie

Spécialité : Mathématiques et leurs interactions

Estimateurs adaptatifs avec parcimonie structurée

Directeur de thèse : *Taoufik Sassi*
Co-directeur de thèse : *Jalal Fadili*
Co-encadrant de thèse : *Christophe Chesneau*

Jury

Arnak Dalalyan	Professeur	ENSAE-CREST	(Rapporteur)
Theofanis Sapatinas	Professeur	University of Cyprus	(Rapporteur)
Fabienne Comte	Professeur	Université Paris Descartes	
Guillaume Lecué	CR CNRS	École Polytechnique	
Erwan Le Pennec	Professeur	École Polytechnique	
Taoufik Sassi	Professeur	Université de Caen	(Directeur)
Jalal Fadili	Professeur	ENSICAEN	(Co-directeur)
Christophe Chesneau	MCF	Université de Caen	(Co-encadrant)

Remerciements

En tout premier lieu, je tiens à remercier chaleureusement mon directeur de thèse Taoufik Sassi pour m'avoir donné l'opportunité de faire cette thèse et pour m'avoir laissé une grande liberté dans mon travail. Je lui suis reconnaissant pour sa disponibilité, ses conseils, sa confiance et la bienveillance qu'il a toujours manifestée à mon égard.

Je remercie sincèrement mon co-directeur, Jalal Fadili, pour ses nombreux conseils avisés, ses encouragements, sa rigueur et son soutien constant qui m'ont permis d'effectuer cette thèse dans de très bonnes conditions.

Mes plus sincères remerciements vont également à Christophe Chesneau qui a très largement contribué à l'encadrement de cette thèse et qui pendant ces trois années a été mon collaborateur et aussi mon ami. Je lui exprime ma gratitude pour sa gentillesse, son soutien constant, sa rigueur scientifique, sa disponibilité et son aide précieuse qui, tout au long de ces trois années, ont largement contribué à l'avancée de ce travail. Merci Christophe !

Je suis très reconnaissant envers Arnak Dalalyan et Theofanis Sapatinas pour avoir accepté de rapporter ma thèse.

Je suis très honoré que Fabienne Comte, Guillaume Lecué et Erwan Le Pennec aient accepté de faire partie de mon jury de thèse.

Un grand merci à Fabienne Comte pour m'avoir accepté comme collaborateur et pour ses conseils et son soutien. De nombreuses contributions dans ce manuscrit sont le fruit de cette collaboration.

J'ai beaucoup apprécié la collaboration avec Maher Kachour, que je remercie pour son enthousiasme.

J'adresse mes remerciements à tous les membres du LMNO et du GREYC. Je souhaite également remercier l'ensemble des doctorants du LMNO et du GREYC.

Je remercie aussi le personnel de la bibliothèque universitaire ainsi que l'ensemble de l'équipe administrative du laboratoire et plus particulièrement Sonia qui s'est toujours montrée aimable et serviable les nombreuses fois où je suis allé la solliciter. Je remercie aussi les administrateurs système et réseaux du labo et plus particulièrement Ludo avec qui j'ai beaucoup apprécié échanger pendant ces trois ans.

Également, je remercie tous les enseignants qui ont contribué à ma formation durant mes années d'études.

Je remercie aussi mes amis et j'ai une pensée toute particulière pour Catherine. Merci pour votre soutien.

Enfin, je tiens à remercier ma famille, qui a toujours su m'apporter un soutien sans faille.

Table des matières

Introduction générale

Cette thèse a pour objet l'étude de certaines propriétés statistiques de différents estimateurs adaptatifs, dans le cadre de modèles originaux, à l'interface entre les domaines de la statistique mathématique, de la théorie de l'approximation et du traitement du signal et de l'image.

Depuis l'avènement de la *révolution numérique*, l'interaction mutuelle entre la communauté mathématique et celle du traitement du signal et de l'image est à l'origine d'une multitude d'avancées fondamentales et appliquées. En effet, l'étude de problèmes pratiques réels issus de domaines d'application variés permet d'identifier différents problèmes théoriques et modèles mathématiques à la fois originaux, novateurs et pertinents. Réciproquement, l'analyse mathématique de ces modèles donne parfois naissance à des méthodes algorithmiques nouvelles et applicables pour de nombreux problèmes. Parmi les avancées, le développement des ondelettes durant les années 1980-1990, aujourd'hui devenu un outil incontournable en traitement du signal et en statistique, constitue un exemple de collaboration fructueuse au sein de la communauté scientifique.

La parcimonie est devenue indéniablement un concept majeur à la fois en statistique mathématique et en traitement des signaux, et son utilisation comme représentation dans une base adaptée a été fructueuse dans plusieurs domaines tels que la régression, la sélection de variables, la théorie de l'apprentissage et dans les domaines de la compression, de la séparation de sources ou encore du débruitage. En témoigne la littérature abondante sur le sujet ainsi que les conséquences dans la vie pratique (e.g., le standard JPEG2000). Le choix d'une représentation appropriée des signaux, des images ou plus généralement des fonctions est souvent crucial pour résoudre des tâches spécifiques. En effet, la manière dont on présente l'information va affecter fortement ce que l'on peut faire avec. De ce point vue, analyser, reconstruire et représenter des fonctions arbitraires à l'aide de fonctions élémentaires (ou atomes) constitue une démarche primordiale.

Dans ce document, nous nous plaçons dans le cadre de l'estimation non-paramétrique, dans lequel les bases d'ondelettes permettent d'exploiter pleinement la notion de représentation parcimonieuse (ou creuse). Pour les modèles standards (e.g. densité, régression, etc), grâce à des propriétés bien particulières, les bases d'ondelettes ont ouvert le chemin des représentations parcimonieuses des signaux en statistique, auxquelles la théorie de l'échantillonnage compressé (*compressed sensing*), récemment popularisée par les travaux de Candes, a apporté une dimension supplémentaire. En effet, une représentation en ondelettes d'une fonction est dite parcimonieuse, c'est-à-dire qu'une grande partie des coefficients de la décomposition dans une base d'ondelettes sont très proches de zéros, et donc qu'un petit nombre de coefficients d'ondelettes concentrent l'essentiel de l'information. On dit qu'elle est compressible s'il existe une base dans laquelle, la fonction est représentée de façon parcimonieuse. Cette notion est intimement liée à celle d'approximation non-linéaire.

Dans une procédure d'estimation non-paramétrique, la projection sur une base d'ondelettes, associée à une approximation non-linéaire de type seuillage, a permis le développement d'estimateurs adaptatifs, dans le sens où ils ne dépendent pas de la régularité de la fonction estimée. Pour de nombreux problèmes, ces estimateurs, basées sur le seuillage des coefficients d'ondelettes dont la valeur absolue est supérieure à un seuil fixé, atteignent des vitesses de convergences (quasi-)minimax sur une large gamme de fonctions. Au début des années 90, Donoho et Johnstone ont élaboré les premières constructions adaptatives utilisant des techniques de seuillage en ondelettes.

Objectifs de la thèse

Lorsque nous sommes confrontés à des modèles plus complexes, par exemple la modélisation de signaux multicanaux, il semble naturel de modifier ces procédures d'estimation standards, tout en conservant la nature de la base considérée afin de toujours bénéficier des propriétés intéressantes de celle-ci. Cependant, cela peut poser des questions à la fois théoriques et numériques. Tout au long de cette thèse, nous tentons d'apporter des éléments de réponse à certaines de ces questions.

La parcimonie structurée tente d'aller plus loin en exploitant des connaissances *a priori* supplémentaires sur la structure du signal, ou plutôt de sa représentation dans un espace transformé donné (e.g. ondelettes), à estimer. Par exemple, les coefficients d'ondelettes peuvent se regrouper pour former des blocs. Ainsi, plusieurs auteurs ont proposé des estimateurs étendant les propriétés connues pour la parcimonie individuelle au cas de la parcimonie groupée (i.e. par blocs). Le seuillage par blocs, a été introduit en 1985 par Efroimovič dans le cadre de l'estimation d'une densité par un estimateur de type séries orthogonales. Il a été adapté en 1994 au contexte de l'analyse en ondelettes par Kerkyacharian, Picard et Tribouley.

Cette thèse se focalise sur l'apport de la parcimonie structurée pour la construction d'estimateurs adaptatifs, avec diverses applications : problèmes inverses et densité.

Problèmes inverses

Les problèmes inverses sont présents dans de nombreux domaines scientifiques : mathématique, physique, économie, etc. On les rencontre notamment au sein de nombreuses branches issues de ces disciplines (e.g., l'électromagnétisme, la thermique, l'astronomie, la dynamique des fluides, la chimiométrie, etc) dans lesquelles l'on peut envisager divers problèmes inverses. Par exemple, reconstruire l'état passé d'un système connaissant son état actuel ou encore l'estimation de certaines quantités d'intérêt du système. Typiquement, estimer les conditions initiales ou les conditions aux limites dans un système d'équations aux dérivées partielles ou l'estimation des dérivées d'un signal bruité qui peuvent être utilisées pour détecter des informations ou des caractéristiques importantes d'un système.

Estimation de densité

L'estimation de la densité de probabilité d'une variable aléatoire est utilisée dans de nombreux domaines scientifiques. A partir d'un ensemble de données d'observations, l'estimation de densité permet de produire une description de ces données. En considérant que ces données ont été généré par un certain processus, le but de l'estimation de densité est alors de construire un estimateur décrivant ce processus. Un tel problème d'estimation se pose dans de nombreuses situations, il arrive que la densité inconnue soit celle

d'une transformation de une ou plusieurs variables aléatoires. Nous devons donc l'estimer en prenant en compte la structure des observations. Notons qu'il existe de nombreux types de transformation (e.g., maximum, minimum, somme, etc) correspondant à différents modèles. Les domaines d'application couvrent l'hydrologie, l'ingénierie de fiabilité, l'assurance, etc. De nombreuses quantités d'intérêt en sciences actuarielles ou financière impliquent des sommes de variables aléatoires. Par exemple, dans le modèle de risque individuel, le montant total des créances sur un portefeuille de contrats d'assurance est modélisé comme la somme de toutes les revendications sur les politiques individuelles. Par conséquent, les fonctions de densité de probabilité des sommes de variables aléatoires sont d'un intérêt particulier. En règle générale, ces fonctions ne sont pas disponibles sous une forme analytique. Ainsi, le recours aux méthodes d'estimation non-paramétrique s'avère être indispensable.

Cette thèse a pour but d'apporter une contribution à l'étude minimax d'estimateurs par seuillage avec parcimonie structurée, pour certains problèmes inverses et modèles de densités.

Contributions et organisation du document

Ce document est composé de deux parties distinctes et organisé en six chapitres, dont cinq rédigés en anglais. Les résultats présentés dans les chapitres de cette thèse ont fait l'objet de plusieurs publications dans des actes de congrès et des revues internationales. La liste exhaustive des travaux est portée à la suite du chapitre 6. Enfin, tous nos travaux ont été effectués avec le soucis de proposer des algorithmes de résolution performants en pratique et faciles à mettre en oeuvre. Chacun des problèmes étudiés est donc accompagné de simulations numériques. Le plan du manuscrit est le suivant :

Partie I : La première partie est consacrée à l'étude d'estimateurs de seuillage adaptatifs avec parcimonie structurée pour la résolution de certains problèmes inverses mal-posés. Les résultats, que nous présentons dans cette première partie, ont fait l'objet de plusieurs publications : une première publication [73] dans *Electronic journal of Statistics* et une autre [74] dans *IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*.

Chapitre 1 : Le Chapitre 1 introduit brièvement le cadre et les outils mathématiques utilisés tout au long de cette thèse. Puis, nous rappelons quelques notions sur les ondelettes qui serviront dans la suite de la thèse.

Chapitre 2 : Ce chapitre introduit précisément le principal objet de cette thèse, à savoir, les estimateurs de seuillage adaptatifs avec parcimonie structurée. On s'intéresse à l'estimation minimax des dérivées $d-$ième d'une fonction à partir d'observations issues d'un modèle de convolution multicanal hétéroscédastiques. Nous construisons un estimateur adaptatif basé sur un seuillage en ondelettes par bloc. En adoptant l'approche minimax pour l'erreur quadratique moyenne intégrée (MISE), nous prouvons que l'estimateur atteint une vitesse de convergence quasi-optimale sur des boules de Besov. Nous avons ainsi établi que l'estimateur du seuillage par bloc de James-Stein atteint les vitesses quasi-minimax (à un facteur log près) sur ces boules. Cette étude est complétée par des simulations numériques en accord avec les résultats théoriques obtenus. L'estimateur proposé présente des résultats se situant au niveau de l'état de l'art (en terme de PSNR) comparé aux estimateurs classiquement utilisés dans ce contexte.

Chapitre 3 : Ce chapitre vient compléter et enrichir le précédant d'un point de vue

numérique. Nous proposons, pour la résolution de problèmes inverses avec bruit additif Gaussien blanc, une procédure adaptative d'estimation permettant de sélectionner les paramètres de l'estimateur de manière optimale. L'approche consiste à écrire le problème dans le domaine des ondelettes, puis à sélectionner à la fois la taille des blocs et le paramètre de seuillage à chaque niveau de résolution en minimisant l'estimateur de Stein sans biais du risque (SURE). L'algorithme résultant est simple à mettre en œuvre et rapide. Nous présentons des expériences de simulation pour appuyer nos conclusions.

Partie II : Cette seconde partie est dédiée à l'étude d'estimateurs adaptatifs de densité dans le cadre de différents modèles statistiques originaux. Les Chapitres 4,5 et 6 ont fait l'objet de trois publications : un article accepté pour publication [24] à *The Canadian Journal of Statistics*, un autre accepté pour publication [27] à *Journal of Statistical Theory and Practice* et un troisième [72] accepté pour publication dans *Communications in Statistics - Simulation and Computation*.

Chapitre 4 : Le problème d'estimation non-paramétrique de densité pour un modèle de densité pondérée, incluant une large gamme de problèmes (e.g. maximum ou minimum d'un nombre fixe ou aléatoire de variables aléatoires *i.i.d.*), fait l'objet du Chapitre 4 . Plus précisément, nous étudions l'estimation d'une densité pondérée prenant la forme $g = w(F)f$, où f désigne une densité inconnue, F la fonction de répartition associée, et w un poids connu (non négatif). Nous construisons ici un nouvel estimateur non-paramétrique adaptatif pour g basé sur approche plug-in et un seuillage dur en ondelettes par bloc. Pour une large classe de modèles, nous montrons qu'il atteint des vitesses rapides de convergence sous le risque L_p avec $p \geq 1$, sur des boules de Besov. Notre estimateur représente une solution intéressante pour l'estimation de densité et permet d'obtenir de bons résultats par rapport aux méthodes de l'état de l'art. Nous présentons également une étude de simulation pour illustrer les performances de notre approche.

Chapitre 5 : Les deux derniers chapitres sont consacrés au même problème, l'estimation du $m-$ème produit de convolution d'une densité avec elle même. En particulier, dans le Chapitre 5, on observe la somme de m variables aléatoire *i.i.d.* et nous considérons un estimateur de seuillage dur, à notre connaissance jamais utilisé dans ce contexte, puis nous montrons qu'il atteint une vitesse de convergence rapide sur des boules de Besov standards. A l'aide d'une simulation, nous comparons les performances de notre estimateur à celles de deux estimateurs à noyaux. En particulier, pour une grande variété de fonctions, nous montrons que les performances pratiques de notre estimateur de seuillage dépassent celles des deux estimateurs à noyau considérés.

Chapitre 6 : Ce chapitre complète, étend et améliore les résultats du chapitre précédant. Nous proposons une procédure adaptative au moyen d'outils différents de ceux utilisés dans le reste de la thèse, à savoir, un estimateur à noyau définit à partir d'outils d'analyse de Fourier et reposant, pour le choix de la fenêtre, sur la méthode dite de Goldenshluger-Lepski. Cet estimateur a la particularité d'atteindre une vitesse de convergence rapide pour une large classe de densités. Les vitesses de convergence sont déterminés sous des hypothèses faibles sur sur la densité inconnue. Les résultats théoriques obtenus sont illustrés à partir de simulation numérique et mettent en évidence l'efficacité de la procédure de sélection de fenêtre. Enfin, nous proposons une application à des données réelles issues de réclamations d'assurance.

Première partie

Problèmes inverses et estimation fonctionnelle adaptative

Chapitre 1

Cadre et outils mathématiques

Sommaire

1.1 Cadre et outils mathématiques

1.1.1 Analyse multirésolution et ondelettes

Nous utiliserons dans toute cette thèse des bases d'ondelettes orthogonales. La nature de celles-ci changera suivant le modèle statistique traité.

La définition de ces ondelettes est basée sur celle des approximations (ou analyses) multirésolutions.

Définition 1.1.1. *(analyse multirésolution). Une suite* $\{V_j\}_{j\in\mathbb{Z}}$ *de sous-espaces fermés* $\{V_j\}_{j\in\mathbb{Z}}$ *de* $L^2(\mathbb{R})$ *est une approximation multirésolution si elle vérifie les* 6 *propriétés suivantes :*

$$\forall (i,j) \in \mathbb{Z}, \quad f(x) \in V_j \Leftrightarrow f(x - 2^j k) \in V_j, \tag{1.1}$$

$$\forall j \in \mathbb{Z}, \quad V_{j+1} \subset V_j, \tag{1.2}$$

$$\forall j \in \mathbb{Z}, \quad f(x) \in V_j \Leftrightarrow f\left(\frac{x}{2}\right) \in V_{j+1}, \tag{1.3}$$

$$\lim_{j\to+\infty} V_j = \bigcap_{j\in\mathbb{Z}} V_j = \{0\}, \tag{1.4}$$

$$\lim_{j\to-\infty} V_j = \overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}). \tag{1.5}$$

$$\exists \phi \in V_0 \text{ tel que } \{\phi(x-n)\}_{n\in\mathbb{Z}} \text{ soit une base orthonormée de } V_0. \tag{1.6}$$

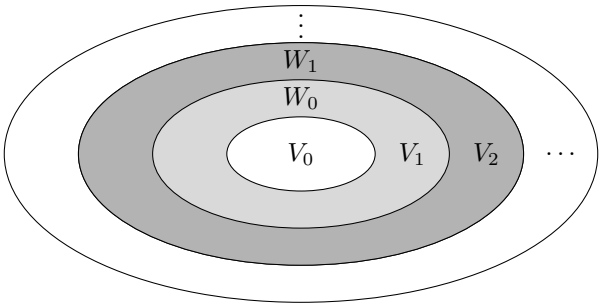


FIGURE 1.1 – Espaces multirésolutions emboîtés.

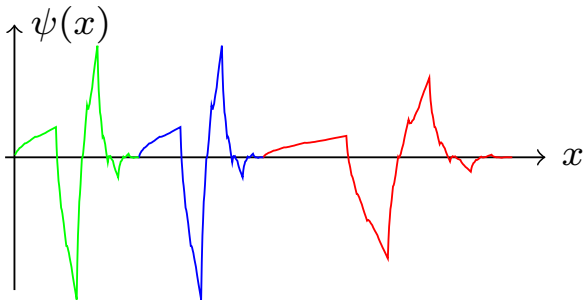


FIGURE 1.2 – Exemple d'une ondelette de Daubechies dilatée et translatée.

Ces propriétés définissent une analyse multirésolution dyadique sur $L^2(\mathbb{R})$. L'idée est de projeter un signal $f(x) \in L^2(\mathbb{R})$ appartenant à un espace V_j sur un sous-espace V_{j+1} et un sous-espace W_{j+1} dans le but de réduire la résolution de moitié.

L'analyse multirésolution, introduite par Meyer et Mallat [66], est un outil puissant de traitement du signal qui permet de décomposer un signal à plusieurs échelles (ou résolutions) et de le reconstruire à partir des éléments de cette décomposition. Elle correspond au calcul d'un ensemble de coefficients qui constituent une représentation plus ou moins précise d'un signal. Ainsi, les ondelettes permettent de représenter les détails gagnés lors du passage d'une résolution à la résolution plus fine suivante.

1.1.2 Les ondelettes à support compact

Dans les chapitres 4 et 5 nous travaillons avec des bases d'ondelettes à support compact sur un intervalle $[-b, b]$. Plus précisément, nous considérons les ondelettes de Daubechies db_{2N} (voir la FIGURE 1.2).

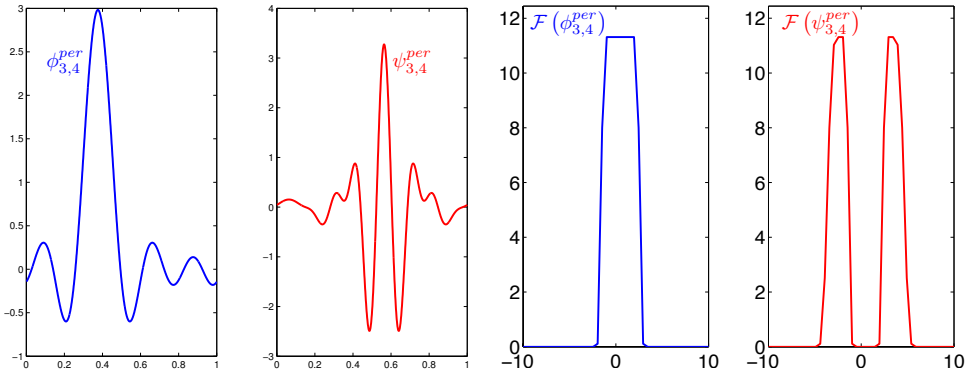
Définition 1.1.2 (base d'ondelettes sur un intervalle compact)**.** *Soit $N \geq 1$. Soient ϕ une fonction d'échelle et ψ une ondelette mère N-régulière toutes deux à support compact. On pose :*

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k).$$

Alors il existe un entier τ tel que, pour tout entier $j_1 \geqslant \tau$, la famille

$$\mathcal{B} = \{\phi_{j_1,k}, k \in \mathbb{Z};\ \psi_{j,k}, j \in \mathbb{N} - \{0, \ldots, j_1 - 1\}, k \in \mathbb{Z}\}, \tag{1.7}$$

forme une base orthonormée de $\mathbb{L}^2([-b, b])$.



(a) (b)

FIGURE 1.3 – Exemple d'une fonction d'échelle et d'une ondelette de Meyer périodisées (a) et leurs transformées de Fourier (b).

Pour les définitions de fonction d'échelle, d'ondelette mère et de N-régularité, voir par example [66].

D'autres exemples de bases d'ondelettes sur un intervalle $[-b, b]$ peuvent être utilisés.

1.1.3 Les bases d'ondelettes d'Yves Meyer

Dans les chapitres 2 et 3, nous considérons une base d'ondelettes orthonormale générée par dilatation et translation d'une ondelette "père" (ou fonction d'échelle) ϕ et d'une ondelette "mère" ψ de Meyer. Les caractéristiques des ces ondelettes sont :

— les transformées de Fourier de ϕ et ψ sont à support compact. Plus précisément, on a

$$\begin{cases} \text{supp}\left(\mathcal{F}(\phi)\right) \subset [-4\pi 3^{-1}, 4\pi 3^{-1}], \\ \text{supp}\left(\mathcal{F}(\psi)\right) \subset [-8\pi 3^{-1}, -2\pi 3^{-1}] \cup [2\pi 3^{-1}, 8\pi 3^{-1}], \end{cases} \tag{1.8}$$

où supp désigne le support, pour tout $h \in \mathbb{L}^2_{per}([0,1])$, $\mathcal{F}(h)$ la transformée de Fourier de h définie par

$$\mathcal{F}(h)(x) = \int_0^1 h(x) e^{-2i\pi xy} dy, \quad x \in \mathbb{R}.$$

— (ϕ, ψ) est r-régulière pour un choix de r tel que $r \in \mathbb{N}$, i.e. $\phi \in \mathcal{C}^r$, $\psi \in \mathcal{C}^r$ et, pour tout $u \in \{0, \dots, r\}$,

$$\int_{-\infty}^{\infty} x^u \psi(x) dx = 0. \tag{1.9}$$

Une conséquence de (1.8) et (1.9) est que, pour tout $m \in \mathbb{N}$ et tout $u \in \{0, \dots, r\}$,

$$\sup_{x \in \mathbb{R}} (|\phi^{(u)}(x)| \left(x^2 + 1\right)^m) < \infty, \quad \sup_{x \in \mathbb{R}} (|\psi^{(u)}(x)| \left(x^2 + 1\right)^m) < \infty. \tag{1.10}$$

Dans les chapitres 2 et 3, nous utilisons des ondelettes périodisées sur l'intervalle $[0,1]$ (voir FIGURE 1.3). Pour tout $x \in [0,1]$, pour $j \geq 0$ et $k \in \{0, \dots, 2^j - 1\}$ fixé, on pose

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$$

les fonctions de la base d'ondelettes, et

$$\phi_{j,k}^{\text{per}}(x) = \sum_{\ell \in \mathbb{Z}} \phi_{j,k}(x-l), \quad \psi_{j,k}^{\text{per}}(x) = \sum_{\ell \in \mathbb{Z}} \psi_{j,k}(x-l),$$

les fonctions de base périodisées. Il existe un entier τ tel que la collection ζ définie par

$$\zeta = \left\{ \phi_{\tau,k}^{\text{per}}(\cdot),\ k \in \{0,\dots,2^\tau - 1\};\ \ \psi_{j,k}^{\text{per}}(\cdot),\ j \geq \tau,\ k \in \{0,\dots,2^j-1\} \right\}$$

constitue une base orthonormée de $\mathbb{L}^2_{\text{per}}([0,1])$. Dans la suite, l'exposant "per" sera supprimé pour ne pas alourdir les notations.

Alors, pour tout $j_1 \geq \tau$, une fonction $h \in \mathbb{L}^2_{\text{per}}([0,1])$ peut être développée en série d'ondelettes

$$h(x) = \sum_{k=0}^{2^{j_1}-1} \alpha_{j_1,k}\phi_{j_1,k}(x) + \sum_{j=j_1}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x), \quad x \in [0,1],$$

où

$$\alpha_{j_1,k} = \int_0^1 h(t)\overline{\phi_{j_1,k}}(t)dt, \quad \beta_{j,k} = \int_0^1 h(t)\overline{\psi_{j,k}}(t)dt. \tag{1.11}$$

Pour plus de détails sur les ondelettes de Meyer et les décompositions en ondelettes, voir par exemple [29], [101] et [102].

1.2 Estimation par seuillage

Le choix d'une représentation appropriée d'une fonction est souvent cruciale pour résoudre un problème donné. On peut par exemple décomposer une fonction inconnue f que l'on cherche à estimer en séries polynômiale ou encore en série de Fourier. En générale, si la fonction f présente des singularités, on utilise plutôt un développement en série d'ondelettes. Si le nombre de coefficients proche de zéro de la décomposition est faible, on parle de représentation creuse ou parcimonieuse de la fonction f relativement à la base considérée.

1.2.1 Approximation linéaire

Dans le cadre de l'estimation de densité, les premiers estimateurs linéaires en ondelettes ont été introduits par [42] et [57, 58]. L'approximation linéaire d'une fonction f consiste à ne conserver qu'un nombre fini de coefficients d'ondelettes de la décomposition en ondelettes. Autrement dit, une approximation linéaire $\hat{f}_M^l$ d'une fonction f appartenant à un certain espace de Hilbert H projette la fonction dans une base orthonormée $B = \{g_m\}_{m\in\mathbb{N}}$ de H sur les M premiers termes du développement :

$$\hat{f}_M^l = \sum_{m=0}^{M-1} < f, g_m > g_m,$$

Si la régularité de f est suffisante et connue, on peut généralement construire des estimateurs performants en minimisant l'erreur d'approximation linéaire. Lorsque l'on considère une base d'ondelettes, on peut montrer que l'erreur d'approximation est minimale lorsque la fonction f appartient à un espace de Sobolev. Cependant, les M premiers termes de la décomposition ne sont pas toujours les plus appropriés pour estimer f.

1.2.2 Approximation non-linéaire

L'approximation adaptative par seuillage des coefficients d'ondelettes constitue une alternative intéressante à l'approximation linéaire. Les premiers estimateurs de seuillage en ondelettes ont été introduits par [38, 39] pour le modèle de régression à pas équidistants et [41] dans le cadre de l'estimation de densités. Une telle approximation vise à sélectionner les meilleurs coefficients d'ondelettes dans la base d'ondelettes considérée de manière à décrire la fonction f avec un minimum de coefficients. Cette sélection s'opère par seuillage, c'est à dire en ne conservant que les M coefficients dont la magnitude dépasse un certain seuil. On obtient un estimateur $\hat{f}_M^{nl}$ de la forme

$$\hat{f}_M^{nl} = \sum_{(j,k)\in I_M} < f, \psi_{j,k} > \psi_{j,k},$$

où I_M désigne les indices des M coefficients d'ondelettes de plus grande amplitude $| < f, \psi_{j,k} > |$.

De manière analogue, nous pouvons montrer que l'erreur d'approximation est minimale lorsque f appartient à un espace de Besov.

1.2.3 Boules de Besov

Il existe différents moyens de de mesurer la régularité de fonctions, le plus souvent au moyen d'un espace de régularité approprié et d'une norme associée. Des exemples classiques sont les espaces de Sobolev et les espaces de Besov. Ces espaces permettent de caractériser la vitesse à laquelle une fonction f peut-être approcher par un estimateur en ondelettes. En effet, les espaces Besov sont des espaces fonctionnels très larges qui décrivent les propriétés de régularité des fonctions et un de leur principale intérêt est qu'ils peuvent être caractérisé en termes de coefficients d'ondelettes :

Soit $M > 0$, $s > 0$, $p \geq 1$ et $r \geq 1$. On dit qu'une fonction h appartient à la boule de Besov $B_{p,r}^s(M)$ si et seulement si il existe une constante $M^* > 0$ (qui dépend de M) tel que les coefficients d'ondelettes associés (1.11) satisfont

$$2^{\tau(1/2-1/p)} \left(\sum_{k=0}^{2^\tau - 1} |\alpha_{\tau,k}|^p \right)^{1/p} + \left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*. \tag{1.12}$$

avec un paramètre de régularité $s > 0$, et les paramètres de norme : $0 < p \leq \infty$ et $0 < r \leq \infty$. Pour un choix particulier de paramètres s, p et r, ces ensembles contiennent les boules de Hölder et de Sobolev.

Nous renvoyons le lecteur à [69] et [37] pour une introduction détaillée à la théorie de l'approximation non-linéaire.

Chapitre 2

Block thresholding for wavelet-based estimation of function derivatives from a heteroscedastic multichannel convolution model

We observe n heteroscedastic stochastic processes $\{Y_v(t)\}_v$, where for any $v \in \{1, \dots, n\}$ and $t \in [0,1]$, $Y_v(t)$ is the convolution product of an unknown function f and a known blurring function g_v corrupted by Gaussian noise. Under an ordinary smoothness assumption on $g_1, \dots, g_n$, our goal is to estimate the d-th derivatives (in weak sense) of f from the observations. We propose an adaptive estimator based on wavelet block thresholding, namely the "BlockJS estimator". Taking the mean integrated squared error (MISE), our main theoretical result investigates the minimax rates over Besov smoothness spaces, and shows that our block estimator can achieve the optimal minimax rate, or is at least nearly-minimax in the least favorable situation. We also report a comprehensive suite of numerical simulations to support our theoretical findings. The practical performance of our block estimator compares very favorably to existing methods of the literature on a large set of test functions.

Sommaire

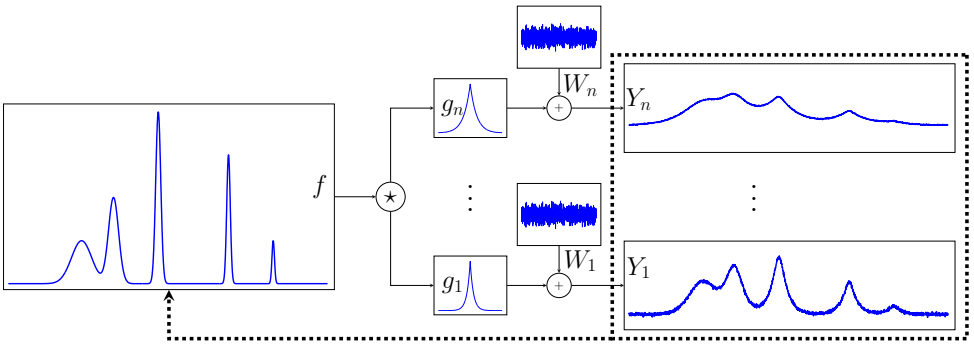


FIGURE 2.1 – Heteroscedastic multichannel convolution model.

2.1 Problem statement and motivation

Suppose that we observe n stochastic processes $Y_1(t), \dots, Y_n(t)$, $t \in [0,1]$ where, for any $v \in \{1, \dots, n\}$,

$$dY_v(t) = (f \star g_v)(t)dt + \epsilon dW_v(t), \qquad t \in [0,1], \qquad n \in \mathbb{N}^*, \tag{2.1}$$

$\epsilon > 0$ is the noise level, $(f \star g_v)(t) = \int_0^1 f(t-u)g_v(u)du$ denotes the convolution product on $[0,1]$, $W_1(t), \dots, W_n(t)$ are n unobserved independent standard Brownian motions, for any $v \in \{1, \dots, n\}$, $g_v : [0,1] \to \mathbb{R}$ is a known blurring function and $f : [0,1] \to \mathbb{R}$ is the unknown function that we target. We assume that f and $g_1, \dots, g_n$ belong to $\mathbb{L}^2_{\text{per}}([0,1]) = \{h;$ h is 1-periodic on $[0,1]$ and $\int_0^1 h^2(t)dt < \infty\}$. The focus of this chapter is to estimate f and its derivatives $f^{(d)}$ (to be understood in weak or distributional sense) from $Y_1(t), \dots, Y_n(t)$, $t \in [0,1]$. This is in general a severely ill-posed inverse problem. Application fields cover biomedical imaging, astronomical imaging, remote-sensing, seismology, etc. This list is by no means exhaustive.

In the following, any function $h \in \mathbb{L}^2_{\text{per}}([0,1])$ can be represented by its Fourier series

$$h(t) = \sum_{\ell \in \mathbb{Z}} \mathcal{F}(h)(\ell) e^{2i\pi\ell t}, \quad t \in [0,1],$$

where the equality is intended in mean-square convergence sense, and $\mathcal{F}_\ell(h)$ denotes the Fourier series coefficient given by

$$\mathcal{F}(h)(\ell) = \int_0^1 h(t)e^{-2i\pi\ell t}dt, \qquad \ell \in \mathbb{Z},$$

whenever this integral exists. The notation $\overline{\cdot}$ will stand for the complex conjugate.

2.2 Relation to prior work

There is an extensive statistical literature on wavelet-based deconvolution problems. For obvious space limitations, we only focus on some of them.

In the special case where $g_1 = \cdots = g_n$, (2.1) reduces to the form

$$d\widetilde{Y}(t) = (f \star g_1)(t)dt + \epsilon n^{-1/2} d\widetilde{W}(t), \qquad t \in [0,1], \tag{2.2}$$

where $\widetilde{Y}(t) = (1/n)\sum_{v=1}^{n} Y_v(t)$, and $\widetilde{W}(t) = (1/n^{1/2})\sum_{v=1}^{n} W_v(t)$ is standard Brownian motion. In such a case, (2.2) becomes the standard deconvolution which attracted attention of a number of researchers spanning a wide range of fields including signal processing and statistics. For instance, wavelet-based estimators of f have been constructed and their asymptotic performance investigated in a number of papers, see e.g. [41, 14, 15, 16, 21, 54]. When $g_1, \ldots, g_n$ are not necessarily equal, estimators of f and their minimax rates under the mean integrated squared error (MISE) over Besov balls were proposed in [33, 78, 79, 80]. These authors develop wavelet thresholding estimators (hard thresholding in [33, 80] and block thresholding in [78, 79]) under various assumptions on $g_1, \ldots, g_n$ (typically, ordinary smooth and super-smooth case, or boxcar blurring functions).

Estimating the derivative of a function on the basis of noisy and blurred observations is of paramount importance in many fields such as signal processing, control or mathematical physics. For instance detecting the singularities of f or characterizing its concavity or convexity properties is a longstanding problem in signal processing. The estimation of the derivatives from noisy solutions of ordinary or partial differential equations is typical in many areas of mathematical physics such as astronomy or fluid dynamics. In the case where $d = 0$, several examples of recovering initial/boundary conditions from observations of a noisy and blurred solution of a PDE (*e.g.*, parabolic, hyperbolic or elliptic PDE) are given in [78]. For higher-order derivatives (typically $d = 1$ or $d = 2$), there are also physical applications where such a model occurs. We mention for instance frequency self-deconvolution encountered in the field of chemometrics. In this context, the first and the second derivatives can be used to detect important information or features in the raw spectra ; see e.g. [4]). The wavelet estimator developed in the chapter could be an interesting alternative to commonly used methods in this area.

The derivatives estimation have already been investigated from several standard nonparametric models. If we only restrict the review to wavelet methods, we refer to [9, 23] for model (2.2) and to [85, 17, 18] for density estimation problems.

In this chapter, considering an appropriate ordinary smoothness assumption on the functions $g_1, \ldots, g_n$, we develop an adaptive wavelet-based block estimator $\widehat{f^{(d)}}$ of $f^{(d)}$ from (2.1), $d \in \mathbb{N}$. It is constructed using a periodized Meyer wavelet basis and a particular block thresholding rule which goes by the the name of BlockJS ; see [8] for the original construction of BlockJS in the standard Gaussian noise model, and [13, 9, 97, 25] for further developments on BlockJS. Adopting the minimax approach under the MISE over Besov balls, we investigate the upper bounds of our estimator. This is featured in Theorem 2.5.1. We prove that the rates of our estimator are nearly optimal by establishing a lower bound as stated in Theorem 2.5.2.

Our work is related to some prior art in the literature. To the best of our knowledge, the closest ones are those of [78, 79]. For the case where $d = 0$ and the blurring function is ordinary-smooth or super-smooth, [78, 79, Theorems 1 and 2] provide the upper and lower bounds of the MISE over Besov balls for a block hard thresholding estimator from the functional deconvolution model [1]. These bounds match ours but only for $d = 0$. In

1. This is a more general model which reduces to the multichannel deconvolution model when observed at a finite number of distinct points, see [78, Section 5] for further details.

this respect, our work goes one step further as it tackles the estimation (with a different wavelet estimator) of f and its derivatives. As far as the methods of proof are concerned, we use similar tools (concentration and moment inequalities as well as the general result in [25]) as theirs for the upper bound, but the proof of the lower bounds are different. However unlike [78], we only cover the ordinary smooth convolution, while their results apply also to the super-smooth case. On the practical side, for $d = 0$, we will show in Section 2.6 that BlockJS behaves better than block hard thresholding [79, (2.9)] over several test functions that contain different degrees of irregularity.

The chapter is organized as follows. Section 2.3 gives a brief account of periodized Meyer wavelets and Besov balls. Section 2.4 states ordinary smoothness assumption on $g_1, \ldots, g_n$, and then constructs the BlockJS-based estimator. The minimax upper and lower bounds of this estimator are investigated in Section 2.5. Section 2.6 describes and discusses the simulation results, before drawing some conclusions in Section 2.7. The proofs are deferred to Section 2.8 awaiting inspection by the interested reader.

2.3 Wavelets and Besov balls

2.3.1 Periodized Meyer Wavelets

We consider an orthonormal wavelet basis generated by dilations and translations of a "father" Meyer-type wavelet ϕ and a "mother" Meyer-type wavelet ψ. These wavelets enjoy the following features.

- They are smooth and frequency band-limited, i.e. the Fourier transforms of ϕ and ψ have compact supports with

$$\begin{cases} \text{supp}\,(\mathcal{F}(\phi)) \subset [-4\pi 3^{-1}, 4\pi 3^{-1}], \\ \text{supp}\,(\mathcal{F}(\psi)) \subset [-8\pi 3^{-1}, -2\pi 3^{-1}] \cup [2\pi 3^{-1}, 8\pi 3^{-1}], \end{cases} \tag{2.3}$$

  where supp denotes the support.
- The functions (ϕ, ψ) are C^∞ as their Fourier transforms have a compact support, and ψ has an infinite number of vanishing moments as its Fourier transform vanishes in a neighborhood of the origin :

$$\int_{-\infty}^{+\infty} t^u \psi(t) dt = 0, \quad \forall\, u \in \mathbb{N}. \tag{2.4}$$

- If the Fourier transforms of ϕ and ψ are also in C^m for a chosen $m \in \mathbb{N}$, then it can be easily shown that ϕ and ψ obey

$$|\phi(t)| = O\left((1+|t|)^{-m-1}\right), \qquad |\psi(t)| = O\left((1+|t|)^{-m-1}\right) \tag{2.5}$$

  for every $t \in \mathbb{R}$.

For the purpose of this work, we use the periodized wavelet bases on the unit interval. For any $t \in [0,1]$, any integer j and any $k \in \{0, \ldots, 2^j - 1\}$, let

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \qquad \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$$

be the elements of the wavelet basis, and

$$\phi_{j,k}^{\text{per}}(t) = \sum_{\ell \in \mathbb{Z}} \phi_{j,k}(t-\ell), \qquad \psi_{j,k}^{\text{per}}(t) = \sum_{\ell \in \mathbb{Z}} \psi_{j,k}(t-\ell),$$

their periodized versions. There exists an integer j_* such that the collection

$$\left\{\phi^{\text{per}}_{j_*,k}(\cdot),\ k \in \{0,\dots,2^{j_*}-1\}; \psi^{\text{per}}_{j,k}(\cdot),\ j \geq j_*, k \in \{0,\dots,2^j-1\}\right\}$$

forms an orthonormal basis of $\mathbb{L}^2_{\text{per}}([0,1])$. In what follows, the superscript "per" will be dropped from ϕ^{per} and ψ^{per} to lighten the notation.

Let $l \geq j_*$, any function $h \in \mathbb{L}^2_{\text{per}}([0,1])$ can be expanded into a wavelet series as

$$h(t) = \sum_{k=0}^{2^l-1} \alpha_{l,k}\phi_{l,k}(t) + \sum_{j=l}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(t), \qquad t \in [0,1],$$

where

$$\alpha_{l,k} = \int_0^1 h(t)\overline{\phi}_{l,k}(t)dt, \qquad \beta_{j,k} = \int_0^1 h(t)\overline{\psi}_{j,k}(t)dt. \tag{2.6}$$

See [69, Vol. 1 Chapter III.11] for a detailed account on periodized orthonormal wavelet bases.

2.3.2 Besov balls

Let $0 < M < \infty$, $s > 0$, $1 \leq p, r \leq \infty$. Among the several characterizations of Besov spaces for periodic functions on $\mathbb{L}^p([0,1])$, we will focus on the usual one based on the corresponding coefficients in a sufficiently q-regular (periodized) wavelet basis ($q = \infty$ for Meyer wavelets). More precisely, we say that a function h belongs to the Besov ball $\mathbf{B}^s_{p,r}(M)$ if and only if $\int_0^1 |h(t)|^p dt \leq M$, and there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (2.6) satisfy

$$2^{j_*(1/2-1/p)}\left(\sum_{k=0}^{2^{j_*}-1}|\alpha_{j_*,k}|^p\right)^{1/p} + \left(\sum_{j=j_*}^{\infty}\left(2^{j(s+1/2-1/p)}\left(\sum_{k=0}^{2^j-1}|\beta_{j,k}|^p\right)^{1/p}\right)^r\right)^{1/r} \leq M^*, \tag{2.7}$$

with a smoothness parameter $0 < s < q$, and the norm parameters p and r. Besov spaces capture a variety of smoothness features in a function including spatially inhomogeneous behavior, see [69].

2.4 The deconvolution BlockJS estimator

2.4.1 The ordinary smoothness assumption

In this study, we consider an orthonormal wavelet basis generated by dilations and translations of a "father" Meyer-type wavelet ϕ and a "mother" Meyer-type wavelet ψ defined in 2.3.1. We focus on the following particular ordinary smoothness assumption on $g_1,\dots,g_n$. We assume that there exist three constants, $c_g > 0$, $C_g > 0$ and $\delta > 1$, and n positive real numbers $\sigma_1,\dots,\sigma_n$ such that, for any $\ell \in \mathbb{Z}$ and any $v \in \{1,\dots,n\}$,

$$c_g\frac{1}{(1+\sigma_v^2\ell^2)^{\delta/2}} \leq |\mathcal{F}(g_v)(\ell)| \leq C_g\frac{1}{(1+\sigma_v^2\ell^2)^{\delta/2}}. \tag{2.8}$$

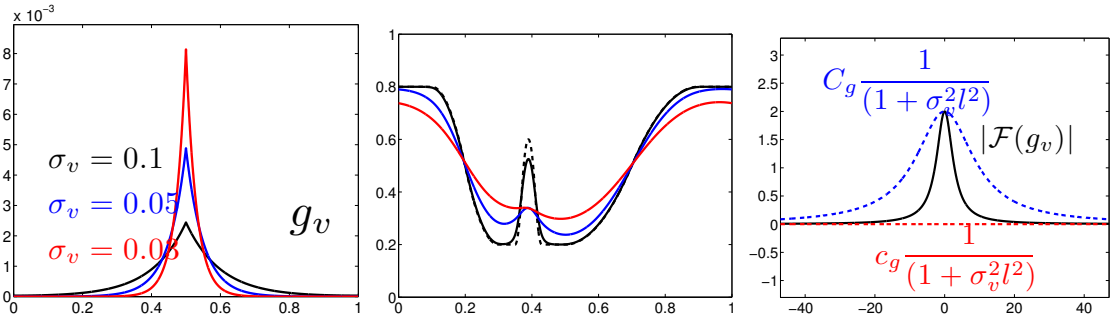


FIGURE 2.2 – Illustration of the *ordinary smooth* assumption for the Laplace distribution.

This assumption controls the decay of the Fourier coefficients of $g_1, \dots, g_n$, and thus the smoothness of $g_1, \dots, g_n$. It is a standard hypothesis usually adopted in the field of nonparametric estimation for deconvolution problems. See e.g. [81], [45] and [54].

Example 2.4.1. *Let $\tau_1, \dots, \tau_n$ be n positive real numbers. For any $v \in \{1, \dots, n\}$, consider the square-integrable 1-periodic function g_v defined by*

$$g_v(t) = \frac{1}{\tau_v} \sum_{m \in \mathbb{Z}} e^{-|t+m|/\tau_v}, \qquad t \in [0,1].$$

Then, for any $\ell \in \mathbb{Z}$, $\mathcal{F}(g_v)(\ell) = 2\left(1 + 4\pi^2 \ell^2 \tau_v^2\right)^{-1}$ and (2.8) is satisfied with $\delta = 2$ and $\sigma_v = 2\pi\tau_v$.

In the sequel, we set

$$\rho_n = \sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{\delta}}. \tag{2.9}$$

For a technical reason that is not restrictive at all (see Section 2.8), we suppose that $\rho_n \geq e$ and $\lim_{n\to\infty} (\ln \rho_n)^v \rho_n^{-1} = 0$ for any $v > 0$.

■ 2.4.2 BlockJS estimator

We suppose that $f^{(d)} \in \mathbb{L}^2_{\text{per}}([0,1])$ and that the ordinary smoothness assumption (2.8) holds, where δ refers to the exponent in the assumption. We are ready to construct our adaptive procedure for the estimation of $f^{(d)}$.

Let $j_1 = \lfloor \log_2(\log \rho_n) \rfloor$ be the coarsest resolution level, and $j_2 = \lfloor (1/(2\delta + 2d + 1)) \log_2(\rho_n / \log \rho_n) \rfloor$, where, for any $a \in \mathbb{R}$, $\lfloor a \rfloor$ denotes the integer part of a. For any $j \in \{j_1, \dots, j_2\}$, let $L = \lfloor \log \rho_n \rfloor$ be the block size.

Let $A_j = \{1, \dots, \lfloor 2^j L^{-1} \rfloor\}$ be the set indexing the blocks at resolution j. For each j, let $\{B_{j,K}\}_{K \in A_j}$ be a uniform and disjoint open covering of $\{0, \dots, 2^j - 1\}$, i.e. $\bigcup_{K \in A_j} B_{j,K} = \{0, \dots, 2^j - 1\}$ and for any $(K, K') \in A_j^2$ with $K \neq K'$, $B_{j,K} \cap B_{j,K'} = \varnothing$ and $\text{Card}(B_{j,K}) = L$, where $B_{j,K} = \{k \in \{0, \dots, 2^j - 1\};\ (K-1)L \leq k \leq KL - 1\}$ is the Kth block.

We define the Block James-Stein estimator (BlockJS) of $f^{(d)}$ by

$$\widehat{f^{(d)}}(t) = \sum_{k=0}^{2^{j_1}-1} \widehat{\alpha}_{j_1,k} \phi_{j_1,k}(t) + \sum_{j=j_1}^{j_2} \sum_{K \in A_j} \sum_{k \in B_{j,K}} \widehat{\beta}^*_{j,k} \psi_{j,k}(t), \quad t \in [0,1], \tag{2.10}$$

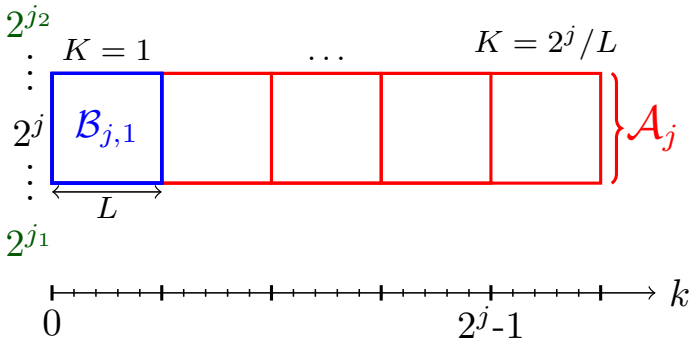


FIGURE 2.3 – Illustration of the blocks partition at a given scale.

where for any resolution j and position $k \in B_{j,K}$ within the Kth block, the wavelet coefficients of $f^{(d)}$ are estimated via the rule

$$\widehat{\beta}^*_{j,k} = \widehat{\beta}_{j,k}\left(1 - \frac{\lambda\epsilon^2\rho_n{}^{-1}2^{2j(\delta+d)}}{\frac{1}{L}\sum_{k\in B_{j,K}}|\widehat{\beta}_{j,k}|^2}\right)_+,$$

with, for any $a \in \mathbb{R}$, $(a)_+ = \max(a, 0)$, $\lambda > 0$, and $\widehat{\alpha}_{j_1,k}$ and $\widehat{\beta}_{j,k}$ are respectively the empirical scaling and detail coefficients, defined as

$$\widehat{\alpha}_{j_1,k} = \frac{1}{\rho_n}\sum_{v=1}^{n}\frac{1}{(1+\sigma_v^2)^\delta}\sum_{\ell\in\mathcal{D}_{j_1}}(2\pi i\ell)^d\frac{\overline{\mathcal{F}(\phi_{j_1,k})}(\ell)}{\mathcal{F}(g_v)(\ell)}\int_0^1 e^{-2\pi i\ell t}dY_v(t)$$

and

$$\widehat{\beta}_{j,k} = \frac{1}{\rho_n}\sum_{v=1}^{n}\frac{1}{(1+\sigma_v^2)^\delta}\sum_{\ell\in\mathcal{C}_j}(2\pi i\ell)^d\frac{\overline{\mathcal{F}(\psi_{j,k})}(\ell)}{\mathcal{F}(g_v)(\ell)}\int_0^1 e^{-2\pi i\ell t}dY_v(t).$$

Notice that thanks to (2.3), for any $j \in \{j_1, \dots, j_2\}$ and $k \in \{0, \dots, 2^j - 1\}$

$$\begin{cases}\mathcal{D}_{j_1} = \operatorname{supp}\left(\mathcal{F}\left(\phi_{j_1,k}\right)\right) \subset [-4\pi 3^{-1}2^{j_1}, 4\pi 3^{-1}2^{j_1}],\\ \mathcal{C}_j = \operatorname{supp}\left(\mathcal{F}\left(\psi_{j,k}\right)\right) \subset [-8\pi 3^{-1}2^{j}, -2\pi 3^{-1}2^{j}] \cup [2\pi 3^{-1}2^{j}, 8\pi 3^{-1}2^{j}].\end{cases} \tag{2.11}$$

2.5 Minimaxity results of BlockJS over Besov balls

2.5.1 Minimax upper-bound for the MISE

Theorem 2.5.1 below determines the rates of convergence achieved by $\widehat{f^{(d)}}$ under the MISE over Besov balls.

Theorem 2.5.1. *Consider the model (2.1) and recall that we want to estimate $f^{(d)}$ with $d \in \mathbb{N}$. Assume that (ϕ, ψ) satisfy (2.5) for some $m \geq d$ and (2.8) is satisfied. Let $\widehat{f^{(d)}}$ be the estimator defined by (2.10) with a large enough λ. Then there exists a constant $C > 0$ such that, for any $M > 0$, $p \geq 1$, $r \geq 1$, $s > 1/p$ and n large enough, we have*

$$\sup_{f^{(d)}\in\mathbf{B}^s_{p,r}(M)}\mathbb{E}\left(\int_0^1\left(\widehat{f^{(d)}}(t) - f^{(d)}(t)\right)^2 dt\right) \leq C\varphi_n,$$

where

$$\varphi_n = \begin{cases} \rho_n^{-2s/(2s+2\delta+2d+1)}, & \text{if } p \geq 2, \\ (\log \rho_n/\rho_n)^{2s/(2s+2\delta+2d+1)}, & \text{if } p \in [1,2), s > (1/p - 1/2)(2\delta + 2d + 1). \end{cases}$$

Theorem 2.5.1 will be proved using the more general theorem [25, Theorem 3.1]. To apply this result, two conditions on the wavelet coefficients estimator are required : a moment condition and a concentration condition. They are established in Propositions 2.8.2 and 2.8.3, see Section 2.8 .

Remark 2.5.1. *Theorem 2.5.1 can be generalized by allowing the decay exponent δ of $g_1, \dots, g_n$ to vary across the channels. More precisely, our ordinary assumption (2.8) is uniform over the channels, in the sense that we assumed that the decay exponent δ is the same for all v. Assume now that there exist three constants, $c_g > 0$, $C_g > 0$, and $2n$ positive real numbers $\sigma_1, \dots, \sigma_n, \delta_1, \dots, \delta_n$ with $\min(\delta_1, \dots, \delta_n) > 1/2$ such that, for any $\ell \in \mathbb{Z}$ and any $v \in \{1, \dots, n\}$,*

$$|\mathcal{F}(g_v)(\ell)| \leq C_g \frac{1}{(1 + \sigma_v^2 \ell^2)^{\delta_v/2}}.$$

Set

$$\omega = \min(\delta_1, \dots, \delta_n).$$

Let us now define the BlockJS estimator of $f^{(d)}$ as in Section 2.4.2 but with the weights

$$\rho_n = \sum_{v=1}^{n} \frac{1}{(1 + \sigma_v^2)^{\delta_v}},$$

resolution levels $j_1 = \lfloor \log_2(\log \rho_n) \rfloor$, $j_2 = \lfloor (1/(2\omega + 2d + 1)) \log_2(\rho_n / \log \rho_n) \rfloor$, and

$$\widehat{\beta}^*_{j,k} = \widehat{\beta}_{j,k} \left(1 - \frac{\lambda \epsilon^2 {\rho_n}^{-1} 2^{2j(\omega+d)}}{\frac{1}{L} \sum_{k \in B_{j,K}} |\widehat{\beta}_{j,k}|^2} \right)_+,$$

$$\widehat{\alpha}_{j_1,k} = \frac{1}{\rho_n} \sum_{v=1}^{n} \frac{1}{(1 + \sigma_v^2)^{\delta_v}} \sum_{\ell \in \mathcal{D}_{j_1}} (2\pi i \ell)^d \frac{\overline{\mathcal{F}(\phi_{j_1,k})}(\ell)}{\mathcal{F}(g_v)(\ell)} \int_0^1 e^{-2\pi i \ell t} dY_v(t)$$

and

$$\widehat{\beta}_{j,k} = \frac{1}{\rho_n} \sum_{v=1}^{n} \frac{1}{(1 + \sigma_v^2)^{\delta_v}} \sum_{\ell \in \mathcal{C}_j} (2\pi i \ell)^d \frac{\overline{\mathcal{F}(\psi_{j,k})}(\ell)}{\mathcal{F}(g_v)(\ell)} \int_0^1 e^{-2\pi i \ell t} dY_v(t).$$

Then Theorem 2.5.1 holds with the convergence rate

$$\varphi_n = \begin{cases} \rho_n^{-2s/(2s+2\omega+2d+1)}, & \text{if } p \geq 2, \\ (\log \rho_n/\rho_n)^{2s/(2s+2\omega+2d+1)}, & \text{if } p \in [1,2), s > (1/p - 1/2)(2\omega + 2d + 1). \end{cases}$$

Of course, this generalization encompasses the statement of Theorem 2.5.1. This generalized result also tells us that the adaptivity of the estimator makes its performance mainly influenced by the best g_i, that is the one with the smallest δ_i, which is a nice feature.

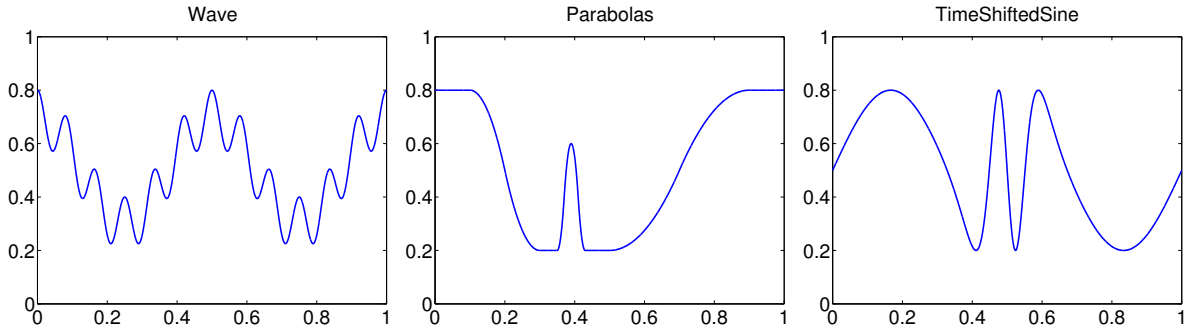


FIGURE 2.4 – Original Signals (a) : Wave. (b) : Parabolas. (c) : TimeShiftedSine.

2.5.2 Minimax lower-bound for the MISE

We now turn to the lower bound of the MISE to formally answer the question whether φ_n is indeed the optimal rate of convergence or not. This is the goal of Theorem 2.5.2 which gives a positive answer.

Theorem 2.5.2. *Consider the model (2.1) and recall that we want to estimate $f^{(d)}$ with $d \in \mathbb{N}$. Assume that (2.8) is satisfied. Then there exists a constant $c > 0$ such that, for any $M > 0$, $p \geq 1$, $r \geq 1$, $s > 1/p$ and n large enough, we have*

$$\inf_{\widetilde{f^{(d)}}} \sup_{f^{(d)} \in \mathbf{B}^s_{p,r}(M)} \mathbb{E}\left(\int_0^1 \left(\widetilde{f^{(d)}}(t) - f^{(d)}(t) \right)^2 dt \right) \geq c\varphi_n^*,$$

where

$$\varphi_n^* = (\rho_n^*)^{-2s/(2s+2\delta+2d+1)}, \qquad \rho_n^* = \sum_{v=1}^{n} \sigma_v^{-2\delta}.$$

and the infimum is taken over all the estimators $\widetilde{f^{(d)}}$ of $f^{(d)}$.

It can then be concluded from Theorem 2.5.1 and Theorem 2.5.2 that the rate of convergence φ_n achieved by $\widehat{f^{(d)}}$ is near optimal in the minimax sense. Near minimaxity is only due to the case $p \in [1, 2)$ and $s > (1/p - 1/2)(2\delta + 2d + 1)$ where there is an extra logarithmic term.

2.6 Simulations results

In the following simulation study we consider the problem of estimating one of the derivatives of a function f from the heteroscedastic multichannel deconvolution model (2.1). Three test functions ("Wave", "Parabolas" and "TimeShiftedSine", initially introduced in [67]) representing different degrees of smoothness were used (see FIGURE 2.4). The "Wave" function was used to illustrate the performance of our estimator on a smooth function. Note that the "Parabolas" function has big jumps in its second derivative.

We have compared the numerical performance of BlockJS to state-of-the-art classical thresholding methods of the literature. In particular we consider the block estimator of [79] and two term-by-term thresholding methods. The first one is the classical hard thresholding and the other one corresponds to the non-negative garrote (introduced in wavelet estimation by [48]). In the sequel, we name the estimator of [79] by 'BlockH', the one of [48] by

'TermJS' and our estimator by 'BlockJS'. For numerical implementation, the test functions were finely discretized by taking T equispaced samples $t_i = i/T \in [0,1]$, $i = 0, \dots, T-1$. The deconvolution estimator was efficiently implemented in the Fourier domain given that Meyer wavelets are frequency band-limited. The performance of the estimators are measured in terms of peak signal-to-noise ratio (PSNR $= 10 \log_{10} \frac{\max_{t_i \in [0,1]} |f^{(d)}(t_i)|^2}{\sum_{i=0}^{T-1} (f^{(d)}(t_i) - f^{(d)}(t_i))^2 / T}$) in decibels (dB)). For any $v \in \{1, \dots, n\}$, the blurring function g_v is that of Example 2.4.1 and was used throughout all experiments.

■ 2.6.1 Monochannel simulation

As an example of homoscedastic monochannel reconstruction (i.e. $n = 1$), we show in FIGURE 2.5 estimates obtained using the BlockJS method from $T = 4096$ equispaced samples generated according to (2.1) with blurred signal-to-noise ratio (BSNR) of 25 dB (BSNR $= 10 \log_{10} \frac{\sum_{i=0}^{T-1} (f \star g_v)(t_i)^2}{T\epsilon^2}$ dB). For $d = 0$, the results are very effective for each test function where the singularities are well estimated. The estimator does also a good job in estimating the first and second-order derivatives, although the estimation quality decreases as the order of the derivative increases. This is in agreement with the predictions of the minimaxity results. We then have compared the performance of BlockJS with BlockH. The blurred signals were corrupted by a zero-mean white Gaussian noise such that the BSNR ranged from 10 to 40 dB. The PSNR values averaged over 10 noise realizations are depicted in FIGURE 2.6 for $d = 0$, $d = 1$ and $d = 2$ respectively. One can see that our BlockJS thresholding estimator produces quite accurate estimates of f, f' and f'' for each test function. These results clearly show that our approach compares favorably to BlockH and that BlockJS has good adaptive properties over a wide range of noise levels in the monochannel setting.

■ 2.6.2 Multichannel simulation

A first point we would like to highlight is the fact that some choices of $\sigma_1, \dots, \sigma_n$ can severely impact the performance of the estimators. To illustrate this, we show in FIGURE 2.7 an example of first derivative estimates obtained using BlockJS from $n = 10$ channels with $T = 4096$ samples and noise level corresponding to BSNR$= 25$ dB, for $\sigma_v = v$ (dashed blue) and σ_v randomly generated in $(0, +\infty)$ (solid blue). With σ_v randomly generated, we can observe a significant PSNR improvement up to 6.85 dB for the first derivative of TimeShiftedSine. Note that this improvement is marginal (about 0.60 dB) for the most regular test signal (i.e. Wave).

We finally report a simulation study by quantitatively comparing BlockJS to the other thresholding estimators described above. For each test function, we generated $T = 4096$ equispaced samples on $[0,1]$ according to (2.1) with varying number of channels ranging from $n = 10$ to 100.

TABLE 2.1 summarizes the results. It shows in particular that BlockJS consistently outperforms the other methods in almost all cases in terms of PSNR. As expected and predicted by our theoretical findings, on the one hand, the performance gets better as the number of channels increases. On the other hand, it degrades with increasing noise level and/or d. Indeed, the derivatives estimation for BSNR$= 10$ dB is rather difficult to estimate, especially for functions having highly irregular derivatives such as "Parabolas" (which has big jumps in its second derivative, see FIGURE 2.5(d)).

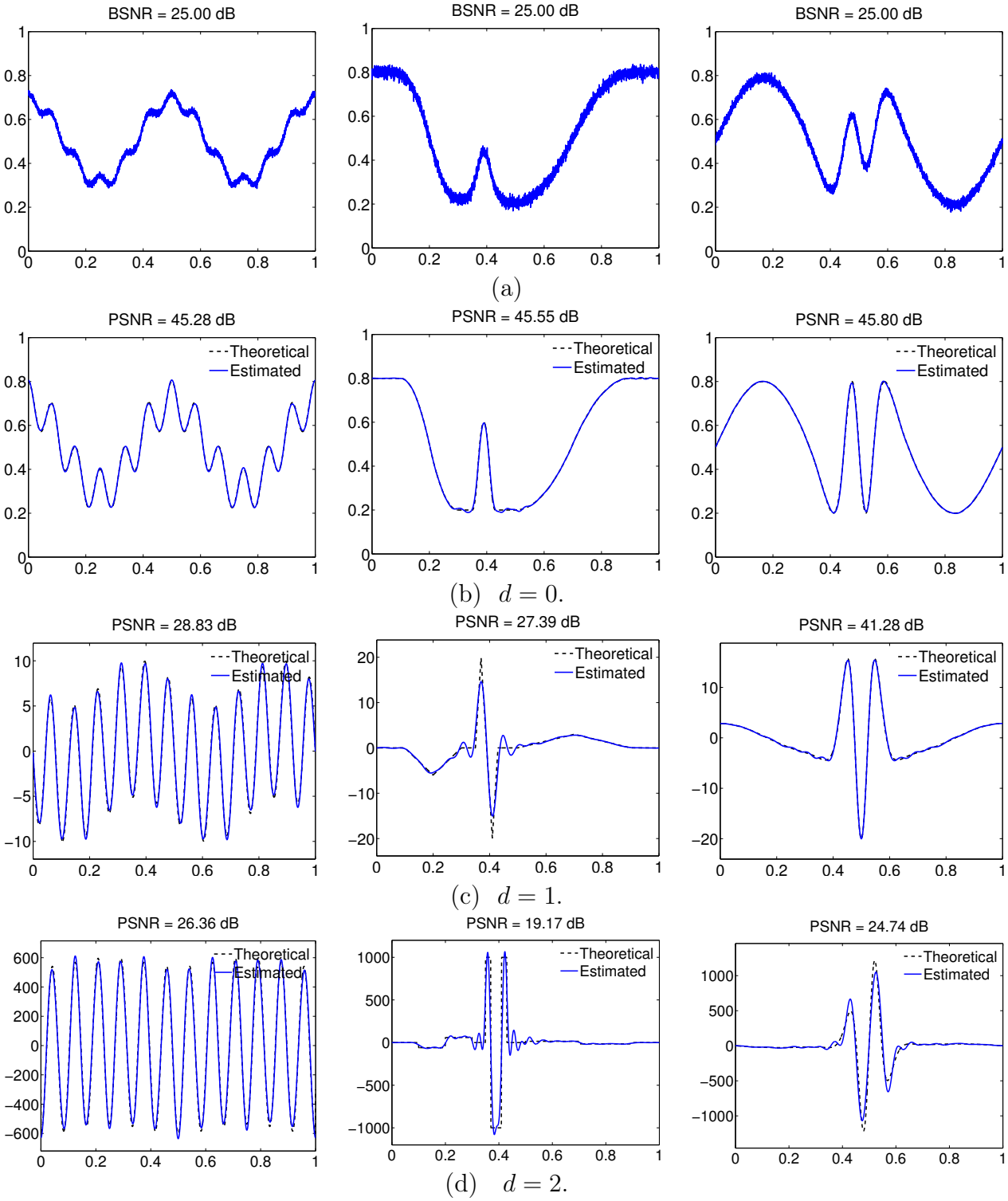


FIGURE 2.5 – Original (dashed) and estimated function/derivatives (solid) using the BlockJS estimator applied to noisy blurred observations shown in (a). (b) : $d = 0$. (c) : $d = 1$ (d) : $d = 2$. From left to right Wave, Parabolas and TimeShiftedSine.

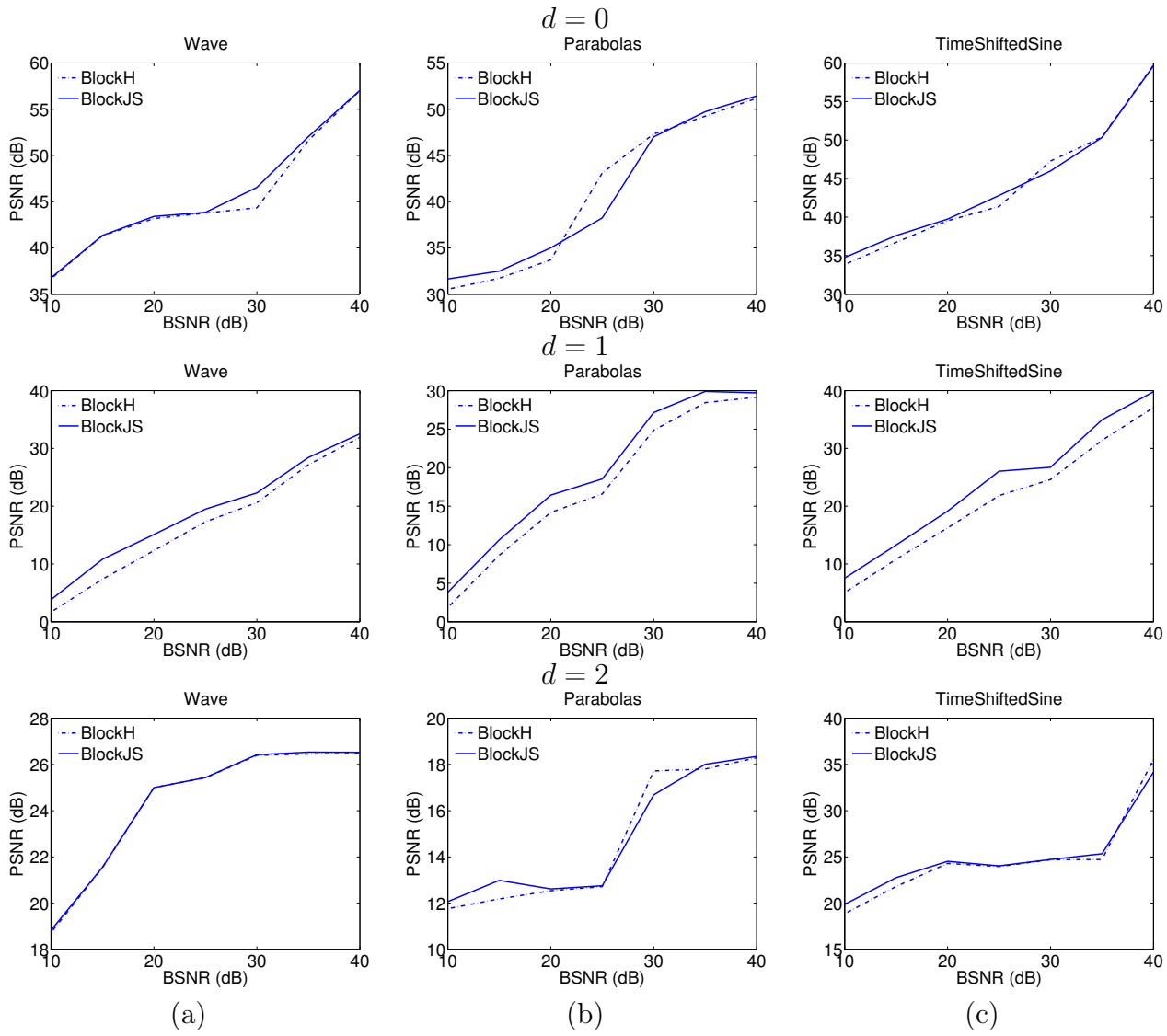


FIGURE 2.6 – Averaged PSNR values as a function of the input BSNR from 10 replications of the noise. (a) : Wave. (b) : Parabolas. (c) : TimeShiftedSine. From top to bottom $d = 0, 1, 2$.

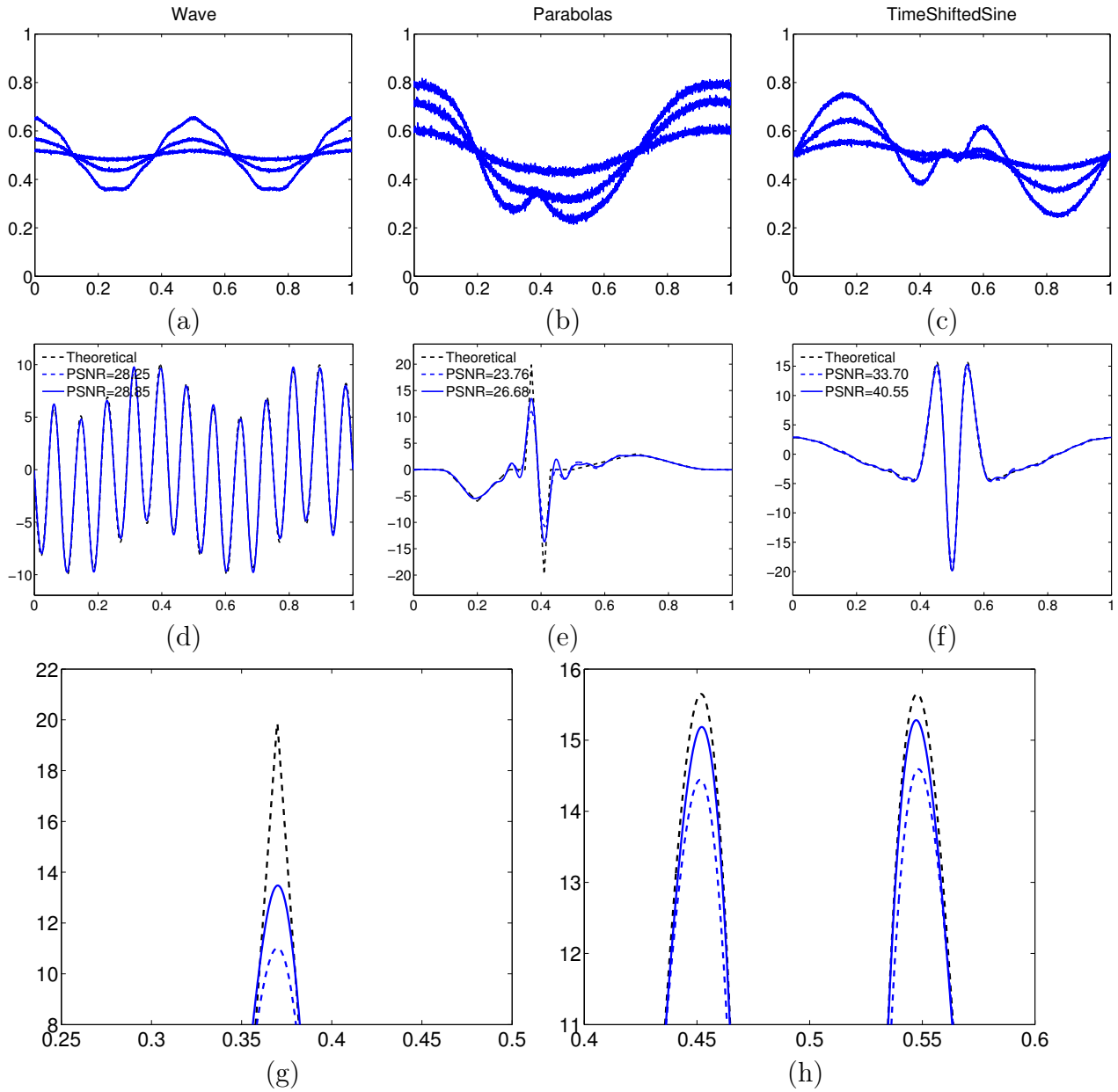


FIGURE 2.7 – Original functions (dashed black) and the estimate for $\sigma_v = v$ (dashed blue) and σ_v randomly generated (solid blue) with $n = 10$ channels. (a)-(c) : noisy blurred observations (3 channels out of 10 shown). (d)-(f) BlockJS estimates of the first derivative. Zoom on the estimates (g) : Parabolas, (h) : TimeShiftedSine.

BSNR= 40 dB

	$d=0$				$d=1$				$d=2$			
n	10	20	50	100	10	20	50	100	10	20	50	100
Wave												
BlockJS	57.42	66.40	66.83	74.62	42.64	43.58	43.94	50.10	22.04	30.34	33.88	36.69
BlockH	57.34	66.31	66.78	74.72	42.43	43.57	43.07	50.07	22.66	29.14	33.78	36.65
TermJS	52.51	61.94	64.86	73.57	40.64	41.56	33.90	49.30	19.17	29.57	30.87	33.20
TermH	50.97	52.72	55.69	74.73	31.39	33.92	37.20	39.48	17.17	28.82	33.89	35.61
TimeShiftedSine												
BlockJS	65.11	65.58	68.47	71.17	42.16	46.25	46.85	49.53	41.09	42.67	43.07	46.35
BlockH	62.11	62.29	62.71	70.43	41.97	45.57	45.75	49.35	39.20	40.55	41.97	42.00
TermJS	64.01	64.73	66.13	68.21	40.96	42.13	43.63	43.98	38.57	41.17	41.73	45.15
TermH	61.84	62.05	67.12	68.80	41.22	42.12	43.32	44.69	39.39	39.81	39.24	43.04
Parabolas												
BlockJS	56.18	57.42	58.10	58.40	29.66	29.76	31.04	31.40	20.63	21.88	21.93	21.90
BlockH	55.92	57.25	57.70	58.10	29.69	29.67	29.91	29.94	20.74	21.09	21.99	21.73
TermJS	54.33	57.05	57.96	58.24	29.21	29.38	29.65	29.79	20.57	20.93	20.91	21.51
TermH	54.59	56.96	57.88	58.10	29.60	29.60	29.90	29.94	20.51	20.96	20.84	21.49

BSNR= 25 dB

	$d=0$				$d=1$				$d=2$			
n	10	20	50	100	10	20	50	100	10	20	50	100
Wave												
BlockJS	44.04	51.93	52.47	59.76	30.14	30.90	30.90	35.93	26.83	26.83	26.98	27.00
BlockH	42.12	51.35	51.82	59.73	28.69	28.69	28.72	35.26	26.89	26.85	26.98	26.48
TermJS	41.67	48.10	49.44	52.59	28.49	28.69	28.72	28.71	25.25	26.62	26.78	26.94
TermH	40.95	49.03	50.22	55.04	28.69	28.69	28.72	28.72	25.28	26.85	26.98	26.86
TimeShiftedSine												
BlockJS	51.85	52.33	55.66	60.49	39.48	41.46	41.88	41.92	27.39	28.72	29.05	35.95
BlockH	52.93	51.51	55.82	60.35	38.68	41.24	41.87	41.84	26.68	29.36	29.54	35.15
TermJS	47.19	47.83	54.45	56.63	29.46	41.34	41.85	41.79	23.66	25.84	25.95	27.58
TermH	47.54	47.47	54.44	59.63	31.42	41.03	40.75	41.79	23.66	25.69	25.91	30.67
Parabolas												
BlockJS	47.81	49.88	52.90	54.40	25.52	26.11	28.74	29.57	17.48	18.24	18.96	20.62
BlockH	47.74	49.23	52.00	53.62	25.43	25.16	29.55	29.69	17.08	18.92	19.00	20.63
TermJS	44.52	49.80	50.80	52.79	24.84	25.78	25.71	25.73	16.39	18.58	19.00	20.57
TermH	43.74	48.84	51.35	53.24	24.86	25.17	25.10	27.98	16.49	18.64	18.95	20.34

BSNR= 10 dB

	$d=0$				$d=1$				$d=2$			
n	10	20	50	100	10	20	50	100	10	20	50	100
Wave												
BlockJS	35.82	43.46	43.91	45.95	26.96	27.34	28.22	28.45	19.02	25.12	26.16	26.78
BlockH	34.08	43.41	43.83	44.60	26.90	27.34	28.19	28.41	19.21	25.16	26.06	26.68
TermJS	33.30	43.16	43.24	44.48	26.67	27.25	28.14	28.36	18.32	18.89	22.87	26.65
TermH	33.22	40.16	39.32	44.27	26.77	27.35	28.15	28.39	18.61	19.07	18.95	26.68
TimeShiftedSine												
BlockJS	38.69	39.28	41.24	48.38	30.12	38.85	39.26	40.19	22.58	22.84	23.37	23.53
BlockH	38.66	38.97	40.12	45.34	26.66	38.22	39.22	39.08	22.58	22.73	23.33	23.48
TermJS	38.41	38.68	40.31	41.45	25.68	29.24	36.79	36.68	22.52	22.82	15.27	23.26
TermH	38.23	38.44	39.07	41.46	26.98	28.10	35.18	36.98	22.49	22.72	15.47	17.56
Parabolas												
BlockJS	35.26	37.04	40.64	44.74	22.03	24.56	24.77	25.56	12.76	12.79	13.09	13.29
BlockH	34.12	35.53	39.29	43.41	22.22	24.54	24.47	25.00	12.64	12.76	12.78	12.77
TermJS	34.29	35.18	39.38	41.37	21.70	24.51	24.63	24.96	12.56	12.76	12.79	12.79
TermH	33.27	34.53	39.45	42.21	21.88	23.98	24.21	24.86	12.70	12.76	12.79	12.79

TABLE 2.1 – Comparison of average PSNR in decibels (dB) over 10 realizations of the noise for $d = 0$, $d = 1$ and $d = 2$. From top to bottom BSNR= $40, 25, 10$ dB.

2.7 Conclusion and perspectives

In this work, an adaptive wavelet block thresholding estimator was constructed to estimate one of the derivative of a function f from the heteroscedastic multichannel deconvolution model. Under ordinary smooth assumption on $g_1, \dots, g_n$, it was proved that it is nearly optimal in the minimax sense. The practical comparisons to state-of-the art methods have demonstrated the usefulness and the efficiency of adaptive block thresholding methods in estimating a function f and its first derivatives in the functional deconvolution setting.

It would be interesting to consider the case where g_v are unknown, which is the case in many practical situations. Another interesting perspective would be to extend our results to a multidimensional setting. These aspects need further investigations that we leave for a future work.

2.8 Proofs

In the following proofs, c and C denote positive constants which can take different values for each mathematical term.

2.8.1 Preparatory results

In the three following results, we consider the framework of Theorem 2.5.1 and, for any integer $j \geq j_*$ and $k \in \{1, \dots, 2^j - 1\}$, we set $\alpha_{j,k} = \int_0^1 f^{(d)}(t)\overline{\phi}_{j,k}(t)dt$ and $\beta_{j,k} = \int_0^1 f^{(d)}(t)\overline{\psi}_{j,k}(t)dt$, the wavelet coefficients (2.6) of $f^{(d)}$.

Proposition 2.8.1 (Gaussian distribution on the wavelet coefficient estimators)**.** *For any integer $j \geq j_*$ and $k \in \{0, \dots, 2^j - 1\}$, we have*

$$\widehat{\alpha}_{j,k} \sim \mathcal{N}\left(\alpha_{j,k}, \epsilon^2 \frac{1}{\rho_n^2} \sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{2\delta}} \sum_{\ell \in \mathcal{D}_j} (2\pi\ell)^{2d} \frac{|\mathcal{F}(\phi_{j,k})(\ell)|^2}{|\mathcal{F}(g_v)(\ell)|^2}\right)$$

and

$$\widehat{\beta}_{j,k} \sim \mathcal{N}\left(\beta_{j,k}, \epsilon^2 \frac{1}{\rho_n^2} \sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{2\delta}} \sum_{\ell \in \mathcal{C}_j} (2\pi\ell)^{2d} \frac{|\mathcal{F}(\psi_{j,k})(\ell)|^2}{|\mathcal{F}(g_v)(\ell)|^2}\right).$$

Démonstration. Let us prove the second point, the first one can be proved in a similar way. For any $\ell \in \mathbb{Z}$ and any $v \in \{1, \dots, n\}$, $\mathcal{F}(f \star g_v)(\ell) = \mathcal{F}(f)(\ell)\,\mathcal{F}(g_v)(\ell)$. Therefore, if we set

$$y_{\ell,v} = \int_0^1 e^{-2\pi i \ell t} dY_v(t), \qquad e_{\ell,v} = \int_0^1 e^{-2\pi i \ell t} dW_v(t),$$

It follows from (2.1) that

$$y_{\ell,v} = \mathcal{F}(f)(\ell)\,\mathcal{F}(g_v)(\ell) + \epsilon e_{\ell,v}. \tag{2.12}$$

Note that, since f is 1-periodic, for any $u \in \{0, \dots, d\}$, $f^{(u)}$ is 1-periodic and $f^{(u)}(0) = f^{(u)}(1)$. By classical properties of the Fourier series, for any $\ell \in \mathbb{Z}$, we have $\mathcal{F}(f^{(d)})(\ell) =$

$(2\pi i\ell)^d \mathcal{F}(f)(\ell)$. The Parseval theorem gives

$$\begin{aligned}\beta_{j,k} &= \int_0^1 f^{(d)}(t)\overline{\psi}_{j,k}(t)dt = \sum_{\ell\in\mathcal{C}_j} \mathcal{F}(f^{(d)})(\ell)\overline{\mathcal{F}\left(\psi_{j,k}\right)}(\ell) \\ &= \sum_{\ell\in\mathcal{C}_j} (2\pi i\ell)^d \, \mathcal{F}(f)(\ell)\overline{\mathcal{F}\left(\psi_{j,k}\right)}(\ell).\end{aligned}$$

Using (2.12), we have

$$\begin{aligned}\widehat{\beta}_{j,k} &= \frac{1}{\rho_n}\sum_{v=1}^n \frac{1}{(1+\sigma_v^2)^\delta} \sum_{\ell\in\mathcal{C}_j} (2\pi i\ell)^d \frac{\overline{\mathcal{F}\left(\psi_{j,k}\right)}(\ell)}{\mathcal{F}(g_v)(\ell)} \, \mathcal{F}(f)(\ell)\, \mathcal{F}(g_v)(\ell) \\ &+ \epsilon\frac{1}{\rho_n}\sum_{v=1}^n \frac{1}{(1+\sigma_v^2)^\delta} \sum_{\ell\in\mathcal{C}_j} (2\pi i\ell)^d \frac{\overline{\mathcal{F}\left(\psi_{j,k}\right)}(\ell)}{\mathcal{F}(g_v)(\ell)} e_{\ell,v} \\ &= \sum_{\ell\in\mathcal{C}_j} (2\pi i\ell)^d \, \mathcal{F}(f)(\ell)\overline{\mathcal{F}\left(\psi_{j,k}\right)}(\ell) \\ &+ \epsilon\frac{1}{\rho_n}\sum_{v=1}^n \frac{1}{(1+\sigma_v^2)^\delta} \sum_{\ell\in\mathcal{C}_j} (2\pi i\ell)^d \frac{\overline{\mathcal{F}\left(\psi_{j,k}\right)}(\ell)}{\mathcal{F}(g_v)(\ell)} e_{\ell,v} \\ &= \beta_{j,k} + \epsilon\frac{1}{\rho_n}\sum_{v=1}^n \frac{1}{(1+\sigma_v^2)^\delta} \sum_{\ell\in\mathcal{C}_j} (2\pi i\ell)^d \frac{\overline{\mathcal{F}\left(\psi_{j,k}\right)}(\ell)}{\mathcal{F}(g_v)(\ell)} e_{\ell,v}.\end{aligned}$$

Since $\{e^{-2\pi i\ell\cdot}\}_{\ell\in\mathbb{Z}}$ is an orthonormal basis of $\mathbb{L}^2_{\text{per}}([0,1])$ and $W_1(t),\dots,W_n(t)$ are i.i.d. standard Brownian motions, $\left(\int_0^1 e^{-2\pi i\ell t}dW_v(t)\right)_{(\ell,v)\in\mathbb{Z}\times\{1,\dots,n\}}$ is a sequence of i.i.d. random variables with the common distribution $\mathcal{N}(0,1)$. Therefore

$$\widehat{\beta}_{j,k} \sim \mathcal{N}\left(\beta_{j,k}, \epsilon^2\frac{1}{\rho_n^2}\sum_{v=1}^n \frac{1}{(1+\sigma_v^2)^{2\delta}} \sum_{\ell\in\mathcal{C}_j} (2\pi\ell)^{2d} \frac{|\,\mathcal{F}\left(\psi_{j,k}\right)(\ell)|^2}{|\,\mathcal{F}(g_v)(\ell)|^2}\right).$$

Proposition 2.8.1 is proved. □

Proposition 2.8.2 (Moment inequalities)**.**

— *There exists a constant $C>0$ such that, for any integer $j\geq j_*$ and $k\in\{0,\dots,2^j-1\}$,*

$$\mathbb{E}\left(|\widehat{\alpha}_{j_1,k}-\alpha_{j_1,k}|^2\right) \leq C\epsilon^2 2^{2(\delta+d)j_1}\rho_n^{-1},$$

— *There exists a constant $C>0$ such that, for any integer $j\geq j_*$ and $k\in\{0,\dots,2^j-1\}$,*

$$\mathbb{E}\left(|\widehat{\beta}_{j,k}-\beta_{j,k}|^4\right) \leq C\epsilon^4 2^{4(\delta+d)j}\rho_n^{-2}.$$

Démonstration. Let us prove the second point, the first one can be proved in a similar way. Let us recall that, by Proposition 2.8.1, for any $j\in\{j_1,\dots,j_2\}$ and any $k\in\{0,\dots,2^j-1\}$, we have

$$\widehat{\beta}_{j,k}-\beta_{j,k} \sim \mathcal{N}\left(0,\rho_n^{-2}\sigma_{j,k}^2\right), \tag{2.13}$$

where

$$\sigma_{j,k}^2 = \epsilon^2\sum_{v=1}^n \frac{1}{(1+\sigma_v^2)^{2\delta}} \sum_{\ell\in\mathbb{Z}} (2\pi\ell)^{2d} \frac{|\,\mathcal{F}\left(\psi_{j,k}\right)(\ell)|^2}{|\,\mathcal{F}(g_v)(\ell)|^2}. \tag{2.14}$$

Due to (2.8) and (2.11), for any $v \in \{1, \dots, n\}$, we have

$$\begin{aligned}\sup_{\ell \in \mathcal{C}_j} \left(\frac{(2\pi\ell)^{2d}}{|\mathcal{F}(g_v)(\ell)|^2} \right) &\leq C \sup_{\ell \in \mathcal{C}_j} \left((2\pi\ell)^{2d} \left(1 + \sigma_v^2 \ell^2\right)^{\delta} \right) \\ &\leq C(1+\sigma_v^2)^{\delta} \sup_{\ell \in \mathcal{C}_j} \left((2\pi\ell)^{2d} \left(1 + \ell^2\right)^{\delta} \right) \\ &\leq C(1+\sigma_v^2)^{\delta} 2^{2(\delta+d)j}. \end{aligned} \tag{2.15}$$

It follows from (2.15) and the Parseval identity that

$$\begin{aligned}\sigma_{j,k}^2 &\leq \epsilon^2 \sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{2\delta}} \sup_{\ell \in \mathcal{C}_j} \left(\frac{(2\pi\ell)^{2d}}{|\mathcal{F}(g_v)(\ell)|^2} \right) \sum_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})(\ell)|^2 \\ &\leq C\epsilon^2 2^{2(\delta+d)j} \left(\sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{\delta}} \right) \left(\sum_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})(\ell)|^2 \right) \\ &= C\epsilon^2 2^{2(\delta+d)j} \rho_n \int_0^1 |\psi_{j,k}(t)|^2 dt = C\epsilon^2 \rho_n 2^{2(\delta+d)j}. \end{aligned} \tag{2.16}$$

Putting (2.13), (2.14) and (2.16) together, we obtain

$$\mathbb{E}\left(|\widehat{\beta}_{j,k} - \beta_{j,k}|^4 \right) \leq C(\epsilon^2 2^{2(\delta+d)j} \rho_n \rho_n^{-2})^2 = C\epsilon^4 2^{4(\delta+d)j} \rho_n^{-2}.$$

Proposition 2.8.2 is proved. □

Proposition 2.8.3 (Concentration inequality)**.** *There exists a constant $\lambda > 0$ such that, for any $j \in \{j_1, \dots, j_2\}$, any $K \in \mathcal{A}_j$ and n large enough,*

$$\mathbb{P}\left(\left(\sum_{k \in B_{j,K}} |\widehat{\beta}_{j,k} - \beta_{j,k}|^2 \right)^{1/2} \geq \lambda 2^{(\delta+d)j} (\log \rho_n / \rho_n)^{1/2} \right) \leq \rho_n^{-2}.$$

Démonstration. We need the Tsirelson inequality stated in Lemma2.8.1 below.

Lemma 2.8.1 ([28])**.** *Let $(\vartheta_t)_{t \in D}$ be a centered Gaussian process. If*

$$\mathbb{E}\left(\sup_{t \in D} \vartheta_t \right) \leq N, \qquad \sup_{t \in D} \mathbb{V}(\vartheta_t) \leq V$$

then, for any $x > 0$, we have

$$\mathbb{P}\left(\sup_{t \in D} \vartheta_t \geq x + N \right) \leq \exp\left(-\frac{x^2}{2V} \right).$$

For the sake of notational clarity, let

$$V_{j,k} = \widehat{\beta}_{j,k} - \beta_{j,k}.$$

Recall that, by Proposition 2.8.1, we have $V_{j,k} \sim \mathcal{N}\left(0, \rho_n^{-2}\sigma_{j,k}^2\right)$, where $\sigma_{j,k}^2$ is given in (2.14). Let $\mathbb{B}(1)$ the unit 2-norm ball in $\mathbb{C}^{\mathrm{Card}(B_{j,K})}$, i.e. $\mathbb{B}(1) = \{a \in \mathbb{C}^{\mathrm{Card}(B_{j,K})};\ \sum_{k \in B_{j,K}} |a_k|^2 \leq$

1}. For any $a \in \mathbb{B}(1)$, let $Z(a)$ be the centered Gaussian process defined by

$$\begin{aligned} Z(a) &= \sum_{k \in B_{j,K}} a_k V_{j,k} \\ &= \epsilon \frac{1}{\rho_n} \sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{\delta}} \sum_{\ell \in \mathcal{C}_j} (2\pi i \ell)^d \frac{e_{\ell,v}}{\mathcal{F}(g_v)(\ell)} \sum_{k \in B_{j,K}} a_k \overline{\mathcal{F}(\psi_{j,k})}(\ell). \end{aligned}$$

By a simple Legendre-Fenchel conjugacy argument, we have

$$\sup_{a \in \mathbb{B}(1)} Z(a) = \left(\sum_{k \in B_{j,K}} |V_{j,k}|^2 \right)^{1/2} = \left(\sum_{k \in B_{j,K}} |\widehat{\beta}_{j,k} - \beta_{j,k}|^2 \right)^{1/2}.$$

Now, let us determine the values of N and V which appeared in the Tsirelson inequality.

Value of N. Using the Jensen inequality and (2.16), we obtain

$$\begin{aligned} \mathbb{E}\left(\sup_{a \in \mathbb{B}(1)} Z(a) \right) &= \mathbb{E}\left(\left(\sum_{k \in B_{j,K}} |V_{j,k}|^2 \right)^{1/2} \right) \leq \left(\sum_{k \in B_{j,K}} \mathbb{E}\left(|V_{j,k}|^2 \right) \right)^{1/2} \\ &\leq C \left(\rho_n^{-2} \sum_{k \in B_{j,K}} \sigma_{j,k}^2 \right)^{1/2} \leq C \left(\rho_n^{-2} \epsilon^2 \rho_n 2^{2(\delta+d)j} \operatorname{Card}(B_{j,K}) \right)^{1/2} \\ &\leq C \epsilon 2^{(\delta+d)j} \left(\log \rho_n / \rho_n \right)^{1/2}. \end{aligned}$$

Hence $N = C\epsilon 2^{(\delta+d)j} (\log \rho_n / \rho_n)^{1/2}$.

Value of V. Note that, for any $(\ell, \ell') \in \mathbb{Z}^2$ and any $(v, v') \in \{1, \dots, n\}^2$,

$$\mathbb{E}\left(e_{\ell,n} \overline{e}_{\ell',v'} \right) = \begin{cases} 1 & \text{if } \ell = \ell' \text{ and } v = v', \\ 0 & \text{otherwise.} \end{cases}$$

It then follows that

$$\begin{aligned}
\sup_{a\in\mathbb{B}(1)} \mathbb{V}(Z(a)) &= \sup_{a\in\mathbb{B}(1)} \mathbb{E}\left(\left|\sum_{k\in B_{j,K}} a_k V_{j,k}\right|^2\right) \\
&= \sup_{a\in\mathbb{B}(1)} \mathbb{E}\left(\sum_{k\in B_{j,K}}\sum_{k'\in B_{j,K}} a_k\overline{a}_{k'} V_{j,k}\overline{V}_{j,k'}\right) \\
&= \epsilon^2\rho_n^{-2} \sup_{a\in\mathbb{B}(1)} \sum_{k\in B_{j,K}}\sum_{k'\in B_{j,K}} a_k\overline{a}_{k'} \sum_{\ell\in\mathcal{C}_j}\sum_{\ell'\in\mathcal{C}_j}\sum_{v'=1}^{n}\sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{\delta}}\frac{1}{(1+\sigma_{v'}^2)^{\delta}}\times \\
&\quad \frac{(2\pi i\ell)^d}{\mathcal{F}(g_v)(\ell)}\overline{\mathcal{F}(\psi_{j,k})}(\ell)\frac{\overline{(2\pi i\ell')^d}}{\overline{\mathcal{F}(g_{v'})}(\ell')}\mathcal{F}(\psi_{j,k'})(\ell')\mathbb{E}\left(e_{\ell,v}\overline{e}_{\ell',v'}\right) \\
&= \epsilon^2\rho_n^{-2} \sup_{a\in\mathbb{B}(1)} \sum_{k\in B_{j,K}}\sum_{k'\in B_{j,K}} a_k\overline{a}_{k'} \sum_{\ell\in\mathcal{C}_j}\sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{2\delta}}\frac{(2\pi\ell)^{2d}}{|\mathcal{F}(g_v)(\ell)|^2}\overline{\mathcal{F}\left(\psi_{j,k}\right)}(\ell)\mathcal{F}(\psi_{j,k'})(\ell) \\
&= \epsilon^2\rho_n^{-2} \sup_{a\in\mathbb{B}(1)} \sum_{\ell\in\mathcal{C}_j}\sum_{v=1}^{n} \frac{1}{(1+\sigma_v^2)^{2\delta}}\frac{(2\pi\ell)^{2d}}{|\mathcal{F}(g_v)(\ell)|^2}\left|\sum_{k\in B_{j,K}} \overline{a}_k\,\mathcal{F}\left(\psi_{j,k}\right)(\ell)\right|^2. \qquad (2.17)
\end{aligned}$$

For any $a \in \mathbb{B}(1)$, the Parseval identity and the fact that $\{\psi_{j,k}\}_{k=0,\dots,2^j-1}$ are orthonormal yields

$$\begin{aligned}
\sum_{\ell\in\mathcal{C}_j}\left|\sum_{k\in B_{j,K}} \overline{a}_k\,\mathcal{F}(\psi_{j,k})(\ell)\right|^2 &= \sum_{\ell\in\mathcal{C}_j}\left|\mathcal{F}\left(\sum_{k\in B_{j,K}} \overline{a}_k\psi_{j,k}\right)(\ell)\right|^2 \\
&= \int_0^1\left|\sum_{k\in B_{j,K}} \overline{a}_k\psi_{j,k}(t)\right|^2 dt \\
&= \sum_{k\in B_{j,K}} |a_k|^2 \le 1. \qquad (2.18)
\end{aligned}$$

Piecing (2.15), (2.17) and (2.18) together, we get

$$\begin{aligned}
\sup_{a\in\mathbb{B}(1)} \mathbb{V}(Z(a)) &\le C\epsilon^2\rho_n^{-1}2^{2(\delta+d)j} \sup_{a\in\mathbb{B}(1)} \sum_{\ell\in\mathcal{C}_j}\left|\sum_{k\in B_{j,K}} \overline{a}_k\,\mathcal{F}\left(\psi_{j,k}\right)(\ell)\right|^2 \\
&\le C\epsilon^2\rho_n^{-1}2^{2(\delta+d)j}.
\end{aligned}$$

Hence it is sufficient to take $V = C\epsilon^2\rho_n^{-1}2^{2(\delta+d)j}$.

Taking λ large enough and $x = 2^{-1}\lambda\epsilon 2^{(\delta+d)j}(\log\rho_n/\rho_n)^{1/2}$, the Tsirelson inequality (see

Lemma 2.8.1) yields

$$\begin{aligned}
&\mathbb{P}\left(\left(\sum_{k\in B_{j,K}}|V_{j,k}|^2\right)^{1/2} \geq \lambda\epsilon 2^{(\delta+d)j}(\log\rho_n/\rho_n)^{1/2}\right)\\
&\leq \mathbb{P}\left(\left(\sum_{k\in B_{j,K}}|V_{j,k}|^2\right)^{1/2} \geq 2^{-1}\lambda\epsilon 2^{(\delta+d)j}(\log\rho_n/\rho_n)^{1/2}+N\right)\\
&= \mathbb{P}\left(\sup_{a\in\mathbb{B}(1)} Z(a)\geq x+N\right) \leq \exp\left(-x^2/(2V)\right) \leq \exp\left(-C\lambda^2\log\rho_n\right)\\
&\leq \rho_n^{-2}.
\end{aligned}$$

Proposition 2.8.3 is proved. □

■ 2.8.2 Proof of Theorem 2.5.1

Démonstration. Plugging Propositions 2.8.2 and 2.8.3 into [25, Theorem 3.1], we end the proof of Theorem 2.5.1. □

■ 2.8.3 Proof of Theorem 2.5.2

Démonstration. Let us now present a consequence of the Fano lemma.

Lemma 2.8.2. *Let $m\in\mathbb{N}^*$ and A be a σ-algebra on the space Ω. For any $i\in\{0,\dots,m\}$, let $A_i\in A$ such that, for any $(i,j)\in\{0,\dots,m\}^2$ with $i\neq j$,*

$$A_i\cap A_j=\varnothing.$$

Let $(\mathbb{P}_i)_{i\in\{0,\dots,m\}}$ be $m+1$ probability measures on (Ω,A). Then

$$\sup_{i\in\{0,\dots,m\}}\mathbb{P}_i\left(A_i^c\right)\geq\min\left(2^{-1},\exp(-3e^{-1})\sqrt{m}\exp(-\chi_m)\right),$$

where

$$\chi_m=\inf_{v\in\{0,\dots,m\}}\frac{1}{m}\sum_{\substack{k\in\{0,\dots,m\}\\k\neq v}}K(\mathbb{P}_k,\mathbb{P}_v),$$

A^c denotes the complement of A in Ω and K is the Kullbak-Leibler divergence defined by

$$K(\mathbb{P},\mathbb{Q})=\begin{cases}\int\ln\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right)d\mathbb{P} & \text{if } \mathbb{P}\ll\mathbb{Q},\\ \infty & \text{otherwise.}\end{cases}$$

The proof of Lemma 2.8.2 can be found in [35, Lemma 3.3]. For further details and applications of the Fano lemma, see [97].

In what follows, we distinguish ψ and ψ^{per} (see Section 2.3.1). For any integrable function h on $\mathbb{R}$, we set

$$\mathcal{F}_*(h)(\ell)=\int_{-\infty}^{\infty}h(t)e^{-2i\pi\ell t}dt,\qquad \ell\in\mathbb{Z}.$$

Consider the Besov balls $\mathbf{B}^s_{p,r}(M)$ (see (2.7)). Let j_0 be an integer suitably chosen below. For any $\varepsilon = (\varepsilon_k)_{k\in\{0,\dots,2^{j_0}-1\}} \in \{0,1\}^{2^{j_0}}$ and $d \in \mathbb{N}^*$, set

$$h_\varepsilon(t) = M_* 2^{-j_0(s+1/2)} \sum_{k=0}^{2^{j_0}-1} \varepsilon_k \frac{1}{(d-1)!} \sum_{l\in\mathbb{Z}} \int_{-\infty}^{t+l} (t+l-y)^{d-1} \psi_{j_0,k}(y)dy,\ t \in [0,1],$$

and, if $d = 0$, set $h_\varepsilon(t) = M_* 2^{-j_0(s+1/2)} \sum_{k=0}^{2^{j_0}-1} \varepsilon_k \psi^{\mathrm{per}}_{j_0,k}(t), t \in [0,1]$. Notice that, owing to (2.5), h_ε exists. Moreover, it is 1-periodic. Using the Cauchy formula for repeated integration, we have

$$h^{(d)}_\varepsilon(t) = M_* 2^{-j_0(s+1/2)} \sum_{k=0}^{2^{j_0}-1} \varepsilon_k \psi^{\mathrm{per}}_{j_0,k}(t), \qquad t \in [0,1].$$

So, for any $j \geq j_*$ and any $k \in \{0,\dots,2^j-1\}$, the (mother) wavelet coefficient of $h^{(d)}_\varepsilon$ is

$$\beta_{j,k} = \int_0^1 h^{(d)}_\varepsilon(t)\overline{\psi}^{\mathrm{per}}_{j,k}(t)dt = \begin{cases} M_*\varepsilon_k 2^{-j_0(s+1/2)}, & \text{if } j = j_0, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore $h^{(d)}_\varepsilon \in \mathbf{B}^s_{p,r}(M)$. The Varshamov-Gilbert theorem (see [97, Lemma 2.7]) asserts that there exist a set $E_{j_0} = \left\{\varepsilon^{(0)},\dots,\varepsilon^{(T_{j_0})}\right\}$ and two constants, $c \in]0,1[$ and $\alpha \in]0,1[$, such that, for any $u \in \{0,\dots,T_{j_0}\}$, $\varepsilon^{(u)} = (\varepsilon^{(u)}_k)_{k\in\{0,\dots,2^{j_0}-1\}} \in \{0,1\}^{2^{j_0}}$ and any $(u,v) \in \{0,\dots,T_{j_0}\}^2$ with $u < v$, the following inequalities hold :

$$\sum_{k=0}^{2^{j_0}-1} |\varepsilon^{(u)}_k - \varepsilon^{(v)}_k| \geq c2^{j_0}, \qquad T_{j_0} \geq e^{\alpha 2^{j_0}}.$$

Considering such a set E_{j_0}, for any $(u,v) \in \{0,\dots,T_{j_0}\}^2$ with $u \neq v$, we have by orthonormality of the collection $\{\psi_{j_0,k}\}_{k=0,\dots,2^{j_0}-1}$

$$\begin{aligned}
&\left(\int_0^1 \left(h^{(d)}_{\varepsilon^{(u)}}(t) - h^{(d)}_{\varepsilon^{(v)}}(t)\right)^2 dt\right)^{1/2} \\
&= M_* c 2^{-j_0(s+1/2)} \left(\sum_{k=0}^{2^{j_0}-1} \left|\varepsilon^{(u)}_k - \varepsilon^{(v)}_k\right|^2\right)^{1/2} \\
&= M_* c 2^{-j_0(s+1/2)} \left(\sum_{k=0}^{2^{j_0}-1} \left|\varepsilon^{(u)}_k - \varepsilon^{(v)}_k\right|\right)^{1/2} \\
&\geq 2\delta_{j_0},
\end{aligned}$$

where

$$\delta_{j_0} = M_* c^{1/2} 2^{j_0/2} 2^{-j_0(s+1/2)} = M_* c^{1/2} 2^{-j_0 s}.$$

Using the Markov inequality, for any estimator $\widetilde{f^{(d)}}$ of $f^{(d)}$, we have

$$\delta^{-2}_{j_0} \sup_{f^{(d)}\in\mathbf{B}^s_{p,r}(M)} \mathbb{E}\left(\int_0^1 \left(\widetilde{f^{(d)}}(t) - f^{(d)}(t)\right)^2 dt\right) \geq \sup_{u\in\{0,\dots,T_{j_0}\}} \mathbb{P}_{h_{\varepsilon^{(u)}}}(A^c_u) = p,$$

where

$$A_u = \left\{ \left(\int_0^1 \left(\widetilde{f^{(d)}}(t) - h^{(d)}_{\varepsilon^{(u)}}(t) \right)^2 dt \right)^{1/2} < \delta_{j_0} \right\}$$

and $\mathbb{P}_f$ is the distribution of model (2.1). Notice that, for any $(u,v) \in \{0,\dots,T_{j_0}\}^2$ with $u \neq v$, $A_u \cap A_v = \varnothing$. Lemma 2.8.2 applied to the probability measures $\left(\mathbb{P}_{h_{\varepsilon^{(u)}}}\right)_{u\in\{0,\dots,T_{j_0}\}}$ gives

$$p \geq \min\left(2^{-1}, \exp(-3e^{-1})\sqrt{T_{j_0}}\exp(-\chi_{T_{j_0}})\right), \tag{2.19}$$

where

$$\chi_{T_{j_0}} = \inf_{v\in\{0,\dots,T_{j_0}\}} \frac{1}{T_{j_0}} \sum_{\substack{u\in\{0,\dots,T_{j_0}\}\\ u\neq v}} K\left(\mathbb{P}_{h_{\varepsilon^{(u)}}}, \mathbb{P}_{h_{\varepsilon^{(v)}}}\right).$$

Let's now bound $\chi_{T_{j_0}}$. For any functions f_1 and f_2 in $\mathbb{L}^2_{\mathrm{per}}([0,1])$, we have

$$\begin{aligned} K\left(\mathbb{P}_{f_1}, \mathbb{P}_{f_2}\right) &= \frac{1}{2\epsilon^2}\sum_{v=1}^n \int_0^1 \left((f_1 \star g_v)(t) - (f_2 \star g_v)(t)\right)^2 dt \\ &= \frac{1}{2\epsilon^2}\sum_{v=1}^n \int_0^1 \left(((f_1 - f_2) \star g_v)(t)\right)^2 dt. \end{aligned}$$

The Parseval identity yields

$$\begin{aligned} K\left(\mathbb{P}_{f_1}, \mathbb{P}_{f_2}\right) &= \frac{1}{2\epsilon^2}\sum_{v=1}^n \sum_{\ell\in\mathbb{Z}} |\,\mathcal{F}((f_1 - f_2) \star g_v)(\ell)|^2 \\ &= \frac{1}{2\epsilon^2}\sum_{v=1}^n \sum_{\ell\in\mathbb{Z}} |\mathcal{F}(f_1 - f_2)(\ell)|^2\, |\,\mathcal{F}(g_v)(\ell)|^2. \end{aligned}$$

So, for any $(u,v) \in \{0,\dots,T_{j_0}\}^2$ with $u \neq v$, we have

$$K\left(\mathbb{P}_{h_{\varepsilon^{(u)}}}, \mathbb{P}_{h_{\varepsilon^{(v)}}}\right) = \frac{1}{2\epsilon^2}\sum_{v=1}^n \sum_{\ell\in\mathbb{Z}} |\mathcal{F}\left(h_{\varepsilon^{(u)}} - h_{\varepsilon^{(v)}}\right)(\ell)|^2\, |\,\mathcal{F}(g_v)(\ell)|^2. \tag{2.20}$$

By definition, for any $(u,v) \in \{0,\dots,T_{j_0}\}^2$ with $u \neq v$ and $\ell \in \mathbb{Z}$, we have

$$\begin{aligned} &\mathcal{F}\left(h_{\varepsilon^{(u)}} - h_{\varepsilon^{(v)}}\right)(\ell) \\ &= M_* 2^{-j_0(s+1/2)} \sum_{k=0}^{2^{j_0}-1} \left(\varepsilon_k^{(u)} - \varepsilon_k^{(v)}\right) \times \\ &\frac{1}{(d-1)!}\sum_{l\in\mathbb{Z}} \mathcal{F}\left(\int_{-\infty}^{\cdot+l} (\cdot + l - y)^{d-1}\psi_{j_0,k}(y)dy\right)(\ell) \\ &= M_* 2^{-j_0(s+1/2)} \sum_{k=0}^{2^{j_0}-1} \left(\varepsilon_k^{(u)} - \varepsilon_k^{(v)}\right) \times \\ &\frac{1}{(d-1)!}\mathcal{F}_*\left(\int_{\infty}^{\cdot} (\cdot - y)^{d-1}\psi_{j_0,k}(y)dy\right)(\ell). \end{aligned} \tag{2.21}$$

Let, for any $k \in \{0, \dots, 2^{j_0} - 1\}$,

$$\theta_k(t) = \frac{1}{(d-1)!} \int_{-\infty}^{t} (t-y)^{d-1} \psi_{j_0,k}(y) dy, \qquad t \in [0,1].$$

Then, for any $u \in \{0, \dots, d\}$, thanks to (2.4), $\lim_{|t|\to\infty} \theta_k^{(u)}(t) = 0$. Consequently, for any $\ell \in \mathbb{Z}$,

$$(2\pi i \ell)^d \mathcal{F}_*(\theta_k)(\ell) = \mathcal{F}_*\left(\theta_k^{(d)}\right)(\ell) = \mathcal{F}(\psi_{j_0,k}^{\mathrm{per}})(\ell).$$

So, for any $\ell \in \mathcal{C}_{j_0}$ (excluding 0), (2.21) implies that

$$\begin{aligned} &\mathcal{F}(h_{\varepsilon^{(u)}} - h_{\varepsilon^{(v)}})(\ell) \\ &= M_* 2^{-j_0(s+1/2)} \sum_{k=0}^{2^{j_0}-1} \left(\varepsilon_k^{(u)} - \varepsilon_k^{(v)}\right) \frac{1}{(2\pi i \ell)^d} \mathcal{F}\left(\psi_{j_0,k}^{\mathrm{per}}\right)(\ell), \end{aligned} \tag{2.22}$$

which entails in particular that $\operatorname{supp}\left(\mathcal{F}(h_{\varepsilon^{(u)}} - h_{\varepsilon^{(v)}})\right) = \mathcal{C}_{j_0}$. This in conjunction with equalities (2.20) and (2.22) imply that

$$\begin{aligned} &K\left(\mathbb{P}_{h_{\varepsilon^{(u)}}}, \mathbb{P}_{h_{\varepsilon^{(v)}}}\right) \\ &= \frac{M_*^2}{2\epsilon^2} 2^{-2j_0(s+1/2)} \sum_{v=1}^{n} \sum_{\ell \in \mathcal{C}_{j_0}} \left| \sum_{k=0}^{2^{j_0}-1} \left(\varepsilon_k^{(u)} - \varepsilon_k^{(v)}\right) \mathcal{F}\left(\psi_{j_0,k}^{\mathrm{per}}\right)(\ell) \right|^2 \frac{1}{(2\pi\ell)^{2d}} |\mathcal{F}(g_v)(\ell)|^2. \end{aligned} \tag{2.23}$$

By assumptions (2.8) and (2.11), for any $v \in \{1, \dots, n\}$,

$$\begin{aligned} \sup_{\ell \in \mathcal{C}_{j_0}} \left(\frac{1}{(2\pi\ell)^{2d}} |\mathcal{F}(g_v)(\ell)|^2 \right) &\le C \sup_{\ell \in \mathcal{C}_{j_0}} \left(\frac{1}{(2\pi\ell)^{2d}} \left(1 + \sigma_v^2 \ell^2\right)^{-\delta} \right) \\ &\le C \sigma_v^{-2\delta} \sup_{\ell \in \mathcal{C}_{j_0}} \left(\ell^{-2(\delta+d)}\right) \le C \sigma_v^{-2\delta} 2^{-2j_0(\delta+d)}. \end{aligned} \tag{2.24}$$

Moreover, the Parseval identity and orthonormality of the family $\left\{\psi_{j_0,k}^{\mathrm{per}}\right\}_{k=0,\dots,2^{j_0}-1}$ imply that

$$\begin{aligned} &\sum_{\ell \in \mathcal{C}_{j_0}} \left| \sum_{k=0}^{2^{j_0}-1} \left(\varepsilon_k^{(u)} - \varepsilon_k^{(v)}\right) \mathcal{F}\left(\psi_{j_0,k}^{\mathrm{per}}\right)(\ell) \right|^2 \\ &= \int_0^1 \left| \sum_{k=0}^{2^{j_0}-1} \left(\varepsilon_k^{(u)} - \varepsilon_k^{(v)}\right) \psi_{j_0,k}^{\mathrm{per}}(t) \right|^2 dt = \sum_{k=0}^{2^{j_0}-1} \left(\varepsilon_k^{(u)} - \varepsilon_k^{(v)}\right)^2 \le 2^{j_0}. \end{aligned} \tag{2.25}$$

It follows from (2.23), (2.24) and (2.25) that

$$K\left(\mathbb{P}_{h_{\varepsilon^{(u)}}}, \mathbb{P}_{h_{\varepsilon^{(v)}}}\right) \le C 2^{-2j_0(s+1/2)} 2^{-2j_0(\delta+d)} 2^{j_0} \sum_{v=1}^{n} \sigma_v^{-2\delta} = C \rho_n^* 2^{-2j_0(s+1/2+\delta+d)} 2^{j_0}.$$

Hence

$$\begin{aligned} \chi_{T_{j_0}} &= \inf_{v \in \{0,\dots,T_{j_0}\}} \frac{1}{T_{j_0}} \sum_{\substack{u \in \{0,\dots,T_{j_0}\} \\ u \neq v}} K\left(\mathbb{P}_{h_{\varepsilon^{(u)}}}, \mathbb{P}_{h_{\varepsilon^{(v)}}}\right) \\ &\le C \rho_n^* 2^{-2j_0(s+1/2+\delta+d)} 2^{j_0}. \end{aligned} \tag{2.26}$$

Putting (2.19) and (2.26) together and choosing j_0 such that

$$2^{-j_0(s+1/2+\delta+d)} = c_0(\rho_n^*)^{-1/2},$$

where c_0 denotes a well chosen constant, for any estimator $\widetilde{f^{(d)}}$ of $f^{(d)}$, we have

$$\begin{aligned}\delta_{j_0}^{-2} \sup_{f^{(d)} \in \mathbf{B}^s_{p,r}(M)} \mathbb{E}\left(\int_0^1 \left(\widetilde{f^{(d)}}(t) - f^{(d)}(t)\right)^2 dt\right) &\geq c \exp\left((\alpha/2)2^{j_0} - Cc_0^2 2^{j_0}\right) \\ &\geq c,\end{aligned}$$

where

$$\delta_{j_0} = c2^{-j_0 s} = c(\rho_n^*)^{-s/(2s+2\delta+2d+1)}.$$

This complete the proof of Theorem 2.5.2. □

Chapitre 3

Adaptive parameter selection for block wavelet-thresholding deconvolution

In this chapter, we propose a data-driven block thresholding procedure for wavelet-based non-blind deconvolution. The approach consists in appropriately writing the problem in the wavelet domain and then selecting both the block size and threshold parameter at each resolution level by minimizing Stein's unbiased risk estimate. The resulting algorithm is simple to implement and fast. Numerical illustrations are provided to assess the performance of the estimator.

Sommaire

3.1 Problem statement and motivating example

Let $A : H \to K$ be a known linear operator between the Hilbert spaces H and K endowed with the corresponding inner products and associated norms. Assume that we observe the data Y according to

$$Y = Af + \epsilon W, \tag{3.1}$$

where $f \in H$ is the unknown function we wish to recover, $\epsilon > 0$ is the noise level and W is a zero-mean Gaussian white noise. We aim to estimate a function f from the observations Y.

In this chapter, we consider the problem of deconvolution, but other examples can also be considered as well, e.g., density with error in variable, tomography, etc., if the operator possesses K the so-called Wavelet-Vaguelette Decomposition (WVD) of [41]. The degradation process can be modelled as a blurring operator A which is characterised

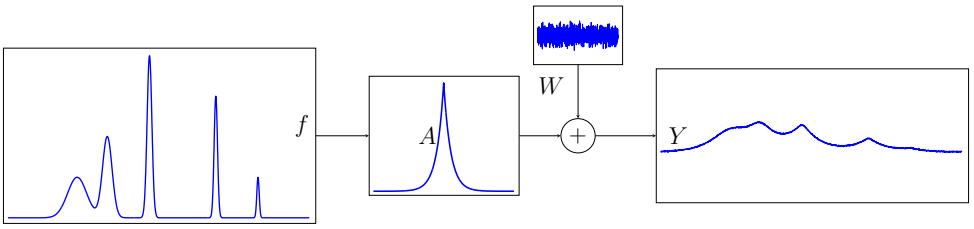


FIGURE 3.1 – Illustration of the deconvolution problem.

by the kernel g (i.e. also often referred to as a point spread function (PSF) or impulse response)

$$Af : \Omega \to \mathbb{R}$$
$$t \mapsto Af(t) = g \circledast f(t) = \int_{\Omega} g(s) f(t-s) ds, \quad t \in \Omega$$

where $\Omega = [0,1]$ and both f and g belong to $H = \mathbb{L}^2_{\text{per}}(\Omega) = \{h;\ h \text{ is 1-periodic on } \Omega \text{ and } \int_{\Omega} |h(t)|^2 dt < \infty\}$. Deconvolution is a challenging ill-posed inverse problem.

3.2 Overview of previous work

There is already an extensive literature on deconvolution for the case where A is known (see, e.g., [41, 14, 15, 16, 21, 53]).

In this work, we consider adaptive block estimators based on the Wavelet-Vaguelette Decomposition (WVD) of [41]. Minimaxity of James-Stein block thresholding over Besov balls with fixed block size has been established for deconvolution in [26]. The performance of Block thresholding estimators strongly depends on the block size L and threshold level λ. In the nonparametric regression setting, in order to choose these two key parameters in an optimal way, [11] proposed an adaptive James-Stein block thresholding estimator whose parameters minimize the Stein's unbiased risk estimate (SURE) and established its minimax rates of convergence under the mean squared error over Besov balls. [53] described a SURE-based level-dependent term-by-term thresholding estimator for a class of linear inverse problems possessing a WVD. Recently, the SURE has proved to be a powerful tool for signal/image restoration [65, 82, 98]. For image denoising problems, a Stein risk estimator have been proposed in [83] to both adapt the block-sparsity structure and the threshold.

In this work, we extend several of the previous works to solve linear inverse problems, mainly deconvolution here, with wavelet block thresholding estimators. Our estimator is constructed using a periodized Meyer wavelet basis and two popular block thresholding estimators ; namely Block-Soft and Block James-Stein (JS) (defined in the sequel). The corresponding optimal block size and threshold parameter are determined by minimizing the SURE after rewriting the model appropriately in the wavelet domain.

The chapter is organized as follows. Section 3.3 briefly reviews wavelets and Besov balls and describes the block thresholding-based deconvolution estimator. Section 3.4 discusses numerical results, before conclusions are drawn in Section 3.5.

3.3 Nonlinear estimation via Block thresholding

3.3.1 Smoothness of the kernel g

In this study, we focus on the standard context of ordinary smoothness of g. That is, that there exist three constants, $c_g > 0$, $C_g > 0$ and $\delta > 1$ such that, for any $\ell \in \mathbb{Z}$,

$$c_g \frac{1}{(1+\ell^2)^{\delta/2}} \leq |g_\ell| \leq C_g \frac{1}{(1+\ell^2)^{\delta/2}}. \tag{3.2}$$

This assumption controls the decay of the Fourier coefficients of g. It is a standard hypothesis usually adopted in the field of nonparametric estimation for deconvolution problems, see e.g. [9, 41, 81].

3.3.2 Wavelet deconvolution in the Fourier domain

We set, $f_\ell = \langle f, e_\ell \rangle, W_\ell = \langle W, e_\ell \rangle$, where W_ℓ are i.i.d. standard (complex-valued) Gaussian random variables. Then, re-writing the model (3.1) in the Fourier domain

$$y_\ell = f_\ell g_\ell + \epsilon W_\ell, \quad \ell \in \mathbb{Z}. \tag{3.3}$$

Moreover, we can compute the wavelet transform via the Parseval formula

$$\beta_{j,k} = \sum_{\ell \in \mathbb{Z}} f_\ell (\psi_{j,k})_\ell,$$

where $(\cdot_{j,k})_\ell = \langle \cdot_{j,k}, e_\ell \rangle$, Then, combining the latter with (3.3), we get

$$\sum_{\ell \in \mathcal{C}_j} \frac{y_\ell}{g_\ell} (\psi_{j,k})_\ell = \sum_{\ell \in \mathcal{C}_j} f_\ell (\psi_{j,k})_\ell + \epsilon \sum_{\ell \in \mathcal{C}_j} \frac{e_\ell}{g_\ell} (\psi_{j,k})_\ell \ .$$

Thus, we can perform wavelet transform and deconvolution simultaneously and (3.3) reduces to a general sequence model with correleted noise

$$z_{j,k} = \beta_{j,k} + \epsilon W_{j,k}, \tag{3.4}$$

where $W_{j,k} = \sum_{\ell \in \mathcal{C}_j} \frac{e_\ell}{g_\ell} (\psi_{j,k})_\ell \sim \mathcal{N}\left(0, \epsilon^2 \sum_{\ell \in \mathcal{C}_j} \frac{|(\psi_{j,k})_\ell|^2}{|g_\ell|^2}\right)$.

Let us denote by σ_j the standard deviations of the noise process $W_{j,k}$ and let $\lambda_j = \lambda \sigma_j$ be a sequence of thresholds to be applied to the wavelet coefficients at level j. In view of (2.8), we have at level j that $\sigma_j \sim \epsilon 2^{j\delta}$. We define the estimator of the wavelet coefficients of f as

$$\widehat{\beta}^*_{j,k} = \eta^*(z_{j,k}, \lambda_j), \tag{3.5}$$

where η^* is either Block-Soft or Block-JS thresholding defined in (3.7).

3.3.3 Block deconvolution estimator

Let the observed signal Y be the vector of n observations $Y_1, \ldots, Y_n$. Let $j_1 = \lfloor \ln_2(\ln n) \rfloor$ be the coarsest resolution level, and $j_2 = \lfloor (1/(2\delta+1)) \ln_2(n/\ln n) \rfloor$, where, for any $a \in \mathbb{R}$, $\lfloor a \rfloor$ denotes the integer part of a. For any $j \in \{j_1, \ldots, j_2\}$, let $L \geq 1$ be the block size.

Let $A_j = \{1, \ldots, \lfloor 2^j L^{-1} \rfloor\}$ be the set indexing the blocks at resolution j. For each j, let $\{b_{j,K}\}_{K \in A_j}$ be a uniform and disjoint open covering of $\{0, \ldots, 2^j - 1\}$, i.e. $\bigcup_{K \in A_j} b_{j,K} =$

$\{0,\dots,2^j-1\}$ and for any $(K,K')\in A_j^2$ with $K\neq K'$, $b_{j,K}\cap b_{j,K'}=\varnothing$ and $\mathrm{Card}(b_{j,K})=L$, where $b_{j,K}=\{k\in\{0,\dots,2^j-1\};\ (K-1)L\leq k\leq KL-1\}$ is the Kth block (see Chapter 2 FIGURE 2.3).

We define our block estimator $\widehat{f}_{\lambda,L}$ of f by

$$\widehat{f}_{\lambda,L}(t)=\sum_{k=0}^{2^{j_1}-1}\widehat{\alpha}_{j_1,k}\phi_{j_1,k}(t)+\sum_{j=j_1}^{j_2}\sum_{K\in A_j}\sum_{k\in b_{j,K}}\widehat{\beta}^*_{j,k}\psi_{j,k}(t), \tag{3.6}$$

$t\in[0,1]$, where for any resolution level j and position $k\in b_{j,K}$ within the Kth block, the wavelet coefficients of f are estimated via (3.5) with one of the rules

$$\begin{aligned}\eta^{\mathrm{Soft}}(z_{j,k},\lambda)&=\left(1-\frac{\lambda}{\|z_{b_{j,K}}\|_2}\right)_+ z_{j,k},\\ \eta^{\mathrm{JS}}(z_{j,k},\lambda)&=\left(1-\frac{\lambda^2}{\|z_{b_{j,K}}\|_2^2}\right)_+ z_{j,k},\end{aligned} \tag{3.7}$$

with, $\|z_{b_{j,K}}\|_2^2=\sum_{k\in b_{j,K}}|z_{j,k}|^2$, and $(a)_+=\max(a,0)$. $\widehat{\alpha}_{j_1,k}=\sum_{\ell\in\mathcal{D}_{j_1}}\frac{y_\ell}{g_\ell}(\phi_{j,k})_\ell$ is the empirical approximation coefficient.

Note that (3.6) is simple to implement and fast (thanks to the Fast Fourier Transform algorithm) which allows us to perform the selection procedure in a reasonable time.

Block thresholding estimators depend on both the block size L and threshold parameter λ which mainly determine the performance of the resulting estimator. So it is important to select these parameters in an optimal way.

■ 3.3.4 Unbiased risk estimation for automatic parameter selection

Since the discrete wavelet transform is orthogonal, the risk of an estimator $\widehat{f}$ of f will be the same as that of its wavelets coefficients. The risk $R(\widehat{f},f)=\mathbb{E}(\|\widehat{f}-f\|^2)$ will measure the estimation error in the Mean Square Error sense. Since the blocks are also non-overlapping, an estimator of the risk is derived by summing the risk estimates over blocks and resolution levels j.

More precisely, since each of the thresholding rules in (3.7) is weakly differentiable with an essentially bounded gradient, Stein lemma allows to get an estimator of the risk in each block, which solely depends on the observation. Denoting $z_{b_{j,K}}=(z_{j,k})_{k\in b_{j,K}}$, the SURE on each block $b_{j,K}$ is given by

$$\begin{aligned}J^{\mathrm{Soft}}(z_{b_{j,K}},\lambda_j,\sigma_j)&=L\sigma_j^2+(L\|z_{b_{j,K}}\|^2-2L\sigma_j^2)\mathbb{1}_{\{\|z_{b_{j,K}}\|<\lambda_j\}}\\&\quad+(L\lambda_j^2-2\sigma_j^2(L-1)\frac{\lambda_j}{\|z_{b_{j,K}}\|})\mathbb{1}_{\{\|z_{b_{j,K}}\|\geqslant\lambda_j\}}\\ J^{\mathrm{JS}}(z_{b_{j,K}},\lambda_j,\sigma_j)&=L\sigma_j^2+(L\|z_{b_{j,K}}\|^2-2L\sigma_j^2)\mathbb{1}_{\{\|z_{b_{j,K}}\|<\lambda_j\}}\\&\quad+\frac{L\lambda_j^2-2\sigma_j^2(L-2)}{\|z_{b_{j,K}}\|^2/\lambda_j^2}\mathbb{1}_{\{\|b_{j,K}\|\geqslant\lambda_j\}},\end{aligned}$$

where $\mathbb{1}_{\{.\}}$ denotes the indicator function of the event in its argument. Thus, the overall SURE of $\widehat{\beta}^*_{j,k}$ is computed by simple summation

$$J^*(\{z_{j,k}\}_{j,k},\lambda)=\sum_{j=j_1}^{j_2}\sum_{K\in A_j}\sum_{k\in b_{j,K}}J^*(z_{b_{j,K}},\lambda_j,\sigma_j)\ . \tag{3.8}$$

This is indeed an unbiased estimator of the risk. Consequently, our approach can be used as a principled way to objectively choose the optimal parameters λ and L that minimize (3.8). The deconvolution algorithm can be summarized as follows :

Algorithm 1 deconvolution algorithm

Parameters : The observed blurred and noisy signal Y, the PSF g.

Initialization :

— Block size : $L \geq 1$.

— Coarsest decomposition scale $j_1 = \lfloor \ln_2(\ln n) \rfloor$.

Step one : Fourier and wavelet domain.

— Apply an inverse filtering on the data using FFT (i.e. $1/g_\ell$).

— Perform one Forward wavelet transform : $z_{j,k} = \sum_{\ell \in \mathcal{C}_j} \frac{y_\ell}{g_\ell} (\psi_{j,k})_\ell$.

— Compute and store the variance σ_j^2.

Step two : Compute the risk.

— For each possible dyadic block $b_{j,K}$ compute $J^*(z_{b_{j,K}}, \lambda_j, \sigma_j)$ and sum over the scale to get $J^*(\{z_{b_{j,K}}\}, \lambda_j, \sigma_j)$.

Step three : Optimal parameters.

— Get the optimal parameters λ and L by a grid search.

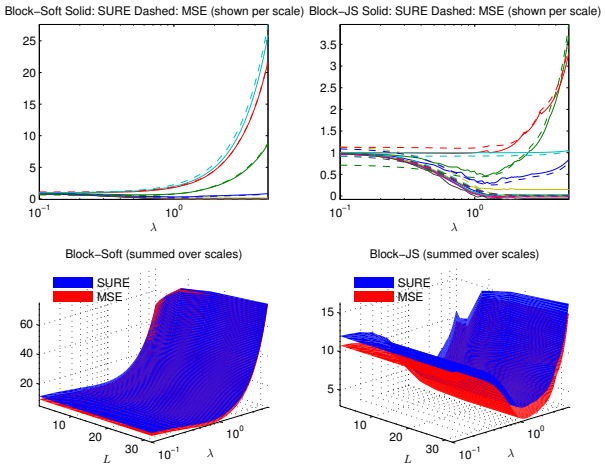
Output : Get deconvolved signal $\widehat{f}_{\lambda,L}$ at the optimal λ and L.

3.4 Simulation experiments

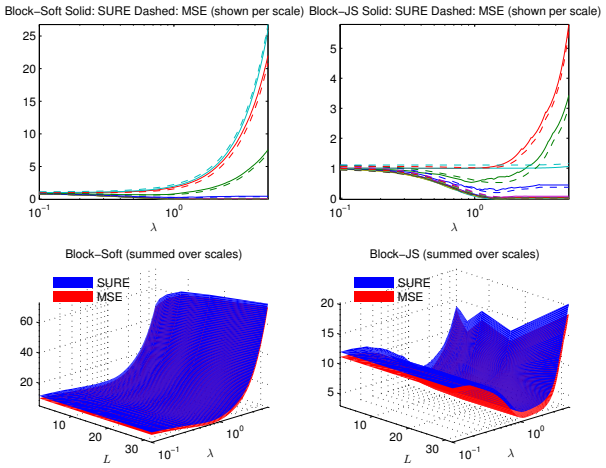
Three test functions, representing different degrees of irregularity, were used. The signals have been blurred with a Laplacian PSF and were corrupted by a zero-mean white Gaussian noise. FIGURE 3.2 depicts the SURE and true risk curves as a function of the block size and the threshold λ, for each thresholding rule, plotted at each scale with the optimal block size (1st row) and summed over scales for each block size used (2nd row). It can be observed that the SURE gives a very reliable estimate for the risk, and in turn, also a high-quality estimate of the optimal λ. The restoration quality can be assessed both visually and quantitatively in FIGURE 3.3, which shows the restored signals at the optimally chosen λ and L, with an input BSNR = 30 dB.

3.5 Conclusion

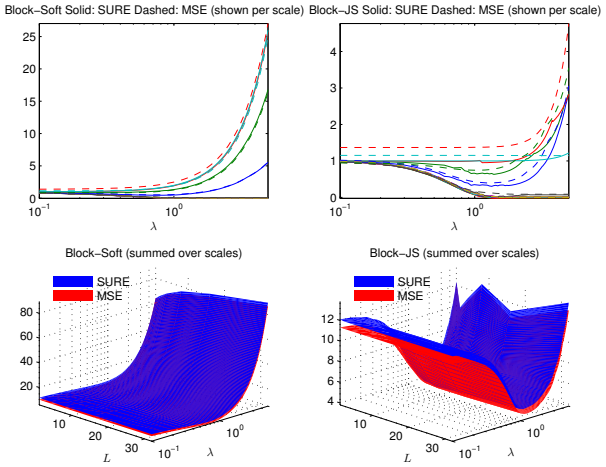
In this chapter, a data-driven block wavelet-based deconvolution estimator was presented. Its usefulness has been illustrated on automatic selection of the both the block size L and threshold level λ for signal deconvolution. Although we focused on one-dimensional convolution, the approach can handle other operators A and easily extended to a multi-dimensional setting (e.g. images). We reported some simulation experiments to support our findings. Our efforts are now directed towards establishing the theoretical minimax performance of this estimator.



(a) Piece-Polynomial

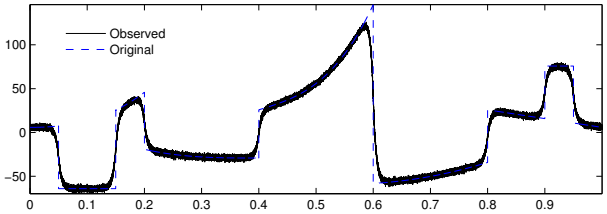


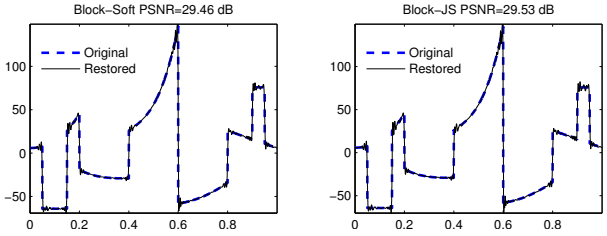
(b) Piece-Regular



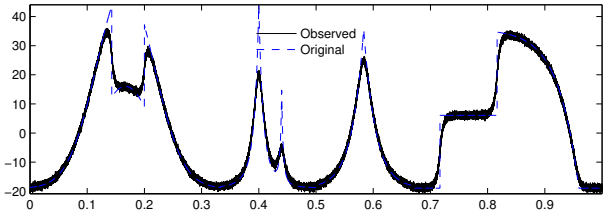
(c) Bumps

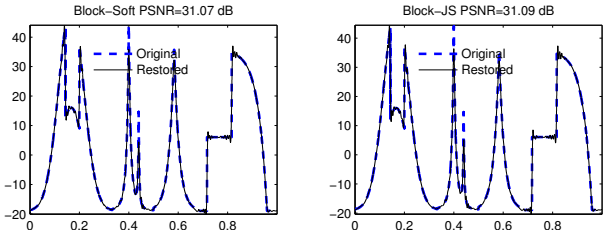
FIGURE 3.2 – SURE and risk as a function of the threshold for Soft (left) and JS (right) block thresholding both at each scale with the optimal block size (top) and summed over scales for each block size used (bottom).



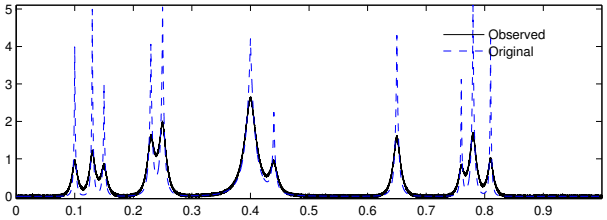


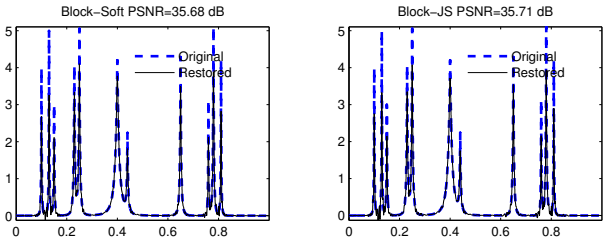
(a) Piece-Polynomial





(b) Piece-Regular





(c) Bumps

FIGURE 3.3 – Top panel : Original and observed signal (top) ; Bottom panels : Block-Soft (left) and Block-JS (right) estimates at the optimal λ and L.

Deuxième partie

Estimation adaptative de densite dans différents modèles

Chapitre 4

On adaptive wavelet estimation of a class of weighted densities

We investigate the estimation of a weighted density taking the form $g = w(F)f$, where f denotes an unknown density, F the associated distribution function and w is a known non-negative weight. Such a class encompasses many examples, including those arising in order statistics or when g is related to the maximum or the minimum of N (random or fixed) independent and identically distributed (*i.i.d.*) random variables. We here construct a new adaptive non-parametric estimator for g based on a plug-in approach and the wavelets methodology. For a wide class of models, we show that it attains fast rates of convergence under the $\mathbb{L}_p$ risk with $p \geq 2$ over Besov balls. Our estimator is also simple to implement and fast. We also report an extensive simulation study to support our findings.

Sommaire

4.1 Problem statement and motivation

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a real random variable with unknown density f and Y be a random variable having the unknown weighted density

$$g(x) = w(F(x))f(x), \qquad x \in \mathbb{R}, \tag{4.1}$$

where w denotes a known non-negative weight and F denotes the distribution function of f. The goal we pursue here is to estimate g from a n-*i.i.d.* sample $X_1, \ldots, X_n$ of X.

Such an estimation problem arises in many situations, typically when g is related to the maximum [1] of N *i.i.d.* random variables, where N is a discrete random number in $\mathbb{N}^*$ which

1. Since $\min(X_1, \ldots, X_N) = -\max(-X_1, \ldots, -X_N)$ the results can be easily reformulated for the sample minimum.

is independent of the X_i's. Application fields cover hydrology, meteorology, reliability, investment, management science, insurance business, etc.. For example, when the X_i are non-negative, the random variable $Y = \max(X_1, \ldots, X_N)$ (or $Y = \min(X_1, \ldots, X_N)$) arises naturally in reliability theory as the lifetime of a parallel (series) system with a random number N of identical components with lifetimes $X_1, \ldots, X_N$.

To make things clearer to the reader, we next give some illustrative examples.

Example 4.1.1 (Order statistics). *Let $X_1, \ldots, X_m$ be m i.i.d. random variables with absolutely continuous distribution function F and probability density function (pdf) f. Let $X_{(1)} \leq \ldots \leq X_{(m)}$ denote the corresponding order statistics. Then, the pdf $g_{X(j)}$ of the j-th order statistic is*

$$g_{X(j)}(x) = \frac{m!}{(j-1)!(m-j)!}(F(x))^{j-1}(1-F(x))^{m-j}f(x), \qquad x \in \mathbb{R}.$$

Thus, $X_{(m)}$, for example, is the random variable representing the largest observation of a sample of n and corresponds to the sample maximum and the density $g_{X(m)}$ of $X_{(m)} = \max(X_1, \ldots, X_m)$ is given by

$$g_{X(m)}(x) = m\,(F(x))^{m-1}\,f(x), \qquad x \in \mathbb{R}.$$

The aim is to estimate $g_{X(j)}$ from a n-i.i.d. sample $X_1, \ldots, X_n$ of X.

Example 4.1.2 (Maximum of a random number N of *i.i.d.* random variables). *Let X be a random variable with density f, $\{X_i\}_{i\in\mathbb{N}^*}$ be a sequence of i.i.d. random variables with density f and N be a discrete random variable taking values in $\mathbb{N}^*$ with a known probability mass function. Then the density of $Y = \max(X_1, \ldots, X_N)$ is*

$$g(x) = w(F(x))f(x), \qquad x \in \mathbb{R}, \tag{4.2}$$

where

$$w(u) = \sum_{k=1}^{\infty} k u^{k-1}\mathbb{P}(N = k), \qquad u \in [0,1].$$

The goal is again to estimate g from a n-i.i.d. sample $X_1, \ldots, X_n$ of X.

Example 4.1.3 (Pile-up model). *Let us now present the "pile-up model". Let $\{Y_i\}_{i\in\mathbb{N}^*}$ be a sequence of i.i.d. random variables with density g, N be a discrete random variable in $\mathbb{N}^*$ as in the previous example, and let $X = \min(Y_1, \ldots, Y_N)$ with density f. Then the density of Y_1 is*

$$g(x) = w(F(x))f(x), \qquad x \in \mathbb{R},$$

where

$$w(u) = \frac{1}{M'(M^{-1}(1-u))}, \qquad u \in [0,1],$$

$M(u) = \mathbb{E}(u^N)$ and $M'(u) = \mathbb{E}(Nu^{N-1})$. We are seeking an estimate of g from a n-i.i.d. sample $X_1, \ldots, X_n$ of X.

4.2 State-of-the-art

Some distributional properties of the maximum and minimum of random variables have been extensively studied in the literature (see, e.g., [86], [93] and [94]). In addition, the literature on order statistics contains a huge work about the maximum. In the context of extreme value theory, various statistical properties and (real data) applications can be found in [1], [64] and the references therein.

The estimation of the density function of the maximum of two independent random variables has been considered by [19] via kernel methods. The Pile-up model has been considered by [31] via model selection methods.

In this chapter, we develop a new *non-linear* adaptive estimator for g in model (4.1) based on a plug-in method, wavelets and the block thresholding rule introduced by [8]. Wavelet-based thresholding estimators are attractive for non-parametric function estimation because of their virtues from the viewpoints of spatial adaptivity, computational efficiency and asymptotic optimality properties. In the case of simple density estimation, wavelet thresholding is probably one of the most attractive nonlinear methods. We refer to e.g., [5], [51] and [99] for a detailed discussion of the performances of wavelet estimators and some of their advantages over traditional methods such as kernel-based or projection estimators.

We here explore the theoretical performance of our estimator under the $\mathbb{L}_p$ risk with $p \geq 2$ over a very rich class of function spaces, namely Besov spaces. Sharp rates of convergence are obtained. Application of our estimator to Example 4.1.2 above is described in detail. Finally, extensive simulation experiments are carried out to illustrate the practical performance of our estimator. In particular, the numerical tests indicate that our block thresholding estimator, which is simple to implement and fast, compares very favorably to standard kernel-based methods.

4.3 Wavelet estimators

First of all, we briefly recall some key facts on wavelets and Besov spaces that will be essential to us in the sequel. Then we develop our nonlinear adaptive wavelet block thresholding estimator.

4.3.1 Wavelets and Besov balls

Let $b > 0$, $p > 0$ and $\mathbb{L}_p([-b,b]) = \left\{h : [-b,b] \to \mathbb{R};\ \|h\|_p^p = \int_{-b}^{b} |h(x)|^p dx < +\infty\right\}$.

For the purposes of this chapter, we use compactly supported wavelet bases on $[-b,b]$. More precisely, we consider the Daubechies family db_{2T} with the scaling and wavelet functions ϕ and ψ, where $R \geq 2$ is a fixed integer, see e.g., [66]. Define the scaled and translated version of the ϕ and ψ

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \qquad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k).$$

Then there exists an integer τ and a set of consecutive integers Λ_j such that $\mathrm{Card}(\Lambda_j) = C2^j$ for a $C > 0$ and, for any integer $\ell \geq \tau$, the collection

$$\mathcal{B} = \{\phi_{\ell,k},\ k \in \Lambda_\ell;\ \psi_{j,k};\ j \in \mathbb{N} - \{0,\ldots,\ell-1\},\ k \in \Lambda_j\},$$

is an orthonormal basis of $\mathbb{L}_2([-b,b])$.

Consequently, for any integer $\ell \geq \tau$, any $h \in \mathbb{L}_2([-b,b])$ can be expanded on $\mathcal{B}$ as

$$h(x) = \sum_{k\in\Lambda_\ell} \alpha_{\ell,k}\phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty}\sum_{k\in\Lambda_j} \beta_{j,k}\psi_{j,k}(x),$$

where

$$\alpha_{\ell,k} = \int_{-b}^{b} h(x)\phi_{\ell,k}(x)dx, \qquad \beta_{j,k} = \int_{-b}^{b} h(x)\psi_{j,k}(x)dx. \tag{4.3}$$

As is traditional in the wavelet estimation literature, we will investigate the performance of our estimator by assuming that the unknown density f belongs to a Besov ball $B^s_{q,r}(M)$. The Besov norm for a function can be related to a sequence space norm on its wavelet coefficients. More precisely, for $M > 0$, $s \in (0,T)$, $q \geq 1$ and $r \geq 1$, a function $h \in \mathbb{L}_p([-b,b])$ belongs to $B^s_{q,r}(M)$ if and only if there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (4.3) satisfy

$$\left(\sum_{k\in\Lambda_\tau} |\alpha_{\tau,k}|^q\right)^{1/q} + \left(\sum_{j=\tau}^{\infty}\left(2^{j(s+1/2-1/q)}\left(\sum_{k\in\Lambda_j}|\beta_{j,k}|^q\right)^{1/q}\right)^r\right)^{1/r} \leq M^*,$$

with the usual modifications if $q = \infty$ or $r = \infty$.

In this expression, s is a smoothness parameter and q and r are norm parameters. They include many traditional smoothness spaces such as Hölder and Sobolev spaces. A comprehensive account on Besov spaces can be found in e.g., [36, 69, 51].

■ 4.3.2 Plug-in block wavelet estimator

Let us consider the general statistical framework described in Section 4.1 with a *n-i.i.d.* sample $X_1,\ldots,X_n$ of X. First of all, we investigate the estimation of f via the so-called wavelet block hard thresholding estimator. We suppose that $\text{supp}(f) \subseteq [-b,b]$ with $b > 0$.

Let $p \geq 2$, and j_1 and j_2 be the integers corresponding to the finest and coarsest scales defined as

$$j_1 = \lfloor (p/2)\log_2(\log n)\rfloor, \qquad j_2 = \lfloor \log_2(n/\log n)\rfloor,$$

where $\lfloor a\rfloor$ denotes the whole number part of $a \in \mathbb{R}^+$. For any integer $j \in \{j_1,\ldots,j_2\}$, let A_j and $U_{j,K}$ be given such that $(U_{j,K})_{K\in A_j}$ form a partition of Λ_j and all the $U_{j,K}$'s are defined with $L = \lfloor (\log n)^{p/2}\rfloor$ consecutive integers k. In a nutshell, at each scale j, each $U_{j,K}$ is the set containing position indices of the wavelet coefficients inside block $K \in A_j$. Note that the number of elements in A_j is proportional to $\lfloor 2^j/L\rfloor$.

We define the wavelet block hard thresholding estimator of f by

$$\hat{f}(x) = \sum_{k\in\Lambda_{j_1}} \hat{\alpha}_{j_1,k}\phi_{j_1,k}(x) + \sum_{j=j_1}^{j_2}\sum_{K\in A_j}\sum_{k\in U_{j,K}} \hat{\beta}_{j,k}\mathbb{1}_{\left\{\left(\sum_{k\in U_{j,K}}|\hat{\beta}_{j,k}|^p/L\right)^{1/p} \geq \kappa n^{-1/2}\right\}}\psi_{j,k}(x),\ x \in [-b,b], \tag{4.4}$$

where $\mathbb{1}_{\{.\}}$ denotes the indicator function, and

$$\hat{\alpha}_{j_1,k} = \frac{1}{n}\sum_{i=1}^{n}\phi_{j_1,k}(X_i), \qquad \hat{\beta}_{j,k} = \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(X_i)$$

and $\kappa > 0$ is a threshold parameter to be discussed later.

The estimator $\hat{f}$ was initially developed by [8] for the regression model under the $\mathbb{L}_2$-risk with equispaced deterministic samples. The $\mathbb{L}_p$ risk version was studied in [84] for the standard density estimation problem and by [22] for the biased density estimation problem.

The idea underlying $\hat{f}$ (4.4) is to operate a group/block selection : it keeps intact the large groups of unknown wavelet coefficients of f (4.3) and removes the others. Wavelet block thresholding is one of the most attractive non-linear thresholding methods, since it is both numerically straightforward to implement and asymptotically optimal for a large variety of Sobolev or Besov classes. Detailed references on the subject for various models include, but are not limited to, [8, 10], [63, 62], [84] and [21, 22].

The performance of Block thresholding estimators depends on the threshold level κ. In the non-parametric regression setting, in order to choose this key parameter, [11] proposed an adaptive James-Stein block thresholding estimator whose parameters (including the threshold) minimize the Stein's unbiased risk estimate (SURE) and established its minimax rates of convergence under the mean squared error over Besov balls. Other selection strategies have been developed in the literature (see e.g. [71] which considered wavelet estimators based on cross-validation to choose the thresholding parameter in practice). In this work, we focus on the universal threshold proposed by [40]. The reason for this choice is twofold. First, it is the one consistent with the theoretical convergence rates established in Section 4.4. Secondly, it allows to remain fair when comparing to the other methods of the literature tested in Section 4.5.

Finally, plugging (4.4) into (4.1) leads to the following estimator of g :

$$\hat{g}(x) = w(\hat{F}(x))\hat{f}(x), \qquad x \in [-b, b], \tag{4.5}$$

where

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}. \tag{4.6}$$

The rest of the chapter explores the theoretical and practical performances of $\hat{g}$.

4.4 Estimator convergence rates

In this section, we discuss the asymptotic properties of the proposed estimator. Rates of $\mathbb{L}_p$ convergence are investigated under the following assumptions :

(A.1) Compact support : $\text{supp}(f) \subseteq [-b, b]$ with $b > 0$.

(A.2) Uniform boundedness : there exists a constant $C_1 > 0$ such that

$$\sup_{x \in [-b,b]} f(x) \leq C_1 \ . \tag{4.7}$$

(A.3) Lipschitz continuity of w : there exist two constants $\theta \in (0, 1]$ and $C_2 > 0$ such that

$$|w(u) - w(v)| \leq C_2|u - v|^{\theta}, \qquad \text{for all } (u, v) \in [0, 1]^2, \tag{4.8}$$

Assumption A.1 is a usual one in the wavelet density estimation framework (see e.g. [40]). Extension to non-compactly supported densities might be possible and ideas from [56] might be inspiring, although these authors considered a model different from

ours, and their results were valid only for the case where f is in the Hölder class. Such an extension is however beyond the scope of this study and we leave it for a future work.

Theorem 4.4.1 studies the $\mathbb{L}_p$ risk of $\hat{g}$ (4.5) over Besov balls and Assumptions A.1-A.3 on f and w.

Theorem 4.4.1. *Consider the general statistical framework described in Section 4.1 (the estimation of g (4.1) is of interest). Suppose that Assumptions A.1-A.3 hold. Let $p \geq 2$ and $\hat{g}$ be given by (4.5). Then, for any $f \in B^s_{q,r}(M)$, $q \geq 1$, $r \geq 1$ and $s \in (1/q, T)$, there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\|\hat{g} - g\|_p^p\right) \leq C \max(\varphi_{n,p}, n^{-\theta p/2}),$$

where

$$\varphi_{n,p} = \begin{cases} n^{-sp/(2s+1)}, & \text{if } q \geq p, \\ \left(\frac{\log n}{n}\right)^{sp/(2s+1)}, & \text{if } \{p > q,\ qs > (p-q)/2\}, \\ \left(\frac{\log n}{n}\right)^{(s-1/q+1/p)p/(2(s-1/q)+1)}, & \text{if } \{qs < \frac{(p-q)}{2}\} \text{ or } \{qs = \frac{(p-q)}{2},\ p \leq q/r\}, \\ \left(\frac{\log n}{n}\right)^{(s-1/q+1/p)p/(2(s-1/q)+1)} (\log n)^{p-q/r}, & \text{if } \{qs = (p-q)/2,\ p > q/r\}. \end{cases} \tag{4.9}$$

Note that the rate $\varphi_{n,p}$ in (4.9) is the near optimal (or optimal in some regimes) one in the minimax sense for f. See, e.g., [40] and [51].

In the case $\theta \geq \max\left(2s/(2s+1), 2(s-1/q+1/p)p/(2(s-1/q)+1)\right)$ (as $\theta = 1$), the near optimality of the estimator of f is transferred to that of g through the plug-in principle. In the other case, the regularity of w deteriorated the rate of convergence; it becomes $n^{-\theta p/2}$.

The proof of Theorem 4.4.1 uses a suitable decomposition of the $\mathbb{L}_p$ risk and capitalizes on results on the performances of $\hat{f}$ (4.4) and $\hat{F}$ (4.6) established in [22].

Proof of Theorem 4.4.1. Observe that

$$\hat{g}(x) - g(x) = w(\hat{F}(x))(\hat{f}(x) - f(x)) + f(x)(w(\hat{F}(x)) - w(F(x))).$$

By Assumption A.3 implying $\sup_{x\in[0,1]} w(x) \leq C$, together with Assumption A.2, we have

$$\begin{aligned} |\hat{g}(x) - g(x)| &\leq C(|\hat{f}(x) - f(x)| + |w(\hat{F}(x)) - w(F(x))|) \\ &\leq C(|\hat{f}(x) - f(x)| + |\hat{F}(x) - F(x)|^\theta). \end{aligned}$$

By the Jensen inequality, we have

$$\mathbb{E}(\|\hat{g} - g\|_p^p) \leq C\left(\mathbb{E}(\|\hat{f} - f\|_p^p) + \mathbb{E}(\|\hat{F} - F\|_{\theta p}^{\theta p})\right).$$

It follows from [22, Theorem 4.1 with $w(x) = 1 = \mu$] that

$$\mathbb{E}(\|\hat{f} - f\|_p^p) \leq C\varphi_{n,p},$$

where $\varphi_{n,p}$ is given by (4.9).

By Assumption A.1, we have

$$\mathbb{E}(\|\hat{F}-F\|_{\theta p}^{\theta p}) \leq C \sup_{x\in[-b,b]} \mathbb{E}\left(|\hat{F}(x)-F(x)|^{\theta p}\right).$$

In order bound the last term, note that

$$\hat{F}(x)-F(x) = \frac{1}{n}\sum_{i=1}^{n} U_i(x),$$

with $U_i(x) = \mathbb{1}_{\{X_i\leq x\}} - F(x)$, $U_1(x),\ldots,U_n(x)$ are *i.i.d.* with $\mathbb{E}(U_1(x))=0$, $|U_1(x)|\leq 2$ and $\mathbb{E}((U_1(x))^2)\leq 4$.

The case $\theta < 2/p$ *:* it follows from the Hölder inequality that

$$\mathbb{E}\left(|\hat{F}(x)-F(x)|^{\theta p}\right) \leq \left(\mathbb{E}\left((\hat{F}(x)-F(x))^2\right)\right)^{\theta p/2} = \left(\frac{1}{n}\mathbb{V}(U_1(x))\right)^{\theta p/2} \leq Cn^{-\theta p/2}.$$

The case $\theta \geq 2/p$ *:* the Rosenthal inequality (see [88]) applied with the exponent $\theta p \geq 2$ yields

$$\mathbb{E}\left(|\hat{F}(x)-F(x)|^{\theta p}\right) = n^{-\theta p}\mathbb{E}\left(\left|\sum_{i=1}^{n} U_i(x)\right|^{\theta p}\right) \leq Cn^{-\theta p}n^{\theta p/2} \leq Cn^{-\theta p/2}.$$

Combining the inequalities above, we obtain the desired result, i.e.,

$$\mathbb{E}(\|\hat{g}-g\|_p^p) \leq C(\varphi_{n,p}+n^{-\theta p/2}) \leq C\max(\varphi_{n,p}, n^{-\theta p/2}).$$

□

■ 4.4.1 An illustrative application

Let's recall Example 4.1.2, where $\{X_i\}_{i\in\mathbb{N}^*}$ is a sequence of *i.i.d.* random variables with pdf f and N be a discrete random variable of values in $\mathbb{N}^*$ independent of this sequence. The density of $Y=\max(X_1,\ldots,X_N)$ is given by (4.2).

Suppose that Assumptions A.1-A.2 hold. Thus, several examples for the distribution of N can be considered.

(a) Degenerate distribution. $\mathbb{P}(N=m)=1$. Then

$$w(u) = mu^{m-1}, \qquad u\in[0,1], \tag{4.10}$$

(b) Geometric distribution. $N\sim G(\eta)$ ($\mathbb{P}(N=k)=\eta(1-\eta)^{k-1}$, $k\in\mathbb{N}^*$). Then

$$w(u) = \frac{\eta}{(1-u(1-\eta))^2}, \qquad u\in[0,1], \tag{4.11}$$

(c) Poisson plus 1 distribution. $N=P+1$ with $P\sim\mathcal{P}(\lambda)$ ($\mathbb{P}(N=k)=e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!}$, $k\in\mathbb{N}^*$). Then

$$w(u) = e^{-\lambda}e^{\lambda u}(1+\lambda u), \qquad u\in[0,1].$$

Remark 4.4.1. *In examples (a)–(c) above it is clear that Assumption A.3 is satisfied with $\theta = 1$; more precisely, in example (a), we have $C_2 = m(m-1)$; in example (b), we have $C_2 = 2(1-\eta)/\eta^2$; in example (c), we have $C_2 = \lambda(2+\lambda)$.*

In this context, Theorem 4.4.1 can be applied. Let $p \geq 2$ and $\hat{g}$ be the estimator given in (4.5). Then, for any $f \in B^s_{q,r}(M)$, $q \geq 1$, $r \geq 1$ and $s \in (1/q, T)$ there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\|\hat{g} - g\|_p^p\right) \leq C\varphi_{n,p},$$

where $\varphi_{n,p}$ is given by (4.9).

Remark 4.4.2. *Taking $m = 2$, the obtained rate is similar to the one attained by the kernel estimators developed by [19]; the only difference is the extra-logarithmic term $(\log n)^{2s/(2s+1)}$. However, unlike kernel estimators [19], our procedure $\hat{g}$ is adaptive and our rate of convergence holds for a wider class of functions f including Hölder class, Sobolev class, etc..*

4.5 Simulation results

We now illustrate these theoretical results by a simulation study within the context described in Section 4.4.1. That is, we consider the problem of estimating the density g of the maximum of a random number N of *i.i.d.* random variables. From a reliability study standpoint, this problem corresponds to a parallel system with N identical components. Thereby, we have considered two numerical examples. They complement the asymptotic results of 4.4.1.

Computational aspects. In the sequel, we will refer to our adaptive wavelet estimator (4.5) simply as Block. We have compared its performance to alternatives from the literature on several densities. We have considered the uniform distribution, as well as a family of normal mixture densities ("SeparatedBimodal", "Kurtotic" and "StronglySkewed", initially introduced in [68]) representing different degrees of smoothness (see Figure 4.1). We assumed that the density function f of the X_i's has a compact support included in $[-b, b]$. We have used the formulae given by [68] to simulate such densities so that

$$\min_l(\mu_l - 3\sigma_l) = -3, \quad \max_l(\mu_l + 3\sigma_l) = 3,$$

where $l = 1, \ldots, d$ with d the number of densities in the mixture (see, [68, Section 4, Table 1], for the values of the parameters). Thereby, it is very unlikely to have values outside the interval $[-4, 4]$ and we loose little by assuming compact support . The same kind of assumption was made in the context of wavelet density estimation by [100]. In order to simplify the presentation of the results, one can simply rescale the data such that they fall into $[-b, b]$ (which covered the full range of all observed data). Thus, the density was evaluated at $T = 2^J$ equispaced points $t_i = 2ib/T$, $i = -T/2, \ldots, T/2 - 1$ between $-b$ and b , where J is the index of the highest resolution level and T is the number of discretization points. The primary level $j_1 = 3$, $T = 512$ and the Symmlet wavelet with 6 vanishing moments were used throughout all experiments. All simulations have been implemented under Matlab.

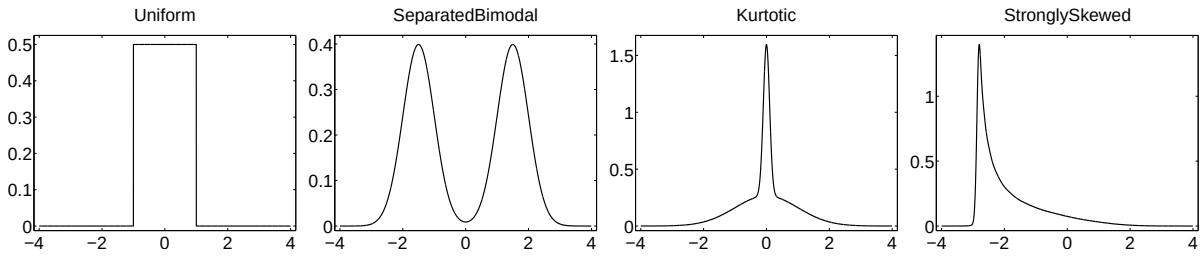
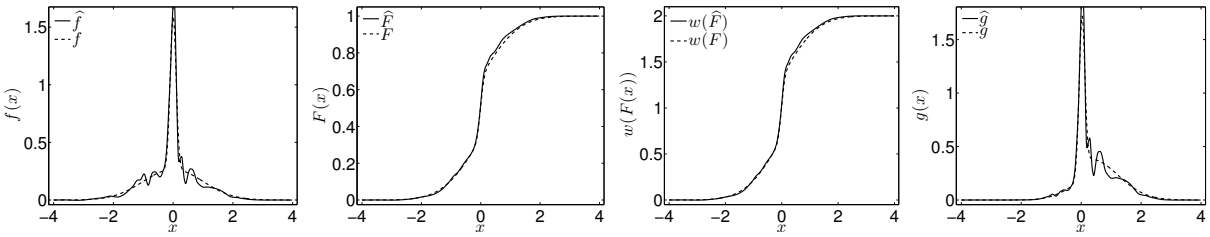


FIGURE 4.1 – Test densities.



(a) (b) (c) (d)

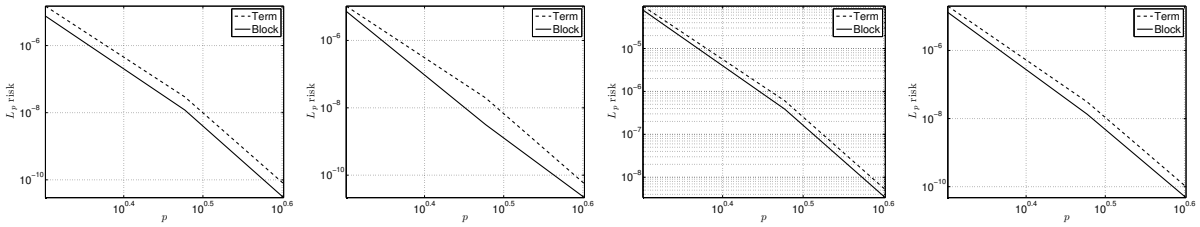
FIGURE 4.2 – Typical reconstructions from a single simulation with $n = 1000$ for the Kurtotic density. The dashed line depicts the original density and the solid one depicts its wavelet block estimate. (a) : $\hat{f}(x)$. (b) : $\hat{F}(x)$. (c) : $w(\hat{F}(x))$. (d) : $\hat{g}(x) = w(\hat{F}(x))\hat{f}(x)$.

Results and discussion. In order to illustrate Theorem 4.4.1, we study the influence of p on the numerical performances of the Block and the term-by-term ($L = 1$) thresholding estimator. Let us first consider a parallel system with $m = 2$ identical independent components. Then, the corresponding weighted function is (4.10) and the goal is to estimate g in (4.2) from $X_1, \ldots, X_n$ sample simulated from one of the test densities. A typical example of estimation for the Kurtotic density (for $p = 2$), with $n = 1000$ is given in Figure 4.2. The mean $\mathbb{L}_p$ risk of $\hat{g}$ i.e., $R_p(\hat{g}, g) = (1/T)\sum_{i=-T/2}^{T/2-1} |\hat{g}(t_i) - g(t_i)|^p$, is obtained with 50 samples for $n = 1000$, and it is plotted as a function of p in Figure 4.3. As predicted by Theorem 4.4.1, the larger p, the smaller $\mathbb{L}_p$ risk of $\hat{g}$. We can see that our estimation procedure provides better results than the term-by-term thresholding ($L = 1$) in all cases. In particular, the risk improvement achieved by the block estimators upon the term-by-term estimator is significant for the non-smooth Uniform density. This is in agreement with the predictions of our theoretical findings.

To conclude this first example, we illustrate from a single simulation, the fact that the parameters dictated by the theory yield the expected performance. We display in Figure 4.4 the empirical $\mathbb{L}_2$ risk as a function of the threshold level κ, where the vertical dashed line represents the universal threshold. One can see that the minimum of the $\mathbb{L}_2$ risk is close to the universal threshold for all test densities, thus supporting the choice dictated by our theoretical procedure, although derived in an asymptotic setting.

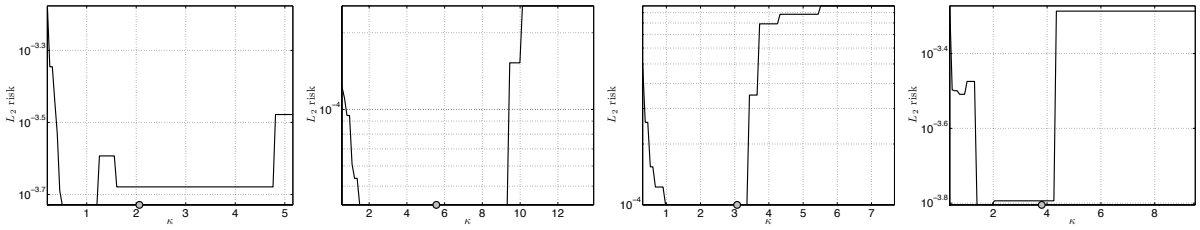
In our second example, the adaptive estimator described in Section 4.4.1 is tested when N follows a Geometric distribution, so that the weight function is that given by (4.11). This example is devoted to a simulation study comparing the performance of the block hard thresholding estimator with that of the traditional kernel defined as follows

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{4.12}$$



(a) Uniform (b) SeparatedBimodal (c) Kurtotic (d) StronglySkewed

FIGURE 4.3 – The influence of p in the numerical values of the $\mathbb{L}_p$ risk (in a log-log scale) of Block (solid) and term-by-term (dashed) thresholding ($L = 1$).



(a) Uniform (b) SeparatedBimodal (c) Kurtotic (d) StronglySkewed

FIGURE 4.4 – $\mathbb{L}_2$ risk as a function of the threshold level κ (in a semi-log scale), the gray circle represents the universal threshold.

where the positive kernel K satisfies $\int K(x)dx = 1$ and the smoothing parameter h is known as the bandwidth.

Many procedures of bandwidth selection for kernel density estimation have been developed in the literature (see, e.g., [96]). We use least-squares cross-validation (LSCV) ([89], [6]) where the bandwidth is defined as

$$h_{\text{LSCV}} = \underset{h}{\text{argmin}} \int_{-b}^{b} \hat{f}_h(x)^2 dx - 2n^{-1} \sum_{i=1}^{n} \hat{f}_{-i}(X_i),$$

and $\hat{f}_{-i}$ is the *leave-one out* kernel estimator constructed from the data without the observation X_i. It is motivated by the fact that for independent data

$$\text{LSCV}(h) = \int_{-b}^{b} \hat{f}_h(x)^2 dx - 2n^{-1} \sum_{i=1}^{n} \hat{f}_{-i}(X_i)$$

is an unbiased estimator of $\text{MISE}(h) = \int_{-b}^{b} f^2(x)dx$. One frequently used cross-validation (CV) procedure is the K-fold CV (as described e.g. in [52, Section 7.10]) in which the data set $X_1, \ldots, X_n$ is randomly partitioned into K approximately equal-sized and non-overlapping subsets $S1, \ldots, S_K$. To obtain the bandwidths h_{LSCV}, we have performed a 10-fold CV, using a Gaussian kernel, with a simple "rule-of-thumb" pilot bandwidth h_{ROT}. Figure 4.5(b) contains a plot of the LSCV function versus the kernel bandwidth h and Figure 4.5(c) the estimated MISE as a function of h. For each density, it is clear from this figure (Figure 4.5(b)), that the value of h_{LSCV} is the unambiguous minimizer of $\text{LSCV}(h)$. We see that h_{LSCV} provides a decent approximation, close to h_{MISE} for all test densities. For the StronglySkewed density, the bandwidth which minimizes $\text{MISE}(h)$ in this case is $h_{\text{MISE}} = 0.039$ and $h_{\text{LSCV}} = 0.044$. In this case, for the Uniform density, $h_{\text{MISE}} = h_{\text{LSCV}} = 0.051$.

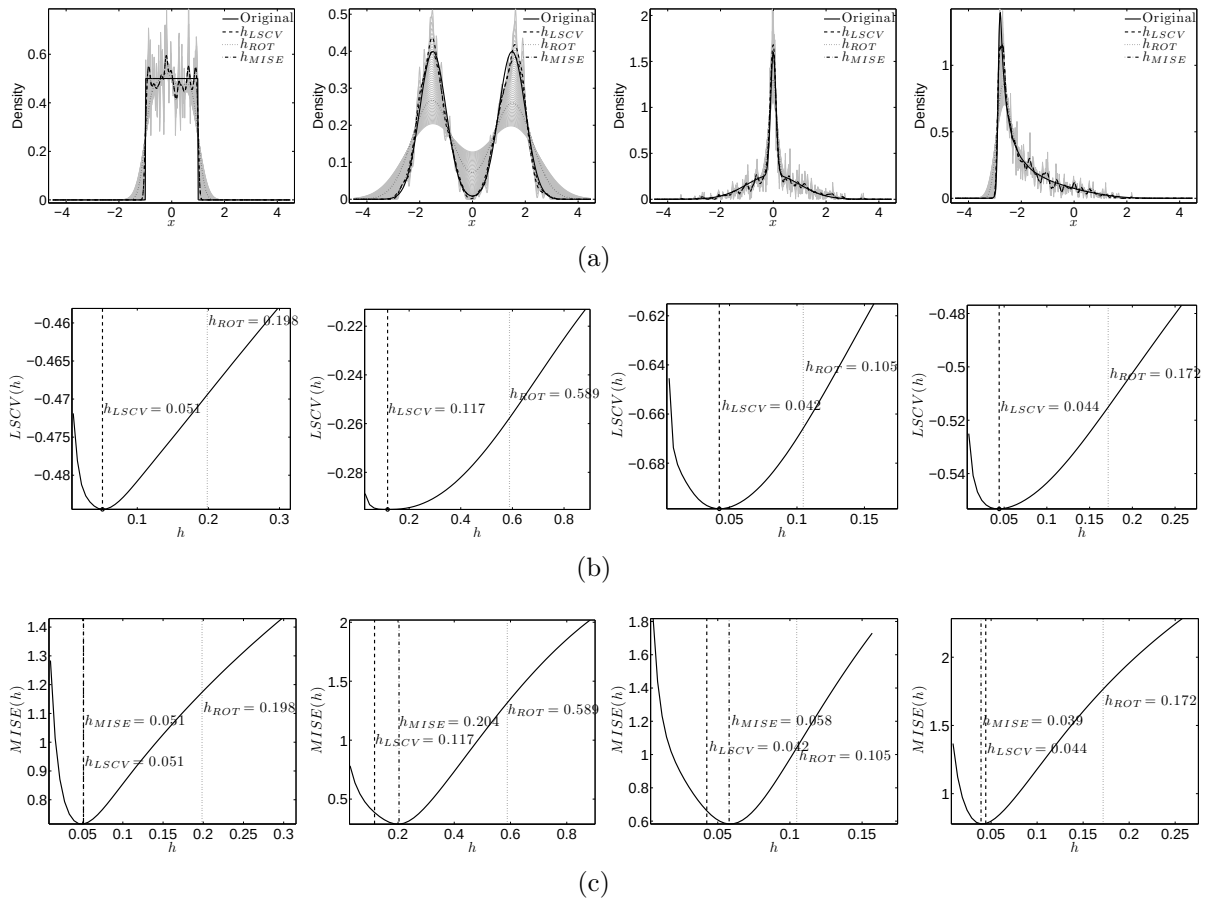
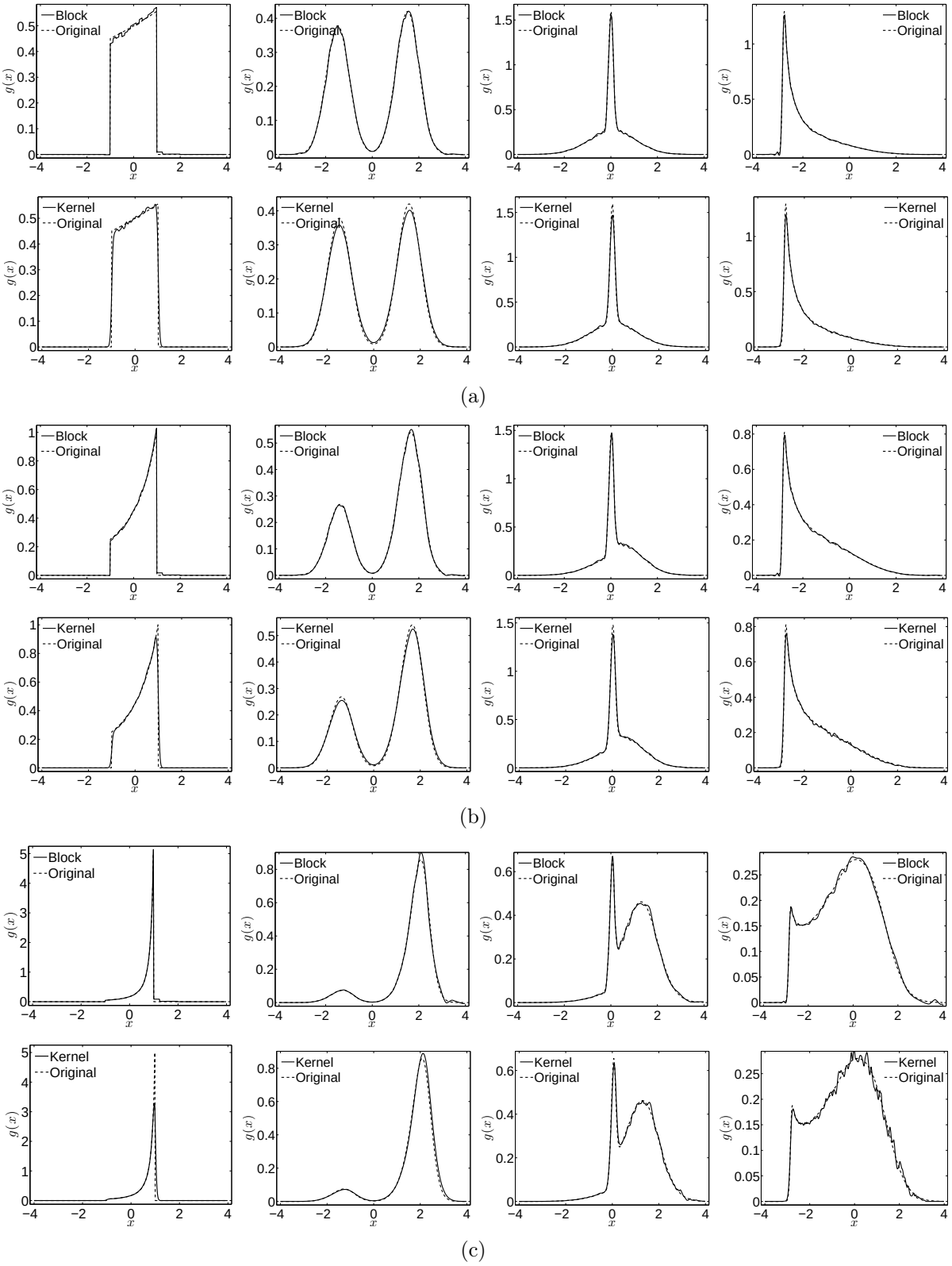


FIGURE 4.5 – (a) : Density estimates. (b) : Graph of the the LSCV function versus the kernel bandwidth h for each of the tested densities, the vertical dashed lines represent the value of h that minimizes LSCV(h). (c) : The solid line depicts the estimated MISE as a function of h, the vertical dashed-dotted lines represent the true MISE-minimizing bandwidth h_{MISE} and the vertical dotted lines represent the pilot bandwidth h_{ROT}.

We then compared the performance of the Block estimator $\hat{g}$ with that of the plug-in kernel estimator, say $\hat{g}_{\mathrm{LSCV}}$, given by $\hat{g}_{\mathrm{LSCV}} = w(\hat{F}(x))\hat{f}_{\mathrm{LSCV}}(x)$, where $\hat{F}$ is defined by (4.6) and $\hat{f}_{\mathrm{LSCV}}$ is given by (4.12) with h_{LSCV}. Figure 4.6 shows the results of $\hat{g}$ and $\hat{g}_{\mathrm{LSCV}}$ for $N \sim G(\eta)$, with $\eta = 0.9$, $\eta = 0.5$ and $\eta = 0.1$ respectively. Table 4.1 presents the MISE for samples sizes $n = 200, 1000, 2000$ and 5000. For virtually all cases, the Block estimator consistently showed lower $\mathbb{L}_2$ risk than $\hat{g}_{\mathrm{LSCV}}$, with the exception of the (very smooth) SeparatedBimodal density for which the kernel estimator performs slightly better. This comes at no surprise given that this density is very smooth. Additionally, small discrepancies in the estimate of f may lead to substantial discrepancies for the estimate of g at the locations overweighted by $w(\hat{F}(\cdot))$. It turns out that this is the case for the Geometric distribution where the weights evolve in $O(1/\eta)$ at high values of x, and thus the discrepancies in $\hat{g}$ increase as η gets smaller. However, the kernel estimator $\hat{g}_{\mathrm{LSCV}}$ seems to be more concerned (see, Figure 4.6(c)), confirming that Block generally provides a better estimate of f. Furthermore, as expected, for both methods, and in all cases, the MISE is decreasing as the sample size increases. Without any prior smoothness knowledge on the unknown density, the Block estimator provides very competitive results in comparison to $\hat{g}_{\mathrm{LSCV}}$.



(a)

(b)

(c)

FIGURE 4.6 – Original densities (dashed), Block thresholding estimator $\hat{g}$ (solid) (1st row), kernel estimator $\hat{g}_{\mathrm{LSCV}}$ (solid) (2nd row) from 50 replications of $n = 1000$ samples $X_1, \ldots, X_n$. From left to right Uniform, SeparatedBimodal, Kurtotic and StronglySkewed. $N \sim G(\eta)$, with (a) $\eta = 0.9$, (b) $\eta = 0.5$ and (c) $\eta = 0.1$.

TABLE 4.1 – 1000×MISE values from 50 replications of sample sizes $n = 200, 1000, 2000$ and 5000, when N follows a Geometric distribution of parameter η.

	Uniform											
	$\eta = 0.9$				$\eta = 0.5$				$\eta = 0.1$			
n	200	1000	2000	5000	200	1000	2000	5000	200	1000	2000	5000
Block	50.04	10.49	7.03	4.18	75.90	11.15	7.61	4.85	102.12	19.15	15.05	13.39
Kernel	47.31	13.76	11.37	9.32	77.52	17.62	14.99	12.60	109.00	70.50	63.24	55.94
	SeparatedBimodal											
	$\eta = 0.9$				$\eta = 0.5$				$\eta = 0.1$			
n	200	1000	2000	5000	200	1000	2000	5000	200	1000	2000	5000
Block	19.63	8.53	6.29	3.69	20.47	9.09	6.89	3.90	62.15	15.27	11.08	7.12
Kernel	14.56	6.30	4.98	3.54	16.30	7.00	5.32	3.82	44.01	13.26	10.86	7.79
	Kurtotic											
	$\eta = 0.9$				$\eta = 0.5$				$\eta = 0.1$			
n	200	1000	2000	5000	200	1000	2000	5000	200	1000	2000	5000
Block	49.73	11.31	8.03	5.52	52.37	11.93	8.04	5.57	98.03	18.08	9.22	7.01
Kernel	45.12	12.27	8.00	5.83	49.43	13.03	8.44	6.13	117.37	21.97	13.04	9.66
	StronglySkewed											
	$\eta = 0.9$				$\eta = 0.5$				$\eta = 0.1$			
n	200	1000	2000	5000	200	1000	2000	5000	200	1000	2000	5000
Block	43.73	9.91	7.78	5.12	36.39	8.69	7.02	4.70	88.49	10.12	9.03	5.93
Kernel	47.71	10.57	8.13	5.97	44.65	11.16	8.68	6.24	128.30	19.62	15.54	10.86

Chapitre 5

On a plug-in wavelet estimator for convolutions of densities

The nonparametric estimation of the m-fold convolution power of an unknown function f is considered. We introduce an estimator based on a plug-in approach and a wavelet hard thresholding estimator. We explore its theoretical asymptotic performances via the mean integrated squared error assuming that f has a certain degree of smoothness. Applications and numerical examples are given for the standard density estimation problem and the deconvolution density estimation problem.

Sommaire

5.1 Problem statement and motivations

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, f be an unknown function related to n *i.i.d.* random variables $Z_1, \ldots, Z_n$ and $m \geq 2$ be a fixed integer. We aim to estimate the m-fold convolution power of f

$$g(x) = \star^m f(x) = \int \ldots \int f(x - u_2 - \ldots - u_m) f(u_2) \ldots f(u_m) du_2 \ldots du_m \tag{5.1}$$

from $Z_1, \ldots, Z_n$. For the case where $m = 2$, note that $\star^2 f(x) = (f \star f)(x) = \int f(x - t) f(t) dt$.

Probably the most famous example concerns the density estimation problem where f is the density of Z_1 and g is the density of $S = \sum_{i=1}^{m} Z_i$. Many quantities of interest in actuarial or financial sciences involve sums of random variables. For example, in the individual risk model, the total amount of claims on a portfolio of insurance contracts is modelled as the sum of all claims on the individual policies. Therefore, probability density functions of sums of random variables are of particular interest. A typical example is the sum of insurance claims, where $(Z_i)_{i=1,\dots,n}$ are individual insurance claims and $S = \sum_{i=1}^{m} Z_i$ is the sum of m claims and m could be interpreted as the expected number of claims in a specified period (e.g. one month). As an example, we refer to [47] which studied the total charges for female patients admitted to the Wisconsin Hospital for circulatory disorders during a year. Another detailed application in the field of health insurance can be found in [77]. Methods and results can be found in [47], [90], [2], [3], [87], [91, 92], [43] and [49]. In particular, [90] have introduced the natural plug-in estimator $\hat{g} = \star^m \hat{f}$, where $\hat{f}$ denotes a kernel estimator.

5.2 Wavelet estimators

First of all, we introduce some basics on wavelets. Then we develop our wavelet hard thresholding estimator and define the Besov balls.

5.2.1 Basics on wavelets

Let $N \geq 1$ be an integer, and ϕ and ψ be the initial wavelet functions of the Daubechies wavelets $db2N$. These functions have the particularity to be compactly supported and $\mathcal{C}^\nu$ where ν is an integer depending on N.

From these wavelet two functions, we define $\phi_{j,k}$ and $\psi_{j,k}$ by

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \qquad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k).$$

Then there exists an integer τ such that, for any integer $\ell \geq \tau$, the collection

$$\mathcal{B} = \{\phi_{\ell,k},\ k \in \mathbb{Z};\ \psi_{j,k};\ j \in \mathbb{N} - \{0, \dots, \ell - 1\},\ k \in \mathbb{Z}\},$$

is an orthonormal basis of $\mathbb{L}_2(\mathbb{R}) = \{h : \mathbb{R} \to \mathbb{R};\ \int h^2(x)dx < \infty\}$.

Suppose that $h \in \mathbb{L}_2(\mathbb{R})$ and $\text{supp}(h) \subseteq [-T, T]$, where $\text{supp}(h)$ denotes the support of h and $T > 0$. Then, for any integer $\ell \geq \tau$, we can write

$$h(x) = \sum_{k \in \Lambda_\ell} \alpha_{\ell,k}\phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k}\psi_{j,k}(x), \tag{5.2}$$

where $\alpha_{j,k}$ and $\beta_{j,k}$ are the wavelet coefficients of h defined by the integrals :

$$\alpha_{j,k} = \int h(x)\phi_{j,k}(x)dx, \qquad \beta_{j,k} = \int h(x)\psi_{j,k}(x)dx \tag{5.3}$$

and Λ_j is a set of consecutive integers with a length proportional to 2^j. For details about wavelet basis, we refer to [29] and [66].

5.2.2 Estimators

Let us consider the general estimation problem described in Section 5.1 and suppose that $f \in \mathbb{L}_2(\mathbb{R})$ and $\text{supp}(f) \subseteq [-T, T]$. We expand the unknown function f on $\mathcal{B}$ as (5.2). Let $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ be estimators of the wavelet coefficients $\alpha_{j,k} = \int f(x)\phi_{j,k}(x)dx$ and $\beta_{j,k} = \int f(x)\psi_{j,k}(x)dx$ respectively.

Following the general approach of [59], we suppose that there exist three constants $C > 0$, $\kappa > 0$ and $\delta > 0$ such that $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ satisfy, for any $j \in \{\tau, \dots, j_1\}$,

(i)

$$\mathbb{E}(|\hat{\alpha}_{j,k} - \alpha_{j,k}|^{\upsilon}) \leq C 2^{\upsilon\delta j} \left(\frac{\ln n}{n}\right)^{\upsilon/2}, \tag{5.4}$$

(ii)

$$\mathbb{E}(|\hat{\beta}_{j,k} - \beta_{j,k}|^{\upsilon}) \leq C 2^{\upsilon\delta j} \left(\frac{\ln n}{n}\right)^{\upsilon/2} \tag{5.5}$$

(iii)

$$\mathbb{P}\left(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \frac{\kappa}{2} 2^{\delta j} \sqrt{\frac{\ln n}{n}}\right) \leq C \left(\frac{\ln n}{n}\right)^{4}, \tag{5.6}$$

where $\upsilon = 4m - 4 \geq 4$ and j_1 is the integer satisfying

$$\left(\frac{n}{\ln n}\right)^{1/(2\delta+1)} < 2^{j_1+1} \leq 2\left(\frac{n}{\ln n}\right)^{1/(2\delta+1)}.$$

For a wide variety of models, one may construct $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ satisfying (5.4), (5.5) and (5.6). The parameter δ plays a major role for some inverse problems ; in the standard statistical model (density,...), it is often equal to 0. Examples are given in Sections 5.5 and 5.5.

Then we define the hard thresholding estimator $\hat{f}$ by

$$\hat{f}(x) = \sum_{k \in \Lambda_\tau} \hat{\alpha}_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k \in \Lambda_j} \hat{\beta}_{j,k} \mathbf{1}_{\left\{|\hat{\beta}_{j,k}| \geq \kappa 2^{\delta j}\sqrt{\ln n/n}\right\}} \psi_{j,k}(x), \tag{5.7}$$

where, for any random event $\mathcal{A}$, $\mathbf{1}_{\mathcal{A}}$ is the indicator function on $\mathcal{A}$.

The idea of the hard thresholding rule in (5.7) is to make a term-by-term selection : only the "large" unknown wavelet coefficients of f which contain its main characteristics are estimated. Details can be found in e.g. [5], [51] and [99].

Using (5.7), we consider the following plug-in estimator for $g = \star^m f$:

$$\hat{g}(x) = \star^m \hat{f}(x). \tag{5.8}$$

To study its asymptotic performance, we need some smoothness assumptions on f. In this study, as usual in wavelet estimation, we suppose that f belongs to Besov balls defined below.

5.2.3 Besov balls

Let h be a function such that $h \in \mathbb{L}_2(\mathbb{R})$ and $\text{supp}(h) \subseteq [-T,T]$. We say that $h \in B^s_{p,r}(M)$ with $M > 0$, $s > 0$, $p \geq 1$ and $r \geq 1$ if there exists a constant $C > 0$ such that (5.3) satisfy

$$\left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k \in \Lambda_j} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq C.$$

In this expression, s is a smoothness parameter and p and r are norm parameters. We consider such Besov balls essentially because of their executional expressive power. In particular, they contain the Hölder and Sobolev balls. See e.g. [69] and [51, Chapter 9].

5.3 Upper bound

Theorem 5.3.1 below investigates the rates of convergence for $\hat{g}$ (5.8) under the MISE over Besov balls.

Theorem 5.3.1. *Consider the estimation problem and notations of Section 5.1. Suppose that* $\text{supp}(f) \subseteq [-T,T]$, *where* $T > 0$ *is a fixed constant, and there exists a constant* $C > 0$ *such that* $\int |f(x)|^{4m-4} dx \leq C$. *Let* $\hat{g}$ *be* (5.8) *(under* (5.4), (5.5) *and* (5.6)*). Suppose that* $f \in B^s_{p,r}(M)$ *with* $r \geq 1$, $\{p \geq 4$ *and* $s > 0\}$ *or* $\{p \in [1,4)$ *and* $s > \max((2\delta+1)/p, (4/p-1)(\delta+1/2)\}$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx \right) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+2\delta+1)}.$$

The proof of Theorem 5.3.1 uses a suitable decomposition of the MISE and a result on the rates of convergence of $\hat{f}$ under the $\mathbb{L}_p$-risk with $p \in \{2,4\}$ over Besov balls derived from [59]. Let us mention that the control of both of the $\mathbb{L}_2$-risk (MISE) and $\mathbb{L}_4$-risk motivates the consideration of the wavelet hard thresholding estimator in the definition of $\hat{g}$ (5.8) instead of other wavelet estimators (as the block thresholding ones introduced by [8]).

Theorem 5.3.1 shows that under mild assumptions on

— the model : only (5.4), (5.5) and (5.6) are required,

— f : only $\text{supp}(f) \subseteq [-T,T]$, $\int |f(x)|^{4m-4} dx \leq C$, $f \in B^s_{p,r}(M)$ with $r \geq 1$, $\{p \geq 4$ and $s > 0\}$ or $\{p \in [1,4)$ and $s > \max((2\delta+1)/p, (4/p-1)(\delta+1/2)\}$ (including the inhomogeneous zone of the Besov balls corresponding to $p \in [1,2)$) are required,

the estimator $\hat{g}$ attains a "fast" rate of convergence. "Fast" in the sense that it is close to the parametric rate $1/n$. However, we do not claim it to be optimal in the minimax sense. This point will be discussed for two particular density models in the next section.

It is important to mention that the rate of convergence $(\ln n/n)^{2s/(2s+2\delta+1)}$ is the near optimal one in the minimax sense for $\hat{f}$ (not $\hat{g}$) under the MISE over Besov balls for various standard nonparametric setting (density model, nonparametric regression model, deconvolution density model,...). See e.g. [51], [45] and [97].

5.4 Application I : the density model

5.4.1 Upper bound

We observe n *i.i.d.* random variables $Z_1, \ldots, Z_n$ with common unknown density f. For a fixed integer $m \geq 2$, let $S = \sum_{i=1}^{m} Z_i$ and g be the density of S. The goal is to estimate g from $Z_1, \ldots, Z_n$. As mentioned in Section 5.1, such a problem has already been considered with kernel-type estimators and various settings by e.g. [47], [90], [2], [3], [91, 92] and [43].

Proposition 5.4.1 below investigates the rates of convergence of $\hat{g}$ (5.8) constructed from a specific wavelet hard thresholding estimator $\hat{f}$ under the MISE over Besov balls.

Proposition 5.4.1. *Consider the standard density model and the associated notations. Suppose that* $\text{supp}(f) \subseteq [-T, T]$, *where* $T > 0$ *is a fixed constant, and there exists a constant* $C > 0$ *such that* $\sup_{x \in \mathbb{R}} f(x) \leq C$.

Let $\hat{g}$ *be* (5.8) *with* $\delta = 0$,

$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j,k}(Z_i), \qquad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \psi_{j,k}(Z_i). \tag{5.9}$$

Suppose that $f \in B^s_{p,r}(M)$ *with* $r \geq 1$, $\{p \geq 4$ *and* $s > 0\}$ *or* $\{p \in [1, 4)$ *and* $s > \max(1/p, (2/p - 1/2)\}$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx \right) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+1)}.$$

As noted in Section 5.1, the rate of convergence $(\ln n/n)^{2s/(2s+1)}$ is a bit slower than the one reached by the kernel estimator of [90] (i.e. the parametric rate $1/n$). The larger s is, the closer they are. The main contribution of Proposition 5.4.1 concerns the assumptions on f : we do not need to have f four times differentiable with fourth derivative continuous, $\sup_{x \in \mathbb{R}} |f^{(j)}(x)| < \infty$ for $j \in \{0, 1, 2, 3, 4\}$, f'' and $f^{(4)}$ are integrable as in [90, Theorem 3] ; the assumption that f belongs to Besov balls includes a wide class of functions which does not satisfy such assumptions.

Remark that $(\ln n/n)^{2s/(2s+1)}$ is the "near optimal" rates of convergence in the minimax sense for the standard density estimation problem for $\hat{f}$ under the MISE over Besov balls. See [40, Theorems 2 and 3].

5.4.2 Simulation results

In the following simulation study, we have analyzed the performances of our adaptive wavelet estimation procedure on a family of normal mixture densities ("SeparatedBimodal", "Outlier" and "DiscreteComb', initially introduced in [68]) representing different degrees of smoothness (see Fig. 2.4). We have adapted the formulae given by [68] to simulate such densities by choosing the parameters so that

$$\min_l (q\mu_l - 3q\sigma_l) = -3q, \quad \max_l (q\mu_l + 3q\sigma_l) = 3q$$

where $l = 1, \ldots, p$ with p the number of densities in the mixture and q a scaling parameter. Thereby, it is very unlikely to have values outside the interval $[-4q, 4q]$ $(\subseteq [-T, T])$ and we lose little by assuming compact support (see Fig. 5.1).

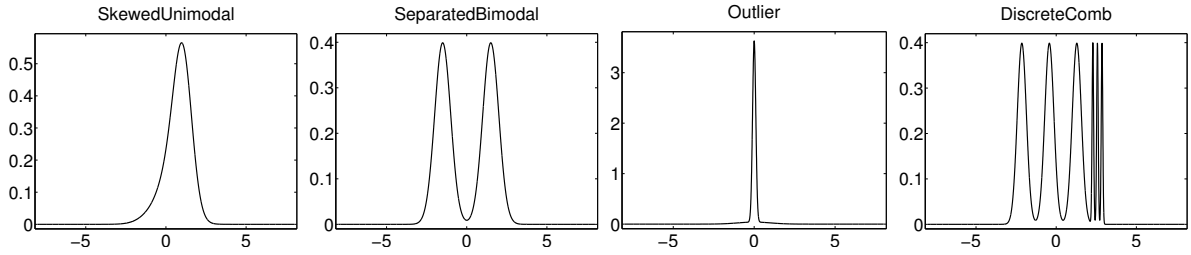


FIGURE 5.1 – Theoretical densities from $n = 1000$ samples $Z_1, \ldots, Z_n$.

Since our estimation method is adaptive, we have chosen a predetermined threshold κ (universal thresholding, see e.g. [40]) for all the tests and the Symmlet wavelet with 6 vanishing moments was used throughout all experiments. The finest resolution level j_1 used in all our simulations was chosen to be the maximum resolution level allowed by the discretization. For each density, $n = 1000$ independent samples were generated and the MISE was approximated as an average of the Integrated Squared Error (ISE) over 100 replications. The m-fold convolution product of $\hat{f}$ with itself defined by (5.8) can be efficiently computed numerically using the fast Fourier transform, thanks to the convolution theorem. Typical reconstructions from a single simulation are depicted in Fig. 5.2 for $m = 1$, $m = 2$ and $m = 3$ respectively. One can see that our adaptive hard thresholding estimator is very effective to estimate each of the nine densities.

Then, we have compared the performance of our adaptive wavelet estimator to those of two different kernel-based estimators. The first one, presented in [90], is based on convolving kernel density estimators : $\hat{g} = \star^m \hat{f}$, where $\hat{f}$ denotes a kernel estimator. The other one, introduced by [47], is the Frees type local U-statistic estimator defined as follow

$$\hat{g}(x) = \frac{1}{\binom{n}{m} b} \sum_{(n,m)} K\left(\frac{x - h(Z_{i_1}, \ldots, Z_{i_m})}{b}\right), \tag{5.10}$$

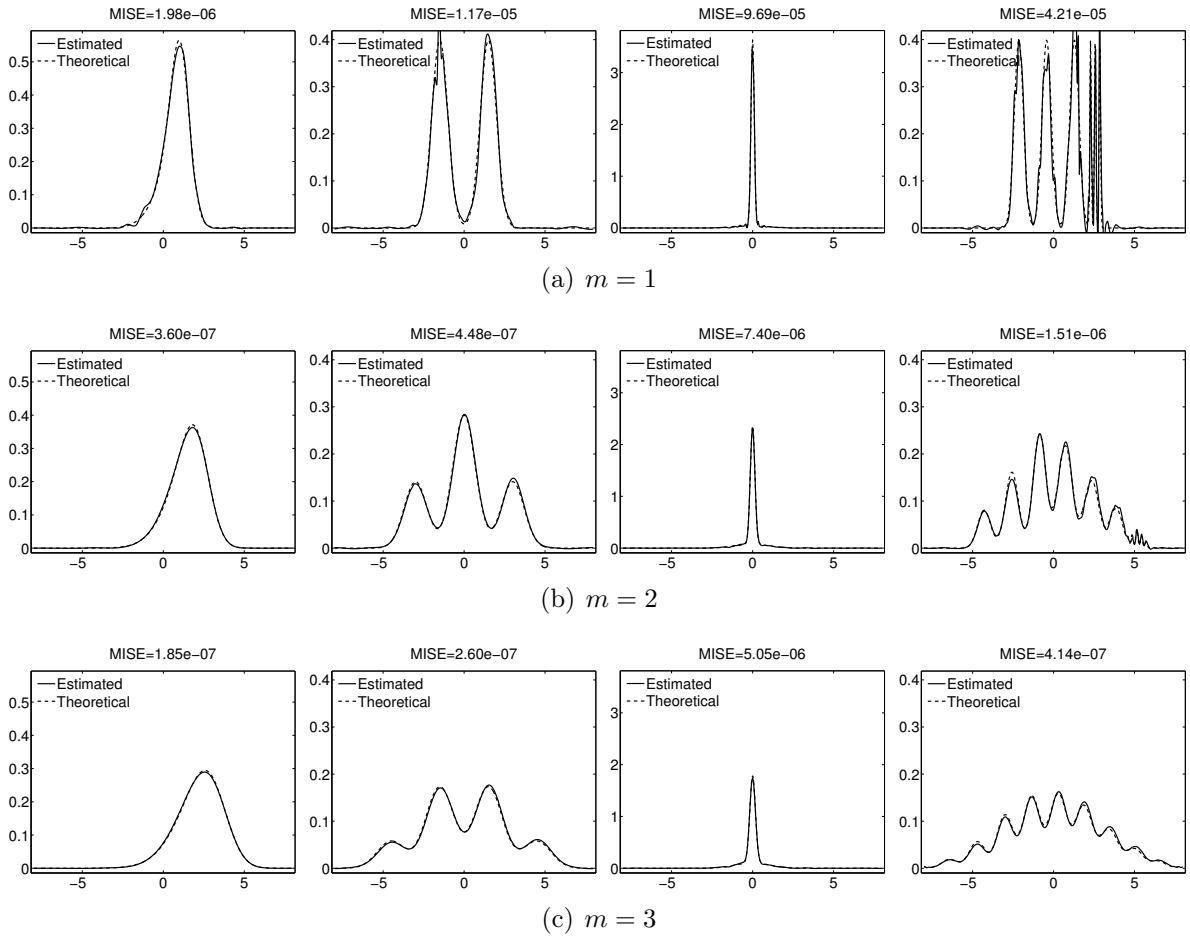
where b is the bandwidth or smoothing parameter, K is a kernel function and $\sum_{(n,m)}$ denotes summation over all $\binom{n}{m}$ subsets. Recall that we have focused here on the interesting case where $h(Z_1, \ldots, Z_m) = \sum_{i=1}^m Z_i$ (see [47] for applications).

In the sequel, we name the estimator of [90] by 'Kernel', the one of [47] by 'Frees' and our estimator by 'Wavelet'.

In the case of *i.i.d.* random variables, the choice of the kernel is not crucial for density estimation. However, it is well known that the choice of the bandwidth is very important. Many procedures of bandwidth selection for density estimation have been developed in the literature (details can be found in [70] where several methods are compared). Here, for both kernel-based estimators, we have been focused on a global bandwidth selector : the *rule of thumb* (ROT) bandwidth selector (see e.g. [96]). [2] derived the asymptotic mean integrated square error and the optimal bandwidth for the Frees estimator (5.10). Thanks to [2, Theorem 2.2] with $r = 2$ and the Gaussian kernel, the optimal bandwidth is given by

$$b_{ROT} = 1.06 \min(\hat{\sigma}, Q/1.34) \binom{n}{m}^{-1/5} \tag{5.11}$$

where $\hat{\sigma}$ is the sample standard deviation and Q is the interquartile range. This choice was motivated by the major drawback of the Frees estimator which is the computation time required to evaluate it (see Table 5.2).



(a) $m = 1$

(b) $m = 2$

(c) $m = 3$

FIGURE 5.2 – Original densities (dashed) and our wavelet hard thresholding estimator $\hat{g}$ (solid) from only one repetition of $n = 1000$ samples $Z_1, \ldots, Z_n$. From left to right SkewedUnimodal, SeparatedBimodal, Outlier and DiscreteComb. (a) : $m = 1$, (b) : $m = 2$, (c) : $m = 3$.

We evaluated the three procedures on small to medium sample. Each method was applied to $n = 10, 20, 50, 100$ data points of each of the densities. All experiments were conducted using a Gaussian kernel for both kernel-based methods. The MISE from 100 repetitions are tabulated in Table 5.1. It shows that none of the methods clearly outperforms the others in all cases. However, our estimator is slightly better than the others in many cases. Table 5.2 reports the average execution times in seconds for $m = 2$ and $m = 3$. For the Frees estimator, the computational cost increases dramatically as far as the sampling parameter n increases and during the computation to estimate the density of the sum of more than two *i.i.d.* random variables. From a practical point of view, unlike Frees's estimator, methods based on a plug-in approach can easily be computed for $m > 3$ and larger samples.

We conclude this section by a comparison to the natural kernel plug-in estimator of [90] on larger samples ($n = 1000, 2000, 5000$). Table 5.3 summarizes the results. Our wavelet method clearly outperforms the kernel one for all tests densities and all sample.

Remark 5.4.1. *We propose to illustrate here our proposed estimator on the positive half-line for an heavy-tailed density function, the strongly skewed density. This density departs in the direction of skewness and was chosen to resemble to lognormal which is commonly*

TABLE 5.1 – $1e4\times$MISE values from 100 replications for each method. From top to bottom SkewedUnimodal, SeparatedBimodal, Outlier, AsymmetricClaw for $m = 2$ (left) and $m = 3$ (right).

1.0e-04×	*SkewedUnimodal*, $m = 2$					*SkewedUnimodal*, $m = 3$				
n	10	20	50	100	n	10	20	50	100
Wavelet	89.52	32.10	11.47	6.90	Wavelet	73.02	28.63	10.44	6.44
Kernel	55.76	25.78	10.64	6.97	Kernel	52.16	23.88	9.79	6.49
Frees	80.56	32.09	11.52	7.08	Frees	85.79	31.78	10.66	6.60
	SeparatedBimodal, $m = 2$					*SeparatedBimodal*, $m = 3$				
n	10	20	50	100	n	10	20	50	100
Wavelet	79.17	38.08	12.30	6.32	Wavelet	52.36	27.10	8.60	4.39
Kernel	57.97	45.59	36.57	35.20	Kernel	36.65	25.25	15.87	14.31
Frees	60.45	33.23	13.01	7.03	Frees	54.06	26.52	8.80	4.41
	Outlier, $m = 2$					*Outlier*, $m = 3$				
n	10	20	50	100	n	10	20	50	100
Wavelet	542.4	257.4	82.61	49.81	Wavelet	525.6	251.9	81.27	48.37
Kernel	597.1	248.1	88.78	57.39	Kernel	750.8	264.4	84.16	53.29
Frees	602.8	265.6	86.42	49.87	Frees	691.8	271.2	84.81	49.41
	DiscreteComb, $m = 2$					*DiscreteComb*, $m = 3$				
n	10	20	50	100	n	10	20	50	100
Wavelet	87.35	42.13	14.20	7.40	Wavelet	46.14	21.72	7.36	3.86
Kernel	44.62	31.72	26.51	24.55	Kernel	27.03	14.97	9.65	7.55
Frees	50.74	32.05	20.92	14.23	Frees	37.25	16.87	6.28	3.40

TABLE 5.2 – Execution times in seconds for $m = 2$ and $m = 3$ (from only one realization). The algorithms were run under Matlab with an Intel Core 2 duo 3.06GHz CPU, 4Gb RAM.

	$m = 2$					$m = 3$				
n	10	20	50	100	n	10	20	50	100	200
Wavelet	0.01	0.01	0.01	0.02	Wavelet	0.01	0.01	0.01	0.02	0.02
Kernel	0.04	0.04	0.04	0.05	Kernel	0.04	0.04	0.04	0.05	0.05
Frees	0.17	0.22	0.53	1.38	Frees	0.37	1.30	9.62	143	2037

used in insurance application (for example for fire insurance). Typical reconstructions from a single simulation are depicted in Figure 5.3 for $m = 1$, $m = 2$ and $m = 3$.

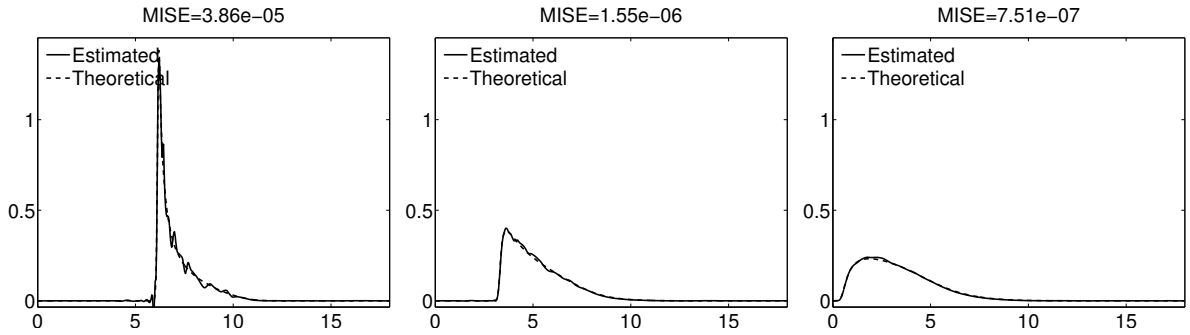
5.4.3 Application to insurance data

The sum of m *i.i.d.* random variables $(Z_i)_{i=1,\dots,n}$ plays an important role in many insurance problems. A classical example is the sum of insurance claims, where Z_i are individual insurance claims, $S = Z_1 + \dots Z_m$ is the sum of m claims and m could be interpreted as the expected number of claims in a specified period of time (e.g. one year). In order to illustrate the real-life applicability of our results, we consider the hospital data example, which was introduced by Frees. The data plotted in Figure 5.4(a) consist

TABLE 5.3 – $1e6\times$MISE values from 100 replications. From top to bottom SkewedUnimodal, SeparatedBimodal, Outlier and AsymmetricClaw for $m = 2$ (left) and $m = 3$ (right).

$1.0e\text{-}06\times$	*SkewedUnimodal*, $m = 2$		
n	1000	2000	5000
Wavelet	25.32	5.62	2.90
Kernel	27.46	10.09	4.86
	SeparatedBimodal, $m = 2$		
n	1000	2000	5000
Wavelet	19.94	9.12	4.02
Kernel	509.5	394.9	258.8
	Outlier, $m = 2$		
n	1000	2000	5000
Wavelet	148.1	39.62	21.46
Kernel	182.6	54.94	34.86
	DiscreteComb, $m = 2$		
n	1000	2000	5000
Wavelet	16.77	10.64	3.68
Kernel	541.8	521.0	474.1

	SkewedUnimodal, $m = 3$		
n	1000	2000	5000
Wavelet	24.35	5.24	2.30
Kernel	26.16	8.34	3.53
	SeparatedBimodal, $m = 3$		
n	1000	2000	5000
Wavelet	14.26	7.75	2.83
Kernel	224.5	185.3	129.3
	Outlier, $m = 3$		
n	1000	2000	5000
Wavelet	139.5	37.05	24.18
Kernel	153.6	45.32	32.97
	DiscreteComb, $m = 3$		
n	1000	2000	5000
Wavelet	9.48	4.90	1.74
Kernel	125.3	118.6	111.2



(a) $m = 1$ (b) $m = 2$ (c) $m = 3$

FIGURE 5.3 – Original density (dashed) and our wavelet hard thresholding estimator $\hat{g}$ (solid) from only one repetition of $n = 2000$ samples $Z_1, \ldots, Z_n$.

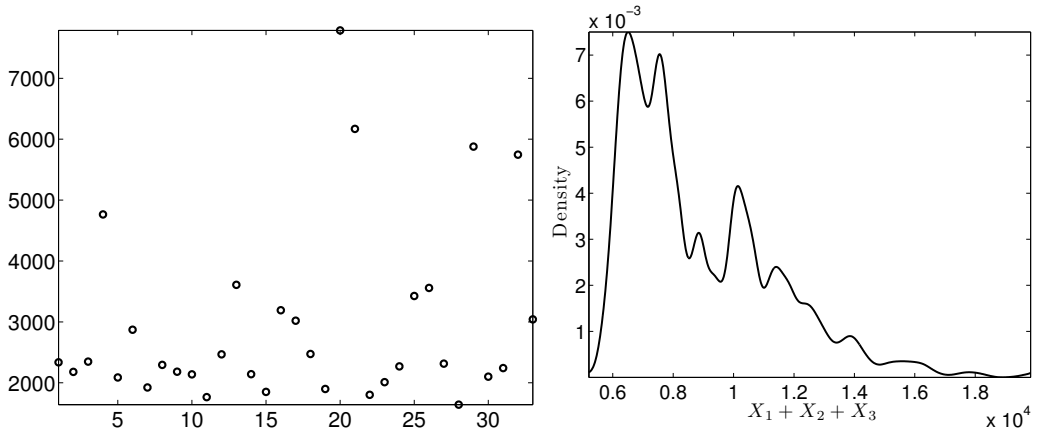
of measurements of the 1989 total charges for 33 patients at a Wisconsin Hospital. Each patient was female, age 30–49, and admitted to the hospital for circulatory disorders. Figure 5.4(b)–(c) depicts density estimate of the sum of claims for $m = 3$.

5.5 Application II : the deconvolution density model

5.5.1 Upper bound

We observe n *i.i.d.* random variables $Z_1, \ldots, Z_n$ where, for any $i \in \{1, \ldots, n\}$,

$$Z_i = X_i + \epsilon_i, \tag{5.12}$$



(a) (b)

FIGURE 5.4 – (a) : 1989 total hospital charges (in dollars) for 33 females aged 30-49 hospitalized for circulatory disorders from a Wisconsin Hospital (see [47]). (b) : $\hat{g}$ for $m = 3$.

$X_1, \ldots, X_n$ are *i.i.d.* random variables and $\epsilon_1, \ldots, \epsilon_n$ are *i.i.d.* random variables. Classically, $X_1, \ldots, X_n$ are measurements of some characteristic of interest contaminated by noise represented by $\epsilon_1, \ldots, \epsilon_n$. For any $i \in \{1, \ldots, n\}$, X_i and ϵ_i are independent. The density of X_1 is unknown and denoted f, whereas the one of ϵ_1 is known and denoted h. For a fixed integer $m \geq 2$, let $S = \sum_{i=1}^{m} X_i$ and g be the density of S. The goal is to estimate g from $Z_1, \ldots, Z_n$. This problem can be viewed as a generalization of the standard deconvolution density one which corresponds to $m = 1$. See e.g. [12], [44], [46], [81], [45], [7], [32], [34] and [60]. However, to the best of our knowledge, the general problem i.e. with $m \geq 2$ is a new challenge.

Proposition 5.5.1 below investigates the rates of convergence of $\hat{g}$ (5.8) constructed from a specific wavelet hard thresholding estimator $\hat{f}$ under the MISE over Besov balls.

Proposition 5.5.1. *Consider* (5.12) *and the associated notations. We define the Fourier transform of an integrable function* u *by* $\mathcal{F}(u)(x) = \int_{-\infty}^{\infty} u(y)e^{-ixy}dy$, $x \in \mathbb{R}$. *The notation* $\overline{\cdot}$ *will be used for the complex conjugate.*

Suppose that $\text{supp}(f) \subseteq [-T, T]$, *where* $T > 0$ *is a fixed constant, and there exist three constants* $C > 0$, $c > 0$ *and* $\delta > 1$ *such that*

$$\sup_{x \in \mathbb{R}} h(x) \leq C, \qquad |\mathcal{F}(h)(x)| \geq \frac{c}{(1+x^2)^{\delta/2}}, \qquad x \in \mathbb{R}. \tag{5.13}$$

Let $\hat{g}$ *be* (5.8) *with*

$$\hat{\alpha}_{j,k} = \frac{1}{2\pi n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\phi_{j,k})}(x)}{\mathcal{F}(h)(x)} e^{-ixZ_i} dx \tag{5.14}$$

and

$$\hat{\beta}_{j,k} = \frac{1}{2\pi n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\psi_{j,k})}(x)}{\mathcal{F}(h)(x)} e^{-ixZ_i} dx. \tag{5.15}$$

Suppose that $f \in B^s_{p,r}(M)$ *with* $r \geq 1$, $\{p \geq 4$ *and* $s > 0\}$ *or* $\{p \in [1, 4)$ *and* $s > \max((2\delta+1)/p, (4/p-1)(\delta+1/2)\}$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx\right) \leq C \left(\frac{\ln n}{n}\right)^{2s/(2s+2\delta+1)}.$$

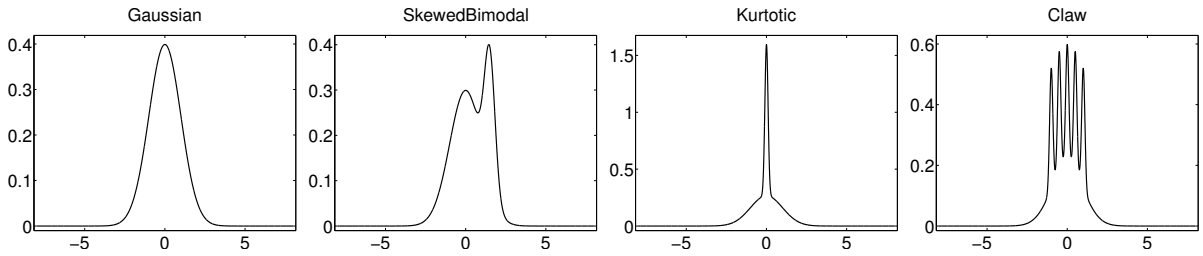


FIGURE 5.5 – Theoretical densities.

To the best of our knowledge, there is no asymptotic result for kernel estimators in this deconvolution setting. Proposition 5.5.1 provides a first theoretical result on the possible achievable rate of convergence for an estimator of g under the MISE over Besov balls. Let us mention that the rate of convergence $(\ln n/n)^{2s/(2s+2\delta+1)}$ corresponds to the "near optimal" one in the minimax sense for $\hat{f}$ under the MISE over Besov balls. See [45, Theorem 2].

5.5.2 Simulation results

In this simulation, $n = 1000$ samples $Z_1, \ldots, Z_n$ were generated according to model (5.12) and we considered Laplace errors (which respect the standard ordinary smooth assumption). The data sets used in this deconvolution study are also normal mixture densities (see [68] for formulae of these densities) different representing degrees of smoothness.

Fig. 5.6 shows the results of $\hat{g}$ for $m = 1$, $m = 2$ and $m = 3$ respectively. Clearly, for these nine densities, even if the estimation problem becomes harder our adaptive hard thresholding estimator is very effective.

5.6 Conclusion and perspectives

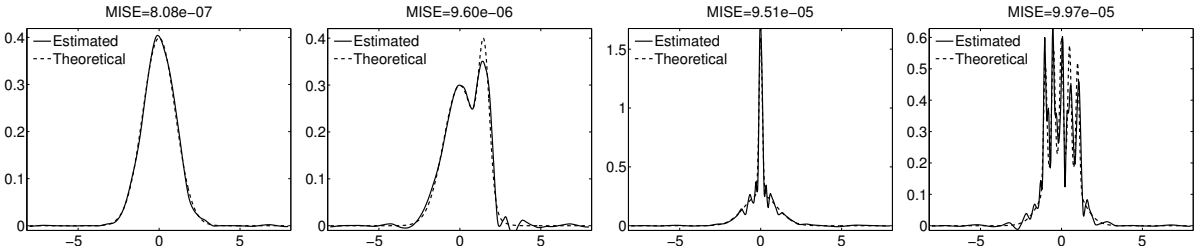
The agreement of our simulations with our theoretical findings show the relevance of our estimator in the context of two classical density estimation problems. The practical comparisons to state-of-the art methods such as the estimator of [47] or the one of [90] have demonstrated the usefulness and the efficiency of adaptive thresholding methods in estimating densities of the sum of random variables. It would be interesting to include both theoretical and practical comparisons with other wavelet thresholding estimators as the block thresholding one (see e.g. [8] and [25]). Another theoretical challenge is to determine the optimal lower bounds under the MISE over Besov balls. These aspects need further investigations that we leave for a future work.

5.7 Proofs

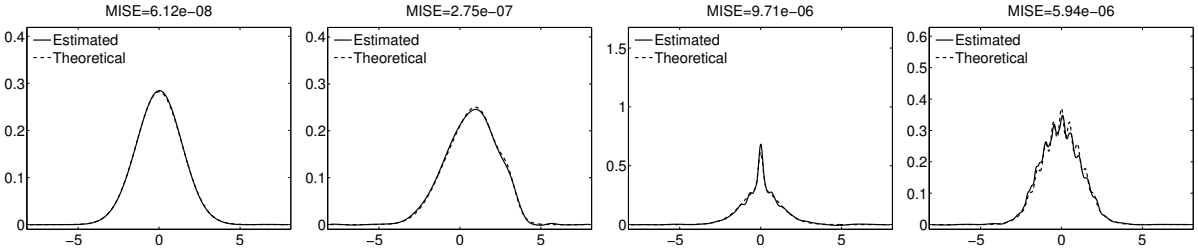
Proof of Theorem 5.3.1. Let us define the Fourier transform of a function $u \in \mathbb{L}_1(\mathbb{R})$ by

$$\mathcal{F}(u)(y) = \int_{-\infty}^{\infty} u(x)e^{-iyx}dx, \qquad y \in \mathbb{R}.$$

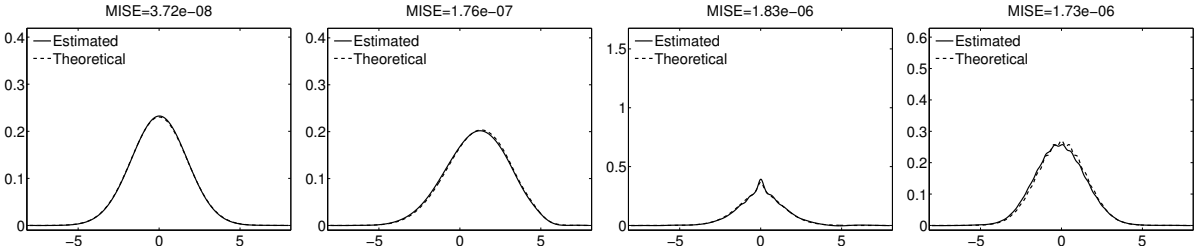
By definition of $\hat{g}$ and g, we have $\mathcal{F}(\hat{g})(y) = \mathcal{F}(\star^m \hat{f})(y) = (\mathcal{F}(\hat{f})(y))^m$ and $\mathcal{F}(g)(y) = \mathcal{F}(\star^m f)(y) = (\mathcal{F}(f)(y))^m$. Owing to the previous equalities and the Parseval theorem, we



(a) $m = 1$



(b) $m = 2$



(c) $m = 3$

FIGURE 5.6 – Original densities (dashed) and our wavelet hard thresholding estimator $\hat{g}$ (solid) from one realization of $n = 1000$ observations $Z_1, \ldots, Z_n$ generated according to (5.12), where $h(x) = (1/2)e^{-|x|}$, $x \in \mathbb{R}$. (a) : $m = 1$. (b) : $m = 2$. (c) : $m = 3$. (note that (5.13) is satisfied with $\delta = 2$).

obtain

$$\int |\hat{g}(x) - g(x)|^2 dx = \frac{1}{2\pi} \int |(\mathcal{F}(\hat{f})(y))^m - (\mathcal{F}(f)(y))^m|^2 dy. \tag{5.16}$$

Now remark that, for any $(u, v) \in \mathbb{C}^2$, the factor theorem yields : $u^m - v^m = (u - v) \sum_{k=0}^{m-1} v^k u^{(m-1)-k}$. It follows from $\binom{m-1}{k} \geq 1$, $k \in \{0, \ldots, m-1\}$, and the binomial theorem that

$$\begin{aligned} |u^m - v^m| &\leq |u - v| \sum_{k=0}^{m-1} |v|^k |u|^{(m-1)-k} \leq |u - v| \sum_{k=0}^{m-1} \binom{m-1}{k} |v|^k |u|^{(m-1)-k} \\ &= |u - v|(|u| + |v|)^{m-1} \leq |u - v|(|u - v| + 2|v|)^{m-1}. \end{aligned} \tag{5.17}$$

Using (5.16), (5.17), the linearity of $\mathcal{F}$, the inequalities : $|\mathcal{F}(\hat{f}-f)(y)| \leq \int |\hat{f}(x)-f(x)|dx$, $|\mathcal{F}(f)(y)| \leq \int |f(x)|dx \leq C$, and the Parseval theorem, we have

$$\begin{aligned}
&\int |\hat{g}(x) - g(x)|^2 dx \\
&\leq \frac{1}{2\pi} \int |\mathcal{F}(\hat{f})(y) - \mathcal{F}(f)(y)|^2 \left(|\mathcal{F}(\hat{f})(y) - \mathcal{F}(f)(y)| + 2|\mathcal{F}(f)(y)| \right)^{2m-2} dy \\
&= \frac{1}{2\pi} \int |\mathcal{F}(\hat{f} - f)(y)|^2 \left(|\mathcal{F}(\hat{f} - f)(y)| + 2|\mathcal{F}(f)(y)| \right)^{2m-2} dy \\
&\leq C \left(\int |\hat{f}(x) - f(x)|dx + 1 \right)^{2m-2} \int |\hat{f}(x) - f(x)|^2 dx.
\end{aligned}$$

Noticing that the last term is the product of two random variables, the Cauchy-Schwarz inequality yields

$$\mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx \right) \leq CIJ, \tag{5.18}$$

where

$$I = \sqrt{\mathbb{E}\left(\left(\int |\hat{f}(x) - f(x)|dx + 1 \right)^{4m-4} \right)}, \qquad J = \sqrt{\mathbb{E}\left(\left(\int |\hat{f}(x) - f(x)|^2 dx \right)^2 \right)}.$$

Let us now bound I and J, in turn.

Upper bound for I. Using $|x+y|^a \leq 2^{a-1}(|x|^a + |y|^a)$, $(x,y) \in \mathbb{R}^2$, $a \geq 1$, the Hölder inequality and $\text{supp}(f) \subseteq [-T,T]$, we have

$$I \leq C\sqrt{\mathbb{E}\left(\int |\hat{f}(x) - f(x)|^{4m-4} dx \right) + 1}.$$

Using the definition of $\hat{f}$ (5.7) and the wavelet expansion of f we have

$$\mathbb{E}\left(\int |\hat{f}(x) - f(x)|^{4m-4} dx \right) \leq C(E + F + G),$$

where

$$F = \int \mathbb{E}\left(\left(\sum_{j=\tau}^{j_1} \sum_{k \in \Lambda_j} (\hat{\beta}_{j,k} \mathbf{1}_{\left\{ |\hat{\beta}_{j,k}| \geq \kappa 2^{\delta j} \sqrt{\ln n/n} \right\}} - \beta_{j,k}) \psi_{j,k}(x) \right)^{4m-4} \right) dx$$

$$E = \sum_{k \in \Lambda_\tau} \mathbb{E}(|\hat{\alpha}_{\tau,k} - \alpha_{\tau,k}|^{4m-4}), \qquad G = \int \left(\sum_{j=j_1+1}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k} \psi_{j,k}(x) \right)^{4m-4} dx$$

Using (5.5) and $\int |f(x)|^{4m-4} dx \leq C$, we obtain

$$E + G \leq C.$$

Now observe that

$$|\hat{\beta}_{j,k} \mathbf{1}_{\left\{ |\hat{\beta}_{j,k}| \geq \kappa 2^{\delta j} \sqrt{\ln n/n} \right\}} - \beta_{j,k}| \leq |\hat{\beta}_{j,k} - \beta_{j,k}| + |\beta_{j,k}|,$$

The unconditional property of the wavelet basis (see [59, Subsection 4.2]) gives

$$F \leq C(G+H)$$

where

$$G = \int \mathbb{E}\left(\left(\sum_{j=\tau}^{j_1}\sum_{k\in\Lambda_j}(\hat{\beta}_{j,k}-\beta_{j,k})\psi_{j,k}(x)\right)^{4m-4}\right)dx, \quad H = \int\left(\sum_{j=\tau}^{j_1}\sum_{k\in\Lambda_j}\beta_{j,k}\psi_{j,k}(x)\right)^{4m-4}dx.$$

As $\int |f(x)|^{4m-4}dx \leq C$, we have

$$H \leq C.$$

Arguing similarly to [59, Theorem 5.1] and using (5.5), we obtain

$$G \leq C\left(\frac{\ln n}{n}\right)^{2m-2} 2^{(1+2\delta)j_1(2m-2)} \leq C.$$

Therefore

$$I \leq C. \tag{5.19}$$

Upper bound for J. It follows from the Hölder inequality and $\mathrm{supp}(f) \subseteq [-T,T]$ that

$$J \leq C\sqrt{\mathbb{E}\left(\int |\hat{f}(x)-f(x)|^4dx\right)}.$$

We now need the following result.

Theorem 5.7.1 ([59]). *Let $\theta > 1$, f be a function such that $f \in \mathbb{L}_2(\mathbb{R})$ and $\mathrm{supp}(f) \subseteq [-T,T]$ with $T>0$, and $\hat{f}$ be (5.7) under (5.4), (5.5) with $\upsilon = 2\theta$ and (5.6) with $(\ln n/n)^\theta$ instead of $(\ln n/n)^4$. Then, for any $r \geq 1$, any $\{p \geq \theta$ and $s>0\}$ or any $\{p \in [1,\theta)$ and $s > \max((2\delta+1)/p, (\theta/p-1)(\delta+1/2)\}$, there exists a constant $C>0$ such that*

$$\sup_{f\in B^s_{p,r}(M)} \mathbb{E}\left(\int |\hat{f}(x)-f(x)|^\theta dx\right) \leq C\left(\frac{\ln n}{n}\right)^{\theta s/(2s+2\delta+1)}.$$

Theorem 5.7.1 can be proved using arguments similar to [59, Theorem 5.1] and [21, Theorem 4.2].

In light of Theorem 5.7.1 with $\theta = 4$, we have

$$J \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+2\delta+1)}. \tag{5.20}$$

Putting (5.18), (5.19) and (5.20) together, we obtain

$$\mathbb{E}\left(\int |\hat{g}(x)-g(x)|^2dx\right) \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+2\delta+1)}.$$

Theorem 5.3.1 is proved.

□

This ends the proof of Theorem 5.3.1.

□

Proof of Proposition 5.4.1. Owing to [40, Subsection 5.1.1, (16) and (17)], under the assumptions $\text{supp}(f) \in [-T, T]$, the estimators $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ (5.9) satisfy (5.4), (5.5) and (5.6) with $\delta = 0$. The proofs are based on the decomposition :

$$\hat{\beta}_{j,k} - \beta_{j,k} = \frac{1}{n}\sum_{i=1}^{n} U_i, \qquad U_i = \psi_{j,k}(Z_i) - \beta_{j,k},$$

where $U_1, \ldots, U_n$ are i.i.d. with $\mathbb{E}(U_1) = 0$, $|U_1| \leq C2^{j_1/2} \leq (n/\ln n)^{1/2}$ and $\mathbb{E}(U_1^2) \leq C$. Then (ii) follows from the Rosenthal inequality and (iii) from the Bernstein inequality. The point (i) is similar to (ii) but with ϕ instead of ψ. The rest of the proof follows from Theorem 5.3.1.

□

Proof of Proposition 5.5.1. Owing to [45, E. Proof of Theorem 7], under the assumptions $\text{supp}(f) \subseteq [-T, T]$ and (5.13), the estimators $\hat{\alpha}_{j,k}$ (5.14) and $\hat{\beta}_{j,k}$ (5.15) satisfy (5.4), (5.5) and (5.6) with the same δ. The proofs are based on the decomposition :

$$\hat{\beta}_{j,k} - \beta_{j,k} = \frac{1}{n}\sum_{i=1}^{n} V_i, \qquad V_i = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\psi_{j,k})}(x)}{\mathcal{F}(h)(x)} e^{-ixZ_i} dx - \beta_{j,k},$$

where $V_1, \ldots, V_n$ are i.i.d. with $\mathbb{E}(V_1) = 0$, $|V_1| \leq C2^{j_1(\delta+1/2)} \leq (n/\ln n)^{1/2}$ and $\mathbb{E}(V_1^2) \leq C2^{2\delta j}$. Then (ii) follows from the Rosenthal inequality and (iii) from the Bernstein inequality. The point (i) is similar to (ii) but with ϕ instead of ψ. We obtain the desired result via Theorem 5.3.1.

□

Chapitre 6

Fast nonparametric estimation for convolutions of densities

This chapter is concerned with the problem of estimating the convolution of densities. We propose an adaptive estimator based on kernel methods, Fourier analysis and the Lepski method. We study the $\mathbb{L}_2$-risk properties of the estimator. Fast and new rates of convergence are determined for a wide class of unknown functions. Numerical illustrations, on both simulated and real data, are provided to assess the performance of our procedure.

Sommaire

6.1 Introduction

6.1.1 Problem statement and motivations

Many quantities of interest in actuarial or financial sciences involve sums of random variables. For example, in the individual risk model, the total amount of claims on a portfolio of insurance contracts is modeled as the sum of all claims on the individual policies. Therefore, probability density functions of sums of random variables are of particular

interest. Typically, such functions are not available in a closed form. Hence, in order to compute functionals of sums of random variables, estimation methods are often used.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, f the unknown density function of a random variable $X : \Omega \to \mathbb{R}$, n a positive integer, $X_1, \ldots, X_n$ a n *i.i.d.* sample of X, m a positive integer and g the m-fold convolution defined by

$$g(x) = \star_m f(x) = \int \ldots \int f(x - x_2 - \ldots - x_m) f(x_2) \ldots f(x_m) dx_2 \ldots dx_m. \tag{6.1}$$

Let us note that g is the density of the random variable $S = \sum_{v=1}^{m} X_v$. We aim to estimate g from $X_1, \ldots, X_n$. Sums of random variables occur in many situations in insurance and finance, for example, such a sum appears when considering the aggregate claims $X_1 + \ldots + X_m$ of an insurance portfolio during a certain reference period (e.g. a month or a year). From an actuarial point of view, one is often interested in the density function of S and m could be interpreted as the expected number of claims in a specified period. Moreover, the random variable S arises naturally in reliability theory as the lifetime of a system of m identical components. A detailed application in the field of health insurance can be found in [77]. Other applications and examples, are indicated in [47]. [103] considers the estimation of sums of functions of observable and unobservable variables and provides examples in the field of data confidentiality problems and network. In the case of dependence between the random variables X_v, [20] study the problem of capital allocation between risks when the sum of losses is bounded.

■ 6.1.2 Relation to prior work

The most famous nonparametric procedures are those developed by [47] and [90]. [47] constructed a kernel based estimator using the random variables $S_\sigma = \sum_{v=1}^{m} X_{\sigma(v)}$, where $\sigma \in \{\text{combinations of } m \text{ elements } (\sigma(1), \ldots, \sigma(m)) \text{ of } \{1, \ldots, n\}\}$. [90] explored another point of view : the plug-in estimator $\hat{g} = \star_m \hat{f}$, where $\hat{f}$ denotes a standard kernel estimator for f. These two estimators enjoy good mean integrated squared error (global $\mathbb{L}_2$ -risk) and asymptotic normality properties (see, e.g., [47], [90], [2], [3], [87], [91, 92], [43] and [49]). In particular, [90, Theorems 3 and 4]) show that $\hat{g} = \star_m \hat{f}$ attains the parametric rate of convergence $1/n$ under the global $\mathbb{L}_2$-risk (or mean integrated square error MISE) if $m \geq 2$, f four times differentiable with fourth derivative continuous, $\sup_{x \in \mathbb{R}} |f^{(j)}(x)| < \infty$ for $j \in \{0, 1, 2, 3, 4\}$ and the functions f'' and $f^{(4)}$ are integrable. More recently, [75] and mainly [76] proved convergence in law results for kernel or wavelet estimators of the convolution of two densities under rather weak assumptions, to which it is worth comparing ours. In this study we introduce a new kernel estimator $\hat{g}_h$, where h denotes the bandwidth. It is based on Fourier analysis following the spirit of, e.g., [44] and [12] for the standard deconvolution density problem "$f \star k$", where k denotes a known density. In the first part, we establish sharp upper bounds for the pointwise and global $\mathbb{L}_2$-risks of our estimator. Rates of convergence are determined under mild assumptions on f ; we only suppose that f belongs to Sobolev balls with smoothness parameter $\beta > 0$. We show the influence of β on the performances of $\hat{g}_h$ and the best possible bandwidth choice. In particular, for each risks, we determine a constant $\upsilon_m > 0$ such that our estimator attains the parametric rate of convergence $1/n$ for $\beta > \upsilon_m$. When $\beta \in (0, \upsilon_m)$, this result does not hold ; we exhibit a new rate of convergence depending on β. To the best of our knowledge, this phenomenon was never shown before. Let us mention that $\hat{g}_h$ is not adaptive since the best h depends on the unknown β. However, for real data applications, such prior knowledge is not available.

This point motivates a second part devoted to the adaptive estimation of g. Focusing on the global $\mathbb{L}_2$-risk, we develop a "Lepski method" (see, e.g., [61] and [50]) constructing an efficient estimator $\hat{h}$ of the best bandwidth h for $\hat{g}_h$, whatever the smoothness of g. This yields an adaptive estimator $\hat{g}_{\hat{h}}$. We then study its rates of convergence under the global $\mathbb{L}_2$-risk and assuming that f belongs to Sobolev balls with smoothness parameter $\beta > 0$. To be more specific, we prove that, if $m \geq 2$ and $\beta > 1/(2m(m-1))$, $\hat{g}_{\hat{h}}$ attains the parametric rate of convergence $1/n$. When $\beta \in (0, 1/(2m(m-1))]$, its rate of convergence is determined by β. These asymptotic results are very sharp since they are close to those obtained by $\hat{g}_h$ with the best nonadaptive h. The only difference is an extra logarithmic term when $\beta \in (0, 1/(2m(m-1))]$. We illustrate its performances via a simulation study and a real-data example is also provided to illustrate the application of the proposed estimator in a realistic situation.

The rest of the chapter is organized as follows. Section 6.2 introduces our estimation procedure. The results are presented in Section 6.3. Simulated examples as well as a real-data application are provided in Section 6.4. The proofs are postponed to Section 6.6.

6.2 Estimation procedure

6.2.1 Notations

For any $a \in \mathbb{R}$, we set $[a]$ the integer part of a and, for any $(a, b) \in \mathbb{R}^2$, we denote $a \wedge b = \min(a, b)$.

For any $p \geq 1$, we set

$$\mathbb{L}_p(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{R};\ ||f||_p = \left(\int_{-\infty}^{\infty} |f(x)|^p dx \right)^{1/p} \right\}.$$

Assuming that $f \in \mathbb{L}_1(\mathbb{R})$, we define the Fourier transform of f by

$$f^*(t) = \int_{-\infty}^{\infty} f(x) e^{-ixt} dx, \qquad t \in \mathbb{R}. \tag{6.2}$$

Remark 6.2.1. *Let us recall that, if $f \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, the Fourier inverse formula yields that f is the inverse Fourier transform of f^* and can be written as*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixy} f^*(y) dy, \qquad x \in \mathbb{R}. \tag{6.3}$$

Let $L > 0$ and $\alpha > 0$. We define the Sobolev space $\mathcal{S}(\alpha, L)$ by

$$\mathcal{S}(\alpha, L) = \left\{ f \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R}), \qquad \int_{-\infty}^{\infty} (1 + x^2)^{\alpha} |f^*(x)|^2 dx \leq L \right\}.$$

6.2.2 Estimator

We consider the kernel estimator for f :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{v=1}^{n} K\left(\frac{x - X_v}{h} \right),$$

where K denotes the sinus cardinal kernel, i.e.,

$$K(x) = \frac{\sin(\pi x)}{\pi x}, \qquad x \in \mathbb{R}, \tag{6.4}$$

and h is a positive real number (called the "bandwidth").

Remark 6.2.2. *Let us mention that other choices of kernel functions K are possible (see, e.g., [97]). In this study we focus our attention on the sinus cardinal kernel because it is compactly supported in the Fourier domain. Indeed, we have $K^*(t) = \mathbf{1}_{[-\pi,\pi]}(t)$, $t \in \mathbb{R}$.*

In view of (6.2), a natural plug-in estimator for f^* is given by :

$$\hat{f}_h^*(t) = \int_{-\infty}^{\infty} \hat{f}_h(x) e^{-ixt} dx = \tilde{f}^*(t) \mathbf{1}_{[-\pi/h,\pi/h]}(t),$$

where $\tilde{f}^*$ denotes the empirical Fourier transform of f, i.e.,

$$\tilde{f}^*(t) = \frac{1}{n} \sum_{v=1}^{n} e^{-itX_v}. \tag{6.5}$$

Using the Fourier inverse formula (6.3) and the standard convolution equality : $(\star_m f)^*(t) = (f^*(t))^m$, observe that

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\star_m f)^*(t) e^{itx} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} (f^*(t))^m e^{itx} dt. \tag{6.6}$$

Another plug-in in (6.6) yields the following estimator for g :

$$\hat{g}_h(x) = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} (\tilde{f}^*(t))^m e^{itx} dt. \tag{6.7}$$

■ 6.2.3 A selection method for h

Let us now develop a bandwidth selection based on the so-called "Lepski method" and the global $\mathbb{L}_2$-risk of $\hat{g}_h$. First of all, let us define

$$\hat{g}_{h,h'}(x) = \frac{1}{2\pi} \int_{-\pi/(h\vee h')}^{\pi/(h\vee h')} (\tilde{f}^*(t))^m e^{itx} dt, \qquad x \in \mathbb{R},$$

$$V(h) = \kappa \frac{2\pi (\log(n))^m}{h n^m}$$

with $\kappa > 0$, and

$$A(h) = \sup_{h' \in \mathcal{H}_n} \left(\|\hat{g}_{h,h'} - \hat{g}_{h'}\|_2^2 - V(h') \right)_+ .$$

Then we consider the following estimator for the optimal bandwidth h of $\hat{g}_h$ under the global $\mathbb{L}_2$-risk :

$$\hat{h} = \underset{h \in \mathcal{H}_n}{\operatorname{argmin}} (A(h) + V(h)), \tag{6.8}$$

where

$$\mathcal{H}_n = \left\{ h_k, \ \frac{1}{h_k} = k \in \{1, 2, \ldots, n^m - 1, n^m\} \right\}.$$

Using $\hat{h}$ (6.8) and $\hat{g}_h$ (3.3), we consider the plug-in estimator $\hat{g}_{\hat{h}}$ for g.

Further details about the Lepski method can be found in, e.g., [61], [50] and [30].

6.3 Results

6.3.1 Performances of $\hat{g}_h$ under the pointwise $\mathbb{L}_2$-risk

Let us start by the following remark.

Remark 6.3.1. *If f is a density such that $f \in \mathcal{S}(\beta, L)$, then $\sup_{u\in\mathbb{R}}(1+u^2)^\beta |f^*(u)|^2 := B < +\infty$ (since $u \mapsto (1+u^2)^\beta |f^*(u)|^2$ is continuous and integrable over $\mathbb{R}$), and*

$$\int (1+u^2)^{m\beta}|g^*(u)|^2 du \le B^{m-1}\int (1+u^2)^\beta |f^*(u)|^2 du < B^{m-1}L,$$

so that $g \in \mathcal{S}(\alpha, M)$ with $\alpha = m\beta$ and $M = B^{m-1}L$. As m if fixed, we can still consider that α can be small (but then β is even smaller).

Proposition 6.3.1 below investigates the performance of $\hat{g}_h$ under the pointwise $\mathbb{L}_2$-risk.

Proposition 6.3.1. *Consider the model described in Section 6.1 and let $\hat{g}_h$ be given by (3.3).*

Upper bound for the pointwise $\mathbb{L}_2$-risk. Assume that $f^*, g^* \in \mathbb{L}_1(\mathbb{R})$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left((\hat{g}_h(x) - g(x))^2\right) \le C\left(\frac{1}{h^2 n^m} + \frac{1}{n} + \left(\int_{|t|\ge \pi/h} |g^*(t)|dt\right)^2\right).$$

Rate of convergence. Assume that $f^*, g^* \in \mathbb{L}_1(\mathbb{R})$ and $f \in S(\beta, L)$ for $\beta, L > 0$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left((\hat{g}_h(x) - g(x))^2\right) \le C r_n, \tag{6.9}$$

where

Case	**(i)** $m \ge 2,\ \beta \ge 1/m$	**(ii)** $m \ge 2,\ \beta \ge 1/(m(m-1))$	**(iii)** $\beta \in (0, 1/(m(m-1)))$
h	$O(n^{-1/2})$	$O(n^{-m/(2(m\beta+1))})$	$O(n^{-m/(2(m\beta+1))})$
r_n	n^{-1}	n^{-1}	$n^{-m^2\beta/(m\beta+1)}$

In **(i)**, observe that $\hat{g}_h$ is adaptive (since $h = O(n^{-1/2})$ does not depend on β) and attains the parametric rate of convergence $1/n$. The cases **(ii)** and **(iii)** are complementary. In **(ii)**, $\hat{g}_h$ is non-adaptive but still attains $1/n$. In **(iii)**, $1/n$ does not hold and $\hat{g}_h$ is still non-adaptive.

Note that we recover the regularity threshold $1/2$ when $m = 2$, that is given in [76], see his Propositions 4 and 5. But [76] studies convergence in law in $C_0(\mathbb{R})$, the space of bounded continuous functions on $\mathbb{R}$ that vanish at infinity, while we only consider here pointwise risk. On the other hand, we find out that the regularity threshold for the parametric rate is $1/(m(m-1))$ and obtain nonparametric rates below the threshold, which is new.

6.3.2 Performances of $\hat{g}_h$

Proposition 6.3.2 below investigates the performance of $\hat{g}_h$ under the global $\mathbb{L}_2$-risk.

Proposition 6.3.2. *Consider the model described in Section 6.1 and let $\hat{g}_h$ be defined by (3.3).*

Upper bound for the global $\mathbb{L}_2$-risk. Assume that $g^* \in \mathbb{L}_2(\mathbb{R})$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\|\hat{g}_h - g\|_2^2\right) \leq C\left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t|\geq \pi/h} |g^*(t)|^2 dt\right).$$

Rates of convergence. Assume that $f \in S(\beta, L)$ for $\beta, L > 0$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\|\hat{g}_h - g\|_2^2\right) \leq C r_n,$$

where

Case	**(iv)** $m \geq 2,\ \beta \geq 1/(2m)$	**(v)** $m \geq 2,\ \beta \geq 1/(2m(m-1))$	**(vi)** $\beta \in (0, 1/(2m(m-1)))$
h	$O(n^{-1})$	$O(n^{-m/(2m\beta+1)})$	$O(n^{-m/(2m\beta+1)})$
r_n	n^{-1}	n^{-1}	$n^{-2m^2\beta/(2m\beta+1)}$

The proof of Proposition 6.3.2 is based on a suitable decomposition of the global $\mathbb{L}_2$-risk, moments inequality for (6.5) and technical elements related to the Fourier analysis.

As in Proposition 6.3.1, Proposition 6.3.2 shows that, in **(iv)**, the procedure $\hat{g}_h$ is adaptive and attains the parametric rate of convergence $1/n$. The cases **(v)** and **(vi)** are complementary. In **(v)**, $\hat{g}_h$ is non-adaptive but still attains $1/n$. In **(vi)**, the rate of convergence $1/n$ is deteriorated and $\hat{g}_h$ is still non-adaptive.

When $m = 1$ is considered in **(vi)**, the rate of convergence becomes the standard one for the density estimation problem via kernel method, i.e., $r_n = n^{-2\beta/(2\beta+1)}$ (see, e.g., [97]).

Note that the results of Proposition 6.3.2 on the integrated error exhibit rates and bandwidth choices which are different from pointwise ones, under weak conditions. For instance, the parametric rate of convergence $1/n$ for the MISE is also attained by the kernel method of [90] (see [90, Theorems 3 and 4]). However, it is established with more restrictive conditions on f, compared to **(iv)**, i.e., f four times differentiable with fourth derivative continuous, $\sup_{x\in\mathbb{R}} |f^{(j)}(x)| < \infty$ for $j \in \{0, 1, 2, 3, 4\}$ and the functions f'' and $f^{(4)}$ are integrable. Nevertheless, similar results, under weak conditions, may be deduced from the main Theorem in [76].

Naturally, the performances of $\hat{g}_h$ deeply depends on h. We see in Propositions 6.3.1 and 6.3.2 that its optimal value depends on the unknown smoothness parameter β. In order to estimate it efficiently via the observations, a selection method was proposed in Section 6.2.3. The next section is devoted to the performances of the resulting adaptive estimator $\hat{g}_{\hat{h}}$.

6.3.3 Performances of $\hat{g}_{\hat{h}}$

Theorem 6.3.1 below explores the performances of $\hat{g}_{\hat{h}}$ under the global $\mathbb{L}_2$-risk.

Theorem 6.3.1. *Consider the model described in Section 6.1. Let $\hat{g}_{\hat{h}}$ be defined by (3.3) with $h = \hat{h}$ from (6.8).*

Upper bound for the global $\mathbb{L}_2$-risk. Assume that $g^* \in \mathbb{L}_2(\mathbb{R})$. Then there exist two constants κ and $C > 0$ such that

$$\mathbb{E}(\|\hat{g}_{\hat{h}} - g\|_2^2) \leq C \left(\inf_{h \in \mathcal{H}_n} \left(\frac{1}{hn^m} + \int_{|t| \geq \pi/h} |g^*(t)|^2 + V(h) \right) + \frac{1}{n} \right).$$

Rates of convergence. Assume that $f \in S(\beta, L)$ for $\beta, L > 0$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\|\hat{g}_{\hat{h}} - g\|_2^2\right) \leq C r_n,$$

where

Case	**(vii)** $m \geq 2,\ \beta > 1/(2m(m-1))$	**(viii)** $\beta \in (0, 1/(2m(m-1))]$
r_n	n^{-1}	$\left(\frac{\log(n)}{n}\right)^{2m^2\beta/(2m\beta+1)}$

Theorem 6.3.1 shows that our adaptive procedure $\hat{g}_{\hat{h}}$ attains similar rates of convergence to those attained by $\hat{g}_h$ with the optimal non-adaptive h. The only difference is a logarithmic loss for $\beta \in (0, 1/(2m(m-1))]$.

6.4 Numerical experiments

For simplicity in the following simulation study, we have added the superscript m on g to denote the m-th convolution power. We consider the problem of estimating the density g_m of the sum of a fixed number m of *i.i.d.* random variables, emphasizing the case $m = 2$. We demonstrate the usefulness of the adaptive kernel estimator $\hat{g}_{m,\hat{h}}$on both simulated and real data examples. The numerical experiments have been carried out using Matlab.

6.4.1 Computational aspects

A computationally-efficient procedure for kernel density estimation has been given by [95, 96], with extensions by [55]. The main idea of this approach is to express kernel estimator as a convolution of the data with the kernel and use a Fast Fourier Transform (FFT) to perform the convolution (see e.g. [96] Section 3.5). The estimator proposed in this chapter is based on the same idea, therefore FFT-based method are well suited for its implementation. The resulting estimator $\hat{g}_{m,h}$ is simple to implement and fast which allows us to perform the selection procedure in a reasonable time for various sample sizes.

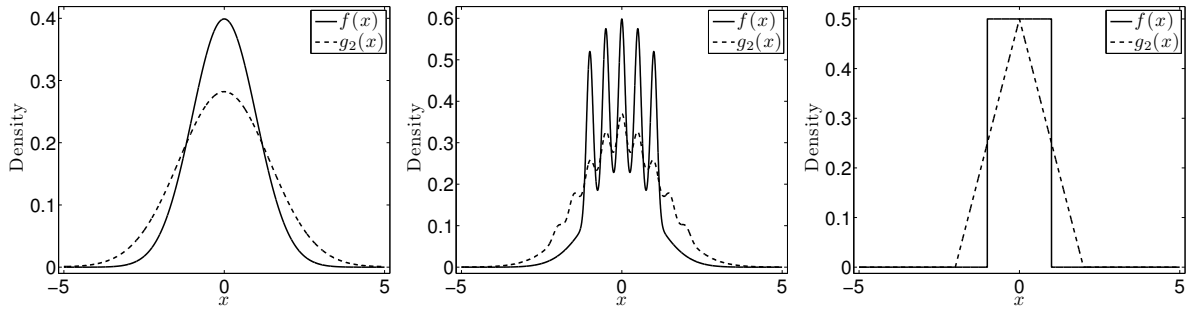
The original observations were generated from an *i.i.d.* sample of random variables $X_1, \ldots, X_n$. Three test functions, representing different degrees of smoothness were used (see Figure 6.1),

(a) the standard normal distribution $\mathcal{N}(0, 1)$,

(b) the Claw distribution $\frac{1}{2}\mathcal{N}(0, 1) + \sum_{l=0}^{4} \mathcal{N}\left(l/2 - 1, (1/2)^2\right)$,

(c) the uniform distribution $\mathcal{U}(-1, 1)$.

For numerical implementation, we consider an interval $[a, b]$ that cover the range of the data and the density estimates were evaluated at $M = 2^r$ equally spaced points $t_i = a + (b-a)/M$, $i = 0, 1, \ldots, M-1$, between a and b, with $r = 8$, $b = -a = 5$ and



(a) Gaussian (b) Claw (c) Uniform

FIGURE 6.1 – Test densities f (solid) and g_2 (dashed) (the 2-fold convolution power of f).

M is the number of discretization points. The constant κ is taken equal to 1 and the normalized sinc Kernel (6.4) were used throughout all experiments. Our results are based on $N = 500$ simulated data sets and the MISE was approximated by the average of the Integrated Squared Error (ISE) over the N replicates.

6.4.2 Bandwidth selection procedure

We have considered the case where $m = 2$ to illustrate the finite sample behavior of the adaptive estimator $\hat{g}_{m,\hat{h}}$, constructed with the data-driven bandwidth selection procedure described in the Section 6.2. Note that, although we focused on $m = 2$ the approach can easily handle other values of m.

In order to reduce, the computational cost of the selection procedure and improve the numerical efficiency of the resulting estimator, we have reduced the cardinality of the finite set of bandwidth $\mathcal{H}_n$. Indeed, a proper choice of an initial grid $\mathcal{H}_n$ allows to reduce considerably the computational time involved in the bandwidth selection procedure and consequently in the estimation of $\hat{g}_{m,\hat{h}}$. Thus, in each case, the grid of h values that we have considered consisted of 100 values from $0.2\hat{h}_0$ to $1.2\hat{h}_0$ where $\hat{h}_0 = n^{-m/5}$ denotes a pilot bandwidth generally selected from a reference rule, like Silverman's rule of thumb (see e.g. [96]). Thus, in practice, for each value of $h \in \mathcal{H}_n$, the function $A(h) + V(h)$ has been computed from a single simulation and then, we have minimized (6.8) numerically over the grid $\mathcal{H}_n$. This step can be easily and quickly computed (thanks to the FFT-algorithm). The results are depicted in Figures 6.2 and 6.3 for n-samples of sizes 1000 et 2000. Figures 6.2(b) 6.3(b) also contain a plot of the function $A(h) + V(h)$ against the smoothing parameter h and a vertical shift of the curve MISE(h) is also overlayed for visualization purposes. For each density, it is clear from the figures, that the value of $\hat{h}$ is the unambiguous minimizer of $A(h) + V(h)$. One notes that selection is already very effective for $n = 1000$, $\hat{h}$ provides a decent approximation, fairly close to h_{MISE} and even closer when the sample size n grows (i.e. $n = 2000$), for all test densities. When $n = 1000$, for the Gaussian density, the bandwidth which minimizes MISE(h) in this case is $h_{\text{MISE}} = 0.12520$ and $\hat{h} = 0.12543$. In this case, for the Claw and the uniform densities, we obtained respectively $h_{\text{MISE}} = 0.12517, \hat{h} = 0.12540$ and $h_{\text{MISE}} = 0.12547, \hat{h} = 0.12542$. Note that, the selection procedure also give good results on smaller sample sizes (e.g. for $n = 100$ and the Gaussian density, $h_{\text{MISE}} = 0.3980$ and $\hat{h} = 0.3888$). Therefore, without any prior smoothness knowledge on the unknown density, our adaptive estimator is very effective to estimate each of the three densities.

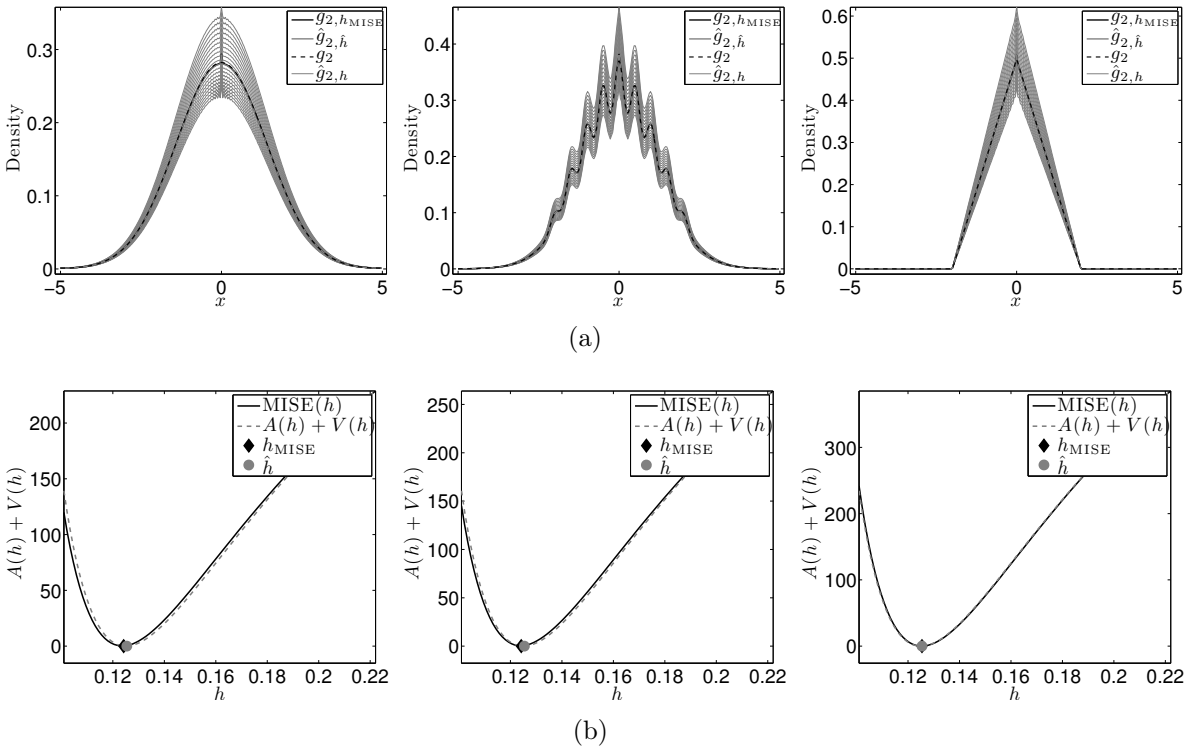


FIGURE 6.2 – (a) : True density (dotted), density estimates (gray) and sample of 20 estimates (thin gray) out of 100 proposed to the selection algorithm obtained with a sample of $n = 1000$ data. (b) : Graph of the the function $A(h) + V(h)$ against the smoothing parameter h and (shifted) MISE(h). The gray diamond represents the global minimizer of $A(h)+V(h)$ and the gray circle represents the global minimizer of MISE(h).

TABLE 6.1 – 10000×MISE values from 500 replications. From top to bottom Gaussian, Claw and Uniform densities for $m = 2$. The numbers in parenthesis show the performance of [90]

n	500		1000		2000	
Gaussian	0.0186	(0.0345)	0.0065	(0.0188)	0.0028	(0.0101)
Claw	0.0291	(0.1535)	0.0114	(0.1261)	0.0056	(0.1114)
Uniform	0.0250	(0.1301)	0.0104	(0.0867)	0.0072	(0.0589)

Finally, we have compared the performance of our adaptive estimator to those of a different kernel-based estimator presented in [90], i.e. $\hat{g} = \star_m \hat{f}$, where $\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$, is a standard kernel estimator and K is a kernel function. All experiments were conducted using a Gaussian kernel and we have been focused on a global bandwidth selector : the *rule of thumb* (ROT) bandwidth selector (see e.g. [96]). Thus, the optimal bandwidth is given by $h_{\mathrm{ROT}} = 1.06\min(\hat{\sigma}, Q/1.34)n^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation and Q is the interquartile range. TABLE 6.1 summarizes the results. It shows in particular that our adaptive estimator systematically outperforms the estimator [90]. As expected and predicted by our theoretical findings, the performance gets better as the sample sizes increases.

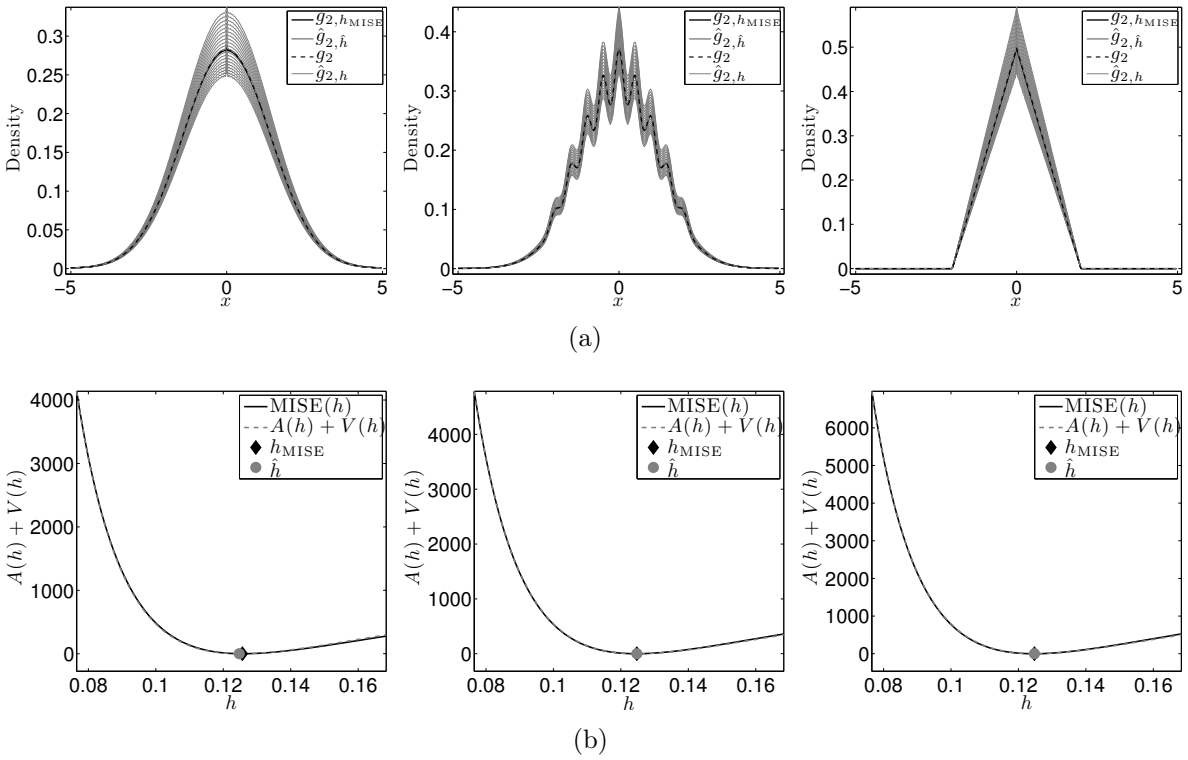


FIGURE 6.3 – (a) : True density (dotted), density estimates (gray) and sample of 20 estimates (thin gray) out of 100 proposed to the selection algorithm obtained with a sample of $n = 2000$ data. (b) : Graph of the function $A(h) + V(h)$ against the smoothing parameter h and (shifted) MISE(h). The black diamond represents the global minimizer of $A(h)+V(h)$ and the gray circle represents the true MISE-minimizing bandwidth h_{MISE}.

Real data example

In this section a real data application of the bandwidth selection method is given. In insurance, a typical example is the sum of insurance claims, where $(X_v)_{v=1,\dots,n}$ are individual insurance claims and $S = \sum_{v=1}^{m} X_v$ is the sum of m claims and m could be interpreted as the expected number of claims in a specified period (e.g. one month). In this illustration, we consider the hospital data example, reported and analyzed in [47] and we discuss the case $m = 2$. The data in Table 6.2 consist of measurements of the 1989 total charges for 33 patients at a Wisconsin Hospital

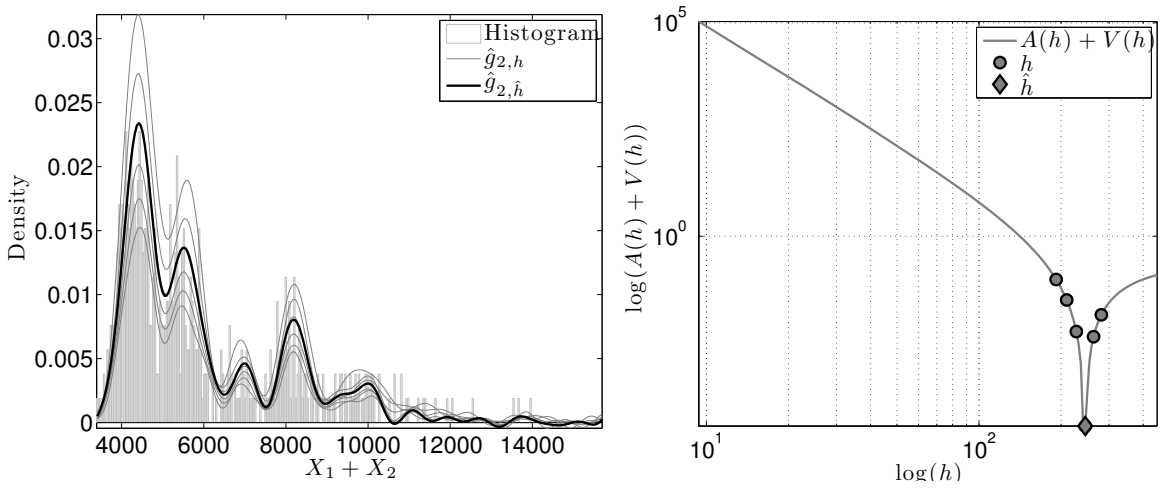
The bandwidth was also chosen by grid search minimization, over an equally spaced grid of 100 values from $h_{\min} = 9$ to $h_{\max} = 450$. Figure 5.4(a) depicts density estimate and show histogram of the sum of claims for $m = 2$. Figure 6.4(b) contain a plot of the function $A(h)+V(h)$ against the smoothing parameter h and the red circle represents the global minimizer $\hat{h}$ of $A(h) + V(h)$. The density estimates depicted by the gray lines in Figure 5.4(a) are based on the bandwidths marked by the gray circles in Figure 6.4(b).

■ 6.5 Conclusion and perspectives

We have constructed a new adaptive estimator $\hat{g}_{\hat{h}}$ for g using kernel methods, Fourier analysis and the Lepski method. Theorems 6.3.2 and 6.3.1 show the good performances of $\hat{g}_{\hat{h}}$ in terms of rates of convergence under the global $\mathbb{L}_2$-risk over $S(\beta, L)$. The agreement

TABLE 6.2 – 1989 total hospital charges (in dollars) for 33 females aged 30-49 hospitalized for circulatory disorders from a Wisconsin Hospital (see [47]).

2337	2179	2348	4765	2088	2872	1924	2294	2182	2138	1765
2467	3609	2141	1850	3191	3020	2473	1898	7787	6169	1802
2011	2270	3425	3558	2315	1642	5878	2101	2242	5746	3041



(a) (b)

FIGURE 6.4 – (a) Our estimator $\hat{g}_{2,\hat{h}}$ (black) for the sum of two claims and a sample of 5 estimates $\hat{g}_{2,h}$ (thin gray) out of 100 used to the selection algorithm. (b) : Graph of $A(h) + V(h)$ against the smoothing parameter h (in a log-log scale). The black diamond represents the global minimizer $\hat{h}$ of $A(h) + V(h)$ and the gray circles represent the values of h corresponding to the $\hat{g}_h$ plot in (a)

of our simulations with our theoretical findings show that our estimator is quite effective on both simulated and real data sets. Possible perspectives of this work are to

— remove the extra logarithmic term in the case $\beta \in (0, 1/(2m(m-1))]$ via another adaptive estimator,

— extend our estimation procedure to another problem as the one of [104], i.e., estimate the density g of $X_1 + Y_1$ from n *i.i.d.* observations $X_1, \ldots, X_n$ with density f_1 and n *i.i.d.* observations $Y_1, \ldots, Y_n$ with density f_2,

All these points need further investigations that we leave for future works.

6.6 Proofs

In the following, the quantity C denotes a generic constant that does not depend on n. Its value may change from one term to another (it can depend on the fixed m).

6.6.1 Intermediary results

Proposition 6.6.1. *Let f^* be (6.2) and $\tilde{f}^*$ be (6.5). For any positive integer v, there exists a constant $C > 0$ such that*

$$\sup_{t\in\mathbb{R}} \mathbb{E}\left(|\tilde{f}^*(t) - f^*(t)|^{2v}\right) \leq C\frac{1}{n^v}.$$

Proof of Proposition 6.6.1. First of all, let us recall the Rosenthal inequality (see [88]).

Lemma 6.6.1 (Rosenthal's inequality)**.** *Let n be a positive integer, $p \geq 2$ and $U_1, \ldots, U_n$ be n zero mean independent random variables such that $\sup_{v\in\{1,\ldots,n\}} \mathbb{E}(|U_v|^p) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\sum_{v=1}^{n} U_v\right|^p\right) \leq C\left(\sum_{v=1}^{n}\mathbb{E}\left(|U_v|^p\right) + \left(\sum_{v=1}^{n}\mathbb{E}\left(U_v^2\right)\right)^{p/2}\right).$$

Let $t \in \mathbb{R}$. We have

$$\tilde{f}^*(t) - f^*(t) = \frac{1}{n}\sum_{v=1}^{n} U_v(t), \qquad U_v(t) = e^{-itX_v} - \mathbb{E}(e^{-itX_1}).$$

Note that $U_1(t), \ldots, U_n(t)$ are *i.i.d.* with $\mathbb{E}(U_1(t)) = 0$, $|U_1(t)| \leq 1$ and $\mathbb{E}(|U_1(t)|^2) \leq 1$. For any positive integer v, the Rosenthal inequality yields the existence of a constant $C > 0$ such that

$$\mathbb{E}\left(|\tilde{f}^*(t) - f^*(t)|^{2v}\right) = \frac{1}{n^{2v}}\mathbb{E}\left(\left|\sum_{v=1}^{n} U_v(t)\right|^{2v}\right) \leq C\frac{1}{n^v}.$$

Proposition 6.6.1 is proved.

□

Lemma 6.6.2. *Let $(u, v) \in \mathbb{C}^2$ such that $|u| \leq 1$ and $|v| \leq 1$. Then, for any integer $m \geq 1$, we have*

$$|u^m - v^m| \leq |u - v|^m + D_m|u - v||v|,$$

with $D_m = (3^m - 2^m - 1)/2$.

Proof of Lemma 6.6.2. For $m = 1$, the desired inequality is obviously satisfied with $D_m = 0$. Let us now investigate the case $m \geq 2$. The binomial theorem yields

$$\begin{aligned} u^m - v^m &= \sum_{k=0}^{m-1}\binom{m}{k}v^k(u - v)^{m-k} \\ &= (u - v)^m + (u - v)v\sum_{k=0}^{m-2}\binom{m}{k+1}v^k(u - v)^{m-2-k}. \end{aligned}$$

The triangular inequality and $|u| \leq 1$ and $|v| \leq 1$ give

$$|u^m - v^m| \leq |u - v|^m + D_m|u - v||v|,$$

with

$$D_m = 2^{m-2}\sum_{k=0}^{m-2}\binom{m}{k+1}2^{-k} = \frac{1}{2}(3^m - 2^m - 1).$$

Lemma 6.6.2 is proved.

□

6.6.2 Proof of the main results

Proof of Proposition 6.3.1

Upper bound for the pointwise L_2 risk. Set

$$\overline{g}_h(x) = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} g^*(t)e^{itx}dt = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} (f^*(t))^m e^{itx}dt, \qquad x \in \mathbb{R}. \tag{6.10}$$

Using the elementary inequality $(x+y)^2 \le 2(x^2+y^2)$, $(x,y) \in \mathbb{R}^2$, we obtain

$$\mathbb{E}\left((\hat{g}_h(x) - g(x))^2\right) \le 2(P+Q), \tag{6.11}$$

where

$$P = \mathbb{E}\left((\hat{g}_h(x) - \bar{g}_h(x))^2\right), \qquad Q = (\bar{g}_h(x) - g(x))^2.$$

Upper bound for Q. The Fourier inverse formula yields

$$Q = \left(\frac{1}{2\pi}\int_{|t|\ge\pi/h} g^*(t)e^{itx}dt\right)^2 \le \left(\frac{1}{2\pi}\int_{|t|\ge\pi/h} |g^*(t)|dt\right)^2. \tag{6.12}$$

Upper bound for Q. We have

$$\begin{aligned} P &= \mathbb{E}\left(\left(\frac{1}{2\pi}\int_{-\pi/h}^{\pi/h}((\tilde{f}^*(t))^m - (f^*(t))^m)e^{itx}dt\right)^2\right) \\ &\le \mathbb{E}\left(\left(\frac{1}{2\pi}\int_{-\pi/h}^{\pi/h}|(\tilde{f}^*(t))^m - (f^*(t))^m|dt\right)^2\right) \\ &\le \left(\frac{1}{2\pi}\int_{-\pi/h}^{\pi/h}\left(\mathbb{E}\left(|(\tilde{f}^*(t))^m - (f^*(t))^m|^2\right)\right)^{1/2}dt\right)^2, \end{aligned} \tag{6.13}$$

where the last line follows from the Fubini theorem (write the squared integral as a multiple integral) and the Cauchy-Schwarz inequality.
It follows from the inequalities $|\tilde{f}^*(t)| \le 1$, $|f^*(t)| \le ||f||_1 = 1$, Lemma 6.6.2 and the elementary inequality $(x+y)^2 \le 2(x^2+y^2)$, $(x,y) \in \mathbb{R}^2$, that

$$|(\tilde{f}^*(t))^m - (f^*(t))^m|^2 \le 2\left(|\tilde{f}^*(t) - f^*(t)|^{2m} + D_m^2|\tilde{f}^*(t) - f^*(t)|^2|f^*(t)|^2\right),$$

with $D_m = (3^m - 2^m - 1)/2$.
Proposition 6.6.1 implies the existence of a constant $C > 0$ such that

$$\begin{aligned} &\mathbb{E}\left(|(\tilde{f}^*(t))^m - (f^*(t))^m|^2\right) \\ &\le 2\left(\mathbb{E}\left(|\tilde{f}^*(t) - f^*(t)|^{2m}\right) + D_m^2|f^*(t)|^2\mathbb{E}\left(|\tilde{f}^*(t) - f^*(t)|^2\right)\right) \\ &\le C\left(\frac{1}{n^m} + \frac{1}{n}|f^*(t)|^2\right). \end{aligned} \tag{6.14}$$

By (6.13), (6.14), the elementary inequalities $\sqrt{|x+y|} \le \sqrt{|x|}+\sqrt{|y|}$, and $(x+y)^2 \le 2(x^2+y^2)$, $(x,y) \in \mathbb{R}^2$, and $\int_{-\pi/h}^{\pi/h} |f^*(t)|dt \le ||f^*||_1$, we have

$$P \le C \left(\int_{-\pi/h}^{\pi/h} \left(\frac{1}{n^m} + \frac{1}{n}|f^*(t)|^2 \right)^{1/2} dt \right)^2$$
$$\le C \left(\frac{1}{n^m h^2} + \frac{1}{n} \left(\int_{-\pi/h}^{\pi/h} |f^*(t)|dt \right)^2 \right) \le C \left(\frac{1}{n^m h^2} + \frac{1}{n} \right). \tag{6.15}$$

Combining (6.11), (6.12) and (6.15), we obtain

$$\mathbb{E}\left((\hat{g}_h(x) - g(x))^2 \right) \le C \left(\frac{1}{h^2 n^m} + \frac{1}{n} + \left(\int_{|t| \ge \pi/h} |g^*(t)|dt \right)^2 \right). \tag{6.16}$$

Rates of convergence. Since $\int_{-\infty}^{\infty} |f^*(t)|^2 (1+t^2)^\beta dt \le L$, recalling that this implies that $|f^*(t)|^2(1+t^2)^\beta \le B$, we have

$$\left(\int_{|t| \ge \pi/h} |g^*(t)|dt \right)^2 = \left(\int_{|t| \ge \pi/h} |f^*(t)|^m dt \right)^2$$
$$= \left(\int_{|t| \ge \pi/h} \left(|f^*(t)|(1+t^2)^{\beta/2} \right)^2 \left(|f^*(t)|(1+t^2)^{\beta/2} \right)^{m-2} (1+t^2)^{-m\beta/2} dt \right)^2$$
$$\le (1+(\pi/h)^2)^{-m\beta} B^{m-2} L^2 \le C h^{2m\beta}. \tag{6.17}$$

(i) $m \ge 2$, $\beta \ge 1/m$ *:* The inequalities (6.16) and (6.17) and $h = O(n^{-1/2})$ give

$$\mathbb{E}\left((\hat{g}_h(x) - g(x))^2 \right) \le C \left(\frac{1}{n^{m-1}} + \frac{1}{n} + n^{-m\beta} \right) \le C \frac{1}{n}.$$

(ii) $m \ge 2$, $\beta \ge 1/(m(m-1))$ *:* It follows from (6.16) and (6.17) and $h = O(n^{-m/(2(m\beta+1))})$ that

$$\mathbb{E}\left((\hat{g}_h(x) - g(x))^2 \right) \le C \left(n^{-m^2\beta/(m\beta+1)} + \frac{1}{n} + n^{-m^2\beta/(m\beta+1)} \right) \le C \frac{1}{n}.$$

(iii) $\beta \in (0, 1/(m(m-1)))$ *:* Similarly to the previous case, we have

$$\mathbb{E}\left((\hat{g}_h(x) - g(x))^2 \right) \le C \left(n^{-m^2\beta/(m\beta+1)} + \frac{1}{n} + n^{-m^2\beta/(m\beta+1)} \right) \le C n^{-m^2\beta/(m\beta+1)}.$$

The proof of Proposition 6.3.1 is complete. □

Proof of Proposition 6.3.2.

Upper bound for the global $\mathbb{L}_2$-risk. Recall that $\bar{g}_h$ is defined in equation (6.10). The elementary inequality $(x+y)^2 \le 2(x^2+y^2)$, $(x,y) \in \mathbb{R}^2$, yields

$$\mathbb{E}\left(\|\hat{g}_h - g\|_2^2 \right) \le 2(R+S), \tag{6.18}$$

where

$$R = \mathbb{E}\left(\|\hat{g}_h - \bar{g}_h\|_2^2 \right), \qquad S = \|\bar{g}_h - g\|_2^2.$$

Upper bound for R. The Parseval theorem gives

$$R = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} \mathbb{E}\left(|(\tilde{f}^*(t))^m - (f^*(t))^m|^2 \right) dt \tag{6.19}$$

Owing to the Parseval theorem, we have

$$\int_{-\pi/h}^{\pi/h} |f^*(t)|^2 dt \leq ||f^*||_2^2 = 2\pi ||f||_2^2 \leq C. \tag{6.20}$$

It follows from (6.19), (6.14) and (6.20) that

$$R \leq C \left(\frac{1}{hn^m} + \frac{1}{n} \int_{-\pi/h}^{\pi/h} |f^*(t)|^2 dt \right) \leq C \left(\frac{1}{hn^m} + \frac{1}{n} \right). \tag{6.21}$$

Upper bound for S. Using the Parseval theorem, we get

$$S = \int_{|t| \geq \pi/h} |g^*(t)|^2 dt. \tag{6.22}$$

Combining (6.18), (6.21) and (6.22), we obtain

$$\mathbb{E}\left(\|\hat{g}_h - g\|_2^2 \right) \leq C \left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t| \geq \pi/h} |g^*(t)|^2 dt \right). \tag{6.23}$$

Rates of convergence. Let us recall that, since $f \in S(\beta, L)$, then $g \in S(\alpha, M)$ with $\alpha = m\beta$ and $M = B^{m-1}L$ (see Remark 6.3.1). Therefore

$$\int_{|t| \geq \pi/h} |g^*(t)|^2 dt \leq \left(1 + (\pi/h)^2\right)^{-m\beta} \int_{|t| \geq \pi/h} (1 + t^2)^{m\beta} |g^*(t)|^2 dt \leq CLh^{2m\beta}. \tag{6.24}$$

(iv) $m \geq 2$ and $\beta \geq 1/(2m)$: It follows from (6.23) and (6.24) with $h = O(n^{-1})$ that

$$\mathbb{E}\left(\|\hat{g}_h - g\|_2^2 \right) \leq C \left(\frac{1}{n^{m-1}} + \frac{1}{n} + n^{-2m\beta} \right) \leq C \frac{1}{n}.$$

(v) $m \geq 2$ and $\beta \geq 1/(2m(m-1))$: The choice $h = O(n^{-m/(2m\beta+1)})$ in the bounds (6.23) and (6.24) yields

$$\mathbb{E}\left(\|\hat{g}_h - g\|_2^2 \right) \leq C \left(n^{-2m^2\beta/(2m\beta+1)} + \frac{1}{n} + n^{-2m^2\beta/(2m\beta+1)} \right) \leq C \frac{1}{n}.$$

(vi) $\beta \in (0, 1/(2m(m-1)))$: Proceeding as for the previous point, we obtain

$$\mathbb{E}\left(\|\hat{g}_h - g\|_2^2 \right) \leq C \left(n^{-2m^2\beta/(2m\beta+1)} + \frac{1}{n} + n^{-2m^2\beta/(2m\beta+1)} \right) \leq C n^{-2m^2\beta/(2m\beta+1)}.$$

This ends the proof of Proposition 6.3.2.

□

Proof of Theorem 6.3.1.

Upper bound for the global $\mathbb{L}_2$-risk. Let $h \in \mathcal{H}_n$ be fixed. The Minkowski inequality and the elementary inequality $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$, $(x,y,z) \in \mathbb{R}^3$, yield

$$\mathbb{E}\left(\|\hat{g}_{\hat{h}} - g\|_2^2\right) \leq 3(\mathbb{E}\left(\|\hat{g}_{\hat{h}} - \hat{g}_{h,\hat{h}}\|_2^2\right) + \mathbb{E}\left(\|\hat{g}_{h,\hat{h}} - \hat{g}_h\|_2^2\right) + \mathbb{E}\left(\|\hat{g}_h - g\|_2^2\right)).$$

By definition of $A(h)$, we have

$$\mathbb{E}\left(\|\hat{g}_{\hat{h}} - \hat{g}_{h,\hat{h}}\|_2^2\right) \leq \mathbb{E}(A(h)) + \mathbb{E}(V(\hat{h})),$$

the definition of $A(\hat{h})$ yields

$$\mathbb{E}\left(\|\hat{g}_{h,\hat{h}} - \hat{g}_h\|_2^2\right) \leq \mathbb{E}(A(\hat{h})) + V(h)$$

and Proposition 6.3.2 gives

$$\mathbb{E}\left(\|\hat{g}_h - g\|_2^2\right) \leq C\left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t|\geq \pi/h} |g^*(t)|^2 dt\right).$$

Therefore, owing to the definition of $\hat{h}$, we have

$$\begin{aligned}
&\mathbb{E}\left(\|\hat{g}_{\hat{h}} - g\|_2^2\right) \\
&\leq 3(\mathbb{E}(A(h)) + \mathbb{E}(V(\hat{h}) + A(\hat{h})) + V(h)) + C\left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t|\geq \pi/h} |g^*(t)|^2 dt\right) \\
&\leq 6(\mathbb{E}(A(h)) + V(h)) + C\left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t|\geq \pi/h} |g^*(t)|^2 dt\right), \qquad (6.25)
\end{aligned}$$

where

$$\mathbb{E}(A(h)) = \mathbb{E}\left(\sup_{h' \in \mathcal{H}_n} \left(\|\hat{g}_{h,h'} - \hat{g}_{h'}\|_2^2 - V(h')\right)_+\right). \qquad (6.26)$$

Upper bound for $\mathbb{E}(A(h))$. Let us introduce the following functions :

$$\overline{g}_h(x) = \frac{1}{2\pi}\int_{-\pi/h}^{\pi/h} g^*(t)e^{itx}dt, \qquad x \in \mathbb{R}$$

and

$$\overline{g}_{h,h'}(x) = \frac{1}{2\pi}\int_{-\pi(1/h\wedge 1/h')}^{\pi(1/h\wedge 1/h')} g^*(t)e^{itx}dt, \qquad x \in \mathbb{R}.$$

Observe that

$$\hat{g}_{h'} - \hat{g}_{h,h'} = \hat{g}_{h'} - \overline{g}_{h'} - (\hat{g}_{h,h'} - \overline{g}_{h,h'}) + \overline{g}_{h'} - \overline{g}_{h,h'}.$$

Using again the Minkowski inequality and the elementary inequality $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$, $(x,y,z) \in \mathbb{R}^3$, we obtain

$$||\hat{g}_{h'} - \hat{g}_{h,h'}||_2^2 \leq 3\left(||\hat{g}_{h'} - \overline{g}_{h'}||_2^2 + ||\hat{g}_{h,h'} - \overline{g}_{h,h'}||_2^2 + ||\overline{g}_{h'} - \overline{g}_{h,h'}||_2^2\right).$$

Set $D = [-\pi/h', -\pi(1/h \wedge 1/h'))] \cup [\pi(1/h \wedge 1/h'), \pi/h']$. The Parseval theorem yields

$$\|\overline{g}_{h'} - \overline{g}_{h,h'}\|_2^2 = \frac{1}{2\pi}\int_D |g^*(t)|^2 dt \leq \int_{|t|\geq \pi/h} |g^*(t)|^2 dt = \|\overline{g}_h - g\|_2^2.$$

In the same way, we prove that

$$\|\hat{g}_{h,h'} - \overline{g}_{h,h'}\|_2^2 \leq \|\hat{g}_{h'} - \overline{g}_{h'}\|_2^2.$$

Then

$$\|\hat{g}_{h'} - \hat{g}_{h,h'}\|_2^2 \leq 6\|\hat{g}_{h'} - \overline{g}_{h'}\|_2^2 + 3\|g - \overline{g}_h\|_2^2. \tag{6.27}$$

It follows from (6.26), (6.27) and the Parseval theorem that

$$\mathbb{E}(A(h)) \leq 6U + 3\int_{|t|\geq \pi/h} |g^*(t)|^2 dt, \tag{6.28}$$

where

$$U = \mathbb{E}\left(\sup_{h'\in\mathcal{H}_n} (\|\hat{g}_{h'} - \overline{g}_{h'}\|_2^2 - V(h')/6)_+\right).$$

Upper bound for U. Let us consider the random event

$$\Omega(u) = \left\{|\tilde{f}^*(u) - f^*(u)| \leq \tau\sqrt{\log(n)/n}\right\}$$

Let us set

$$v_n = \frac{h'}{2\pi}V(h') = \kappa\frac{(\log(n))^m}{n^m}.$$

Owing to the Parseval theorem and Lemma 6.6.2, we have

$$\begin{aligned}
&\|\hat{g}_{h'} - \overline{g}_{h'}\|_2^2 - V(h')/6 = \frac{1}{2\pi}\int_{-\pi/h'}^{\pi/h'} |(\tilde{f}^*(t))^m - (f^*(t))^m|^2 dt - V(h')/6 \\
&\leq \frac{1}{\pi}\Bigg(\int_{-\pi/h'}^{\pi/h'} \left(|\tilde{f}^*(t) - f^*(t)|^{2m}\mathbf{1}_{\Omega(t)} - v_n/12\right) dt + \int_{-\pi/h'}^{\pi/h'} |\tilde{f}^*(t) - f^*(t)|^{2m}\mathbf{1}_{\Omega^c(t)} dt \\
&+ D_m^2\int_{-\infty}^{\infty} |\tilde{f}^*(t) - f^*(t)|^2 |f^*(t)|^2 dt\Bigg).
\end{aligned}$$

Therefore

$$U \leq C(R + S + T), \tag{6.29}$$

where

$$R = \sum_{k=1}^{n^m} \mathbb{E}\left(\left(\int_{-\pi/h_k}^{\pi/h_k} \left(|\tilde{f}^*(t) - f^*(t)|^{2m}\mathbf{1}_{\Omega(t)} - v_n/12\right) dt\right)_+\right),$$

$$S = \sum_{k=1}^{n^m} \int_{-\pi/h_k}^{\pi/h_k} \mathbb{E}\left(|\tilde{f}^*(t) - f^*(t)|^{2m}\mathbf{1}_{\Omega^c}(t)\right) dt$$

and

$$T = \int_{-\infty}^{\infty} \mathbb{E}\left(|\tilde{f}^*(t) - f^*(t)|^2\right) |f^*(t)|^2 dt.$$

Evaluation of R. On $\Omega(t)$, we have

$$|\tilde{f}^*(t) - f^*(t)| \leq \tau\sqrt{\frac{\log(n)}{n}}.$$

Therefore, taking τ large enough, we have

$$\int_{-\pi/h_k}^{\pi/h_k} \left(|\tilde{f}^*(t) - f^*(t)|^{2m}\mathbf{1}_{\Omega(t)} - v_n/12\right) dt \leq 0.$$

Hence

$$R = 0. \tag{6.30}$$

Upper bound for S. Following Neumann (1997), we apply Bernstein Inequality and get that

$$\begin{aligned}\mathbb{P}(\Omega(t)^c) &= \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}(e^{-itX_k} - \mathbb{E}(e^{-itX_k}))\right| \geq \tau\sqrt{\log(n)/n}\right) \\ &\leq 2(\exp(-\tau^2\log(n)/4) + \exp(-3\tau\sqrt{n\log(n)}/4)).\end{aligned}$$

Thus, taking $\tau \geq 2 + p$ implies

$$\mathbb{P}(\Omega^c(t)) \leq Cn^{-p}. \tag{6.31}$$

Applying the Cauchy-Schwarz inequality, Lemma 6.6.1 with $v = 2m$, (6.31) with $p = m + 2$ and $1/h_k \leq n^m$, we obtain

$$\begin{aligned}S &\leq \sum_{k=1}^{n^m}\int_{-\pi/h_k}^{\pi/h_k}\left(\mathbb{E}(|\tilde{f}^*(t) - f^*(t)|^{4m})\right)^{1/2}(\mathbb{P}(\Omega^c(t)))^{1/2}\,dt \\ &\leq C\frac{1}{n^{m+p/2}}\sum_{k=1}^{n^m}\frac{1}{h_k} \leq C\frac{1}{n^{(p-m)/2}} = C\frac{1}{n}.\end{aligned} \tag{6.32}$$

Note that we get $\kappa \geq 4 + m$ for the constant in the definition of $V(h)$.

Upper bound for T. Using Lemma 6.6.1 with $v = 1$ and applying the Parseval theorem, we get

$$T \leq C\frac{1}{n}||f^*||_2^2 \leq C\frac{1}{n}||f||_2^2 \leq C\frac{1}{n}. \tag{6.33}$$

Putting (6.29), (6.30), (6.32) and (6.33) together, we have

$$U \leq C\frac{1}{n}. \tag{6.34}$$

Combining (6.28) and (6.34), we get

$$\mathbb{E}(A(h)) \leq C\left(\frac{1}{n} + \int_{|t|\geq\pi/h}|g^*(t)|^2dt\right). \tag{6.35}$$

Finally, from (6.25) and (6.35), we obtain

$$\mathbb{E}(\|\hat{g}_{\hat{h}} - g\|_2^2) \leq C\left(\frac{1}{hn^m} + \int_{|t|\geq\pi/h}|g^*(t)|^2 + V(h) + \frac{1}{n}\right). \tag{6.36}$$

We obtain the desired result by taking the infimum of h over $\mathcal{H}_n$.

Rates of convergence.

(vii) $m \geq 2$, $\beta > 1/(2m(m-1))$ *:* It follows from (6.36) and (6.24) with $h = [n^{-m/(2m\beta+1)}]$ that

$$\mathbb{E}(\|\hat{g}_{\hat{h}} - g\|_2^2) \leq C\left(n^{-2m^2\beta/(2m\beta+1)} + (\log(n))^m n^{-2m^2\beta/(2m\beta+1)} + \frac{1}{n}\right) \leq C\frac{1}{n}.$$

(viii) $\beta \in (0, 1/(2m(m-1))]$ *:* Putting (6.36) and (6.24) together with

$$h = \left[\left(\frac{(\log(n))^m}{n}\right)^{m/(2m\beta+1)}\right]$$

we obtain

$$\mathbb{E}(\|\hat{g}_{\hat{h}} - g\|_2^2) \leq C\left(\left(\frac{(\log(n))^m}{n}\right)^{2m^2\beta/(2m\beta+1)} + \frac{1}{n}\right) \leq C\left(\frac{(\log(n))^m}{n}\right)^{2m^2\beta/(2m\beta+1)}.$$

Theorem 6.3.1 is complete.

□

Production scientifique

Publications

Revues internationales avec comité de lecture

[1] F. Navarro, C. Chesneau, J. Fadili et T. Sassi. Block thresholding for wavelet-based estimation of function derivatives from a heteroscedastic multichannel convolution model. *Electronic Journal of Statistics*, Vol. 7, pp. 428-453, 2013.
http://hal.archives-ouvertes.fr/docs/00/80/04/96/PDF/dec-multi.pdf

[2] C. Chesneau, M. Kachour et F. Navarro. A note on the adaptive estimation of a quadratic functional from dependent observations. *İSTATİSTİK*, Vol. 6(1), pp. 10-26, 2013.
http://jtsa.ieu.edu.tr/index.php/istatistik/article/download/2/pdf

[3] C. Chesneau, F. Comte et F. Navarro. Fast nonparametric estimation for convolutions of densities. *The Canadian Journal of Statistics*, Vol. 41(4), pp. 617–636, 2013.
http://hal.archives-ouvertes.fr/docs/00/80/29/53/PDF/estimconvfinal.pdf

[4] C. Chesneau, M. Kachour et F. Navarro. On the estimation of density-weighted average derivative by wavelet methods under various dependence structures. Accepté pour publication dans la revue : *Sankhya*.
http://hal.archives-ouvertes.fr/docs/00/82/83/46/PDF/average3-dens-wav.pdf

[5] C. Chesneau et F. Navarro. On a plug-in wavelet estimator for convolutions of densities. Accepté pour publication à : *Journal of Statistical Theory and Practice*.
http://hal.archives-ouvertes.fr/docs/00/80/78/86/PDF/dec-sum-final_1_.pdf

[6] F. Navarro, C. Chesneau et J. Fadili. On adaptive wavelet estimation of a class of weighted densities. Accepté pour publications à : *Communications in Statistics - Simulation and Computation*.
http://hal.archives-ouvertes.fr/docs/00/79/55/19/PDF/weig-dens-est.pdf

Communication internationale avec actes et comité de lecture

[7] F. Navarro, J. Fadili et C. Chesneau. Adaptive parameter selection for block wavelet-thresholding deconvolution. *IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*.

Bibliographie

[1] K. Adamidis, T. Dimitrakopoulou, and S. Loukas. On an extension of the exponential-geometric distribution. *Statist. Probab. Lett.*, 73(3) :259–269, 2005.

[2] I. A. Ahmad and Y. Fan. Optimal bandwidth for kernel density estimators of functions of observations. *Statist. Probab. Lett.*, 51 :245–251, 2001.

[3] I. A. Ahmad and A. R. Mugdadi. Analysis of kernel density estimation of functions of random variables. *J. Nonparametric Statistics*, 15 :579–605, 2003.

[4] A. Al-Mbaideen and M. Benaissa. Frequency self deconvolution in the quantitative analysis of near infrared spectra. *Analytica Chimica Acta*, 705(1-2) :135–148, 2011.

[5] A. Antoniadis. Wavelets in statistics : a review (with discussion). *J. Italian Statistical Society*, 6(2) :97–144, 1997.

[6] M. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2) :353–360, 1984.

[7] C. Butucea and C. Matias. Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11 :309–340, 2005.

[8] T. Cai. Adaptive wavelet estimation : A block thresholding and oracle inequality approach. *The Annals of Statistics*, 27 :898–924, 1999.

[9] T. Cai. On adaptive wavelet estimation of a derivative and other related linear inverse problems. *Journal of Statistical Planning and Inference*, 108 :329–349, 2002.

[10] T. Cai. On block thresholding in wavelet regression : Adaptivity, blocksize and threshold level. *Statistica Sinica*, 12 :1241–1273, 2002.

[11] T. Cai and H. Zhou. A data-driven block thresholding approach to wavelet estimation. *Ann. Stat.*, 37 :569–595, 2009.

[12] R. J. Caroll and P. Hall. Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83 :1184–1186, 1988.

[13] L. Cavalier. Penalized blockwise stein's method, monotone oracles and sharp adaptive estimation. *Math. Methods of Stat.*, 10 :247–282, 2001.

[14] L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24, 2008.

[15] L. Cavalier, Y. Golubev, and A. B. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Probability Theory and Related Fields*, 48(3) :426–446, 2004.

[16] L. Cavalier and A. B. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123 :323–354, 2002.

[17] Y.P. Chaubey and H. Doosti. Wavelet based estimation of the derivatives of a density for m-dependent random variables. *Journal of the Iranian Statistical Society*, 4(2) :97–105, 2005.

[18] Y.P. Chaubey, H. Doosti, and B.L.S. Prakasa Rao. Wavelet based estimation of the derivatives of a density with associated variables. *International Journal of Pure and Applied Mathematics*, 27(1) :97–106, 2006.

[19] S.-M. Chen and Y.-S. Hsu. Kernel density estimations for maximum of two independent random variables. *J. Nonparametr. Statist.*, 16(6) :901–924, 2004.

[20] U. Cherubini, S. Mulinacci, and S. Romagnoli. On the distribution of the (un)bounded sum of random variables. *Insurance : Mathematics and Economics*, 48 :56–63, 2011.

[21] C. Chesneau. Wavelet estimation via block thresholding : A minimax study under the L^p risk. *Statistica Sinica*, 18(3) :1007–1024, 2008.

[22] C. Chesneau. Wavelet block thresholding for density estimation in the presence of bias. *Journal of the Korean Statistical Society*, 39(1) :43–53, 2010.

[23] C. Chesneau. Wavelet estimation of the derivatives of an unknown function from a convolution model. *Current Developments in Theory and Applications of Wavelets*, 4(2) :131–151, 2010.

[24] C. Chesneau, F. Comte, and F. Navarro. Fast nonparametric estimation for convolutions of densities. *The Canadian Journal of Statistics*, 41(4) :617–636, 2013.

[25] C. Chesneau, M. J. Fadili, and J.-L. Starck. Stein block thresholding for image denoising. *Applied and Computational Harmonic Analysis*, 28(1) :67–88, 2010.

[26] C. Chesneau, M.J. Fadili, and J.-L. Starck. Stein block thresholding for wavelet-based image deconvolution. *Elec. J. Statistics*, 4 :415–435, 2010.

[27] C. Chesneau and F. Navarro. On a plug-in wavelet estimator for convolutions of densities. *Journal of Statistical Theory and Practice*, in press.

[28] C. Cirelson, I. Ibragimov, and V. Sudakov. *Norm of Gaussian sample functions.* Springer Verlag, 1976.

[29] A. Cohen, I. Daubechies, B. Jawerth, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24(1) :54–81, 1993.

[30] F. Comte and V. Genon-Catalo. Convolution power kernels for density estimation. *Journal of Statistical Planning and Inference*, 142 :1698–1715, 2012.

[31] F. Comte and T. Rebafka. Adaptive density estimation in the pile-up model involving measurement errors. *Electron. J. Statist.*, 6 :2002–2037, 2012.

[32] F. Comte, Y. Rozenholc, and M.-L. Taupin. Penalized contrast estimator for adaptive density deconvolution. *The Canadian Journal of Statistics*, 34 :431–452, 2006.

[33] D. De Canditiis and M. Pensky. Simultaneous wavelet deconvolution in periodic setting. *Scandinavian Journal of Statistics*, 33 :293–306, 2006.

[34] A. Delaigle and I. Gijbels. Estimation of boundary and discontinuity points in deconvolution problems. *Statistica Sinica*, 16 :773 –788, 2006.

[35] R. DeVore, G. Kerkyacharian, D. Picard, and V. Temlyakov. On mathematical methods of learning. *Foundations of Computational Mathematics*, special issue for S. Smale 6(1) :3–58, 2006.

[36] R. DeVore and V. Popov. Interpolation of besov spaces. *Trans. Amer. Math. Soc.*, 305(1) :397–414, 1988.

[37] Ronald A. DeVore. Nonlinear approximation. *ACTA NUMERICA*, 7 :51–150, 1998.

[38] D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81 :425–455, 1994.

[39] D. Donoho and I. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90 :1200–1224, 1995.

[40] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2) :508–539, 1996.

[41] D.L. Donoho. Nonlinear solution of inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2 :101–126, 1995.

[42] P. Doukhan and J. Leaon. Déviation quadratique d'estimateur de densité par projection orthogonale. *Comptes Rendus Acad. Sciences Paris (A)*, 310 :424–430, 1990.

[43] J. Du and A. Schick. Root-n consistency and functional central limit theorems for estimators of derivatives of convolutions of densities. *Internat. J. Statist. Management Systems*, 2 :67–87, 2007.

[44] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3) :1257–1272, 1991.

[45] J. Fan and J.Y. Koo. Wavelet deconvolution. *IEEE transactions on information theory*, 48 :734–747, 2002.

[46] J. Fan and Y. Liu. A note on asymptotic normality for deconvolution kernel density estimators. *Sankhya*, 59 :138–141, 1997.

[47] E. Frees. Estimating densities of functions of observations. *J. Amer. Statist. Assoc.*, 89(426) :517–525, 1994.

[48] H.Y. Gao. Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics*, 7 :469–488, 1998.

[49] E. Giné and D. M. Mason. On local u-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, 35 :1105–1145, 2007.

[50] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation : oracle inequalities and adaptive minimax optimality. *Annals of Statistics*, 39 :1608–1632, 2011.

[51] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelet, Approximation and Statistical Applications*. Springer Verlag, New York 129, 1998.

[52] T. Hastie, R. Tibshirani, and J. H. Friedman. *Elements of Statistical Learning : Data Mining, Inference and Prediction*. Springer-Verlag, New York, second edition, 1986.

[53] I. Johnstone. Wavelet shrinkage for correlated data and inverse problems : Adaptivity results. *Statistica Sinica*, 9 :51–83, 1999.

[54] I. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo. Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society, Serie B*, 66(3) :547–573, 2004.

[55] M. C. Jones and H. W. Lotwick. Remark as r50. a remark on algorithm as 176. kernal density estimation using the fast fourier transform. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33 :120–122, 1984.

[56] A. Juditsky and S. Lambert Lacroix. On minimax density estimation on $\mathbb{R}$. *Bernoulli*, 10(2) :187–220, 2004.

[57] G. Kerkyacharian and D. Picard. Density estimation in besov space. *Statistics and Probability Letter*, 13 :15–24, 1992.

[58] G. Kerkyacharian and D. Picard. Density estimation by kernel and wavelets methods : optimality of besov spaces. *Statistics and Probability Letter*, 18 :327–336, 1993.

[59] G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2) :283–344, 2000.

[60] C. Lacour. Rates of convergence for nonparametric deconvolution. *C. R. Acad. Sci. Paris Ser. I Math.*, 342 :877–882, 2006.

[61] O. V. Lepski. On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and its Applications*, 35 :454–466, 1990.

[62] L. Li. On the block thresholding wavelet estimators with censored data. *Journal of Multivariate Analysis*, 99(8) :1518–1543, 2008.

[63] L. Li and Y. Xiao. On the minimax optimality of block thresholded wavelet estimators with longmemory data. *Journal of Statistical Planning and Inference*, 137(9) :2850–2869, 2008.

[64] F. Louzada, E. Bereta, and M. Franco. On the distribution of the minimum or maximum of a random number of *i.i.d.* lifetime random variables. *Applied Mathematics*, 3(4) :350–353, 2012.

[65] F. Luisier, T. Blu, and M. Unser. A new sure approach to image denoising : Inter-scale orthonormal wavelet thresholding. *IEEE Trans. Image Process.*, 16 :593–606, 2007.

[66] S. Mallat. *A Wavelet Tour of Signal Processing : The Sparse Way, 3rd edition.* Academic Press, 2008.

[67] J.S. Marron, S. Adak, I.M. Johnstone, M.H. Neumann, and P. Patil. Exact risk analysis of wavelet regression. *J. Comput. Graph. Statist.*, 7 :278–309, 1998.

[68] J.S. Marron and M.P. Wand. 'exact mean integrated squared error. *Ann. Statist.*, 20(2) :712–736, 1992.

[69] Y. Meyer. *Wavelets and Operators.* Cambridge University Press, Cambridge, 1992.

[70] A. R. Mugdadi and I. Ahmad. A bandwidth selection for kernel density estimation of functions of random variables. *Computational Statistics and Data Analysis*, 47 :49–62, 2004.

[71] G. Nason. Wavelet shrinkage using cross-validation. *Wavelet shrinkage using cross-validation*, 50 :463–479, 1996.

[72] F. Navarro, C. Chesneau, and J. Fadili. On adaptive wavelet estimation of a class of weighted densities. *Communications in Statistics - Simulation and Computation*, in press.

[73] F. Navarro, C. Chesneau, J. Fadili, and T. Sassi. Block thresholding for wavelet-based estimation of function derivatives from a heteroscedastic multichannel convolution model. *Electronic journal of statistics*, 7 :428–453, 2013.

[74] F. Navarro, J. Fadili, and C. Chesneau. Adaptive parameter selection for block wavelet-thresholding deconvolution. *IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*, in press.

[75] R. Nickl. Donsker-type theorems for nonparametric maximum likelihood estimators. *Probability Theory and Related Fields*, 138 :411–449, 2007.

[76] R. Nickl. On convergence and convolutions of random signed measures. *Journal of Theoretical Probability*, 22 :38–56, 2009.

[77] H. H. Panjer and G. E. Willmot. *Insurance Risk Models.* Society of Actuaries, Schaumburg, 1992.

[78] M. Pensky and T. Sapatinas. Functional deconvolution in a periodic setting : Uniform case. *The Annals of Statistics*, 37(1) :73–104, 2009.

[79] M. Pensky and T. Sapatinas. On convergence rates equivalency and sampling strategies in a functional deconvolution model. *The Annals of Statistics*, 38(3) :1793–1844, 2010.

[80] M. Pensky and T. Sapatinas. Multichannel boxcar deconvolution with growing number of channels. *Electronic Journal of Statistics*, 5 :53–82, 2011.

[81] M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27 :2033–2053, 1999.

[82] J.C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A sure approach for digital signal/image deconvolution problems. *IEEE Trans. Signal Process.*, 57 :4616–4632, 2009.

[83] G. Peyré, J. Fadili, and C. Chesneau. Adaptive structured block sparsity via dyadic partitioning. pages 303–307, 2011.

[84] D. Picard and K. Tribouley. Adaptive confidence interval for pointwise curve estimation. *Annals of Statistics*, 28(1) :298–335, 2000.

[85] B.L.S. Prakasa Rao. Nonparametric estimation of the derivatives of a density by the method of wavelets. *Bull. Inform. Cyb.*, 28 :91–100, 1996.

[86] K. Raghunandanan and S.A. Patil. On order statistics for random sample size. *Stat. Neerl.*, 26(4) :121–126, 1972.

[87] B. L. S. Prakasa Rao. Moment inequalities for supremum of empirical processes of u-statistic structure and application to density estimation. *J.Iran. Statist. Soc.*, 3 :59–68, 2004.

[88] H.P. Rosenthal. On the subspaces of $\mathbb{L}^p$ $(p \geq 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8(3) :273–303, 1970.

[89] M. Rudemo. Empirical choice of histograms and kernel density estimates. *Scand. J. Statist.*, 9(2) :65–78, 1982.

[90] A. Saavedra and R. Cao. On the estimation of the marginal density of a moving average process. *Canad. J. Statist.*, 28 :799–815, 2000.

[91] A. Schick and W. Wefelmeyer. Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.*, 16 :925–935, 2004.

[92] A. Schick and W. Wefelmeyer. Root n consistent density estimators of convolutions in weighted $\mathbb{L}_1$-norms. *J. Statist. Plann. Inference*, 137 :1765–1774, 2007.

[93] M. Shaked. On the distribution of the minimum and of the maximum of a random number of *i.i.d.* random variables. *Statistical Distributions in Scientific Work*, pages 363–380, 1997.

[94] M. Shaked and T. Wong. Stochastic comparisons of random minima and maxima. *J. Appl. Prob.*, 34(2) :420–425, 1997.

[95] B. W. Silverman. Algorithm as 176. kernel density estimation using the fast fourier transform. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 31 :93–99, 1982.

[96] B. W. Silverman. *Density estimation : for statistics and data analysis.* Chapman & Hall, 1986.

[97] A. B. Tsybakov. *Introduction à l'estimation non paramétrique.* Springer Verlag, Berlin, 2004.

[98] S. Vaiter, C. Deledalle, G. Peyré, J. Fadili, and C. Dossal. Local behavior of sparse analysis regularization : Applications to risk estimation. *to appear in Appl. Comput. Harmon. Anal.*, 2012.

[99] B. Vidakovic. *Statistical Modeling by Wavelets.* John Wiley & Sons, Inc., New York, 1999.

[100] B. Vidakovic and M. Vannucci. Preventing the dirac disaster : Wavelet based density estimation. *J. Italian Statistical Society*, 6(2) :145–159, 1997.

[101] G. Walter. *Wavelets and other orthogonal systems in applications.* CRC Press, Boca Raton, 1994.

[102] A. Zayed. Characterization of analytic functions in term of their wavelet coefficients. *Complex Variables*, 29 :265–276, 1996.

[103] C.-H. Zhang. Estimation of sums of random variables : examples and information bounds. *Annals of Statistics*, 33(5) :2022–2041, 2007.

[104] Z. T. Zijaeva. The estimation of the density of a sum of random variables. *Izv. Akad. Nauk. USSR Ser. Fiz.-Mat. Nauk*, 95(3) :13–15, 1975.

Table des figures

Liste des tableaux

Estimateurs adaptatifs avec parcimonie structurée

Résumé de la thèse : Cette thèse présente de nouvelles procédures statistiques dans un cadre non-paramétrique et étudie à la fois leurs propriétés théoriques et empiriques. Nos travaux portent sur deux sujets différents qui ont pour point commun l'observation indirecte du paramètre fonctionnel inconnu. La première partie est consacrée à une procédure d'estimation de type seuillage par blocs pour le modèle de bruit blanc gaussien. Nous nous intéressons à l'estimation adaptative d'un signal f et des ses dérivées à partir de n versions floutées et bruitées de ce signal et prouvons que notre estimateur atteint une vitesse de convergence (quasi-)optimale sur une large classe de boules de Besov. Puis, nous proposons une procédure adaptative d'estimation permettant de sélectionner les paramètres de l'estimateur de manière optimale en minimisant un estimateur sans biais du risque. Dans la deuxième partie, nous nous plaçons dans le cadre du modèle de densité au sein duquel la fonction inconnue subit une transformation donnée avant d'être observée. Nous construisons et étudions un estimateur adaptatif basé sur une approche plug-in et un seuillage dur en ondelettes par blocs. Enfin, nous étudions le problème d'estimer la convolution d'ordre m d'une densité à partir d'observations i.i.d. tirées dans la loi sous-jacente. Nous proposons un estimateur adaptatif fondé sur des noyaux, de l'analyse de Fourier et la méthode de Lepski. Nous étudions son risque quadratique et des vitesses nouvelles et rapides sont obtenues, pour une vaste classe de fonctions inconnues. Chaque méthode d'estimation étudiée est analysée numériquement par simulations, à la fois sur données simulées et sur données réelles.

Adaptive estimators with structured sparsity

Thesis summary : This thesis presents new statistical procedures in a non-parametric framework and studies both their theoretical and empirical properties. Our works are devoted to two different topics which have in common the indirect observation of the unknown functional parameter. The first part is dedicated to a block thresholding estimation procedure for the Gaussian white noise model. We focus on the adaptive estimation of a signal f and its derivatives from n blurred and noisy versions of the signal and prove that our estimator achieve the optimal minimax rate over a wide class of balls Besov. Then, we propose an adaptive estimation procedure for selecting the parameters of the estimator in an optimal way by minimizing an unbiased estimator of risk. In the second part, we consider the context of the density model in which the unknown function undergoes a given transformation before being observed. We construct and study an adaptive estimator based on a on a plug-in approach and the wavelets methodology. Finally, we focus on the problem of estimating the convolution of densities. We propose an adaptive estimator based on kernel methods, Fourier analysis and the Lepski method. We study the L2-risk properties of the estimator. Fast and new rates of convergence are determined for a wide class of unknown functions. Each estimation method is numerically analyzed by simulation, both simulated and real data.

Discipline : Mathématiques et leurs interactions.

Indexation RAMEAU : Statistique non-paramétrique ; Estimation, Théorie de l' ; Ondelettes ; Convolutions (mathématiques) ; Convergence (mathématiques) ; Problèmes inverses.

Indexation libre : Estimation non-paramétrique ; Estimation adaptative ; Représentations parcimonieuses ; Ondelettes ; Seuillage par blocs ; Théorie minimax ; Vitesse de convergence ; Problèmes inverses.

Laboratoire de Mathématiques Nicolas Oresme,CNRS UMR 6139, Université de Caen Basse-Normandie, BP 5186, 14032 CAEN Cedex, FRANCE